universität wien

MASTERARBEIT

Titel der Masterarbeit

„Screening for Novel Targets of the MAPK Hog1 using Quantitative Proteomics“

verfasst von

Natalie Romanov, BSc.

angestrebter akademischer Grad

Master of Science (MSc)

Wien, 2014

Studienkennzahl lt. Studienblatt:	A 066 834
Studienrichtung lt. Studienblatt:	Molekulare Biologie
Betreut von:	Ao.Univ.Prof.Gustav Ammerer

Table of Contents

I. Introduction

I.1. Biological Background: The Search for Hog1 Targets

I.1.1. Yeast osmoregulation

Maintaining cellular water balance proves to be a matter of life and death for cells. For instance, when thriving on nutrient-rich fruits, yeasts might eventually experience hyperosmotic stress as the extracellular sugar concentration reaches its saturation point. This causes efflux of water from the cell and hence cell shrinkage. On the other hand, the same fruit where the yeast cell has so far prospered might be subject to rainfall which will lead to water influx and cell swelling (hypo-osmotic stress). In either case control of the water balance is crucial for the cell to uphold its internal biochemical processes, regular transmembrane transport and turgor pressure (defined as the difference in hydrostatic pressure between the inside and outside of the cell [3]). *Saccharomyces cerevisiae* turned out to be particularly well suited for the studies of the underlying signaling pathways of osmoregulation, as it is a genetically tractable system and can be conveniently exposed to osmotic stress in the laboratory.

Still, the capability to cope with osmotic challenges is a necessary cell feature in general and potentially contributes to a number of human diseases as well, especially inflammatory disorders [4]. The tissue lining of the human gut, for example, routinely experiences hyperosmotic stress; affected cells have developed many adaptive strategies to avoid cell shrinkage, oxidative stress, DNA damage, protein carbonylation, and all other deadly features that come along with increased osmolarity. Thus, osmoadaptation is a common strategy to survive and function under potentially cytotoxic conditions. Yet another example illustrating the huge impact of hyperosmotic stress on cells is provided by the food industry and salt-based food preservation: Many molds and yeasts can only be inhibited by 50-60% sugar concentration in the environment [5], which is analogous to more than 15% salt concentration. Such concentrations of osmolytes clearly have an impact on food quality in general, and are suggested to change the scale of antibiotic resistance of food-related

pathogens, such as *Escheria coli*, *Salmonella typhimurium* and *Staphylococcus aureus* [15]. Thus, osmoregulation is an up-to-date and important field of study.

Although fragmented, the present picture of the yeast *S.cerevisiae* osmoregulation is extensive [6,7,8,9,68]. Basically it suggests four main steps necessary for a cell to re-establish homeostasis after hyperosmotic insult [10].

(1) The cell accumulates compatible osmolytes [9] by closing specific channels and thereby retaining osmotic strength. Interestingly, the act of closing is mainly controlled by PTMs of specific channel regulators [11].
(2) Osmolyte production is primarily stimulated by transient post-translational modifications (PTM) of the enzymes involved in their biosynthesis [9, 12,13].
(3) Changes in the transcriptional pattern with regard to genes involved in osmolyte synthesis and many other important physiological adaptations provide a sustained cellular response to conquer hyperosmotic conditions in the long run [9].
(4) To allow for a robust and sensitive response, cells have installed transient negative feedback loops to hinder repetitive stimulation of signaling pathways by prolonged hyperosmosis: Additionally, this feedback also functions in adequate termination of signaling [10,14].

Until recently, studies have focused mostly on the analysis of the transcriptional response (3), which- as current modeling analyses suggest [14, 73]- does not suffice for the immediate cell survival upon hyperosmotic stress. Rather, this immediate response is likely to be regulated by PTMs of key enzymes that are not involved in transcriptional regulation. The rapid alteration of the activity of osmolyte-producing enzymes is one example of this short-term adaptation in yeast [14, 73]. The main regulator governing these crucial modification events during the hyperosmotic-stress-response in yeast *S.cerevisiae* is the Hog1 MAP kinase.

I.1.2. MAPK signaling pathways

In eukaryotes, the response to cellular stress in general is regulated by tightly ordered and controlled cascades called mitogen-activated protein kinase (MAPK) pathways. Among signaling cascades, this type of pathway has received considerable attention as they are evolutionary conserved from yeast to mammals; indeed, a substantial number of reviews

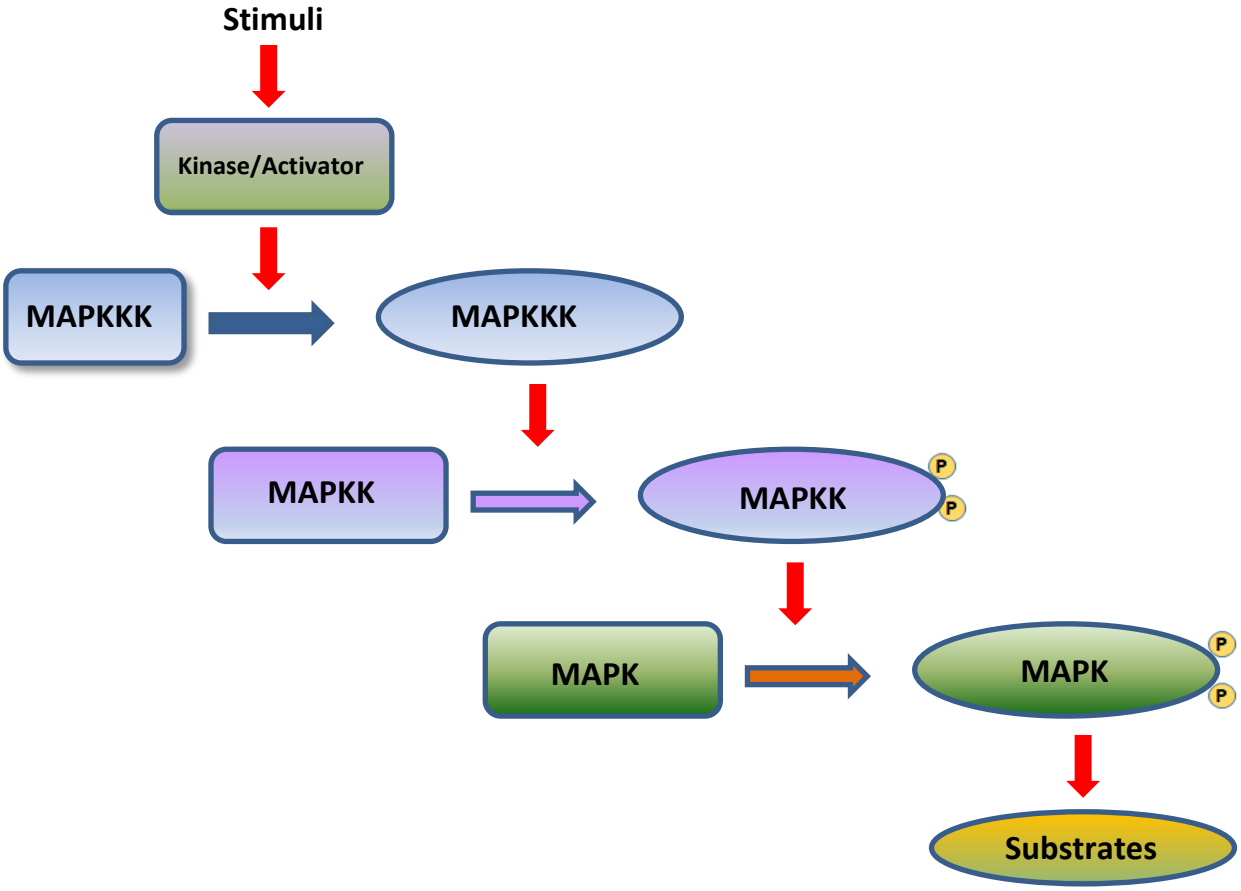


***Figure 1**. Scheme showing the MAPK module: The MAPKKK gets activated (not necessarily by phosphorylation) and activate the MAPKK, which in turn phosphorylates the MAPK at two characteristic residues.*

can be consulted for further reading [7,38,39,40]. A typical MAPK cascade is composed of three major components that act in series (see *Figure 1*). When activated a so-called MAPKK kinase (MAPKKK) phosphorylates a MAPK kinase (MAPKK) on conserved serine and threonine residues, which, in turn, stimulates a MAP kinase (MAPK) by dual phosphorylation of the conserved Thr and Tyr residues in the TXY motif within the activation loop. Once a MAPK is activated it can phosphorylate its targets on serines/threonines next to a proline (ST/TP-motifs). The advantage of this typical cascade-architecture is the fact that it can be regulated at multiple steps and allows for signal amplification. MAPK pathways are highly intertwined between each other with shared components, cross-pathway inhibition mechanisms and kinetic insulation [39,40]. Individual pathway constructions then allow regulation of response strength in a sensitive manner that is dependent on the different, and even conflicting signals the cell may receive [41].

In yeast there are five MAPK modules [9,42] that have been classified: The Cell Wall Integrity Pathway (CWI) controls cellular response upon cell wall deterioration, temperature shifts, etc. [26, 43,44]. The filamentous growth pathway [45,46] and the sporulation pathway [42] play an important role in starved haploid and diploid cells. The pheromone responding pathway [45,47] controls the morphological changes necessary for mating of yeast cells, and finally, the hyperosmotic glycerol (HOG) pathway- a homologue to the mammalian p38 MAPK pathway- is activated upon hyperosmotic stress. Among all these MAPK pathways, the HOG pathway proves to be particularly well suited to study signaling mechanisms and dynamics, since it can be reliably activated by increased medium osmolarity.

In this report I focus particularly on the HOG pathway and the identification of possible downstream targets of the corresponding MAPK Hog1. So far, the literature lacks a comprehensive large-scale description of direct Hog1 targets, especially those involved in cytoplasmic events. In this report we provide a workflow to elucidate and validate these previously unrecognized targets of Hog1.

I.1.3. The HOG MAPK Signaling Pathway

When challenged with hyperosmotic shock, cells experience a rapid loss in volume due to water efflux whereupon two distinct but redundant branches of the HOG pathway are activated [6,7,8,9]. Generally spoken, each branch has its own MAPKKK that activates a common MAPKK and MAPK that represent the integration nodes for the incoming signals. *Figure 2* illustrates the activation of the individual MAPKKKs, for each separate branch: One branch closely resembles two-component systems in bacteria and plants in that it uses a histidine-aspartate phosphorelay module involving the protein kinases Sln1, Ypd1 and the cytoplasmic response regulator Ssk1 [49] (see *Figure 2*). Under normal osmotic conditions Sln1 keeps itself autophosphorylated at its histidine kinase domain and phosphorylates its downstream relay partner Ypd1, which, in turn, phosphorylates Ssk1. Default phosphorylation of Ssk1 in turn leads to the inhibition of the redundant MAPKKKs Ssk2 and Ssk20 [38,48,49], and thus represents a negative regulation system. In response to hyperosmotic stress, however, Sln1 is inactivated through an unknown mechanism [6] and cannot phosphorylate Ypd1 anymore. Hence, Ssk1 also loses its phosphate moiety and the MAPKKKs Ssk2 and Ssk20 are no longer inhibited. They then autophosphorylate and activate

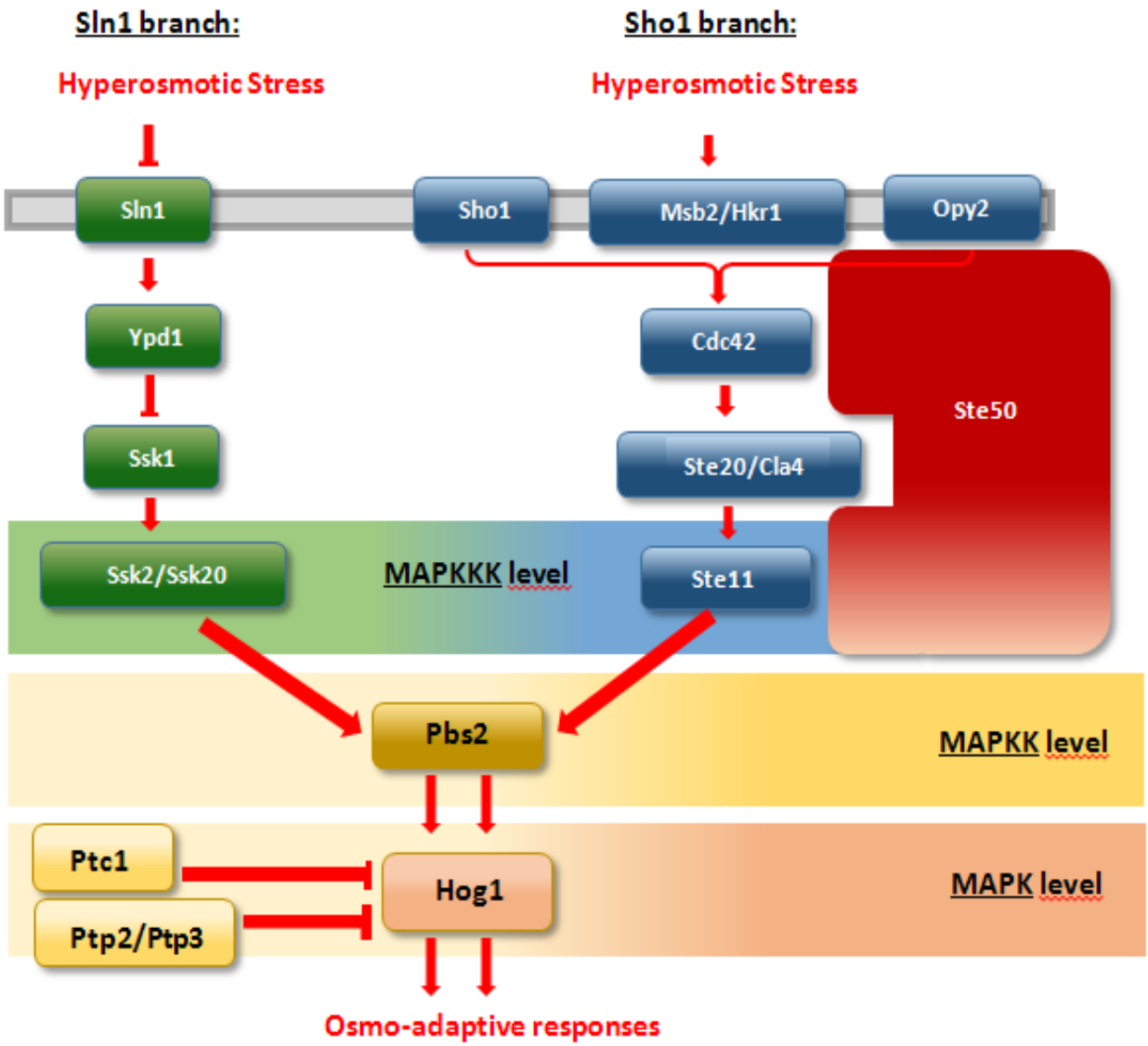


***Figure 2**. (A) A schematic diagram of the HOG MAPK pathway. The two separate branches are colored differently. Ste50 can also be part of other pathways and is not specific for the HOG pathway and is therefore colored differently. The differently colored bars symbolize the different MAPK modules/levels the pathway can be split up into.*

the subsequent MAPKK. Since this kind of branch architecture allows rapid sensing of normal osmolarity conditions, it has also been proposed as a fast-acting negative feedback mechanism for osmosensing [50]. The second branch is comprised of the MAPKKK Ste11 and its scaffolding element Ste50 that are also shared with other MAPK pathways. The most current hypothesis suggests that signal transduction is induced by two putative osmosensors Msb2 and Hkr1 [51] that interact with a co-osmosensor Sho1 through an unknown mechanism. The concerted action of Sho1 and these osmosensors leads to the activation of the kinase Ste20 by association with the membrane-bound small G-protein Cdc42 [52] (see

Figure 2). It is assumed that the whole protein machinery assembles at the membrane and the MAPKKK Ste11 is recruited there. This membrane localization is due to the Ste50 scaffold protein, which forms a stable complex with Ste11, as well as Cdc42, Sho1 and Opy2, a small membrane anchor protein. Upon assembly, the MAPKKK Ste11 is phosphorylated and activates the subsequent MAPKK. Interestingly, the binding sites and affinity of the corresponding proteins to Ste50 may vary under different stress conditions, allowing a fine-tuned response already at the upstream part of the MAPK pathway [53]. Although binding and phosphorylation of Ste11 is important, it is notable that overexpression of constitutively active Ste11 is sufficient to activate the Hog1 MAPK pathway [54].

The node of integration of the described Sln1 and Sho1 branches is the MAPKK Pbs2 (see *Figure 2*), which becomes anchored to the membrane by Sho1 and phosphorylated by the respective MAPKKKs, Ssk2/20 and Ste11, on the conserved residues *Ser514* and *Thr518* [55]. Mutating one of these residues to alanine indeed renders the cell osmosensitive [56]. Once phosphorylated, Pbs2 activates Hog1 by dual phosphorylation (Hog1pp) on the conserved *Thr174* and *Tyr176* residues [55]. When the amount of phosphorylated Hog1 surpasses a certain threshold, a significant portion of it is translocated to the nucleus. Tracing Hog1pp-accumulation via a fused GFP-reporter reveals that as osmo-adaptation proceeds, Hog1 is exported back to the cytoplasm and its overall activity diminishes [6]**.** More specifically, Hog1pp is dephosphorylated by nuclear phosphatases (PTCs and PTPs) after restoration of osmotic balance [105], which occurs after 20 to 30 minutes depending on the severity of the shock. Thus, cell swelling due to glycerol accumulation triggers a signaling termination of the Hog1 pathway by dephosphorylation and nuclear export events [57]. Overall these observations raise the question as to how osmolyte balance is achieved.

I.4. Osmolyte Accumulation: Glycerol and Trehalose

It is generally agreed that short-term adaptation is mainly aimed at the accumulation of glycerol and other osmolytes to lower the osmotic potential [3], thereby promoting water influx. Although upregulation of inorganic ions would serve the same purpose, an increased exposure of proteins and nucleic acids to ions may lead to destabilization and breakage, respectively [20]. Organic osmolytes, on the other hand, are small molecules that do not have any damaging impact on cellular macromolecules and are therefore also termed

"compatible solutes". Most renowned osmolytes are polyols (glycerol, sorbitol, inositols, etc.), small carbohydrates (trehalose, e.g.), amino acids and their derivatives (e.g. taurine), and methylamines (e.g. glycine betaines). They are highly soluble, neutral at physiological pH (either zwitterionic or achieving neutrality by corresponding ions) and prove effective in stabilizing proteins [21]. There have been various hypotheses put forward to explain their action at a biophysical level [22]: The so-called water replacement theory, for example, suggests that proteins are usually stabilized by their surrounding hydration shell which consists of water molecules forming hydrogen bonds around the molecule. When water is removed, however, this hydration layer is destabilized and may lead to denaturation of the protein [23]; hence the need for an adequate water molecule replacement. Another hypothesis relies on observations made with trehalose and its capability to hold biomolecules in a vitreous state. It is the only sugar that forms amorphous, non-hygroscopic crystals that remain stable during high temperatures; when rehydrated, however, molecules captured in this transient glass-like formation return to their original conformation [24,25]. In addition to these mechanisms, there are other organic osmolytes that have special properties and are utilized for antioxidation purposes [21].

Upon hyperosmotic shock *S.cerevisiae* mainly produces the three-carbon polyol glycerol and the disaccharide trehalose; however, glycerol has been described as the most compatible osmolyte for the growth of yeast in osmostress [9]. In concert with the water replacement theory, glycerol is believed to stabilize the native conformations of proteins as it is usually excluded from the hydration sphere [26]. The enzymes directly responsible for glycerol synthesis, such as the glycerol-3-phosphate dehydrogenase (Gpd1) and the glycerol-3-phosphatases Gpp1 and Gpp2, have been found to be transcriptionally upregulated under osmotic stress conditions [12,13,60]. Furthermore, rapid accumulation of glycerol is also achieved by appropriate regulation of specific channels, such as the proton symporter Sugar Transporter-Like protein Stl1 and the aquaporin Fps1 that mediate the transport of glycerol across the cytoplasmic membrane [27].

Yeast stress tolerance has also been observed to be correlated with trehalose accumulation. The disaccharide is produced from covalent linking of two molecules of glucose-6-phosphate (G6P) to trehalose-6-phosphate by the respective synthase Tps1, and subsequent dephosphorylation by Tps2. When dispensable, trehalose is degraded in the cytoplasm by

neutral trehalases Nth1 and 2. The rapid increase of trehalose levels upon heat shock has mainly been attributed to the increase in levels of both Tps1 and Tps2 [28,29], along with the substrates required for its synthesis, suggesting a distinct role of trehalose under stressful cellular conditions. Furthermore, some evidence in the literature indicates that Tps1 enzymatic activity increases during heat shock as well [30]. Notably, the activity of Nth1 is greater during recovery from stress [31], suggesting a distinct role of trehalose under stressful cellular conditions. This leveling out of trehalose is critical during recovery as very high levels of the osmolyte may lead to stabilization of non-native protein conformations thereby preventing appropriate refolding and functioning of proteins [32,33,34,35]. In addition, it has also been shown that trehalose serves to stabilize the membrane structure during processes involving dehydration (e.g. spore formation) [36]. When deprived of water, the lipid bilayer undergoes lipid phase transitions which render membranes non-elastic and easily breakable [37]. Although the function of trehalose is not fully understood, it is a widespread assumption that it is somehow involved in protecting and preserving the original membrane structure [37]. Thus, following up the role of trehalose in yeast osmoregulation might be quite intriguing, as hyperosmotic shock is ultimately coupled to dehydration processes.

How are the osmolyte levels regulated? For glycerol, there are three known mechanisms: 1) transcriptional upregulation of genes involved in biosynthesis, 2) Activation of enzymes involved in biosynthesis by post-translational modification [112] and 3) closing of channels. In all 3 of these types of processes Hog1 plays a role. I will next focus on the transcriptional response of the MAP kinase (1).

I.5. The Hog1-Dependent Transcriptional Response

Measuring the levels of gene expression using microarrays at different time points after an osmotic shock [58] revealed that similar to the phosphorylation profile of Hog1, the transcriptional response is transient with the period and the amplitude of the response strongly correlating to the stress level [6]. Usually, phosphorylation of Hog1 increases rapidly after hyperosmotic shock, reaches its maximum level after 5 minutes and then gradually decreases to basal levels due to several negative feedback mechanisms [59]. A strongly correlating pattern was found with the stress-induced transcriptional response: Several

studies identified 200-400 upregulated genes, comprising genes for glycerol production, ion homeostasis, redox metabolism and general stress response genes [58,60,61]. Also many genes are downregulated as they encode proteins necessary for ribosome biogenesis, the translational machinery and glycolysis [58] to boost cell's capacity for stress survival.

One well-characterized function of Hog1 in the transcriptional regulation during osmostress is its action through several unrelated transcription factors that control specific gene sets [62]. These transcriptional regulators include Hot1, Smp1, Msn1, Msn2/Msn4 and Sko1 [63,64,65]. Importantly, activated Hot1 [64] initiates glycerol production by expression of *GPD1*, *GPD2*, *GPP1*, and *GPP2*, encoding glycerol-3-phosphate dehydrogenase and glycerol-3-phosphatase, respectively. Though the Hog1 kinase seems to exert its major effects through phosphorylation, interesting observations suggested that Hog1 may also alter the gene expression pattern by mechanisms other than phosphorylation. With Hot1, for example, it has been shown that replacing Hog1-dependent phosphorylation sites with alanine does not change the transcriptional activity of Hot1. However, it has been demonstrated by Alepuz *et al.* (2003) [64] that the tight association of Hog1 with the large subunit of the RNA Pol II and subsequent tethering to the promoter vicinity where Hot1 constitutively resides, is sufficient to induce transcription of *GPD1*. With other transcription factors, such as Sko1, Hog1 yet again shows another distinct behavior: In the absence of stress Sko1 acts as a transcriptional repressor by recruiting co-repressor complex Ssn6-(Cyc8)-Tup1 to the respective promoters (e.g. *GRE2*). During elevated osmolarity, however, Hog1 phosphorylates Sko1 and thereby reduces its affinity to the co-repressor complex, literally turning it into a transcriptional activator [66]. Additionally Hog1 directly recruits the chromatin remodelling complexes and histone modifiers SAGA, Mediator and SWI/SNF to osmostress-inducible promoters [67].

Besides regulation of transcription factors, the binding of Hog1 to chromatin-regions other than promoters proves to play an essential role in transcriptional regulation as well. Recently it has been shown that Hog1 interacts with the RSC chromatin-remodeling complex, which modifies nucleosome structure [69]. Indeed, a compromised RSC complex elicits osmosensitivity [69], suggesting an underestimated role of Hog1 in transcriptional elongation. This is further confirmed by Proft *et al.* [70] who demonstrate that Hog1 is traveling along the coding region within the RNA Pol II elongation complex, which is

dependent on Hog1 recruitment to the promoter region in the first place. Again the adequate recruitment of Hog1 and the RNA Pol II to the promoter requires a structural chromatin remodeling which is mediated by the Rpd3 histone deacetylase (HDAC) complex. Hog1 presumably binds to Rpd3L, thereby ensuring that local chromatin structure is suited for proper harboring of the RNA Pol II [71].

I.6. Transcription-Independent Effects of Hog1

Although microarray analysis elicited a large number of genes regulated upon hyperosmotic shock, deletion mutants of the corresponding genes do not necessarily display an osmosensitive phenotype [72] under moderate stress conditions. Thus, before transcriptional induction, Hog1 mediates other short-term adaptation processes to guarantee immediate cell survival when sudden osmotic shifts occur [72]. Ultimately, these processes should be aimed at glycerol accumulation [60], re-establishment of ionic balance [77] and prevention of any energy-consuming cellular process such as cell proliferation and protein synthesis. The notion of such a short-term adaptation mediated by Hog1 phosphorylation of cytoplasmic targets has been fueled by the work of Westfall *et al.* [73] who show that a membrane-tethered version of the Hog1 kinase is sufficient for cell survival under moderate hyperosmotic conditions. Furthermore, they show that indeed no Hog1-dependent transcription occurs in the corresponding mutant. While the transcriptional response has been shown to be necessary for long-term adaptation, the early cytoplasmic response to hyperosmotic stress appears to be biologically more relevant at the stress onset [72]. These swift regulatory steps are primarily characterized by the Hog1-mediated phosphorylation of protein factors at S/T-P sites, the typical Hog1 kinase motif.

For example, the Hog1/mediated phosphorylation of Rck2 (a kinase related to mammalian MAPK-activated protein kinases) leads to inhibition of protein biosynthesis, perhaps via regulation of the elongation factor 2 (EF-2) [72,74]. Furthermore, Hog1 resumes protein translation after a transient pause in protein synthesis. Besides the biosynthetic machinery, cell-cycle progression is also affected by Hog1 activity leading to transiently interrupted G1/S and G2/M-transitions, as well as mitotic exit [75,78].

A particularly illustrative example of the short-term effect mediated by Hog1 is the decrease of glycerol efflux by the glycerol channel Fps1 closure [11]. The channel mediates the passive

diffusion of glycerol in both directions through the plasma membrane, thereby ensuring cellular homeostasis [11]. After hyperosmotic insult the activity of Fps1 diminishes to allow glycerol to accumulate inside the cell [90]. Rgc1 and Rgc2 (regulator of glycerol channel) are positive regulators of Fps1 and recent studies suggest that Hog1 may negatively regulate Fps1 by inhibiting these regulators through phosphorylation at S/T-P motifs [11]. Once the cell has accumulated glycerol, Fps1 activity increases again to release glycerol and therefore turgor pressure [90]; it is still not quite clear whether the negative feedback is provided by the glycerol accumulation *per se* or yet again by the phosphorylation/dephosphorylation events on the cytoplasmic targets Rgc1/2. This example only displays a small fraction of the system-wide implications of Hog1-mediated phosphorylations.

One of the most intriguing non-transcriptional responses within the Hog-pathway, however, encompasses the restoration of the ionic balance. A systematic approach revealed that the higher the salt concentration in the environment, the slower the nuclear localization of Hog1, thus the transcriptional response becomes delayed [76]. Then the question arises: How does the severity of the osmotic stress (measured by the amount of ions in the environment) influence the nuclear accumulation of Hog1? It appears that an increased ion concentration in the nucleus brings about the release of transcription factors from the chromatin and that upon reduction of these ions the proteins can re-associate with the chromatin [77]. Therefore, it seems like Hog1-dependent maintenance of ion balance in the cell should precede Hog1 nuclear localization. Indeed, Hog1 induces the phosphorylation of at least two proteins located at the plasma membrane, the Nha1 Na+/H+ antiporter and the Tok1 potassium channel [6,77]. As re-assocation of transcription factors to the chromatin is delayed in a *nha1Δ* mutant in the same way as in a *hog1Δ* mutant, the intracellular Na+ concentration seems to be primarily responsible for the perturbations in the nucleus upon hyperosmotic stress. Also the ATPase sodium pump Ena1 mediating Na+ and Li+ efflux has been shown to become upregulated upon hyperosmotic stress in a Hog1-dependent manner [77].

Although genetic and biochemical analyses has provided us with extraordinarily precise information on essential players in the Hog pathway, questions still arise regarding the dynamics of the pathway itself. The discovery of Hog1 modulation of various overlooked cellular processes has definitely taken the understanding of the osmostress response to a

new level, e.g. endocytosis [79] and glycerol channels [11]. The search for novel Hog1 substrates is far from complete and furthermore, it is clear that high-throughput methods are needed to screen the cellular proteome for potential targets that get phosphorylated by Hog1.

To this end, we make use of mass spectrometry for phosphopeptide identification and assessment of the regulatory event in dependence of Hog1 activity.

I.2. Methodic Issues: Mass Spectrometry

I.2.1. Basics of Mass Spectrometry

Recent developments in quantitative mass spectrometry proved quite effective for high-throughput screening (analogous to genetic screening termed “shotgun”) of the proteome under various conditions.

In a classical identification approach, proteins are digested with a protease that cuts at specific sites to produce smaller peptides. These peptides are separated on a reversed phase-HPLC (high-pressure liquid chromatography) according to their hydrophobicity. The LC (liquid chromatography) is coupled online to a mass spectrometer capable of precursor fragmentation. The information gathered in these MS/MS (MS2) scans is interpreted by search algorithms and translated into peptide sequences which are then assigned to proteins.

At the interface of LC and mass spectrometer the peptides are ionized and transferred into a gaseous state. In the so-called ESI-source (electrospray ionization) the peptides eluting from

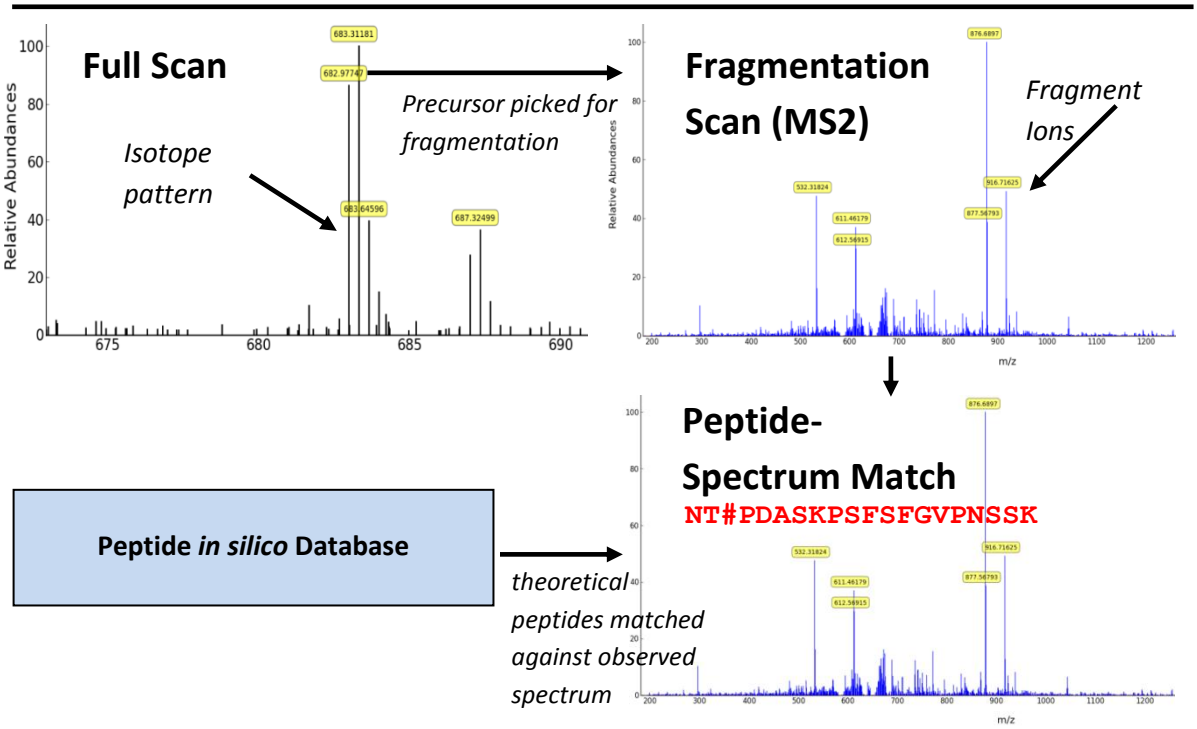


Figure 3. *Representation of MS data in context of data interpretation: A full scan provides highly accurate mass data with charge-dependent isotope patterns for observed m/z-values. MS2 spectra provide sequence information of the most abundant precursor peptides, linking intensity values to the actual identification of the peptide. The fragment spectrum typically contains a series of b- and y-ions that can be matched to a theoretical peptide spectrum (PSM).*

the LC get positively charged, the solvent evaporates, leading to single, charged molecules required for MS analysis. In high-resolution mode spectra are obtained as a result of exact frequency measurements, reflecting specific mass/charge-ratios (*m/z*-values) and their intensities (MS1). In order to obtain the amino acid sequence of the precursor ions, peptides with specific *m/z*-value are isolated and broken into fragments by colliding them with an inert gas (an MS2-spectrum based on collision-induced dissociation (CID) or higher-energy collisional dissociation (HCD) is generated). Typically, the peptide bonds are most sensitive to this kind of fragmentation, resulting in a distinct series of so-called *b*- and *y*- fragment ions [92]. Depending on the amino acid sequence of the peptides, these primary fragment ions are accompanied by "satellite" ions that derive from the neutral losses of ammoniac or water of the amino acids' side chains.

In MS1-based quantification, high-resolution power of mass-spectrometers provides the information on peptide signal intensities used for quantification. The MS2-spectrum is needed for the identification process, as depicted in *Figure 3*. Theoretical peptides generated from *in silico* digestion are matched against observed fragmentation spectra (MS2). If most of the experimental ion pattern fits the theoretical pattern, a so-called peptide-spectral match (PSM) occurs. How is the quality of such a match evaluated?

I.2.2. Statistical Validation of MS-Dataset and Estimating the False Discovery Rate

Scoring Algorithms. All search engines, such as SEQUEST© and Mascot©, provide a score of how likely a particular spectrum is to match to a particular peptide in the database provided. They do so by scoring algorithms that reflect the strength of PSMs or how likely such matches would have arisen by coincidence (see *Figure 3*). SEQUEST©, for example, calculates the cross-correlation score (XCorr) for all candidate peptides suggested from the database. For this purpose, the intensity of the peaks in the experimental spectrum are normalized and low-intensity peaks are removed, resulting in a processed spectrum A. The theoretical spectrum B is created by *in silico* fragmentation of candidate peptides. The XCorr correlation function basically counts the number of fragment ions that spectrum A shares with spectrum B, allowing for small differences in *m/z*-values; up to a certain length the XCorr shows length-dependency [94]. As precursor charge features affect the stability of the ion and hence the fragmentation efficiency of the peptide in the mass spectrometer, the

XCorr also depends on the charge state of the peptide. We demonstrated these two issues with parts of our shotgun data in the course of our research work (data not shown).

False Discovery Rate. Nonetheless there are a number of reasons for wrong peptide-spectrum assignment: 1) the spectrum might be of low quality, 2) its corresponding peptide is not included in the used database, 3) its peptide might be modified so that the peak *m/z*-values are shifted such as to match another theoretical spectrum and 4) imperfect scoring function. While MS software provide powerful algorithms, the overlap between the scores for correct and incorrect PSMs can be quite substantial [95]. This can either lead to elimination of true positive identifications (when minimizing the number of false positives) or a large number of false positives (by maximizing the yield of true positives). Thus, the software's peptide identification results should be statistically validated by estimating a so-called false discovery rate (FDR). Usually this is done by sorting PSMs by their scores and then choosing a score threshold where high-quality data can be guaranteed within a certain probability. The FDR is then defined as the ratio between the false PSMs and the total number of PSMs above the score threshold (see *Figure 4*).

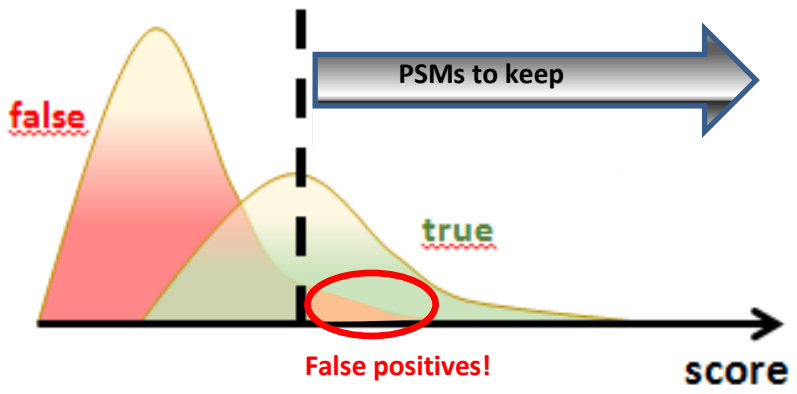


Figure 4. *Integrating the areas above the given score threshold (black line) and calculating the corresponding ratio (false/true) gives the FDR. The encircled area gives the number of false positives that surpasses the given score threshold.*

Estimating the FDR with the Target-Decoy Method. One possibility to distinguish between correct and incorrect peptide identifications is provided by the so-called target-decoy strategy. Basically, one creates a fake set of decoy peptide sequences by reversing the existing protein sequences in the search database and appending it to the set of the original target sequences. The resulting composite database would be twice the size and MS/MS-spectra can be searched against it. The core idea of the strategy is that, assuming the decoy sequences are clearly labeled, the number of spectra matched to decoys should definitely be

an incorrect match and thus, should give an estimate of the number of PSMs that are *probably* not correct in the target database [95,96].

However, for the successful application of this procedure, several requirements should be fulfilled [95]: First, the target and decoy database should be alike in some parameters, such as distribution of amino acids and average peptide length, especially considering that XCorr distributions are length-dependent [94]. At the same time, however, the databases should not contain overlapping sequences to guarantee that matches to decoy sequences are indeed incorrect. Second, target and decoy false positives are equally likely: Low XCorr scores of a spectrum increase the chance of an incorrect peptide assignment to a target or decoy peptide in the same way. In this work, we made use of the reverse-target-decoy approach to estimate the PSM-FDR at 1% for the entire dataset.

Q-Values Needed to Estimate the FDR over several experiments. Though the FDR is score-dependent, it does not show any linear correlation: For two significantly different scores the FDR can be the same. Moreover, a higher score threshold does not necessarily mean that the FDR would be smaller [97]. Therefore, an FDR threshold *per se* is not a reliable criterion for appropriate data handling [97].

A widely used method in approaching this problem is to assign each PSM event a statistical confidence measure (q-value) [98], which indicates how likely this event is to be true given a null hypothesis (random match). A set of PSMs is ranked according to individual XCorrs. Then for each example in the ranked list, a threshold is set such that all the examples above it are considered "accepted" and all those below are considered "rejected". The unique q-value associated with this particular example is then determined by the fractions of positives (spectrum has been matched to sequence of the target database) and negatives (spectrum has been matched to sequence of the decoy database) found above this threshold (see *Figure 5*).

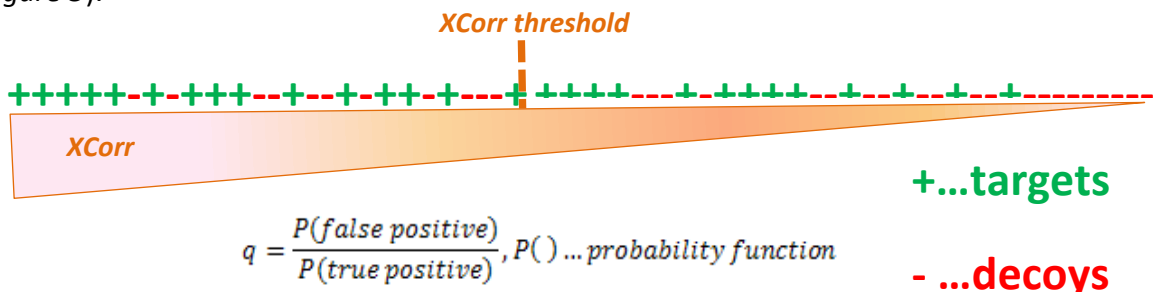


***Figure 5**. Definition of the q-value and illustration of the q-value concept.*

The q-value may therefore be referred to as the minimum FDR-threshold at which a given PSM is accepted and hence makes it a score-dependent parameter [97]. Matches are only considered statistically significant when they are below a certain user-specified q-value threshold and should be selected for further biological analysis. We set this value at 1%.

In large-scale biological screens, such as provided by our lab, the introduction of the q-value is advantageous since it is associated with a unique PSM and provides comparability between two differential experiments. We will therefore stick to the notion of a q-value threshold in this report.

I.2.3. Analysis of Phosphorylations Using Mass Spectrometry

Since our primary focus in this work has been phosphoproteomics, there are some important considerations that one has to take into account: If a given peptide is phosphorylated, phosphate moieties of a peptide become released in the fragmentation procedure and generate *m/z*-values distinct from an unphosphorylated counterpeptide. However, a broad range of challenges in phosphopeptide identification emerge in connection with the physical properties of phosphopeptides *per se* [80,81]. Phosphopeptides are difficult to analyze because they are (a) prone to biased fragmentation and (b) are hardly detectable in complex samples without enrichment techniques due to ion suppression [93]. There has been considerable effort in the past few years to circumvent these problems. Various phosphopeptide enrichment techniques such as IMAC, TiO_2, etc. were invented and the instruments were fine-tuned to be even more sensitive to modification-bearing peptides [82,100]. Moreover, in view of the above described scoring algorithms, proper score assignment has been acknowledged to be essential for phosphorylated peptides: At the same time, contaminants and protease precursors are given high-scoring results, peptides derived from low-abundance proteins and poor fragmenting characteristics are not identified at high confidence level in the SEQUEST algorithm [80,99]. Thus, to prioritize those underestimated identifications, manual evaluation of interesting peptides is sometimes necessary. In this report we also made use of this expertise (see *III.5*).

There are other difficulties that need to be addressed during phosphorylation analysis: First, a peptide might be phosphorylated at multiple residues. Even when using modern phosphorylation site allocation algorithms such as PhosphoRS [111], phosphorylation site assignments cannot always be resolved to a single residue. A series of possibly

phosphorylated amino acids greatly diminishes the efficiency of phosphorylation allocation if no informative ions are found. Second, one has to bear in mind that proteolytic digests of samples are often incomplete to a certain degree, resulting in peptides with intact cleavage sites (missed cleavage). As with additional variable modifications (methionine oxidation), missed cleavage sites increase the number of peptide variants. What is more important, however, is the fact that these missed cleavages might occur because of nearby phosphorylation sites. If a phosphorylation site is located near a cleavage site, the protease might not be able to attack the site due to the large interfering phosphate moiety of the neighboring amino acid.

All described factors as well as some ion properties (peptide charge) enlarge the number of possible peptide species for one phosphorylation site, which affects phosphorylation data analysis. Therefore, the concept of "phosphorylation site groups" (PSGs) was introduced in our research group: Basically, a phosphorylation site group is a group of peptides harboring one particular phosphorylation site or a specific set of phosphorylation sites (see *Figure 6*). Hence, it discards any information on missed cleavages, charges and variable modifications (usually methionine oxidation), compressing this information to the phosphorylation site only (that has been found with variable peptide species). If the latter phosphorylation site is found with another phosphorylation site in tandem, however, we would consider that species to be a separate double-phosphorylated PSG. At the level of SILAC-quantification values, the different members of one PSG should be ideally similar.

Additionally to the issues described above, phosphorylation is often a sub-stoichiometric process, meaning that not all given peptide copies are found in the phosphorylated state at a given time point [81,99]. The resulting ratio of phosphopeptides versus the unphosphorylated counter-peptides could be used as a quantitative measure of peptide regulation (personal communication): If the phosphopeptide is usually not as abundant as the unphosphorylated variant, but gets up-regulated in the course of stressed conditions while the unphosphorylated variant is downregulated, it would be indicative of a regulatory event and would even disclose information on peptide stoichiometry.

Following up such stoichiometric changes in phosphorylation and quantifying them proves valuable for a better understanding of the dynamics of signaling pathways. Several methods exist to do this within an MS-framework, for example label-free quantifications, iTRAQ and SILAC. In the following chapter I will focus on SILAC quantification.

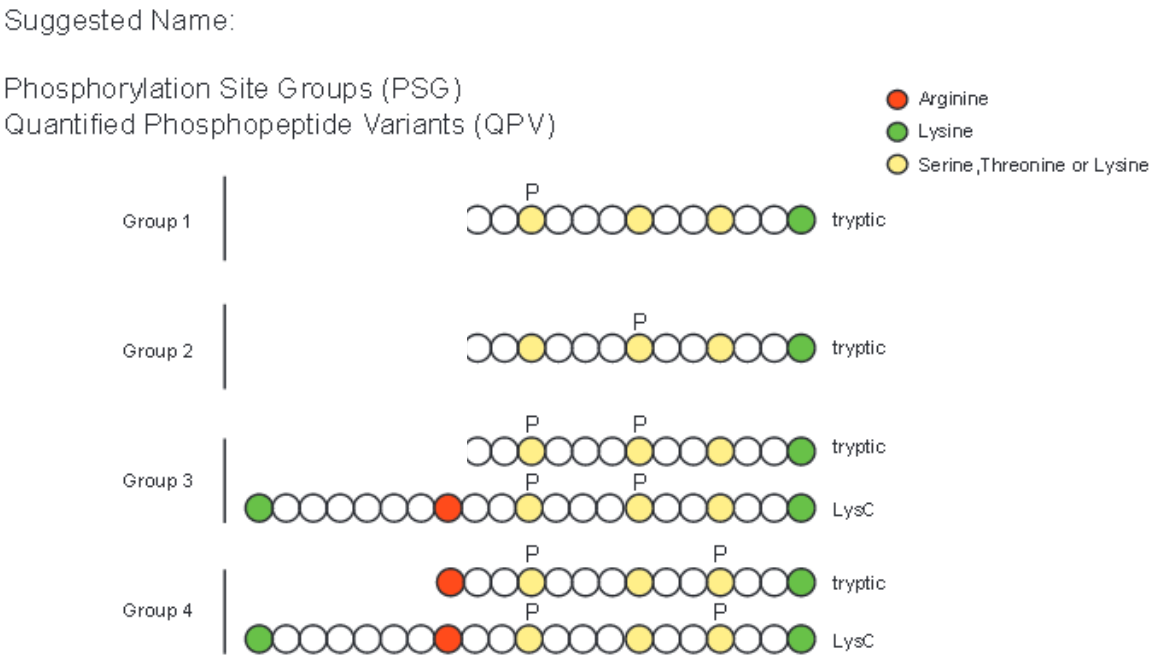


***Figure 6**. (© Wolfgang Reiter) Illustration of the phosphorylation site group (PSG) concept.*

I.2.4. Quantitative MS and Phosphoproteomics

One method to describe quantitatively whether a given phosphorylation event is regulated or not, is the SILAC technology (stable isotope labeling of amino acids in cell culture), as described by Ong *et al.* [84]. The method is based on the principle that peptides, which are labeled with different stable isotopes (^{12}C and ^{13}C, subsequently called "light" and "heavy"), are chemically identical but differ by a certain mass increment. Thus, in an experiment, two cell populations are grown in culture media that are identical except for the supplemented labeled amino acids (lysine and arginine, respectively) that are incorporated into all newly synthesized proteins. As this incorporation is efficient and reproducible and no chemical difference between the isotopes is observed, the cells should behave exactly like the population grown in the presence of the normal amino acid [101]. When exposing one culture to a specific condition and merging these two cultures during harvest, the corresponding Light/Heavy-ratio (subsequently called "L/H-ratio") of peptides of the resulting sample enables direct quantitation of relative protein expression levels and the degree of modifications produced by this condition. Of course, SILAC can be used in different experimental contexts, in shotgun as well as in target validation approaches.

I.2.5. Shotgun Proteomics

Analogous to shotgun-sequencing of DNA, shotgun proteomics or discovery-based proteomics starts with the proteins in the mixture being digested and the resulting peptides separated by liquid chromatography. Tandem mass spectrometry is then used to identify most of the present peptide species.

Identification of usually low abundant phosphopeptides requires highly sensitive methods for isolation, detection and quantification. An inherent problem of shotgun MS is that no reproducibility can be generated, mainly because not all peptide species in a sample can be recorded in MS. Redundant peptides are picked up by the mass spectrometer for further fragmentation. Yet again, this very fragmentation does not necessarily give rise to good MS spectra that are clearly assigned to peptides by search engines. Küster *et al.* [83] even makes the suggestion that only 10% of the peptides submitted to mass-spectrometry analysis give rise to adequate peptide identification. At the same time though, it is clear that "re-measurement" mode will be established sooner or later with data and techniques accumulating, which will eventually lead to increasingly comprehensive and also faster and more accessible whole-proteome quantification.

One particularly important methodological strategy to increase reproducibility by sample complexity reduction is the enrichment strategy via strong cation exchange chromatography (SCX) [82,100]. In this technique, binding to the SCX column is dependent on electrostatic interaction between the negatively charged resin and the positively charged peptides. When the sample is loaded under strong acidic conditions (pH~2.7), carboxylates are rendered neutral whereas phosphate groups retain their negative charge; hence, the overall charge of the peptide is reduced. A gradient of increasing salt concentration then allows phosphopeptides to elute earlier relative to non-phosphorylated peptides. Combined enrichment techniques yield a larger reproducibility and a large number of phosphorylation sites identified, both in a shotgun as well as targeted proteomics approach.

I.2.6. Targeted Validation Approach

Targeted proteomics with MS. Nowadays, large sample quantities and multi-dimensional fractionation strategies allow the generation of massive data sets in a single mass spectrometry shotgun analysis. However, one has to be aware of the inherent boundaries of

this discovery approach, namely, the failure to detect scarcely abundant peptides and the insufficient overlap of discovered peptides between two biological replicates [2]. Much of this irreproducibility is due to the operating mode of the mass spectrometer. In general, the instrument continuously repeats a cycle consisting of full-scan mass spectrum and subsequent fragmentation of a given number of most abundant peaks for identification. The semi-automated peak selection process is not reproducible *per se*. Thus, the data provided by shotgun analyses should be treated as a means to extract a targeted group of proteins that may show an interesting biological feature (modification) in regard to the conditions tested. It is difficult, if not impossible, however, to make any conclusions about its biological relevance at the level of a single protein without any further validation steps. Targeted studies encompass biochemical and genetic validations, as well as mass spectrometry instrumentation to directly analyze purified proteins [79]. Usually, this targeted proteomics method aims at reducing sample complexity for more sensitive detection and quantification of pre-selected components [2]. Thereby it not only confirms results from large-scale experiments, but also expands the number of identified peptides for a selected protein.

To follow up regulatory events at the level of post-translational modifications of a single protein, we applied a modified variant of the HB-tag for tandem affinity purifications under denaturing conditions [79,85,86]. Stringent purification conditions involving 8M urea and 6M guanidium guarantee the inhibition of enzymatic activity that would otherwise modify the *in vivo* post-translational modification (PTM) pattern. The modified tag (HTB_{eaq}) consists of three elements: two poly-histidine domains, followed by three TEV cleavage sites and a bacterially derived biotinylation signal peptide which recruits biotin molecules to the tag. This tag allows a pre-purification by Ni^{2+} chelate chromatography, followed by binding to streptavidin agarose beads. The strong interaction between biotin and streptavidin withstands the stringent wash conditions needed to remove any non-specific binding of unwanted proteins, thereby significantly reducing the sample complexity. This kind of analysis therefore allows purifying the protein of interest and acquiring a full picture of its potential phosphorylation sites and enhanced sequence coverage.

Validation of kinase-substrate interactions. An essential issue in phospho-proteomics has been raised by Bodenmiller *et al.* [88] who performed a label-free quantification study of phosphorylation patterns in differentially perturbed *S.cerevisiae* strains. These strains

contained either gene deletion mutants of some non-essential kinase or phosphatase, or essential kinases that can be inactivated by cell-permeable inhibitors ("analog-sensitive" kinase strains (*as*), [102]). Analyzing the phosphoproteome alterations in the yeast mutants revealed that the absence of a particular kinase or phosphatase usually affects far more proteins than its immediate downstream targets due to secondary effects.

To circumvent this issue we defined a strict set of criteria for a potential kinase target: First, if known, one should particularly look for phosphorylated kinase/phosphatase motifs. Second, if a pathway is inducible, particular phosphorylations are upregulated under certain conditions where the kinase in question has its highest activity. Third, this upregulation should not occur if the kinase is specifically inhibited or removed. Finally, mass spectrometry analysis does not provide any information on *direct* kinase-substrate interaction; hence, ideally an alternative technique, e.g. protein-protein interaction assay, should be applied for validation of a transiently occurring interaction between the kinase and an interesting group of targets that has shown up in the shotgun experiment.

In the work presented here, we made use of a recently developed protein-proximity assay called "M-track" [1] to confirm a direct kinase-substrate interaction. The assay is based on tagging, with the bait fused to a histone lysine methyltransferase (HKMT) and the prey fused to the enzyme's substrate (starting 21-aa-region of histone3 containing lysine 9) (see *Figure 7*). Once in close proximity, the prey gets stably methylated which can be detected by methylation-specific antibodies. Notably, a positive assay does not necessarily imply that the kinase in question actually *phosphorylates* the substrate it transiently interacts with; however, it reinforces the argument that an observed regulatory phosphorylation event observed in the MS shotgun approach, is due to a direct kinase-substrate interaction.

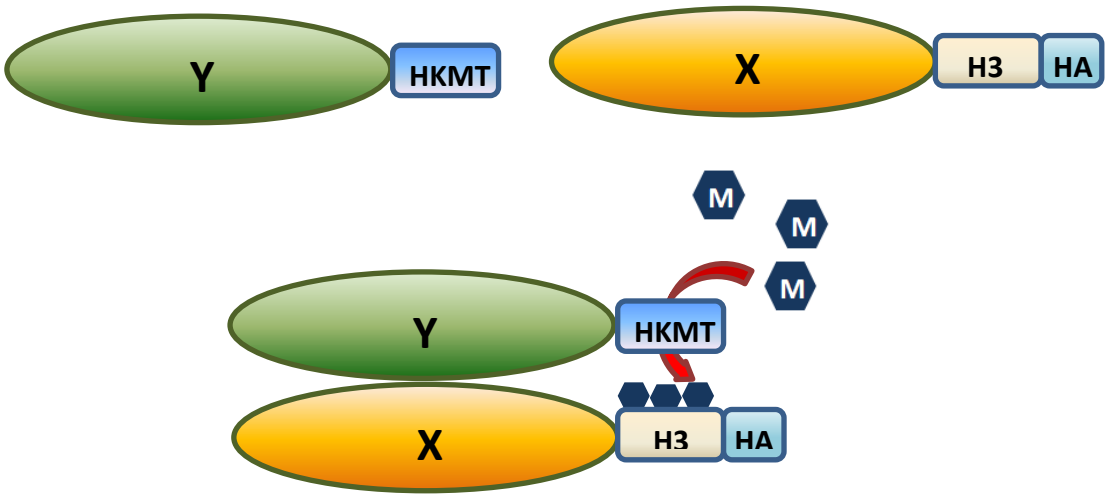


Figure 7. *Illustration of the M-track system with random interacting proteins X and Y that are tagged correspondingly. The dark blue hexagons present the methylation groups which are transferred by HKMT activity to H3 if X and Y interact with each other.*

II. Project Outline

II.1. The Search for Target of Hog1

A broader overview of the system-wide implications of hyperosmotic stress in yeast has so far only been provided by Soufi *et al.* [89] who identified a number of proteins that become phosphorylated upon hyperosmotic stress using a quantitative MS approach. Their SILAC experimental set-up captured all proteins that would change in abundance or in regulation of a particular phosphosite after the cell has been exposed to increased osmolarity for 5 and 20 minutes, respectively. Although the screen provides a huge amount of data, it fails to pinpoint at targets that are clearly Hog1-dependent, as no experiments in which Hog1 was deactivated were included in their study. Nevertheless, the authors observe a number of S/T-P motifs that are upregulated in phosphorylation upon stress induction. However, at the same time they find several other sequence motifs that become phosphorylated upon hyperosmotic stress. Bioinformatic analysis of these sequence motifs suggested that more kinases are involved in the regulation of phosphorylation patterns upon osmotic disbalance than previously anticipated [89]. Therefore it still remains elusive to what extent the proteome is directly influenced by Hog1 activity.

II.2. Project Design

This thesis aims at extending the list of potential Hog1 targets by a specifically designed experimental pipeline that allows distinguishing between direct and indirect Hog1-mediated phosphorylation events. First, we employed stable isotope labeling of amino acids (SILAC) in combination with high-resolution mass spectrometry to screen for phosphorylations that become upregulated during hyperosmotic stress. Furthermore, using an analogue sensitive allele of Hog1, we tried to identify the proteins whose phosphorylation patterns are dependent on Hog1. Those proteins that display a phosphorylated S/T-P site that fulfills the described criteria, would then qualify as potential Hog1 downstream targets. To further

differentiate these candidates into direct and indirect substrates, we employed the M-track system to detect transient interactions between the Hog1 kinase and its alleged substrate. The resulting collection of novel direct protein targets provides an input for further studies where the behavior of the Hog pathway is linked to other cellular processes during hyperosmotic stress. Furthermore, we were interested in the role of trehalose in osmoregulation and exemplify the target validation approach with Tsl1, the large subunit of the trehalose-synthase complex.

II.3. The Original Hog1 Screen Dataset

The starting point of the project presented here was an extensive dataset that has been previously generated by our research group. The primary aim of that MS-shotgun screen was to follow up changes in the phosphorylation patterns of the yeast proteome upon hyperosmotic conditions in *wildtype* cells and also in different mutants of the Hog1 MAPK pathway to be able to extract potential Hog1 substrates. For this purpose, various shotgun-experiments (listed in *Table I*) were designed, employing SILAC technology and high-resolution mass spectrometry. Usually the ^{13}C-labeled-population (subsequently referred to as “heavy”) served as the unstressed control, whereas the ^{12}C-labeled one (“light”, respectively) was stressed with 0.5M NaCl for a given amount of time.

After harvesting equal amounts of ^{13}C and ^{12}C-labelled cells, proteins were extracted and digested with trypsin. To obtain more insight into the phosphoproteome, samples were further processed using previously described phosphopeptide enrichment protocols, such as immobilized metal affinity chromatography (IMAC) and TiO_2 in the presence of 2,5-dihydroxybenzoic acid. The resulting probe was then analyzed by online LC-MS/MS in a so-called collision-induced decay (CID) activation mode and mapped against a yeast protein database (2007) with the *BioWorks (©Thermo Fisher Scientific)* algorithm. Nevertheless, it is important to note that *Thermo Scientific© BioWorks* is an out-of-date platform used for MS-data evaluation. Peaks were analyzed within a retention time window of a tenth of a minute, allowing for mass deviations of up to 0.01Da. It is important to note that this corresponds to a mass accuracy that is five times higher than in modern MS instrumentation. Peptide ratios were calculated as the relative difference between the peptide intensities observed in light states and the heavy control (subsequently referred to as “Light/Heavy”-ratio (L/H)).

Experiment Names	Heavy Conditions (control)	Light Conditions
wt+-0min stress	wildtype, no stress	wildtype, no stress
wt+-5min stress	wildtype, no stress	wildtype+0.5M NaCl for 5min
wt+-10min stress	wildtype, no stress	wildtype+0.5M NaCl for 10min
wt+-15min stress	wildtype, no stress	wildtype+0.5M NaCl for 15min
wt+-20min stress	wildtype, no stress	wildtype+0.5M NaCl for 20min
wt+-30min stress	wildtype, no stress	wildtype+0.5M NaCl for 30min
Δhog1+-0min stress	Δhog1, no stress	Δhog1, no stress
Δhog1+-5min stress	Δhog1, no stress	Δhog1+0.5M NaCl for 5min
Δhog1+-10min stress	Δhog1, no stress	Δhog1+0.5M NaCl for 10min
Δhog1+-15min stress	Δhog1, no stress	Δhog1+0.5M NaCl for 30min
Δhog1Δste11+-5min stress	Δhog1 Δste11, no stress	Δhog1 Δste11+0.5M NaCl for 5min
Δhog1 Δste11+-10min stress	Δhog1 Δste11, no stress	Δhog1 Δste11+0.5M NaCl for 10min
Δpbs2+-5min stress	Δpbs2, no stress	Δpbs2+0.5M NaCl for 5min
hog1*as*+inh, +-5min stress	hog1*as*+ *as*-specific inhibitor, no stress	hog1*as*+ *as*-specific inhibitor+ 0.5M NaCl for 5min
hog1*as*+inh, +-10min stress	hog1*as*+ *as*-specific inhibitor, no stress	hog1*as*+ *as*-specific inhibitor+ 0.5M NaCl for 10min
hog1*as*+inh, +-30min stress	hog1*as*+ *as*-specific inhibitor, no stress	hog1*as*+ *as*-specific inhibitor+ 0.5M NaCl for 30min
hog1+inh, +-5min stress	hog1+hog-specific inhibitor, no stress	hog1+hog-specific inhibitor+ 0.5M NaCl for 5min
hog1+inh, +-10min stress	hog1+hog-specific inhibitor, no stress	hog1+hog-specific inhibitor+ 0.5M NaCl for 10min
hog1+inh, +-30min stress	hog1+hog-specific inhibitor, no stress	hog1+hog-specific inhibitor+ 0.5M NaCl for 30min
hog1*as*+-inh, +10min stress	hog1*as*, no inhibitor, + 0.5M NaCl for 10min	hog1*as*+ *as*-specific inhibitor, + 0.5M NaCl for 10min

Table I. *The original experimental setup for the analysis of potential Hog1 targets.*

A significant part of the experiments aimed at identifying proteins that change their phosphorylation pattern during osmotic stress: To this end, light cell populations treated with 0.5M NaCl were harvested at different time points to resolve the timely pattern of Hog1 activity through the progression of the stress response. Several studies have shown a rapid increase of the phosphorylation of the Hog1 activation site (TGY), peaking at approximately 5min of osmostress stimulation and then again decreasing to basal levels within 30 minutes [59,104]. Accordingly, time-points in the stress-experimental set were set at 0, 5, 10, 20 and 30 minutes allowing monitoring of the activity kinetics of the kinase [105]. The list of *stress-induced* phosphorylations identified within this dataset is both extensive and coherent with the literature descriptions of known targets: 551 upregulated phosphogroups were identified within 289 proteins, out of which 196 phosphogroups harbored an S/T-P MAPK motif (corresponding to 131 proteins). The screen identified various prominent hallmarks of the Hog1 pathway that could be used as markers for the quality of the dataset. These included upstream components of the HOG MAPK pathway: For instance, the regulatory site *serine-248* (*S248*) of the MAPKK Pbs2 [6,104,106] was found seven times more abundant in the stressed sample than in the untreated one. Additionally, several described S/T-P sites of the Sho1-pathway-related MAPKKK Ste20 *(T203,S502)* were found to be upregulated due to the stress condition [107]. Finally, double phosphorylation of the *TGY*-motif (*T174* and *Y176*) of the MAPK Hog1 itself was found to be 50 times induced.

Protein Candidates for Target Validation	Description	Hog1 Dependency
Tsl1	*Large subunit of trehalose 6-phosphate synthase/phosphatase complex; TSL1 has a paralog, TPS3, that arose from the whole genome duplication*	Yes
Ede1	*Key endocytic protein involved in a network of interactions with other endocytic proteins, binds membranes in a ubiquitin-dependent manner, may also bind ubiquitinated membrane-associated proteins*	Yes
Sla1	*Cytoskeletal protein binding protein; required for assembly of the cortical actin cytoskeleton; interacts with proteins regulating actin dynamics and proteins required for endocytosis*	No
Bul1	*Ubiquitin-binding component of the Rsp5p E3-ubiquitin ligase complex; disruption causes temperature-sensitive growth*	Yes
Pan1	*Part of actin cytoskeleton-regulatory complex Pan1p-Sla1p-End3p; associates with actin patches on cell cortex; promotes protein-protein interactions essential for endocytosis*	Yes
Rgc1	*Pleckstrin homology domain containing protein; proposed to function as a glycerol channel activator; RGC1 has a paralog, ASK10*	Yes
Rgc2/Ask10	*Component of RNA polymerase II holoenzyme; phosphorylated in response to oxidative stress; has a role in destruction of Ssn8p; proposed to function in activation of the glycerol channel Fps1p*	Yes
Fen1 (?)	*Fatty acid elongase, involved in sphingolipid biosynthesis; acts on fatty acids of up to 24 carbons in length*	Yes
Other candidates	*Reg1?*	

Table II. *Protein Candidates selected from the original screen for further target validation using HTB tandem affinity purification and subsequent mass spectrometry analysis.*

The upregulation of this Hog1-peptide is a good indicator for the activation of the Hog1 pathway (L/H-ratio~ 50 after 5 minutes of stress, leveling down to 20 after 15 minutes) [6,104].

Finally, some renowned downstream targets of the MAPK were also detected in this screen (see *Supplementary Table I*), such as the transcriptional repressor Sko1 (L/H for *T226* approximately ~13) [63,66] and the previously described target kinase Rck2, which is involved in the translational regulation during osmotic stress conditions [74].

Furthermore, the endocytic factor Pan1 was among the group of newly identified putative Hog1 substrates. Phosphorylation at several S/T-P motifs of Pan1 became induced in response to hyperosmotic stress. Interestingly, the involvement of endocytic factors has not been described in the framework of the yeast osmotic stress response before. When following up *in vivo* modification events of purified Pan1, Reiter *et al.* [79] confirmed that this endocytic factor is indeed a Hog1 target.

Additionally, Hog1-mediated phosphorylation of the glycerol channel regulators Rgc1 and Rgc2 has also been observed for the first time in these shotgun experiments. This has led to a follow-up study that has been recently published [11].

Analogously, other potential Hog1-substrates identified in this screen were tagged and purified for further target validation using mass spectrometry (see *Table II*). For example, some phosphogroups of the large subunit of trehalose synthase (Tsl1) also displayed a significant up-regulation under stress conditions (*S147/S155* (covered by sequence *IAS#PIQHEHDS#GSR*), *S161* (covered by sequence *IAS#PIQQQQQDPTRNLLK*)) in our screen. Interestingly, phosphopeptides of the neutral trehalase (Nth1) have also been found to be regulated upon hyperosmotic stress, providing indication of a possible link between trehalose synthesis and Hog1 activity. As mentioned above, trehalose is an important compatible osmolyte and several studies pointed at its involvement in conquering major turgor pressure imbalances in the first seconds of osmotic stress.

Here I will present a separate *in vivo* study of the phosphorylation pattern of Tsl1 in *wildtype* and *hog1Δ* mutants and I will demonstrate that Tsl1 is indeed regulated by Hog1 (see *III.4).*

II.4. MS-shotgun pitfalls exemplified on the original Hog1 Screen Dataset

Even though the original dataset led to the identification of novel Hog1 substrates, the overall output of the study was rather unsatisfying. This was mainly due to three main problems: First, when comparing datasets, it was often observed that the same S/T-P sites became upregulated in *wildtype* as well as in *hog1Δ* mutants. This indicates a Hog1-independent regulation of these sites upon stress induction, which might be explained by the activity of other independent kinases (as suggested by Soufi *et al.* [89]) or crosstalk-activity (see *Figure 8*). Therefore, even though many S/T-P sites were found to be upregulated, no Hog1-dependency could be confirmed for many other potential candidate proteins, such as, for example, for the endocytotic factor Sla1 (see *Supplementary Databases: HTB-purifications*).

Second, when datasets were compared, only a low overlap of quantified PSMs (and therefore S/T-P-carrying peptides) was found. Therefore, many S/T-P sites were either covered in one or the other experimental set-up. However, identification of a true Hog1 target requires a distinct upregulation upon stress in the *wildtype*, which should not be observed with the *hog1Δ* mutant. With several S/T-P sites found upregulated upon hyperosmotic stress in the wildtype that have not been covered in the mutant experiment, it is likely that there are a number of additional, yet unidentified Hog1 targets missing.

Third, as mentioned above there are also several S/T-P sites that have only been found in the mutant/Hog1-inhibited experiments, showing a 1:1 ratio when logarithmic growth and stress treatment are compared. Without the information on a corresponding (upregulatory) phosphorylation event in the *wildtype*-experiment, one cannot say whether this ratio is Hog1-dependent or –independent. A differently designed experimental set-up would have probably been more suitable. For instance, comparing *wildtype* versus the *hog1Δ* cells under stressed conditions would directly show a ratio for all sites that are Hog1-dependent.

If unstressed, this set-up allows to study the "housekeeping" phosphorylation activity of Hog1, e.g. its inhibition of other MAPK pathways that would otherwise get activated and

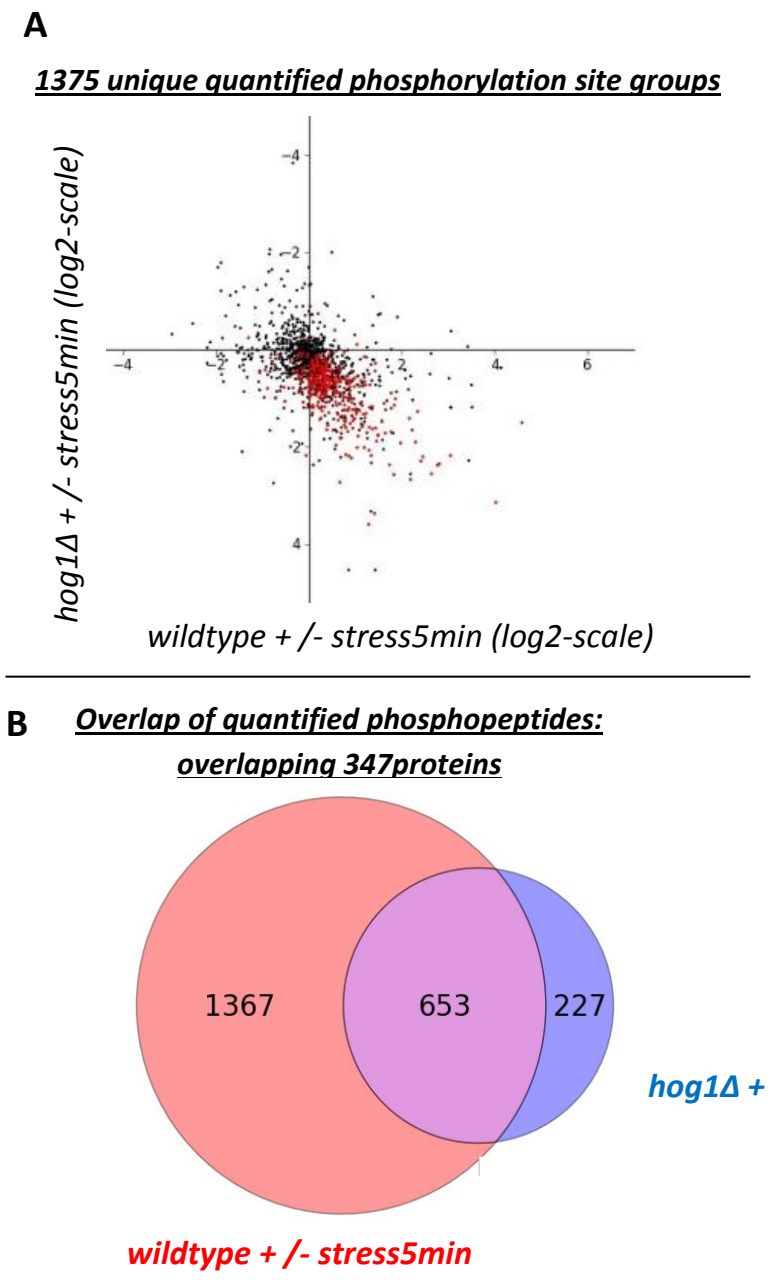


Figure 8. *A) Scatterplot of phosphogroups that have been found quantified in wildtype+- stress and hog1Δ+-stress experiments, evaluated by Bioworks ©. Shown are target sites that become elevated upon osmotic stress in both mutant and wildtype. The black dots corresponding to all quantified phosphorylation site groups and the red dots specifically to the S/T-P-phosphorylated quantified phosphorylation site groups. Since they are upregulated in both experiments, they would not correspond to direct Hog1 targets as they would increase in abundance even in the absence of Hog1. The figure does not display any trend or correlation that might facilitate or hint at possible Hog1 MAPK targets. B) Venn diagram showing the overlap of quantified phosphogroups (1375 as shown in A) between wildtype+-stress experiments and hog1Δ+-stress experiments.*

interfere with quantification in the previously described mutant experiments. One particularly well studied MAPK cross-talk of Hog1 is the interaction with the Kss1 MAPK module (filamentous growth pathway) [109]. If both SILAC cultures get osmo-stressed, however, we should observe a subset of cellular proteins that additionally become activated by elevated Hog1 kinase activity. In terms of SILAC quantifications, target phosphorylation sites would be downregulated (decreased L/H-ratio).

III. Results

III.1. The Reevaluated Dataset is not Ideal for Hog1 Target Search

As described above, detection of Hog1 targets by using the MS dataset generated and analyzed with *Bioworks©* (evaluation performed by Ilse Dohnal) proved to be difficult. To accurately define a target, the two datasets "*stressed versus unstressed wildtype (5min)*" and "*stressed versus unstressed Δhog1 (5min)*" should ideally overlap for a high number of peptides. However, the datasets displayed a low peptide overlap (see *Figure 8/B, Introduction*), severely restraining any adequate search for Hog1 stress targets. We figured that there are two possible explanations for these unsatisfying results: (1) the applied MS-data evaluation strategy and algorithm used for it was inadequate and/or (2) there have been variations in the sample quality itself.

Furthermore, *Bioworks©* did not allow for accurate SILAC quantification as the algorithm would not search for complementary isotope patterns, but for the precursor mass only. This fact severely restrains the search for reliable Hog1 targets and gives rise to false positive Hog1 hits as well.

Overlap of peptides between Soufi and Bioworks©:

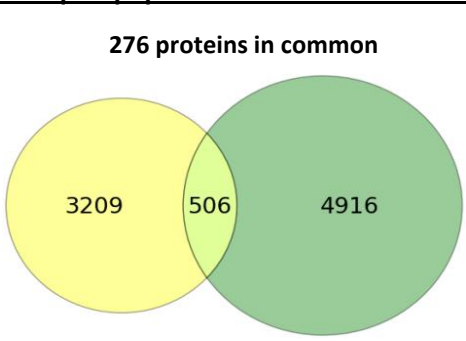


Figure 9. *Venn diagrams presenting the overlap of peptide-protein groups. Comparison of our Bioworks© dataset (yellow set) with published dataset by Soufi et al. (green set).*

Given the fact that the original dataset as such has already given rise to newly identified Hog1 targets, for example Pan1 and Ede1 (Reiter *et al.*[79]), we wanted to see whether we could enhance the overlap of peptides and S/T-P phosphorylation sites between the two corresponding datasets by re-analzying the MS rawfiles. If increased, this would help to shed light on other potential Hog1 targets that have so far gone amiss.

Such a re-evaluation might also lead to an increased overlap between our dataset and the dataset published by Soufi *et al.* [89], who performed experiments similar to the condition that we chose for experiment "*stressed versus unstressed wildtype (5min)*". So far, only 10.3% of the peptides detected by Soufi *et al.* were also discovered in our dataset (see *Figure 9*). Note, however, that Soufi *et al.* performed protein digestion with LysC.

We therefore decided to re-evaluate the raw-files with the *Thermo Scientific Proteome Discoverer©*. *Proteome Discoverer©* is the further developed version of *Bioworks©,* also

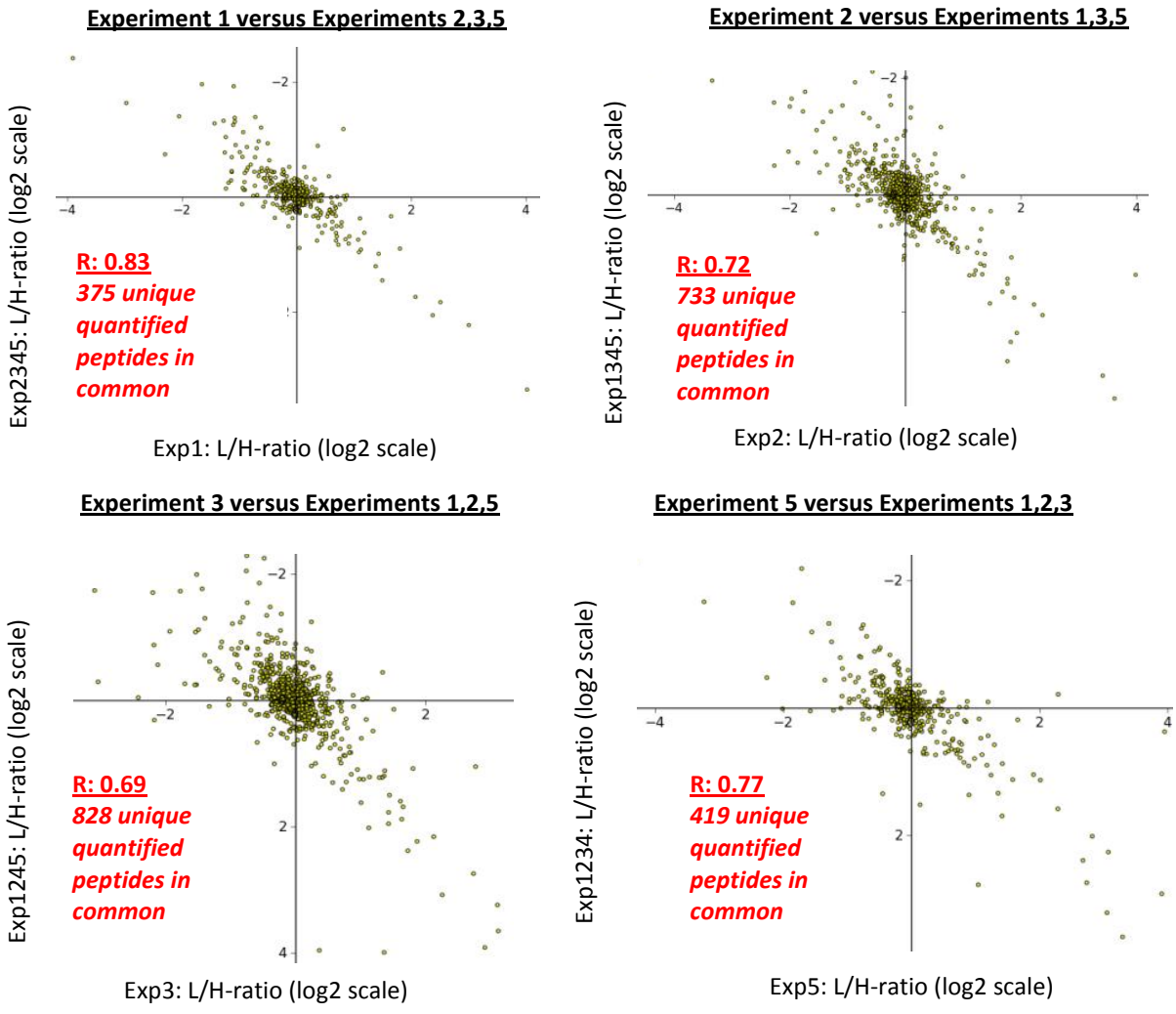


Figure 10. *Quantification reproducibility between biological replicates of the experiment type 'stressed versus unstressed wildtype (5min)'. The ratio scatter plots show protein abundance change after 5 min of hyperosmotic stress as measured in each experiment and plotted versus the combined ratio of the remaining four experiments. The combined ratio is the geometric mean of the log2-transformed ratio values. The Pearson correlation coefficient is shown in the bottom left corner of each plot.*

employing the SEQUEST algorithm, but with more accurate quantification principles applied. Correct PSMs identified in *Bioworks©* should therefore also be identified by the *Proteome Discoverer©*, whereas the false positive rate should decrease due to the developed algorithm.

To assess whether the newly evaluated list of protein identifications and ratios provides statistically reliable data, we first checked the ratio reproducibility of the relevant experiments at the level of identified peptides (see *Figure 10*). For the exemplary data, the Pearson correlation coefficient for peptides ranged from 0.7 to 0.8 for four out of five experiments, indicating a highly consistent dataset. Only experiment 4 was not coherent with the other biological replicates and was therefore removed from the dataset for further analysis.

For an appropriate comparison between the original dataset as analyzed by *Bioworks©* and the re-evaluated dataset, it is important to keep the following considerations in mind: (1) Contrary to protein-peptide groups, only peptide *sequences* (without protein assignment) were considered due to ambiguous protein-peptide assignments in *Bioworks©*. These sequences would also not hold any information as to whether it contains isoleucine or leucine. (2) Any site information on individual phosphopeptides was skipped since no adequate phosphorylation site allocation tool (such as phosphoRS [111]) was implemented in *Bioworks©*. (3) As *Bioworks©* filtering was performed using a strict score-cutoff at different charge states, these filtering settings were also applied to the re-evaluated dataset for comparison purposes only (see *Figure 11/B*). (4) Charge states of peptides equal to 5 or above were not considered as the *Bioworks©* dataset does not contain them. Grouping all four experiments in which *wildtype* cells have been exposed to hyperosmotic stress according to these criteria revealed 72.5% of all unique sequences detected by *Bioworks©* to be also covered by the *Proteome Discoverer©* (see *Figure 11/A*, *Supplementary Table II: Reevaluated Data 2007*). Interestingly, *Proteome Discoverer©* assigned 1,804 peptide sequences anew which have not been observed with *Bioworks©.* Note that these additional assignments are primarily due to the less strict FDR filtering in the re-evaluated list with score thresholds calculated at different charge states.

The removal of phosphorylation site information from individual phosphopeptides (sequence and number of phosphorylation as information left) contributed significantly to the overlap enlargement. Due to the lack of scan information, the increased overlap can

Re-evaluation Analysis: „ *stressed versus unstressed wildtype (5min)* " in *Bioworks©* and *Proteome Discoverer©*

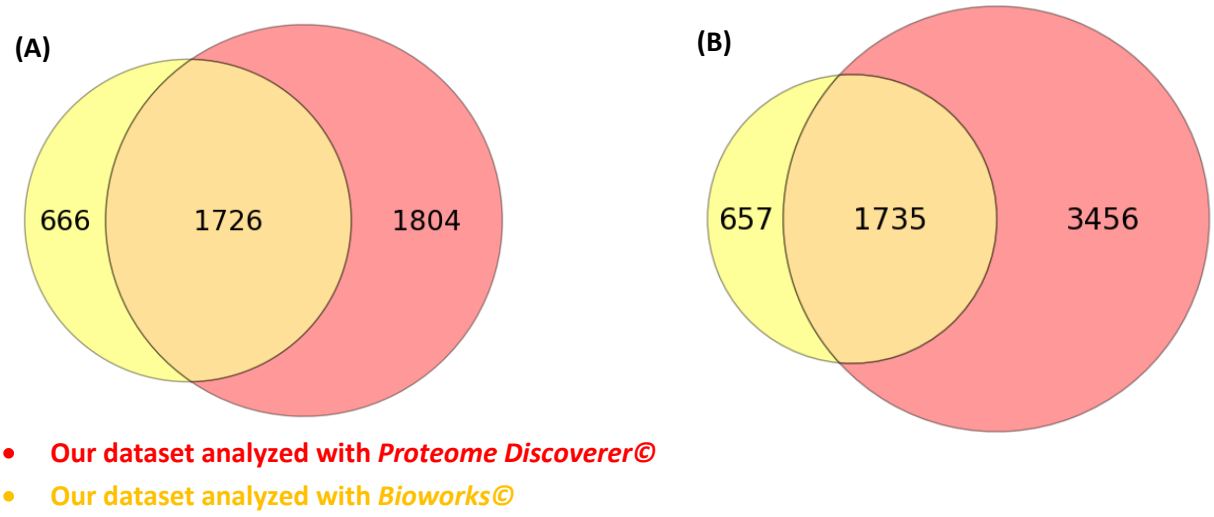


Figure 11. *Venn diagrams presenting the overlap of peptide sequences without phosphorylation site information, leucine/isoleucine distinction and charges above 5. Comparison of our Bioworks© dataset (yellow set) with re-evaluated dataset as analyzed with Proteome Discoverer©. (A) filtering criteria for Discoverer© dataset according to q-value calculation (q-value<1), (B) filtering criteria according to Bioworks© dataset (see (1) in text).*

merely hint at a differential phosphorylation site allocation applied by the two search engines. For instance, a peptide of the protein kinase Rck2 was found to be double-phosphorylated at *S516* and *S520* with *Bioworks©*. This peptide likely corresponds to the Rck2 peptide discovered in the re-evaluated set, where the phosphorylation sites have been allocated to *S515* and *S520*. Such considerations are especially important when it comes to multiple phosphorylated peptides and peptide species with many residues that have the potential of becoming phosphorylated.

657 peptide sequences that had originally been identified by *Bioworks©*, have not been covered by *Proteome Discoverer©* (see *Figure 11/B*). Surplus peptides are not unexpected due to additional features that differentiate the two search engines, for example, the difference of handling MS3 scans. While *Proteome Discoverer©* assigns MS3 scans with a precursor mass tolerance of 7ppm, *Bioworks©* did so with a spacious mass window of 2 Da. The latter is the cause for high variation in actual spectral identification between the two datasets.

Finally, a slightly different yeast database (fasta-file) was used in the *Proteome Discoverer©*, that has additional contaminant entries. Hence, *Bioworks©* was evaluating a set of PSMs slightly distinct from the one analyzed by *Proteome Discoverer©* which is important for the

eventual FDR filtering. Because of the issues described and others, following up individual scans proves difficult and will not be discussed in this report.

The re-evaluated datasets were assessed in regard to (1) quantification credibility and (2) peptide overlap between datasets for Hog1 target search. For (1) I exemplarily describe the double-phosphorylated Tsl1 peptide *IAS#PIHEHDS#GSR* covering the phosphorylation sites *S147* and *S155* which has been shown to have the highest L/H-ratio in the *Bioworks©* dataset (L/H~596.36). With *Proteome Discoverer©* only the site *S147* could be assigned confidently in a double phosphorylated variant *IAS#PIHEHDS(?)GS(?)R*. The light peptide has been fragmented, but no heavy complementary peptide could be matched for quantification. While *Bioworks©* assigned an implausible SILAC ratio of 596.36 to the peptide, *Proteome Discoverer©* does not perform a quantification due to the absence of a reliable heavy isotope pattern. Since the elution profile does not display any peptide interfering with a potential heavy peptide signal, it is likely that heavy peptide analogue is not present in the sample in an amount detectable by mass spectrometry. With *Proteome Discoverer©,* we could also detect a single-phosphorylated Tsl1 peptide variant

Re-evaluation Analysis: overlap of *quantified* peptides between „ *stressed versus unstressed wildtype (5min)*" and "*stressed versus unstressed Δhog1 (5min)*"

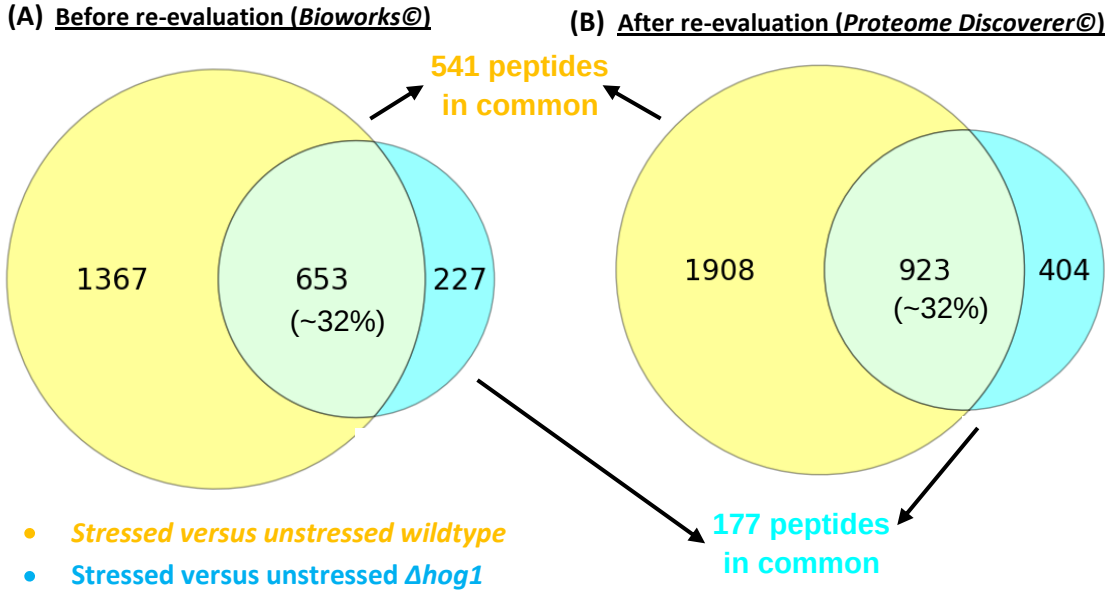


Figure 12. *Venn diagrams presenting the overlap of protein-peptide groups between experiments "stressed versus unstressed wildtype (5min)" (yellow set) and stressed versus unstressed Δhog1 (5min)" (blue set). (A) Comparison of datasets before re-evaluation, hence with Bioworks© datasets, (B) comparison of datasets after re-evaluation, hence with Proteome Discoverer© datasets.*

IAS#PIQHEHDSGSR (*S147*), which has not been covered by *Bioworks©*. This site increased in abundance in response to hyperosmotic stress conditions in the re-evaluated dataset (L/H~4.2).

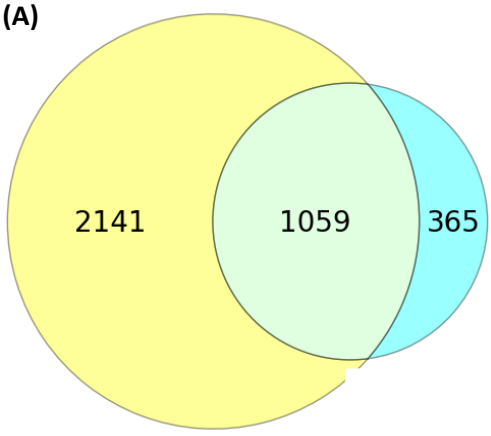
Next remains the question whether the overlap between the two relevant datasets "*stressed versus unstressed wildtype (5min)*" and "*stressed versus unstressed Δhog1 (5min)*" (2) increased by re-evaluation with *Proteome Discoverer©*. In terms of quantified peptides, both datasets increased in the number of peptides after the re-evaluation procedure, with 541 (19%) quantified peptides joining the "*stressed versus unstressed wildtype (5min)*" and 177 (13%) joining the "*stressed versus unstressed Δhog1 (5min)*" experiment (see *Figure 12*).

			Discoverer: Average L/H ratio	
Protein name	Phospho group	SP/TP site	stressed vs. unstressed wildtype (5min)	stressed vs. unstressed *Δhog1* (5min)
SVL3	471(S)	yes	3.12	1.76
SEC16	706(S)	yes	2.35	1.57
VIP1	54(S)	yes	2.35	1.87

Figure 13. *Table shows three sites that become upregulated in response to hyperosmotic stress and do not if Hog1 is knocked out.*

Re-evaluation Analysis: overlap of quantified peptides between *„stressed versus unstressed wildtype (5min)"* and *"stressed versus unstressed Δhog1 (5min)"* in *MaxQuant* searches

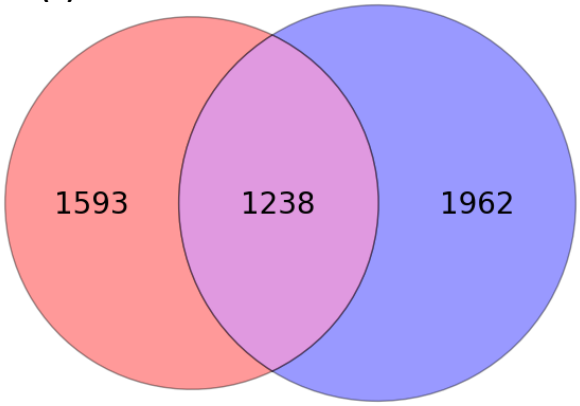
(A)



- ***Stressed versus unstressed wildtype***
- **Stressed versus unstressed *Δhog1***

Re-evaluation Analysis: overlap of *quantified* peptides between *„ stressed versus unstressed wildtype (5min)"* datasets evaluated by *Proteome Discoverer©* and *MaxQuant*

(B)



- ***"stressed vs. unstressed wildtype"* dataset analyzed with *Proteome Discoverer©***
- ***"stressed versus unstressed wildtype"* dataset analyzed with *MaxQuant***

Figure 14. *(A) Venn diagram presenting the overlap of protein-peptide groups between experiments "stressed versus unstressed wildtype (5min)" (yellow set) and stressed versus unstressed Δhog1 (5min)" (blue set) as evaluated by MaxQuant. (B) Overlap of quantified peptides between dataset "stressed versus unstressed wildtype (5min)" analyzed with Proteome Discoverer© (red set) and with MaxQuant (blue set).*

However, the relative overlap between the two datasets remained the same; only 32% of quantified peptides found in "*stressed versus unstressed wildtype (5min)*" were effectively covered in the complementary dataset. At the level of S/T-P sites we only found three phosphorylation sites to fit the regulation pattern expected for putative Hog1 targets (see *Figure 13*). None of these has been described as hallmarks for Hog1 activation, however.

We figured that an additional re-evaluation using *MaxQuant 1.3.0.5* might increase the low overlap of quantified peptides. Quantified peptides retrieved from corresponding *MaxQuant* searches of the most representative experiment ("*stressed versus unstressed wildtype (5min)"*) covered 44% of the peptides quantified by *Proteome Discoverer©* (see *Figure 14/B*). Again, these discrepancies occur most likely due to different scoring, FDR calculation and filtering. When again comparing the datasets "*stressed versus unstressed wildtype (5min)*" and "*stressed versus unstressed Δhog1 (5min)*" for overlapping quantified peptides, *MaxQuant* did not improve the relative overlap (33%, see *Figure 14/A*).

Hence, our re-evaluation strategy unfortunately did not provide sufficient overlap between the datasets "*stressed versus unstressed wildtype (5min)*" and "*stressed versus unstressed Δhog1 (5min)*" as would be needed for an acceptable search for Hog1 targets. We therefore decided not to follow up potential substrates in the recycled datasets, but to generate new reliable MS data. Note that the experiment "*stressed versus unstressed Δhog1 (5min)*" raises some difficulties in interpretation, as has been discussed in *III.4*. For the new experiments we hence deployed the hog1*as*-allele along with a highly specific hog1*as*-specific inhibitor to circumvent these problems. Finally, we ventured that in order to increase the set of potential substrates (found in the overlap of corresponding datasets), sample complexity must be drastically reduced so as to allow for more identifications in general.

III.2. Design of New Shotgun Experiments to Expand the Dataset

II.2.1. Implementation of the SCX technology for Complexity Reduction

To do so, we utilized the strong cation exchange chromatography technology. By user-dependent fractionation it allows reduction of sample complexity and ion suppression, leading to increased reproducibility and improved spectral quality.

Basically, there were two types of experiments: In the first experiment, the light *wildtype* population was exposed to five minutes of osmotic stress (0.5M NaCl), in contrast to the

Experiment Name	Heavy (control) conditions	Light conditions	# of biological replicates
Stress response experiment	*wildtype*, no stress	*wildtype*, treated with 0.5M NaCl for 5min	2
Hog1 basal activity	Hog1*as*, no inhibitor, unstressed	Hog1*as*, treated with *as*-specific inhibitor, unstressed	2
Hog1 stress activity/5min	Hog1*as*, no inhibitor, treated with 0.5M NaCl for 5min	Hog1*as*, treated with *as*-specific inhibitor, treated with 0.5M NaCl for 5min	2
Hog1 stress activity/10min	Hog1*as*, no inhibitor, treated with 0.5M NaCl for 10min	Hog1*as*, treated with *as*-specific inhibitor, treated with 0.5M NaCl for 10min	2

Table III. *New Experimental Setup for the analysis of potential Hog1 targets.*

heavy labeled *wildtype* control which remained untreated and thereby served as a control (see *Table III*). We will further refer to that experiment as the "*stress response experiment*". The second experiment then exposed only the light population of *hog1as* cells to inhibitor treatment for ten minutes and to subsequent zero, five or ten NaCl stress, whereas the complementary heavy population was exposed to osmotic stress for the corresponding time. These three experiments will be further referred to as "*Hog1 basal activity experiment*", and "*Hog1 stress activity/5min (10min)*", respectively (see *Table III*). For each of the four

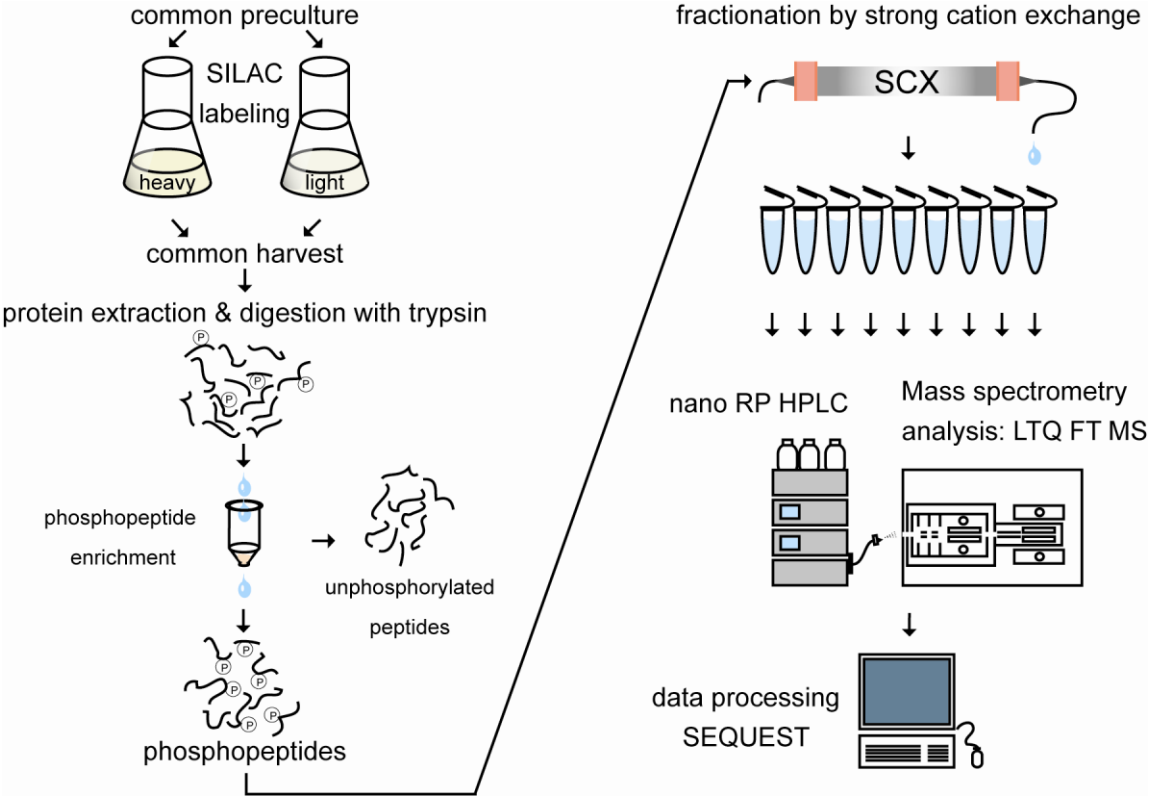


Figure 15. *Quantitative proteomics workflow (Figure kindly provided by Dr. Wolfgang Reiter©). Control and treated cells are labeled by heavy and light versions of arginine and lysine, harvested and lyzed together and then digested by trypsin. Phosphopeptides are first enriched by TiO_2 chromatography and then fractioned by SCX. Individual and pooled fractions are separated on a nanoflow-HPLC and directly measured in a mass spectrometer. Data processing was performed with the SEQUEST algorithm (Proteome Discoverer©); additionally, we also used MaxQuant 1.3 to evaluate raw spectra.*

experiment types, two biological replicates were made.

Protein extraction from SILAC-encoded yeast cell populations were generated by a TriZol purification protocol. After each purification, proteins (approximately 10mg/sample) were trypsinized and then selectively enriched for phosphopeptides with TiO_2. The enrichment throughout all experiments generally ranged from 70% to 80%. The TiO_2-enriched samples were then separated by SCX chromatography with fractions collected every two minutes in an 80min-gradient (40 fractions per run in total). Based on peak intensities in the UV spectra (λ=240nm), fractions containing less peptides were pooled. All fractions were then desalted, reduced in volume and injected for HPLC-MS/MS-analysis (see *Figure 15*). The MS-data was then analyzed with *Thermo Scientific Proteome Discoverer©*. In total approximately 20 fractions were generated for each experiment, as exemplified in *Figure 16*.

Additionally, we also applied SCX-stage tips and manually fractionated the sample by elution with different saline concentrations (from 10mM NaCl to 500mM NaCl for total elution of the sample from the column). Thereby twelve SCX-fractions were produced. This method was performed in order to compare different SCX enrichment protocols and was only applied for one experiment (second replicate of the experiment "*Hog1 stress activity/10min*" (see *Table IV*)).

All resulting fractions were controlled for phosphopeptide enrichment efficiency (1), peptide overlap between individual fractions (2) and ratio reproducibility (3). While analyzing (1) I generally observed similar patterns for phosphopeptide enrichment in all experiments. First, the ratio of the phosphopeptides relative to the total amount of peptides increased with each fraction (maximum almost 100%) until it drops to a saturation of only ~75% (see *Figure 16/A*). This behavior might be due to the change in the buffer system during the SCX-run. Interestingly, in later fractions the abundance of phosphopeptides, especially multiply phosphorylated species, increases (see *Figure 16/A*). It might be interesting to analyze which types of phosphopeptides are specifically enriched in the fractions before and after the change in the buffer system. Neighboring SCX-fractions were checked for an overlap in peptide identifications (2) to confirm that our pooling/fractionating strategy was efficient (see *Figure 16/B*). For SCX-fractions that have been particularly enriched in phosphopeptides, the overlap encompassed about 30% of the peptides identified in the two neighboring fractions, respectively. The remaining ~70% of peptides was not covered in the neighboring fraction. A similar pattern was observed in the other experiments as well. Next

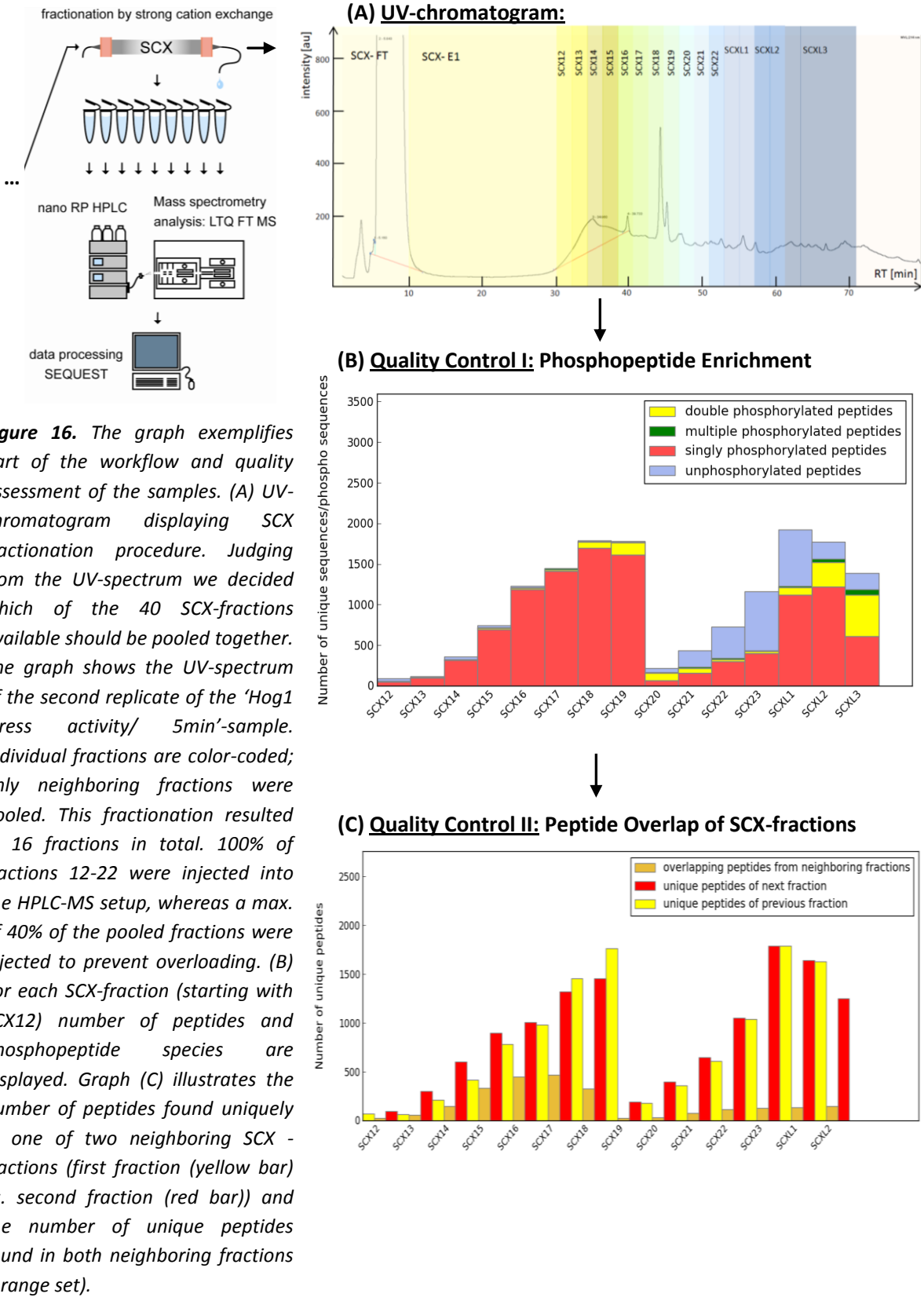


Figure 16. *The graph exemplifies part of the workflow and quality assessment of the samples. (A) UV-chromatogram displaying SCX fractionation procedure. Judging from the UV-spectrum we decided which of the 40 SCX-fractions available should be pooled together. The graph shows the UV-spectrum of the second replicate of the 'Hog1 stress activity/ 5min'-sample. Individual fractions are color-coded; only neighboring fractions were pooled. This fractionation resulted in 16 fractions in total. 100% of fractions 12-22 were injected into the HPLC-MS setup, whereas a max. of 40% of the pooled fractions were injected to prevent overloading. (B) For each SCX-fraction (starting with SCX12) number of peptides and phosphopeptide species are displayed. Graph (C) illustrates the number of peptides found uniquely in one of two neighboring SCX - fractions (first fraction (yellow bar) vs. second fraction (red bar)) and the number of unique peptides found in both neighboring fractions (orange set).*

we assessed SILAC ratio reproducibility of the biological replicates for each experiment (3). For peptide ratios, we observe Pearson correlation coefficients ranging from 70-90%. The

high level of correlation remains conserved throughout the quantified phosphorylation site groups as well, indicating that introducing the definition of PSGs does not significantly influence the quality of SILAC quantification reproducibility (see *Figure 17*). One can clearly

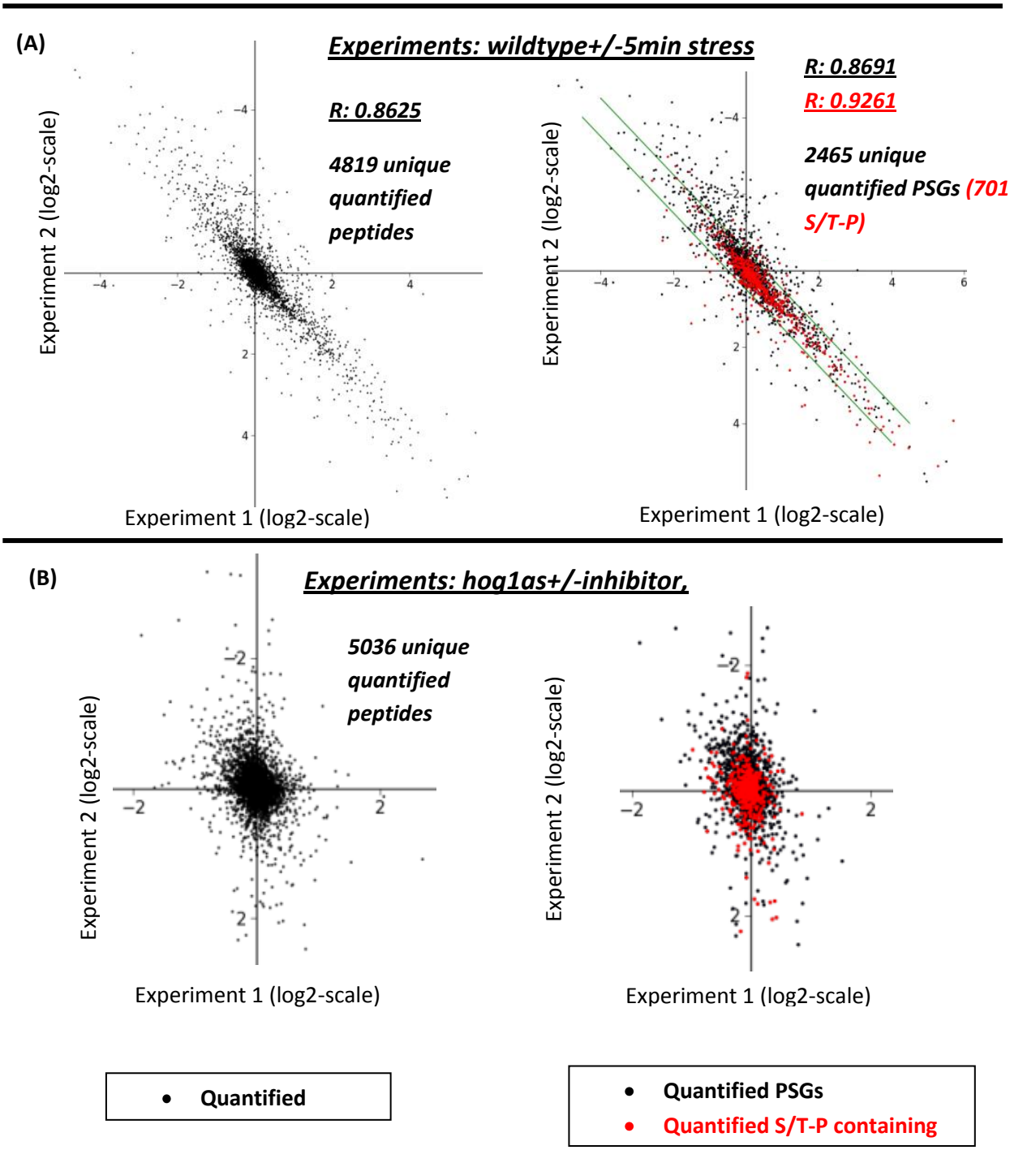
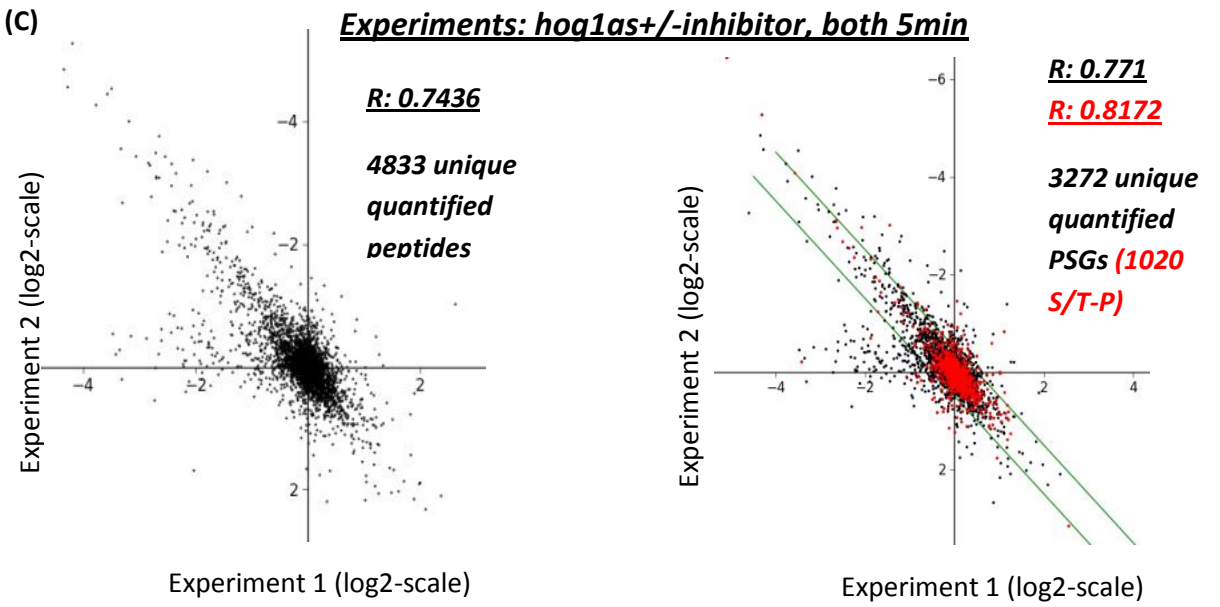
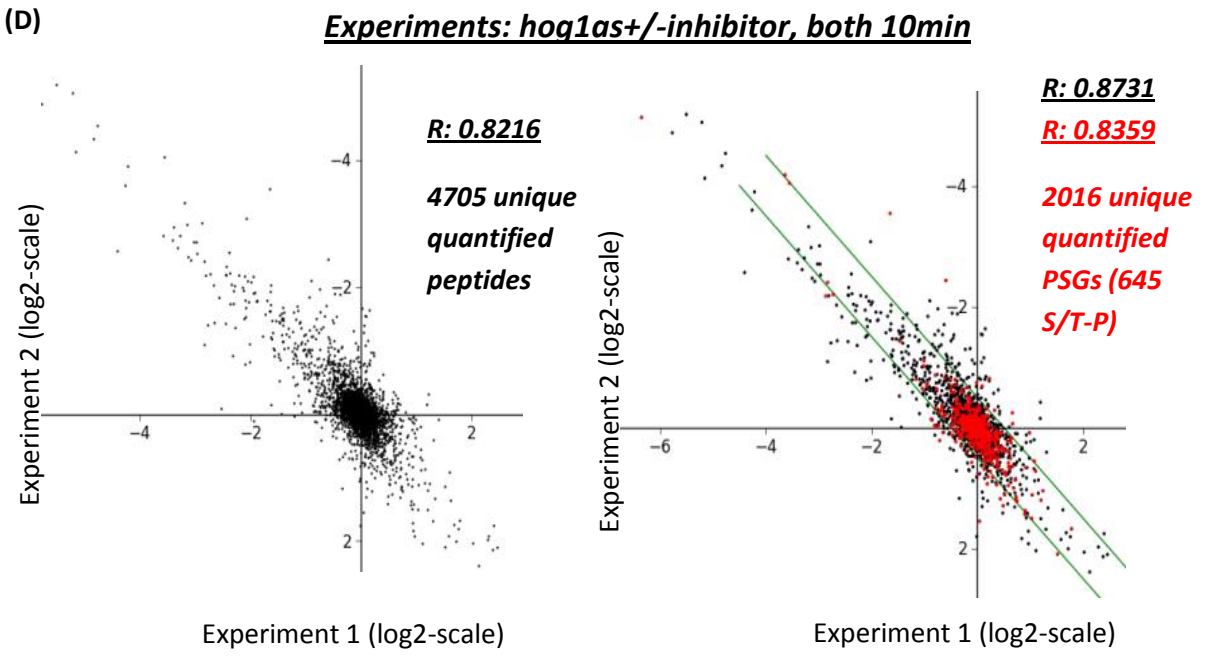


Figure 17. *In each box the SILAC quantification reproducibility between two biological replicates is illustrated at both peptide and phosphorylation site group levels. (A) Stress response experiment, (B) Hog1 basal activity experiment. Only minor changes occur in terms of SILAC ratio perturbations in this experiment; hence no linear Pearson correlation coefficient is depicted. (C) (next page) Hog1 stress activity/5min experiment, (D) (next page) Hog1 stress activity/10min experiment.*





see from *Figure 17/B* that the experiment *Hog1 basal activity* does not display significant ratio changes. This is evidence that the Hog1*as*-inhibitor is highly specific and does not induce secondary effects in the short exposure duration to the inhibitor, as they might occur with *Δhog1*-strains.

HCD versus CID. To decide whether to use the collision-induced dissociation (CID) or the higher-energy collision dissociation (HCD) mode for MS-measurement, we analyzed five fractions of the second replicate of the *stress response experiment* in these two modes. Basically, in the CID-mode, the ions are accelerated by an electrical potential in the ion trap to collide with neutral molecules. The bond breakage and fragmentation of peptides

characteristically results in a *b*-and *y*-series (see *II.1*). In the HCD-mode, in contrast to CID, fragmentation takes place external to the ion trap. The resulting HCD spectra do not suffer from the low mass cutoff of resonant-excitation CID, and provide more high-resolution data due to the different readout [114]. There was no significant deviation at the level of SILAC quantification (see *Figure R18/A*). CID-analyses provided larger data sets than HCD (see *Figure 18/B*) due to faster data acquisition. We were also interested in whether HCD-spectra indeed allow for more reliable phosphorylation site allocation than CID-spectra. In order to test that, we defined the phosphosite confidence on the basis of the phosphoRS-algorithm in *Proteome Discoverer©* [111]. "High" confidence signifies that all phosphosites have been assigned a phosphoRS-probability of more than 75%. If at least one of the sites has a probability of less than 75%, the phosphopeptide is given a "low" confidence assignment. For HCD we found more high-confident data in contrast to CID. With low–confident data, CID provided significantly more results. Thus, we can confirm even with the low amount of data that we used for this comparison that HCD indeed provides more high-confident data. However, since the peptide overlap between CID and HCD is considerable (>50%) and CID provides more spectra in general, we measured our samples in CID-mode.

In total we collected 421,680 confident PSMs, matching to 40,886 peptides. These correspond to 3,885 yeast proteins, making up nearly 65% of the yeast proteome. According to the SEQUEST-phosphoRS analysis with *Thermo Scientific Proteome Discoverer©*, 28,081

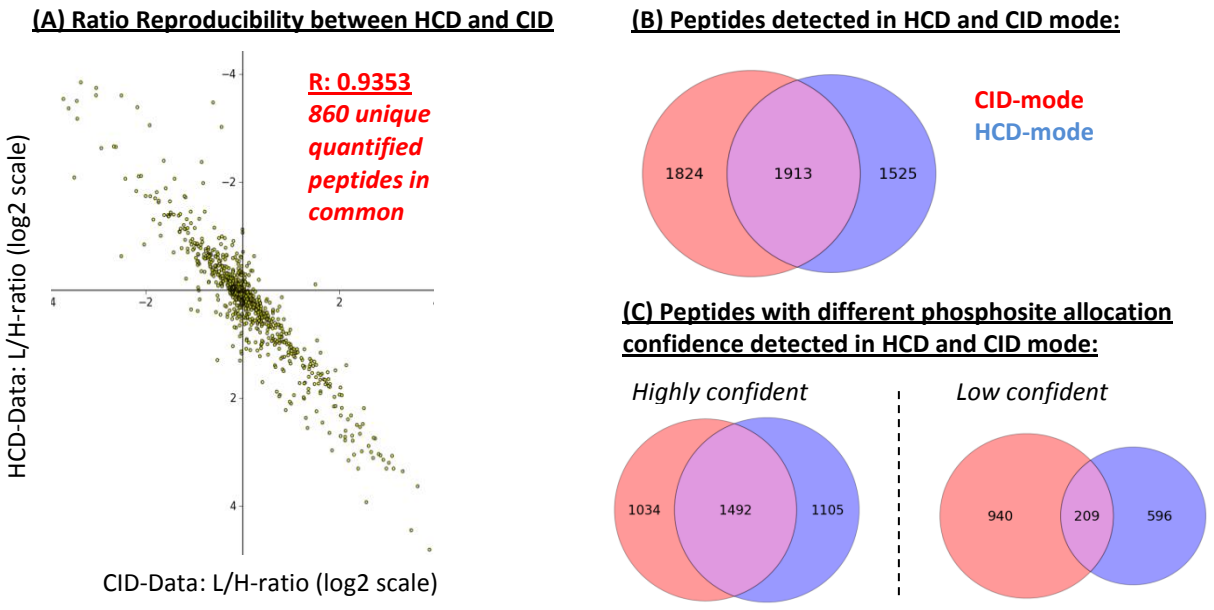


Figure 18. *The graph exemplifies part of the workflow and quality assessment of the samples. (A) Ratio Reproducibility between HCD-measured and CID-measured dataset. (B) Peptide Overlap between the two measurements. (C) Peptides with different phosphosite-allocation confidence detected in HCD- and CID-mode.*

phosphopeptides were reliably detected. As one can see from *Table IV*, however, the number of identified phosphopeptides severely varies between the individual biological experiments. Interestingly, the number of identified proteins hosting at least one phosphorylation site relative to the total amount of proteins identified in the respective sample, remains at 60-80%.

The phosphopeptide-enrichment effect of our approach in comparison to the non-enriched sample was significant as illustrated in *Figure 19*. Whereas the number of unphosphorylated species covered by the non-enriched sample run (200ng injected in total) surmounted the number of unphosphorylated peptides of all SCX-fractions taken together two-fold (see *Figure 19/A*). At the phosphopeptide level it only made out 0.68% of the number of phosphopeptides found in all SCX fractions (see *Figure 19/B*). This again underlines that phosphopeptides are sub-stoichiometric in non-enriched samples and are prone to ion suppression due to the large number of unphosphorylated counterparts. Using SCX-technology together with TiO_2 therefore proves imperative in generating a representative set of phosphopeptides. Along with more identifications, the overlap between the *stress response experiment* and *Hog1 stress activity/5min* also increased (see *Figure 19/C*). This increased overlap now allows the search for Hog1 targets in a systematic manner.

Finally, for the *stress response experiment* we also applied the *MaxQuant 1.3.0.5* search engine to further compare our dataset with the published dataset by Soufi *et al.* (see *Figure*

Proteome Discoverer© Data					
	number of PSMs	number of peptides	number of phosphopeptides (relative to peptide number)	number of proteins	number of phosphoproteins (relative to protein number)
Stress response experiment 1	54,482	18,884	12,574 (67%)	2,708	1,941 (72%)
Stress response experiment 2	78,821	19,309	12,658 (66%)	2,874	2,086 (73%)
Hog1 basal activity experiment 1	64,082	17,485	10,809 (62%)	2,690	1,885 (70%)
Hog1 basal activity experiment 2	31,536	11,266	4,586 (41%)	2,325	1,345 (58%)
Hog1 stress activity/5min experiment 1	60,398	18,277	10,419 (57%)	2,766	1,849 (67%)
Hog1 stress activity/5min experiment 2	50,924	14,460	11,914 (82%)	2,392	1,937 (81%)
Hog1 stress activity/10min experiment 1	28,434	8,704	4,721 (54%)	1,931	1,227 (64%)
Hog1 stress activity/10min experiment 2	50,846	12,803	5,324 (41%)	2,394	1,335 (56%)

Table IV. *Confident numbers of each biological experiment in the newly designed setup.*

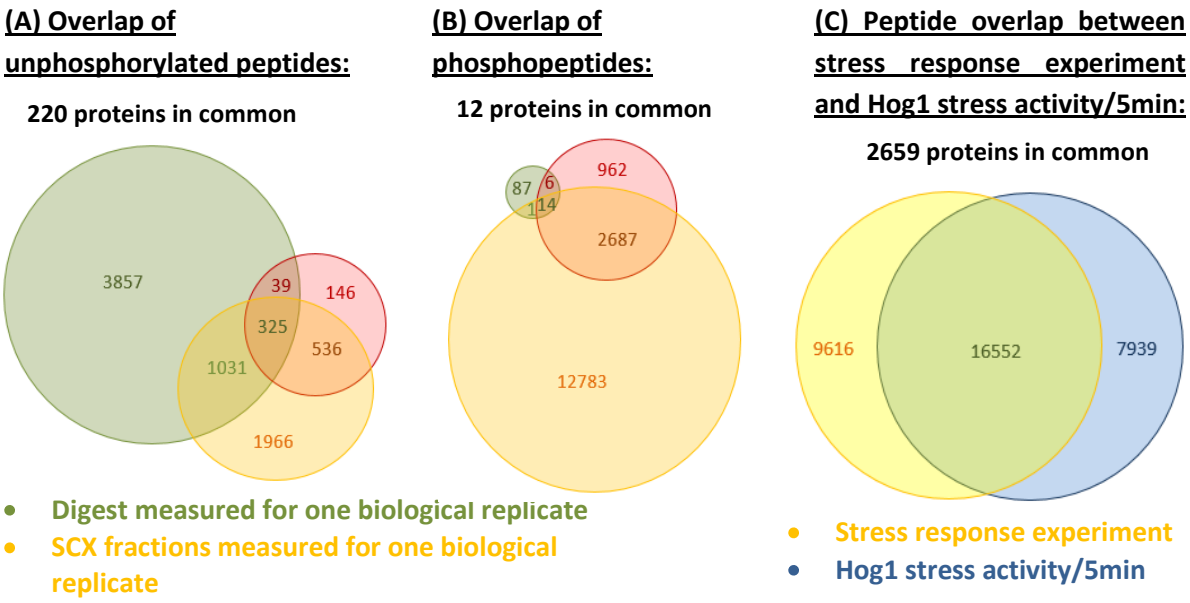


Figure 19. *Venn diagrams presenting unphosphorylated peptides(A) and phosphopeptides (B) for the stress response experiment. The green set represents the non-enriched sample, the red set the TiO_2-enriched samples and the yellow set the SCX-fractons. Graph (C) shows the peptide overlap between the stress response experiment and Hog1 stress activity/5min.*

R20). Approximately 13% of the peptides identified by Soufi *et al.* were also identified by us; however, most strikingly, the group of peptides not covered by Soufi *et al.* increased substantially.

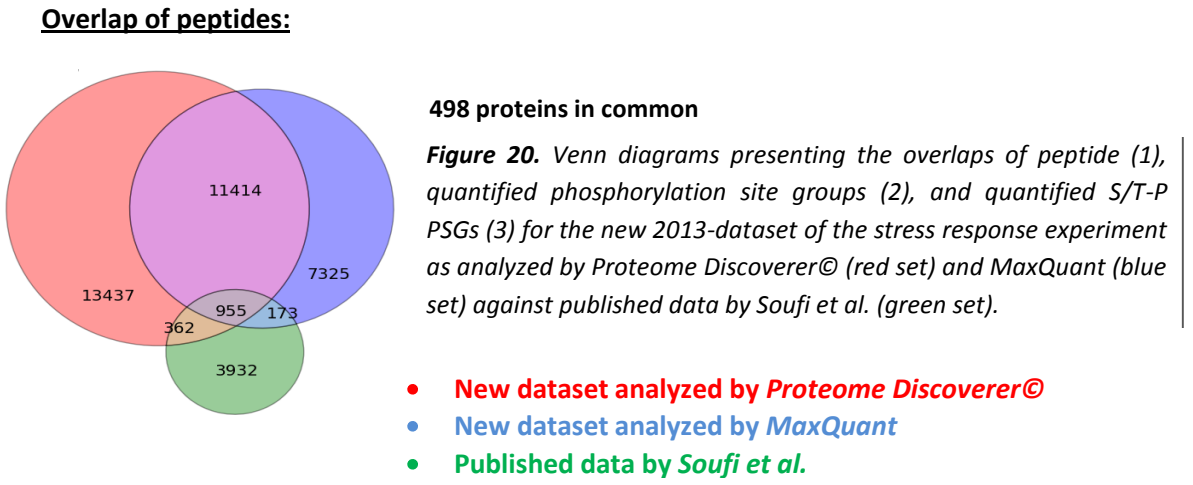


Figure 20. *Venn diagrams presenting the overlaps of peptide (1), quantified phosphorylation site groups (2), and quantified S/T-P PSGs (3) for the new 2013-dataset of the stress response experiment as analyzed by Proteome Discoverer© (red set) and MaxQuant (blue set) against published data by Soufi et al. (green set).*

III.3. New Dataset Provides a Range of Potential Hog1 Kinase Substrates

As it has been previously described in the *Project Outline*, we were interested in searching for a set of phosphorylation site groups that become up-regulated in response to high

osmolarity and down-regulated if the Hog1 kinase is specifically inhibited. We found 229 PSGs to fit this pattern with 40 of them being phosphorylated at S/T-P motifs. These 40 PSGs make up a total of 36 proteins that might be potential target candidates of the Hog1 kinase (see *Table V)*. Among them are representative Hog1 targets that have been previously described in the literature, e.g. the transcription factors Sko1 and Hot1, as well as the glycerol channel regulator Rgc2/Ask10 and the protein kinase Rck2 [6].

III.3.1. A Number of Positive Markers ensures Screen Validity.

To verify the experimental conditions that were chosen for our screen, we first controlled whether the activation loop of Hog1 becomes modified in response to osmotic stress at the residues *T174* and *Y176*. We found the double-phosphorylated variant to increase in abundance more than 40 times after five minutes exposure to hyperosmotic conditions (see *Figure 21*). Note that this activation was not observed in the similar dataset published by Soufi *et al.* [89]. The double-phosphorylated peptide was also detected in the experiments *Hog1 stress activity/ 5min* and *Hog1 stress activity/10min* with L/H-ratios of about 1.5. As it has been described by Macia *et al.* [50], inhibition of the activity of Hog1 increases its phosphorylation in both the absence and presence of hyperosmotic stress. Adding the inhibitor to cells that were subsequently subjected to hyperosmotic stress prevents dephosphorylation of Hog1*as*. This behavior could be attributed to negative feedback of Hog1 on the signaling branch and on itself. Accordingly, the L/H-ratio of the Hog1-phosphgroup covering *T174/Y176* should be slightly reduced, as observed in our dataset (*Figure 21*).

There are also other members of the Hog1-MAPK signaling available that could serve as positive markers for the response to hyperosmolarity, such as the MAPKK Pbs2 and the membrane-bound anchor protein Opy2. The phosphorylations residing on the S/T-P sites *S248* and *S514 (*see *Introduction)* of Pbs2, for example, have been described as a good indication for the hyperosmotic stress response in yeast [56]. For *S514*, we found the corresponding phosphopeptide upregulated seven-fold in response to hyperosmotic stress and no abundance change in the experiment *Hog1 stress activity/5min*. The latter indicates that we accurately stressed both samples in the corresponding experiment. In concert with the described crosstalk-mechanism of Hog1 with other MAPK-pathways in *S.cerevisiae* [6,46], we also detected the double phosphorylation of the activation loop (*T183/Y185*) of

Markers for components of the Hog1 pathway or stress signaling

Decreased in abundance under specified conditions

Increased in abundance under specified conditions

Protein Markers	Site Markers
Hog1	T174/Y176
Hot1	S153
Kss1	T183/Y185
Msn2	S201
Msn4	T142
Nha1	T765,S768
Pbs2	S514
Rck2	S519
Smp1	S348,S357,T365,S376
Ste20	-
Ste50	S202, T244, S248, T341

			Discoverer: Average L/H ratio			
Protein name	Phospho group	SP/TP site	stress response experiment	Hog1 basal activity experiment	Hog1 stress activity/ 5min	Hog1 stress activity/ 10min
HOG1	174(T)		3.66	1.99	2.18	1.87
HOG1	176(Y)		1.65	1.82	0.81	0.86
HOG1	174(T),176(Y)		43.50		1.56	1.43
HOT1	136(S)		3.33			
HOT1	146(S)		0.43			
HOT1	150(T)				0.48	
HOT1	153(S)	S/T-P	3.88	0.75	0.39	
HOT1	155(S)			0.65	0.33	
KSS1	183(T)			1.41		
KSS1	185(Y)				1.32	
KSS1	183(T),185(Y)		4.47		1.34	
MSN2	201(S)			0.77	0.59	0.84
MSN2	288(S)		0.97	0.91	0.76	0.84
MSN2	304(S)		1.43	1.06	0.65	
MSN2	339(S)	S/T-P	2.73			
MSN2	633(S)		0.96	0.95	0.88	0.91
MSN4	138(S),142(T)	S/T-P			1.07	
MSN4	142(T)	S/T-P	27.46		0.81	1.00
MSN4	294(S)		0.34	1.05		
MSN4	319(S)	S/T-P	0.98	0.89	2.45	2.25
MSN4	479(T)	S/T-P	4.04		0.88	
MSN4	537(S)	S/T-P	3.86		2.64	
NHA1	568(S)		1.19	1.04	0.45	0.41
NHA1	765(T)	S/T-P	1.49	1.08	1.23	1.16
NHA1	767(T),768(S)		0.39	0.92		0.79
NHA1	768(S)				1.11	
PBS2	514(S)	S/T-P	7.90	1.23	1.16	1.31
RCK2	350(T)	S/T-P	5.48	0.67	0.20	
RCK2	379(T)	S/T-P	55.24		0.02	0.02
RCK2	515(S),520(S)	S/T-P	16.21			
RCK2	520(S)	S/T-P	23.36		0.04	0.17
SMP1	425(S)	S/T-P		0.50		
STE20	552(T),562(S)	S/T-P	0.68	0.80		0.94
STE20	546(T),562(S)	S/T-P	0.80	1.06	0.05	
STE20	492(S)	S/T-P	0.86	1.19	2.11	3.33
STE20	512(T),517(S)	S/T-P	1.49		1.01	
STE50	202(S)	S/T-P	1.84	0.95	0.90	0.90
STE50	241(T),244(T)	S/T-P		0.45		
STE50	244(T)	S/T-P	17.64		0.06	
STE50	244(T),248(S)	S/T-P		0.51	0.29	0.54
STE50	248(S)	S/T-P		0.50	0.75	
STE50	341(T)	S/T-P	23.63			

Figure 21. *Protein and site markers for the activation of the Hog1 pathway under osmotic stress conditions. Sites have been found in the Hog1 literature; due to space limits not all table entries for the Hog1-pathway-markers are illustrated. An increased L/H-ratio (>2) is displayed in green, whereas the decreased L/H-ratio is shown in red (<0.5). No color change indicates that no significant change has occurred in either direction. If no ratio is displayed in the field, the corresponding site has either not been identified in the experiment or identification has not passed filtering.*

the MAPK Kss1 to be upregulated under hyperosmotic stress conditions (L/H-ratio of about 4). This kinase is involved in the control of filamentous growth and the pheromone response. Interestingly, we found Ste50, an adaptor protein involved in mating response, filamentous growth and osmostress pathway to show Hog1-dependency at the S/T-P site *T244* (*Figure 21*). This is indicative of the Hog1-dependent negative feedback mechanism. Additional S/T-P sites of Ste50, *S248* and *T341*, however, showed no confident Hog1-dependency: *S248* displayed a significantly decreased abundance under inhibitor conditions, but was not covered in the *stress response experiment*. Phosphorylation at *T341*, on the contrary, was highly abundant in the stress response experiment, but was not quantified in any inhibitor-experiment.

Likewise, we also detected a number of established downstream Hog1-targets, for example, the Hog1-regulated transcriptional repressor Sko1, and the transcription factors Hot1. Additionally we observe stress-dependent phosphorylation of the transcription factors Smp1 and Msn2/Msn4. For Sko1 the residues *S108*, *T113* and *S126* were previously characterized as Hog1-dependent since they would eliminate any phosphorylation mobility shift when mutated [63,66]. We could detect the Sko1 peptide harboring double phosphorylation at *S108* and *T113*, PTIIS#PPILT#PGGSK in both the *stress response experiment* and *Hog1 stress activity/5min/10min* experiments (see *Figure 22*). Additionally, we identified *T215* as a potential Hog1 target phosphorylation site on Sko1 which has so far escaped any observation in *in vitro* kinase assays (see *Figure 22*).

			Discoverer: Average L/H ratio			
Protein name	Phospho group	SP/TP site	stress response experiment	Hog1 basal activity experiment	Hog1 stress activity/ 5min	Hog1 stress activity/ 10min
SKO1	7(S)		1.01			
SKO1	54(S)		2.99	0.87	0.69	0.28
SKO1	65(S)				0.24	
SKO1	65(S),71(S)		1.28			
SKO1	94(S),105(T),117(S)	S/T-P	7.14			
SKO1	94(S),113(T)	S/T-P		0.24		
SKO1	105(T)			0.50	9.67	
SKO1	108(S),113(T)	S/T-P	1.25		0.15	
SKO1	113(T),117(S)	S/T-P	0.96			
SKO1	117(S)				19.92	
SKO1	215(T)	S/T-P	15.40	0.97	0.02	0.02
SKO1	223(T)		21.52			
SKO1	325(S)	S/T-P	1.03	0.83	0.71	0.69
SKO1	399(S)		0.87	0.62	0.33	0.37
SKO1	558(S)		1.94	0.75	0.51	
SKO1	588(S)		5.19			
SKO1	591(S)			0.76		

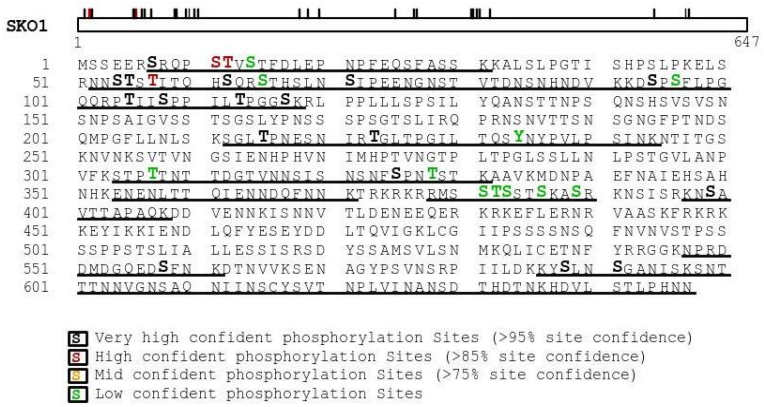


Figure 22. *Sko1 data screenshot from shotgun dataset and phosphorylation map of Sko1 as covered by the shotgun experiment. An increased L/H-ratio (>2) is displayed in green, whereas the decreased L/H-ratio is shown in red (<0.5). For example, for T215 that means that the phosphorylation at this residue is upregulated ~15-fold in the stress response experiment, whereas it is downregulated ~50-fold in the experiment testing Hog1 stress activity/5min.*

With Hot1 we observed the phosphorylation at S/T-P site *S153* is becoming three times more abundant when cells are exposed to hyperosmotic stress, but diminishing three-fold if Hog1 is specifically inhibited. As put forward by Alepuz *et al.* [64], Hog1-mediated phosphorylation of Hot1 does not constitute the determining mechanism for transcriptional activation of Hot1 target genes. In agreement with previous studies, our data also shows that *S153* is indeed phosphorylated by Hog1 under stress conditions (neutral L/H-ratios if no stress is applied to inhibitor-treated culture). All these findings exemplify that not all Hog1-mediated phosphorylation events are physiologically relevant. This example illustrates the necessity to check for phenotypes by mutation of corresponding sites to alanine.

Remarkably, the transcription factor Smp1 displayed the only phosphorylated S/T-P site (*S425*) decreasing in abundance in the Hog1 basal activity experiment (see *Figure 21*); *S425* is not detected under stress-conditions however.

With the stress-responsive transcription factor Msn2 we only observe one S/T-P motif *S339* to be significantly stress-dependent (see *Figure 21*), which has not been described as such so far. However, it is difficult to assess whether this is due to Hog1 kinase activity since confirming data from the hog1*as* experiments is missing. We also looked at the known PKA sites of Msn2 in order to determine whether a potential PKA interference might occur during osmostress conditions; the relevant sites for the regulation of Msn2 localization *S288*, *S304* and *S633* have been covered in our dataset. None of the phosphorylation at these sites changes in abundance during hyperosmotic stress and do not exhibit any inhibitor-sensitive regulation pattern which fits to previously described data [113]. Msn4, the paralogue of Msn2, shows a rather interesting regulatory pattern: For example, the S/T-P site *T142* has been found to become nearly 30-fold more abundant in response to hyperosmotic stress, but does not show any regulation in the hog1*as* experiment (see *Supplementary Table III*).

The glycerol channel regulator proteins Rgc1 and Rgc2/Ask10 were recently identified as target proteins of the Hog1-pathway [11]. Here we again identify Rgc2/Ask10 as potential Hog1 targets, with three S/T-P PSGs (*T808*, *S944/S947* and *T1017/S1021*) becoming increasingly abundant during hyperosmotic stress conditions and less abundant if Hog1 is inhibited. A recent publication on the phosphoregulation of Rgc2 [11] did not include any quantitative information on the phosphorylated residue *T808*; only the unphosphorylated counterpart was described to become upregulated under hyperosmotic stress. Here we provide the missing evidence that *T808* is indeed a potential Hog1 target site (see *Figure 23*),

			Discoverer: Average L/H ratio			
Protein name	Phospho group	SP/TP site	stress response experiment	Hog1 basal activity experiment	Hog1 stress activity/ 5min	Hog1 stress activity/ 10min
RGC1	537(S)	S/T-P		1.03		
RGC1	774(T)	S/T-P	5.42	0.91	0.32	
RGC1	817(T)	S/T-P	1.63	0.29	0.24	0.18
RGC1	966(S),969(S)	S/T-P		0.72		
RGC1	966(S),975(S)	S/T-P	0.78	0.50		
RGC1	969(S)	S/T-P	0.97			
RGC1	975(S)	S/T-P	9.67		0.07	0.08
RGC2	808(T)	S/T-P	2.62	1.06	0.14	0.16
RGC2	944(S),947(S)	S/T-P	2.97			
RGC2	944(S),948(S)	S/T-P			0.11	
RGC2	948(S)	S/T-P	0.66		0.17	0.33
RGC2	1017(T),1020(S)	S/T-P			1.02	
RGC2	1017(T),1021(S)	S/T-P		0.58	1.07	1.19
RGC2	1020(S)	S/T-P	0.29			
RGC2	1020(S),1021(S)	S/T-P	2.93			
RGC2	1021(S)	S/T-P	0.16	1.03	1.35	

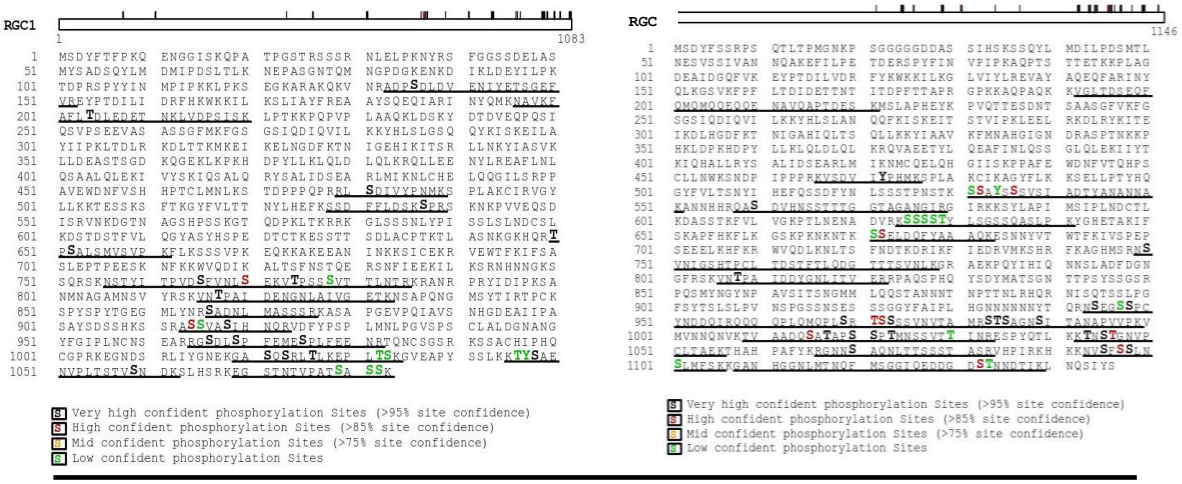


Figure R23. *Rgc1 and Rgc2 data is provided by the shotgun experiments. In both proteins phosphorylations are predominantly found in the C-terminal part of the protein.*

as it was found upregulated three fold in response to hyperosmotic stress and downregulated three-fold if Hog1 is inhibited.

In total, we can draw the following two conclusions: First, samples were harvested under hyperosmotic stress conditions; and second, the Hog1-pathway was efficiently inhibited. We thereby ensured experimental quality.

III.3.2. Phosphorylation Site Groups display various SILAC ratio combinations.

Next we define the criteria for the identification of novel putative Hog1 targets. Our concept is illustrated in *Figure 24*. The SILAC ratios of the phosphorylation site groups quantified in *the stress response experiment*, as well as in the experiments testing *Hog1 stress activity/5min* and *10min*, respectively, are plotted against each other. The data points widely disseminate along the x-axis (*stress response experiment*) from -6 to 6 (log2-scale), but do not significantly spread along the y-axis (*Hog1 stress activity/5min (10min)*). Most data-

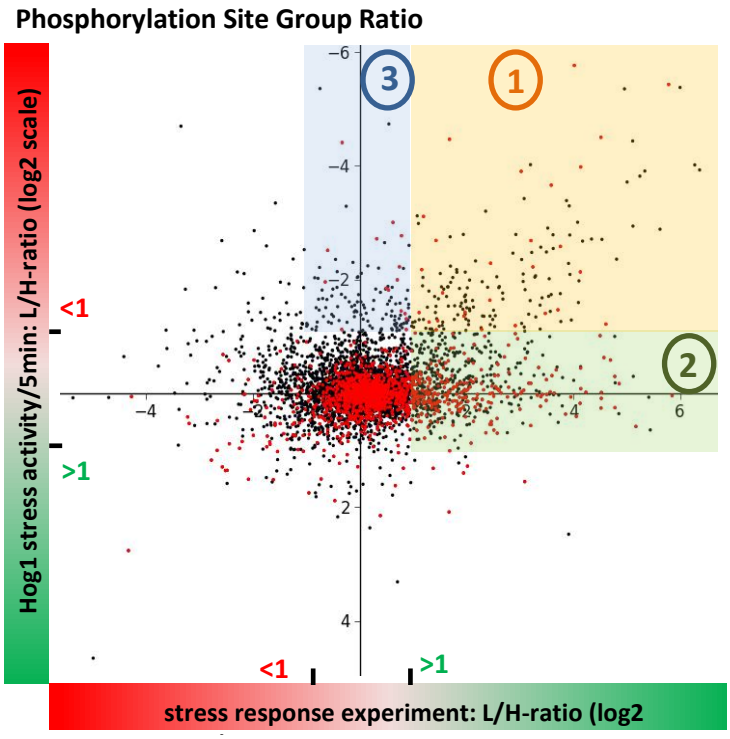


Figure 24. *Scatterplot displaying dataset from the stress response experiment (combined ratios from both biological replicates) and the dataset 'Hog1 stress activity/5min'. Black dots represent all data-points (unique quantified PSGs) found in both experiments; the red dots represent the S/T-P-containing PSGs. The scatterplot is divided into four main fields according to the SILAC ratio combinations: Field 1, Field 2 and Field 3.*

- ***5392 unique quantified PSGs both experiments have in common***
- ***1566 unique quantified PSGs containing S/T-P sites both experiments have in common***

points are shifted to the right side of the x-axis, and are hence upregulated in response to hyperosmotic stress. This overall distribution is not unexpected since hyperosmotic stress provokes a wide range of responses, dependent and independent of Hog1. The experiment testing *Hog1 stress activity/5min (10min)*, however, is confined to the inhibition of a specific pathway and is expected to induce fewer changes than observed with the *stress response experiment*.

We next defined the L/H-ratio of 2 as a conservative cutoff. Ratios above 2 are considered to be significantly up-regulated. Accordingly, a cutoff of 0.5 is defined for significant down-regulation. Setting up these ratio cutoffs results in the definition of distinct data fields that can be characterized individually. In this report I will focus on three fields designated in *Figure 24*.

Field 1: This dataset includes 185 phosphorylation site groups (corresponding to 146 proteins in total) that show an increased abundance under elevated hyperosmotic conditions and a decreased abundance if Hog1 is inhibited in stressed cells. Hence, the data field encompasses potential direct (SP/TP-sites) Hog1 targets and sites that are indirectly regulated by the MAPK (all other motifs found in this group). Remarkably, we found only 33 S/T-P phosphorylation site groups (30 proteins) to be represented in this data field, suggesting high specificity for this experimental set-up. Including the experiment *Hog1 stress*

activity /10min expands the group to 40 phosphorylation site groups and 36 proteins in total.

Since the data field contains such a small number of S/T-P-PSGs, we could not perform any statistical analysis to assess protein-enrichment in pre-defined GO-terms (see *Materials & Methods*). We observed four endocytosis-related proteins (see *Figure 25*), such as Akl1, a serine-threonine kinase involved in endocytosis and actin cytoskeleton organization. This protein kinase has been linked to the actions of Ede1, which, together with Pan1 and Sla1, constitutes a key endocytic protein factor [7].

Not surprisingly, proteins highly specific for the osmotic stress response in yeast were also found, such as the transcription factors Hot1, Sko1, the scaffold protein Ste50, the kinase Rck2 and the glycerol channel regulators Rgc1/2. Proteins that fall into the functional GO-category "protein phosphorylation" were also detected in **Field 1** with kinases such as the MAPKKKs Bck1/2. We also observe theses kinases to contain residues that become increasingly phosphorylated upon hyperosmotic stress in a Hog1-independent manner as well. With the PSGs *T765* of Bck1 and *S258* of Bck2 showing Hog1-dependency, these sites should be validated to provide evidence for a potential involvement of Hog1 in the control of the PKC pathway. Strikingly, we found two phosphatase systems to be putative target points for Hog1, one being the PP1-system (type 1 serine/threonine protein phosphatase with Glc7 as the catalytic subunit) involving regulatory proteins Reg1 and Gip3, the other one being the PPZ1/2-system (serine/threonine protein phosphatase Z) that is involved in the regulation of potassium transport.

Screening the experiment *Hog1 stress activity/10min* for phosphorylation site groups that significantly decrease in abundance (where they usually show induction by hyperosmotic

GO-term distribution of proteins in Field1 phosphorylated at an S/T-P motif

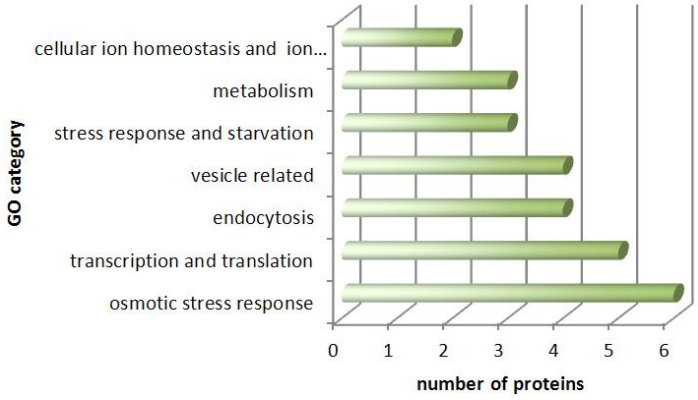


Figure 25. *Functional grouping of S/T-P containing phosphor-proteins discovered in Field 1 according to pre-defined GO-terms. Proteins are predominantly associated with the categories "osmotic stress response" and "transcription and translation".*

stress) revealed seven additional putative Hog1 targets. Four of these do not have any SGD (*Saccharomyces Genome Database*)-ascribed function yet, such as YMR124W, YML081W, YLR257W and Aim21, which seems to play a role in cytoskeletal organization. The remaining three potential Hog1 substrates are the ubiquitin-binding component of the E3-ubiquitin ligase complex Bul1, the transcriptional repressor and activator Sfl1, and the subunit of the SAGA transcriptional regulatory complex Spt20. *Table V* presents the finalized version of the list of potential Hog1 targets, including their ratios, the number of peptide variants detected, and the number of quantified PSMs throughout all experiments. For a more detailed view including all intensities, scans and q-values assigned to the different phosphorylation site groups, the report also provides a *Supplementary Table IV*.

Notably, the list includes protein substrates that might be regulated by the MAPK Hog1 in a direct or indirect manner. From our MS-data it is not clear whether Hog1 influences a kinase or phosphatase system that might induce changes in downstream phosphorylation patterns as well. In fact, when analyzing the GO-terms for all misregulated proteins throughout the entire dataset (432 in total), we observed a significant enrichment in the functional class "protein phosphorylation" (p-value<0.01), as well as in "pseudohyphal growth" (p-value<0.05), suggesting wide-spread effects the Hog1-pathway activation has in the cell in response to hyperosmotic stress (see *Figure 26*). We assume that the 80 PSGs observed in **Field 1** that are not phosphorylated at S/T-P motifs are such indirectly controlled target sites (see *Supplementary Table V*). Notably, a number of kinases fall into that range, for example the protein kinase involved in the response to DNA damage, Alk1. Phosphorylation at its site *S212* becomes 50-times more abundant in response to hyperosmotic stress and 50-times less abundant in the experiment *Hog1 stress activity/10min*. We also observe such changes

GO-term distribution of all (432) misregulated protein throughout entire MS-dataset

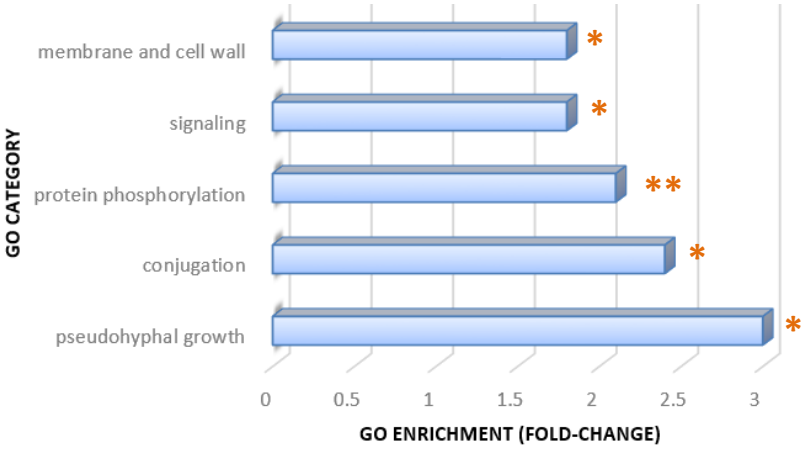


Figure 26. *Functional GO-terms for all proteins containing at least one misregulated phosphorylation site group. The GO-categories have been pre-defined and modified in our research group. Enrichment occurs if the p-value as calculated with Fisher's exact test reaches a value below 0.05 (one star: 0.01<p-value<0.05, two stars: p<0.01)*

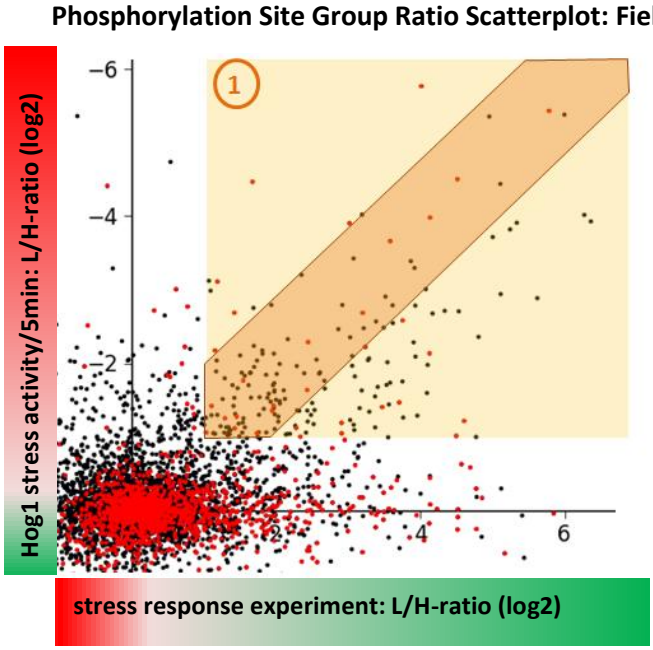


Figure 27. *Scatterplot displaying dataset from the stress response experiment and the dataset 'Hog1 stress activity/5min'. Black dots represent all data-points (unique quantified PSGs) found in both experiments; the red dots represent the S/T-P-containing PSGs. Field 1 is displayed with the darker shade designating the range where we assume the S/T-P containing PSGs to be directly controlled by Hog1 (confident set).*

- ***unique quantified PSGs both experiments have in common***
- ***unique quantified PSGs containing S/T-P sites both experiments have in common***

in phosphorylation abundance for Psk1, which is an S/T-protein kinase involved in sugar flux control. It might be interesting to follow up a potential link between these indirectly regulated targets and Hog1 activity.

In regard to S/T-P-phosphorylations, we venture that individual SILAC ratios obtained from the two different experiments might provide a hint as to whether a site is more likely to be directly regulated by Hog1 only or not. In *Figure 26* we designate part of **Field 1** where L/H-ratios of the PSGs behave linearly within a certain deviation range (+/-1 (log2-scale)). *T350* of the protein kinase Rck2, for instance, falls into this range as it is upregulated ~5 times in the *stress response experiment*, and displays the same-fold downregulation in the experiment testing *Hog1 stress activity/5min*. *S258* of Bck2, on the other hand, does not fit in this linear range since it is strongly upregulated in response to hyperosmotic stress (L/H-ratio ~ 22.5), but downregulated two-fold if Hog1 is inhibited. Assuming no quantification error, Bck2 is probably an integration node for several kinase inputs during hyperosmotic stress conditions. According to these criteria, we designated PSGs that are more likely to be controlled directly by the MAPK Hog1, in *Table V* for illustration purposes. In total, 13 PSGs fall into the confident linear range that we arbitrarily defined.

Different ratio distributions for individual candidate PSGs render interpretation rather difficult. It is therefore important to keep possible validation strategies in mind even at the level of data evaluation. Thus, before experimentally validating putative Hog1 targets as described in the *Project Outline*, we could assess the credibility of the SILAC ratios

Set of Confident Hog1 Targets:

				Discoverer: Average L/H ratio							
Protein name	Phospho group	PhosphoRS Confidence	SP/TP site	stress response experiment	# of peptide variants	# of PSMS quantified	Hog1 basal activity experiment	Hog1 stress activity/ 5min	Hog1 stress activity/ 10min	# of peptide variants	# of PSMS quantified
AIM21	476(S)	high	S/T-P	93.37	1	1			0.03	1	3
AKL1	971(S)	confident	S/T-P	2.76	1	2		0.41		1	1
ARE2	175(S)	low	S/T-P	12.03	1	1	0.50	0.08	0.07	2	3
ART5	158(S)	very high	S/T-P	13.73	2	3	1.21	0.17	0.17	2	4
ASK10	808(T)	very high	S/T-P	2.62	1	8	1.06	0.14	0.16	1	11
BCK1	765(T)	very high	S/T-P	7.54	1	3		0.44		1	2
BCK2	258(S)	confident	S/T-P	22.67	1	1		0.50		1	1
BOI1	647(T)	high	S/T-P	7.55	1	1	1.08	0.48	0.51	1	1
BUL1	107(T),111(T)	very high	S/T-P	25.96	1	1		0.67	0.37	1	1
CHS5	373(T)	very high	S/T-P	11.70	1	3	0.95	0.36	0.25	1	4
ENT3	320(T)	very high	S/T-P	11.14	1	1		0.38		1	1
ERG11	458(S)	very high	S/T-P	4.88	1	2	0.95	0.46	0.40	1	2
GAL11	750(T)	high	S/T-P	5.08	1	1		0.47	0.65	1	1
GET2	45(S)	very high	S/T-P	17.56	1	1		0.23	0.19	1	1
GIP3	260(S),264(T)	very high	S/T-P	2.06	1	1		0.47		1	2
HAL5	63(S),66(S)	low	S/T-P	2.29	1	1		0.12		1	1
HOT1	153(S)	high	S/T-P	3.88	1	1	0.75	0.39		2	3
PPZ2	310(S)	very high	S/T-P	3.35	1	2	1.26	0.48		1	1
RCK2	350(T)	very high	S/T-P	5.48	1	2	0.67	0.20		1	2
RCK2	379(T)	very high	S/T-P	55.24	1	2		0.02	0.02	1	3
RCK2	520(S)	very high	S/T-P	23.36	2	4		0.04	0.17	2	5
REG1	898(S)	very high	S/T-P	3.18	1	3		0.05		1	3
RGC1	774(T)	very high	S/T-P	5.42	1	1	0.91	0.32		1	4
RGC1	975(S)	very high	S/T-P	9.67	4	5		0.07	0.08	2	3
ROD1	602(S),608(S)	very high	S/T-P	2.17	1	1		0.37		1	1
SFL1	556(S)	very high	S/T-P	7.43	2	3		1.00	0.35	1	1
SKG6	632(S)	very high	S/T-P	2.47	1	1	0.85	0.42		1	1
SKO1	215(T)	very high	S/T-P	15.40	1	3	0.97	0.02	0.02	1	2
SPT20	516(T)	very high	S/T-P	7.07	2	3	0.88	0.56	0.45	1	2
SSK2	54(S),57(S)	low	S/T-P	2.24	1	1		0.22		1	1
STE50	244(T)	very high	S/T-P	17.64	1	1		0.06		1	1
TIF4632	196(T)	very high	S/T-P	2.34	1	4	0.38	0.29		1	2
TOD6	152(T)	very high	S/T-P	24.46	1	1		0.42		1	4
VAS1	294(S)	very high	S/T-P	2.74	1	3		0.46	0.40	1	1
VPS53	790(S)	high	S/T-P	9.24	1	1		0.15	0.17	2	5
VPS9	375(S)	very high	S/T-P	3.95	1	3	0.80	0.37	0.26	1	2
YLR257W	7(S)	high	S/T-P	2.53	1	1	0.95		0.50	1	1
YLR257W	8(T)	high	S/T-P	2.49	1	3	0.81	0.61	0.50	1	3
YML081W	603(S)	very high	S/T-P	7.78	1	1	0.83	0.51	0.50	1	1
YMR124W	209(S)	high	S/T-P	9.49	1	1		0.21		1	1

PSGs falling into the linear range | **Increased in abundance under specified conditions** | **Decreased in abundance under specified conditions**

Table V. *List of potential Hog1 substrates as extracted from all shotgun experiments "Hog1 stress activity" and overlapped with the data from the stress response experiment. The graph displays the number of peptide variants for the corresponding PSG, and the number of PSMs quantified. The phosphoRS-confidence level informs on the level of confidence assigned to the phosphorylation site(s) (median confidence taken for multiple entries).*

considering parameters, such as confidence of identification, phosphorylation assignment confidence, and intensities.

Field 2: This data field includes 600 phosphorylation site groups (334 proteins in total), 225 of them containing S/T-P motifs, that become increased in abundance in response to hyperosmotic stress but are not influenced by the inhibition of the MAPK Hog1. Phosphorylated S/T-P sites that fall into this category are phosphorylated by stress-dependent kinases other than Hog1. Interestingly, we found a large number of kinases in this field that show an increase of phosphorylation during hyperosmotic conditions, such as for example PKC-signaling related kinases Bck1/2, the glucose-repressible kinase Rim15 and the upstream serine/threonine kinase for the Snf1 complex, Sak1. The list also includes the kinase Ptk2 pivotal for re-balancing the ion transport across the plasma membrane. In coherence with the fact that Pbs2 is not influenced by the active MAPK Hog1 [6], we also found Pbs2 to be represented in **Field 2**. Finally, the adenylate cyclase Cyr1, which is required for cAMP production and cAMP-dependent protein kinase signaling (PKA) becomes also phosphorylated under osmostress-conditions, as well as Sok2, a nuclear protein that plays a regulatory role in the same pathway. Hyperosmotic stress conditions therefore also activate different signaling pathways, considerably complicating the search for protein substrates of Hog1 only. According to Reiter *et al.* [79] a number of Hog1-independent S/T-P sites become phosphorylated under stress conditions, which is coherent with the large number of S/T-P phosphosites detected by Soufi *et al.* [89].Our data supports this view as it includes kinases that target S/T-P motifs when activated. For instance, we found the MAPK Kss1 that specifically targets S/T-P sites to become activated in the *stress response experiment* via phosphorylation of the *TY*-motif (see *Figure 21*), confirming previous observations.

Along the lines presented by Reiter *et al.* [79], several S/T-P motifs on Pan1 become

			Discoverer: Average L/H ratio			
Protein name	Phospho group	SP/TP site	stress response experiment	Hog1 basal activity experiment	Hog1 stress activity/ 5min	Hog1 stress activity/ 10min
PAN1	748(S)	S/T-P	2.51		1.10	1.03
PAN1	993(T)	S/T-P	23.04		1.06	
PAN1	1003(S)	S/T-P	2.10	0.91	0.97	0.88
PAN1	1321(T)	S/T-P	16.66		1.32	

Figure 28. *Pan1-phosphorylation site groups at S/T-P motifs covered in Field 2 of the entire shotgun-experiment.*

phosphorylated in response to hyperosmotic stress in a Hog1-independent way (see *Figure 28).* We found *S748*, *T993* (*S995*) and *S1003* in **Field 2**. The only Hog1-dependent site of Pan1, namely *T1225*, was not included in this field. Even without using a target validation approach, **Field 2** allows for distinguishing kinase dependencies of individual phosphorylation sites if the same protein occurs in other data fields as well.

			Discoverer: Average L/H ratio			
Protein name	Phospho group	SP/TP site	stress response experiment	Hog1 basal activity experiment	Hog1 stress activity/ 5min	Hog1 stress activity/ 10min
SSD1	231(S)	S/T-P	1.63	0.92	0.21	0.28
GAL11	783(S),793(T)	S/T-P	1.41	0.76	0.28	
STE20	546(T),562(S)	S/T-P	0.80	1.06	0.05	
UPC2	122(T)	S/T-P	1.58	0.52	0.28	0.32
DNA2	13(S),17(S)	S/T-P	0.64		0.26	0.35
CRP1	182(T)	S/T-P	1.63	0.40	0.39	0.32
SWE1	294(S)	S/T-P	1.72	1.21	0.37	
SKO1	108(S),113(T)	S/T-P	1.25		0.15	
ASK10	948(S)	S/T-P	0.66		0.17	0.33
MSB1	863(S)	S/T-P	1.83		0.40	
YNL050C	107(T)	S/T-P	1.72		0.15	
VPS5	252(S)	S/T-P	1.83		0.46	
SSK2	74(S)	S/T-P	1.54		0.12	
RTG3	227(S)	S/T-P	0.79	0.94	0.46	0.70
RGC1	817(T)	S/T-P	1.63	0.29	0.24	0.18

Figure 29. *Phosphorylated S/T-P motifs as observed in Field 3 that seem to be Hog1-dependent under non-stressed conditions.*

Field 3: 143 phosphorylation site groups (15 S/T-P sites and a total of 115 proteins) are assigned to this data field. Under hyperosmotic stress conditions these sites are not upregulated; however, if Hog1 is inhibited, they show Hog1-dependency. This regulatory pattern could be generated by three possible scenarios: (1) PSGs are phosphorylated by the MAPK under non-stressed and stressed conditions in the same way, which might be an explanation for recurring Hog1 target proteins that fall into this field (Rgc1, Rgc2/Ask10, Sko1, and Ste20, see *Figure 29*). (2) Phosphorylations are mediated by Hog1-dependent but stress-independent kinases, and (3) the data might include quantification errors. To rule out the third possibility, we compared **Field 3** with the data obtained from the experiment *Hog1 basal activity*. Only the glycerol channel regulator Rgc1 and the DNA-binding protein Crp1 showed down-regulation under non-stressed conditions as well if Hog1 was specifically inhibited (see *Figure 29*). Those proteins that do not show an overlap with the experiment *Hog1 basal activity* should be analyzed more closely to exclude the possibility of SILAC quantification errors.

III.3.3. Potential Hog1 Targets Missed.

The definition of fields as described above necessitates that the corresponding PSGs were quantified in both the *stress response experiment* and the experiment testing *Hog1 stress activity/5min (10min)*. We observe 44 S/T-P containing PSGs (39 proteins) however, that show significant down-regulation if Hog1 is inhibited, but were not quantified in the *stress*

response experiment (see *Table VI*). Due to this lack of this information, these sites are not included in **Field 1**. In terms of gene ontology, we see most of the proteins falling into the category of *stress response and starvation* and *transcription/ translation* (see *Figure 30*). Affected proteins are, for instance, the isoform of Ppz2, Ppz1, which displays a three-fold down-regulation in both experiments testing *Hog1 stress activity/5min (10min)*, respectively. Also, the nucleoporin protein Nup2, which is involved in nucleo-cytoplasmic shuttling and chromatin organization, is present in the list of these potential Hog1 targets. Notably, we also detected Erb2, a constituent of the pre-ribosomal particles required for maturation of the large ribosomal subunit, to be downregulated 100-fold in the experiment testing *Hog1 stress activity/5min*. Due to the strict identification filtering, we might hence miss out on a number of interesting targets (32 additional proteins to the existing Hog1 list).

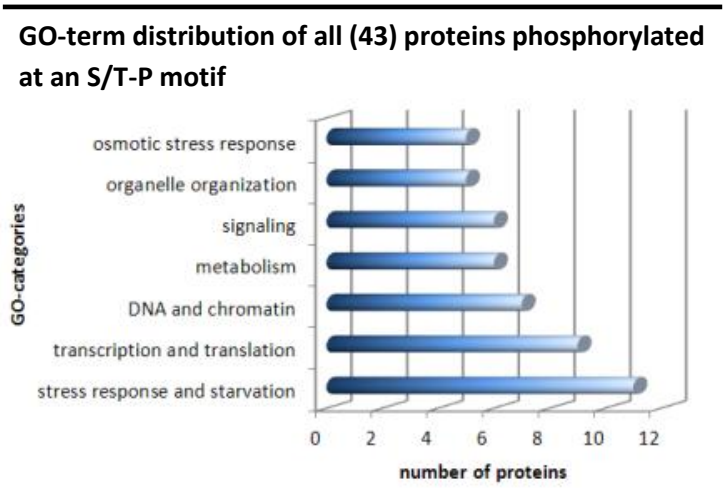


Figure 30. *Functional GO-terms for S/T-P phosphoproteins that are down-regulated in the experiment testing Hog1 stress activity/5min (10min), but have not been detected in the stress response experiment.*

Notably, we also observed the opposite case where S/T-P sites were up-regulated in response to hyperosmotic stress, but not confidently covered in the complementary *Hog1 stress activity* dataset. Tsl1, which displayed strongly regulated S/T-P sites in the original dataset, was partially affected. *S77, S147* and *S161* showed increased phosphorylation in response to hyperosmotic stress, but only *S147* has been detected in the experiment testing *Hog1 stress activity*. Thus, the lack of required experimental data makes it difficult to rule it out as a potential Hog1 substrate or vice versa. We therefore decided to focus our attention on that protein and to investigate its potential involvement in the Hog1-dependent osmotic stress response. To do so, we applied a targeted validation approach first to complete the phosphorylation map of the protein.

			Discoverer: Average L/H ratio		
Protein name	Phospho group	SP/TP site	Hog1 basal activity experiment	Hog1 stress activity/ 5min	Hog1 stress activity/ 10min
ARE2	176(S)	S/T-P	0.51	0.08	0.07
ASK10	944(S),948(S)	S/T-P		0.11	
BUD6	352(T),353(S)	S/T-P		0.41	1.15
ECM25	463(T),476(S)	S/T-P		0.21	
ECM25	463(T),477(S)	S/T-P		0.18	0.18
ERB1	713(S)	S/T-P		0.02	
ERG27	345(T)	S/T-P		1.45	0.35
FUN30	451(S),455(T)	S/T-P		0.35	
GAL11	398(T)	S/T-P		0.39	
HOS4	688(T),690(S)	S/T-P	0.19		
ISW2	1079(T)	S/T-P			0.43
ITR1	12(T),35(S)	S/T-P		0.41	
KIN1	973(S),976(T)	S/T-P		0.53	0.39
KIN1	973(S),979(S)	S/T-P	0.80	0.51	0.32
LCB4	454(S)	S/T-P	0.81	0.45	
MPS2	147(T),149(S)	S/T-P	0.20		
MRH1	289(S),299(S)	S/T-P	0.97	0.33	
MRL1	334(S),341(S)	S/T-P		0.45	
NOP4	657(T)	S/T-P	0.79	0.49	0.75
NUP2	361(T)	S/T-P		0.41	
PAM1	711(S)	S/T-P		0.44	
PMS1	110(S),116(S)	S/T-P	0.39		
PPZ1	49(S),67(T)	S/T-P		0.36	0.31
RIM15	1542(S),1557(S)	S/T-P		0.37	0.73
RLM1	374(S),381(S)	S/T-P		0.48	
ROD1	602(S),613(S)	S/T-P	0.79	0.40	0.27
SAP1	406(T)	S/T-P		0.24	
SEC27	829(T)	S/T-P		0.77	0.43
SKO1	94(S),113(T)	S/T-P	0.24		
SMP1	425(S)	S/T-P	0.50		
SOG2	214(T)	S/T-P		0.16	
SPA2	970(S),979(S)	S/T-P		0.54	0.44
STE20	448(T)	S/T-P		0.36	
STE20	546(T),552(T),562(S)	S/T-P			0.46
STE50	244(T),248(S)	S/T-P	0.51	0.29	0.54
STE50	241(T),244(T)	S/T-P	0.45		
SWA2	24(T)	S/T-P		0.49	
SWI6	178(S),182(T)	S/T-P		0.35	
SYP1	373(T),375(T)	S/T-P			0.17
TCO89	104(S),112(T),115(S)	S/T-P		0.49	
TIF4631	183(T)	S/T-P			0.48
TOD6	333(S),339(Y),341(S)	S/T-P		0.10	
YCL012C	122(T)	S/T-P	0.85	0.64	0.41

Table VI. *Potential Hog1 targets that are not covered in the stress response experiment. Entries that are highlighted are proteins that have not been observed in Field 1.*

III.4. Target Validation Experiments using the HTB-Tag

This report so far has presented the design and output of the discovery-based proteomics approach to identify novel targets of the MAPK Hog1. Notably, this kind of proteomics approach (shotgun proteomics) is limited by a number of factors, such as the protein interference problem [80,83] and random sampling, which usually result in decreased reproducibility. In contrast, the analysis of a specifically preselected protein- termed “target validation approach” [2] - allows the generation of more reliable, quantitative and sensitive data.

Once a protein phosphorylation site in the shotgun MS-dataset that seems to be regulated in a Hog1-dependent way is discovered, a potential target is defined. For further studies the completion of the entire phosphorylation map of the protein is of advantage. For these

purposes, we used an MS-targeted validation approach. We applied tandem affinity purification under denaturing conditions [79,85,86], thereby assuring that post-translational modifications are not altered during extract preparation and sample processing. The protein of interest was genetically tagged with an HTB_{eaq}-tag [79] which is a modified variant of the HB-tag [86], containing poly-histidine, a TEV-protease cleavage site, a biotinylation signal and the AQUA sequence, allowing subsequent absolute quantification in MS if desired [6]. We set up an SQL database to store all relevant experimental data for easier handling and analysis. Using a phosphorylation visualization script it is equally easy to construct a phosphorylation and protein coverage map for each individual protein under different conditions (script kindly provided by David Hollenstein). So far, the database includes Ede1, Bul1, Sla1, Tsl1, Msn2, Atg1, Atg13, Atg19, Atg9, Atg34, Rgc1, and Rgc2-data (see *Supplementary Databases: HTB-purifications*.

III.4.1. Purifying Tsl1 reveals its Hog1-dependency.

This approach was also used for Tsl1. We were highly intrigued by the high induction of the Tsl1 peptide under hyperosmotic conditions found in the original shotgun dataset, but could not verify any Hog1-dependency with the new MS-data. To analyze whether it is a true target of the MAPK or not, we followed up on phosphorylation pattern changes of the specifically HTB-purified Tsl1. We analyzed the samples in a Velos Orbitrap instrument; in a total of 12 experiments, 60% of the Tsl1 protein sequence (against 4% covered by the entire set of shotgun experiments) was covered and 19 phosphorylation sites were identified with

Phosphopeptides:		L/H normalized		**Unphosphorylated peptides:**	L/H normalized	
Sequence	**Protein Site**	**TSL1 wt +- NaCl**	**TSL1wt vs. TSL1hog1as + NaCl**	**Sequence**	**TSL1 wt +- NaCl**	**TSL1wt vs. TSL1hog1as + NaCl**
ALS#NNISQESLVAPAPEQGVPPAISR	S49	9.733	0.236	ALSNNISQESLVAPAPEQGVPPAISR	no ID	no ID
ALSNNIS#QESLVAPAPEQGVPPAISR	S53	n.q.	0.591			
ALS#NNIS#QESLVAPAPEQGVPPAISR	S49/56	n.q.	n.q			
ALS#NNIS#QESLVAPAPEQGVPPAISR	S53/S56	n.q.	0.450			
ALSNNISQES#LVAPAPEQGVPPAISR	S56	0.825	0.753			
SATRS#PSAFNR	S77	2.889	0.803	SATRSPSAFNR	no ID	no ID
FFS#PSSNIPTDR	S135	18.316	0.453	FFSPSSNIPTDR	medium(H/L:0.769)	0.807
FFSPS#S#NIPTDR	S137/S138	18.331	not found			
IAS#PIQHEHDSGSR	S147	4.594	0.335	IASPIQHEHDSGSR	0.487	1.001
SLLVHSLLNNT#SQTSLEGPNNHIVTPK	T191	n.q.	0.374	SLLVHSLLNNTSQTSLEGPNNHIVTPK	0.669	0.946
AGNRPT#SAATSLVNR	T215 or S216	8.998		AGNRPTSAATSLVNR	1.047	0.761
AGNRPT#SAATSLVNR	T215		0.387			
YSELADISSSETSSQHNESDPDDLTTAPDEEYVS#DLEM*DDAkQDYK	S303	0.802	0.769	YSELADISSSETSSQHNESDPDDLTTAPDEEYVSDLEM*DDAkQDYK	no ID	no ID
AGHAIVYGDATSTY#AK	Y1077	0.486	no ID	AGHAIVYGDATSTYAK	0.820	0.782

* oxidation

x manual quanification

n.q. no quantification (either heavy or light form not detected)

Table VII. *Table summarizing results of the Tsl1-HTB purification.*

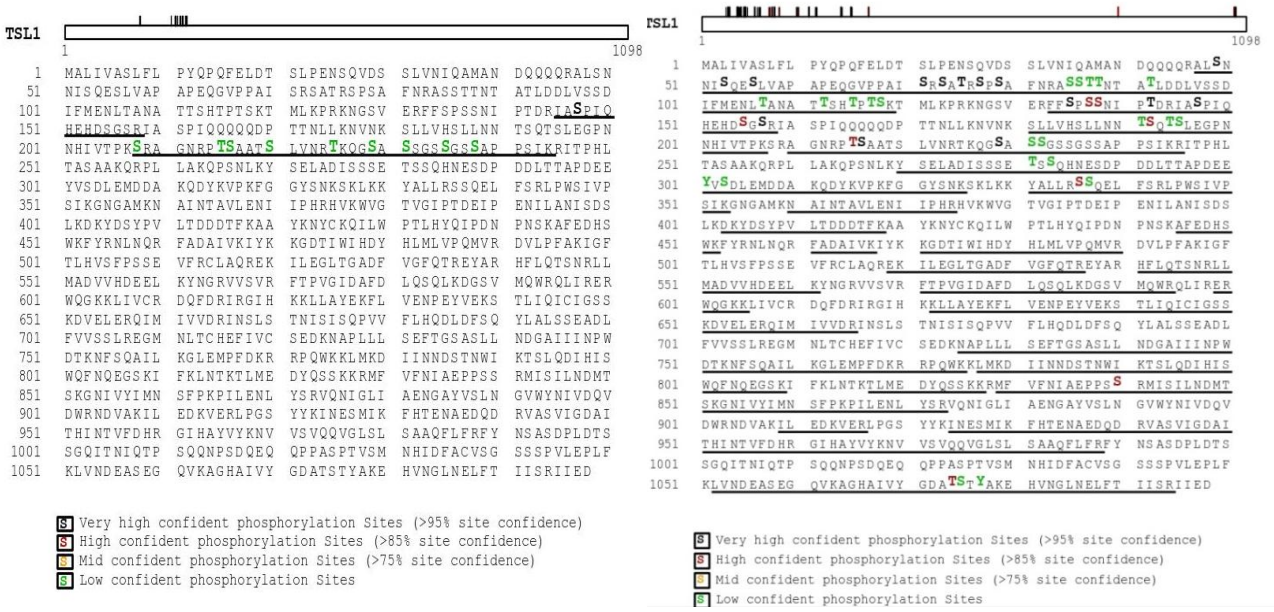


Figure 31. *Comparison of protein coverage map of Tsl1 as covered (A) in shotgun conditions (throughout all replicates of all experiments described in III.3, 4% coverage) and (B) in HTB-purifying conditions (60% coverage).*

high confidence (75%, see *Figure 31*). Based on database entries in PhosphoGrid (www.phosphogrid.org,14/06/11), 12 of the phosphorylation sites have not been described previously. Interestingly, we localized most phosphorylation sites in the N- terminal part of the protein.

Three of the identified phosphorylation sites lie within an S/T-P MAPK consensus motif, namely *S77*, *S135* and *S147*, two of which we have already discovered in our shotgun approach described above (*S77* and *S147*).

Next we made use of the SILAC technology to identify sites of Tsl1 that become altered in the phosphorylation status upon exposure to hyper-osmotic stress. We additionally performed SILAC experiments to identify phosphorylation sites that are susceptible to inactivation of the Hog1 kinase, either by applying an analogue sensitive allele of kinase in combination with a specific as-inhibitor or by using a *hog1Δ* strain. S/T-P sites that are found to become induced in response to hyperosmotic stress and are susceptible to Hog1 inhibitor treatment (or to deletion of the kinase) were considered putative Hog1 substrates. Our analysis identified five phosphorylation sites that become upregulated 2.5-fold after a 5-minute exposure to osmotic stress (see *Table R5*). These sites include the phosphoserines of the S/T –P MAPK consensus motifs *S77, S135* and *S147*. The two remaining up-regulated sites, namely *S49* and *T215*, do not fit to any known kinase motif. We also detected one site that becomes dephosphorylated in response to external osmolarity; *T1077*. Interestingly,

with the exception of *T1077*, all regulated sites were susceptible to inhibitor treatment. This strongly indicates that phosphorylation at these sites is dependent on Hog1, indicating that Hog1 regulates the phosphorylation pattern of Tsl1 in a direct as well as indirect manner. When looking at the corresponding unphosphorylated counter-peptides, we only measured an approximately two-fold downregulation for the peptide *IASPIQHEHDSGSR*. All other unphosphorylated counterpeptides were either not found to be regulated (*S49*) or not covered by our analysis (*T191*, *S215*, *T1077*). Similar observations were made in the inhibitor-treated samples. Therefore, we predict that only a very minor percentage of Tsl1 is phosphorylated at these sites before stress exposure, and also after osmo-stress induction indicating substoichiometric levels of phosphorylation.

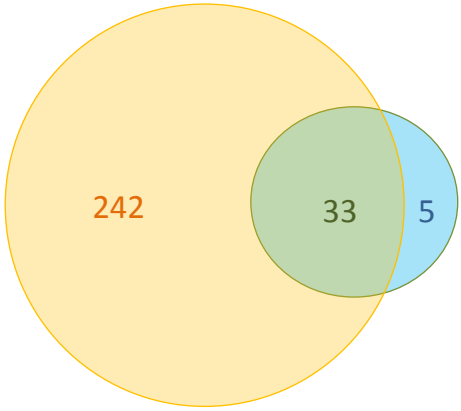
Results obtained with *hog1Δ* strain were less conclusive since we observed a general decrease in the abundance of all peptides independent of whether they are phosphorylated or unphosphorylated. We concluded that expression levels of Tsl1 are strongly dependent on the kinase. Taken together, our analysis confirmed that two S/T-P motifs of Tsl1, namely *S135* and *S147* are most likely directly regulated by the Hog1 MAPK. Thus, Tsl1 seems to be a Hog1-dependent target.

We next looked for a reasonable explanation as to why specific Tsl1-peptides could not be detected throughout all new shotgun-experiments (encompassing 8 replicates in total). For example, we could not detect the peptide harboring *S147* in the new *stress response experiment* and therefore retrieved corresponding SILAC ratios from the former MS-dataset generated in our research lab (see *III.1*). This is valid since the former *stress response* dataset passed our sample quality control. Ultimately, there are many possible answers for the missing detection, ranging from peptide detectability in general, to ion suppression. For now, we will focus on three major issues in this report: 1) Was Tsl1 not detected due to restricted peptide detectability in general? 2) Are Tsl1-peptides only detected if the background signal is severely reduced? 3) Is Tsl1 not found due to ion suppression and/or identification failure?

Regarding (1), we did not observe any tendency as to whether any specific region of the protein becomes more preferentially covered, which allows the following conclusions: First, Tsl1 does not seem to include extensive peptide stretches that are not accessible for tryptic digestion, and second, most peptides of Tsl1 have ionization properties that allow for compatibility with mass spectrometry analysis. Notably, this rules out the possibility that

Tsl1 sequence coverage remained low in the shotgun-experiments due to low ion detectability. Two peptide stretches (*M1-R46, F989-K1050*) limit the total coverage, however (see *Figure 31*): Due to the lack of arginine and lysine, the preferred cleavage sites of trypsin, the corresponding peptides are too large to be detected by high-resolution MS instrumentation. Sequence coverage could be complemented by using alternative peptidases, such as chymotrypsin, or by inserting artificial arginines and lysines into specific protein parts, provided that the protein function remains intact. In summary, the enhanced coverage of Tsl1 in the targeted validation approach suggests that Tsl1-peptides are detectable in an LC-MS/MS-setup.

With (2) we assume that the absence of Tsl1-identifications in the discovery-based proteomics approach is due to increased sample complexity. In the shotgun approach the background is loaded with a variety of proteins, whereas by purifying a protein, sample complexity is reduced and the intensities of noise signals are diminished. Hence, we assumed that these different background matrices may favor detectability of different peptides in general. To test this hypothesis, we analyzed our entire HTB-purification database to see whether we could detect some major differences in the types of peptides detected in one MS-approach versus the other for a given protein. *Figure 32* illustrates all peptides we detected from one protein only which we chose as a representative example. Note that in other experiments all peptides of a protein detected in the shotgun approach would usually be detected in the targeted validation approach as well. In *Figure 32* the five surplus peptides detected in the shotgun sample were missed-cleaved variants of the peptides covered by the targeted validation approach (four out of five were concerned). One peptide detected in the shotgun approach was the oxidized variant of the peptide detected in the other approach. Therefore, our analysis allows the conclusion that neither of the two



Overlap of Peptides detected in different experimental conditions:

- **Protein-specific peptides found in HTB-purified sample**
- **Protein-specific peptides found in shotgun sample**

Figure 32. *Comparison of peptides detected under shotgun conditions (yellow set) and HTB-purifying conditions (blue set).*

experimental approaches preferentially enrich for distinct peptide species. Hence, we rule out the possibility that Tsl1-peptides could not be found in the shotgun-approach due to decreased peptide detectability.

Finally, we concluded that Tsl1 might have not been detected due to lack of fragmentation (ion suppression) and/or non-qualitative peptide-assignment (3). I will further elaborate on that issue in the next chapter and on possible strategies to approach the general problem of peptide MS-identification.

III.5. MS-Data Evaluation- facing challenges

So far, I have primarily discussed the confidently assigned phosphorylation site groups and characterized Hog1-dependency based on distinct phosphorylation pattern changes under different conditions. However, there are two important issues that need to be considered for comprehensive MS-data validation: First, we might miss a great number of Hog1-regulated phosphorylation sites since the corresponding peptides have not been covered throughout all relevant shotgun experiments. There are many general explanations for the problem as to why peptides are not detected during an MS-run: 1) The peptide is not MS-amenable since it is either too long, not retained on the HPLC or singly/highly charged. We referred to this topic already as *peptide detectability*. 2) The peptide might carry a further modification that is disregarded in the search parameter; this can lead to wrong peptide-spectrum assignments. 3) There has been no qualitative identification, meaning that the spectrum did not pass FDR-filtering, or 4) the precursor has not been selected for fragmentation since the signal was too low. Finding a means to extract such low-abundant peptide-precursors from the MS-raw file might prove useful in extending the number of potentially regulated sites. From other MS-projects in our research group it has become increasingly clear that searching for specific peptides that have not been confidently identified by *Proteome Discoverer©* requires lower FDR-stringency and manual inspection of MS-spectra. Usually, we generate an extracted-ion-chromatogram (XIC) representing m/z-values (precursor mass) of interest for a chromatographic run [116]. Basically, the C-12-peak intensity within a mass tolerance and time window around a particular precursor mass is plotted at every retention time point where it occurs.

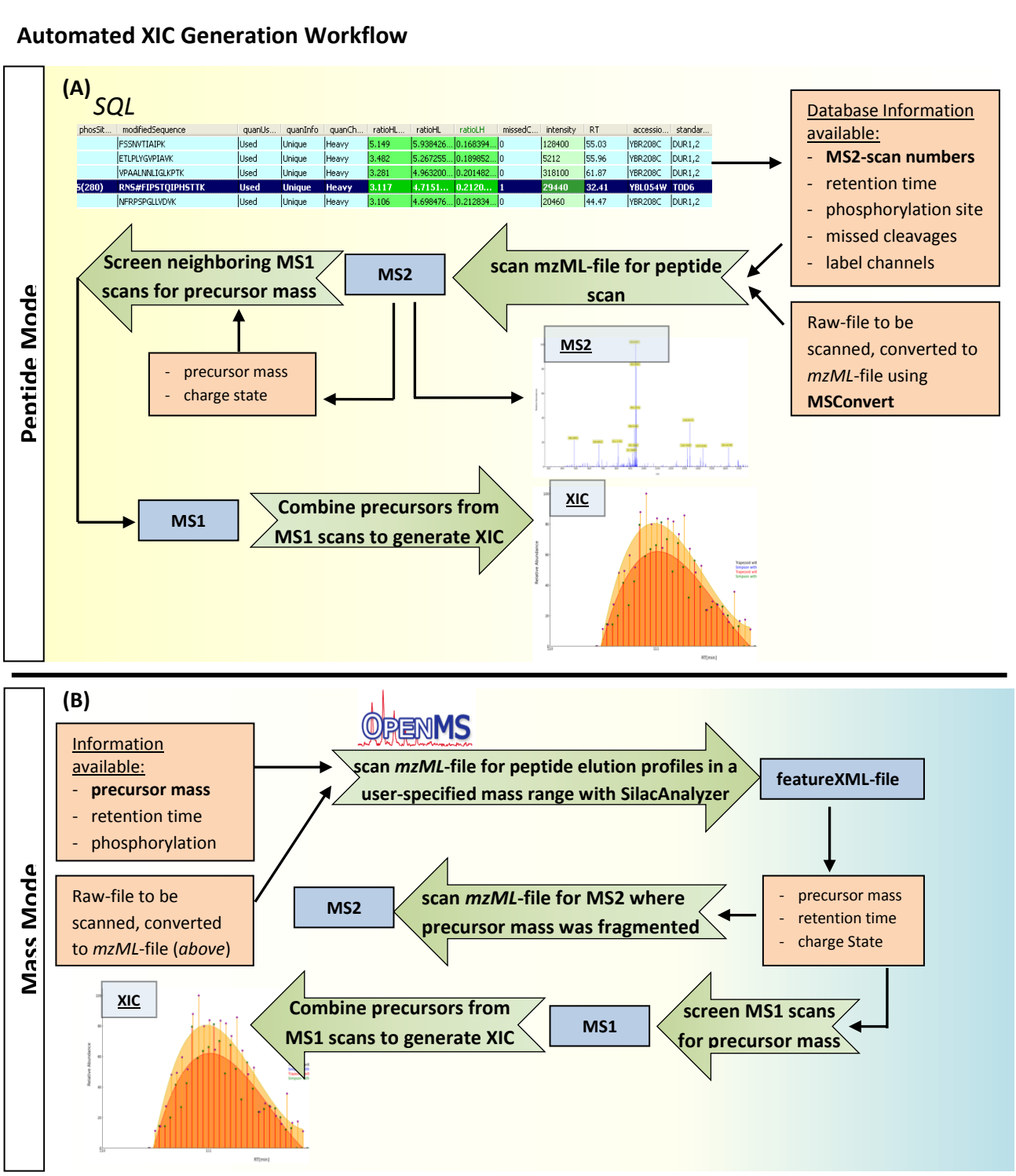


Figure 33. *Workflow Scheme for the two operating modes,* `peptide` *(A) and* `mass` *(B). Both (A) and (B) can generate XICs the design of which can be specified by the user. The inputs for (A) and (B) are different: For (A) it is assumed that the precursor of interest was identified by the search engine used and that corresponding database information is available. Since the scan numbers associated with each identification are MS2-scan numbers, the raw-files (converted to an mzML file for easier searching) have to be scanned for the corresponding MS2-scan. Once detected, the neighboring full scans are screened for the precursor mass to generate the correct XIC. In (B) the user specifies a precursor-mass range (or retention time window, etc.). Using the SilacAnalyzer-node the open-source software OpenMS screens for peptide elution profiles that fit the user-specified criteria and resemble SILAC pairs.*

The second issue regarding the validation of the Hog1 shotgun proteomics screen is related to SILAC ratio credibility. Again, we generate XICs to ensure *Proteome Discoverer©* did not commit any serious quantification error.

Note, however, that manual validation of peptide assignments to spectra by visual inspection is very time-consuming and simply not feasible with large datasets containing tens of thousands of spectra. Therefore, I set up a computational framework that allows visualizing and analyzing scans, precursor masses and retention time windows much faster without using the *XCalibur Qualbrowser* (see *Materials & Methods*). It allows different input files (huge peptide lists) for search queries and user-dependent visualization design. In this chapter I will tackle the above described issues with the computational set-up developed during my thesis work.

III.5.1. Post-processing of MS-data to identify the previously undetected

One important consideration that I took into account while designing the computational framework was the fact that peptides that have remained undetected in one experiment, might be confidently identified in another MS-experiment. Hence, additional information on peptide parameters is available, such as retention time, precursor mass, isotope patterns and fragmentation spectra. Using these parameters one can devise a computational set-up to potentially find peptides that otherwise would have gone amiss (see *Figure 33*).

As illustrated in the scheme in *Figure 33*, there are two modes our set-up operates in, depending on the user input: `peptide` (A) and `mass` (B). If a particular PSM has not passed filtering, we could still retrieve all the associated information from our SQL-database (see *Materials & Methods*). In the peptide-mode, the corresponding rawfile (see *Figure 33*) is searched for the correct MS2-spectra, and neighboring MS1-scans where the specified precursor mass occurs. By extracting a binary array of mass- and intensity-values, we can draw precursor intensities over retention time and thereby generate an extracted ion chromatogram (XIC). Additionally, the set-up allows the drawing of individual MS2-spectra if needed and isotope patterns of corresponding precursors.

The other strategic mode assumes that there has been no identification of the peptide in question; to potentially find it, the precursor mass and/or retention time window are provided as input data. Looking specifically for SILAC precursor pairs, we then applied the Silac-Analyzer node in *OpenMS* to scan rawfiles for peptide elution profiles in a given

precursor-mass range and stored the data as *featureXML*-files. These files include SILAC pairs that fit to the user-defined precursor-mass within a certain mass-error and again the corresponding XICs can be extracted. Parallel to the MS1-screen, it is also possible to look for MS2 scans where the given precursor mass has been fragmented. Ideally, by changing to the `peptide`-mode and searching through the unfiltered dataset stored in the SQL-database.

Both modes of the workflow are based on *Python* scripts I designed during my research work. Basically, there are two main scripts named *mzMLParserPeptides.py* and *mzMLParserMasses.py* that realize the two modes described. With these scripts information can be manually specified or retrieved from the SQL-database (A) or the *featureXML*-file (B). Screening the database for specified peptides has been integrated into the script *mzMLParserPeptides.py* as a separate function. In order to retrieve information from the *featureXML*-file and suitably store them in a fast-accessible dictionary, the script *XMLParser.py* was designed. Finally, I designed three scripts for the generation of MS2-spectral graphs, isotope patterns and XICs that have been integrated in the corresponding mode-class. The source codes can be found in the *Supplementary Materials*.

Using the computational workflow, we proceeded to investigate why Tsl1 was not found in the new *stress response experiments* (see *III.4*) using the `mass`-mode as well as `peptide`-mode. To search for missed potential Hog1-targets, on the other hand, we chose to examine the phosphatase Ppz1 and the nucleoporin Nup2 more thoroughly in the `peptide`-mode of the workflow.

III.5.2. The sites *S49/T67* of Ppz1 are not upregulated in response to hyperosmotic stress

As illustrated in *III.3*, the phosphatase Ppz1, homologue to the Hog1-target Ppz2, hosts a tryptic peptide that is phosphorylated at the residues *S49* and *T67*. This double-phosphorylated peptide has not been quantified in the *stress response experiment*, but displays a three-fold downregulation in the experiments testing *Hog1 stress activity/5min (10min)*, respectively. Hence, proving an increased abundance of the peptide in question, *S#LPSSSTTNTNSNVPDPST#PSKPNLEVNHQR*, in response to hyperosmotic stress is critical to verify that Ppz1 is a Hog1 target.

Using our workflow, we first scanned all confident entries in the SQL-database for the presence of this Ppz1-peptide in the *stress response experiment*. Interestingly, we found *Proteome Discoverer©* to identify the peptide confidently two times in the light form only.

XIC-Analysis of Ppz1-peptide *S#LPSSSTTNTNSNVPDPST#PSKPNLEVNHQR* by precursor-extraction (`peptide` mode)

Peptide identified & quantified/ Database Information

Stress response experiment:
Protein: Ppz1
Peptide:
S#LPSSSTTNTNSNVPDPST#PSKPNLEVNHQR
Number of *confident* identifications/experiment: 2
Number of quantifications: 0
Precursor Mass: 1155.8497
Charges: 3/3
Retention Time [min]: 86.39/78.31
Intensities: 87330/16370
Labeling Channel: Light/Light
Phosphorylation Sites: S49/T67 (all)
Scores: 4.71/3.96
Ratio L/H: n.q./n.q.
Scans: 12610/10027
Rawfiles: silac01/SCX23, silac02/SCX22

Hog1 stress activity/5min (10min):
Protein: Ppz1
Peptide:
S#LPSSSTTNTNSNVPDPST#PSKPNLEVNHQR
Number of *confident* identifications: 4
Number of quantifications: 2
Precursor Mass: 1155.8497
Charges: 3/3
Retention Time [min]: 94.08/79.62
Intensities: 25820/40240
Labeling Channel: Heavy/Light/Light
Phosphorylation Sites: S49/T67 (all)
Scores: 5.19/4.03
Ratio L/H: **0.36**/0.31
Scans: 13016/13051/15447
Rawfiles: 5min-silac01/SCX2021, 10min-silac01/SCX4

Light precursor: 1155.8497
Heavy precursor: 1157.95
Identification cycle

Valid Identification/MS2 spectra

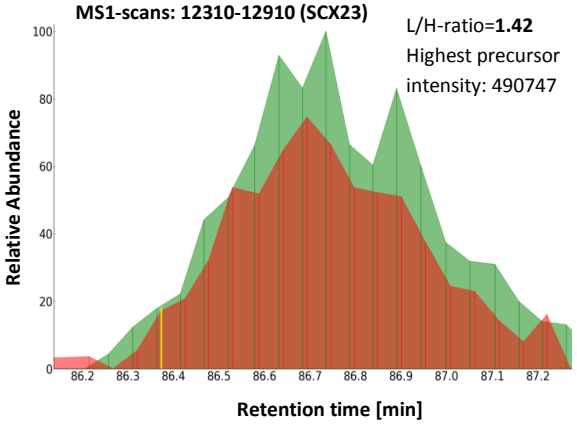
(A) Top10-method (visualization of fragment ions for MS2 spectra comparison)

stress response experiment : scan 12610		*stress response experiment* : scan 10027		control: *Hog1 stress activity/5min:* scan 13016	
masses	relative intensities	masses	relative intensities	masses	relative intensities
892.6	100.00	892.7	100.00	898.7	100.00
998.5	58.01	998.7	66.95	1066.8	57.85
1084.6	87.71	1063.0	65.90	1004.7	57.57
1090.5	89.11	710.2	40.66	1088.9	53.36
949.5	77.59	1090.8	40.55	849.7	37.74
1138.2	99.16	843.7	34.86	1094.9	34.25
1078.9	87.05	1111.7	32.49	716.2	33.92
1088.8	91.38	1084.7	30.00	955.7	22.25
834.6	98.85	893.6	28.88	840.7	20.35
883.7	96.29	1079.0	24.40	1372.6	20.19

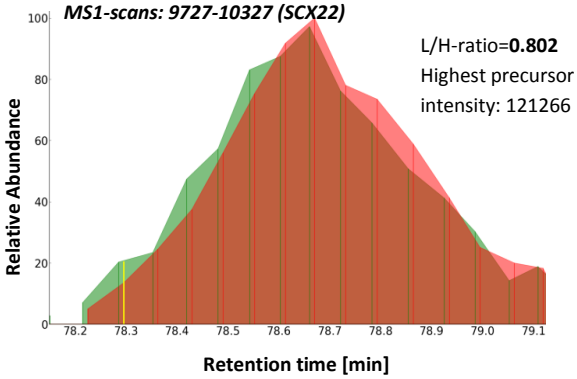
fragment ions detected in control and at least one stress response spectra

Quantification by XIC-Analysis

(B) Detection in *stress response experiment 1*
MS1-scans: 12310-12910 (SCX23)



(C) Detection in *stress response experiment 2*
MS1-scans: 9727-10327 (SCX22)



(D) Detection in *Hog1 stress activity/5min*
MS1-scans: 12716-13351 (SCX2021)

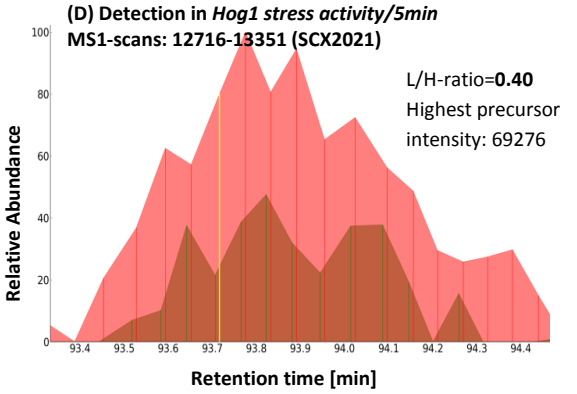


Figure 34. *The graph exemplifies the workflow with the Ppz1-peptide S#LPSSSTTNTNSNVPD-PST#PSKPNLEVNHQR. Given the peptide information from different biological experiments (stress response experiment and Hog1 stress activity/5min), MS2 and MS1-scans were extracted. In neither of the two replicates of the stress response experiment the peptide was quantified by Proteome Discoverer©. Only in Hog1 stress activity/5min experiment the peptide has been quantified. (A) Ten MS2-fragment ions with the highest intensities are displayed. The identification in the Hog1 stress activity/5min serves as a control. In the graphs (B),(C) and (D) XIC-extraction for the first precursor isotope is shown. The yellow line indicates the scan cycle where the precursor has been identified.*

Although this might in principle be indicative of the complementary heavy peptide not being present under hyperosmotic conditions, one needs to prove that assumption by analyzing the corresponding XIC. We submitted the corresponding scan numbers to the workflow to easily retrieve the corresponding MS2-spectra and MS1-isotope patterns of the precursor masses over retention time. To test whether the information recovered in our peptide-mode-workflow is correct, we compared the results with the confident peptide-information we gained from the experiments *Hog1 stress activity/5min (10min)* (see *Figure 34).*

Since *Proteome Discoverer©* assigns the MS2 spectra in both cases confidently to the peptide, it is valid to extract the corresponding precursor masses from the MS2-scans and to compare them between different experiments and rawfiles. We proceeded to screen MS1 scans for the precursor mass of the peptide with a stringent mass deviation of 3ppm. The normalized XIC-area ratio calculated for the precursor in the experiment *Hog1 stress activity/5min* (L/H~0.4) corresponded to the L/H-ratio calculated by *Proteome Discoverer©* (L/H~0.36) (see *Figure 34/D*). Therefore, our method to extract precursor and calculated peak area gives comparable results to the search engine quantification algorithm.

Extracting the same precursor value from the two replicates of the *stress response experiment*, however, revealed that there is no considerable difference between the light

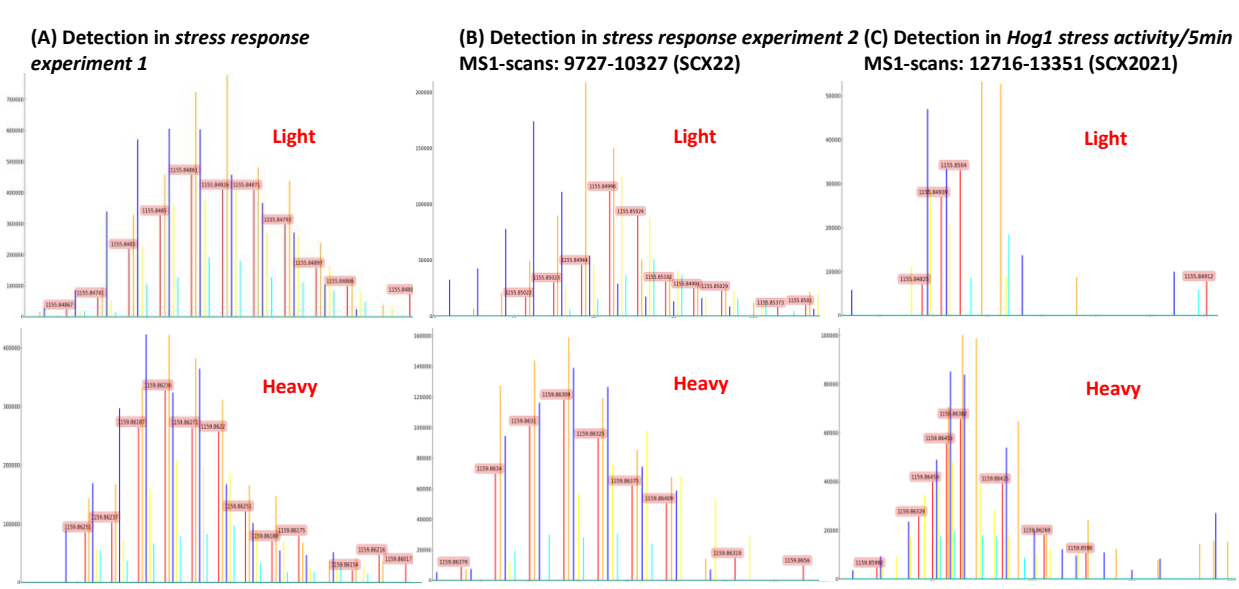


Figure 35. *Isotope Patterns of the peptide precursor in different experiments: (A) stress response experiment, first biological replicate, (B) stress response experiment, second biological replicate, and (C) Hog1 stress activity/5min. The read peak signifies the first isotope of the pattern, the orange one the second, the blue one the third, etc.*

and heavy peptide intensities, respectively (see *Figure 34/B&C*). To again make sure that this is no artifact of the new computational workflow, we checked the same precursor in the corresponding raw-files by manual inspection (data not shown). Furthermore, I propose an additional control to assess peptide elution quality by extracting isotope patterns for the light and heavy precursor, respectively (see *Figure 35*). Interestingly, in the *stress response experiment* where *Proteome Discoverer©* performed no quantification for the peptide in question, we extracted for both light and heavy precursors isotope patterns within 3ppm mass error. The complete isotope patterns suggest a qualitative peptide elution profile in general. With the experiment *Hog1 stress activity/5min,* we did not observe such a peptide elution profile for the light precursor (see *Figure 35/C1*). These isotope patterns cannot provide an explanation as to why there has been no accurate quantification performed in *the stress response experiment*. Our XIC-generation, on the other hand, allows a general conclusion: The (normalized) data indicates that the double-phosphorylated peptide is not upregulated in response to hyperosmotic stress. Hence, according to the field-definition described in *III.3*, this peptide would fall into **Field 3**, meaning that it might be controlled by Hog1 in a stress-independent manner.

III.5.3. *T361* of Nup2 cannot be confidently detected in the *stress response experiment*

All other potentially missed Hog1 targets could be analyzed in the same manner, such as for instance the site *T361* of the nucleoporin Nup2. The site has been previously described by Regot *et al.* [115] to be induced 10-fold in response to hyperosmotic stress. We did not find any confident quantification or identification in our *stress response experiment*, but a 2.5-fold downregulation in the experiment *Hog1 stress activity/5min* (see *Table VI*).

When screening our entire (unfiltered) peptide database (see *Materials & Methods*), we found the peptide to be identified once in the *stress response experiment* with an *XCorr*-value of 3.08 (see *Figure 36*). Hence it did not pass the q-value filtering (FDR<1%). Again, we extracted the peptide-related MS2- and MS1-scans from the corresponding rawfiles. The same peptide that was confidently identified and quantified in the experiment *Hog1 stress activity/5min* was subjected to the same data processing and served as a control. The ratio calculated by *Proteome Discoverer*© fits the normalized area ratio calculated by my script (L/H~0.48).

XIC-Analysis of Nup2-peptide *NT#PDASKPSFSFGVPNSSK* by precursor-extraction (peptide mode)

Peptide identified & quantified/ Database Information

Hog1 stress activity/5min:
Protein: Nup2
Peptide: NT#PDASKPSFSFGVPNSSK
Number of *confident* identifications/experiment: 1
Number of quantifications: 1
Precursor Mass: 682.97589
Charges: 3
Retention Time [min]: 112.79
Intensities: 21410
Labeling Channel: Light
Phosphorylation Sites: T361
Scores: 3.77
Ratio L/H: 0.41
Scans: 15079
Rawfiles: silac01/SCX1819

Stress response experiment (unfiltered):
Protein: Nup2
Peptide: NT#PDASKPSFSFGVPNSSK
Number of identifications/experiment: 1
Number of quantifications: 0
Precursor Mass: 682.97589
Charges: 3
Retention Time [min]: 96.14
Intensities: 48650
Labeling Channel: Light
Phosphorylation Sites: T361
Scores: 3.08
Ratio L/H: n.q.
Scans: 13927
Rawfiles: silac02/SCX22

Valid Identification/MS2 spectra comparison:

(A) Top10-method (visualization of fragment ions for MS2 spectra

stress response experiment: scan 15079		control: *Hog1 stress activity/5min:* scan 13927	
masses	relative intensities	masses	relative intensities
876.7	100.00	876.7	100.00
532.2	41.67	916.7	49.24
877.5	30.82	532.3	47.56
611.6	21.70	877.6	38.68
612.7	20.81	611.5	37.03
916.6	16.64	612.6	29.73
688.4	13.77	533.3	16.15
735.1	13.38	671.5	16.07
533.3	11.50	770.6	15.46
665.3	10.80	673.7	14.72

fragment ions detected in control and stress response spectrum

Quantification by XIC-Analysis

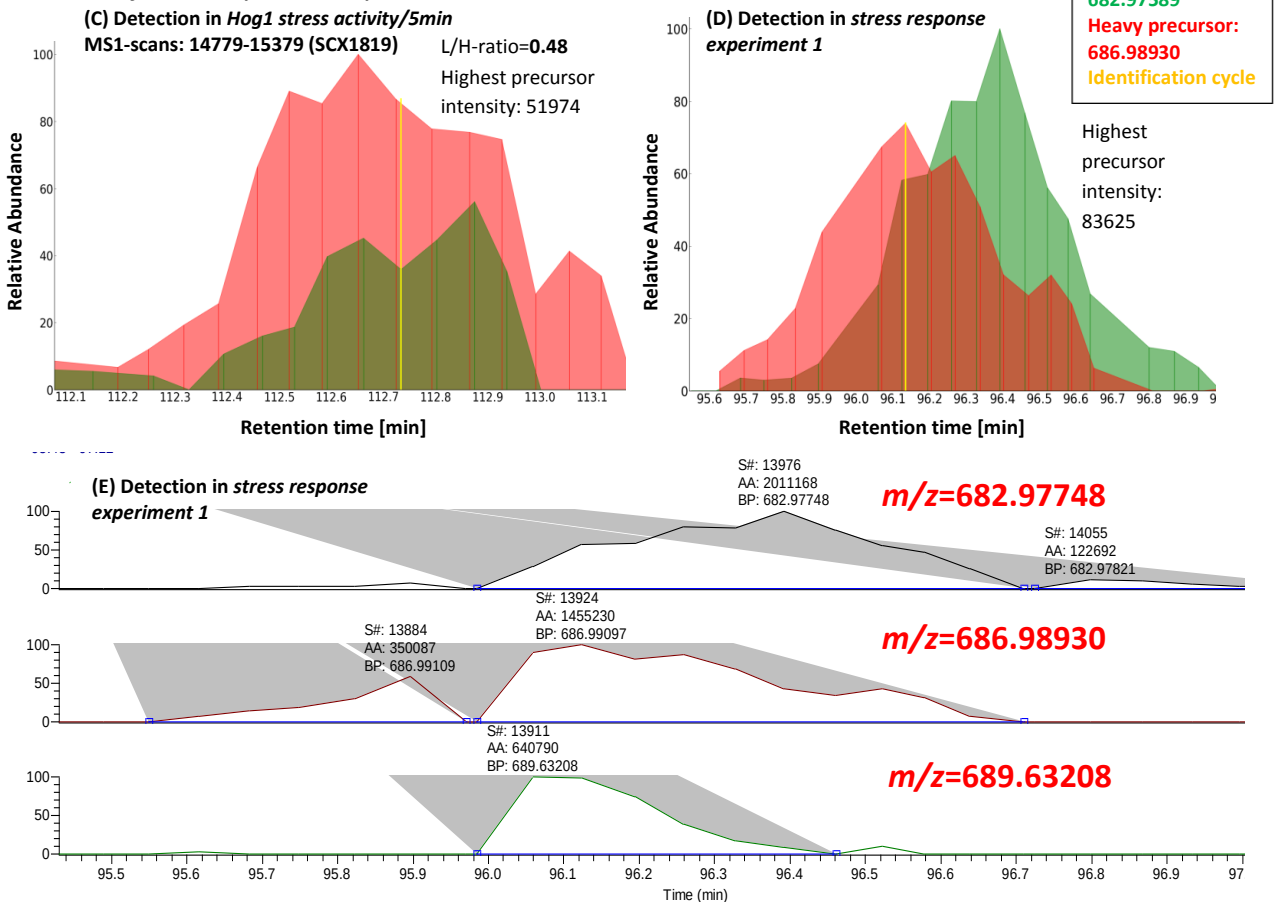


Figure 36. *The graph exemplifies the workflow with the Nup2-peptide NT#PDASKPSFSFGVPNSSK. The peptide has not been confidently identified in the stress response experiment; the experiment Hog1 stress activity/5min serves as a control. (A) Ten MS2-fragment ions with the highest intensities are displayed. Comparing the top100 fragment ions using a mass error of 0.8Da revealed an overlap of 73 ions (data not shown). In the graphs (C) and (D) XIC-extraction for the first precursor isotope (C-12) is shown for the respective experiments. The yellow line indicates the full scan where the precursor has been identified. (E) XICs as extracted for peptide-related precursors masses using Thermo© XCalibur Qualbrowser.*

Interestingly, when visualizing the XIC for the same precursor mass in the *stress response experiment*, we observe a slight retention time shift between the heavy and light peptide versions which in principle should not occur, as they should not differ in their chromatographic behavior (see *I.2.*). To ensure that this is no script artifact, we also manually visualized the XICs (using *Thermo© XCalibur Qualbrowser*, see *Figure 36/E*). The shift between the light and heavy peptide species remained, however. To find an explanation for this unexpected observation, we checked whether (a) this shift would still occur if the ion chromatogram extraction is restricted to 2ppm-precursor mass tolerance and (b) whether this shift also occurs with a differently-charged precursor that should elute in the same retention time window.

The results for (a) and (b) were unsatisfying, since we still observed a retention-time-shift with 2ppm-precursor mass tolerance and the presumed double-charged peptide (data not shown). We then ventured that this peptide behavior might be caused by co-eluting peptides of the same precursor mass. Isolating the peptide isotope patterns from individual MS1 scans might give an additional hint as to whether peptide identification has been the

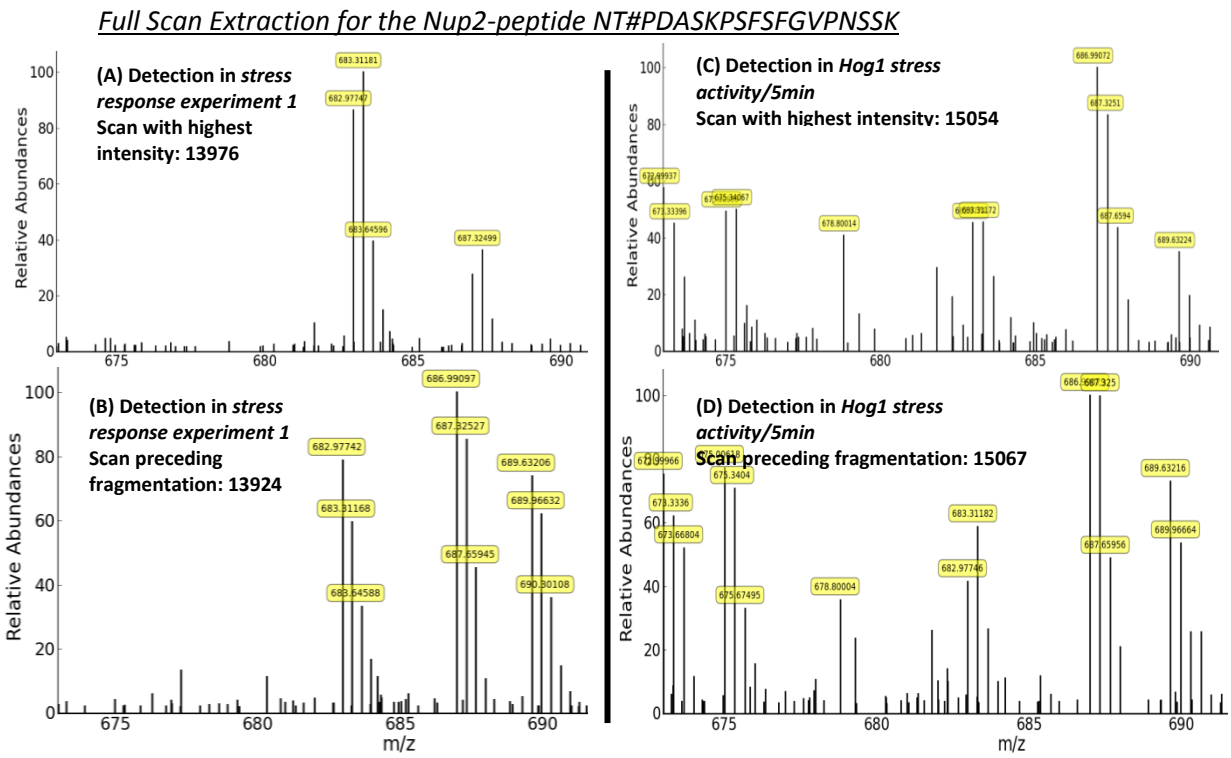


Figure 37. *Full scan snapshots of specific full scans. (A) Full scan with highest precursor intensity in the control experiment Hog1 stress activity/5min. (B) Full scan preceding precursor fragmentation in the control experiment Hog1 stress activity/5min. (C) Full scan with highest precursor intensity in the stress response experiment. (D) Full scan preceding precursor fragmentation in the stress response experiment.*

same in the control (*Hog1 stress activity/5min*) as well as in the *stress response experiment*. The isotope patterns between the experiments were in fact distinct from each other (data not shown). The precursor isotope pattern observed in the *stress response experiment* allows the surmise that a peptide other than *NT#PDASKPSFSFGVPNSSK* is observed in the affected retention time window. My script additionally allows visualizing a snapshot of a specified full scan. We reasoned that if something is co-eluting with the light or the heavy precursor in the *stress response experiment*, we could probably visualize it in the full scan. Hence, we extracted the full scan snapshots for both experiments for the full scan preceding precursor fragmentation (yellow line in *Figure 36*) and the full scan with the highest precursor intensity. Interestingly, we observed the peptide elution profiles for the light and heavy precursor to be remarkably distinct in the *stress response experiment* (see *Figure 37/A&B*) and no other peptide species to interfere with their elution. Therefore, we could exclude latter as the cause for the retention time shift observed in the XIC.

This leads us to our final conclusion that the light and heavy peaks actually harbor differently phosphorylated species of the same Nup2-peptide and that it fails to quantify distinct peptide species due to insufficient chromatographic resolution. Clearly, in the first part of the heavy XIC (see *Figure 36/D*), the heavy precursor would predominate over the light precursor. A few milliseconds afterwards, however, the light precursor dominates over the heavy peptide variant. Considering that the Nup2-peptide *NT#PDASKPSFSFGVPNSSK* contains six residues that might be potentially phosphorylated, we likely observe a number of phosphorylated species.

In total, analyzing the peptide-XIC in the *stress response experiment* did not yield any information as to whether the phosphorylation site *T361* is either up-or down-regulated in response to hyperosmotic stress. Note, however, that the computational workflow has provided a handful of tools that are useful for troubleshooting (such as demonstrated) and a possible explanation as to why *Proteome Discoverer©* does not provide quantification in this case. For further analysis of *T361* of Nup2, we would consider a targeted validation approach using longer chromatographic gradients.

III.5.3. Analyzing *S147* of Tsl1 in more detail

As described in *III.4,* we have established *S135* and *S147* of Tsl1 to be putative target sites of the MAPK Hog1. We next aimed at understanding why we could not detect or quantify any

corresponding peptide in the *stress response experiment* of our recent approach. As mentioned in *III.4,* the corresponding SILAC ratios were retrieved from the original *stress response* dataset, in which the Tsl1-peptide harboring *S147* was detected (see *III.1*). Thus, we asked ourselves why we could not repeatedly detect it in the new *stress response experiment* and thereby confirm the SILAC ratios as illustrated in *Figure 38*. As described in *III.4*, we have ruled out poor Tsl1-peptide detectability as the cause for failed detection. We assume that the latter is most probably linked to non-occurring precursor fragmentation or non-qualitative peptide assignments.

We did find the peptide harboring *S147* in the shotgun-experiments *Hog1 stress*

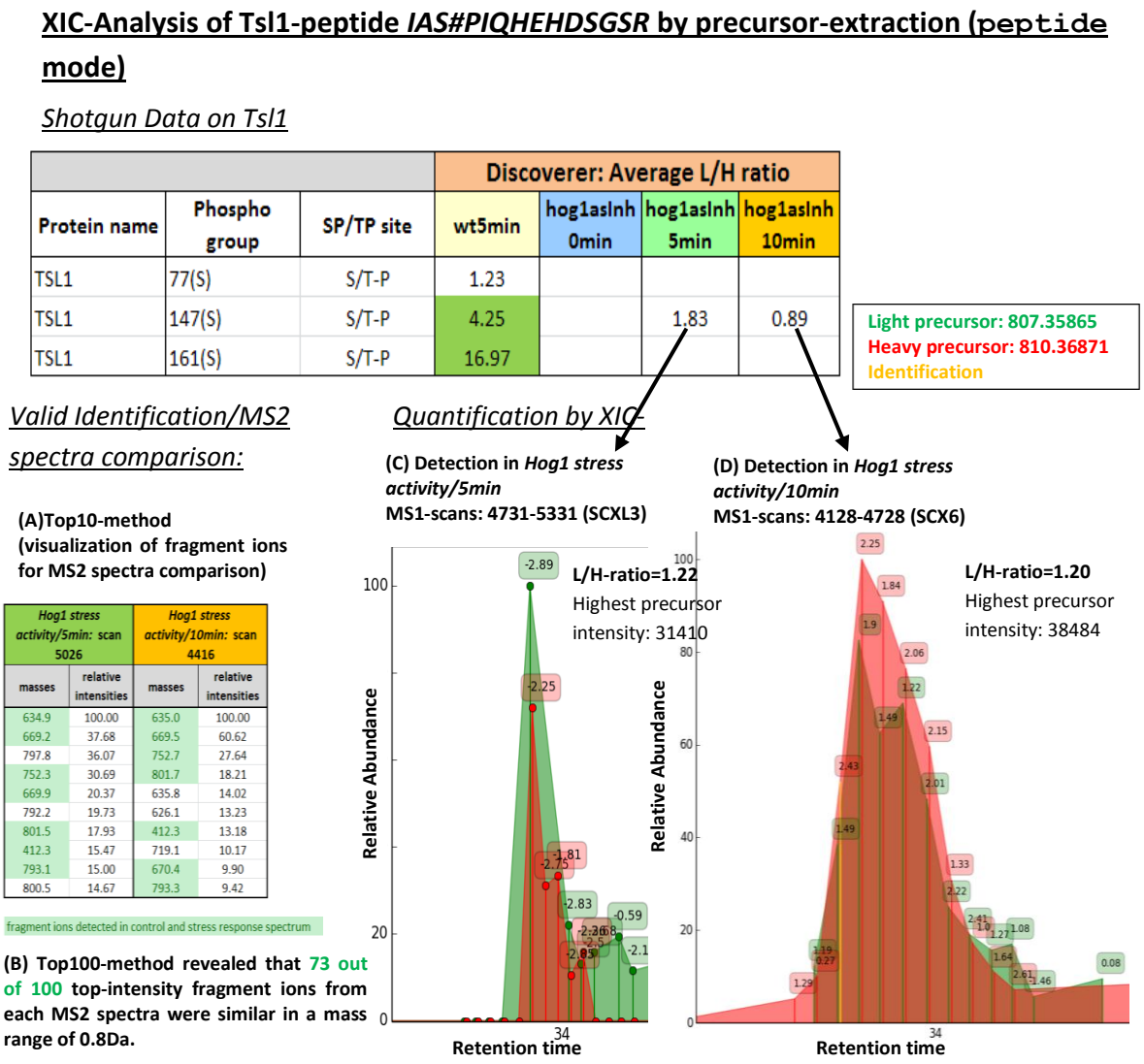


Figure 38. *The graph exemplifies the workflow with the Tsl1-peptide IAS#PIQHEHDSGSR. (A) Ten MS2-fragment ions with the highest intensities are displayed for the experiments Hog1 stress activity/5min and 10 min. Comparing the top100 fragment ions using a mass error of 0.8Da revealed an overlap of 73 ions (data not shown). In the graphs (C) and (D) XIC-extraction for the first precursor isotope (C-12) is shown for the respective experiments.*

activity/5min (10min), however. In this respect, we were interested in why the phosphorylation at *S147* indicated no Hog1-dependency in the experiment *Hog1 stress activity/5min*. Keep in mind that we previously identified the site *S147* as a Hog1-dependent site in the targeted validation approach.

To tackle these issues, we first employed the `peptide` mode of our computational workflow to screen our unfiltered database for Tsl1-peptides. The peptide *IAS#PIQHEHDSGSR* harboring the site *S147* was identified and quantified twice in the experiment *Hog1 stress activity/5min (10min)*, respectively. We extracted the corresponding XICs to assess peptide abundance change under these conditions and to verify the SILAC ratios as calculated by *Proteome Discoverer©*. For the respective experiments we calculated and normalized the area ratios. In both cases the L/H-ratio was 1.2, thus, according to XIC-extraction, the phosphorylation at *S147* does not become less abundant if Hog1 is specifically inhibited. However, is the peptide identification consistent with the identification in the targeted validation approach? Identification spectra of the peptide left no doubt in the targeted validation as well as in the shotgun approach that peptide-spectrum assignment was accurate (data not shown). Thus, further investigation of Tsl1 would be necessary to define its Hog1-dependency in an MS-setup with more biological replicates.

Next, we used the `mass` mode to possibly detect a peptide elution profile in our new *stress response experiment* to confirm the data generated from the former dataset. As described in *Figure 33*, we used the Silac-Analyzer-node in OpenMS to search for peptide elution profiles (termed “features”) in a specific mass/charge window. For the peptide *IAS#PIQHEHDSGSR* we set a window of *m/z*=800-850. We generated 42 featureXML-files in total from the two replicates of the *stress response experiment*. Using a mass tolerance of 3ppm, we then searched for peptide elution profiles that most likely resemble the Tsl1-peptide. Interestingly, we often found the light precursor 807.35865 as the heavy part of a presumed SILAC pair. In the first shotgun experiment, for example, this occurred 13 times along with the presumed light precursor with *m/z*=804.34531. In case the heavy precursor part got fragmented, this would provide an indirect explanation for the failed detection of the Tsl1-peptide since the light precursor would fall into the dynamic exclusion window. However, OpenMS also detected SILAC features that resemble the peptide elution profile of the peptide *IAS#PIQHEHDSGSR*. In the first replicate of the *stress response experiment* this was

Isotope Pattern Analysis of Tsl1-peptide *IAS#PIQHEHDSGSR* by precursor-extraction (`mass-mode`)

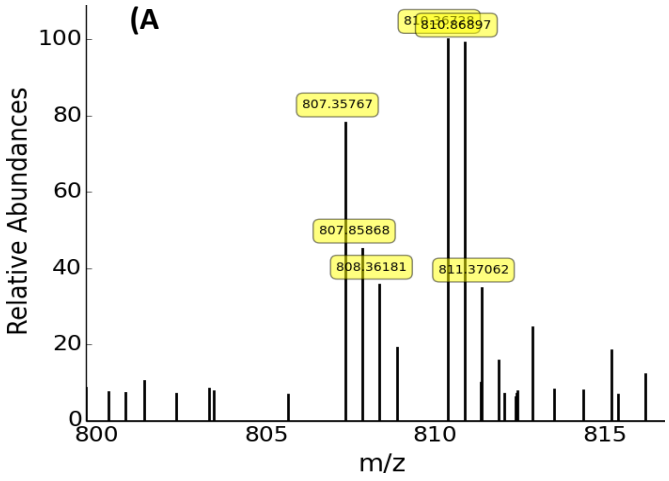


Figure 39. *(A) Control Isotope Pattern of the peptide IAS#PIQHEHDSGSR from the targeted validation approach. The peptide has a charge of 2. (B) displays the results of the isotope patterns as detected by the Silac-Analyzer (OpenMS©) for the light and heavy precursor, respectively. The columns marked in light green fit to double-charged isotope patterns and are hence the most likely peptide elution profiles to correspond to the peptide IAS#PIQHEHDSGSR.*

(B

	stress response experiment 1				stress response experiment 2					
	SCX15	807.361-810.37106	SCX18	807.35655-810.36662	SCX13	807.35913-810.36919	SCX14	807.35979-810.36988	SCX23	807.35676-810.36683
Light Isotope Pattern (m/z)		807.361 808.361		807.3565/807.35669 807.85652/807.85669 808.35652/808.35669 808.85653		807.35919 808.35912		807.35949/807.359871 807.85949/807.859871 808.35949/808.359871 808.85987/809.35987		807.35676 808.35676
Heavy Isotope Pattern (m/z)		810.3711 811.3711		810.36658/810.36675 810.86658/810.86675 811.36658/811.36675 811.86659		810.36919 811.36919		810.36956/810.36994 810.86956/810.86994 811.36956/811.36994 811.86994/812.36994		810.36682 811.36682

observed two times in the SCX-fractions 15 and 18. In the second replicate, the supposed peptide was observed three times in SCX13, SCX14 and SCX23. Since the peptide elution profiles were detected in different SCX-fractions, it is important to assess them via visualization and direct comparison with the identified Tsl1-peptide. For that purpose we use the peptide detected in the targeted validation approach as a control (see *Figure 39/A*) and first assessed which detected isotope patterns in principle would fit to the correct elution profile. The features detected in SCX18 in the first replicate and SCX14 in the second replicate might correspond to the peptide *IAS#PIQHEHDSGSR* (see *Figure 39/B*). To examine these peptide elution profiles more thoroughly, we checked whether any fragmentation has occurred in the concerned retention time window. Naturally the fragmentation spectrum would provide the ultimate proof whether the peptide elution profile indeed corresponds to *IAS#PIQHEHDSGSR*. We found the light precursor 807.35865 to be fragmented four times in the SCX18-fraction of the *stress response experiment 1*. The heavy precursor was not subject to fragmentation. In the SCX14-fraction of the *stress response experiment 2,* the light precursor was fragmented twice and the heavy peptide was again not picked for fragmenta-

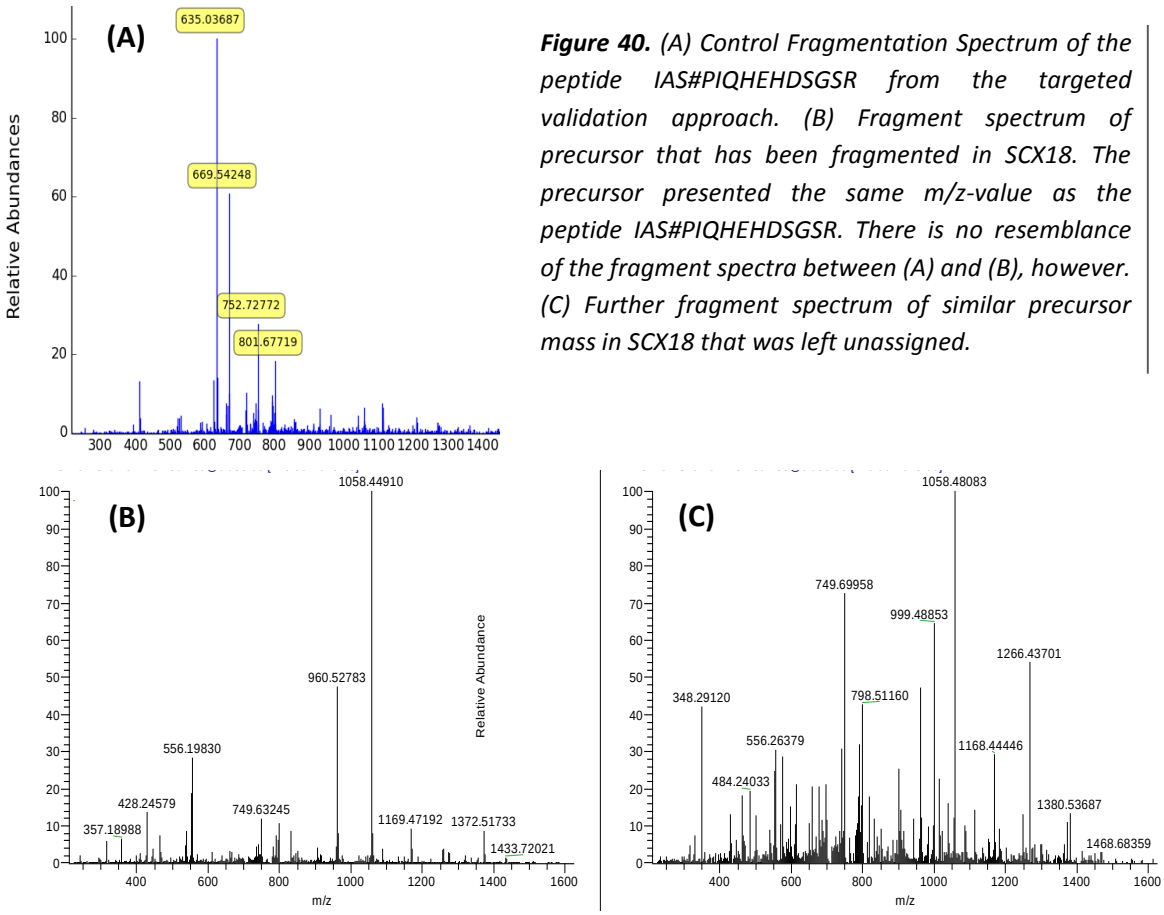


Figure 40. *(A) Control Fragmentation Spectrum of the peptide IAS#PIQHEHDSGSR from the targeted validation approach. (B) Fragment spectrum of precursor that has been fragmented in SCX18. The precursor presented the same m/z-value as the peptide IAS#PIQHEHDSGSR. There is no resemblance of the fragment spectra between (A) and (B), however. (C) Further fragment spectrum of similar precursor mass in SCX18 that was left unassigned.*

tion. However, when comparing these fragmentation spectra with the confident spectra of the targeted validation approach, we could clearly see that different peptide species were fragmented (see *Figure 40*). In SCX18, the spectrum (*Figure 40/B*) was assigned to the heavy version of the TIF6-associated peptide *LQDAQPESIS#GNLR*; in SCX14 (*Figure 40/C*) the fragmentation spectra remained unassigned. Our computational tools have therefore ruled out the features that could pass as an elution profile of *IAS#PIQHEHDSGSR* to originate from this peptide. Although the `mass` mode of our workflow did not bear fruit in this case, we believe that the *Python* tools developed in the course of this thesis project provide a productive asset for future MS-experiments and sophisticated data analysis.

IV. Discussion

Saccharomyces cerevisiae reacts to increased external osmolarity by activation of the high-osmolarity glycerol (HOG) Mitogen-activated Protein (MAPK) pathway, a homologue to the mammalian p38 MAPK-pathway. The key regulating MAP kinase Hog1 has proven to be essential for cellular survival under hyperosmotic stress conditions in yeast. However, the changes its activity induces in cellular physiology in response to hyperosmotic stress has not been fully recorded yet. Studying its effects in a tractable model organism such as yeast should prove useful to understand the workings of mammalian MAPK-pathways in general which are often involved in human disease, for example. Around ten years ago it was widely believed that all MAPK-pathways elicit alteration in cellular physiology solely via transcriptional regulation: Essentially the activated MAPK would be translocated to the nucleus where it induces the transcription of specific genes. In the case of Hog1, these are e.g. genes involved in the glycerol machinery. This transcriptional activation *per se* would suffice to conquer hyperosmotic stress. This notion, however, was undermined by several studies in yeast: Westfall *et al.* [73] demonstrated that if the MAPK is tethered to the cell plasma membrane and thus could physically not move into the nucleus, cells were still capable of surviving severe hyperosmotic shock. Moreover, other yeast studies suggested that a number of cytoplasmic proteins, for example the aquaglyceroprotein Fps1, might be directly regulated by the MAPK Hog1 [11]. Members of our research lab hence proceeded to set up a shotgun-study to capture other potential cytoplasmic targets that are regulated by Hog1 under hyperosmotic stress conditions. Elucidating these targets would then give a clarified picture of the extent of the regulatory role of the MAPK Hog1. The shotgun data had to be extended due to the technical limitations of the previous approach by recent experiments which are discussed in detail in this thesis.

IV.1. Reliable MS-Dataset to Extract Confident Set of Hog1 Targets. Small is Beautiful.

In this report, we provide a comprehensive proteomic MS-shotgun analysis to screen for novel potential targets of Hog1. To ensure a specific search and avoid secondary effects from

knockout-mutant strains, we employed a Hog1*as*-alleles and a novel *as*-specific inhibitor [102,103]. Although mutant strains carrying conditional temperature-sensitive alleles could potentially provide a similar cell state, these alleles require analysis under non-physiological conditions and a longer time period to become inactivated. The inhibitor, on the other hand, exhibits its effects rather quickly, in a dose-dependent manner (total or partial inhibition of kinase possible) and without activating other non-Hog-related pathways. However, there are some issues to be aware of when using analog-sensitive inhibitors for a given kinase: For one thing, optimizing kinase specificity and membrane permeability remains challenging, as well as estimating inhibitor dissociation rates after being attached to the kinase for some period of time. For the Hog1*as* inhibitor these parameters were optimized to allow selective targeting of the Hog1 MAPK even in complex extracts with multiple kinase representatives [102,103].

In total we were able to generate four different MS-datasets (with two replicates each) and confine them to 36 proteins that we confidently defined as Hog1 targets (see *Table R3*). We identified 44 additional protein candidates (see *Table R4*) that might be regulated by Hog1, but require further analysis to substantiate this assumption.

To our knowledge, this is the first mass-spectrometry screen that integrates two different experimental conditions; one putting forward phosphorylation pattern changes that occur upon exposure to hyperosmotic stress and the other putting forward the effect on these phosphorylation sites if the MAPK Hog1 is specifically inhibited. Filtering S/T-P-motifs for these criteria resulted in the confident set of Hog1 target proteins. Although we assume the observed changes in phosphorylation to be reliable evidence already, we currently undertake a further validation step for these putative 36 targets by measuring their ability to interact with Hog1 using a protein interaction assay (M-track). Therefore, this report provides one of the most thoroughly-analyzed and best-controlled studies for the search of possible targets of a specific kinase.

Though carefully scrutinized, it is important to keep in mind how likely it is that false positive information might slip through. From the biological point of view, this would be the case if a) conditions were not properly chosen or b) the phosphorylations at S/T-P motifs are indirectly controlled by Hog1. For a) we have confirmed sample quality regarding all relevant aspects by identifying protein hallmarks associated with the activation of the Hog1 pathway and biological processes of the stress response in general. Furthermore, we have generated

biological replicates that are highly reproducible in terms of SILAC ratios and partly implemented a dataset that has been previously generated by our group for comparison. We figure b) unlikely to occur since we consider two independent experimental dimensions and a highly conserved MAPK-motif. Thus, the biological aspect most likely does not allow false positives to occur in our confident set of target proteins. Regarding the MS-technical aspects, there are some considerations to be pointed out, however: Basically, there are two kinds of errors that might occur with the MS-analysis, one being a false peptide-spectrum-assignment and the other being erroneous SILAC quantification. We circumvented false identifications by strict FDR-filtering using q-value calculation according to peptide charge state. As shown in *III.5*, we would thereby even miss some high-scoring spectra, leaving the window open for a maximum of 1% potential false positive peptide assignments. Notably, such filtering criteria could not be applied to SILAC-quantifications as there is no statistical method to estimate the number of false quantifications. These might, for example, occur if a peptide is identified in the tail of its elution peak (after a local peak minimum has occurred already). Clearly, potentially erroneous SILAC quantification represents the only potential gap for false leads to occur even in the confident protein set we established.

IV.2. Coupling Automated Analysis with Manual Validation

We hence set out to close this gap by manually validating the corresponding spectra to ensure quality of SILAC quantification. Since this task is considerably time-intensive and not easily feasible with the massive MS-datasets, I designed *Python* scripts to visualize user-defined extracted ion chromatograms (XICs). Additionally to XIC-visualization, my script provides parameters such as mass deviation for each specified precursor, the scan numbers (MS2 and MS1) and the identification-associated full scan. Moreover, it is also possible to specifically extract a full scan to assess potential co-elution of other precursors that lie in the specified mass window. Note, however, that the scripts only provide a visualization tool and do not perform reliable area quantifications. This is due to the extracting mode for XIC-generation: Whereas *Proteome Discoverer©* considers the entire peptide isotope pattern for quantification, my script depicts individual peaks of the isotope pattern (C-12, C-13-1, C-13-2, etc.). Thus, using the tool one could therefore only approximate the ratio calculated by the

search engine, which we reckon will suffice in ensuring SILAC ratio credibility of our confident protein target set.

Employing my visualization tool, we further addressed the question as to whether our confident set of identified Hog1 target proteins is only a small part of the actual number of targets. After all, we observed 44 additional proteins to harbor phosphorylations at S/T-P motifs that show decreased abundance if Hog1 is specifically inhibited but could not confidently recover these sites in the *stress response experiment*. Naturally, we were interested in retrieving the missing information from our unfiltered dataset and in assessing a possible increase in abundance in response to hyperosmotic stress. Due to time reasons, this approach was so far used for three proteins, namely Ppz1, Nup2 and Tsl1. For the affected sites *S45/T67* of Ppz1 we could reliably observe that the corresponding peptide does not increase in abundance under hyperosmotic stress conditions. The phosphorylation at the Nup2-site *T361*, on the other hand, remained un-quantified; using the computational tools, however, we could pinpoint at chromatographic resolution being the major limit to this peptide quantification. Finally, I also used the tool to analyze peptide features deduced with OpenMS to study potentially missed Tsl1-peptide elution profiles. When refined and tuned for broader features, this computational workflow as illustrated in *III.5* might prove useful in the day-to-day analysis as conducted in our research lab and in the mass spectrometry facility. For instance, an analysis to confirm potential links to Hog1 activity under hyperosmotic stress conditions, should in principle be applied to all 44 proteins that have remained unquantified or unidentified in the *stress response experiment*.

IV.3. Possible Hog1-Dependent Scenarios in Response to Hyperosmotic Stress

For now, I will stick to the list of 36 confident Hog1 targets that are currently tested for protein interaction with the MAPK. Positive M-track results will provide the ultimate proof for an involvement in the Hog1-MAPK-pathway and form the basis for further mutational studies. To this list of proteins we also added the large subunit of the trehalose synthase complex Tsl1. This is due to our observation in a MS-target validation procedure of two sites being regulated in a Hog1-dependent manner. From the biological perspective, it is indeed realistic that Tsl1 is implicated in the Hog1-regulatory mechanism as it has been shown that

trehalose levels increase at the onset of the yeast osmotic stress response (personal communication). Additionally, we observe the neutral trehalase Nth1 to display Hog1-dependent alterations of the phosphorylation pattern (personal communication) after five minutes of hyperosmotic stress (see *Figure 41*). This again provides supporting information as to the putative involvement of trehalose as a compatible osmolyte during hyperosmotic stress in yeast.

		Discoverer: Average L/H ratio			
Protein name	Phospho group	wt5min	hog1asInh 0min	hog1asInh 5min	hog1asInh 10min
NTH1	23(S)	0.08			
NTH1	58(T)			0.75	1.35
NTH1	60(S)	0.07	0.92	0.75	0.82
NTH1	83(S)	0.26	0.89	0.81	0.61
NTH1	87(T)	0.35			
NTH1	91(S)		0.94		

Figure 41. *Nth1 data from the shotgun experiments. The N-terminal part of the protein has been primarily covered; no S/T-P site could be detected, however.*

Interestingly, we also identified the phosphatase Ppz2 to contain a phosphorylated S/T-P site that displays Hog1-dependency. The phosphatase has been shown to negatively regulate the potassium channel Trk1, thereby determining the upper limits of potassium accumulation [117,118, 119,120]. It does so by modulating the phosphorylation state of Trk1 [121], which would then presumably pass to its activated state or control transporter trafficking more stably, as it has been suggested for other yeast transporter proteins [122]. Further genetic evidence indicates that Trk1 is- reciprocal to Ppz1- activated by the protein kinase Hal5. The critical point is that we identified both Hal5 and the homologue of Ppz1, Ppz2, to contain Hog1-dependent S/T-P-sites and could confidently assign them to our list of 36 confident Hog1 targets (see *Figure 42/A*). Moreover, we also observed significant changes in the phosphorylation pattern of Trk1 at residues that are not located within an S/T-P motif (see *Figure 42/B*). This would suggest Hog1-dependent indirect control, putting forward Hal5 and Ppz1/2 as potential (Hog1-dependent) key regulators. Notably, a similar scenario has already been described by Levin *et al.* [11], where the glycerol channel regulators Rgc1/2 were demonstrated to be phosphorylated by Hog1 and to exert the effect on the glycerol channel Fps1. Hog1 is physically recruited to Fps1, however. To verify whether this interaction mode is also valid for the regulation of the yeast potassium transporter, it is necessary to confirm a direct interaction between the MAPK Hog1 and both Trk1-regulators (work in progress).

Furthermore, it is necessary to perform MS-target validation approaches for all participating

(A)			Discoverer: Average L/H ratio			
Protein name	Phospho group	SP/TP site	wt5min	hog1asInh 0min	hog1asInh 5min	hog1asInh 10min
HAL5	19(S)		0.45			
HAL5	63(S)	S/T-P	5.71		1.57	
HAL5	63(S),66(S)	S/T-P	2.29		0.12	
HAL5	63(S),68(S)	S/T-P	2.44		1.09	
HAL5	229(S),233(S)		0.34		0.94	
HAL5	233(S)		0.50	0.93	0.72	0.84
HAL5	273(S)		0.41	0.64	0.55	0.50
HAL5	273(S),280(S)		3.03		0.55	
HAL5	277(S),280(S)		16.10		0.85	0.48
HAL5	280(S)		12.97	0.67	0.73	0.60
HAL5	391(S),395(S)		0.47			
HAL5	395(S)		2.06	1.11	1.02	1.04
PPZ1	31(T),34(S),37(S)		5.95			
PPZ1	235(T)		10.26			
PPZ1	249(S)		0.03	1.72	25.22	26.30
PPZ1	253(S)		0.11	2.11		18.67
PPZ1	260(S),265(S)	S/T-P	16.79			
PPZ1	283(S),284(T)		0.42	0.74		
PPZ1	310(S)		0.22	1.41		

(B)		Discoverer: Average L/H ratio			
Protein name	Phospho group	wt5min	hog1asInh 0min	hog1asInh 5min	hog1asInh 10min
TRK1	12(T)	3.01		0.87	
TRK1	414(S)	2.96	0.97	0.28	0.28
TRK1	412(S)	2.89	0.81	0.27	0.34

Figure 42. *(A) Dataset for regulated sites of the protein kinase Hal5 and the phosphatase Ppz1. (B) Dataset for the regulated sites of the potassium transporter Trk1.*

key players of the proposed mechanism. Thereby all potential phosphorylation sites would be captured and accurately quantified so as to allow for defined and clean point-mutational studies. Finally, these studies might provide an explanation as to why and how the potassium transporter Trk1 integrates input signals from two widely differing branches that might be controlled by the same key regulator Hog1 (see *Figure 43*).

IV.4. Fine-tuning of the Hog1-Response by Phosphorylation Mechanisms

A number of targets that fall into our confident protein set were transcription-associated proteins, such as the Hog1-related protein hallmarks, Sko1 and Hot1. For these proteins we could again confirm Hog1-dependency for sites that have been previously described in the literature [6] and additionally could use them as a quality control for our screen. As transcription-associated proteins actually represent a considerable part of our confident protein set, I take this opportunity to return to the question that has been partly raised in this chapter before. Does the MAPK-induced transcription of stress-responsive genes play the most essential role in the yeast hyperosmotic stress response? As our set of target proteins confidently covers nucleic factors, our data confirms the notion that transcription

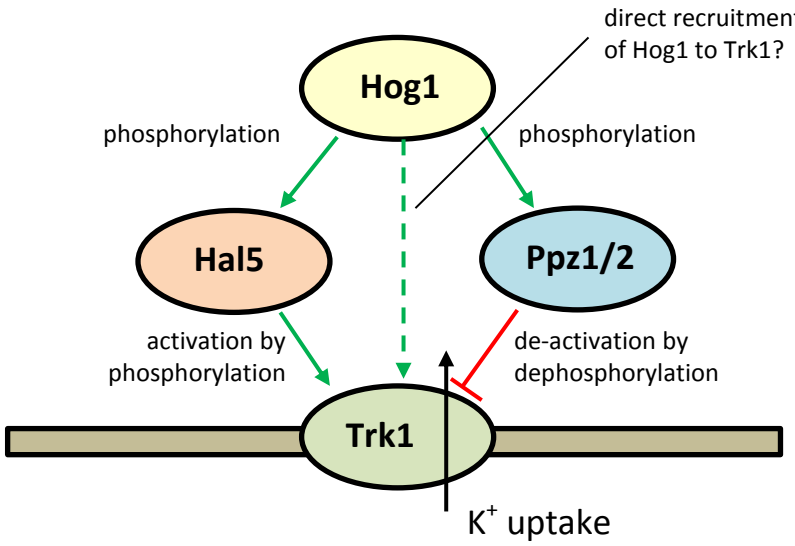


Figure 43. *Proposed scheme of action of the Hog1-regulated Ppz1-Hal5-Trk1-system.*

is indeed strongly influenced during the hyperosmotic stress response. The importance of Hog1-associated transcriptional regulation has been outlined in the recently published paper by Babazadeh *et al.* [123]. According to the authors, transcriptional activation of the two genes *GPD1* and *GPP1* is the only essential regulatory mechanism needed for the cell to survive sudden osmotic shock. Since our confident set of target proteins, however, cover both nucleic as well as cytoplasmic proteins, there also has to be a valid explanation for latter targets to be Hog1-regulated. We therefore put forward the natural assumption that the yeast osmotic stress response might require both nucleic as well as cytoplasmic regulatory mechanisms in different contexts. Transcriptional regulation *per se* is definitely important to uphold cellular levels of glycerol to conquer enduring osmotic challenges and severe insults. Note, however, that such dramatic changes might rarely be the case in the natural environment yeast cells usually prosper in. It would be highly inefficient and resource-consuming for the yeast cell to fully turn on glycerol production each time osmotic conditions are slightly shifted in the environment. We generally assume that the way in which yeast strains have been stressed in our research lab, as well as in the work of Babazadeh *et al.* [123] might give a slightly distorted picture of the mechanisms that are essential in the yeast osmotic stress response. This way the activation of the glycerol machinery might appear as the core element of the hyperosmotic stress response, whereas under slightly increased external osmolarity the phosphorylation of cytoplasmic targets would already be sufficient for cellular survival. Indeed, it might provide the basis for a fine-tuned physiology to allow an efficient resource-saving reaction to environmental changes. I therefore strongly believe that all identified targets in the confident set, be it nucleic or cytoplasmic, contribute to a well-orchestrated response to hyperosmotic stress.

V. Materials & Methods

Yeast Strain and Plasmid construction. W303-1A SILAC (*lys1::kanMX arg4::kanMX CAN1 leu2-3, 112 trp1-1 ura3-1 ade2-1 his3-11*) *TSL1*-HTBeaq (*Tsl1-HTBeaq::hpxMX*) was obtained by transformation of PCR amplifications of the HTB*eaq* tagging cassette using plasmid *pWR168 (*HTB*eaq::hphMX)* (Reiter *et al.* [79]) and primers *TSL1*-tag-fw: ATGTAAATGGGTTAAACGAACTTT TCACGATCATTTCAAGAATCATTGAAGATTCCGGTTCTGCTGCTAG and *TSL1*-tag-re: GACCTTACGAA GCAAAAATGAATAATAGCTACCCATATACATAATGTATAAGACCTCGAGGCCAGAAGAC into W303-1A SILAC MatA. *Hog1Δ* deletion strains were produced by transformation with deletion cassettes (*hog1::HIS3*). W303-1A SILAC *TSL1*-HTBeaq hog1-*as* was obtained by a cross with strain W303 *hog1-as*.

Cell culture, labeling and harvesting. The yeast *Saccharomyces cerevisiae* strain W303-SILAC was used in this study which contains a lysine- as well as arginine-deficient gene which makes it compatible with the SILAC methodology. SILAC technology was applied as described by Ong *et al.* [84]: Full medium-pre-culture (LT media + 2%glucose + amino acid dropout solution), lacking both lysine and arginine, was divided into two flasks, one containing the ^{12}C-lysine/arginine, the other one ^{13}C-labeled lysine/arginine. To optimize a regular 1:1 ratio for ^{12}C:^{13}C, we performed inoculation at a minimal optical density, letting cultures grow at 30°C to an OD_{600} of 1.0 over 12 generations. Cells were harvested by filtration (0.45µm nitrocellulose membrane used) and frozen in liquid N_2.

Shotgun approach: Yeast cell extract preparation and TRIzol purification. TRIzol purification was performed according to manufacturer's protocol (*Invitrogen*) with a few modifications: Cell pellets were resuspended in TRIzol reagent (*Invitrogen*) and cell breakage was performed by FastPrep treatment (*FP120* shaker: 3 cycles: 45 secs, Level 6) with glass beads. Insoluble material was removed by centrifugation (12.000xg, 15mins, 4°C) and RNA was extracted by the addition of chloroform and subsequent centrifugation (12000xg, 15mins, 4°C). To separate DNA from protein in the remaining organic- and interphase, ethanol was added for 2-3 mins

DNA was sedimented by centrifugation (2000xg, 5mins, 4°C). The supernatant was mixed with isopropanol to precipitate proteins. The protein was washed 3 times with 0.3M guanidine-HCl in 95% ethanol, dried in a speed vac and resuspended in 50mM ammonium bicarbonate (ABC) buffer containing 8M urea. Urea concentration of the extract was shifted to 6M; protein from whole cell extracts (Bradford measurement: 8mg) was subjected to reduction by dithiothreitol (DTT, 160µg) at 56°C for 30 minutes. After alkylation of reduced cysteine thiols with iodoacetamide (IAA, 800µg) for 30 minutes, proteins were digested with LysC (*Roche*, 80µg) for 2 hours at 30°C. Urea concentration was reduced to 0.8M by dilution with 50mM ABC prior to further proteolytic digestion with trypsin (recombinant, proteomics grade, *Roche*; 133µg) overnight. The resulting peptides were desalted using Strata-X 33µm Polymeric Sorbent (8B-S100-TAK columns, *Phenomenex*) using 70% acetonitrile/0.1% formic acid for elution. The samples were then lyophilized for a few days. We kept track of the peptide concentration by regularly measuring aliquots with a UV spectrometer (spectrum range of 240-340nm).

Shotgun approach: Phosphopeptide enrichment and fractionation. To enrich phosphopeptides, samples were incubated with TiO_2 resin [87] and 0.8M phtalic acid solution; the amount of TiO2 resin was adjusted to the peptide concentration (1.25mg of TiO_2 /3.5mg yeast protein extract). Captured phosphopeptides were then eluted with 0.3M ammoniumhydroxide and desalted with C18 Sep-Pak cartridges (*Waters*). Subsequent fractionation was performed using strong cation exchange (SCX) chromatography, either with SCX-StageTips (*Glygen*) or with 1ml Resource S column (*GE healthcare*) installed in a nano-HPLC machine (*Ultimate 3000*, Dionex, Sunnyvale CA, USA). In latter procedure samples were injected using SCX Buffer A (5mM NaH_2PO_4,30% acetonitrile (ACN), pH 2.7). Peptides bound to the column were separated by a linear gradient of sodium dihydrogen phosphate in 30% acetonitrile and 0.1% tri-fluoroacetic acid (TFA). Based upon UV measurements, some fraction containing low amount of peptides were pooled which resulted in 12-20 fractions. Each elution sample was adjusted by TFA to pH 2-3 for subsequent desalting (method by Dr. WL Straube) and mass spectrometry measurement.

Target Validation Approach: HTB tandem affinity purification. Cells were resuspended in a denaturing guanidium buffer (6M guanidine hydrochloride, 50mM Tris-HCl (pH~8.8), 5mM NaF, 0.1% Tween-20, 1mM PMSF and protease inhibitor cocktail (*Roche*, 11 873 580 001)) ;

subsequent breakage was achieved by glass beads and FastPrep-treatment at power level 6 for 3 rounds at 45 seconds. After rotating the samples for 20 minutes at room temperature, the cell debris was removed by high-speed centrifugation (14.000rpm) for 20 minutes. Tandem affinity purifications were performed as follows: cell extracts were incubated for 4h with Ni^{2+}-Sepharose beads (*GE Healthcare*, Buckinhamshire, UK, 17-5318-06). The beads were washed twice with urea buffer (8M urea, 50mM Na-PO_4, 300mM NaCl, 0.01% Tween, pH 8) and three times with urea wash buffer pH 6.3 to diminish unspecific background. The enriched proteins were then eluted in three subsequent elution steps using urea buffer pH 4.3 containing 10mM EDTA, allowing for competitive withdrawal of the histidine tag from Ni^{2+} beads. The combined eluates were then adjusted to a pH of 8 with 2M NaOH and incubated overnight with streptavidin-agarose beads (*Thermo Scientific Pierce*, Rockford, IL., USA, 20349). Beads were washed three times using urea wash buffer containing 1% SDS and three times with a standard urea wash buffer pH 8 to remove any residual SDS before mass spectrometry submission.

Target Validation Approach: Sample Preparation before Mass Spectrometric Analysis. Protein samples were delivered to ammoniumbicarbonate (50mM ABC) buffer for trypsin compatibility purposes; disulfide bridges were reduced with dithiothreitol (DTT, 10% w/w of estimated protein amount) for 30 minutes at 56°C and alkylated with iodacetamide (IAA, 50% w/w of estimated protein amount) for 20-25 minutes in the dark. Proteins were digested with trypsin (recombinant, proteomics grade, *Roche*, 3.3% w/w of the estimated protein amount) overnight at 37°C. Digests were stopped by addition of trifluoroacetic acid (TFA, 10%) to approximately pH 3 for the protease to be inactivated.

Target Validation Approach: titanium dioxide chromatography. If needed, phosphopeptide enrichment using titanium dioxide tips (*Glygen*) was performed [87]: TiO_2 tips were washed with 80% acetonitrile and equilibrated with loading buffer (50mg 2,5-dihydroxybenzoic acid (DHB) and 90mg 1-octanesulfonic acid (OSA) diluted in 500µl 40% acetic acid, with 1µl heptafluorobutyric acid). Samples were diluted 1:1 with 2x loading buffer and then loaded onto the microcolumn; after washing with 80% acetonitrile/0.1% TFA, trapped phosphopeptides were eluted in 50mM ammonia phosphate solution (pH 10.5).

Mass Spectrometric analysis. Analysis of the samples was performed using reversed phase nano-HPLC (*Ultimate 3000*, Dionex, Sunnyvale CA, USA) with a trapping column for peptide

concentration and an analytical column (*PepMap C18*) for peptide separation purposes. Concentration of organic solvent (acetonitrile) was increased from 2.5% to 40% in 0.1% formic acid in 60 minutes at a flow rate of 275nl/min. The HPLC was coupled to an LTQ Orbitrap Velos mass spectrometer (*Thermo Scientific ©*) via a nanoelectrospray ion source (*Proxeon*). The configuration of the instrument for SILAC experiments was set as follows: A full scan is provided by the high-resolution Orbitrap with 12 most intense peaks submitted to further fragmentation into the linear ion trap, resulting in MS2 spectra. Xcalibur software 2.1.0 (*Thermo Scientific©*) was used for data acquisition.

Shotgun approach: Identification and quantification of peptides and proteins using *MaxQuant.* MS raw files were processed using *MaxQuant* software 1.3.0.5 and the recently released *1.5* version in annotated experiments. Multiplicity was set to two, matching the number of SILAC labels used ('light', 'heavy') in each experiment with Arg6/Lys6 signifying the 'heavy' labels. Peaks were searched for by the Andromeda search engine against a concatenated forward and reversed version of the yeast protein database (downloaded from the *Saccharomyces Genome Database www.yeastgenome.org*) containing contaminants, such as human keratin and trypsin. Full enzyme specificity was required and up to two missed cleavages were allowed. Trypsin/P was specified as the enzyme used. Cysteine carbamido-methylation *(+57.021464 Da*) was set as a fixed modification. N-acetylation of protein (*+42.0105656 Da*), methionine oxidation (*+15.994915 Da*), serine/threonine/tyrosine phosphorylation *(+79.966331 Da*) were defined as variable modifications. Labeled lysines and arginines were specified as variable modifications as well. Initial mass tolerance for the precursor was set to 3ppm and fragment ion mass tolerance for the CID data was set to 0.8 Da; for HCD data to 20ppm, respectively. We only allowed an FDR below 1% for the entire dataset both at peptide and protein level based on the performed decoy search using the reversed database. if a peptide was identified but not quantified (corresponding SILAC partner not fragmented) this phenomenon was considered in our final result with "*only Light*"/"*only Heavy*".

Shotgun approach: Identification and quantification of peptides and proteins using *Proteome Discoverer©*.

Additionally to *MaxQuant* searches, we loaded the same raw spectra into a specially designed workflow in the *Proteome Discoverer©* software package (*Thermo Scientific©*) with the same modification settings as in *MaxQuant*, linked to the SEQUEST© search algorithm and to the phosphoRS-module [110]. For the adequate FDR-adjustments (below 1% for the entire dataset), the decoy data was extracted from each search and concatenated with the true target list. We performed q-value calculation and saved resulting target and decoy data (q-value>1%) in a separate SQL-database.

Target Validation Approach: Evaluation of Mass Spectrometry Data. For analysis of accumulated spectra the *Proteome Discoverer* software package (*Thermo Scientific©*) was used, as well as the Qualbrowser Software (*Xcalibur*) for manual validation of peak assignment and peptide identification. Our search was restricted to a yeast protein database (downloaded from the *Saccharomyces Genome Database www.yeastgenome.org*) containing keratin and trypsin. A decoy database of reversed peptide sequences was used to estimate the likelihood that a PSM is correct with the FDR set at 1%. For peptide assignment to generated MS2 spectra we applied the SEQUEST algorithm in the *Proteome Discoverer©* environment, defining a 2 or 3ppm mass deviation (high confidence search) and phosphorylation (+79,96 Da) as a dynamic modification. Relative quantification using *Discoverer Quantifier* tool was performed for preliminary SILAC quantification with 0.5 Da deviation; phosphoRS [110] was applied to validate phosphorylation probability of specific sites. In case of manual double-check, manual validation was performed in the Qualbrowser software using 2ppm mass window centered on the exact precursor mass, for XIC (extracted ion chromatogram) generation.

Normalization of Data. Correction factors the experiments were calculated as the geometric mean of the L/H-ratios for all unphosphorylated peptides that contained either zero, one, two, etc. prolines (necessary correction for the signal loss due to arginine-proline conversion). The ratios measured for the phosphopeptides were divided by the correction factor.

Automatic Data Handling and *Python* Scripts. To handle and save output MS-data of the *Proteome Discoverer©* in a high-throughput manner, we used SQL-databases and *Python* pickle dictionaries. With *Python* scripts, we could then easily extract useful information and compare datasets.

Automatic Decoy Extraction from *.msf*-files (*Xtract.py*). The *Thermo Scientific Proteome Discoverer©* does not openly show the decoy data in the standard user interface. This might complicate posterior FDR-calculations that are not performed in the *Proteome Discoverer©*. The software stores its file information and tables in an SQL database format that is accessible via the *Sqlite*-manager. Without opening *Proteome Discoverer©* it is possible to read *.msf*-file formats and retrieve user-relevant information. The script *Xtract.py* allows saving both target and decoy PSMs into an SQL database, independently of the search engine interface.

Peptide-Retention Time-Analysis (*RTAnalyzer.py*). To analyze individual peptide retention-time behavior on reversed-phase liquid chromatography, another tool that visualizes defined peptides in a scale of their retention times and intensity was developed. Thereby, one could observe the behavior of identical peptides throughout a number of SCX fractions or different biological experiments.

Semi-automatic Spectra Inspection (*mzMLParserPeptides.py* and *mzMLParserMasses.py*). MS-rawfiles were converted to mzML-files via *ProteoWizard© MSConvert*-tool. For the `mass`-operating-mode (*mzMLParserMasses.py*), the SilacAnalyzer-node of the *OpenMS©*- software was applied to search for a specified mass-range in the mzML-file. FeatureXML-file format was chosen to save all resulting peptide elution profiles. Profile extraction was performed using the *XMLParser.py* code. All source codes are available in the *Supplementary Materials*.

Functional enrichment analysis. Gene Ontology (GO) annotation of *S.cerevisiae* from SGD (downloaded from *www.yeastgenome.org*) were retrieved and certain terms were restricted to broader annotations. To test whether specific annotation terms are enriched or depleted in a given set of proteins we normalized against the *S.cerevisiae* proteome background that has been quantified in one experiment.

VI. References

„Ich habe mich bemüht, sämtliche Inhaber der Bildrechte ausfindig zu machen und ihre Zustimmung zur Verwendung der Bilder in dieser Arbeit eingeholt. Sollte dennoch eine Urheberrechtsverletzung bekannt werden, ersuche ich um eine Meldung bei mir.“

[1] Zuzuarregui A., Kupka T., Bhatt B., Dohnal I., Mudrak I., Friedmann C., et al. M-Track: detecting short-lived protein-protein interactions in vivo. Nat Methods 2012, 9(6): 594-596
[2] Marx V., Method of the Year: Targeted proteomics. Nature Methods, 2013, vol.10(1),pp.19-22
[3] Gervais P. and Beney L., Osmotic mass transfer in the yeast Saccharomyces cerevisiae. Cell. Mol. Biol., 2001, vol.47, pp. 831-839
[4] Brocker C., Thompson D.C., Vasiliou, V., The role of hyperosmotic stress in inflammation and disease. Biomol Concepts, 2012, vol.3(4), pp.345-364
[5] http://en.wikipedia.org/wiki/Yeast_in_winemaking
[6] Saito H. and Posas F., Response to Hyperosmotic Stress. Genetics, 2012, vol.192, pp.289-318
[7] Miermont A., Uhlendorf J., McClean M., Hersen P., The Dynamical Systems Properties of the Hog Signaling Cascade. Journal of Signal Transduction, 2011, pp.1-12
[8] O'Rourke S.M., Herkowitz I., O'Shea E.K., Yeast go the whole HOG for the hyperosmotic response. TRENDs in Genetics, 2002, vol.18(8): pp.405-412
[9] Hohmann S., Osmotuc stress signaling and osmoadaptation in yeasts. Microbiol. Mol. Biol. Rev. 2002, vol. 66(2), pp.300-372
[10]: Schaber J., Baltanas R., Bush A., Klipp E. et al. Modelling reveals novel roles of two parallel signalling pathways and homeostatic feedbacks in yeast. Mol Syst Biol. 2012, vol. 8(622)
[11]: Lee, J., W. Reiter, I. Dohnal, C. Gregori, S. Beese-Sims, K. Kuchler, G. Ammerer, and D. E. Levin. MAPK Hog1 closes the S. cerevisiae glycerol channel Fps1 by phosphorylating and dsplacing its positive regulators. Genes & Dev. 2013, vol.27: pp.2590-2601.
[12]: Albertyn J., Hohmann S., Thevelein J.M, Prior B.A.: GPD1, which encodes glycerol-3-phosphate dehydrogenase, is essential for growth under osmotic stress in Saccharomyces cerevisiae, and its expression is regulated by the high-osmolarity glycerol response pathway. Mol Cell Biol. 1994, vol. 14(6), pp.4135-4144
[13]: Pahlman A.K., Granath K., Ansell R., Hohmann S., Adler L., The yeast glycerol 3-phosphatases Gpp1p and Gpp2p are required for glycerol biosynthesis and differentially involved in the cellular responses to osmotic, anaerobic, and oxidative stress. J Biol Chem. 2001, vol. 276(5), pp.3555-3563

[14] Mettetal J.T, Muzzey, D., Gomez-Uribe C., van Oudenaarden A., The Frequency Dependence of Osmoadaptation in Saccharomyces cerevisiae. Science 2008, 319 (5862): 482-484
[15] McMahon, M.A.S, Xu, J., Moore, J.E., Blair, I.S., McDowell, D.A., Environmental Stress and Antibiotic Resistance in Food-Related Pathogens. Appl. Environ. Microbiol. January 2007 vol. 73 no. 1 211-217
[16] Olsen J.V. et al.: Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell, 2006, vol. 127(3), pp.635-648
[17] Manning G.: The protein kinase complement of the human genome. Science, 2002,
[18] S.M. Thomas (Annual Review, 1997): Cellular functions regulated by the SRC family kinases
[19] Alberts, 2005, Fifth Edition
[20] Kültz D. and Chakravarty D.: Hyperosmolarity in the form of elevated NaCl but not urea causes DNA damage in murine kidney cells. Proc. Natl. Acad. Sci. USA, 2001, vol.98(4), pp: 1999-2004
[21] Yancey P.H., Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. The Journal of Experimental Biology, 2005, vol.208, pp. 2819-2830
[22] Roser B., Colaco C. A sweeter way to fresh food. New Sci. 1993, vol. 138, pp.25-28
[23] Clegg J.S. The physical properties and metabolic status of Artemia cysts at low water contents: “The water replacement hypothesis.” In Membranes, Metabolism and Dry Organisms; Leopold A.C., Ed.: Cornell University Press: Ithaca, NY, USA, 1985; pp.169-187
[24] Crow J.H., Carpenter J.F., Crowe L.M. The role of vitrification in anhydrobiosis. Annu. Rev. Physiol. 1998, vol.60, pp. 73-103
[25] Green J.L., Angell. C.A., Phase relations and vitrification in saccharide-water solution and the trehalose anomaly. J.Phys.Chem. 1989, vol. 93, pp.2880-2882
[26] Blomberg A., Adler L., Roles of glycerol and glycerol-3-phosphate dehydrogenase (NAD+) in acquired osmotolerance of Saccharomyces cerevisiae. J. Bacteriol. 1989, vol. 171, pp.1087-1092
[27] Ferreira C., van Voorst F., Martins A., et al. A member of the sugar transporter family, Stl1p is the glycerol/H+ symporter in Saccharomyces cerevisiae. Mol Biol Cell, 2005, vol.16, pp.2068-2076
[28] Bell, W., P. Klaassen, M. Ohnacker, T. Boller, M. Herweijer, P. Schoppink, P. Van der Zee, and A. Wiemken. 1992. Characterization of the 56-kDa subunit of yeast trehalose-6-phosphate synthase and cloning of its gene reveal its identity with the product of CIF1, a regulator of carbon catabolite inactivation. Eur. J. Biochem. 209:951–959.
[29] Parrou, J. L., M. A. Teste, and J. Francois. 1997. Effects of various types of stress on the metabolism of reserve carbohydrates in Saccharomyces cerevisiae: genetic evidence for a stress-induced recycling of glycogen and trehalose. Microbiology 143: pp.1891–1900.
[30] Neves, M. J., and J. Francois. 1992. On the mechanism by which a heat shock induces trehalose accumulation in Saccharomyces cerevisiae. Biochem. J. 288: pp. 859–864.

[31] Nwaka, S., and H. Holzer. 1998. Molecular biology of trehalose and the trehalases in the yeast Saccharomyces cerevisiae. Prog. Nucleic Acid Res. Mol. Biol. 58:197–237.
[32] Diamant, S., N. Eliahu, D. Rosenthal, and P. Goloubinoff. Chemical chaperones regulate molecular chaperones in vitro and in cells under combined salt and heat stresses. J. Biol. Chem., 2001,vol. 276: pp.39586–39591.
[33] Sebollela, A., P. R. Louzada, M. Sola-Penna, V. Sarone-Williams, T. Coelho- Sampaio, and S. T. Ferreira. 2004. Inhibition of yeast glutathione reductase by trehalose: possible implications in yeast survival and recovery from stress. Int. J. Biochem. Cell Biol. 36:900–908.
[34] Singer, M. A., and S. Lindquist. 1998. Multiple effects of trehalose on protein folding in vitro and in vivo. Mol. Cell 1: pp. 639–648.
[35] Viner, R. I., and J. S. Clegg. 2001. Influence of trehalose on the molecular chaperone activity of p26, a small heat shock/alpha-crystallin protein. Cell Stress Chaperones 6: pp. 126–135.
[36] Kimura Y., Kawasaki S., Tuchimoto R., Tanaka N., Trehalose biosynthesis in Myxococcus xanthus under osmotic stress and during spore formation. J. Biochem 2014, vol. 155(1), pp.17-24
[37] Nelson, 2006: The Natural Osmolyte Trehalose Is a Positive Regulator of the Heat-Induced Activity of Yeast Heat Shock Transcription Factor
[38] Chen Z., Gibson T.B., Robinson F., Silvestro L., et al. MAP kinases. Chem. Rev. 2001, vol. 101, pp.2449-2476
[39] Schwartz M.A., Madhani H.D., Principles of MAP kinase signaling specificity in Saccharomyces cerevisiae. Annu. Rev. Genetics, 2004, vol.38: pp. 725-748
[40] Behar M., Dohlman H.G., and Elston T.C., Kinetic insulation as an effective mechanism for achieving pathway specificity in intracellular signaling networks. Proc. Natl. Acad. Sci. USA, vol. 104, pp. 16146-16151
[41] Furukawa K., Furukawa T., Hohmann S., Efficient construction of homozygous diploid strains identifies genes required fort he hyper-filamentous phenotype in Saccharomyces cerevisiae. PLoS One, 2011, vol. 6
[42] M.C. Gustin, J. Albertyn, M. Alexander and K. Davenport, “MAP kinase pathway in the yeast Saccharomyces cerevisiae.” Microbiology and Molecular Biology Reviews, vol. 66, no.2, pp. 300-372,2002
[43] K. R. Davenport, M. Sohaskey, Y. Kamada, D.E.Levin and M.C. Gustin, “A second osmosensing signal transduction pathway in yeast: hypotonic shock activates the PKC1 protein kinase-regulated cell integrity pathway.” Journal of Biological Chemistry, vol. 270, no.50, pp. 30157-30161, 1995
[44] H. Martín, J.M. Rodriguez-Pachon, C. Ruiz, C.Nombela and M.Molina, “Regulatory mechanisms for modulation of signaling through the cell integrity Slt2-mediated pathway in Saccharomyces cerevisiae.” Journal of Biological Chemistry, vol. 275, no.2, pp. 1511-1519, 2000
[45] K. Tedford, S.Kim, D. Sa, K.Stevens, and M. Tyers, “Regulation of the mating pheromone and invasive growth responses in yeast by two MAP kinase substrates.” Current Biology, vol. 7, no. 4, pp. 228-238,1997.

[46] F. Banuett, “Signaling in the yeast: an informational cascade with links to the filamentous fungi.” Microbiology and Molecular Biology Reviews, vol. 62, no. 2, pp. 249-274, 1998.
[47] E. A. Elion, “Pheromone response, mating and cell biology.” Current Opinion in Microbiology, vol. 3, no. 6, pp. 573-581, 2000.
[48] Posas, F. and H. Saito, Activation of the yeast SSK2 MAP kinase kinase kinase by the SSK1 two-component response regulator. EMBO J, 1998. 17(5): p. 1385-94.
[49] Posas F. et al., Yeast HOG1 MAP kinase cascade is regulated by a multistep phosphorelay mechanism in the SLN1-YPD1-SSK1 “two-component” osmosensor. Cell, 1996. 86(6): p. 865-875
[50] J. Macia, S.Regot, T. Peeters, N.Conde, R.Solé, and F. Posas, “Dynamic signaling in the Hog1 MAPK pathway relies on high basal signal transduction.” Science Signaling, vol. 2, no. 63, p.ra13, 2009
[51] Tatebayashi, K., et al., Transmembrane mucins Hkr1 and Msb2 are putative osmosensors in the SHO1 branch of yeast HOG pathway. EMBO J., 2007. 26(15): p. 3521-3533
[52] Lamson, R.E., M.J. Winters, and P.M. Pryciak, “Cdc42 regulation of kinase activity and signaling by the yeast p21-activated kinase Ste20.” Mol Cell Biol, 2002. 22(9): p.2939-2951
[53] Yamamoto, K. et al., “Dynamic control of yeast MAP kinase network by induced association and dissociation between the Ste50 scaffold and the Opy2 membrane anchor.” Mol. Cell., 2010. 40(1): p.87-98
[54]: Lamson R.E., S.Takahasi, M.J.Winters, and P.M.Pryciak, Dual role for membrane localization in yeast MAP kinase cascade activation and its contribution to signaling fidelity. Curr. Biology, 2006, vol. 16: pp. 618-623
[55] S.J. Mansour, K.A. Resing, J.M. Candi et al. “Mitogen-activated protein (MAP) kinase phosphorylation of MAP kinase kinase: determination of phosphorylation sites by mass spectrometry and site-directed mutagenesis.” Journal of Biochemistry, vol. 116, no.2, pp.304-314, 1994
[56] Posas, F. and Saito, H. (1997): Osmotic activation of the HOG MAPK pathway via Ste11p MAPKKK: scaffold role of Pbs2p MAPKK. Science 276, 1702-1705
[57] E. Klipp, B. Nordlander, R. Krüger, P. Gennemark, and S. Hohmann, „Integrative model of the response of yeast to osmotic shock.” Nature Biotechnology, vol. 23, no. 8, pp. 975-982, 2005.
[58] A. P. Gasch, P.T. Spellman, C.M. Kao et al., „Genomic expression programs in the response of yeast cells to environmental changes.” Molecular Biology of the Cell, vol. 11, no. 12, pp. 4241-4257, 2000.
[59] Hao, N., M. Behar, S.C. Parnell, M.P. Torres, C.H. Borchers et al., 2007: A systems-biology analysis of feedback inhibition in the Sho1 osmotic-stress-response pathway. Curr. Biol. 17: 659-667
[60] Albertyn J, Hohmann S, Thevelein JM, Prior BA (1994) GPD1, which encodes glycerol-3-phosphate dehydrogenase, is essential for growth under osmotic stress in Saccharomyces

cerevisiae, and its expression is regulated by the high-osmolarity glycerol response pathway. Mol Cell Biol 14:4135-4144

[61] Krantz M, Nordlander B, Valadi H, Johansson M, Gustafsson L, Hohmann S (2004) Anaerobicity prepares Saccharomyces cerevisiae cells for faster adaptation to osmotic shock. Eukaryot Cell 3:1381-1390

[62] Miller, C., B. Schwalb, K. Maier, D. Schulz, S. Dümcke et al., 2011. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Mol. Syst. Biology.7: 458

[63] M. Proft, A. Pascual-Ahuir, E. de Nadal, J. Arno, R. Serrano, and F. Posas, "Regulation of the Sko1 transcriptional repressor by the Hog1 MAP kinase in response to osmotic stress," The EMBO Journal, vol. 20, no. 5, pp.1123-1113, 2001.

[64] P. M. Alepuz, E. de Nadal, M. Zapater, G. Ammerer and F. Posas, "Osmostress-induced transcription by Hot1 depends on a Hog1-mediated recruitment of the RNA Pol II", The EMBO Journal, vol. 22, no. 10, pp. 2433-2442, 2003

[65] M. Rep, M.Krantz, J.M. Thevelein and S. Hohmann, "The transcriptional response of S. cerevisiae to osmotic shock. Hot1p and Msn2p/Msn4p are required for the induction of subsets of high osmolarity glycerol pathway-dependent genes." Journal of Biological Chemistry, vol. 275, no. 12, pp.8290-8300,2000.

[66] Proft M, Struhl K (2002) Hog1 kinase converts the Sko1-Cyc8-Tup1 repressor complex into an activator that recruits SAGA and SWI/SNF in response to osmotic stress. Mol Cell 9:1307-1317

[67] Kobayashi, Y., T. Inai, M. Mizunuma, I. Okada, A. Shitamukai et al., 2008. Identification of Tup1 and Cyc8 mutations defective in the responses to osmotic stress. Biochem. Biophys. Res. Commun 368, 50-55

[68] de Nadal E., G. Ammerer and F. Posas, Controlling gene expression in response to stress. Nat. Rev. Genet. 2011, vol.12, pp.833-845

[69] Mas, G., E. de Nadal, R.Dechant, M.L.Rodriguez de la Concepcion, G. Logie et al., 2009. Recruitment of a chromatin remodeling complex by the Hog1 MAP kinase to stress genes. EMBO J.28: 326-336

[70] Proft. M, G. Mas, E.de Nadal, A. Vendrell, N. Noriega et al., 2006. The stress-activated Hog1 kinase is a selective transcriptional elongation factor for genes responding to osmotic stress. Mol. Cell 23: 241-250.

[71] de Nadal, E. M. Zapater, P.M. Alepuz, L. Sumoy, G.Mas et al., 2004. The MAPK Hog1 recruits Rpd3 histone deacetylase to activate osmoresponsive genes. Nature 427: 370-374.

[72] Warringer J., M.Hult, S.Regot, F. Posas and P. Sunnerhagen, 2010. The HOG pathway dictates the short-term translational response after hyperosmotic shock. Mol. Biol. Cell 21: 3080-3092

[73] Westfall P. J., J.C. Patterson, R. E. Chen and J. Thorner, 2008. Stress resistance and signal fidelity independent of nuclear MAPK function. Proc. Natl. Acad. Sci. USA 105: 12212-12217

[74] Teige M., E. Scheikl, V. Reiser, H. Ruis and G. Ammerer, 2001. Rck2, a member of the calmodulin-protein kinase family, links protein synthesis to high osmolarity MAP kinase signaling in budding yeast. Proc. Natl. Acad. Sci. USA, vol.98, no.10, pp. 5625-5630

[75] Yakoov G., Duch M., Garcia-Rubio J., Clotet J., Jimenez et al., The stress-activated protein kinase Hog1 mediates S-phase delay in response to osmostress. Mol. Biol. Chem. 2009, vol. 20, pp.3572-3582
[76] Van Wuytswinkel O, Reiser V, Siderius M, Kelders MC, Ammerer G, Ruis H, Mager WH (2000) Response of Saccharomyces cerevisiae to severe osmotic stress: evidence for a novel activation mechanism of the HOG MAP kinase pathway. Mol Microbiol 37:382-397
[77] Proft M, Struhl K (2004) MAP kinase-mediated stress relief that precedes and regulates the timing of transcriptional induction. Cell 118:351-361
[78] Clotet J., and F. Posas. Control of cell cycle in response to osmostress: lessons from yeast. Methods Enzymol. 2007, vol.428, pp.63-76
[79] Reiter W., Anrather D., Dohnal I., Pichler P., Veis J., Grotli M., Posas F. and G.Ammerer, Validation of regulated protein phosphorylation events in yeast by quantitative mass spectrometry analysis of purified proteins. Proteomics 2012, vol. 12, pp. 1-14
[80] Aebersold R., M. Mann, Mass spectrometry-based proteomics. Nature Reviews 2003, vol. 422, 198-207
[81] Mann M., Ong, S.E., Gronborg M., Steen H., Jensen O.N., Pandey A., Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. Trends in Biotechnology 2002, vol. 20/6, 261-268
[82] Eyrich B., Sickmann A., Zahedi R.P, Catch me if you can: mass-spectrometry based phosphoproteomics and quantification strategies. Proteomics 2011, vol. 11, 554-570
[83] Bantscheff M., Schirle M., Sweetman G., Rick J., Küster B., Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem 2007, vol. 289, 1017-1031
[84] Ong S.E., Blagoev B., Kratchmarova I., Kristensen D.B., et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 2002, vol. 1, 376-386
[85] Tagwerker C., Flick K., Cui M., Guerrero C. et al., A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivo cross-linking. Mol Cell Proteomics 2008, vol. 5, 737-748
[86] Tagwerker C., Zhang H, Wang X., Larsen L.S., et al., HB tag modules for PCR-based gene tagging and tandem affinity purification in Saccharomyces cerevisiae. Yeast 2006, vol. 23, 623-632
[87] Thingholm T.E., Jörgensen T.J., Jensen O.N., Larsen M.R. Highly selective enrichment of phosphorylated peptides using titanium dioxide. Nat Protoc. 2006, 1(4): 1929-1935
[88] Bodenmiller B., Wanka S., Kraft C., Urban J., et al. Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. Sci Signal. 2010, 3 (153)
[89] Soufi B., Kelstrup C.D., Stoehr G., Fröhlich F., Walther T.C., Olsen J.V., Global Analysis of the yeast osmotic stress response by quantitative proteomics. Mol Biosyst. 2009, 5(11): 1337-1346

[90] Támas MJ, Karlgren S, Bill RM, Hedfalk K, Allegri L, Ferreira M, Thevelein JM, Rydström J, Mullins JGL, Hohmann S. A short regulatory domain restricts glycerol transport through yeast Fps1p. J Biol Chem. 2003;278:6337–6345
[91] Olsen J.V, M.Mann, Status of Large-Scale Analysis of Post-translational Modifications by Mass Spectrometry. December 1, 2013 Molecular & Cellular Proteomics, 12, 3444-3452.
[92] Steen H., Mann M., The ABC's (and XYZ's) of peptide sequencing. Nature Reviews 2004, vol.5, 1-15
[93] Edelmann M.J., Strong Cation Exchange Chromatography in Analysis of Posttranslational Modifications: Innovations and Perspectives. Journal of Biomedicin and Biotechnology 2011
[94] Matthiesen, R., Mass Spectrometry Data Analysis in Proteomics. Methods in Molecular Biology. 2007
[95] Elias J.E., Gygi S.P., Target-Decoy Search Strategy for Mass Spectrometry-Based Proteomics. Methods Mol Biol. 2010; 604: 55-71
[96] Moore R.E., Young M.K., Lee T.D. Qscore: an algorithm for evaluating SEQUEST database search results. J. Am. Soc. Mass Spectrom. 2002;13:378–386.
[97] Käll L., Storey, J.D., MacCoss M.J., Noble W.S., Assigning Significance to Peptides Identified by Tandem Mass Spectrometry Using Decoy Databases. Journal of Proteome Research, 2008. Vol. 7., 29-34
[98] Storey J.D., Tibshirani R. Statistical significance for genome-wide studies. Proc. Natl. Acad. Sci. USA, 2003, vol. 100, pp. 9440-9445
[99] Olsen J.V., Blagoev B., Gnad F., Macek B., Kumar C., Global In Vivo and Site-Specific Phosphorylation Dynamics in Signaling Networks. Cell, 2006, vol. 127, pp.635-648
[100] Nita-Lazar A., Saito-Benz, H., White F.M., Quantitative phosphoproteomics by mass spectrometry: Past, present and future. Proteomics 2008, vol. 8, pp.4433-4443
[101] Mann M., Functional and quantitative proteomics using SILAC. Nature, 2006, vol. 7, pp.1725-1735
[102] Gregan J., Zhang C., Rumpf C., Cipak L., et al. Construction of conditional analog-sensitive kinase alleles in the fission yeast Schizosaccharomyces pombe. Nat Protoc., 2007, vol. 2 (11), pp.2996-3000
[103] Klein M., Morillas M., Vendreli A., et al. Design, Synthesis and Characterization of a Highly Effective Inhibitor for Analog-Sensitive (as) Kinases. PLOS One, 2011, vol. 6(6)
[104] Maeda T., M. Takekawa, and H. Saito, Activation of yeast component system that regulates an osmosensing MAP kinase cascade in yeast. Nature 1995, vol.369, pp.242-245
[105] Wurgler-Murphys S.M., T. Maeda, E.A. Witten, and H.Saito, Regulation of the Saccharomyces cerevisiae HOG1 mitogen-activated protein kinase by the PTP2 and PTP3 protein tyrosine phosphatases. Mol.Cell.Biol. 1997, vol.(17), pp.1289-1297
[106] Maeda, T., M. Takekawa, and H. Saito. Activation of yeast PBS2 MAPKK by MAPKKKs or by binding of an SH3-containing osmosensor. Science 1995, vol.269:pp.554-558.
[107] Oda Y., Huang K., Chait B.T. et al., Accurate quantitation of protein expression and site-specific phosphorylation. Proc. Natl. Acad. Sci. USA, 1999, vol. 96(12), pp.6591-6596
[108] Blethrow J., Zhang C., Shokat K.M., Weiss E.L. Design and use of analog-sensitive protein kinases. Curr. Protoc. Mol. Biol. 2004, vol.18(11)

[109] Pitoniak A., Birkaya B., Dionne H. M., Vadaie N., Cullen P. J.,The signaling mucins Msb2 and Hkr1 differentially regulate the filamentation mitogen-activated protein kinase pathway and contribute to a multimodal response. Mol. Biol. Cell, 2009, vol. 20: pp. 3101–3114
[110] Wepf A., Glatter T., Schmidt A., Aebersold R., Gstaiger M., Quantitative interaction proteomics using mass spectrometry. Nat. Methods. (2009), vol. 6(3): pp.203-205
[111] Taus T., Köcher T., Pichler P., Pascke C., Schmidt A., et al., Universal and confident phosphorylation site localization using phosphoRS. J. Proteome Res. 2011, vol. 10(12): pp.5354-5362
[112] Lee Y., Jeschke G.R., Roelants F. et al., Reciprocal Phosphorylation of Yeast Glycerol-3-Phosphate Dehydrogenases in Adaptation to Distinct Types of Stress. Mol. Cell. Biol. 2012, vol. 32 (22): pp. 4705-4717
[113] Reiter W., Klopf, E., De Wever V., et al. Yeast protein phosphatase 2A-Cdc55 regulates the transcriptional response to hyperosmolarity stress by regulating Msn2 and Msn4-chromatin recruitment. Mol Cell Biol. (2013), vol. 33(5): pp.1057-1072
[114] Jedrchowski M.P., Huttlin E.L, Gygi S., Evaluation of HCD- and CID-type Fragmentation Within their respective Detection Platforms for Murine Phosphoproteomics. Mol. Cell Proteomics (2011), vol. 10(12)
[115] Regot S., de Nadal E., Rodriguez-Navarro S., Gonzalez-Novo A., et al., The Hog1 stress-activated protein kinase targets nucleoporins to control mRNA export upon stress. J Biol.Chem (2013), vol. 288(24): pp. 17384-17398
[116] Murray, K.K., Boyd, R. K., Eberlin, M.N., Langley, G., Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013). Pure and Applied Chemistry (2013): 1.
[117] Merchan, S., Bernal D., Serrano R., Yenush L., Response of the Saccharomyces cerevisiae Mpk1 mitogen-activated protein kinase pathway to increase in internal turgor pressure cause by loss of Ppz protein phosphatases. Eukaryot. Cell 3 (2004): pp.100-107
[118] Ruiz A., del Carmen Ruiz M., et al. The Ppz protein phosphatases regulate Trk-independent potassium influx in yeast. FEBS Lett.578 (2004): pp.58-62
[119] Ruiz A., Yenush L, Arino J., Regulation of ENA1 Na+-ATPase gene expression by the Ppz1 protein phosphatase is mediated by the calcineurin pathway. Eukaryot. Cell 2 (2003): pp.937-948
[120] Yenush L, Mulet J.M, Arino J. and Serrano R., The Ppz protein phosphatases are key regulators of K+ and pH homeostasis: implications for salt tolerance, cell wall integrity and cell cycle progression. EMBO J.21(2002):pp.920-929
[121] Yenush L., Merchan S., Holmes J., Serrano R., pH-Responsive, Posttranslational Regulation of the Trk1 Potassium Transporter by the Type 1-Related Ppz1 Phosphatase. Mol. and Cell. Biology. 25 (2005): pp.8683-8692
[122] Haro R., Rodriguez-Navarro A., Molecular analysis of the mechanism of potassium uptake through the TRK1 transporter of Saccharomyces cerevisiae. Biochim. Biophys. Acta. 1564 (2002): pp. 114-122
[123] Babazadeh, R., Furukawa, T., Hohmann S. and Furukawa, K., Rewiring yeast osmostress signaling through the MAPK network reveals essential and non-essential roles of Hog1 in osmoadaptation. Nature Scientific Reports (2014), vol.4

VII. Appendix

VII.1. Description of Supplementary Materials

Databases available:

- *HTB-purification database (SQL-format)*
- *Shotgun MS- database for Hog1 Screen (SQL-format)*

Tables:

- Supplementary Table I: Hog1 Dataset Analyzed with *Bioworks©*
- Supplementary Table II: Reevaluated Data 2007 (*Proteome Discoverer©)*
- Supplementary Table III: PSG-data generated by new shotgun experiments
- Supplementary Table IV: confident set of Hog1 targets in detailed analysis
- Supplementary Table V: list of indirect Hog1-dependent targets

Python Source Codes:

- *XMLParser.py*
- *Xtract.py*
- *mzMLParserPeptides.py*
- *mzMLParserMasses.py*
- *RTAnalyzer.py*

VII.2. Abstract

The environment of a living cell is defined by a set of physico-chemical and biological parameters, such as temperature, pH, oxygen exposure, nutrient availability and quality, etc. Since these parameters can significantly fluctuate, cells have evolved specialized signaling pathways to respond appropriately to any occurring changes in the environmental conditions. Simple eukaryotes such as the yeast *Saccharomyces cerevisiae* have a wide range of sensors and signaling proteins which initiate processes that re-establish cellular homeostasis in unfavorable or stressful environmental situations. Sensors and signaling proteins are commonly organized in a distinct hierarchy in various signaling cascades. Often an intracellular signaling pathway becomes activated upon external stimuli resulting in adjustments of cellular processes at different levels, ranging from transcriptional regulation of particular genes to post-translational fine-tuning of protein activity. The timing of these particular responses is additionally controlled by feedback loops.

While alterations of the global gene expression profile usually provide the protein set necessary to conquer environmental insults in the long run, transcription *per se* might have fewer implications for the immediate cell survival under stress conditions. It is rather the modulation of activity through post-translational modifications (PTMs) of pre-existing protein factors that gives rise to immediate cellular responses. One particularly wide-spread and important modification is the addition of a phosphate moiety to a particular amino acid residue, termed phosphorylation. This PTM is involved in a number of signaling pathways in bacteria, as well as in eukaryotes.

Although detection of post-translational changes is crucial for the understanding of many cellular responses, a large-scale description of these regulatory events remains challenging. Only recent advances in quantitative mass spectrometry (MS) allow monitoring thousands of modified sites at the same time and to assess whether specific modifying enzymes target certain sites.

In this work we present an approach to study protein regulation at the level of post-translational modifications (PTMs) in yeast in response to hyperosmotic shock. In particular, we tried to identify targets of the main regulatory kinase of the hyperosmotic-stress-response, namely the mitogen-activated protein kinase (MAPK) Hog1. To specify the targets

we integrate two MS shotgun datasets: one showing changes in the phosphorylation pattern that occur upon hyperosmotic stress and a second one showing the effect on these phosphorylation patterns if the MAPK Hog1 is specifically inhibited. We expect possible Hog1 targets to produce a characteristic behavior in their phosphorylation pattern within the two MS datasets. In a second line of evidence, a direct interaction of the MAPK Hog1 with some of these potential substrates was confirmed using a protein-protein interaction assay. Interesting candidates were further analyzed via an MS target validation approach to increase sequence coverage and complete the phosphorylation map. As an example I present results obtained from the analysis of the large subunit of the trehalose synthase complex (Tsl1): The protein contains three Hog1 motifs that get phosphorylated upon osmotic disbalance, suggesting that the disaccharide trehalose might play a role in the first seconds of the osmotic stress response in yeast.

VII.3. Zusammenfassung

Alle Zellen überwachen (und reagieren entsprechend) auf Veränderungen in den extrazellulären Parametern, wie Osmolarität, Temperatur, pH-Wert aber auch auf Schwankungen in der Nährstoffsituation. Eine solche Veränderung verursacht häufig die Aktivierung spezialisierter Signaltransduktionskaskaden. Eukaryoten wie z.B. die Hefe *Saccharomyces cerevisiae* haben eine breite Palette an Sensoren und Signalproteinen, die Prozesse initiieren um das zelluläre Gleichgewicht in Stresssituation wieder in Ordnung zu bringen. Diese Sensoren und Signalproteine sind in typischen Signalkaskaden in einer hierarchischen Ordnung gegliedert. Sobald eine Signalkaskade aktiviert wird, führt dies zu weiteren zellulären Ereignissen, wie Transkription und post-translationale Proteinmodifizierung zur Steuerung von Proteinaktivität. Mitspieler der Kaskade sind über Feedback-Loops miteinander verflochten.

Obzwar die Veränderung der Genexpression eine wesentliche und vor allem dauerhafte Rolle in der zellulären Reaktion auf hyperosmotischen Stress spielt, könnte die Transkription *per se* wenig Einfluss auf das unmittelbare Überleben der Zelle unter gestressten Bedingungen haben. Vielmehr ist es die Modulierung der Proteinaktivität durch post-translationale Modifikationen (PTMs), die die unmittelbare Reaktion auf Stress ermöglicht. Eine besonders verbreitetete Modifizierung, die über Proteinkinasen und Phosphatasen gesteuert wird, ist die Phosphorylierung. Die Identifizierung der post-translationalen Modifikationsstellen an möglichen Zielproteinen und auch die Erfassung der Modifikationskinetik stellt oft eine besondere Herausforderung dar. Erst die jüngsten Entwicklungen in der quantitativen Massenspektrometrie (MS)-Analytik ermöglichen es tausende von Phosphorylierungen und eventuelle Site-Abhängigkeiten von modifizierenden Enzymen festzustellen.

In dieser Arbeit präsentiere ich eine Vorgehensweise Proteinregulation in de Bäckerhefe auf Reaktion von extrazellulären hyperosmotischen Stress über post-translationale Modifikationen (PTM) zu erfassen. Es ging vor allem darum Targetproteine zu identifizieren, die über die mitogen-aktivierte Proteinkinase (MAPK) Hog1 gesteuert werden. Um diese zu spezifizieren, integrierten wir zwei MS-Datensätze: Während ein Datensatz im Allgemeinen Veränderungen im Phosphorylierungsmuster in hyperosmotischen Stress aufzeigt, beinhaltet der andere Datensatz jene Veränderungen im Phosphorylierungsmuster, die duch Inhibition

der MAPK Hog1 hervorgerufen werden. Für ein sicheres Proteintarget von Hog1 erwarten wir uns ein charakteristisches Verhalten des betroffenen Phosphorylierungsmusters in beiden Datensätzen. Weiters testen wir die extrahierten Proteintargets auf Interaktion mit der MAPK über ein Protein-Protein Interaktionsassay. Um eine möglichst vollständige Sequenzabdeckung zu erreichen und alle regulierten Modifikationsstellen eines Proteins abzudecken, werden interessante Kandidaten im Zuge einer „Targeted Proteomics"-Strategie analysiert. So präsentiere ich in dieser Arbeit Tsl1, die große Untereinheit der Trehalose Synthase in der Hefe: Das Protein weist drei Hog1-Motive auf, die unter hyperosmotischen Bedingungen stark phosphoryliert werden, was eventuell auf eine Rolle in der Hog1-gesteuerten zellulären Stressreaktion hinweist.

VII.4. Acknowledgements

In undertaking my master project I was privileged to work with, and enjoy the support of a vast number of colleagues and friends. I would like to take this opportunity to gratefully acknowledge their contributions to the accomplishment of this- I daresay- mighty tome.

My heartfelt thanks to Prof. Gustav Ammerer who has given me the opportunity to work on this exciting project and many others as well. Fortunately for me, he did not push me into a strict working framework, but let me try out various things, from experimental bench-work to actual MS-bioinformatics. I clearly benefited from this multifaceted-view on science and of course from gaining access to such a powerful technology such as mass spectrometry. Gustav always had an open ear for my problems and needs and would give me advises and considerations which I tried to implement with good faith in all my projects. I very much hope that this work will satisfy his expectations.

The key figure who helped me shaping this thesis was without doubt Dr. Wolfgang Reiter. It would be an understatement if I wouldn't thank him for everything. As my guide and teacher he supported me in my scientific projects and helped me to get my thoughts structured. I'm confident that I would have never developed such a scientific attitude without him being my immediate supervisor. Moreover, I appreciate Wolfi as a dear friend of mine, who also made me aware of my failures.

Deserving of special mention is of course Dr. Dorothea Anrather who was training me in the field of mass spectrometry. In the beginning this training was characterized by the phrase "Just look into the raw-file!", but obviously practice makes perfect. I hope that this master thesis also includes some computational implementations she always dreamt of having at hand. Regardless of the scientific aspect, I cherish our friendship, which has developed not only over the analysis of spectra but one glass of wine or two.

I would also like to thank David Hollenstein who supported me with the programming challenges and inspired me during regular discussions at the daily coffee break.

I am grateful to the enjoyable environment provided by my colleagues from the MS-facility, Markus, Sonja and Thomas, and the research group of Prof. Ammerer, Jiri, Syam and Jessica. Thank you for the daily doses of humor, inspiration and the great deal of scientific support. I

would also like to thank members of the IMP-MS-facility, who supported me with MS-measurement arrangements and MS-related advices.

Finally I would like to thank everyone who has gone through the actual endeavor of reading this long-winded thesis in a critical manner. I also thank my family for the emotional support they provided during the writing procedure.

VII.5. Curriculum Vitae: Natalie Romanov, BSc.

Personal Data:

Birthday: 27th December, 1990

Nationality: Austrian

Research-Relevant Employment History:

2012-2014	Technician at Mass Spectrometry Facility in the research lab of Gustav Ammerer

Research-Relevant Education:

2001-2009	Bundesrealgymnasium Perchtoldsdorf
2009-2011	Bachelor studies of Molecular Biology at the University of Vienna • Bachelor Thesis: "Construction of a database for the analysis of the link between protein structure properties and their mRNAs' pyrimidine content " (supervisor: Dr. Bojan Zagrovic) • Passed with distinction (received performance scholarship in 2010 and 2011)
2011-2014	Master studies of Biochemistry at the University of Vienna • Master Thesis: "Screening for Novel Targets of the MAPK Hog1 using Quantitative Proteomics" (supervisor: Prof. Gustav Ammerer)

Teaching & Mentoring:

2013	Tutoring Experience in student courses held by Dr. Wolfgang Reiter at the VBC (Vienna Biocenter, Vienna) • course: "Advanced Genomics and Proteomics in Yeast" • course: "Yeast Genetic and Proteomic Tools"

International Meetings:

2013	Poster presentation at the UnicellSys Network Meeting in Innsbruck