

TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Proteomik und Bioanalytik

# Unraveling signal transduction networks by high-resolution and quantitative mass spectrometry-based proteomics

Maximiliane Hilger

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. D. Langosch

Prüfer der Dissertation:

1. Univ.-Prof. Dr. B. Küster
2. Hon.-Prof. Dr. M. Mann  
(Ludwig-Maximilians-Universität München)
3. Priv.-Doz. Dr. H. Daub

Die Dissertation wurde am 01.09.2010 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 03.11.2010 angenommen.



For my parents



# Table of Contents

Abbreviations .....	7
Prologue of the thesis .....	9
<b>1 Introduction.....</b>	<b>11</b>
1.1 General principles of cell signaling.....	11
1.2 Fundamentals of high-resolution and quantitative mass spectrometry-based proteomics .....	13
1.2.1 Generic workflow of MS-based proteomics .....	15
1.2.2 Quantitative proteomics.....	17
1.2.3 Bioinformatic analysis of quantitative MS data .....	19
<b>2 Global phosphoproteome analysis.....</b>	<b>23</b>
2.1 Mapping of endogenous Ser/Thr/Tyr phosphorylation in <i>Drosophila</i> SL2 cells .....	25
2.2 Global phosphoproteome response to Ptp61F phosphatase knock-down.....	29
<b>3 Interaction proteomics.....</b>	<b>35</b>
3.1 Stimulus specific interaction screens in the Wnt pathway.....	36
<b>4 Perspectives.....</b>	<b>43</b>
<b>5 References.....</b>	<b>45</b>
<b>6 Acknowledgements.....</b>	<b>49</b>
<b>7 Appendix.....</b>	<b>51</b>
7.1 Overview of publications and unpublished projects.....	51
7.2 Publication: A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics .....	53
7.3 Publication: Systems-wide analysis of a phosphatase knock-down by quantitative proteomics and phosphoproteomics.....	63
7.4 Publication: Phosphorylation of SUMO-1 occurs in vivo and is conserved through evolution. ....	79
7.5 Publication: Triple SILAC to determine stimulus specific interactions in the Wnt pathway .....	89
7.6 Curriculum Vitae.....	113



## Abbreviations

Abi	Abl interaction protein
Abl	Abelson
AP	Affinity Purification
APC	Adenomatous Polyposis Coli
BAC	Bacterial Artificial Chromosome
CID	Collisional Induced Dissociation
CK	Casein Kinase
ds	double-stranded
DVL	Disheveled
ECD	Electron Capture Dissociation
EGF	Epidermal Growth Factor
EmPAI	Exponentially modified Protein Abundance Index
ESI	ElectroSpray Ionization
ETD	Electron Transfer Dissociation
FDR	False Discovery Rate
FTICR	Fourier Transform Ion Cyclotron Resonance
FZD	Frizzled
GFP	Green Fluorescent Protein
GO	Gene Ontology
GSK	Glycogen Synthase Kinase
GST	Glutathione S-Transferase
H	Heavy
HCD	Higher-energy Collisional Dissociation
HPLC	High Performance Liquid Chromatography
IMAC	Immobilized Metal Affinity Chromatography
IRMPD	InfraRed MultiPhoton Dissociation
IT	Ion Trap
iTRAQ	isobaric Tag for Relative and Absolute Quantitation
JAK	Janus Kinase
KO	Knock-Out
L	Light
LTQ	Linear Trap Quadrupole
M	Medium
m/z	mass to charge ratio
MALDI	Matrix-Assisted Laser Desorption/Ionization

## Abbreviations

---

MRM	Multiple Reaction Monitoring
MS	Mass Spectrometry
MS/MS	Tandem mass spectrometry
MSA	MultiStage Activation
MS <sup>n</sup>	Multiple MS
p	phospho
PEP	Posterior Error Probability
ppm	parts per million
PQD	Pulsed-Q Dissociation
PTM	Post Translational Modifications
Ptp	Protein tyrosine phosphatase
QQQ	Triple Quadrupole
QUBIC	QUantitative BAC InteraCtomics
RNAi	RNA interference
RTK	Protein Tyrosine Kinase
SCX	Strong Cation eXchange chromatography
Ser	Serine
SILAC	Stable Isotope Labeling by Amino acids in Cell culture
SL2	Schneider Line 2
STAT	Signal Transducer and Activator of Transcription
SUMO	Small Ubiquitin like MODifier
TAP	Tandem Affinity Purification
Thr	Threonine
TMT	Tandem Mass Tag
TOF	Time Of Flight
Tyr	Tyrosine
XIC	eXtracted Ion Current
Y	Tyrosine



## Prologue of the thesis

The overall aim of this thesis was to establish and apply quantitative proteomic, phosphoproteomic and interaction proteomic techniques to address questions in cell signaling.

Phosphorylation of proteins is one of the key mechanisms to transduce signals. While a number of phosphorylation sites have already been very well characterized, the total complexity of phosphorylation, including their number, dynamics and functionality, is far from being completely understood. Global and unbiased analysis of phosphorylation sites by mass spectrometry (MS)-based proteomics can help to fill this gap. Contributing to this goal, a global Ser/Thr/Tyr phosphoproteome of the model organism *Drosophila melanogaster* was characterized in the first project of this thesis. The obtained high-quality SILAC-based phosphoproteome of *Drosophila* Schneider Line 2 (SL2) cells was the basis for a comparison with a human phosphoproteome<sup>1</sup>. The established phosphoproteome analysis in SL2 cells served as starting point for the second thesis project, which focused on the combination of quantitative phosphoproteomics and proteomics with RNA interference (RNAi). In contrast to mammalian systems, transient knock-downs in SL2 cells are very efficient and easy to scale up, which is a prerequisite for global phosphoproteome analysis<sup>2</sup>. The combined approach has the potential of systems-wide knock-down analysis of any protein of interest without the need for laborious generation of gene knock-out (KO) cell lines. In this study, the knock-down of a protein tyrosine phosphatase (Ptp), Ptp61F, was analyzed as a proof of principle. Phosphatases can be crucial for signal termination, which is just as important as their initiation by kinases, but their targets are much less known.

The third project aimed to investigate stimulus specific interactions in the Wnt pathway by a triple SILAC labeling approach. Canonical Wnt signaling is of great interest due to its pivotal role in development, tissue homeostasis and cancer formation<sup>3-7</sup>. Although it is an intensively studied signaling pathway, the dynamics of the main regulating complex are still not fully understood. System-wide and hypothesis-free MS analysis shed light on this complex and at the same time allowed identification of novel components involved in the Wnt signaling.

These three main topics of this cumulative thesis are summarized with respect to their classification into global phosphoproteome and interaction proteomics projects. Side projects are mentioned in the main text and listed in the appendix of the thesis (7.1).

# 1 Introduction

## 1.1 General principles of cell signaling

Communication with the environment is crucial for all organisms to coordinate cellular functions and is mediated by numerous signaling pathways. Their dysregulation is implicated in various diseases such as cancer, inflammation and diabetes. Therefore, in-depth analysis of cell signaling is a large and important research area<sup>8</sup>.

Although the nature of signals and their processing can be very diverse they share general characteristics. Signals need to be specific to lead to a defined response. They must be rapid to initiate, easy to mobilize and reversible. Most signals are registered at the plasma membrane, transmitted into the cell by a signaling cascade until they reach the destination for their final effects. This could for example be the cytoplasm regulating glycogen metabolism or the nucleus controlling gene expression. During their transduction, signals are amplified to generate a large response from a small initial signal<sup>9</sup>. In my research I studied the canonical Wnt pathway, as an example of a ligand-receptor signaling pathway (Figure 1). In this pathway, the signal is initiated by binding of a Wnt ligand to both Frizzled (FZD) and LRP5/6 receptors<sup>3-7</sup>. Consequently, phosphorylation and degradation of  $\beta$ -Catenin, which is the key transcriptional regulator of the pathway, is inhibited. Stabilized  $\beta$ -Catenin accumulates in the cytoplasm and the nucleus, where it eventually coactivates TCF/LEF mediated gene expression<sup>3-7</sup>. The illustration of canonical Wnt signaling in Figure 1 is a typical example for a simplified view of signaling as individual pathways. But in reality signaling is much more complicated. One signal can activate multiple signaling pathways and can have several, diverse outcomes. Different signals can be transmitted via shared signaling components not necessarily leading to the same effect. Pathways can also regulate their own activity by feedback regulation. Moreover, different pathways can interact with each other (“crosstalk”). Consequently, cell signaling is best viewed as complex networks that contains signal divergence, amplification and convergence rather than as individual, linearly pathways<sup>9</sup>.

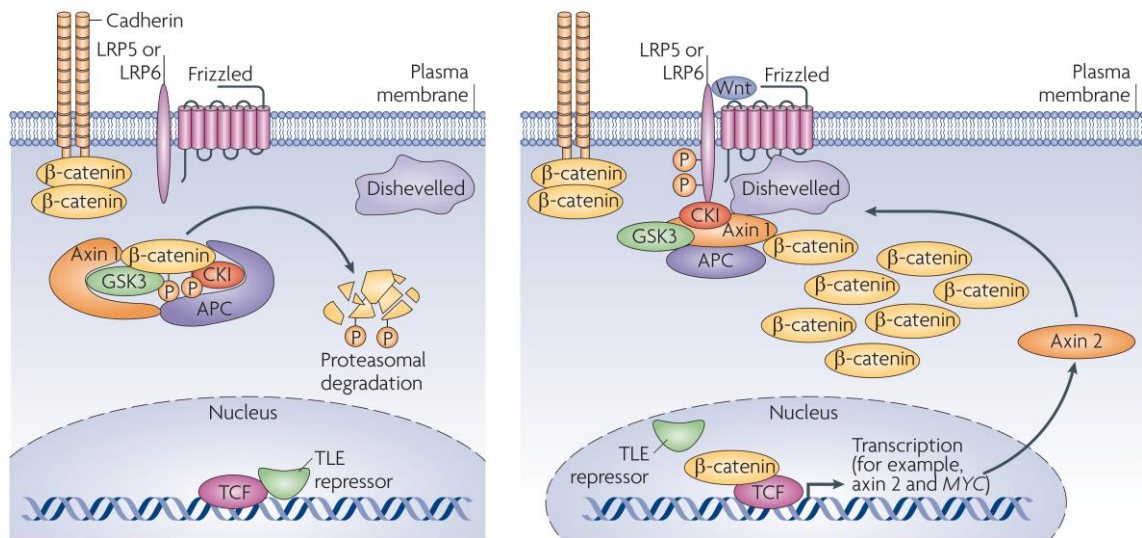


Figure 1 **A selected example of a signaling pathway: canonical Wnt signaling.** The regulation of  $\beta$ -Catenin stability is the key step of the canonical Wnt pathway<sup>3-7</sup>. In the absence of a Wnt ligand,  $\beta$ -Catenin levels are kept low.  $\beta$ -Catenin is continuously phosphorylated within the so called destruction complex that consists of the core components APC and Axin-1, acting as scaffolds, and the kinases GSK-3 $\beta$  and CKI- $\alpha$ . Phosphorylated  $\beta$ -Catenin is ubiquitylated and subsequently degraded by the proteasome. Upon binding of a Wnt ligand to Frizzled and LRP5/6 receptors, function of the destruction complex is disabled in part through relocalization to the plasma membrane and interaction with phosphorylated Dishevelled (DVL). Consequently,  $\beta$ -Catenin accumulates in the cytoplasm and translocated to the nucleus. Here it replaces TLE repressors from TCF/LEF transcription factors and, together with them, coactivates Wnt target genes. From reference<sup>10</sup>.

The activity of signaling network components can be regulated by various mechanisms. Among them are signal induced binding of activators or inhibitors, conformation or localization changes and covalent posttranslational modifications (PTMs) such as Ser, Thr and Tyr phosphorylation, Lys acetylation, Lys and Arg methylation or Lys ubiquitylation. A PTM or the combination of multiple PTMs can control subcellular protein localization, protein activity or protein stability. As an example of a PTM driven mechanism, phosphorylation of  $\beta$ -Catenin within the destruction complex leads to its ubiquitylation and proteasomal degradation in the non-activated Wnt pathway (Figure 1). The destruction complex contains the scaffold proteins APC (Adenomatous Polyposis Coli) and Axin-1 and the kinases glycogen synthase kinase 3 $\beta$  (GSK-3 $\beta$ ) and casein kinase I- $\alpha$  (CKI- $\alpha$ ). CKI- $\alpha$  is the priming kinase that phosphorylates S45 on  $\beta$ -Catenin and thus enables phosphorylation on T41, S37 and S33 by GSK-3 $\beta$ <sup>11</sup>.

PTMs exert their functions by modulating the conformation of signaling molecules or by creating docking sites for upstream and downstream effectors. These proteins recognize PTMs

by specific domains or modules<sup>12,13</sup>. For example, Tyr protein phosphorylation is recognized by SH2 and PTB domains and can induce inter- and intra-molecular interactions. To act as regulating tools, PTMs need to be reversible. Thus, enzymes that add or remove PTMs - like kinases and phosphatases do with phosphorylation - are important members of signaling networks<sup>9,14</sup>.

The study of signal transduction aims to detect signal-induced PTMs, regulated protein-protein interactions and downstream protein expression changes. Traditionally protein or PTM specific antibodies were used almost exclusively for this purpose. However, in the last few years mass spectrometry has become a powerful technology for studying cell signaling. In contrast to antibody-based technologies mass spectrometry is unbiased in that it permits a global system-wide view and enables accurate quantification of proteins and PTMs.

## **1.2 Fundamentals of high-resolution and quantitative mass spectrometry-based proteomics**

Mass spectrometry became a generally useful tool for the analysis of biomolecules in the late 1980s through the invention of the gentle ionization techniques electrospray ionization (ESI)<sup>15</sup> and matrix-assisted laser desorption/ionization (MALDI)<sup>16</sup>. For a long time, proteins were separated by two-dimensional gel electrophoresis and were individually digested and analyzed. Low throughput and low dynamic range allowed the identification of only a few hundred proteins at most, mainly the highest abundant ones. The analysis of complex peptide mixtures - which is a mainstay of MS-based proteomics today - was hampered by the low resolution of available ion trap mass spectrometers. Peptides of similar mass could not be resolved, which led to low identification rates. During the last decades advances in mass spectrometric instrumentation have yielded constantly improved mass accuracy, resolving power, novel fragmentation methods and sequencing speed. Together these developments constituted a breakthrough for the analysis of complex peptide mixtures<sup>17-21</sup> (Table 1). Nevertheless, the dynamic range and the sequencing speed of mass spectrometers still limit the achievable depth of proteome analysis.

**Table 1 Characteristics and performances of commonly used types of mass spectrometers.** Parentheses indicate an optional feature. +, ++, and +++ indicate possible or moderate, good or high, and excellent or very high, respectively. Ion trap (IT), triple quadrupole (QQQ), Fourier transform ion cyclotron resonance (FTICR), multiple MS ( $MS^n$ ), multiple reaction monitoring (MRM), electron capture dissociation (ECD), infrared multiphoton dissociation (IRMPD), higher-energy collisional dissociation (HCD), pulsed-Q dissociation (PQD), electron transfer dissociation (ETD). Adapted from reference<sup>20</sup>.

	IT	QQQ	QQ-TOF	TOF-TOF	LTQ-FTICR	LTQ-Orbitrap	LTQ-Orbitrap Velos
Mass accuracy	Low	Medium	Good	Good	Excellent	Excellent	Excellent
Resolution	Low	Low	Good	High	Very high	Very high	Very high
Sensitivity	Good	High	High	High	Medium	Good	High
Dynamic range	Low	Medium	Medium	Medium	Medium	Medium	Medium
ESI	Yes	Yes	Yes	No	Yes	Yes	Yes
MALDI	(Yes)	No	(Yes)	Yes	(Yes)	(Yes)	(Yes)
MS/MS capabilities	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional capabilities	$MS^n$	Neutral loss, MRM			$MS^n$ , ECD, IRMPD	$MS^n$ , HCD, PQD, (ETD)	$MS^n$ , HCD, PQD, (ETD)
Identification	++	+	++	++	+++	+++	+++
Quantification	+	+++	+++	++	+++	+++	+++
Throughput	+++	++	++	+++	++	+++	+++
PTM detection	+	+	+	+	+	++	+++

To apply MS analysis to biological questions, it is not sufficient to identify peptides and proteins, but it is rather necessary to quantify them. MS analysis is not intrinsically quantitative, therefore development of various methods that enable protein quantification constitute major advances in the field<sup>22-24</sup> (see 1.2.2 for details). Furthermore, the development of software tools for automated data analysis including peptide and protein identification and quantification have revolutionized accuracy, reproducibility and speed of the data analysis<sup>25</sup> (1.2.3).

Together, these achievements allow now rapid and robust analysis of thousands of proteins and even complete detection of small proteomes such as that of *Saccharomyces cerevisiae*<sup>26</sup>. Furthermore, large scale analysis of PTMs<sup>27-29</sup> and interactome studies combining mass spectrometry with affinity purifications (AP-MS)<sup>30-35</sup> can routinely be performed in a quantitative manner. The projects of this thesis span all of these different applications of MS-based proteomics for global analysis of signaling networks.

### 1.2.1 Generic workflow of MS-based proteomics

Today, proteins from complex mixtures are typically analyzed by bottom-up proteomics (also called shotgun proteomics). In this approach, peptides rather than intact proteins are measured by the mass spectrometer. Peptides, in comparison to whole proteins, are easier to solubilize and separate and the interpretation of MS and MS/MS spectra from peptides is easier and less error-prone. Moreover, the identification of few peptides is generally sufficient to unambiguously identify a protein.

The generic workflow of bottom-up proteomics (Figure 2) starts with protein extraction from cells or tissues. Proteins are digested into peptides by sequence specific endoproteases such as trypsin, LysC, GluC or AspN. Trypsin is the most commonly used protease and specifically cleaves C-terminal to arginines and lysines<sup>36</sup>. In an acidic environment these peptides are protonated not only at the N-terminal amine, but also at the C-terminal arginine or lysine leading to favorable ionization and fragmentation characteristics. Peptide mixtures are separated online with the mass spectrometer. Typically nanoflow C<sub>18</sub>-reversed-phase high performance liquid chromatography (HPLC) is used. The quality of the chromatography strongly influences the number of identified proteins and is dependent on the correct selection of bead size and column length. Eluting peptides are electrosprayed into the mass spectrometer. A fine spray of charged droplets is generated in the presence of a strong electric field. As the neutral solvent evaporates and the droplet size decreases, the charge density on the surface increases. Droplets are dispersed by charge repulsion and released, desolvated ions are directed via an orifice into the mass spectrometer<sup>15</sup>. In the MS scan, mass to charge ratio ( $m/z$ ) and relative intensity of coeluting peptides are measured. To gain sequence information, peptides are isolated and fragmented (which happens mainly at the peptide bonds) and the  $m/z$  ratio and intensity of the fragment ions are detected (MS/MS scan). Fragment ion information and parent ion mass are used to assign the corresponding amino acid sequence of the peptide. This is done by searching against an *in-silico* digested protein database that contains all proteins of the organism of interest. Sophisticated software tools enable automated analysis of the peptide sequences and mapping of identified peptides to their corresponding proteins (1.2.3).

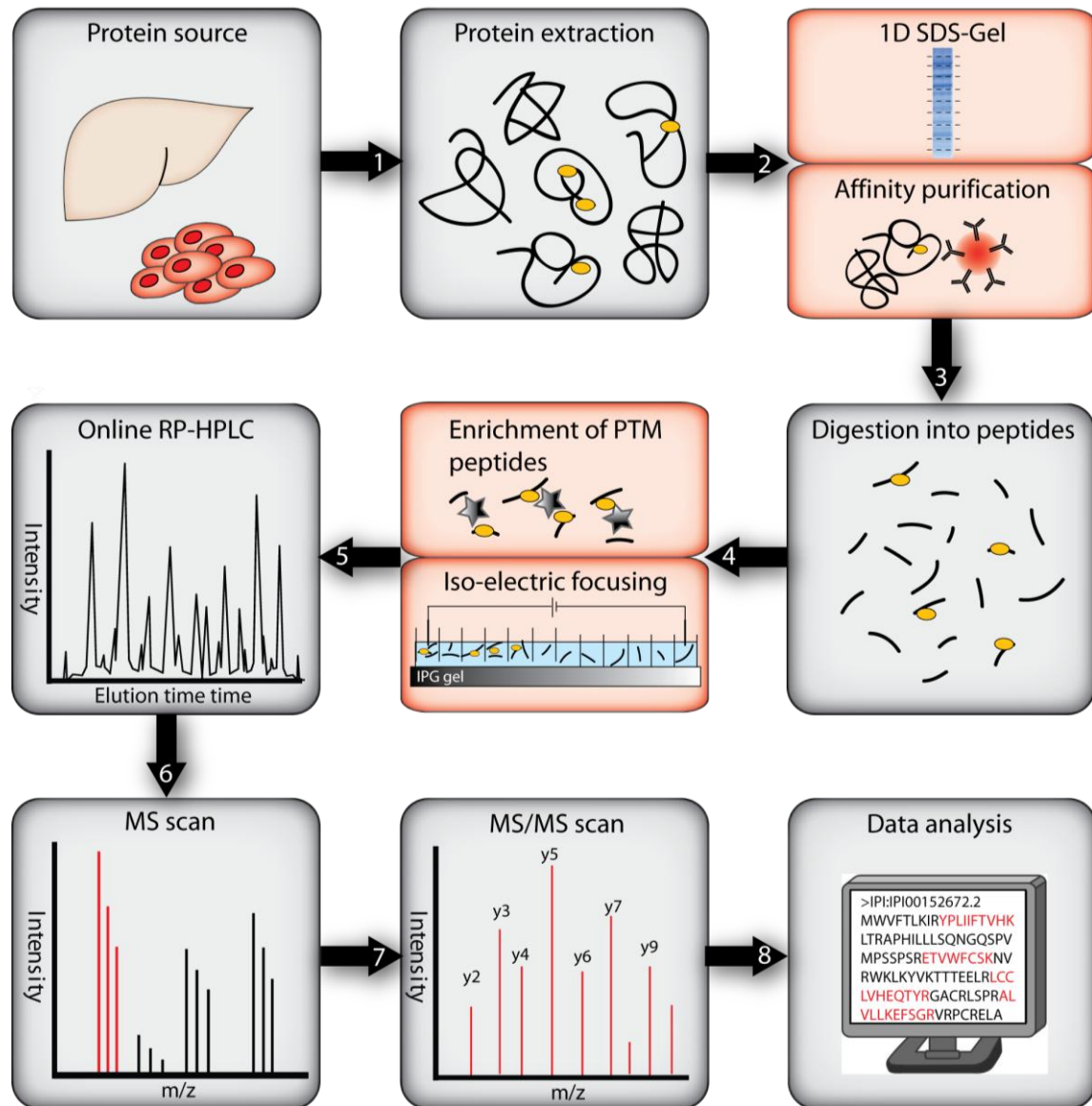


Figure 2 **Schematic representation of a typical workflow of a bottom-up proteomics experiment.** The main steps are protein extraction (1), optionally protein enrichment or separation (2), protein digestion (3), optionally offline peptide fractionation or enrichment (4), online HPLC peptide separation (5), MS (6) and tandem MS (MS/MS) acquisition (7) and data analysis for protein identification and quantification (8).

In addition to this basic procedure further optional steps can be included- depicted as red boxes in Figure 2. To reduce sample complexity, protein fractionation such as one-dimensional SDS gel-electrophoresis<sup>37</sup> or peptide fractionation such as isoelectric focusing<sup>38,39</sup> can be performed. However, a high degree of fractionation quickly inflates the required measuring



time. In interaction studies, the protein of interest and its interactors are enriched by affinity purification after protein extraction. For the global analysis of modified subproteomes such as the phosphoproteome or acetylome, modified peptides are specifically enriched after in solution protein digestion<sup>40-43</sup>.

The two types of mass spectrometers that are most commonly used for bottom-up proteomic applications are actually hybrid instruments: the linear ion trap-Orbitrap (Linear Trap Quadrupole (LTQ)-Orbitrap)<sup>17</sup> and the quadrupole time of flight (TOF) mass analyzers<sup>44</sup>. TOF mass analyzers separate ions spatially and in time. The time ions need to reach the detector is proportional to the square root of their  $m/z$ . The Orbitrap detects the frequency of peptide ions oscillating around a central electrode in an electric field. This frequency is mass dependent and thus the mass spectrum can be calculated from the frequency spectrum by Fourier transformation. TOF and Orbitrap analyzers require pulsed injection of ions that are isolated and accumulated in a quadrupole (or other multi-pole) or an ion trap, respectively. Compared to the TOF instruments, the Orbitrap is superior in mass accuracy (even lower parts per million (ppm)-range) and resolution (routinely 60.000 versus 15.000, although higher values can be achieved by both instrument types). MS/MS scans in TOF instruments involve peptide isolation in a first quadrupole, fragmentation within a second quadrupole used as collision cell and the separation of the fragments in the TOF section. The LTQ-Orbitrap allocates the acquisition of MS and MS/MS to different parts of the hybrid instrument and thereby optimally exploits the complementary features of its coupled mass analyzers. The MS scan is performed in the Orbitrap, which excels in terms of mass accuracy and resolution. The linear ion trap, which is a highly sensitive and very fast mass analyzer, is used for the consecutive MS/MS scans of the typically five to ten most intense peptides. The spatially separated acquisition of MS and MS/MS can be performed in parallel, which moreover results in shorter scan cycle times.

### 1.2.2 Quantitative proteomics

Functional proteomic experiments go beyond proteome mapping and require accurate comparison of different physiological states. However, mass spectrometry is not inherently quantitative since the size of the MS signal is greatly influenced by peptide properties such as size, charge and hydrophobicity. Therefore, generally speaking, only the same peptides can be

compared within or between different samples. Most quantitative MS methods therefore involve stable isotope labeling of proteins or peptides to add a mass tag that can be detected in the mass spectrometer. Differentially labeled samples are analyzed together and are quantified by direct comparison of the corresponding relative signal abundances. This can be either by chemical labeling of digested peptides such as TMT (tandem mass tag)<sup>45</sup> or iTRAQ (isobaric tag for relative and absolute quantitation)<sup>24</sup>. Alternatively, the mass tag can be introduced by metabolic labeling of proteins such as in the <sup>15</sup>N labeling method<sup>46</sup> or the SILAC (stable isotope labeling by amino acids in cell culture) technology<sup>23</sup>.

The iTRAQ method employs isobaric tags such as N-hydroxysuccinimide ester derivatives to label peptide N-termini and the epsilon-amino group of lysine residues. Quantification is based on reporter ions that are generated by fragmentation of the tags during the MS/MS scan. The major advantage of iTRAQ as of other chemical labeling methods is its universal applicability to any type of protein source, cells, tissues and body fluids. Furthermore, complex experimental setups like multi time course experiments are supported by the ability to compare up to 8 samples in the same run. Multi-plex labeling is feasible because differentially iTRAQ labeled peptides have the same mass and do not increase sample complexity. But quantification in the MS/MS also has major disadvantages that limit accurate quantification. It is only based on single or few data points rather than on scans across the whole HPLC peak (1.2.3). Reporter ion signals from co-eluting HPLC peaks cannot be distinguished in the MS/MS spectra and therefore represent a possible error source. Furthermore, quantification accuracy of chemical labeling techniques is in general complicated by potentially varying peptide labeling efficiencies that can result in incomplete labeling.

In the SILAC approach, essential amino acids are substituted by <sup>13</sup>C and <sup>15</sup>N derivatives in the culture media that are incorporated into all proteins during cell growth and division<sup>23,47,48</sup>. Typically lysine and arginine derivatives are used, so that all tryptic peptides - apart from the C-terminal one - carry at least one label. The differentially labeled peptides are distinguished by their defined mass difference in the MS spectrum and are quantified by comparing their relative signal intensities. SILAC is commonly considered the most accurate quantification technique. This is because labeling of the proteins during their synthesis circumvents problems with labeling efficiencies and it allows early mixing of the differential labeled samples, which reduces

quantitation variability. Furthermore, quantification based on the MS scan allows integration of multiple data points per peptide. The major disadvantage of the SILAC method is that labeling is restricted to dividing cells that can be adapted to SILAC media, which contain dialyzed serum to prevent incorporation of non labeled serum amino acids. Therefore, non-dividing primary cells and tissues are generally excluded from SILAC labeling as are dividing cells that require specific growth factors that are missing in dialyzed serum. This problem has been partially circumvented by using an internal SILAC standard generated from one or more SILAC labeled cell lines<sup>49,50</sup>. Moreover, whole animals like mouse<sup>51</sup> and fly<sup>52</sup> have been SILAC labeled and can be applied as internal standards for quantification. An additional limitation of SILAC is the restriction to maximally triple labeling (five-plexing has been reported<sup>53</sup> but is not practical due to isotope overlap). Consequently, multi-plex comparisons need to be split into different experiments sharing a common sample for normalization.

For some applications it can be sufficient to use less accurate quantification methods, such as label-free quantitation methods. Label-free quantitation methods either compare intensities based on the extracted ion current (XIC)<sup>54,55</sup> or number of fragmentation events from individual peptides between non-labeled and separately analyzed samples<sup>56,57</sup>. Without the need for labeling this method is the easiest to implement and is applicable to all protein sources of interest. For example in interaction studies, where affinity purification typically leads to very high protein ratios, label-free quantification can be successfully applied. Spectral counting<sup>56,57</sup> and the related emPAI (exponentially modified protein abundance index) method<sup>58</sup> are even less accurate but are nevertheless still preferable to purely qualitative results.

### **1.2.3 Bioinformatic analysis of quantitative MS data**

Automated data analysis is a prerequisite for MS-based proteomics as a high-throughput technique with the potential of future applications to systems biology. Important steps of MS data analysis are peptide identification and quantification, assembly of this information at the protein level and statistical estimation of false positive identifications as well as determination of significant outliers in quantitative distributions.

For peptide identification, fragment masses and intensities are extracted from the MS/MS scans. Together with the parent peptide mass, derived from the accompanying full scans, they are searched against a sequence database. Depending on the chosen labeling approach, quantification is obtained by comparison of the corresponding signal intensities in the full scan (e.g. for SILAC), between full scans (e.g. for label-free) or within the MS/MS scan (e.g. for iTRAQ). Alternatively, as mentioned above, the number of fragmentation events per peptide can be used to estimate their abundance, referred to as spectral counting<sup>56,57</sup>.

High resolution mass spectrometry and SILAC-based quantification was used for all experiments in this thesis. MaxQuant is a software suit specifically developed for the analysis of this kind of mass spectrometric data<sup>25</sup>. Recently, a novel label-free quantification algorithm has been implemented as well. MaxQuant features fully automated protein identification and quantification using Mascot as the search engine. Recently, a new search engine, termed Andromeda, has been developed and incorporated into the newest versions of MaxQuant. Andromeda, however, has not yet been used in this thesis.

MaxQuant makes use of the high resolution and mass accuracy of the full scan and starts with the detection of isotope clusters and SILAC peptide pairs as three-dimensional objects in  $m/z$  and elution time plane before peptide identification. The peptide mass is precisely determined from all full scans belonging to the 3D peaks as intensity weighted average of all peak centroids. Nonlinear mass recalibration improves mass accuracy to the sub-ppm range. SILAC pairs are quantified considering the complete 3D peak clusters. Those peptides that are not detected as SILAC pairs before the database search can still be re-quantified afterwards. This is often necessary for very extreme ratios where one SILAC partner has a very low signal. Peptide ratios are normalized to eliminate systematic errors e.g. mixing errors by shifting the median of logarithmized ratios to zero. Here it is assumed that most peptide ratios are unchanged and that the bulk of them follow a normal distribution. However, this has to be verified individually for each data set. The Mascot search engine is used to generate a list of peptide candidates for each MS/MS spectrum. A priori knowledge about the SILAC state for many peptides allows searching them with fixed SILAC amino acid modifications, which decreases the search space. MaxQuant uses the information about the SILAC state and the individual peptide mass errors in addition to the pure raw identification score for selecting the

peptide sequence. Together with the high mass precision this enables MaxQuant to reach very high identification rates of typically more than 70% of MS/MS spectra on SILAC peptide pairs. The peptide false discovery rate is estimated by searching against a concatenated target-decoy database (reviewed in <sup>59</sup>). This database contains all normal database entries and their reversed sequence counterparts. By default a 1% false discovery rate (FDR) is applied, which means the peptide list is cut at 1% reverse hits. For this purpose, peptides are ranked according to their individual peptide error probability (posterior error probability (PEP)), which is dependent on the identification score and the peptide length. Identified peptides are further assembled into proteins. If all peptides of one protein are shared with another protein entry in the data base, they are merged into one 'protein group'. Thereby overestimation of protein identifications is prevented. The list of proteins is sorted according to the protein PEP and by default cut at 1% false positives. The protein PEP is calculated as the product of the individual peptide posterior error probabilities. (Note that protein PEPs do not need to be accurate as they are only used to order the protein list.) The protein ratios are calculated as the median of the SILAC peptide ratios; a required minimum number of SILAC ratio counts improve quantification accuracy. The outlier significance calculation follows box plot statistics ("significance A"). Optionally it can be calculated separately for intensity bins ("significance B") to take into account that higher abundant proteins can be quantified more precisely.

Experiments that need to be compared should be analyzed together to achieve the same protein grouping and to reliably control the overall FDR. Quantification of the individual experiments can be maintained or merged within the same MaxQuant analysis. We described in a step by step protocol how to use MaxQuant including its features, output tables and possible error messages with the title: "A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics<sup>60</sup>" (7.2).

The goal of this publication was to document the practical procedure of the freely available MaxQuant software for the analysis of quantitative proteomics data. The protocol aims to guide MaxQuant users through the computational workflow that is required to analyze large-scale double or triple labeling SILAC-based mass spectrometric data acquired with high resolution.

It describes stepwise and in great detail how to obtain the result output from the raw data including the hardware prerequisites. Figure 1 of the paper (reference<sup>60</sup> (7.2)) gives an overview about the MaxQuant workflow. After the data acquisition, which is done by the vendor software of the mass spectrometer, in this case Xcalibur, the “Quant” module of MaxQuant does all possible tasks that can be done without knowing the peptide identity such as the SILAC pair detection and quantification. As next step a search engine performs the MS/MS ion search and associates sequences to fragmentation spectra. The “identify” module subsequently selects the peptide identity and assembles the peptide into protein information. The obtained output files can be browsed and used for downstream bioinformatics. The described MaxQuant version supports LTQ-Orbitrap and LTQ-FT-ICR data and uses Mascot as search engine. It is capable to deal with more than 1,000 LC-MS/MS runs in a single project. Two example raw data sets from already published studies and corresponding result files are provided with the publication, which should allow validating the proper software usage by the MaxQuant user.

The protocol moreover explains how to deal with experiments with pre-fractionation and complex experimental designs, such as replicate experiments with optional interchanged SILAC labels and time course experiments. Furthermore additional information for the identified proteins such as Gene Ontology, KEGG pathway information and Pfam protein domain information is integrated into the output tables and discussed in the protocol.

## 2 Global phosphoproteome analysis

Phosphorylation is the most studied PTM in cell signaling. It is involved in the regulation of most processes of cellular life such as cell growth, proliferation, differentiation, gene expression and metabolism<sup>61</sup>. This highly dynamic and reversible modification is coordinately regulated by kinases that add phosphorylation and phosphatases that in turn remove it. Both enzymes act together in signaling networks to modulate signal intensity and duration. The phosphorylation status of a protein can determine not only its activity but also its localization, stability and interactions. Moreover, many proteins have multiple phosphorylation sites that can have independent, opposing or synergistic effects. Manifesting its global importance in signaling networks, phosphorylation occurs at all levels: at the receptor, within the signaling cascade and at the final destination of the signal. Early estimates predicted that about one-third of mammalian proteins can be phosphorylated<sup>61</sup>. Strikingly, a recent phosphoproteome study of the cell cycle reported a much higher percentage of 70 %<sup>62</sup>. Nevertheless, it still remains elusive how many phosphorylation sites are functionally important.

Individual phosphorylation sites are often studied in candidate-based approaches by *in-vitro* kinase assays or by immune techniques employing phospho-specific antibodies. In contrast, application of MS for the detection of phosphorylation sites benefits from its global and unbiased nature as mentioned for proteome analysis in general. In MS analysis, phosphorylation is detected by its 79.99 Da mass difference in the MS spectra compared to the unmodified peptide and localization of the phosphorylation sites with single amino acid resolution can be derived from the MS/MS spectra. Ser and Thr phosphorylated peptides have the tendency to fragment at the labile phosphoester bond during collisional induced dissociation (CID) in the ion trap. In that case, the loss of phosphoric acid ( $H_3PO_4$ ) from the parent ion (neutral loss of 97.977 Da) often gives rise to the most dominant fragment in the MS/MS spectra, while the informative fragment ions are suppressed or absent<sup>63</sup>. Special fragmentation strategies such as multistage activation (MSA) lead to the fragmentation of the neutral loss peak by additional consecutive activation steps at m/z values pre-calculated to consider one or more neutral losses and different charge stages<sup>64</sup>. Alternatively, ECD and ETD fragmentation methods can be used

that mainly lead to fragmentation in the peptide backbone leaving many post-translational modifications such as phosphorylation intact<sup>65-67</sup>.

Regulatory phosphorylation is often found on proteins with low expression levels and in addition at low stoichiometry, meaning that only a minor amount of the protein is phosphorylated. Therefore, MS analysis of phosphorylation requires efficient phosphopeptide enrichment methods. The most commonly used procedure for global analysis of Ser, Thr and Tyr phosphorylation sites are strong cation exchange chromatography (SCX)<sup>63</sup> combined with metal affinity complexation of the phospho group with titanium dioxide (TiO<sub>2</sub>)<sup>42,43</sup> or immobilized metal affinity chromatography (IMAC)<sup>40</sup>. However, there are also many other phospho-enrichment methods<sup>68</sup>. SCX separates phosphopeptides from non-phosphorylated peptides based on their net charge difference. Most acidified (pH 2.7) tryptic peptides are protonated at the N-terminus and the C-terminal arginine or lysine leading to a net charge of +2. If such peptides are singly phosphorylated, the net charge will be reduced to +1. Consequently, they elute at less stringent elution conditions (lower salt concentration) from the negatively charged resin (e.g. methyl sulfonate). For the same reasons, multiply phosphorylated peptides are not positively charged and are therefore generally contained in the flow-through fraction. As peptides with acidic residues behave similarly in SCX, a second but more specific purification step is necessary. TiO<sub>2</sub>, for example, binds phosphopeptides with higher affinity than any non-phosphopeptide<sup>42,43</sup>. Competing agents such as 2,5-dihydroxybenzoic acid are used to further reduce binding of acidic, non-phosphopeptides<sup>42</sup>. As the two-step phosphopeptide enrichment procedure requires relatively high protein amounts (>10 mg), direct TiO<sub>2</sub> enrichment from in-gel or in-solution digestions can instead be performed for smaller sample amounts.

Quantification of phosphorylation sites is usually based on only one or a few phosphopeptide ratio measurements, making quantification less robust than in the proteome measurements. Therefore, application of accurate quantification methods such as SILAC is beneficial.

Global phosphoproteome analysis can be divided into qualitative phosphoproteome mapping of cells, tissues or body fluids and quantitative, functional phosphoproteome analysis.



In the latter, the emphasis is on the investigation of signal transduction processes e.g. after stimulation of cell signaling or perturbation of a signaling network. The first project of this thesis is an example for the first category, in which the *Drosophila* phosphoproteome is mapped. The second project is an example for a functional phosphoproteome study that characterizes the effect of the knock-down of a phosphatase. Both projects were combined and published with the title “Systems-wide analysis of a phosphatase knock-down by quantitative proteomics and phosphoproteomics”<sup>69</sup> (7.3).

## **2.1 Mapping of endogenous Ser/Thr/Tyr phosphorylation in *Drosophila* SL2 cells**

In this first thesis project, I aimed to establish SILAC and global phosphoproteome analysis in *Drosophila melanogaster* SL2 cells, which was also intended as the basis for a subsequent functional phosphoproteome study. The goal was to map out an in-depth and high-quality *Drosophila* phosphoproteome and to characterize its structural and evolutionary properties.

The *Drosophila* SL2 system was chosen since it is a very well-established model for studying development and cell signaling. Moreover, *Drosophila* is less complex than vertebrates, which should allow deeper coverage of the phosphoproteome – given that the limitations are of a technological nature. As the ultimate goal was a global phosphoproteome analysis of a knock-down (2.2), the lower level of functional redundancy of *Drosophila* is furthermore advantageous to circumvent compensatory effects. Lastly, SL2 cells are a very well-suited system for functional gene knock-down studies because a highly efficient and penetrant transient RNAi can be obtained by a very simple silencing protocol<sup>2</sup>.

Although the mapping phase of the project only required qualitative data, we still chose a SILAC experimental setup to map Ser, Thr and Tyr phosphorylation sites (Figure 3). The heavy SILAC cell population was treated with permeable inhibitors of Ser/Thr phosphatases (calyculin A and deltamethrin) and of Tyr phosphatases (pervanadate) to increase the global level of phosphorylation. This had an analytical rather than a functional goal. It triggers the identification of very low abundant endogenous phosphorylation sites that were enhanced in response to the inhibitor treatment. Under the chosen inhibition conditions, Tyr phosphatase

## 2 Global phosphoproteome analysis

inhibition was more effective than the Ser/Thr one. 88 % of all detected phosphotyrosine (pTyr) sites were changed by more than 1.5-fold upon phosphatase inhibitor treatment, whereas only 44 % of the phosphoserine (pSer) / phosphothreonine (pThr) showed that effect (Supplemental Figure 2 of reference<sup>69</sup>). Importantly, presence of the SILAC partner peptide from the non stimulated condition ensured that the identified sites were endogenous.

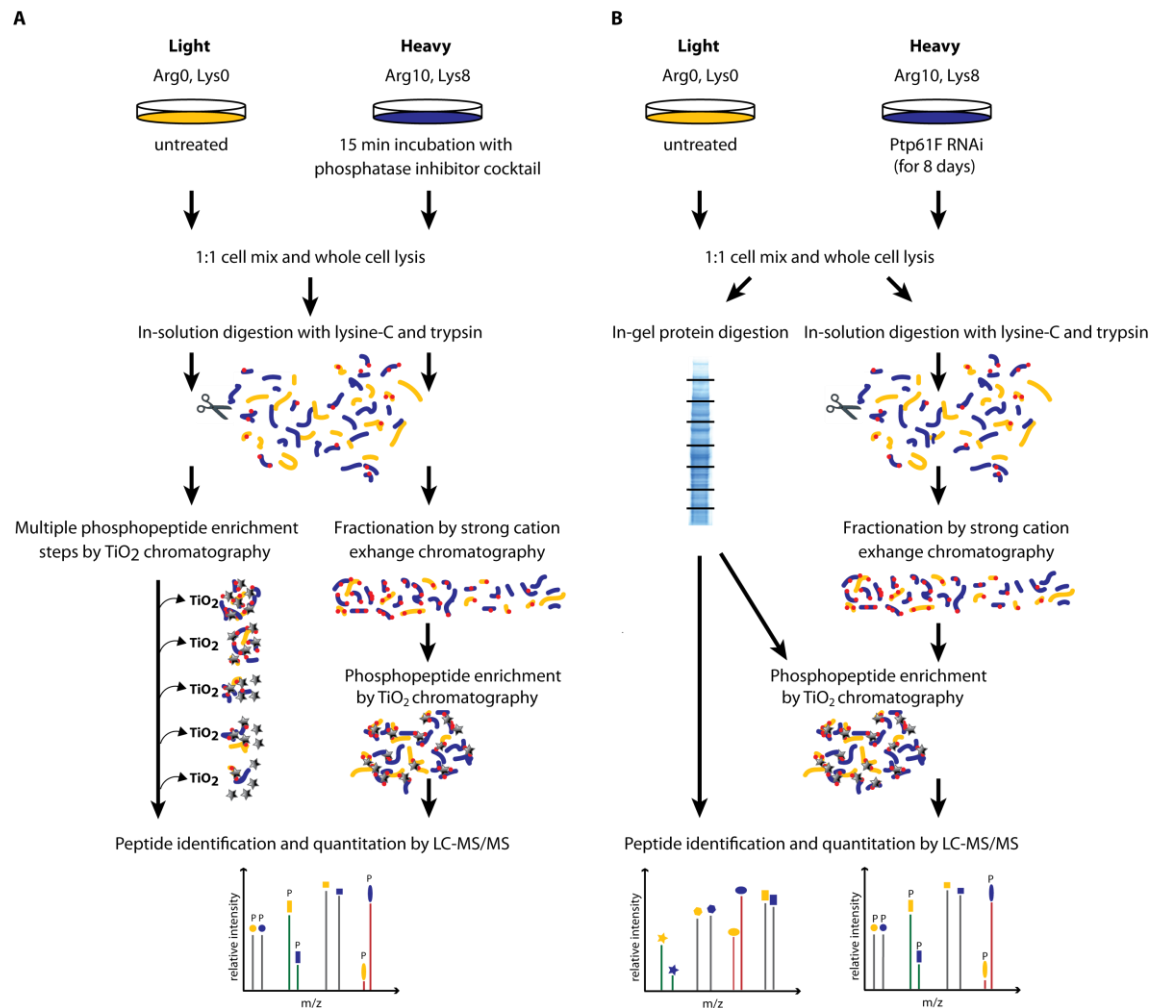




Figure 3 **Experimental workflow for the acquisition of the SL2 phosphoproteome with phosphatase inhibitor treatment (A) or Ptp61F RNAi (B) in the heavy SILAC population.** These procedures are described in more detail in the main text (2.1 and 2.2). • Phosphate group,  peptides,  titansphere (TiO<sub>2</sub>).

Phosphopeptides were enriched both by multiple incubations with TiO<sub>2</sub> (small-scale pre-experiment) and by SCX-TiO<sub>2</sub> chromatography from LysC and trypsin in-solution digested

peptide solution (Figure 3). In total 6,752 high-confidence phosphorylation sites (class I sites: 99 % peptide identification confidence and over 75 % amino acid localization confidence<sup>1</sup>) were detected in technical duplicates. The depth of this phosphoproteome is in the same range of the human HeLa phosphoproteome after epidermal growth factor (EGF) stimulation, for which 6,600 phosphorylation sites had been detected with a greater sampling and measuring effort<sup>1</sup>. This similarity suggested that a technical rather than the biological limit was reached in both cases. In a second global *Drosophila* phosphoproteome study (2.2), I quantified phosphorylation changes upon RNAi-mediated knock-down versus a control. The overlap between the two different *Drosophila* phosphoproteome data sets was only 47 %. However, technical replicates of the phosphatase inhibitor-assisted phosphoproteome measurement showed an overlap of 74 %. This suggested that the phosphatase inhibitor treatment and the knock-down data sets contributed different *Drosophila* phosphoproteome subsets. Therefore, both phosphoproteomes were combined, which led to the identification of 10,043 high-confidence phosphorylation sites on 2,379 phosphoproteins. All phosphorylation sites that are present in the control state reflect the endogenous phosphorylation status of the *Drosophila* SL2 cells, termed “basal phosphoproteome”. This accounts for 9,749 phosphorylation sites. Other *Drosophila* phosphoproteome studies using whole embryos<sup>70</sup> or a different cell line (KC167) under various growth and treatment conditions<sup>71</sup> reported similar numbers of phosphorylation sites. However, only 52 % of the SL2 basal phosphosites overlapped with these studies. This may reflect the different phosphorylation status of the different systems and/or the incompleteness of all three phosphoproteomes. This is why a phosphorylation site predictor was trained that can be used for *in-silico* analysis of phosphorylation of any protein of interest. 90 % of the basal *Drosophila* phosphoproteome sites detected by us were used to train the predictor whose performance was tested with the residual 10 %. This procedure was repeated five times with different partitions (5-fold cross validation). The prediction accuracy was 89.8 % for Ser and 81.1 % for Thr assuming equal numbers of phosphorylated and non-phosphorylated sites. The application of the *Drosophila* predictor to a human dataset<sup>1</sup> showed only a slightly worse performance, which implies high similarity of the human and *Drosophila* phosphoproteomes. This observation was further supported by similarities in the Gene Ontology (GO) enrichment of cellular compartments as well as molecular function and by a significant matching of kinase motifs from

*Drosophila* and human. Conservation analysis of the detected phosphoproteins and all FlyBase proteins known from experimental evidence showed a higher number of orthologous counterparts for phosphoproteins compared to all proteins for all analyzed 36 eukaryotes. This suggests that phosphoproteins are more conserved than non-phosphoproteins. However, it has to be considered that the detected phosphoproteomes usually covers the more abundant proteins, which are themselves more conserved.

SUMO is an example for an evolutionary conserved phosphoprotein. This finding is part of the research article with the title “Phosphorylation of SUMO-1 Occurs in Vivo and Is Conserved through Evolution”<sup>72</sup> (7.4), to which I contributed as coauthor.

SUMO is a small protein that serves as post-translational protein modifier and regulates protein interactions, half-life, transport and activity<sup>73</sup>. The C-terminus of SUMO is conjugated to target lysines mediated by an enzymatic mechanism analogous to ubiquitinylation. The same as other members of the ubiquitin and ubiquitin like protein family SUMO is a modification but also a protein, which can be targeted by PTMs. It is known that SUMO can be ubiquitinated<sup>74</sup> and sumoylated<sup>75,76</sup>. In this study we raised the question if SUMO can be targeted by PTMs not belonging to the ubiquitin/ubiquitin like protein family. We specifically analyzed the phosphorylation status of SUMO by enriching SUMO phosphopeptides with TiO<sub>2</sub> chromatography followed by MS analysis. In contrast to other studies that used tagged and overexpressed SUMO for MS characterization we focused on the endogenous SUMO. In order to obtain sufficient SUMO protein amount HeLa cells were lysed under non-denaturing conditions leading to deconjugation of SUMO from its target proteins. The HeLa protein lysate was separated by one-dimensional electrophoresis and the deconjugated SUMO containing 5-20 kDa region was excised and subjected to in-gel digestion. The resulting peptide mixture was analyzed by LC-MS/MS analysis on the LTQ-Orbitrap. We detected the N-terminal peptide of SUMO-1 lacking the initial methionine and being N-acetylated on Ser2 as described<sup>77</sup>. Moreover we found this peptide in a Ser2 phosphorylated and unphosphorylated version. At next we asked the question whether SUMO-1 is also phosphorylated on Ser2 when conjugated to target proteins. For this purpose a purified His6-SUMO-2 fraction, that did not contain free SUMO-1 but conjugates SUMO-1 as determined by Western Blot, was analyzed by MS and searched for phosphorylation. Again we found a peptide fraction being N-terminally acetylated and

phosphorylated on Ser2. Taken together, both free and conjugated SUMO-1, in this case to SUMO-2, can be targeted by phosphorylation. In both cases Ser2 was phosphorylated to high levels, free SUMO-1 to 35-40 % of and conjugated SUMO-1 to 50 %. Such a high stoichiometry suggests a functional significance of this phosphorylation site. Since Ser2 on the N-terminal tail of SUMO-1 is highly accessible, this phosphoserine could regulate protein-protein interactions. Moreover conservation analysis revealed that Ser2 of SUMO-1 is conserved in the SUMO homolog of *Saccharomyces cerevisiae* and *Drosophila melanogaster* called Smt3. As the human SUMO-1, Smt3 was found N-acetylated and phosphorylated on Ser2 by MS analysis in both species. The *Drosophila* Smt3 was detected in the gel fraction corresponding to 5-20 kDa within my basal *Drosophila* phosphoproteome. In this case the lysis was not performed under non-denaturing conditions but nevertheless enough free Smt3 was available to detect the phosphorylated and non-phosphorylated Ser2 peptide. The evolutionary conservation of the Ser2 phosphorylation site further supports its possible functional relevance.

In conclusion, the basal SL2 phosphoproteome was mapped to considerable depth by the applied phosphoproteomics procedure (Figure 3). Bioinformatic analysis of the basal SL2 phosphoproteome revealed shared common features, kinase motifs and conserved phosphorylation sites compared to human. Together with the lower complexity and the efficient RNAi capabilities *Drosophila* is a well-suited system for the characterization of gene knock-downs on the phosphoproteome level, on which I report next.

## **2.2 Global phosphoproteome response to Ptp61F phosphatase knock-down**

Global phosphoproteome studies of signaling processes have previously mainly focused on the investigation of growth factor or pheromone stimulated systems<sup>1,78</sup>. This second thesis project aimed to study the phosphoproteomic consequences of the depletion of just one member of phosphorylation dependent signaling, the *Drosophila* non-transmembrane protein tyrosine phosphatase Ptp61F. For this purpose, I combined global quantitative phosphoproteomics with RNAi in *Drosophila* SL2 cells. As mentioned above, SL2 cells enable very efficient transient knock-downs using long double-stranded RNA (dsRNA). The RNAi procedure is very simple as

it is sufficient to supplement the dsRNA to the cell culture media<sup>2</sup>, from which it is taken up by receptor-mediated endocytosis. The endogenous RNaseIII dicer processes the long RNA into multiple 21 bp to 23 bp siRNAs that target complementary mRNA sequences for cleavage by the endonuclease Argonaute<sup>279</sup>. Moreover, the reduced compensatory effects obtained in transient knock-downs, the lower extent of off-target effects and the lower functional redundancy of *Drosophila* compared to other vertebrates make it an ideal system to find targets of a phosphatase and of the processes it is involved in. I chose Ptp61F as a knock-down target because is an ortholog of the important mammalian phosphatase PTP1B<sup>80</sup>, which plays a role in diabetes, obesity and cancer<sup>81</sup>. Ptp61F is known as a negative regulator of Janus kinase (JAK)/signal transducer and activator of transcription (STAT) signaling. This pathway is evolutionary conserved and is involved in immune responses, haematopoiesis and cell proliferation<sup>82,83</sup>. Moreover, Ptp61F and the protein tyrosine kinase (PTK) Abelson (Abl) coordinately regulate localization and stability of the Abl interaction protein (Abi) and thereby control lamella formation<sup>84</sup>.

Transient protein knock-down requires inhibition of synthesis of the protein of interest and the complete turnover of protein residues that were expressed before starting the RNAi. The knockdown can affect both protein expression and phosphorylation, thus such an analysis requires parallel acquisition of phosphoproteome and proteome to be able to normalize phosphorylation level to protein level. This way changes in the stoichiometry of the phosphorylation site can be discriminated from absolute phosphorylation changes resulting from altered protein expression.

Preliminary experiments to optimize the target sequence of the dsRNA and the duration of the RNAi experiment were performed. Ptp61F was effectively down-regulated 8 days after incubation with the RNAi probe. As SILAC labeling required the same time frame<sup>85</sup>, labeling and knock-down were started in parallel (Figure 3). The Ptp61F knock-down was performed in the heavy cell population and the light control cells were incubated with a Glutathione S-Transferase (GST) RNAi probe that has no endogenous target. The proteome analysis involved in-gel digestion of 15 slices. For the phosphoproteome measurement, I applied TiO<sub>2</sub> enrichment after SCX peptide fractionation or after SDS-PAGE separation and in-gel digestion. The latter substituted the multiple TiO<sub>2</sub> enrichment steps from the in-solution digest performed in the

previous phosphoproteome experiment because those did not substantially increase the number of unique phosphorylation sites ( $\ll 1\%$ ) (Supplemental Table 8 of reference<sup>69</sup>).

The proteome analysis led to the identification of 4132 proteins (with a minimum of two unique peptides) out of which 3454 proteins were quantified with at least three quantification events per protein. Ptp61F was downregulated to more than 90% but otherwise no drastic proteome changes were detected. This implies that few if any off-target effects were present at least to the depth of the detected proteome. Requiring a minimum protein abundance change of 1.5-fold and a significance B (1.2.3) value of 0.2 % as outlier significance criteria revealed only sixteen upregulated and six downregulated proteins. However, if multiple hypothesis testing is taken into account they do not remain significant.

**Table 2 pSer/pThr/pTyr distribution of the total, normalized and regulated Ptp61F knock-down phosphoproteome.** Asterisk indicates that only phosphorylation sites that are normalized to protein level are considered.

	<b>Total</b>	<b>Normalized to protein level</b>	<b>Regulated*</b>	<b>Up*</b>	<b>Down*</b>
pSer/pThr/ pTyr	6478	4963	217	101	116
pSer	5562 (85.8 %)	4233 (85.3 %)	162 (74.7 %)	69 (68.3 %)	93 (80.2 %)
pThr	820 (12.7 %)	649 (13.1 %)	31 (14.3 %)	11 (10.9 %)	20 (17.2 %)
pTyr	96 (1.5 %)	81 (1.6 %)	24 (11.1 %)	21 (20.8 %)	3 (2.6 %)

In the phosphoproteome analysis in total 6,478 phosphorylation sites were identified. Combined analysis of proteome and phosphoproteome with MaxQuant enabled normalization of 77 % of the phosphorylation sites to their total protein level (Table 1). Of these phosphorylation sites, 4.4 % were regulated upon Ptp61F knock-down (applying the same outlier significance criteria as for the proteome). The percentage of pTyr in the upregulated sites was 20.8 % (21 phosphosites) compared to the total pTyr content of 1.6 % of all normalized phosphorylation sites. These upregulated tyrosine-phosphorylation sites are the potential direct

targets of the protein tyrosine phosphatase. Nineteen of those were detected on singly phosphorylated peptides so that the quantitative SILAC ratios could be uniquely assigned to those pTyr sites. The corresponding 15 proteins are illustrated in Figure 4 and are grouped according to known, novel, and novel conserved pTyr sites.

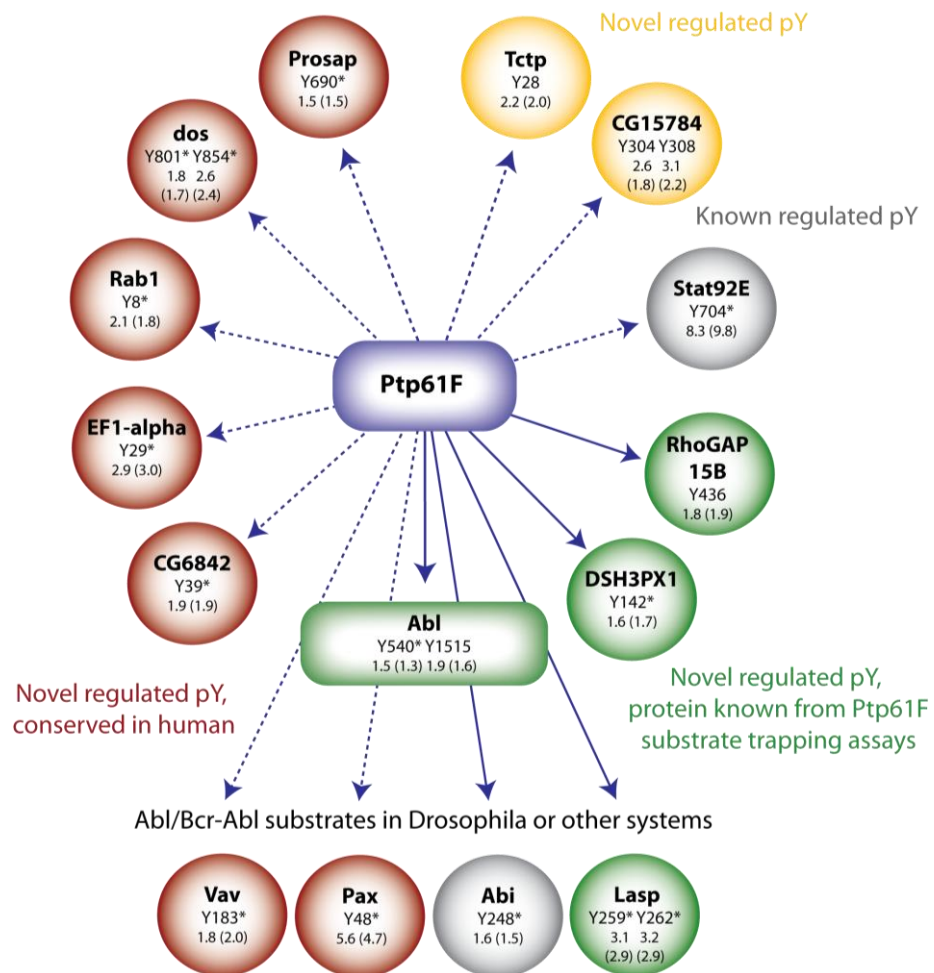


Figure 4 **Potential direct or indirect targets of Ptp61F that show upregulated tyrosine phosphorylation upon Ptp61F RNAi.** H/L phosphopeptide ratios and in brackets normalized ratios to protein level are indicated. Tyrosine (Y) sites with asterisk are conserved in human.

Among the tyrosine phosphorylated proteins I found the positive controls STAT92E (Y704) and Abi (Y248), which are known to be regulated by Ptp61F as mentioned above. A group of upregulated pTyr sites (marked in green in Figure 4) were suggested as Ptp61F targets in a previously published substrate trapping study that employed a catalytically inactive Ptp61F



mutant (D203A) to specifically capture its substrates and MS analysis to identify those<sup>86</sup>. While this approach lacked phosphorylation site localization my study confirmed and site-specifically analyzed these potential targets. The kinase Abl, for example, was found increased in phosphorylation on Tyr540, which is known to increase the kinase activity<sup>87,88</sup>. Moreover its C-terminal autophosphorylation site Tyr1515 was found to be upregulated, further supporting an active kinase state. In agreement with Abl activation, phosphorylation of Vav<sup>89</sup>, Paxillin<sup>90</sup>, Abi<sup>84</sup> and Lasp<sup>91</sup>, which are known substrates of Abl in Drosophila and/or vertebrate systems, was increased upon Ptp61F RNAi in my study. However the approach applied here cannot discriminate between an effect mediated by combined or subsequent action of Abl and Ptp61F from an exclusive action of Abl. More generally it cannot discriminate direct from indirect effects. Its strength is rather to detect the net effect of the phosphatase knock-down in a systems-wide and unbiased manner.

The focal adhesion component Paxillin and the guanine-nucleotide exchange factor Vav link Ptp61F function to cytoskeletal organization. Both proteins were suggested as PTP1B targets on other phosphorylation sites. Paxillin was identified in a PTP1B substrate trapping experiment<sup>92</sup>. The same study detected Vav3 with increased pTyr phosphorylation in a quantitative phosphotyrosine proteomics experiment comparing PTP1B deficient mouse fibroblasts to wild type cells.

Apart from the regulated pTyr sites, numerous pSer and pThr were affected by the PTP61F RNAi - both increasing and decreasing phosphorylation (Table 1). This demonstrates the broad effects of the depletion of just one phosphatase.

In conclusion, the combined quantitative proteomics and phosphoproteomics approach revealed distinct effects of the knock-down of one phosphatase within the global phospho-signaling network. High resolution mass spectrometry and accurate quantification lead to the detection of significant phosphoproteome changes. The results linked Ptp61F to new potential targets and downstream effects. Enrichment in the categories such as cytoskeleton regulation and adhesion signaling further expanded the role of the phosphatase in these processes. This study constitutes proof of principle for the power of combined quantitative proteomics and global phosphoproteomics for systems-wide analysis of signaling network perturbation. While

## 2 Global phosphoproteome analysis

---

performed for one phosphatase, it could conceivably be extended to a systems-wide effort, perhaps including kinases, which would also help distinguishing direct from indirect effects.

### 3 Interaction proteomics

Most biological processes involve protein interactions and many are mediated by multi-protein complexes. The identification of these complex members is crucial to understand underlying molecular mechanisms and regulated physiological functions. In combination with affinity purification, MS-based proteomics can be optimally exploited for protein-protein interaction analysis enabling unbiased and systems-wide analysis of protein interaction networks<sup>93</sup>.

Similar to many other AP-MS studies, one of the two first high-throughput interaction screens of the complete yeast interactome applied tandem affinity purification (TAP) aiming for high purity of enriched complexes<sup>31</sup>. Despite the success of the TAP AP-MS method, especially in yeast, it has some limitations. During the purification procedure contaminations are de-enriched by multiple steps requiring relatively high starting amounts and stringent washing conditions, which can lead to the loss of substoichiometric and low affinity interactors. Even then, in most cases complete removal of contaminations cannot be obtained, especially considering the increasing sensitivity of today's mass spectrometers. In recent years, quantitative proteomics has been employed to distinguish specific interactors from background binders comparing differentially isotope labeled specific bait and control APs<sup>30,94</sup>. In this way, quantitative proteomics obtains high-confidence interaction data and moreover allows the use of low stringent purification procedures that increase the chance of retaining substoichiometric and low affinity interactors. While different types of baits such as tagged, overexpressed proteins and endogenous proteins have been employed in quantitative AP-MS studies, our group established the use of GFP-tagged full-length proteins expressed under control of their endogenous promoters and other endogenous up- and downstream regulatory elements<sup>95</sup>. For that purpose, a bacterial artificial chromosome (BAC) containing the gene of interest is fused to the green fluorescent protein (GFP) and the BAC transgene is stably integrated into the genome of a target cell line<sup>96-99</sup>. This results in bait proteins with endogenous protein expression levels, splice isoforms and post-translational modifications. The combination of quantitative AP-MS and BAC TransgeneOmics was termed QUBIC for QUantitative BAC InteraCtomics<sup>95</sup>.

Importantly, most regulatory protein complexes are dynamic and change in composition and activity in response to cellular stimuli<sup>100</sup>. This is mediated by stimulation-dependent changes in protein conformation, subcellular localization or post translational modification. In the following thesis project, QUBIC is established in the triple encoding SILAC format and used to investigate Wnt stimulus specific interactions of known Wnt pathway components. A paper about this project with the title “Triple SILAC to determine stimulus specific interactions in the Wnt pathway” is currently under revision<sup>101</sup> (7.5).

### **3.1 Stimulus specific interaction screens in the Wnt pathway**

While many interaction studies have applied SILAC in a double labeling format, triple SILAC so far had only been used for differentiation of isoform specific interactors<sup>102</sup> and recently for the analysis of the RNA polymerase complex upon inhibition of transcription<sup>103</sup>. My study aimed to establish a generic QUBIC-based experimental and data analysis procedure for characterizing regulatory protein complexes in a stimulus dependent manner. A triple SILAC format enables to differentiate background binders from specific interactors and in the same measurement constitutive from stimulus dependent interactors. This approach was tested on components of the Wnt pathway, for which various known interaction partners served as positive controls to evaluate the interaction screen. Conversely, the dynamics of the central regulating complex of the Wnt pathway have not yet been fully understood and were addressed by the dynamic interactome screens. Moreover, I aimed to identify novel Wnt pathway components and their interaction dynamics.

The Wnt signaling pathway is one of the most intensely studied ones and of great importance due to its role the regulation of cell fate, proliferation and self-renewal of adult stem and progenitor cells during the entire lifespan of metazoan organisms<sup>3-7</sup>. Dysregulation of the pathway causes severe disorders such as colon cancer. The pathway has been depicted and introduced as an example of signal transduction in the first chapter (Figure 1). The most important complex of the pathway is the cytoplasmic destruction complex, which determines the stability of the transcriptional coactivator  $\beta$ -Catenin. It is composed of the main components APC, Axin-1, GSK-3 $\beta$  and CKI- $\alpha$  and mediates continuous phosphorylation of  $\beta$ -Catenin in the

absence of a Wnt signal. Phosphorylated  $\beta$ -Catenin is ubiquitinated and subsequently undergoes proteasomal degradation. Wnt ligands such as Wnt3a activate the Wnt pathway by inhibition of  $\beta$ -Catenin degradation. Although potential mechanisms for this have been proposed e.g. complex dissociation by partial translocation to the membrane<sup>104</sup> or by Axin-1 degradation<sup>105</sup>, they still remain controversial.

The scaffold destruction complex components APC and Axin-1, DVL2, a mediator of the Wnt signal from the membrane to the destruction complex, as well as CtBP2, a coregulator of Wnt target gene transcription were chosen as bait proteins for the dynamic triple SILAC interaction screens.

The cells expressing the GFP-tagged bait protein were light (L) and medium (M) SILAC labeled, while the control cell line without BAC transgene was heavy (H) SILAC labeled (Figure 5). The light labeled transgenic cells were stimulated with Wnt3a for two hours. To prevent light to heavy or medium to heavy exchange of specific interaction partners during the immunopurification procedure<sup>106,107</sup>, GFP pull-downs were performed separately for each SILAC condition. A monoclonal anti-GFP antibody coupled to magnetic microbeads was used allowing short incubation times of 15 min. Fractionation of the mixed eluates by one-dimensional SDS separation – rather than single-run analysis of an in-solution digest - was necessary to cover the interactome in-depth. Peptide triplets were detected by the LTQ-Orbitrap Velos and can be grouped according to the detected peak pattern. Unspecific background binders to beads or GFP-antibody are present in equal amounts in all the three states. Specific interactors with the GFP-bait show significant peptide ratios between the pull-down with the non-stimulated cell population and the untransfected control and/or between the stimulated cell population and the control. A stimulus dependent interactor moreover has a significant ratio for the peptide intensity of the GFP-bait pull-down from the stimulated cell population compared to the GFP-bait pull-down from the non-stimulated cell population. Several ways to represent these data are explained in detail in the manuscript of this study (7.5). Most importantly, all three possible SILAC ratios of the triplets have to be considered independently. Either two different two-dimensional plots of the ratios or a heat map that shows all three ratios in single picture can be used to visualize the data.

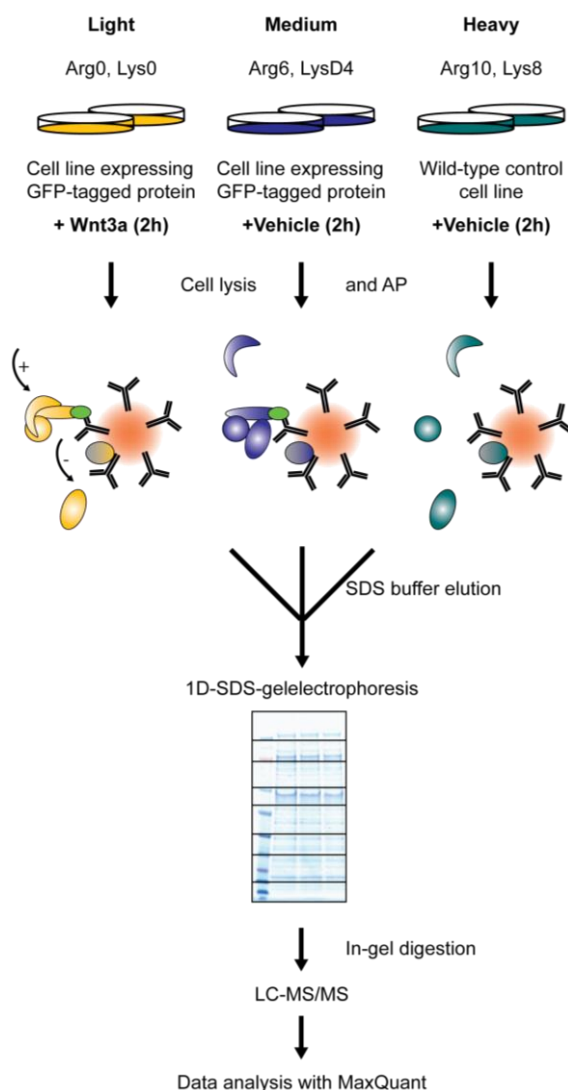
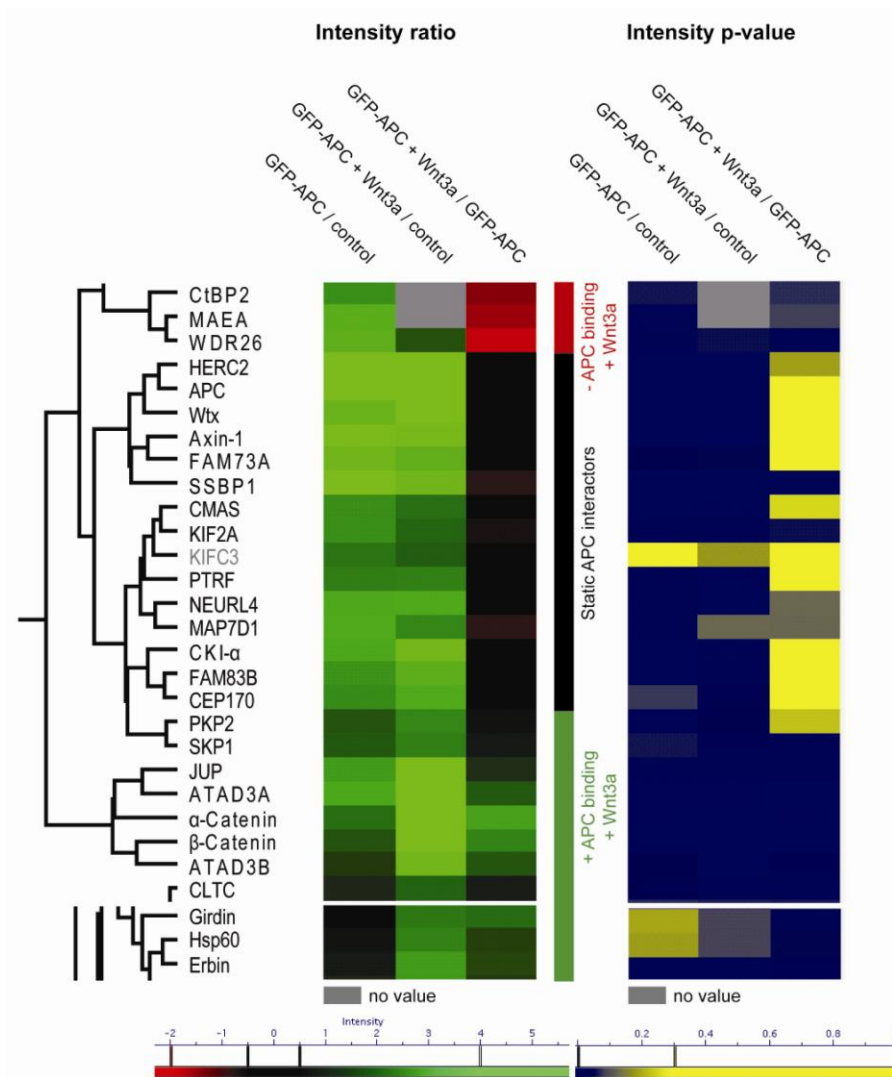


Figure 5 **Experimental workflow for triple SILAC pull-downs to determine Wnt3a dependent interaction dynamics.** The cell line expressing the GFP-tagged protein of interest is light and medium SILAC labeled and the untransfected wild-type control cell line is heavy SILAC labeled. Cells are lysed after two hour treatment with Wnt3a (200 ng/mL) or vehicle solution, respectively. GFP-pull-downs are performed separately for each SILAC state. Eluates are combined, separated on a one-dimensional gel into eight slices and in-gel digested. Resulting peptide mixtures are analyzed by high resolution HPLC -MS/MS on an LTQ-Orbitrap Velos. SILAC ratios are automatically quantified by MaxQuant.

The APC interactome was measured in four biological replicates to assess the biological reproducibility of the APC interactions via a one sample t-test. In total 28 reproducible APC interactors (p value < 0.1 and a minimum ratio of 4) were obtained and are represented by a

heat map of their median ratios (Figure 6). They are assembled into two subgroups by hierarchical clustering of all detected proteins quantified in three replicates.



**Figure 6 Dynamic APC interactome visualized by one-way hierarchical clustering.** The three ratios of the triple SILAC pull-down (median of four experiments) are used to cluster the reproducible APC interactors by one-way hierarchical clustering. A green color value signifies specific binding to APC without Wnt stimulation (first column) or with Wnt stimulation (second column). The third column depicts the SILAC ratio of simulated against unstimulated bait pull-down. In this column, a green color value represents enhanced binding to APC upon Wnt activation and a red color represents reduced binding. Constitutive interactors have no significant ratio and therefore appear in black. Additionally, t-test results for the ratio reproducibility are visualized in the right panel after the clustering process. Proteins with reproducible ratios are in blue and those below the threshold in yellow. All ratio intensities are shown in log scale using the indicated color code.

The Axin-1, CtBP2 and DVL2 interactomes were analyzed in biological duplicates with reverse SILAC labeling. Results of both replicates are hierarchical clustered and visualized as heat maps (Figure 5 of reference<sup>101</sup> (7.5)). In this case, interactor reproducibility is assessed by comparison of the colour coding of the ratios. Reproducible interactors mainly clustered together.

The scaffold destruction complex proteins APC and Axin-1 were chosen as bait proteins to detect novel destruction complex interaction partners and destruction complex interaction dynamics. The comparison of the detected Axin-1 and APC interactomes revealed ten shared interactors (Figure 7). Most of them show highly similar interaction behavior to both APC and Axin making them likely interaction partners of an APC-Axin-1 complex. Among them are the known destruction complex members CKI- $\alpha$  and Wtx, which were not affected in their interaction to APC and Axin-1 by Wnt3a stimulation. APC and Axin-1 also did not show dynamic binding to each other in the reciprocal pulldowns. The destruction complex components GSK-3 $\beta$  and PP2A were only identified as Axin-1 interactors and showed stimulus independent binding to Axin-1. Taken together, these interaction measurements suggest that the destruction complex does not disassemble after Wnt3a ligand binding but rather stays intact while the complex is translocated to the membrane. This model is in agreement with recent observations<sup>108</sup>. Moreover, FAM83B was identified as a novel constitutive interactor of Axin-1 and APC and represents a potential novel destruction complex member.

$\beta$ -Catenin is an example for a stimulus dependent interactor that showed enhanced binding to APC and Axin-1 upon Wnt3a activation. This is most likely simply the result of accumulated  $\beta$ -Catenin upon Wnt pathway stimulation. However, a complex consisting of these three proteins (APC, Axin and  $\beta$ -Catenin) has been reported to localize at the membrane after Wnt3a activation<sup>109</sup>.

Interestingly, Girdin (CCDC88A) was found as a novel dynamic interactor of Axin-1 and APC, whose binding was greatly enhanced upon Wnt3a activation. Moreover, it was identified as a novel interactor in the DVL2 interactome. However, the interaction of Girdin with DVL2 was not Wnt stimulus dependent. These observations suggest the hypothesis that Girdin and DVL2 are located in a pre-formed complex that may recruit the destruction complex (or at least APC-Axin complexes) to the membrane receptors upon Wnt pathway stimulation. DVL2



proteins are known to be important for counteracting destruction complex action upon Wnt stimulation. However, the exact mechanism by which the Wnt signal is transduced - including the complex formation at the membrane, which involves phosphorylated DVL - is still not fully understood<sup>3</sup>. So far, Girdin is a known substrate and regulator of Akt signalling and a paralog of Daple, which interacts with DVL via its Gly-Cys-Val C-terminal motif<sup>10</sup>. Girdin, however, does not have that motif and thus the interaction to DVL must be mediated in a different manner.

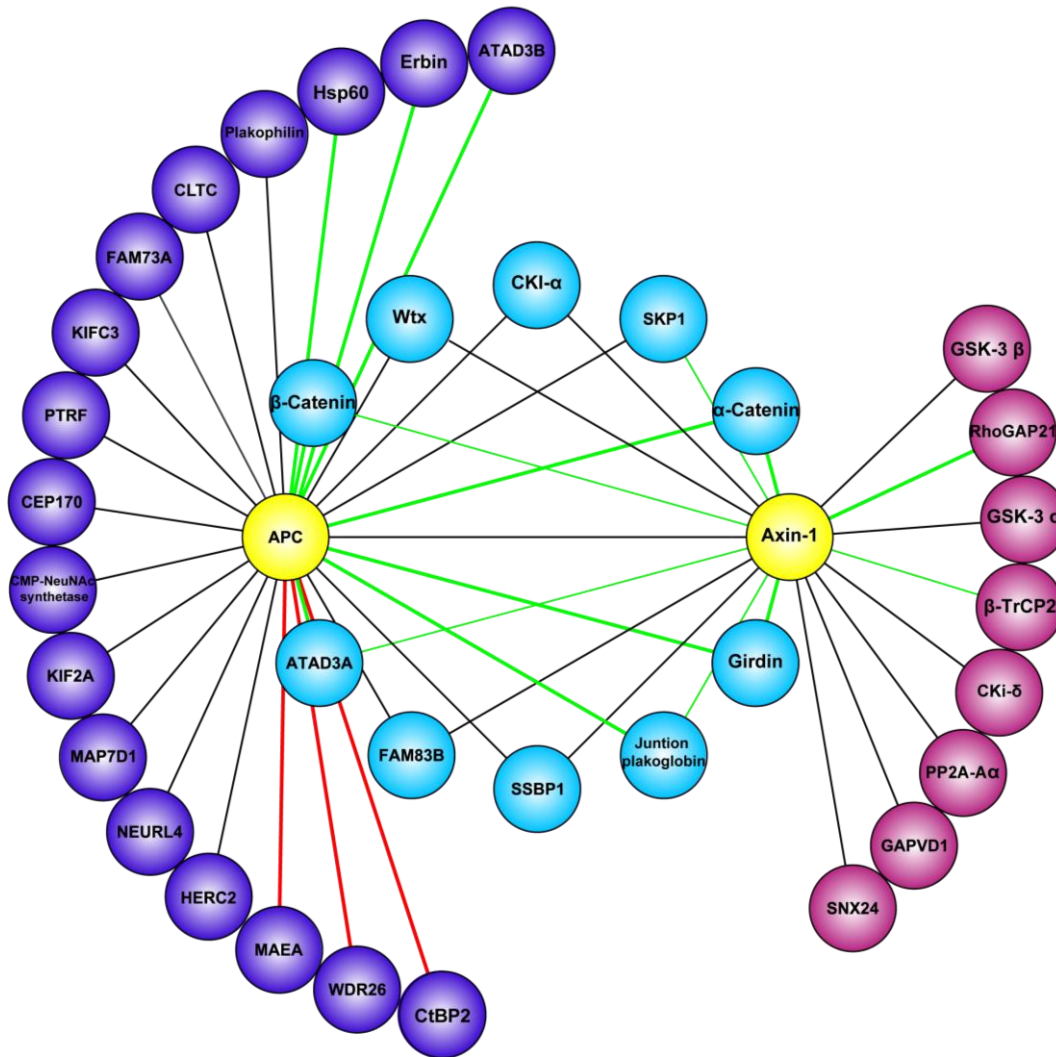


Figure 7 **Overlap of APC and Axin-1 interactomes.** Protein-protein interactions were drawn in Cytoscape. Baits are depicted in yellow, shared APC and Axin-1 interactors in blue and unique interactors for APC and Axin-1 in purple and pink, respectively. Lines represent detected interactions. Green lines indicate enhanced interaction upon Wnt3a activation, while red lines indicate reduced interactions upon Wnt3a activation. Line width reflects the SILAC ratio intensity for the dynamic interactors.

CtBP2 was detected with decreased binding to APC upon Wnt3a activation in the APC interactome. Without Wnt activation CtBP2 is a transcriptional regulator that binds the APC- $\beta$ -Catenin complex in the nucleus<sup>111</sup>. The formation of the CtBP2-APC- $\beta$ -Catenin complex competes with the binding of  $\beta$ -Catenin to the transcription factor TCF and thereby represses Wnt-dependent gene expression. Colorectal cell lines with truncated APC also show diminished binding of APC to CtBP2<sup>111</sup>, which contributes to increased expression of Wnt target genes. My observation that the APC-CtBP2 interaction is decreased upon Wnt activation shows that the truncation of APC is mechanistically equivalent to stimulation by the Wnt ligand in abolishing CtBP2 binding to the C-terminal part of APC. The CtBP2 interactome detected a constitutive interaction of  $\beta$ -Catenin and CtBP2. This is in line with the suggestions that these two proteins function together in both repression<sup>111</sup> and activation of gene expression<sup>112</sup>.

In conclusion, the triple SILAC interaction screen developed here accurately distinguished constitutive from dynamic, signal dependent interactions. In contrast to targeted techniques, such as western blotting, it can both discover and characterize interaction dynamics in the same experiment. This study also showed that interaction proteomics is clearly ready to include a dynamic dimension in protein interaction studies in a highly accurate and generic manner.

## 4 Perspectives

Hopefully, this thesis has demonstrated the power of MS-based proteomics to gain deeper insights into cell signaling networks. My studies led to the identification of novel *Drosophila* phosphorylation sites and Ptp61F targets. For the first, time combined proteome and global phosphoproteome analysis was used as readout of a gene knock-down. Moreover, triple SILAC interaction screens were established that detected novel interactors of regulatory Wnt pathway components and at the same time Wnt stimulation dependent interaction dynamics. Importantly, our experiments provide evidence that the  $\beta$ -Catenin destruction complex remains essentially unchanged upon pathway activation.

Despite the great success proteomic approaches already have for unraveling signal transduction, much remains to be done. Future developments will aim for high-throughput systems biology applications and the use of *in-vivo* material. For these goals improvements may come from the following directions:

Technical advances in mass spectrometry instrumentation will be necessary to improve the dynamic range and sensitivity to be able to reach near complete proteome coverage at the same time aiming for 'low' measurement expenses. The yeast system is the first example for comprehensive proteome analysis<sup>26</sup> that is at the moment being optimized to employ only a single day of measurement time. For this, not only the faster LTQ-Orbitrap Velos with increased sensitivity but also the optimizations of the upfront liquid chromatography will play a major role. These important technical developments will ultimately allow systematic analysis of complete signaling networks e.g. by consecutive KOs of all their components. Due to the low abundance and stoichiometry of regulatory phosphorylation sites, the comprehensive analysis of the phosphoproteome is much more challenging. Moreover, it is still difficult to estimate the total number of phosphorylation sites of a given cell state. While MS developments may promote more comprehensive phosphoproteomic analysis, for systematic analysis of known phosphorylation sites, targeted proteomic approaches might become a solution.

As cell line systems have their limitations in resembling the *in-vivo* signaling situation, the move to *in-vivo* systems is very important and ongoing at the moment. While proteome analysis

even of distinct cell types is possible<sup>113</sup>, in-depth phosphoproteome analysis still requires a substantial amount of starting material that can currently only be derived from large organs or tumor material. Therefore, method development that reduces the required protein amount will be crucial and will be a key to understand cell-type specific signaling.

Protein-protein interactions are essential for all cell signaling processes and the detection of interactomes of whole signaling networks and even of whole organisms will be a goal of high throughput MS-based interaction studies. A second generation of interaction studies should moreover take into account that the formation of protein interaction complexes is dependent on the cell type, developmental stage and stimulation state, as shown here for the Wnt pathway.

Data analysis and mining will remain challenging as mass spectrometry will be used in an even more high-throughput manner. Moreover, the integration of different layers of proteomic information: interactome, proteome, phosphoproteome, acetylome and glycome as well as data from other disciplines such as high-throughput RNAi and mRNA expression data will be important and may be routinely involved for systems-wide characterization of signaling pathways in the future. This will be accompanied by the future use of MS-based proteomics as more and more routine method that can be applied by non-specialist labs.

## 5 References

- 1 J. V. Olsen, B. Blagoev, F. Gnad et al., *Cell* **127** (3), 635 (2006).
- 2 C. A. Worby, N. Simonson-Leff, and J. E. Dixon, *Sci STKE* **2001** (95), PL1 (2001).
- 3 S. Angers and R. T. Moon, *Nature reviews* **10** (7), 468 (2009).
- 4 H. Clevers, *Cell* **127** (3), 469 (2006).
- 5 A. Klaus and W. Birchmeier, *Nat Rev Cancer* **8** (5), 387 (2008).
- 6 B. T. MacDonald, K. Tamai, and X. He, *Developmental cell* **17** (1), 9 (2009).
- 7 R. van Amerongen and R. Nusse, *Development (Cambridge, England)* **136** (19), 3205 (2009).
- 8 T. Hunter, *Cell* **100** (1), 113 (2000).
- 9 John T. Hancock, *Cell Signalling*, Third ed. (Oxford University Press, New York, 2010).
- 10 H. McNeill and J. R. Woodgett, *Nat Rev Mol Cell Biol* **11** (6), 404 (2010).
- 11 C. Liu, Y. Li, M. Semenov et al., *Cell* **108** (6), 837 (2002).
- 12 T. Pawson, *Cell* **116** (2), 191 (2004).
- 13 Y. L. Deribe, T. Pawson, and I. Dikic, *Nature structural & molecular biology* **17** (6), 666 (2010).
- 14 Gerhard Krauss, *Biochemistry of Signal Transduction and Regulation*, 4 ed. (WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2008).
- 15 J. B. Fenn, M. Mann, C. K. Meng et al., *Science (New York, N.Y)* **246** (4926), 64 (1989).
- 16 M. Karas and F. Hillenkamp, *Analytical chemistry* **60** (20), 2299 (1988).
- 17 A. Makarov, E. Denisov, A. Kholomeev et al., *Analytical chemistry* **78** (7), 2113 (2006).
- 18 J. E. Syka, J. A. Marto, D. L. Bai et al., *Journal of proteome research* **3** (3), 621 (2004).
- 19 R. Zubarev and M. Mann, *Mol Cell Proteomics* **6** (3), 377 (2007).
- 20 B. Domon and R. Aebersold, *Science (New York, N.Y)* **312** (5771), 212 (2006).
- 21 J. V. Olsen, J. C. Schwartz, J. Griep-Raming et al., *Mol Cell Proteomics* **8** (12), 2759 (2009).
- 22 S. P. Gygi, B. Rist, S. A. Gerber et al., *Nature biotechnology* **17** (10), 994 (1999).
- 23 S. E. Ong, B. Blagoev, I. Kratchmarova et al., *Mol Cell Proteomics* **1** (5), 376 (2002).
- 24 P. L. Ross, Y. N. Huang, J. N. Marchese et al., *Mol Cell Proteomics* **3** (12), 1154 (2004).
- 25 J. Cox and M. Mann, *Nature biotechnology* (2008).
- 26 L. M. de Godoy, J. V. Olsen, J. Cox et al., *Nature* **455** (7217), 1251 (2008).
- 27 R. Aebersold and M. Mann, *Nature* **422** (6928), 198 (2003).
- 28 K. Schmelzle and F. M. White, *Current opinion in biotechnology* **17** (4), 406 (2006).
- 29 E. S. Witze, W. M. Old, K. A. Resing et al., *Nature methods* **4** (10), 798 (2007).
- 30 B. Blagoev, I. Kratchmarova, S. E. Ong et al., *Nature biotechnology* **21** (3), 315 (2003).
- 31 A. C. Gavin, P. Aloy, P. Grandi et al., *Nature* **440** (7084), 631 (2006).
- 32 T. Kocher and G. Superti-Furga, *Nature methods* **4** (10), 807 (2007).
- 33 M. E. Sowa, E. J. Bennett, S. P. Gygi et al., *Cell* **138** (2), 389 (2009).
- 34 M. Vermeulen, N. C. Hubner, and M. Mann, *Current opinion in biotechnology* **19** (4), 331 (2008).
- 35 A. Wepf, T. Glatter, A. Schmidt et al., *Nature methods* **6** (3), 203 (2009).

## 5 References

---

- 36 J. V. Olsen, S. E. Ong, and M. Mann, *Mol Cell Proteomics* **3** (6), 608 (2004).
- 37 A. Shevchenko, M. Wilm, O. Vorm et al., *Analytical chemistry* **68** (5), 850 (1996).
- 38 P. Horth, C. A. Miller, T. Preckel et al., *Mol Cell Proteomics* **5** (10), 1968 (2006).
- 39 N. C. Hubner, S. Ren, and M. Mann, *Proteomics* **8** (23-24), 4862 (2008).
- 40 L. Andersson and J. Porath, *Analytical biochemistry* **154** (1), 250 (1986).
- 41 C. Choudhary, C. Kumar, F. Gnad et al., *Science (New York, N.Y)* **325** (5942), 834 (2009).
- 42 M. R. Larsen, T. E. Thingholm, O. N. Jensen et al., *Mol Cell Proteomics* **4** (7), 873 (2005).
- 43 M. W. Pinkse, P. M. Uitto, M. J. Hilhorst et al., *Analytical chemistry* **76** (14), 3935 (2004).
- 44 W. Ens and K. G. Standing, *Methods in enzymology* **402**, 49 (2005).
- 45 A. Thompson, J. Schafer, K. Kuhn et al., *Analytical chemistry* **75** (8), 1895 (2003).
- 46 Y. Oda, K. Huang, F. R. Cross et al., *Proceedings of the National Academy of Sciences of the United States of America* **96** (12), 6591 (1999).
- 47 S. E. Ong and M. Mann, *Nature chemical biology* **1** (5), 252 (2005).
- 48 M. Mann, *Nat Rev Mol Cell Biol* **7** (12), 952 (2006).
- 49 T. Geiger, J. Cox, P. Ostasiewicz et al., *Nature methods* **7** (5), 383 (2010).
- 50 Y. Ishihama, T. Sato, T. Tabata et al., *Nature biotechnology* **23** (5), 617 (2005).
- 51 M. Kruger, M. Moser, S. Ussar et al., *Cell* **134** (2), 353 (2008).
- 52 M. D. Sury, J. X. Chen, and M. Selbach, *Mol Cell Proteomics* (2010).
- 53 H. Molina, Y. Yang, T. Ruch et al., *Journal of proteome research* **8** (1), 48 (2009).
- 54 D. Chelius and P. V. Bondarenko, *Journal of proteome research* **1** (4), 317 (2002).
- 55 P. V. Bondarenko, D. Chelius, and T. A. Shaler, *Analytical chemistry* **74** (18), 4741 (2002).
- 56 M. J. MacCoss, C. C. Wu, H. Liu et al., *Analytical chemistry* **75** (24), 6912 (2003).
- 57 H. Liu, R. G. Sadygov, and J. R. Yates, 3rd, *Analytical chemistry* **76** (14), 4193 (2004).
- 58 Y. Ishihama, Y. Oda, T. Tabata et al., *Mol Cell Proteomics* **4** (9), 1265 (2005).
- 59 J. E. Elias and S. P. Gygi, *Nature methods* **4** (3), 207 (2007).
- 60 J. Cox, I. Matic, M. Hilger et al., *Nature protocols* **4** (5), 698 (2009).
- 61 P. Cohen, *Philosophical transactions of the Royal Society of London* **354** (1382), 485 (1999).
- 62 J. V. Olsen, M. Vermeulen, A. Santamaria et al., *Science signaling* **3** (104), ra3 (2010).
- 63 S. A. Beausoleil, M. Jedrychowski, D. Schwartz et al., *Proceedings of the National Academy of Sciences of the United States of America* **101** (33), 12130 (2004).
- 64 M. J. Schroeder, J. Shabanowitz, J. C. Schwartz et al., *Analytical chemistry* **76** (13), 3590 (2004).
- 65 L. M. Mikesch, B. Ueberheide, A. Chi et al., *Biochimica et biophysica acta* **1764** (12), 1811 (2006).
- 66 J. E. Syka, J. J. Coon, M. J. Schroeder et al., *Proceedings of the National Academy of Sciences of the United States of America* **101** (26), 9528 (2004).
- 67 R. A. Zubarev, *Mass spectrometry reviews* **22** (1), 57 (2003).
- 68 Marjo de Graauw (Ed.), *Phospho-Proteomics Methods and Protocols*. (Humana Press, 2009).
- 69 M. Hilger, T. Bonaldi, F. Gnad et al., *Mol Cell Proteomics* **8** (8), 1908 (2009).
- 70 B. Zhai, J. Villen, S. A. Beausoleil et al., *Journal of proteome research* **7** (4), 1675 (2008).

- 71 B. Bodenmiller, J. Malmstrom, B. Gerrits et al., *Molecular systems biology* **3**, 139 (2007).
- 72 I. Matic, B. Macek, M. Hilger et al., *Journal of proteome research* **7** (9), 4050 (2008).
- 73 R. T. Hay, *Molecular cell* **18** (1), 1 (2005).
- 74 C. M. Pickart and D. Fushman, *Current opinion in chemical biology* **8** (6), 610 (2004).
- 75 I. Matic, M. van Hagen, J. Schimmel et al., *Mol Cell Proteomics* **7** (1), 132 (2008).
- 76 M. H. Tatham, E. Jaffray, O. A. Vaughan et al., *The Journal of biological chemistry* **276** (38), 35368 (2001).
- 77 V. Lallemand-Breitenbach, M. Jeanne, S. Benhenda et al., *Nature cell biology* **10** (5), 547 (2008).
- 78 A. Gruhler, J. V. Olsen, S. Mohammed et al., *Mol Cell Proteomics* **4** (3), 310 (2005).
- 79 G. Meister, M. Landthaler, A. Patkaniowska et al., *Molecular cell* **15** (2), 185 (2004).
- 80 J. N. Andersen, R. L. Del Vecchio, N. Kannan et al., *Methods (San Diego, Calif)* **35** (1), 90 (2005).
- 81 N. Dube and M. L. Tremblay, *Biochimica et biophysica acta* **1754** (1-2), 108 (2005).
- 82 G. H. Baeg, R. Zhou, and N. Perrimon, *Genes & development* **19** (16), 1861 (2005).
- 83 P. Muller, D. Kuttenukeuler, V. Gesellchen et al., *Nature* **436** (7052), 871 (2005).
- 84 C. H. Huang, T. Y. Lin, R. L. Pan et al., *The Journal of biological chemistry* **282** (44), 32442 (2007).
- 85 T. Bonaldi, T. Straub, J. Cox et al., *Molecular cell* **31** (5), 762 (2008).
- 86 Y. C. Chang, S. Y. Lin, S. Y. Liang et al., *Journal of proteome research* **7** (3), 1055 (2008).
- 87 M. J. Henkemeyer, R. L. Bennett, F. B. Gertler et al., *Molecular and cellular biology* **8** (2), 843 (1988).
- 88 K. Q. Tanis, D. Veach, H. S. Duewel et al., *Molecular and cellular biology* **23** (11), 3884 (2003).
- 89 F. Bassermann, T. Jahn, C. Miething et al., *The Journal of biological chemistry* **277** (14), 12437 (2002).
- 90 J. M. Lewis and M. A. Schwartz, *The Journal of biological chemistry* **273** (23), 14225 (1998).
- 91 Y. H. Lin, Z. Y. Park, D. Lin et al., *The Journal of cell biology* **165** (3), 421 (2004).
- 92 P. Mertins, H. C. Eberl, J. Renkawitz et al., *Mol Cell Proteomics* (2008).
- 93 A. C. Gingras, M. Gstaiger, B. Raught et al., *Nature reviews* **8** (8), 645 (2007).
- 94 J. A. Ranish, E. C. Yi, D. M. Leslie et al., *Nature genetics* **33** (3), 349 (2003).
- 95 N. C. Hubner, A. W. Bird, J. Cox et al., *The Journal of cell biology* **189** (4), 739 (2010).
- 96 Y. Zhang, F. Buchholz, J. P. Muyrers et al., *Nature genetics* **20** (2), 123 (1998).
- 97 R. Kittler, L. Pelletier, C. Ma et al., *Proceedings of the National Academy of Sciences of the United States of America* **102** (7), 2396 (2005).
- 98 M. Sarov, S. Schneider, A. Pozniakovski et al., *Nature methods* **3** (10), 839 (2006).
- 99 I. Poser, M. Sarov, J. R. Hutchins et al., *Nature methods* **5** (5), 409 (2008).
- 100 S. Charbonnier, O. Gallego, and A. C. Gavin, *Biotechnology annual review* **14**, 1 (2008).
- 101 M. Hilger and M. Mann, *Mol Cell Proteomics* (In revision).
- 102 L. Trinkle-Mulcahy, J. Andersen, Y. W. Lam et al., *The Journal of cell biology* **172** (5), 679 (2006).
- 103 S. Boulon, Y. Ahmad, L. Trinkle-Mulcahy et al., *Mol Cell Proteomics* (2009).
- 104 K. M. Cadigan and Y. I. Liu, *Journal of cell science* **119** (Pt 3), 395 (2006).

## 5 References

---

- <sup>105</sup> J. Mao, J. Wang, B. Liu et al., *Molecular cell* **7** (4), 801 (2001).
- <sup>106</sup> F. Mousson, A. Kolkman, W. W. Pijnappel et al., *Mol Cell Proteomics* **7** (5), 845 (2008).
- <sup>107</sup> X. Wang and L. Huang, *Mol Cell Proteomics* **7** (1), 46 (2008).
- <sup>108</sup> N. Yokoyama, D. Yin, and C. C. Malbon, *Journal of molecular signaling* **2**, 11 (2007).
- <sup>109</sup> J. Hendriksen, M. Jansen, C. M. Brown et al., *Journal of cell science* **121** (Pt 11), 1793 (2008).
- <sup>110</sup> A. Enomoto, J. Ping, and M. Takahashi, *Annals of the New York Academy of Sciences* **1086**, 169 (2006).
- <sup>111</sup> F. Hamada and M. Bienz, *Developmental cell* **7** (5), 677 (2004).
- <sup>112</sup> M. Fang, J. Li, T. Blauwkamp et al., *The EMBO journal* **25** (12), 2735 (2006).
- <sup>113</sup> L. F. Waanders, K. Chwalek, M. Monetti et al., *Proceedings of the National Academy of Sciences of the United States of America* **106** (45), 18902 (2009).



## 6 Acknowledgements

My deepest thank goes to the following people, who have supported me over the years. Without you I would not be where I am now! **Thank you:**

**Matthias** for everything. “Paths are made by walking.” (Kafka) and I can say that the past four years in your lab were definitively the best that I could have taken. I appreciated to be part of your lab that is not only one of the best in the world for proteomics but that is also extremely friendly and a place to meet great people from all over the world. You were always my role model and fascinated me by your excellence, passion and dedication at the same time being a down-to-earth person. Thank you for your support and faith.

**Bernhard** for all your great advice and support in the thesis advisory committee and for being my doctoral thesis advisor.

**Henrik** for being the co-referee for this thesis.

**Dieter Langosch** for being the chair of the thesis.

**Tiziana** for your wonderful guidance in the first year of my PhD and further support in the thesis advisory committee. Thank you for sharing your experience both on the job and on the private life.

All my office mates **Tiziana, Michiel, Falk, Peter, Gustavo, Stefan, Lyris, Chris E., Michael, Dirk, Annette, Suman, Tar, Nina, Mara, Bianca, Marco, Markus and Daniela** for the fantastic time we had together. I enjoyed the scientific discussions and the fun we had.

**Tiziana and Michiel** for creating the best spirit - I can imagine - to get even more fascinated by science. I always highly appreciated that.

**Boris** for being a wonderful teacher - patient and helpful- with all my questions about MS and SCX.

**Korbi, Bianca and Florentine** for all their technical support.

**Jürgen** for all your bioinformatic advice and for programming MaxQuant. I came when the first version of MaxQuant came out. Thank you so much for that! You made my life easier.

**Tami and Nina** for critical reading of the thesis.

## 6 Acknowledgements

---

All colleagues for the nice working atmosphere and helpful discussions. I enjoyed sharing a wonderful part of our lives.

Moreover, my collaborators **Tiziana Bonaldi, Florian Gnad, Jürgen Cox, Ivan Matic, Natalie Krahmer and Tobias Walther** for the great and productive work.

**Mara, Chris E., Tami and Leonie** for being my “kitchen-chefs” for one year. I broadened my horizon and even eat exotic food now.

**Leonie, Mara, Tami, Sara, Stefan, Chris E., Chris K., Bo, Boris, Jürgen, Annette, Bianca, Falk, Marion, Tiziana, Michiel and Nina** for the great time outside the lab. We enjoyed the mountains, parties, beer gardens, cinemas, parks or restaurants. Let’s stay in touch.

**Tiziana, Michiel and Nina** for your friendship. I will never forget you.

**Nina** for intensively sharing the past three years of our lives. You are the best friend I ever had!

**Dennis** for all your moral support that was essential to this thesis. Moreover, thank you for sharing your life with a scientist, which is not always easy. I am nothing without you.

Last but definitively not least I would like to thank my parents, **Marion and Wolfgang Hilger**, and my grandmother, **Anneliese Gensch**, for their great support throughout my whole life and their endless love, in which I mutually return.

## 7 Appendix

### 7.1 Overview of publications and unpublished projects

Published Projects
Cox, J., Matic, I., <b>Hilger, M.</b> , Nagaraj, N., Selbach, M., Olsen, J. V., and Mann, M. (2009) A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. <b>Nature protocols</b> 4, 698-705. (7.2)
<b>Hilger, M.</b> , Bonaldi, T., Gnad, F., and Mann, M. (2009) Systems-wide analysis of a phosphatase knock-down by quantitative proteomics and phosphoproteomics. <b>Mol Cell Proteomics</b> 8, 1908-1920. (7.3)
Matic, I., Macek, B., <b>Hilger, M.</b> , Walther, T. C., and Mann, M. (2008) Phosphorylation of SUMO-1 occurs in vivo and is conserved through evolution. <b>Journal of proteome research</b> 7, 4050-4057. (7.4)

Unpublished Projects
<b>Hilger, M.</b> and Mann, M. Triple SILAC to determine stimulus specific interactions in the Wnt pathway. <i>In revision at Mol Cell Proteomics</i> . (7.5)
Krahmer, N.; Guo, Y.; <b>Hilger, M.</b> ; Lingrell, S.; Heger, K.; Newman, H.W.; Schmid-Supprian, M.; Vance, D.E. ; Mann, M.; Farese, Jr, R.V; Walther, T. Phosphatidylcholine Synthesis for Lipid Droplet Expansion: Targeted Activation of CTP: Phosphocholine Cytidylyltransferase, the Rate-Limiting Enzyme. <i>In revision at Cell</i> .
Krahmer, N.*; <b>Hilger, M.*</b> ; Walther, T. Proteome correlation profiling of lipid droplets reveals proteins with coatomer dependent lipid droplet localization. <i>Manuscript in preparation</i> .



## 7.2 Publication:

### **A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics**

Jürgen Cox, Ivan Matic, **Maximiliane Hilger**, Nagarjuna Nagaraj, Matthias Selbach, Jesper V. Olsen and Matthias Mann  
**Nature protocols** 4, 698-705 (2009)

This paper was published in Nature protocols in 2009 and was a joint effort of Jürgen Cox, Ivan Matic, Nagarjuna Nagaraj, Matthias Selbach, Jesper V. Olsen, Matthias Mann and me. Jürgen Cox is the bioinformatician, who programmed MaxQuant and who is the main author of this publication. Together with Ivan Matic and Nagarjuna Nagaraj I contributed to the structure and content of the paper representing the user's point of view. This was the result of an effort to help the MaxQuant users to better understand the data analysis workflow, features, generated output and common error types.



# A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics

Jürgen Cox<sup>1</sup>, Ivan Matic<sup>1</sup>, Maximiliane Hilger<sup>1</sup>, Nagarjuna Nagaraj<sup>1</sup>, Matthias Selbach<sup>2</sup>, Jesper V Olsen<sup>1</sup> & Matthias Mann<sup>1</sup>

<sup>1</sup>Department for Proteomics and Signal Transduction, Max-Planck Institute for Biochemistry, Martinsried, Germany. <sup>2</sup>Department for Cell Signalling and Mass Spectrometry, Max Delbrück Center for Molecular Medicine, Berlin, Germany. Correspondence should be addressed to J.C. (cox@biochem.mpg.de) or M.M. (mmann@biochem.mpg.de).

Published online 16 April 2009; doi:10.1038/nprot.2009.36

**MaxQuant is a quantitative proteomics software package designed for analyzing large mass spectrometric data sets. It is specifically aimed at high-resolution mass spectrometry (MS) data. Currently, Thermo LTQ-Orbitrap and LTQ-FT-ICR instruments are supported and Mascot is used as a search engine. This protocol explains step by step how to use MaxQuant on stable isotope labeling by amino acids in cell culture (SILAC) data obtained with double or triple labeling. Complex experimental designs, such as time series and drug-response data, are supported. A standard desktop computer is sufficient to fulfill the computational requirements. The workflow has been stress tested with more than 1,000 liquid chromatography/mass spectrometry runs in a single project. In a typical SILAC proteome experiment, hundreds of thousands of peptides and thousands of proteins are automatically and reliably quantified. Additional information for identified proteins, such as Gene Ontology, domain composition and pathway membership, is provided in the output tables ready for further bioinformatics analysis. The software is freely available at the MaxQuant home page.**

## INTRODUCTION

Mass spectrometry-based proteomics<sup>1</sup> has become a very data-intensive science, especially since high-resolution Fourier transform mass spectrometers have become widespread<sup>2,3</sup>. For instance, assume there would be 5,000 high-mass precision instruments, e.g., LTQ-Orbitraps<sup>4,5</sup>, running worldwide without downtime, each performing 10 liquid chromatography/mass spectrometry (LC/MS) runs per day and generating 1 GB of raw data per run. Together, this would result in an annual data production rate of 18 PB (=  $1.8 \times 10^{16}$  bytes), which is more than the four experiments in the large hadron collider at CERN—the prime example for data-intensive experimental science—will produce per year (<http://lcg.web.cern.ch/LCG/>). Without any doubt, this situation calls for solid, efficient and standardized data-processing workflows that are widely applicable in quantitative proteomics. In ref. 6, we describe such a computational platform called MaxQuant, which is targeted at high-resolution quantitative data obtained with stable isotope labeling by amino acids in cell culture (SILAC)<sup>7,8</sup> and which we have already successfully applied to a wide range of biological problems<sup>9–15</sup>. Although the novel algorithmic concepts are explained in detail in ref. 6, here we give the step-by-step instructions how to analyze large-scale proteomic data with MaxQuant. Two example data sets are provided, which allow re-analysis of the data from two recent papers of our laboratory<sup>6,14,16</sup>.

SILAC proteome experiments can be performed in many alternative ways, differing for instance in additional protein or peptide fractionation or in the kind of stable isotope labels used. Pre-fractionation of proteins, e.g., by gel electrophoresis, or a separation of digested peptides, e.g., with immobilized pI strips<sup>17</sup>, is compatible with the MaxQuant computational workflow. SILAC data from double- or triple-labeling<sup>18</sup> experiments can be analyzed. Although a standard procedure is to label arginine and lysine when digesting proteins with trypsin, the kind of amino acids that are labeled as well as the atoms within the amino acids that are

replaced by stable heavy isotopes can be freely configured and may be adapted to any existing combination of labels. In addition to simple treatment/control SILAC comparisons, experiments with a more complex design can be analyzed, such as replicate measurements, SILAC time series with common reference points, label switches, multi-sample comparisons, and more. Furthermore, interaction data in the form of SILAC pulldowns<sup>19</sup> can be analyzed. MaxQuant will in each case automatically assemble a matrix of SILAC ratios with rows corresponding to proteins and columns to different ‘samples’ or ‘conditions’ facilitating cross-experiment comparison of protein ratios. Thus, it is preferable in most situations to analyze data from different samples together in one MaxQuant project, so as to enable comparison of quantitative information directly. This is also a precondition for reliable control of protein false discovery rate (FDR).

**Figure 1** shows an overview of the computational pipeline and the types of files that are exchanged. Raw files are generated by the instrumentation software and transferred to the local computer where they are loaded into the ‘Quant’ module. It performs all tasks that can be done before knowing the identity of peptides. In particular, the assembly of isotope patterns into SILAC pairs is already done here, before the submission of data to a tandem mass spectrometry (MS/MS) search engine. An advanced three-dimensional peak and isotope pattern detection is also carried out in this module. Output files are generated containing processed MS/MS spectra (‘msm’ files) bundled together from all LC/MS runs analyzed together, ready for submission to the Mascot search engine<sup>20</sup>. A parameter file containing the search engine parameters is also created (‘par’ file), facilitating the submission of MS/MS spectra with Mascot Daemon (see below). The ‘Identify’ module takes the search engine results, the raw files (as well as intermediate results from the ‘Quant’ module), performs integration and statistical validation, assembles peptides into proteins, quantifies

proteins and writes out several tables containing the results as tab-separated text files (.txt). These can be uploaded for browsing and downstream bioinformatics analysis into programs such as Microsoft Office Excel, R in conjunction with Bioconductor, Spotfire, Matlab and the like.

The MaxQuant website will be expanded as a repository of tools and documents supporting the use of MaxQuant. Presently, we provide protein sequence databases for the most common organisms in MaxQuant-compatible formats ready for upload to Mascot. In addition, here we describe and provide a program ('Sequence Reverser') that allows creating a MaxQuant-compatible FASTA file given the user's own organism-specific protein list. Furthermore, a list of common contaminants—which can be expanded or configured—is included in Sequence Reverser.

## MATERIALS

### EQUIPMENT

#### Hardware requirements

- A personal computer (PC) with at least 2 GB of RAM ▲ **CRITICAL**. At least a dual-core processor is recommended. Most computational parts scale with the number of available computing cores because of parallelization.
- Local storage is used for all raw files belonging to a project, and about half of this size for intermediate results. An external disc connected through USB 2.0 would be sufficient

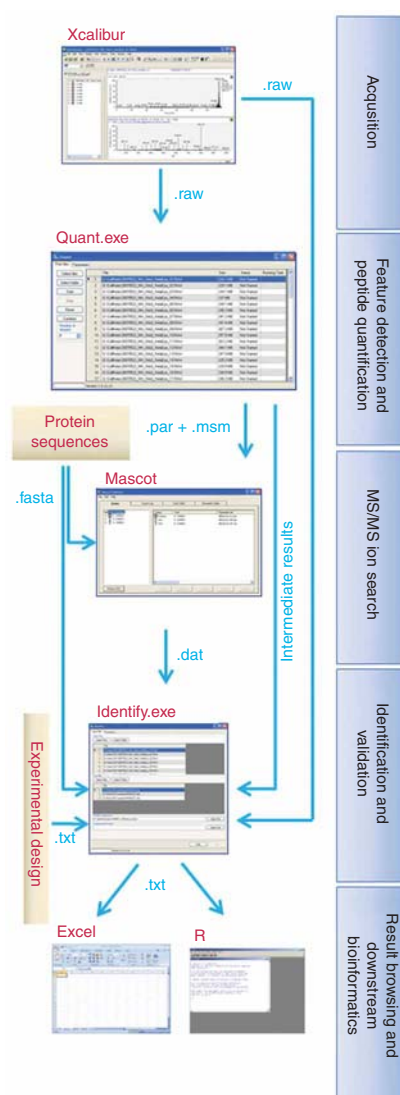
#### Software requirements

- 32 bit versions of Windows XP or Windows Vista operating systems ▲ **CRITICAL** 'Regional and Language Options' have to be set to English.
- .NET Framework 2.0 is to be installed
- Thermo Fisher Scientific Xcalibur software ▲ **CRITICAL** Version must be compatible with your .raw files.
- Access to a Mascot (Matrix Science) server (currently version 2.2)
- Mascot Daemon installed on your local computer. Currently, MaxQuant uses Mascot (version 2.2, Matrix Science) as MS/MS search engine. For convenient and automatic submission of .msm and .par files generated by Quant, we recommend Mascot Daemon, a client application included in the Mascot purchase
- Microsoft Office Excel 2007. Recommended for browsing the result files ▲ **CRITICAL** Older versions of Excel will likely lead to problems due to limitations in allowed numbers of rows and columns.
- Sequence database (see EQUIPMENT SETUP)
- MaxQuant (see EQUIPMENT SETUP)

### EQUIPMENT SETUP

**MaxQuant software installation** Go to MaxQuant home page and navigate to the 'Downloads' section. Please read the software license agreement carefully and stop at this point if you do not agree to its terms and conditions. When downloading MaxQuant you will receive a zipped file containing all necessary binaries and configuration files. Unzip this file (e.g., with WinZip) and store the resulting folder named 'MaxQuant' anywhere on the computer that one is going to use for the computations. No installation script needs to be executed. After saving the 'MaxQuant' folder, one needs to adapt the files in the folder 'MaxQuant\conf' to your local environment. The files 'enzymes,' 'mod\_file,' 'mascot.dat' and 'unimod.xml' are copies of the corresponding files in the configuration folder of Mascot installation. These files have to be identical in the MaxQuant installation and on the Mascot server. Whenever anything is changed in the configuration of the Mascot server, e.g., modifications or sequence databases are added, one has to copy those files from the Mascot server into the MaxQuant installation, because local editing will lead to malfunction. SILAC labels can be configured in the file 'labels.txt,' which is a tab-separated text file and may be edited with Microsoft Office Excel.

**Table 1** contains some examples of definitions for some standard SILAC labels. In the 'Composition' column, the labeled form of the amino acid is specified by its empirical formula. For this purpose, <sup>13</sup>C, <sup>2</sup>H and <sup>15</sup>N are represented by 'Cx,' 'Hx' and 'Nx,' respectively. The 'Mascot Name' column contains the modification as it appears in the 'mod\_file.' Note that in the 'mod\_file' the amino acid type (e.g., 'R') is automatically appended at the end of the name in brackets. 'Short Name' defines how the label is denoted in the graphical user interface of Quant.



**Figure 1** | Overview of the computational workflow. It consists of five steps: the first step—data acquisition—is performed by the vendor software of the mass spectrometer used; in the second step, the 'Quant.exe' module of MaxQuant detects peak features and quantifies peptides; in the third step, a search engine (here Mascot) associates fragment spectra with amino acid sequences; in the fourth step, the 'Identify.exe' module of MaxQuant validates and scores peptide identifications, assembles them to protein identifications and determines protein ratios; and in the fifth step, downstream bioinformatic analysis is performed by general purpose software (spreadsheets), statistical packages or bioinformatic packages.

**Sequence database** We recommend using a species-specific protein sequence database that includes all predicted proteins from an organism with fully sequenced genome. For instance, when analyzing human data, the human International Protein Index database<sup>21</sup> may be used, or, alternatively, all human protein entries contained in Uniprot<sup>22</sup>, including the TrEMBL part, or all ENSEMBL<sup>23</sup> proteins. Redundancy of protein sequences does not matter at this point, as it will be dealt with at a later stage by the software, when assembling the identified peptides to proteins. MaxQuant validates scoring statistics on the basis of the hits to reversed protein entries in a target-decoy database<sup>24</sup>. Therefore, it is necessary to include a reversed version for each original entry in the protein database FASTA file. Reversed entries have to be indicated by a recognizable prefix to the protein ID, e.g., 'REV\_'. In addition, a set of common contaminant proteins can be included, also having a specific prefix, e.g., 'CON\_'. These may, for instance, contain different forms of



# PROTOCOL

**TABLE 1** | The default content of the SILAC label configuration file 'labels.txt'.

Composition	Mascot name	Short name	Amino acid
Cx6H12N4O	Arginine-13C6 (R-13C6) (R)	Arg6	R
Cx6H12Nx4O	Arginine-13C615N4 (R-full) (R)	Arg10	R
C6H8Hx4N2O	Lysine (D4) (K)	Lys4	K
Cx6H12N2O	Lysine-13C6 (K-13C6) (K)	Lys6	K
Cx6H12Nx2O	Lysine-13C615N2 (K-full) (K)	Lys8	K
C6H8Hx3NO	Leucine (D3) (L)	Leu3	L

It may need adaptation depending on the environment.

keratins or of abundant proteins from bovine serum if applicable. Visit the accompanying web page for pre-built MaxQuant-compatible FASTA files for the most common organisms and for downloading the program Sequence Reverser for reversing each entry and adding contaminants of

arbitrary protein sequence collections. The FASTA file should be uploaded to the Mascot server in the normal way as described in the Mascot manual. A copy of exactly the same file needs to be available locally on the computer where MaxQuant is running.

## PROCEDURE

### Preparation of data files

1| Copy all Xcalibur .raw files belonging to a project to a single folder on a local computer. The files may reside on an external hard drive.

▲ **CRITICAL STEP** MaxQuant currently supports only files produced by LTQ-FT-ICR and LTQ-Orbitrap. Additionally, the current version of the Xcalibur software can only open 'raw' files smaller than 2 GB. Therefore, we suggest recording the MS/MS spectra in the centroid mode, which will keep file size sufficiently small for normal length gradients.

2| Make sure that there is sufficient storage space available on the hard disc containing the raw data. In addition to the raw files, there should be at least half of this space available for intermediate results generated by MaxQuant during the calculations.

### Quant.exe: feature detection and peptide quantification

3| Start Quant.exe by double clicking on it. The program is in the 'MaxQuant' folder on your local PC.

4| Go to the 'Raw files' tab. The location of this tab and those of the tabs described in later steps are shown in **Figure 2**.

5| Load the Xcalibur .raw files to be analyzed by clicking the 'Select files' button. Alternately, select all the .raw files in a folder with the 'Select folder' button. The raw files will appear in the main table.

▲ **CRITICAL STEP** All raw files to be analyzed should be in the same folder.

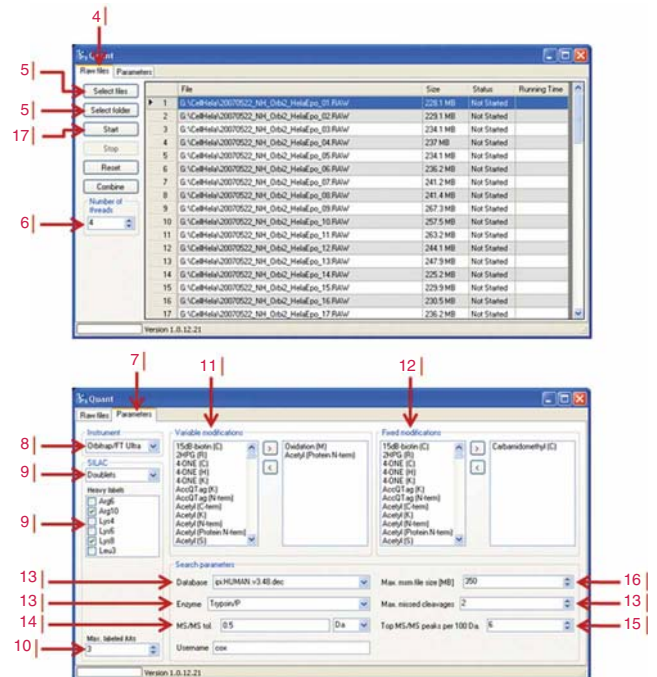
▲ **CRITICAL STEP** MaxQuant's strength is in the analysis of large numbers of LC/MS runs. When analyzing only very few LC/MS runs, the statistical evaluation of peptide identifications may suffer, as a sufficiently high number of identified MS/MS spectra are required to determine histograms used in intermediate steps of the calculation. Although it is possible to analyze even single runs separately with practically usable results, it is recommended to analyze together at least 10 LC/MS runs of reasonably high peptide complexity.

6| Select the number of threads that will be run in parallel by MaxQuant. Each file will be analyzed by one process and the use of multiple processes will considerably shorten the time of analysis. If the number of threads selected is the same or higher than the number of available computing cores, the computer will be very busy and it will hardly be possible to use it for other purposes during the processing time.

7| Go to the 'Parameters' tab.

8| Choose the type of instrument that has produced the files. In the current MaxQuant version, 'FT' and 'Orbitrap/FT Ultra' can be chosen.

9| Select the following types of SILAC experiment:



**Figure 2** | The graphical user interface of the Quant module. The upper panel shows the 'Raw files' tab, whereas in the lower panel, the 'Parameters' tab is shown. The positions in the user interface of Quant.exe that correspond to numbered steps in the procedure are indicated by the step number.

- (a) 'Singlets' if no isotopic labeling was used; peptides and protein are identified and no quantification will be provided.
- (b) 'Doublets' (default selection) in case of a double SILAC labeling; a 'Heavy labels' panel will appear.
- (c) 'Triplets' for a triple SILAC labeling; a 'Medium labels' and a 'Heavy labels' panel will appear.

Select the appropriate labeled amino acids used in the SILAC experiment in the appropriate panels. It is assumed that in the 'Light' SILAC state, all amino acids have a natural isotopic composition. Note that the available labels are read from the labels.txt file in the 'conf' folder mentioned above.

**10|** Specify the maximum number of labeled amino acids ('Max. labeled AAs') a peptide can have to be detected by Quant.

**11|** Add variable modifications by selecting the desired modification on the left of the 'Variable modifications' panel and click on the right arrow button. The modification will then appear in the right panel. Per default oxidation of methionine and N-terminal protein acetylation are used. To remove a modification, select the modification to be removed in the right panel and click on the left arrow button. Variable modifications may or may not be present on a specific residue or a terminus.

**▲ CRITICAL STEP** Modifications related to SILAC labeling must not be specified here, as they are already automatically taken care of by defining the SILAC experiment type in Step 9.

**▲ CRITICAL STEP** Mascot, and consequently MaxQuant, only allows up to nine variable modifications. It is important to note that SILAC modifications (e.g., Arg10 and Lys8) as selected in the SILAC panels are *de facto* variable modifications when the database search is performed on the unpaired isotope patterns. In general, large numbers of variable modifications should be avoided because if they do not occur sufficiently often, they only tend to decrease the number of identifications at fixed FDR, and they will increase search times considerably (combinatorial explosion)<sup>25</sup>.

**12|** Select or deselect, as desired, any 'Fixed modifications' in the same way as described for variable modifications.

Fixed modifications are applied to every occurrence of the specified residue or the terminus. For example, during the digestion reaction, iodoacetamide was used to alkylate cysteines, and to select 'Carbamidomethyl (C)' as fixed modification.

**13|** Choose from the 'Database' box menu the protein sequence database desired to be used in the Mascot search. Select the 'Enzyme' specificity according to the enzyme used during protein digestion. Select the maximum number of missed cleavages a peptide can have in order to be found in the Mascot search.

**▲ CRITICAL STEP** Selections of 'Max.-labeled AAs', 'Enzyme' and 'Max.-missed cleavages' are not independent. For most efficient application of SILAC, the choice of labeled amino acids should coincide with the enzyme specificity. In that case, the maximum number of labeled amino acids should be one more than the number of maximum missed cleavages. We then recommend 3 and 2, respectively.

**14|** Specify the maximum mass deviation allowed for the fragment ions ('MS/MS tol.'). Units can be selected as Dalton (Da) or parts per million (ppm). For a calibrated LTQ, a tolerance of 0.5 Da is recommended<sup>26</sup>. The maximum mass deviation for parent ion masses is determined by MaxQuant on the basis of the achieved mass accuracy and does not need to be specified.

**15|** Select the number of most intense peaks per 100 Da in which Quant will be retained after processing of MS/MS spectra ('Top MS/MS peaks per 100 Da') for the Mascot database search. By default we use six, which is a good balance between scoring correct fragment peaks and suppression of noise.

**16|** Set the size limit in megabytes for the .msm files that Quant creates, containing the processed MS/MS spectra ready for submission to Mascot ('Max. msm file size (MB)'). If a file exceeds this limit, it will be split into two or more parts. A maximum file size of 350 MB is recommended, as for larger files occasional crashes of Mascot searches have been observed. Splitting does not affect results, as Identify.exe assembles data from all Mascot searches.

**17|** Go back to the 'Raw files' tab, press the 'Start' button and wait for the program to finish. A popup will indicate that it is done. In the same folder where the original raw files are stored, Quant will create a folder for each of your raw files containing intermediate results of computations. In addition, a folder 'combined' will be generated, which will, after Quant has finished, contain output files ending with .par and .msm; .msm files contain the processed MS/MS spectra, whereas .par files carry the corresponding parameter settings for the Mascot searches. Depending on the SILAC setting, Quant will create one, three or four kinds of paired .par/.msm files (**Table 2**), which differ in the way in which SILAC labeling-related modifications are treated, either as fixed or as variable modifications. As the SILAC state of many isotope patterns is known earlier, the label modifications are treated as fixed modifications in the Mascot search. For unpaired isotope patterns, the search is done in the conventional way with variable label modifications.

## ? TROUBLESHOOTING

## PROTOCOL

**TABLE 2** | Quant output files for the submission to Mascot.

SILAC	MSM file	PAR file	Description
Singlets	test.iso_0.msm	test.iso.par	MS/MS on single isotope patterns. (No SILAC pair assembly is done)
Doublets	test.iso_0.msm	test.iso.par	MS/MS on single isotope patterns. SILAC labels are treated as variable modifications
	test.sil0_0.msm	test.sil0.par	SILAC state 'light.' No SILAC labels modifications are used
	test.sil1_0.msm	test.sil1.par	SILAC state 'heavy.' Heavy SILAC labels are treated as fixed modifications
Triplets	test.iso_0.msm	test.iso.par	MS/MS on single isotope patterns. SILAC labels are treated as variable modifications
	test.sil0_0.msm	test.sil0.par	SILAC state 'light.' No SILAC label modifications are used
	test.sil1_0.msm	test.sil1.par	SILAC state 'medium.' Medium SILAC labels are treated as fixed modifications
	test.sil2_0.msm	test.sil2.par	SILAC state 'heavy.' Heavy SILAC labels are treated as fixed modifications

MS/MS, tandem mass spectrometry; SILAC, stable isotope labeling by amino acids in cell culture.

### Mascot: MS/MS ion search

**18** | Once Mascot Daemon is launched, click on the 'Task Editor' tab. Insert the name of the task (e.g., 'test sil0') into the field entitled 'Task'.

**19** | Upload one of the .par files in the 'combined' folder into the 'Parameter set' field. Upload the corresponding .msm file(s) by clicking on the 'Add Files...' button or by using drag&drop from an explorer window. Click on the 'Run' button. The display will switch to the 'Status' tab. Once the task is running, the corresponding task icon will change to a clock.

**20** | Repeat Step 19 with the other files until all the searches have been submitted.

### ? TROUBLESHOOTING

**21** | Once the search is finished, expand a task node and click on the result node to see the name of the generated .dat file. There will be one .dat file for each .msm file searched in Mascot.

**22** | Using an internet browser, go the home page of your Mascot server.

**23** | In the address bar of the browser, append '/data' at the end of the address after '.../mascot'. For example, if the address of your Mascot server is http://hansi.biochem.mpg.de:2000/mascot, go to http://hansi.biochem.mpg.de:2000/mascot/data. See the MaxQuant website FAQ for further information.

**24** | Open the folder with the name of the date when the search was performed.

**25** | To save the .dat file, right click and select 'Save Target As...'. It is convenient (but not required) to save them in the combined folder containing the other files of the same project.

### Identify.exe: identification and validation

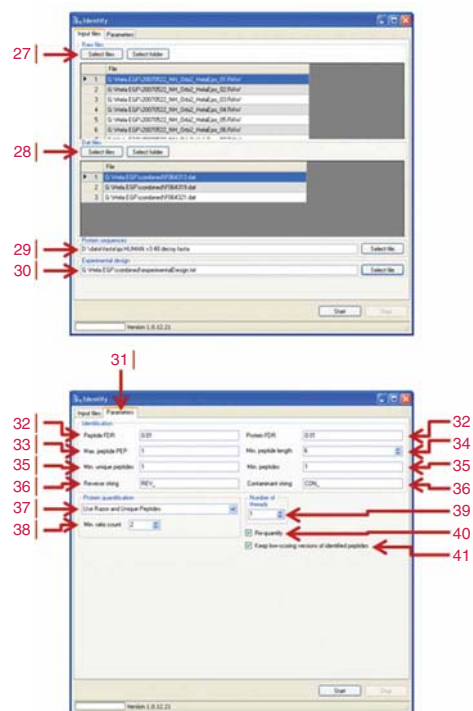
**26** | Locate the program Identify.exe in the 'MaxQuant' folder on the local PC and start it by double clicking it.

**27** | In the 'Input files' tab, upload the .raw files that have been previously run with Quant using either the 'Select files' or the 'Select folder' button. The location of this tab, and the tabs described in later steps, is shown in **Figure 3**.

**28** | Upload the .dat files in the same way.

**29** | Select the appropriate Protein sequences file in the FASTA format.

**▲ CRITICAL STEP** On the local computer, one must have exactly the same sequence database (FASTA) file as used for the Mascot database search. Otherwise, Identify will report an error message and will not be able to finish.



**Figure 3** | The graphical user interface of the Identify module. The upper panel shows the 'Input files' tab, whereas in the lower panel, the 'Parameters' tab is shown. The positions in the graphical interface of the Identify module that correspond to numbered steps in the procedure are indicated by the step number.

## BOX 1 | SPECIFYING THE EXPERIMENTAL DESIGN

To facilitate the inter-sample comparison of protein expression ratios or to assemble data into a specific form—for instance into a time series—the software needs to know which LC/MS runs belong to which ‘time point.’ This can be specified by uploading an experimental design file in Step 30. If no experimental design file is used, SILAC ratios and other protein information are provided as a whole for the entire data set. A template file for the experimental design is created automatically in the ‘combined’ folder. Open this template file in Excel, make appropriate changes and save it as a tab-delimited text file. If the ‘Experiment’ column is filled in, separate average ratios (and other information) will be reported for each different term used in the Experiment column. For instance, suppose one measures a six-point time course with five SILAC experiments by using the time zero sample always as the light state and the samples at five later time points in the heavy states. In addition, one may have separated proteins into 10 gel slices, resulting in 50 LC/MS runs. In this case one would fill the ‘Experiment’ column with five different terms, which would give rise to an expression ratio matrix with five columns. In addition, in the ‘Slice’ column denote the numbers 1–10 regarding which gel slice an LC/MS run corresponds to. Additional statistics on slice-specific identifications of proteins will then be provided, which might, for instance, be useful to detect isoforms differing in molecular weight. In the ‘Invert’ column, one can specify if the ratios originating from certain LC/MS runs should be inverted. This is useful when the labels have been swapped.

- 30| Insert the experimental design file (optional; see **Box 1**).
- 31| Go to the ‘Parameters’ tab.
- 32| Choose the desired FDRs at the protein and peptide levels. For a comprehensive explanation, see ref. 6. The default value is in both cases 0.01 (1%).
- 33| Select the desired maximum posterior error probability, which is the probability of a false hit given the peptide identification score and length of peptides<sup>6</sup>. The default value is 1, corresponding to no additional filtering.
- 34| Set the minimum peptide length. Peptide shorter than the threshold will not be reported nor be considered for protein identification and quantification. Short peptides are usually not unique in the protein database and therefore not statically informative in any case.
- 35| Select the minimum number of unique and total peptides a protein group should have to be considered as identified and reported in the final table.
- 36| Specify the string for reverse and contaminant hits as used in the protein sequences file. We usually use ‘REV\_’ and ‘CON\_’, respectively.
- 37| Specify how the protein ratios will be calculated (‘Protein quantification’). When ‘Use all peptides’ is selected, the quantification is done on all peptides. With ‘Use unique peptides,’ only the peptides unique for that specific protein group are used for quantification. The ‘Use unique and razor peptides’ mode calculates ratios from unique and razor peptides. Razor peptides are non-unique peptides assigned to the protein group with the most other peptides (Occam’s razor principle).
- 38| Select the minimum ratio count for protein quantification. A protein with a lower number of quantified SILAC pairs/triplets will not be used for calculating the protein ratio. As ratio counts are reported in the output files, it is possible to filter the results in downstream applications (e.g., Excel).
- 39| Select the number of parallel threads.
- 40| Tick the ‘Re-quantify’ box if Identify should calculate the ratio for isotopic patterns not assembled in SILAC pairs by Quant. The shapes of the identified isotope pattern will be translated to the place in the  $m/z$  retention time plane, where its missing SILAC partner is expected and intensities will be integrated over these regions. This is particularly helpful for quantifying proteins with very high ratios, e.g., from pulldowns, where one of the SILAC partners is at or below noise level. However, in some cases where extreme ratios are expected, e.g., in incorporation studies, we do not recommend using ‘Re-quantify’.
- 41| If desired, check ‘Keep low-scoring versions of identified peptides.’ If checked, additional MS/MS spectra will be accepted beyond those that have individually passed the criteria for identification. This means that an MS/MS spectrum will be accepted even if its posterior error probability value is not sufficient, as long as the highest scoring peptide sequence for that spectrum has been identified with another MS/MS spectrum. This has no effect on the total number of identified peptide sequences or proteins. This option is particularly beneficial for complex experimental designs, as it increases the likelihood for each protein of finding a ratio in different samples or time points.
- 42| Press the Start button and wait for the program to finish. A popup will appear when it is done.

### ? TROUBLESHOOTING

## PROTOCOL

### ● TIMING

Computation times vary with sample complexity, richness of spectra and LC gradient length. Typical values for 72 LC/MS runs are Steps 1–17: 10 h; Steps 18–25: 4 h; and Steps 26–42: 2 h.

Almost all algorithmic parts of MaxQuant are parallelized and execution is automatically distributed to the number of specified computing cores. Although not tested rigorously, Windows clusters may enhance processing speed significantly. Disk access is likely also limiting and therefore solid-state disks may improve processing speed.

### ? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 3** and in the FAQ on the MaxQuant.org website. For more detailed assistance and troubleshooting with Mascot and Mascot Daemon, refer to the Mascot manual and support.

**TABLE 3** | Troubleshooting table.

Step	Problem	Possible reason	Solution
17	Quant.exe crashes	Maximum file size of 2 GB may be exceeded	Record MS/MS spectra in the centroid mode
20	Quant.exe or Identify.exe crashes Mascot crashes	Storage space may not be enough .msm file may be too big in size	Use an external disc as storage space Use a maximal msm file size of 350 MB or even smaller
42	Identify.exe crashes	FASTA file used for Mascot search missing on your local computer Wrong .dat files uploaded	Copy the exact same to your local computer Check the .dat file number of your 'task' in Mascot Daemon
	Identify.exe reports error of unknown protein ID	Wrong FASTA file uploaded	Use the same database for the Mascot search and Identify.exe
Any	Unforeseen	Unknown	First, double check the FAQ on MaxQuant home page. If still unresolved, describe the problems in the MaxQuant Google group at <a href="http://groups.google.com/group/maxquant-list">http://groups.google.com/group/maxquant-list</a>

MS/MS, tandem mass spectrometry.

## BOX 2 | MAXQUANT OUTPUT TABLES

- **parameters.txt**—summary of parameters used for analysis  
Version number of software used, threshold values used for identification and quantification, mode of quantification and so on
- **msScans.txt**—full scan summary  
Information about the full scan details stored in the raw file including the fill time and total elapsed time, scan numbers, total ion current, dead time and so on
- **msmsScans.txt**—msms scan summary  
Consists of the raw file parameters including the peak type, scan number, total ion count and sequence information wherever there was an identification event
- **allPeptides.txt**—full scan peptides details  
Isotopic clusters, charge state,  $m/z$  values. Mass precision, retention time and whether the peptide is picked up for sequencing or not and so on
- **msms.txt**—summary of scan parameters for identified peptides  
Similar to msmsScans file but restricted to identified peptides only
- **evidence.txt**—one stop master information  
Peptide sequence, protein groups a peptide belong to, modification state, experiments in which peptide was identified, PTM score, mass error calibrated and un-calibrated, MS/MS count, raw file detail, gel slice/fraction and so on
- **peptides.txt**—concise non-redundant list of identified peptide sequences  
Peptide sequence, proteins groups that contain the peptide, modification, miss cleavages, length of peptide, PEP values and so on.  
Identification scores for the best-identified version are displayed
- **modifiedPeptides.txt**—non-redundant list of identified peptide sequences with specific modifications  
Peptide sequence, modification content, PTM score, intensity, Mascot score and so on
- **summary.txt**  
Overall summary of whole analysis including identification success rate for individual raw files, the percentage of isotopic clusters picked up for sequencing and isotope patterns repeatedly sequenced
- **proteinGroups.txt**  
Comprehensive list of identified proteins from the whole analysis, cross-references to various databases including Swissprot, Ensembl, Kegg, GO and so on

## BOX 3 | INTEGRATED PROTEIN ANNOTATION

For the most common organisms, we automatically provide additional protein annotation in the output files. This is based on the mappings of Uniprot IDs to the respective specialized annotation IDs, as they are provided in the latest database dump of Uniprot. The Uniprot IDs for each protein are extracted from the FASTA header by matching the patterns SWISS-PROT:[ID\_1;...;ID\_n] and TREMBL:[ID\_1;...;ID\_n], as they occur, for instance, in the IPI FASTA files. In the proteinGroups.txt table, we report

- Gene Ontology (GO): GO IDs as well as their descriptive names are provided separately for biological process, molecular function and cellular component
- Pfam domain content: Occurrence of protein domain families represented by a hidden Markov model in the Pfam repository is indicated
- Membership of a protein in a KEGG pathway

### ANTICIPATED RESULTS

Identify will create several tables as tab-delimited .txt files that contain the results of the MaxQuant analysis. See **Box 2** for a description of the different output tables. A document with detailed explanation of the columns can be downloaded from the MaxQuant website. Several annotation columns, such as Gene Ontology<sup>27</sup>, Pfam<sup>28</sup> domain content and KEGG<sup>29</sup> pathway membership, are automatically shown in the table of proteins (**Box 3**).

A benchmark data set consisting of 72 LC/MS runs from epidermal growth factor-stimulated HeLa cells used in ref. 6 can be downloaded from <http://www.proteomecommons.org/data/show.jsp?id=7816>. Identification and quantification of more than 4,000 proteins as described in ref. 6 are expected. In another example, Pan *et al.*<sup>16</sup> quantified the proteome of a cell line against primary cells of the same cell type. Furthermore, quantification of more than 4,000 proteins is expected from the analysis of the combined data set, which is also deposited at <http://www.proteomecommons.org>.

**ACKNOWLEDGMENTS** We thank all the other members of the Proteomics and Signal Transduction group for help with the development of MaxQuant. This work was supported by the Max-Planck Society and by the 6th Framework Program of the European Union (Interaction Proteome Grant LSHG-CT-2003-505520 and HEROIC Grant LSHG-CT-2005-018883).

Published online at <http://www.natureprotocols.com/>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

1. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
2. Zubarev, R. & Mann, M. On the proper use of mass accuracy in proteomics. *Mol. Cell. Proteomics* **6**, 377–381 (2007).
3. Mann, M. & Kelleher, N.L. Special feature: Precision proteomics: the case for high resolution and high mass accuracy. *Proc. Natl Acad. Sci. USA* **105**, 18132–18138 (2008).
4. Makarov, A. *et al.* Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* **78**, 2113–2120 (2006).
5. Olsen, J.V. *et al.* Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **4**, 2010–2021 (2005).
6. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
7. Ong, S.E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386 (2002).
8. Mann, M. Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.* **7**, 952–958 (2006).
9. de Godoy, L.M. *et al.* Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455**, 1251–1254 (2008).
10. Bonaldi, T. *et al.* Combined use of RNAi and quantitative proteomics to study gene function in *Drosophila*. *Mol. Cell* **31**, 762–772 (2008).
11. Selbach, M. *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**, 58–63 (2008).
12. Graumann, J. *et al.* Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. *Mol. Cell. Proteomics* **7**, 672–683 (2008).
13. Zanivan, S. *et al.* Solid tumor proteome and phosphoproteome analysis by high resolution mass spectrometry. *J. Proteome Res.* **7**, 5314–5326 (2008).
14. Cox, J. & Mann, M. Is proteomics the new genomics? *Cell* **130**, 395–398 (2007).
15. Schimmel, J. *et al.* The ubiquitin-proteasome system is a key component of the SUMO-2/3 cycle. *Mol. Cell. Proteomics* **7**, 2107–2122 (2008).
16. Pan, C., Kumar, C., Bohl, S., Klingmueller, U. & Mann, M. Comparative proteomic phenotyping of cell lines and primary cells to assess preservation of cell type-specific functions. *Mol. Cell. Proteomics* **8**, 443–450 (2009).
17. Hubner, N.C., Ren, S. & Mann, M. Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis. *Proteomics* **8**, 4862–4872 (2008).
18. Blagoev, B., Ong, S.E., Kratchmarova, I. & Mann, M. Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat. Biotechnol.* **22**, 1139–1145 (2004).
19. Vermeulen, M., Hubner, N.C. & Mann, M. High confidence determination of specific protein-protein interactions using quantitative mass spectrometry. *Curr. Opin. Biotechnol.* **19**, 331–337 (2008).
20. Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 (1999).
21. Kersey, P.J. *et al.* The International Protein Index: an integrated database for proteomics experiments. *Proteomics* **4**, 1985–1988 (2004).
22. UniProt Consortium. The universal protein resource (UniProt). *Nucleic Acids Res.* **36**, D190–D195 (2008).
23. Flicek, P. *et al.* Ensembl 2008. *Nucleic Acids Res.* **36**, D707–D714 (2008).
24. Elias, J.E. & Gygi, S.P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214 (2007).
25. Pevzner, P.A., Mulyukov, Z., Dancik, V. & Tang, C.L. Efficiency of database search for identification of mutated and modified proteins via mass spectrometry. *Genome Res.* **11**, 290–299 (2001).
26. Cox, J., Hubner, N.C. & Mann, M. How much peptide sequence information is contained in ion trap tandem mass spectra? *J. Am. Soc. Mass Spectrom.* **19**, 1813–1820 (2008).
27. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
28. Finn, R.D. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **36**, D281–D288 (2008).
29. Kanehisa, M. *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36**, D480–D484 (2008).

### **7.3 Publication:**

#### **Systems-wide analysis of a phosphatase knock-down by quantitative proteomics and phosphoproteomics.**

**Maximiliane Hilger**, Tiziana Bonaldi, Florian Gnad and Matthias Mann

**Mol Cell Proteomics** 8, 1908-1920 (2009)

This paper was published in Molecular Cellular Proteomics in 2009 and was a joint effort of Tiziana Bonaldi, Florian Gnad, Matthias Mann and me. The paper is the result of the first two main projects of my thesis. I selected all experimental conditions, performed the experiments and analyzed and interpreted the data. Tiziana Bonaldi advised on the biological and mass spectrometric experiments and the data interpretation. Florian Gnad did the main part of the post MaxQuant bioinformatic analysis such as the conservation and gene ontology analysis and generated a Drosophila specific phosphosite predictor.





✂ Author's Choice

# Systems-wide Analysis of a Phosphatase Knock-down by Quantitative Proteomics and Phosphoproteomics<sup>§</sup>

Maximiliane Hilger<sup>‡</sup>, Tiziana Bonaldi<sup>‡§</sup>, Florian Gnad<sup>‡</sup>, and Matthias Mann<sup>‡¶</sup>

Signal transduction in metazoans regulates almost all aspects of biological function, and aberrant signaling is involved in many diseases. Perturbations in phosphorylation-based signaling networks are typically studied in a hypothesis-driven approach, using phospho-specific antibodies. Here we apply quantitative, high-resolution mass spectrometry to determine the systems response to the depletion of one signaling component. *Drosophila* cells were metabolically labeled using stable isotope labeling by amino acids in cell culture (SILAC) and the phosphatase Ptp61F, the ortholog of mammalian PTB1B, a drug target for diabetes, was knocked down by RNAi. In total we detected more than 10,000 phosphorylation sites in the phosphoproteome of *Drosophila* Schneider cells and trained a phosphorylation site predictor with this data. SILAC-based quantitation after phosphatase knock-down showed that apart from the phosphatase, the proteome was minimally affected whereas 288 of 6,478 high-confidence phosphorylation sites changed significantly. Responses at the phosphotyrosine level included the already described Ptp61F substrates Stat92E and Abi. Our analysis highlights a connection of Ptp61F to cytoskeletal regulation through GTPase regulating proteins and focal adhesion components. *Molecular & Cellular Proteomics* 8: 1908–1920, 2009.

Information processing in biological systems relies heavily on activation and inactivation of proteins by phosphorylation. This key post-translational modification is involved in the regulation of most cellular processes and mediates many rapid responses as well as long-term gene expression changes in response to stimuli. Protein kinases and protein phosphatases coordinately regulate this highly dynamic and reversible modification. Phosphorylation is usually studied in a candidate-based approach by *in vitro* kinase assays or by immune techniques employing phospho-specific antibodies. Despite the success of this reduc-

tionist approach, it does not afford a systems-wide observation of the effects upon perturbations of signaling networks.

Recent advances in MS-based<sup>1</sup> proteomics now allow the identification of thousands of phosphorylation sites from complex protein mixtures (1–3). Most large-scale phosphoproteomics studies have been qualitative rather than quantitative; however, isotope-based methods enable precise quantitation of phosphorylation sites between two or more cellular states (4–6). Our group has applied the metabolic labeling technology termed stable isotope labeling by amino acids in cell culture (SILAC) (7) for the quantitative comparison of phosphoproteomes. For example, we quantified phosphorylation dynamics in response to epidermal growth factor stimulation. Out of a measured phosphoproteome of several thousand sites only a minority (about 10%) was regulated by the signal, highlighting the importance of quantitation in pinpointing specific systems responses (8).

*Drosophila* is a well established model system to study key players in cell signaling and development. Genetic studies have been performed for decades whereas more recently also RNA interference (RNAi) has been employed for gene function studies using a highly efficient silencing protocol (9). A further advantage of *Drosophila* as a model system is the lower degree of functional redundancy compared with higher vertebrates while maintaining a high level of conservation of human genes linked to disease (10).

Two large-scale, non-quantitative *Drosophila* phosphoproteome studies were carried out in embryonic Kc167 cells (11) and embryos (12). Both studies identified more than 10,000 sites of the *Drosophila* phosphoproteome.

We have recently adapted the SILAC methodology for quantitative proteomics to *Drosophila*. Schneider line 2 (SL2) cells were treated with either mock dsRNA or dsRNA against ISWI, a component of chromatin remodeling complexes. The

From the <sup>‡</sup>Proteomics and Signal Transduction, Max-Planck Institute for Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany and <sup>§</sup>Experimental Oncology, European Institute of Oncology, Via Adamello 16, 20139 Milano, Italy

✂ Author's Choice—Final version full access.

Received, December 8, 2008, and in revised form, March 26, 2009  
Published, MCP Papers in Press, May 9, 2009, DOI 10.1074/mcp.M800559-MCP200

<sup>1</sup> The abbreviations used are: MS, mass spectrometry; Abl, Ablson kinase; Abi, Abl interacting protein; DHB, 2,5-dihydroxybenzoic acid; GO, Gene Ontology; LC-MS, liquid chromatography-mass spectrometry; MS/MS, tandem mass spectra; RNAi, RNA interference; SCX, strong cation exchange chromatography; SILAC, stable isotope labeling by amino acids in cell culture; TFA, trifluoroacetic acid; RT, reverse transcriptase; SL2, Schneider line 2; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; GST, glutathione S-transferase; dsRNA, double-stranded RNA; DAS, distributed annotation system.

combination of RNAi and SILAC allows the unbiased “phenotypization” of the gene knock-down directly at the proteome level (13).

Here we determined a high-quality basal phosphoproteome in SL2 cells and characterized its structural and evolutionary properties. We compared kinase substrate motives between *Drosophila* and human and trained a *Drosophila* phosphorylation site predictor.

To explore the potential of quantitative phosphoproteomics in a systems-wide manner, we focused on the *Drosophila* non-transmembrane tyrosine phosphatase Ptp61F. This phosphatase is the ortholog of mammalian PTB1B, which is thought to be involved in type 2 diabetes, obesity, and cancer (14), and which is the target of several ongoing drug development projects (15). Ptp61F is a negative regulator of JAK/STAT signaling (16, 17) and, together with the Abl kinase (Abl), involved in the regulation of the Abl interacting protein (Abi) and lamella formation (18). Both PTP1B and Ptp61F are among the best studied protein tyrosine phosphatases in their respective organisms; however the characterization of their substrates is still far from complete. Two recent mass spectrometric studies employed substrate trapping to identify direct substrates of PTP1B and Ptp61F (19, 20). The PTP1B study was combined with phosphotyrosine peptide enrichment, which led to site-specific detection of potential PTP1B targets. PTP1B function was additionally investigated by quantitative phosphotyrosine proteomics comparing wild type and PTP1B-deficient fibroblasts. In contrast, the Ptp61F study identified potential substrates without site-specific information. One of these was PVR, the *Drosophila* homolog of VEGFR and PDGFR, suggesting that Ptp61F - like its mammalian counterpart - counteracts receptor tyrosine kinase signaling. Apart from Abi, further components of the SCAR/WAVE complex as well as its regulatory kinase Abl were identified as potential Ptp61F substrates. This supports an involvement of Ptp61F in the regulation of actin reorganization and remodeling.

To study the role of Ptp61F in a global and unbiased approach we combined global quantitative phosphoproteome analysis with RNA interference. We profiled tyrosine, serine and threonine phosphorylation changes upon ablation of Ptp61F by RNAi. In parallel, we quantified changes in the proteome, which allowed us to normalize changes in phosphorylation sites to corresponding changes at the protein level. Interestingly, we observed increased tyrosine phosphorylation of the protein tyrosine kinase Abl which suggests an enhanced Abl activity upon Ptp61F RNAi. We additionally detected up-regulated phosphotyrosine sites on GTPase regulating proteins (like RhoGAP15B and Vav) and constituents of focal adhesions (like Paxillin and Lasp) which expand the proposed involvement of Ptp61F in the regulation of cytoskeleton organization. Our work represents proof-of-principle that the combination of large-scale phosphoproteomics and a loss-of-function approach can contribute significantly to elu-

cidating the role of key players in phosphorylation-dependent signaling. Importantly, this systems-wide approach measures the net effect of the perturbation on the entire signaling network, without the need to define specific substrate-kinase or substrate -phosphatase relationships or other direct functional mechanisms.

#### EXPERIMENTAL PROCEDURES

**Cell Culture—*Drosophila*** Schneider SL2 cells were a kind gift from Peter Becker’s laboratory. Cells were grown in custom-made Schneider media at 26 °C (13). Schneider media were supplemented by yeast extract and fetal bovine serum dialyzed overnight using a molecular weight cutoff of 1.2 kDa to remove free amino acids. “Heavy” and “light” SILAC media were prepared by adding 0.4 g/liter  $^{13}\text{C}_6^{15}\text{N}_4$  L-arginine (Arg-10) and 1.65 g/liter  $^{13}\text{C}_6^{15}\text{N}_2$  L-lysine (Lys-8) (Sigma Isotec) or the corresponding non-labeled amino acids,  $^{12}\text{C}_6^{14}\text{N}_4$  L-arginine (Arg-0) and  $^{12}\text{C}_6^{14}\text{N}_2$  L-lysine (Lys-0), respectively. Media were always prepared freshly before the experiments. To obtain two labeled cell populations the cells were grown for 8–9 days in heavy or light SILAC media.

**Phosphatase Inhibitor Treatment of SILAC-labeled SL2 Cells—**SILAC-labeled SL2 cells were treated with an inhibitor mixture for 15 min. The mixture contained 0.3 mM Pervanadate (prepared from 100 mM sodium orthovanadate (Sigma-Aldrich) mixed with 100 mM  $\text{H}_2\text{O}_2$ ), 33 nM Calyculin A (Millipore), and 3  $\mu\text{M}$  deltamethrin (Merck).

**RNAi of SILAC-labeled SL2 Cells—** $7.7 \times 10^6$  SL2 cells were re-suspended in 7 ml of heavy or light serum-free SILAC media and seeded in a 75  $\text{cm}^2$  tissue culture flask. 210- $\mu\text{g}$  Ptp61F dsRNA was added to the heavy or 210- $\mu\text{g}$  control GST dsRNA to the light condition, respectively (less than 200 nM). The cells were incubated for one hour at 26 °C and subsequently 14 ml of heavy or light serum-containing SILAC media were added. After 4 days two-third of the SILAC media were refreshed, and after 8 days the cells were harvested. The Ptp61F dsRNAs (forward: 5’-TTAATACGACTCACTATAGGGAGAGGATGTGAATCCGTACGATCAT-3’; reverse: 5’-TTAATACGACTCACTATAGGGAGAGAACTGCATCACTTCACGACTCT-3’) and GST dsRNAs (forward: 5’-TTAATACGACTCACTATAGGGAGAA-TGTCCCTATACTAGGTTA-3’; reverse: 5’-TTAATACGACTCACTATAGGGAGAACGCATCCAGGCACATTG-3’) were synthesized according to Worby *et al.* (9) using MEGAshortscript™ kit (Ambion).

**Preparation of Cell Extract—**Heavy and light SL2 cells were counted, and equal cell numbers were combined. The cell mixture was pelleted by centrifugation at 1000 rpm for 5 min, washed with cold phosphate-buffered saline, and lysed under rotation for 60 min at 4 °C using radioimmune precipitation assay buffer (containing 1% Triton, 0.1% SDS, 0.1% sodium deoxycholate, 140 mM NaCl, 1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride, 1 mM EDTA, 10 mM Tris, pH 8) supplemented with protease inhibitors (complete tablets; Roche Diagnostics) and phosphatase inhibitors (1 mM sodium orthovanadate, 5 mM NaF, 5 mM beta-glycerophosphate). The lysate was cleared by centrifugation, and the resulting supernatant was subjected to digestion.

**Protein Digestion—**For in-gel digestion the whole cell lysate was separated by one-dimensional SDS-PAGE (4–12% Novex mini-gel; Invitrogen) and visualized by colloidal Coomassie staining. For the proteome analysis 100- $\mu\text{g}$  whole cell lysate was separated in two adjacent lanes and identically cut into 15 slices whereas for the phosphoproteome analysis 300  $\mu\text{g}$  of cell lysate was separated in one lane and cut into 8 slices. All gel slices were subjected to in-gel digestion with trypsin (21, 22). Resulting tryptic peptides were extracted with 30% ACN in 3% TFA. For phosphoproteome analysis extracted peptides were subjected to  $\text{TiO}_2$  chromatography. For proteome analysis samples were concentrated until full evaporation of

organic solvent and reconstituted with 1% TFA and 5% ACN. Subsequently peptides were concentrated and desalted on reversed phase C<sub>18</sub> StageTips (23, 24). Peptides were eluted by 30- $\mu$ l buffer B (80% ACN in 0.5% acetic acid) solution into a sample 96-well plate (Abgene), concentrated in a SpeedVac until removal of the organic solvent and reconstituted with a one-to-one mix of buffer A (0.5% acetic acid) and buffer A<sup>+</sup> (2% ACN in 1% TFA).

For in-solution digestion whole cell extracts were mixed with four volumes of ice-cold acetone to precipitate proteins. Precipitated proteins were collected by centrifugation and dissolved in 6 M urea, 2 M thiourea, 10 mM HEPES or 10 mM Tris and 1% *n*-octyl glucoside. 2-mg (phosphoproteome with phosphatase inhibitor glucoside) or 20-mg proteins (phosphoproteome of Ptp61F RNAi) were reduced with 1 mM dithiothreitol for 45 min at RT and then alkylated with 5.5 mM iodoacetamide for 30 min. Reduced and alkylated proteins were digested with 200- $\mu$ g endopeptidase Lys-C (Waco) for 3 h. Subsequently the peptide mixture was diluted with four volumes deionized water and digested with 200- $\mu$ g sequencing grade modified trypsin (Promega) overnight at RT. Trypsin was inactivated by acidification with TFA to pH < 3, and the peptide mixture was subjected to phosphopeptide enrichment.

**Phosphopeptide Fractionation and Enrichment**—Phosphopeptides were fractionated and enriched from in-solution digested peptides by the combination of strong cation exchange chromatography (SCX) and TiO<sub>2</sub> chromatography or by multiple rounds of TiO<sub>2</sub> chromatography in batch mode. Alternatively phosphopeptides were enriched by TiO<sub>2</sub> chromatography from in-gel digested peptide fractions.

SCX was carried out as described (25). The resulting 15 fractions were pooled according to the 230-nm absorbance intensity. For the phosphatase inhibitor phosphoproteome experiment 10 fractions and the flow-through were further enriched for phosphopeptides by TiO<sub>2</sub> chromatography (26). For the Ptp61F RNAi experiment 15 fractions and the flow-through (loading 20 mg at once) and 13 fractions and the flow-through (performing two separate SCX runs with 10 mg each for better separation) were subjected to TiO<sub>2</sub> chromatography (25). For the phosphatase inhibitor phosphoproteome experiment both the TiO<sub>2</sub> beads (kindly provided by GL Sciences) and the samples were pre-incubated with 5 mg/ml 2,5-dihydroxybenzoic acid (DHB) in 80% ACN before the sample, and ~5 mg of the TiO<sub>2</sub> beads were combined and incubated batchwise with end-over-end rotation for 1–3 h. For the Ptp61F phosphoproteome the beads were pre-incubated with 30 mg/ml DHB and added directly to the SCX fraction to reduce the amount of DHB in the samples.

After incubation, the beads were washed twice with 10% ACN in 0.1% TFA and once with 80% ACN in 0.1% TFA solution. Bound peptides were eluted from TiO<sub>2</sub> beads once with 1% ammonium solution, pH 10.5, in 20% ACN and twice with 1% ammonium solution, pH 10.5, in 40% ACN into 20  $\mu$ l of 2% ACN in 1% TFA solution. Eluates were subsequently dried to ~5  $\mu$ l in a SpeedVac and reconstituted with 2% ACN in 1% TFA solution for LC-MS/MS analysis.

**LC-MS/MS Analysis**—The samples were analyzed by a nanoflow HPLC (Agilent 1100 or 1200; Agilent Technologies) coupled on-line via a nano-electrospray ion source (Proxeon Biosystems) to a LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific). Peptide mixtures were loaded onto a C<sub>18</sub>-reversed phase column (15 cm long, 75- $\mu$ m inner diameter, packed in-house with ReproSil-Pur C<sub>18</sub>-AQ 3- $\mu$ m resin) in buffer B with a flow rate of 500 nl/min for 20 min and eluted with a linear gradient from 2% to 40% buffer B (80% ACN and 0.5% acetic acid solution) at a flow rate of 250 nl/min over two hours. After each sample the column was washed with 90% buffer B and re-equilibrated with buffer A.

Mass spectra were acquired in the positive ion mode applying a data-dependent automatic switch between survey scan and tandem mass spectra (MS/MS) acquisition. Proteome samples were analyzed

with a top5 method, acquiring one Orbitrap survey scan in the mass range of *m/z* 300–2000 followed by MS/MS of the five most intense ions in the LTQ. The target value in the LTQ-Orbitrap was 1,000,000 for survey scan at a resolution of 60,000 at *m/z* 400 using lock masses for recalibration to improve the mass accuracy of precursor ions (27). Fragmentation in the LTQ was performed by collision-induced dissociation with a target value of 5,000 ions. Ion selection threshold was 500 counts. Selected sequenced ions were dynamically excluded for 90 s.

For the phosphoproteome acquisition a total scan cycle consisted of three mass ranges, 350–1050, 850–1750, and 350–1750 with MS/MS of the five most intense ions for the first two ranges and MS/MS of the seven most intense ions for the last range. Multi-stage activation was enabled upon detection of a neutral loss of phosphoric acid (97.97, 48.99, or 32.66 amu) (28) for further ion fragmentation. Selected sequenced ions were dynamically excluded for 300 s after sequencing. Otherwise settings were the same as described before for the proteome acquisition.

**Data Analysis**—Mass spectrometric data were analyzed with the in-house developed software MaxQuant (version 1.0.11.5 for the total basal SL2 phosphoproteome and version 1.0.12.33 for the spectra extraction as well as for the proteome and phosphoproteome of Ptp61F RNAi) (29, 30). MS/MS spectra were searched by Mascot (version 2.2.04, Matrix Science) against the FlyBase (version 5.4) (containing 20,822 entries) combined with 175 common contaminants and concatenated with the reversed versions of all sequences. The following parameters were set for the Mascot searches. Trypsin allowing for cleavage N-terminal to proline and cleavage between aspartic acid and proline was chosen as enzyme specificity. Cysteine carbamidomethylation was selected as a fixed modification, whereas protein N-terminal acetylation, methionine oxidation and serine, threonine and tyrosine phosphorylation were selected as variable modifications. Depending on a priori knowledge about the number of Arg and Lys in the precursor ion determined by MaxQuant before the search, Arg-10 and Lys-8 were used as additional fixed or variable modifications. Maximally three (version 1.0.11.5) or two missed cleavages (version 1.0.12.33) and three labeled amino acids were allowed. Initial mass deviation of precursor ion and fragment ions were up to 7 ppm and 0.5 Da, respectively. MaxQuant automatically quantified SILAC peptides and proteins. SILAC protein ratios were calculated as the median of all peptide ratios assigned to the protein. A false discovery rate of 0.01 was required for proteins and peptides with a minimum length of 6 amino acids. In addition a posterior error probability for each MS/MS spectrum below or equal to 0.1 was required. In case the identified peptides of two proteins were the same or the identified peptides of one protein included all peptides of another protein, these proteins (e.g. isoforms and homologs) were combined by MaxQuant and reported as one protein group. Phosphorylation sites were made non-redundant with regards to their surrounding peptide sequence. Nevertheless all alternative proteins that matched a particular phosphosite were reported as one group. The PTM score was used for assignment of the phosphorylation site(s) as described (8). Class I phosphorylation sites are defined by a localization probability of 0.75, and probability localization score difference higher or equal to 5 (8).

**Quantitative RT-PCR**—Total RNA was extracted from 1.5-ml cell suspension 8 days after dsRNA treatment using PrepEase™ RNA spin kit (USB). The protocol included a DNA digestion step with DNase I on the spin column. Reverse transcription was carried out from 2- $\mu$ g total RNA using first strand cDNA synthesis kit (Fermentas) according to manufacturer's protocol with oligo(dT)<sub>18</sub> primers. Quantitative real-time PCR was performed using iQ™ SYBR® Green Supermix (BIO-RAD) in the MyiQ Real-Time PCR System (BIO-RAD). Each reaction mix (25  $\mu$ l) contained 10 ng of cDNA, 12.5  $\mu$ l of SYBER Green Supermix, and 12.5 pmol of forward and reverse primers. Two

different primer pairs were used to analyze Ptp61F mRNA levels (Ptp61F 1 forward: 5'-AGCTGCACGATCCACCTT-3', Ptp61F 1 reverse: 5'-GTGGCAGATTAAGCGATTGGA-3'; Ptp61F 2 forward: 5'-CAGCAGACATGATCCACGA-3', Ptp61F 2 reverse: 5'-TGTCGTCCTCGTCATCGTCAT-3'). GAPDH levels were measured for normalization via  $2^{-\Delta\Delta CT}$  method (31) (GAPDH forward: 5'-ATGAAGGTGGTCTCCAACGC-3'; reverse: 5'-TCATCAGACCCTCGACGATCT-3'). Three technical replicates of each sample (Ptp61F RNAi) and control (GST RNAi) were performed.

**Motif Extraction**—Motif-X was used to derive potentially new phosphorylation motifs *in silico* and to confirm the observations of the  $\chi^2$ -Test. A probability value of 0.0001 was considered significant. In addition, a minimum occurrence of 20 was required to derive a significant consensus sequence. The approach was performed residue-specific on all Class I sites using the entire fly proteome as a background model.

**Phosphorylation Site Prediction**—A support vector machine was trained separately for each amino acid on high-confidence phosphorylation sites as described in Gnad *et al.* (32). The essential feature of each instance that was used for this machine learning approach was the raw sequence: the phosphorylated residue along with its surrounding sequence ( $\pm 6$  residues). This yields 260 dimensions in the feature space reflecting the surrounding amino acids. Each feature dimension presents a certain residue on a certain position within the given 13 amino acid sequence window. To generate a negative set of the same size, we randomly chose sites from fly proteins that were not, to date, detected to be phosphorylated. The positive and negative datasets were split into a training set (90%) and a test set (10%). The parameters  $C$  and  $\sigma$  were optimized by varying them from  $2^{-10}$  to  $2^{10}$  in multiplicative steps of two on the basis of a 5-fold cross validation on the training set. The optimal  $C$  and  $\sigma$  parameters for predicting phosphoserines were 32 and 0.125, respectively. In the case of phosphothreonines the best parameters were 4 and 0.125, respectively. We trained the optimal model for each set of each phospho amino acid separately using the radial basis function.

**Data Availability**—Identified phosphorylated peptides and phosphorylated sites along with further information ranging from evolutionary conservation to structural constraints and Swiss-Prot annotations were uploaded to Phosida, the phosphorylation site database (32). Besides the fly phosphoproteome, Phosida also contains phosphorylation sites of mouse, human, and various prokaryotes and now provides the described *Drosophila*-specific phosphorylation site predictor. Furthermore, the genome annotation section of Phosida links directly to the EnsEMBL database, which displays our proteomic data on the basis of the DAS source management. The setup of Phosida DAS layers for a given gene of interest in EnsEMBL is described in the "Background" section of Phosida and in EnsEMBL.

Moreover, the proteomic data (identified peptides) were uploaded to MAPU - Max-Planck Unified Proteome Database. The detected peptides, which have been assigned to gene entries as described in this study, are also available via the MAPU DAS source in the EnsEMBL database.

## RESULTS AND DISCUSSION

**High-confidence Determination of the *in Vivo* SL2 Phosphoproteome**—Most of the phosphoproteomics experiments are performed in a qualitative manner. Here we employed the SILAC method instead to measure quantitative changes in the levels of *in vivo* phosphorylation sites. In a first SILAC experiment, we compared basal phosphorylation to phosphatase inhibitor mixture-induced phosphorylation. This involves application of cell permeable inhibitors of serine/threonine and

tyrosine protein phosphatases (see below) and has an analytical rather than a functional goal; it is advantageous because we trigger the identification of very low abundant endogenous phosphorylation sites that are up-regulated in response to the inhibitor treatment. Furthermore, the presence and mass spacing of the SILAC pair distinguish peptides from noise and allow counting the number of labeled amino acids, thereby decreasing the search space in database identification (30). The second SILAC experiment in which we quantified changes in the phosphoproteome upon RNAi-mediated knock-down of a phosphatase (see below) has both an analytical and a functional goal: it increases the measured endogenous phosphoproteome, which we subsequently analyzed bioinformatically, and it supplies information on the cellular roles of the phosphatase. Analytically, these two experiments lead to the detection of different sets of the *Drosophila* phosphoproteome (Fig. 1 and supplemental Tables 8 and 9). All phosphorylation sites in either SILAC experiments that are present in the control state (manifest as SILAC pairs) thus reflect the endogenous phosphorylation status of the *Drosophila* SL2 cells here termed "basal phosphoproteome".

First we SILAC-labeled SL2 cells and treated the heavy cell population with an inhibitor mixture against serine/threonine and tyrosine protein phosphatases for 15 min. The two SILAC populations were combined in equal amounts, and their total protein extract was in-solution digested. To enrich phosphopeptides, we employed both multiple rounds of  $TiO_2$  chromatography in batch mode and the combination of SCX followed by  $TiO_2$  chromatography (supplemental Fig. 1). Phosphopeptide fractions were separated by reverse-phase chromatography and on-line detected in a high-resolution linear ion trap-orbitrap instrument (LTQ-Orbitrap) (LC-MS/MS) (Fig. 1A). A phosphorylation site not responding to inhibitor treatment is detected as a phosphopeptide peak pair with the same peak intensities (ratio H/L = 1). An up-regulated phosphorylation site has a ratio H/L > 1 and a down-regulated one H/L < 1. In total, we identified 6,752 high-confidence phosphorylation sites (Class I sites: 99% peptide identification confidence and >75% amino acid localization confidence (8)). Of the phosphotyrosine sites, 88% were regulated upon phosphatase inhibitor treatment whereas only 44% of the phosphoserine/phosphothreonine sites changed by more than 1.5-fold (supplemental Fig. 2). This likely reflects the high efficiency of pervanadate in inhibiting tyrosine phosphatases (33, 34).

Next, we applied quantitative phosphoproteomics to RNAi of the phosphatase Ptp61F by down-regulating its expression in the heavy cell population. Cells were SILAC-labeled and dsRNA treated from the same day because both processes take about the same number of cell doublings and are completely independent. Both in-gel digestion coupled with  $TiO_2$  chromatography and in-solution digestion in combination with SCX and  $TiO_2$  chromatography were performed (supplemental Fig. 3). Since dsRNA treatment for 8 days can also alter the proteome, we additionally analyzed global protein expression

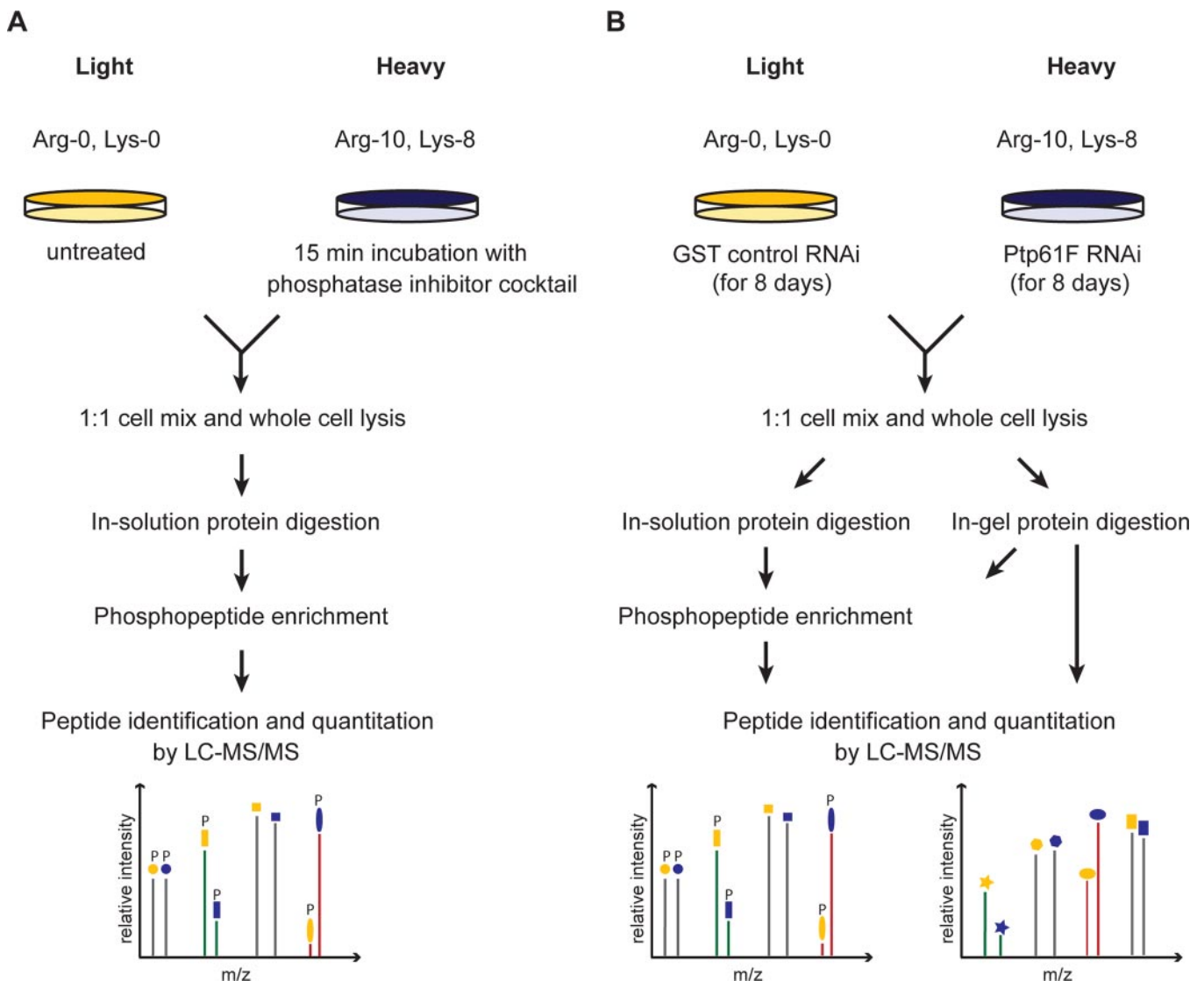


FIG. 1. Basal phosphoproteome of *Drosophila* SL2 cells. A, experimental procedure for quantitative phosphoproteomics with phosphatase inhibitor mixture treatment of the heavy labeled SL2 cell population. Phosphopeptides were enriched by applying two different strategies (supplemental Fig. 1) and analyzed by high-resolution mass spectrometry. B, experimental procedure for quantitative proteomics and phosphoproteomics in response to Ptp61F deletion by RNA interference. For proteome analysis extracted proteins were digested in gel after one-dimensional SDS-gel electrophoresis. For phosphoproteome analysis extracted proteins were both in-gel and in-solution digested and enriched for phosphopeptides (supplemental Fig. 3).

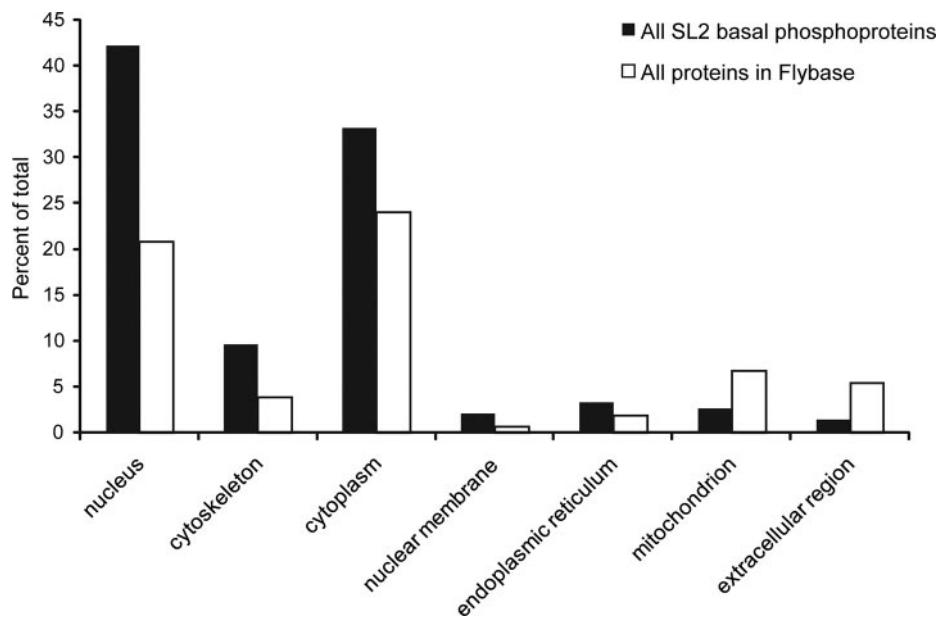
changes (Fig. 1B). This allows normalization of the phosphorylation changes with respect to protein levels.

All quantitative MS data were analyzed together as described (29, 30, 33, 35). In total, we detected 10,043 high-confidence phosphorylation sites on 2,379 phosphoproteins of which 9,749 were quantifiable and represented our basal phosphoproteome (supplemental Table 9). This number is similar to other recent large-scale *Drosophila* phosphoproteome datasets (11, 12). There is a good overlap of phosphorylated proteins (65.9% to Zhai *et al.* 12 and 75.2% to Bodenmiller *et al.* 11) (supplemental Fig. 4A), especially because the other investigations used a different cell line (Kc167) under various conditions or homogenized embryos. However,

while 1,274 phosphorylated proteins were identified in all three studies, 334 phosphoproteins were exclusively found in our study indicating that none of the *Drosophila* phosphoproteomes is exhaustive yet. This observation is even more striking at the level of phosphorylation sites. We detected 4,691 novel phosphorylation sites, while 5,051 sites (51.8%) were already covered by the other two studies (supplemental Fig. 4B).

**Bioinformatics Analysis of the SL2 Phosphoproteome**—Insects are at considerable evolutionary distance to mammals, in which most phosphoproteomic studies have been done. *Drosophila*, in particular, serves as an important model organism, but so far it is not known how closely the phosphoproteome

FIG. 2. Over- and under-represented cellular compartments in the basal phosphoproteome resulting from Gene Ontology analysis.



resembles that of mammals. We therefore investigated the overall properties of our high-confidence SL2 phosphoproteome using Gene Ontology (GO), kinase motif analysis, structural constraints, and evolutionary conservation.

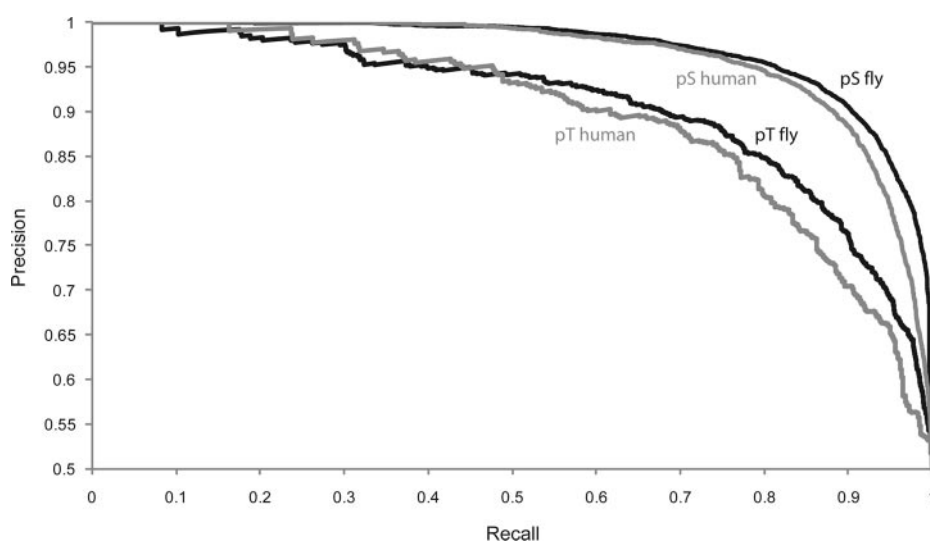
*Overall Properties of the Drosophila Phosphoproteome Are Similar to the Human Phosphoproteome*—GO cellular compartment analysis, using Cytoscape (36) and BiNGO (37), revealed mitochondria and extracellular compartment to be underrepresented while nucleus, cytoplasm, and cytoskeleton were overrepresented (all with  $p < 0.001$ ) (Fig. 2 and supplemental Table 1). Regulatory protein classes such as transcription factors, protein kinases, and GTPase regulating proteins were highly overrepresented in GO molecular function analysis ( $p < 10^{-15}$ ; supplemental Table 1). Furthermore, GO analysis after combining our dataset with that of Bodenmiller *et al.* (11) and Zhai *et al.* (12) (which yielded 4,871 phosphoproteins), showed similar trends, suggesting they are valid for the entire *Drosophila* phosphoproteome (supplemental Table 2). They also mirror recent analyses of a human and mouse cell line (8, 33).

*Phosphorylation Substrates in Drosophila Can Largely be Predicted by Human Kinase Motives*—To ask whether SL2 phosphorylation sites match significantly with known human kinase motifs, we employed the  $\chi^2$ -Test as described in Pan *et al.* (33). The  $\chi^2$ -Test assesses the significance of the variance between observed and expected frequencies. This showed that fly phosphorylation sites matched significantly to most of the known kinase motifs, derived from human cells (supplemental Table 3). As an example, the CDK1 motif p[ST]-P-X-[KR] matches with the surrounding amino acids of identified phosphorylation sites six times more frequently than expected by chance. In the case of human the fold enrichment was twenty.

As described in Schwartz *et al.* (38), Motif-X is an iterative approach to derive significantly overrepresented motifs from large-scale datasets. A motif is built via recursive extraction of significant amino acid combinations in a given foreground set (surrounding sequence of phosphorylation sites) compared with a given background set. The application of Motif-X confirmed the findings of the  $\chi^2$ -Test (supplemental Table 4). For example, the second most significant consensus sequence was the CDK1 motif. We next applied Motif-X to phosphorylation sites detected in human HeLa cells (8). Overall, 27 significant sequence motifs matched exactly with those derived from the *Drosophila* set (supplemental Fig. 5) indicating a high degree of conservation of kinases and their signaling pathways ranging from CDK to ERK. This is in concordance with a previous comparison of various eukaryotic kinomes to the human kinome by Manning *et al.* (39). For example, fly and human share several kinase families involved in immunity, neurobiology as well as basic cellular functions such as the cell cycle (40).

*Phosphorylation Sites Are Largely Confined to Loops and Hinges on the Protein Surface*—Next we assigned secondary structure and accessibility constraints to phosphorylation sites employing the prediction method SABLE 2.0 (41). SABLE assigns the predicted secondary structure (coil, helix, or  $\beta$ -sheet) to each residue of a given input sequence along with calculated accessibility values ranging from 0 (fully buried) to 9 (fully exposed). As in our previous analysis of the human phosphoproteome (32), phosphorylation sites are almost exclusively located in coil regions of the surfaces of proteins (supplemental Fig. 6A). In total, 91% of phosphorylated serines were found to be located in loop regions in comparison to 77% of the serines that were not identified to be phosphorylated on the same proteins. The resulting degrees

FIG. 3. Precision-Recall curves reflecting the accuracy of phosphoserine and phosphothreonine prediction using the fly-specific predictor. The recall reflects the proportion of true positives to the sum of true positives and false negatives, whereas the precision describes the number of true positives out of all predicted positives.



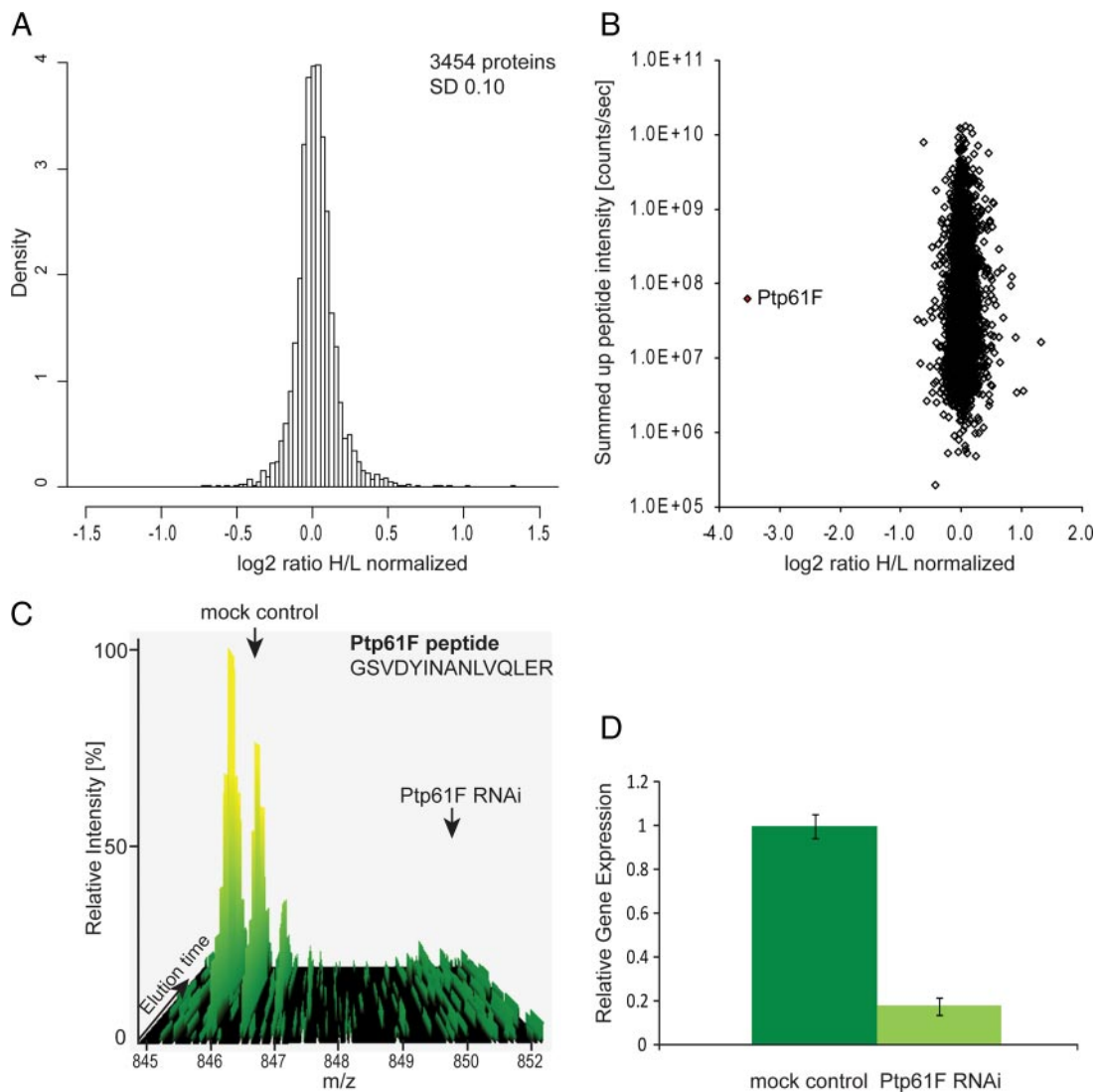
of accessibilities calculated by SABLE 2.0 were also significantly higher for phosphorylated residues compared with their non-phosphorylated counterpart (supplemental Fig. 6B). This is important for kinase and phosphatase substrate accessibility, as well as for subsequent functional effects on the substrate protein.

***Drosophila melanogaster* Phosphorylation Site Predictor**—Because none of the *Drosophila* phosphoproteome datasets reaches completion, we used a machine learning approach to predict phosphorylation events *in silico* (32). We trained a support vector machine on the full dataset of 7,756 Ser(P), 1,427 Thr(P), and 325 Tyr(P) along with their surrounding sequences. Serines, threonines, and tyrosines that have not been proven to be phosphorylated were randomly chosen from the FlyBase database and defined as negative set. We used 90% of the given instances for training and 10% for testing the accuracy of prediction (5-fold cross validation). We found that 89.8% in the serine set and 81.1% in the threonine set were predicted correctly (Fig. 3). The accuracy of the prediction of phosphotyrosines was 63.0%, due to their smaller number in the training set. Note that this accuracy assumes equal numbers of phosphorylated and non-phosphorylated sites; therefore the accuracy for a given residue in the proteome will be lower. The predictor may prove useful to classify potential phosphorylation sites on proteins of interest, for example, for subsequent mutational studies.

Application of the *Drosophila*-specific predictor to the human phospho dataset (8) showed only somewhat inferior performance compared with the predictor trained on human data (Fig. 3). This further supports the conclusion that human and insect phosphoproteome are generally similar overall. The predictor is available online via the Phosida database. Phosida also lists matching kinase motifs for any given predicted and experimentally found phosphorylation site, thus suggesting potential responsible kinases.

**Evolutionary Conservation of the *Drosophila* Phosphoproteome**—We used FlyBase (42) to assign our proteomic data to the genome and Ensembl (43) for chromosomal localization. The SL2 phosphoproteome (quantified and non-quantified sites) encompassed 2,454 gene transcripts, corresponding to 2,411 distinct genes. For annotation of the *Drosophila* genome, we employed the DAS source system using the Proserver technology (44, 45). This combination allows easy checking of individual genes and proteins starting from the genome as well as visualizing overall distribution patterns in the entire genome. As expected, our dataset did not show a specific pattern in the genome (supplemental Fig. 7).

The Ensembl Compara database contains phylogenetic relationships between fly and around 40 eukaryotes ranging from yeast to human. Besides the phylogenetic classifications between homologous proteins, it also stores global cDNA alignments. With assigned gene transcripts in hands, we compared the conservation of detected phosphorylated proteins with the one of all proteins from FlyBase that are defined as “known” due to former experimental evidence. We found that phosphorylated proteins show more orthologs throughout all eukaryotes than all other annotated proteins (supplemental Table 5 and supplemental Fig. 8). For example, around 70% of proteins that are phosphorylated in fly have orthologous counterparts in mosquito; in comparison, around 60% of the other known proteins are homologous to mosquito proteins. The same tendency was found for the homology to more distantly related species such as human or even yeast. The results of the analysis on the inter-species level are in concordance with ones of the intra-species level. We found that more than half (55%) of all fly proteins show paralogs, whereas a mere 27% of the phosphorylated subset shows homology within the same species. The higher proportion of non-phosphorylated proteins showing paralogy reflects faster evolution.



**FIG. 4. RNAi of the tyrosine phosphatase Ptp61F and the effect on the proteome.** *A*, histogram of log<sub>2</sub>-transformed normalized protein ratios for all quantified proteins in the Ptp61F RNAi experiment. *B*, plot of the normalized ratios of all quantified proteins plotted against their summed heavy and light peptide intensities. *C*, MS-based analysis of RNAi depletion: a peptide-pair for Ptp61F from the SILAC-based Ptp61F RNAi proteome experiment, where the phosphatase was depleted in the heavy samples. The Ptp61F peptide (GSVDYINANLVQLER) shows a heavy to light ratio (H/L) of 0.073 corresponding to a more than 10-fold reduction of the protein level. *D*, quantitative RT-PCR of Ptp61F transcript derived from mock control and Ptp61F RNAi cells. Ptp61F mRNA level is reduced to less than 20% by RNA interference. *Error bars* represent the standard deviation of three different technical replicates. GAPDH transcript levels were used for normalization.

In addition, we assessed the preservation of the protein sequence for each homologous protein pair using the rates of synonymous and non-synonymous base pair substitutions ( $d_N/d_S$  ratio) (46). The median  $d_N/d_S$  ratio of orthologous phosphoproteins was found to be 0.25 compared with 0.33 in the case of all homologous fly proteins, indicating that there is negative selection for amino acid changes within the sequence of phosphorylated proteins.

In conclusion, evolutionary analysis indicates strong conservation of phosphoproteins, presumably because of their key regulatory functions; for example, in cell signaling. Phylogenetic relationships and global alignments between homologous proteins are displayed in the Phosida database.

*Proteome-wide Analysis of Phosphatase Ptp61F RNAi*—To in-depth characterize proteome- and phosphoproteome-wide responses of the cell to the depletion of the phosphatase Ptp61F, we combined SILAC-based proteomics with RNA interference. The quantitative MS data of the proteome and phosphoproteome analysis of Ptp61F RNAi (Fig. 1B) were analyzed together as described (29, 30, 33, 35). In total 4,132 proteins were identified with a false discovery rate of 1% and at least two unique peptides; out of these 3,454 proteins were quantified with a minimum of three quantitation events (three SILAC pairs) (supplemental Table 10). The histogram of the normalized H/L SILAC ratios is shown in Fig. 4A and is somewhat narrower than a normal distribution, indicating



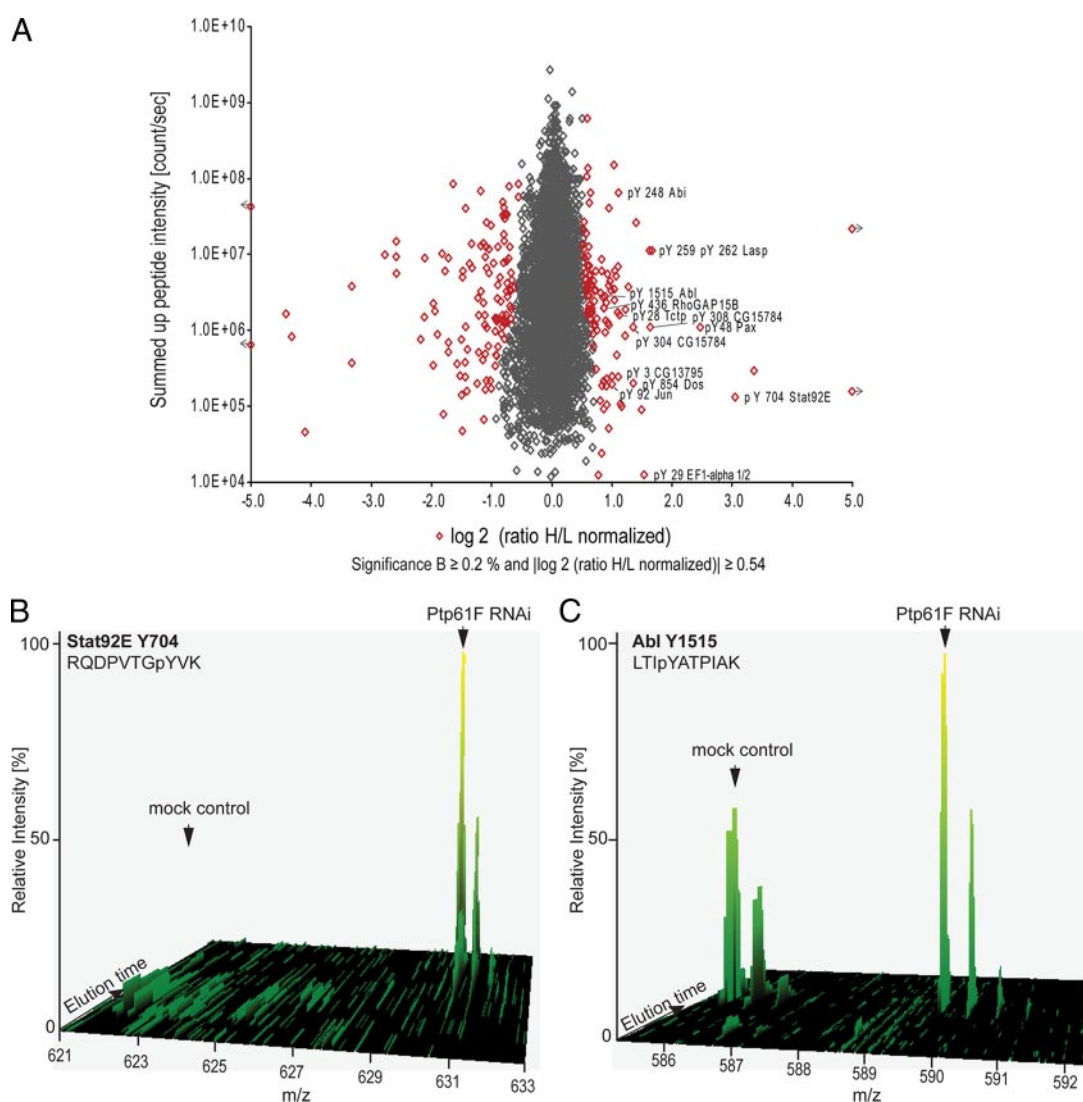


FIG. 5. **Quantitative phosphoproteome upon Ptp61F RNAi.** A, plot of normalized ratios of all quantified phosphorylation sites against their summed heavy and light peptide intensities ( $n = 6537$ ). B, SILAC peptide pair of STAT92E, the most up-regulated phosphotyrosine peptide upon Ptp61F RNAi. C, SILAC peptide- pair of Abl, 1.9-fold up-regulation on a C-terminal autophosphoylation site.

that no gross proteome changes occurred. In Fig. 4B the same ratios are plotted as a function of added peptide signal. The highest significant outlier is the phosphatase, which is down-regulated more than 90% (Fig. 4, B and C). This agrees with 80% decrease of its mRNA in real-time PCR measurements (Fig. 4D).

The tight distribution around the one-to-one ratio indicates the high accuracy of protein quantitation. We defined significantly regulated proteins as those that changed at least 1.5-fold and the fold change of which had an intensity-dependent statistical significance of at least 0.002 (termed “significance B”). Applying this level of stringency the detected proteome changes are restricted to sixteen up-regulated proteins and six down-regulated proteins (supplemental Table 6). This suggests - at least at the level of the

proteome that was quantified in our experiments - that the response to the depletion of the phosphatase has limited effect at the proteome level, and that our RNAi had few if any off-target effects. The proteins found up-regulated are mainly proteins involved in metabolic pathways like the glycolytic enzyme phosphofructokinase or the glutamine synthase 2.

An important purpose of the proteome analysis was the possibility to correct changes at the phosphorylation level by changes at the protein expression level. In our dataset 77% of the phosphorylation sites could be normalized to their corresponding protein expression levels (supplemental Table 10). Interestingly, the remaining phosphorylation sites were not significantly associated with low intensity phosphopeptides. Thus they probably reflect the different selectivity of phos-

phopeptide enrichment compared with the whole proteome measurement.

**Global Phosphoproteome Response to Ptp61F RNAi**—In total 6,478 phosphorylation sites on 1,904 proteins were detected upon Ptp61F RNAi (supplemental Table 10). Applying the same criteria as for the proteome, 288 phosphorylation sites changed significantly, comprising 217 serines, 45 threonines, and 26 tyrosines (Fig. 5A). Thus, in total 4.4% of all detected phosphorylation sites were affected upon depletion of a single tyrosine phosphatase. Direct targets of a tyrosine phosphatase should show up-regulated tyrosine phosphorylation upon phosphatase RNAi. (Note that we use “up-regulated phosphorylation” to denote increase in phosphorylation after accounting for protein expression changes.) We detected 15 proteins with 19 up-regulated tyrosine phosphorylation sites (supplemental Table 7 and Fig. 6). The sole *Drosophila* Stat (STAT92E) serves as positive control because Ptp61F is a known suppressor of STAT92E-dependent transcription (16, 17). The phosphorylation site Tyr-704 of this transcription factor determines its dimerization and accumulation in the nucleus (47, 48). It was found 10-fold up-regulated upon Ptp61F RNAi (Fig. 5B), which represents the most extreme tyrosine responder in the entire dataset.

One category of proteins with up-regulated tyrosine phosphorylation sites are likely direct substrates of Ptp61F because they are also found in a recent substrate trapping assay (20). These are the non-receptor protein tyrosine kinase Abl, its interacting protein Abi (also found by Huang *et al.* (18)), the focal adhesion protein Lasp, the GTPase activating protein RhoGAP15B, and the sorting nexin DSH3PX1. However, the substrate trapping experiment did not identify specific sites, and our experiment now provides this important information (Fig. 6). The kinase Ack was also identified by substrate trapping. The high-confidence tyrosine phosphorylation site Tyr-987 of Ack was up-regulated 1.3-fold but was not statistically significant by our definition (Fig. 5A). This was also the case for the tyrosine phosphorylation sites of SCAR (1.3-fold) and Pvr (1.2-fold), suggesting that our significant thresholds are conservative.

The focal adhesion component Paxillin and the guanine-nucleotide exchange factor Vav are potential novel targets of Ptp61F with a functional link to cytoskeleton organization. In support, mammalian Paxillin was identified in a PTP1B substrate assay experiment (19). Furthermore, the same study found increased tyrosine phosphorylation of mammalian Vav3 comparing wild type and in PTP1B-deficient mouse fibroblasts by quantitative phosphotyrosine proteomics. The tyrosines of the identified Paxillin and Vav phosphorylation site are conserved but not yet described as PTP1B target sites. The translation elongation factor EF1-alpha, the vacuolar protein sorting 4 homolog CG6842, the adaptor protein Dos, the GTPase Rab1, and Prosap are potential novel Ptp61F substrates with conserved tyrosine sites in human (supplemental Fig. 10). The positive regulator of Stat92E-dependent signaling CG15784 and

the guanine-nucleotide exchange factor Tctp are novel potential substrates where either the protein or the detected regulated phosphorylation site is not conserved in human.

Our approach is designed to measure systems-wide effects of pathway perturbation in an unbiased way, rather than to discriminate direct and indirect actions of the phosphatase. In addition to up-regulation of direct targets, changes in the tyrosine phosphorylation levels can also be due to activation of substrate tyrosine kinases. Illustrating this point, we found the direct target Abl with increased phosphorylation on Tyr-540 within the activation loop and on Tyr-1515 within a potential autophosphorylation sequence at the C terminus (Fig. 5C). The activation loop phosphorylation of Abl has been detected in oncogenic Abl kinases *in vivo* (49), and this is the first direct observation of endogenous activation at this site upon Ptp61F depletion. Enhanced tyrosine phosphorylation of Abl in general and activation loop phosphorylation in particular increases its kinase activity (50, 51). Activated Abl propagates the effects of the Ptp61F RNAi to a second layer of effector molecules. Several proteins with up-regulated tyrosine phosphorylation in our experiment are known to be phosphorylated by Bcr-Abl or c-Abl in human cell lines. These include Paxillin (52, 53) and Vav1 (54).

Abi is an interesting example because its phosphorylation is regulated by both Abl and Ptp61F (18). Thus its increased phosphorylation upon Ptp61F RNAi is the combined effect of increased Abl activity and depleted phosphatase. Our *in vivo* screen shows that the net effect is a 1.5-fold up-regulation of Tyr-248. Similarly, the up-regulation of Paxillin and Vav phosphorylation sites, in our experiment could reflect the separate or combined action of Abl and Ptp61F. Since mammalian Paxillin was also identified in a PTP1B substrate assay experiment (19) a dual action of Abl and Ptp61F on Paxillin is likely.

With regards to the up-regulation of several metabolic enzymes at the protein level, it is interesting that Bcr-Abl signaling has been linked to up-regulation of phosphofructokinase expression level (55). Furthermore, Bcr-Abl signaling activates c-Jun NH<sub>2</sub>-terminal kinase (JNK) (56), possibly providing the mechanism for increased threonine phosphorylation on the fly homolog dJun within the transactivation domain (57).

Altogether, we detected 101 up- and 115 down-regulated phosphorylation sites upon Ptp61F RNAi (supplemental Table 7), emphasizing the broad effects of the phosphatase on the phosphoproteome. There are a number of serine/threonine kinases whose phosphorylation level is altered such as Fray (1.6-fold up) and MEKK-1 (1.3-fold up), which although not all passing our threshold for significance could be the mechanistic link to these broad effects. Note that phosphorylation of a kinase can either enhance or inhibit its activity.

**Conclusion**—Here we have, for the first time, combined quantitative proteomics and phosphoproteomics to study the effects of perturbing phosphorylation-dependent signal-

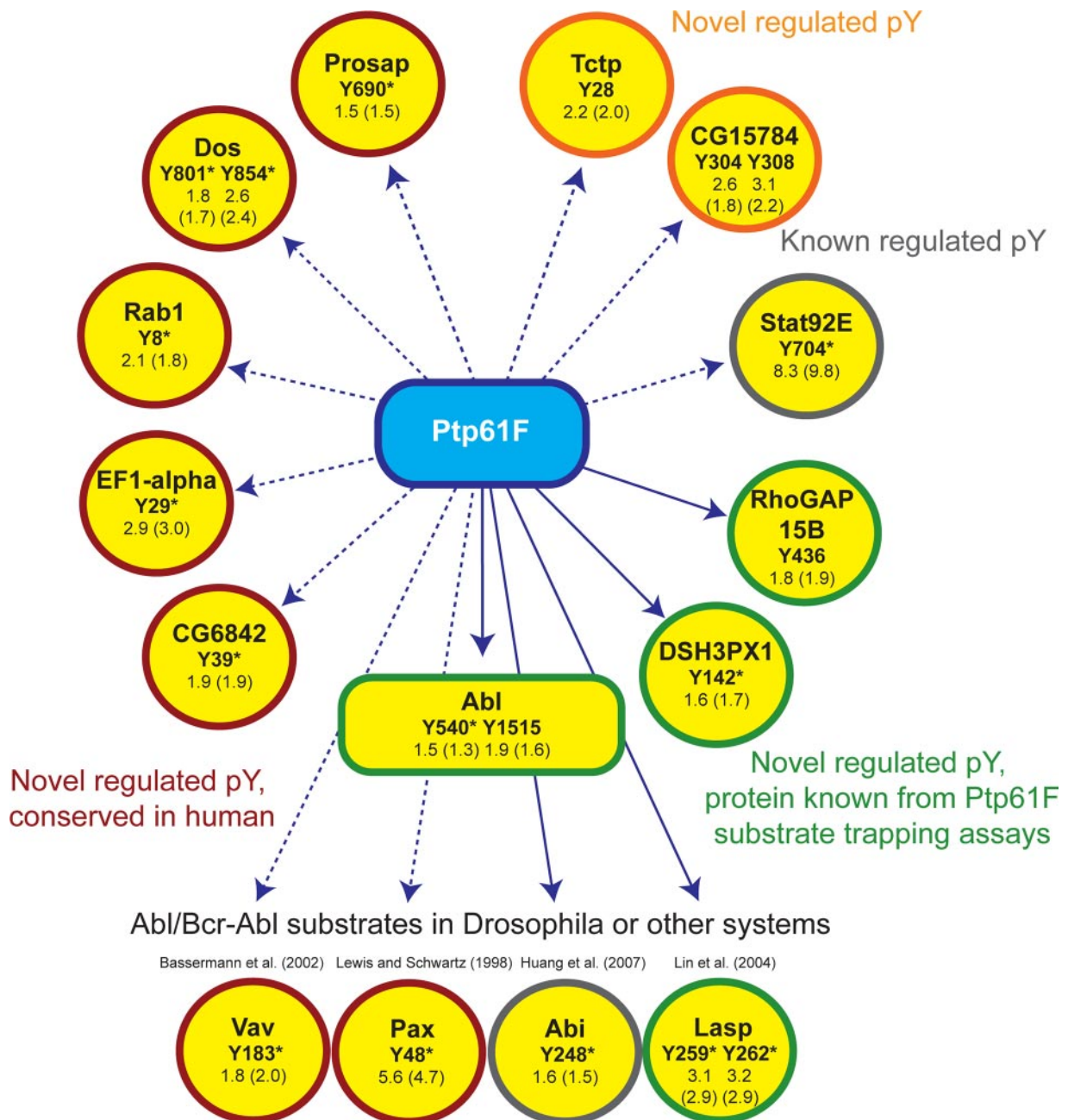


FIG. 6. **Direct and indirect targets of Ptp61F.** The H/L phosphopeptide ratios upon Ptp61F RNAi and in brackets their corresponding values normalized to the protein expression are indicated. Here only phosphotyrosines with normalization and a ratio belonging to a singly phosphorylated peptide are shown. Tyrosines with asterisk are conserved in human (supplemental Fig. 9).

ing network by ablating one of its components. This approach differs from “chemical proteomics” involving ATP-based enrichment or classical *in vitro* approaches in that it measures the net effect of the perturbation on the signaling network and the expressed proteome in a completely generic and hypothesis-free approach. Effects on the phosphorylation pattern are captured at the quantitative phosphoproteomics level, whereas resulting gene expression changes are measured by quantitative proteomics. These

effects are both upstream and downstream of the intermediate output measured by mRNA-based technologies.

We also introduced the concept of parallel acquisition of a proteome and a phosphoproteome for the system-wide analysis of a perturbed system. This is generally necessary in perturbations that last longer than the time required for protein expression changes (generally 20–30 min) because it enables assigning changes to the stoichiometry of the phosphorylation site (phosphorylation change relative to protein

amount) and the absolute phosphorylation changes resulting from a protein expression change. Biologically, the total amount of phosphorylation, the relative fraction of phosphorylation, or both, may be important.

We chose to ablate the *Drosophila* homolog of a disease relevant phosphotyrosine phosphatase, Ptp61F by RNAi. The detected potential substrates were enriched for the categories cytoskeleton regulation and adhesion signaling and thus our data further support and expand a role of Ptp61F in these processes. Our results show a limited effect of the Ptp61F RNAi on the proteome in contrast to a pronounced effect on the phosphoproteome. This indicates that the phosphatase predominantly affects phosphorylation-mediated signaling rather than downstream gene expression. Therefore a microarray experiment may not have been efficient at defining the most pertinent effects of this phosphatase knock-down.

Our proof of principle study clearly demonstrates the power of combined quantitative proteomics and phosphoproteomics to study system-wide effects of signaling network perturbation. With increasing throughput and automation of MS, signaling networks could routinely be probed in this way. Although we used RNAi, other obvious perturbations include gene knock-outs of kinases and phosphatases as well as small molecule inhibitors.

**Acknowledgments**—Juergen Cox and Nadin Neuhauser helped in the analysis of large-scale datasets, and Jesper Olsen advised on the phosphoproteome acquisition. Shubin Ren contributed to the construction of the *Drosophila* phosphorylation site predictor. We thank Reinhard Faessler and Ralf Boettcher of the Department for Molecular Medicine for critical reading and suggestions. We thank Ewan Birney, Andrew Jenkinson, and Albert Bertan of the European Bioinformatics Institute for help with establishing the DAS server and the use of the European Bioinformatics Institute Compara database.

☐ The on-line version of this article (available at <http://www.mcponline.org>) contains supplemental material.

✉ To whom correspondence may be addressed. Ph.: +49-89-8578-2557; Fax: +49-89-8578-2219; E-mail: [mmann@biochem.mpg.de](mailto:mmann@biochem.mpg.de).

#### REFERENCES

- Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**, 198–207
- Schmelzle, K., and White, F. M. (2006) Phosphoproteomic approaches to elucidate cellular signaling networks. *Curr. Opin. Biotechnol.* **17**, 406–414
- Witze, E. S., Old, W. M., Resing, K. A., and Ahn, N. G. (2007) Mapping protein post-translational modifications with mass spectrometry. *Nat. Methods* **4**, 798–806
- Ong, S. E., and Mann, M. (2005) Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **1**, 252–262
- Ficarro, S. B., McClelland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F., and White, F. M. (2002) Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **20**, 301–305
- White, F. M. (2008) Quantitative phosphoproteomic analysis of signaling network dynamics. *Curr. Opin. Biotechnol.* **19**, 404–409
- Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386
- Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006) Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635–648
- Worby, C. A., Simonson-Leff, N., and Dixon, J. E. (2001) RNA interference of gene expression (RNAi) in cultured *Drosophila* cells. *Sci. STKE* 2001, PL1
- Rubin, G. M., Yandell, M. D., Wortman, J. R., Gabor Miklos, G. L., Nelson, C. R., Hariharan, I. K., Fortini, M. E., Li, P. W., Apweiler, R., Fleischmann, W., Cherry, J. M., Henikoff, S., Skupski, M. P., Misra, S., Ashburner, M., Birney, E., Boguski, M. S., Brody, T., Brokstein, P., Celniker, S. E., Chervitz, S. A., Coates, D., Cravchik, A., Gabrielian, A., Galle, R. F., Gelbart, W. M., George, R. A., Goldstein, L. S., Gong, F., Guan, P., Harris, N. L., Hay, B. A., Hoskins, R. A., Li, J., Li, Z., Hynes, R. O., Jones, S. J., Kuehl, P. M., Lemaitre, B., Littleton, J. T., Morrison, D. K., Mungall, C., O'Farrell, P. H., Pickeral, O. K., Shue, C., Voshall, L. B., Zhang, J., Zhao, Q., Zheng, X. H., and Lewis, S. (2000) Comparative genomics of the eukaryotes. *Science* **287**, 2204–2215
- Bodenmiller, B., Malmstrom, J., Gerrits, B., Campbell, D., Lam, H., Schmidt, A., Rinner, O., Mueller, L. N., Shannon, P. T., Pedrioli, P. G., Panse, C., Lee, H. K., Schlapbach, R., and Aebersold, R. (2007) PhosphoPeP—a phosphoproteome resource for systems biology research in *Drosophila* Kc167 cells. *Mol. Syst. Biol.* **3**, 139
- Zhai, B., Villén, J., Beausoleil, S. A., Mintseris, J., and Gygi, S. P. (2008) Phosphoproteome analysis of *Drosophila* melanogaster embryos. *J. Proteome Res.* **7**, 1675–1682
- Bonaldi, T., Straub, T., Cox, J., Kumar, C., Becker, P. B., and Mann, M. (2008) Combined use of RNAi and quantitative proteomics to study gene function in *Drosophila*. *Mol. Cell* **31**, 762–772
- Dubé, N., and Tremblay, M. L. (2005) Involvement of the small protein tyrosine phosphatases TC-PTP and PTP1B in signal transduction and diseases: from diabetes, obesity to cell cycle, and cancer. *Biochim. Biophys. Acta* **1754**, 108–117
- Johnson, T. O., Ermolieff, J., and Jirousek, M. R. (2002) Protein tyrosine phosphatase 1B inhibitors for diabetes. *Nat. Rev. Drug Discov.* **1**, 696–709
- Baeg, G. H., Zhou, R., and Perrimon, N. (2005) Genome-wide RNAi analysis of JAK/STAT signaling components in *Drosophila*. *Genes Dev.* **19**, 1861–1870
- Müller, P., Kutenkeuler, D., Gesellchen, V., Zeidler, M. P., and Boutros, M. (2005) Identification of JAK/STAT signaling components by genome-wide RNA interference. *Nature* **436**, 871–875
- Huang, C. H., Lin, T. Y., Pan, R. L., and Juang, J. L. (2007) The involvement of Abl and PTP61F in the regulation of Abi protein localization and stability and lamella formation in *Drosophila* S2 cells. *J. Biol. Chem.* **282**, 32442–32452
- Mertins, P., Eberl, H. C., Renkawitz, J., Olsen, J. V., Tremblay, M. L., Mann, M., Ullrich, A., and Daub, H. (2008) Investigation of protein tyrosine phosphatase 1B function by quantitative proteomics. *Mol. Cell. Proteomics* **7**, 1763–1777
- Chang, Y. C., Lin, S. Y., Liang, S. Y., Pan, K. T., Chou, C. C., Chen, C. H., Liao, C. L., Khoo, K. H., and Meng, T. C. (2008) Tyrosine phosphoproteomics and identification of substrates of protein tyrosine phosphatase dPTP61F in *Drosophila* S2 cells by mass spectrometry-based substrate trapping strategy. *J. Proteome Res.* **7**, 1055–1066
- Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858
- Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., and Mann, M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1**, 2856–2860
- Rappsilber, J., Mann, M., and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906
- Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670
- Macek, B., Mijakovic, I., Olsen, J. V., Gnad, F., Kumar, C., Jensen, P. R., and Mann, M. (2007) The serine/threonine/tyrosine phosphoproteome of the model bacterium *Bacillus subtilis*. *Mol. Cell. Proteomics* **6**, 697–707

26. Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., and Jørgensen, T. J. (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell. Proteomics* **4**, 873–886
27. Olsen, J. V., de Godoy, L. M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. (2005) Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **4**, 2010–2021
28. Schroeder, M. J., Shabanowitz, J., Schwartz, J. C., Hunt, D. F., and Coon, J. J. (2004) A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. *Anal. Chem.* **76**, 3590–3598
29. Cox, J., and Mann, M. (2007) Is proteomics the new genomics? *Cell* **130**, 395–398
30. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372
31. Livak, K. J., and Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* **25**, 402–408
32. Gnad, F., Ren, S., Cox, J., Olsen, J. V., Macek, B., Oroshi, M., and Mann, M. (2007) PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol.* **8**, R250
33. Pan, C., Gnad, F., Olsen, J. V., and Mann, M. (2008) Quantitative phosphoproteome analysis of a mouse liver cell line reveals specificity of phosphatase inhibitors. *Proteomics* **8**, 4534–4546
34. Deleted in proof
35. Graumann, J., Hubner, N. C., Kim, J. B., Ko, K., Moser, M., Kumar, C., Cox, J., Schöler, H., and Mann, M. (2008) Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. *Mol. Cell. Proteomics* **7**, 672–683
36. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504
37. Maere, S., Heymans, K., and Kuiper, M. (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448–3449
38. Schwartz, D., and Gygi, S. P. (2005) An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol.* **23**, 1391–1398
39. Manning, G., Plowman, G. D., Hunter, T., and Sudarsanam, S. (2002) Evolution of protein kinase signaling from yeast to man. *Trends Biochem. Sci.* **27**, 514–520
40. Bettencourt-Dias, M., Giet, R., Sinka, R., Mazumdar, A., Lock, W. G., Balloux, F., Zafirooulos, P. J., Yamaguchi, S., Winter, S., Carthew, R. W., Cooper, M., Jones, D., Frenz, L., and Glover, D. M. (2004) Genome-wide survey of protein kinases required for cell cycle progression. *Nature* **432**, 980–987
41. Adamczak, R., Porollo, A., and Meller, J. (2005) Combining prediction of secondary structure and solvent accessibility in proteins. *Proteins* **59**, 467–475
42. Grumbling, G., and Strelets, V. (2006) FlyBase: anatomical data, images and queries. *Nucleic Acids Res.* **34**, D484–D488
43. Flicek, P., Aken, B. L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T., Down, T., Dyer, S. C., Eyre, T., Fitzgerald, S., Fernandez-Banet, J., Gräf, S., Haider, S., Hammond, M., Holland, R., Howe, K. L., Howe, K., Johnson, N., Jenkinson, A., Kähäri, A., Keefe, D., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Megy, K., Meidl, P., Overduin, B., Parker, A., Pritchard, B., Prlc, A., Rice, S., Rios, D., Schuster, M., Sealy, I., Slater, G., Smedley, D., Spudich, G., Trevanion, S., Vilella, A. J., Vogel, J., White, S., Wood, M., Birney, E., Cox, T., Curwen, V., Durbin, R., Fernandez-Suarez, X. M., Herrero, J., Hubbard, T. J., Kasprzyk, A., Proctor, G., Smith, J., Ureta-Vidal, A., and Searle, S. (2008) Ensembl 2008. *Nucleic Acids Res.* **36**, D707–D714
44. Dowell, R. D., Jocker, R. M., Day, A., Eddy, S. R., and Stein, L. (2001) The distributed annotation system. *BMC Bioinformatics* **2**, 7
45. Finn, R. D., Stalker, J. W., Jackson, D. K., Kulesha, E., Clements, J., and Pettett, R. (2007) ProServer: a simple, extensible Perl DAS server. *Bioinformatics* **23**, 1568–1570
46. Zhang, Z., Li, J., Zhao, X. Q., Wang, J., Wong, G. K., and Yu, J. (2006) KaKs\_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomic Proteomics Bioinformatics* **4**, 259–263
47. Yan, R., Small, S., Desplan, C., Dearolf, C. R., and Darnell, J. E., Jr. (1996) Identification of a Stat gene that functions in Drosophila development. *Cell* **84**, 421–430
48. Arbouzova, N. I., and Zeidler, M. P. (2006) JAK/STAT signalling in Drosophila: insights into conserved regulatory and cellular functions. *Development* **133**, 2605–2616
49. Woodring, P. J., Hunter, T., and Wang, J. Y. (2003) Regulation of F-actin-dependent processes by the Abl family of tyrosine kinases. *J. Cell Sci.* **116**, 2613–2626
50. Henkemeyer, M. J., Bennett, R. L., Gertler, F. B., and Hoffmann, F. M. (1988) DNA sequence, structure, and tyrosine kinase activity of the Drosophila melanogaster Abelson proto-oncogene homolog. *Mol. Cell. Biol.* **8**, 843–853
51. Tanis, K. Q., Veach, D., Duewel, H. S., Bornmann, W. G., and Koleske, A. J. (2003) Two distinct phosphorylation pathways have additive effects on Abl family kinase activation. *Mol. Cell. Biol.* **23**, 3884–3896
52. Lewis, J. M., and Schwartz, M. A. (1998) Integrins regulate the association and phosphorylation of paxillin by c-Abl. *J. Biol. Chem.* **273**, 14225–14230
53. Salgia, R., Brunkhorst, B., Pisick, E., Li, J. L., Lo, S. H., Chen, L. B., and Griffin, J. D. (1995) Increased tyrosine phosphorylation of focal adhesion proteins in myeloid cell lines expressing p210BCR/ABL. *Oncogene* **11**, 1149–1155
54. Bassermann, F., Jahn, T., Miething, C., Seipel, P., Bai, R. Y., Coutinho, S., Tybulewicz, V. L., Peschel, C., and Duyster, J. (2002) Association of Bcr-Abl with the proto-oncogene Vav is implicated in activation of the Rac-1 pathway. *J. Biol. Chem.* **277**, 12437–12445
55. Hickey, F. B., and Cotter, T. G. (2006) Identification of transcriptional targets associated with the expression of p210 Bcr-Abl. *Eur. J. Haematol.* **76**, 369–383
56. Raitano, A. B., Halpern, J. R., Hambuch, T. M., and Sawyers, C. L. (1995) The Bcr-Abl leukemia oncogene activates Jun kinase and requires Jun for transformation. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 11746–11750
57. Franklin, C. C., Sanchez, V., Wagner, F., Woodgett, J. R., and Kraft, A. S. (1992) Phorbol ester-induced amino-terminal phosphorylation of human JUN but not JUNB regulates transcriptional activation. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 7247–7251



**7.4 Publication:****Phosphorylation of SUMO-1 occurs in vivo and is conserved through evolution.**

Ivan Matic, Boris Macek, **Maximiliane Hilger**, Tobias C. Walther and Matthias Mann

**Journal of proteome research** 7, 4050-4057 (2008)

This paper was published in Journal of proteome research in 2008 and was a joint effort of Ivan Matic, Boris Marek, Tobias C. Walther, Matthias Mann and me. Ivan Matic was the main responsible for the generation and analysis of most of the data. I contributed the *Drosophila* data that demonstrate the evolutionary conservation of the described SUMO/Smt3 phosphorylation.





## Phosphorylation of SUMO-1 Occurs *in Vivo* and Is Conserved through Evolution

Ivan Matic,<sup>†</sup> Boris Macek,<sup>†</sup> Maximiliane Hilger,<sup>†</sup> Tobias C. Walther,<sup>‡</sup> and Matthias Mann<sup>\*,†</sup>

Department of Proteomics and Signal Transduction, Max Planck Institute for Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany, and Organelle Architecture and Dynamics, Max Planck Institute for Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany

Received May 20, 2008

Protein dynamics is regulated by an elaborate interplay between different post-translational modifications. Ubiquitin and ubiquitin-like proteins (Ubls) are small proteins that are covalently conjugated to target proteins with important functional consequences. One such modifier is SUMO, which mainly modifies nuclear proteins. SUMO contains a unique N-terminal arm not present in ubiquitin and other Ubls, which functions in the formation of SUMO polymers. Here, we unambiguously show that serine 2 of the endogenous SUMO-1 N-terminal protrusion is phosphorylated *in vivo* using very high mass accuracy mass spectrometry at both the MS and the MS/MS level and complementary fragmentation techniques. Strikingly, we detected the same phosphorylation in yeast, *Drosophila* and human cells, suggesting an evolutionary conserved function for this modification. The nearly identical human SUMO-2 and SUMO-3 isoforms differ in serine 2; thus, only SUMO-3 could be phosphorylated at this position. Our finding that SUMO can be modified may point to an additional level of complexity through modifying a protein-modifier.

**Keywords:** SUMO-1 • Smt3 • phosphorylation • mass spectrometry • HCD • higher energy dissociation • evolution • conservation

### Introduction

Post-translational modifications (PTMs) are essential in virtually every cellular process and often function by modification-dependent protein interactions.<sup>1,2</sup> Ubiquitin and ubiquitin like proteins (Ubls) covalently modify target proteins by attaching their C-termini to specific substrate lysines. They are related by a common tertiary structure, called the ubiquitin superfold.<sup>3</sup> SUMO modification regulates the function of target proteins by modulating protein interactions, intracellular transport, half-life and activity.<sup>4</sup> There are three distinct human SUMO isoforms (SUMO-1, -2, and -3), whereas a single SUMO, called Smt3, is present in yeast and *Drosophila*. SUMO-2 and SUMO-3 are almost identical to each other and 50% identical to SUMO-1.<sup>5</sup> In contrast to ubiquitin and other Ubls, SUMO proteins contain an unstructured and flexible N-terminal segment that protrudes from the protein core.<sup>6</sup> No functional differences have yet been demonstrated between SUMO-2 and SUMO-3, which differ in just three amino acids, all of them localized on the N-terminal protrusion.

Sumoylation is mediated by an enzymatic mechanism analogous to ubiquitylation involving three enzymes, namely, E1, E2 and E3. SUMO-specific proteases catalyze SUMO precursor processing and deconjugation of SUMO from modified substrates.<sup>7</sup>

Members of the ubiquitin and Ubl family are not only PTMs, but also proteins that can be targeted by PTMs, raising the concept of “modification of a modification”. All seven internal lysines of ubiquitin can be ubiquitylated,<sup>8</sup> of which at least three have different functional consequences. K11 of SUMO-2 and SUMO-3 is modified *in vivo* by SUMO-1 and SUMO-2/3.<sup>9,10</sup> Recently, it was shown that K11, K32 and K41 of SUMO-2 can be modified by ubiquitin.<sup>11,12</sup> This raises the question if SUMO can also be targeted by PTMs not belonging to the ubiquitin/Ubl family.

Here, we use a targeted, mass spectrometric approach to characterize the primary structure of SUMO-1. We report that the SUMO-1 N-terminal arm is phosphorylated and investigate the evolutionary conservation of this PTM.

### Materials and Methods

**Cell Culture and Protein Extraction.** HeLa-S3 suspension cells were grown in RPMI medium containing 10% FBS and 100 units/mL penicillin/streptomycin (Invitrogen), harvested by centrifugation for 5 min at 400g, washed with cold PBS and lysed under nondenaturing conditions with lysis buffer (50 mM HEPES, pH 7.5, 10% glycerol, 150 mM NaCl, 1% Triton X-100, 1 mM EDTA, 1 mM EFTA, 5 mM  $\beta$ -glycerophosphate, 1 mM

\* To whom correspondence should be addressed. Prof. Dr. Matthias Mann, Department of Proteomics and Signal Transduction, Max Planck Institute for Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany. Tel, +49 89-8578-2557; fax, +49 89-8578-2219; e-mail, mmann@biochem.mpg.de.

<sup>†</sup> Department of Proteomics and Signal Transduction, Max Planck Institute for Biochemistry.

<sup>‡</sup> Organelle Architecture and Dynamics, Max Planck Institute for Biochemistry.

**Table 1.** Detected Phosphorylation Sites from *S. cerevisiae* and *Drosophila melanogaster* Smt3 and Human SUMO-1

protein	phosphopeptide sequence	charge	<i>m/z</i>	mass deviation [ppm]	Mascot score	intensity	acquisition method
SUMO-1	Ac-pSDQEAKPSTEDLGDK	2	871.3618	-0.06	36.93	$3.30 \times 10^5$	Full range Top5
	Ac-SDQEAKPSTEDLGDK	2	831.3787	-1.05	16.39	$1.99 \times 10^5$	Full range Top5
	Ac-pSDQEAKPSTEDLGDKK	2	935.4093	-0.19	35.93	$1.04 \times 10^6$	Full range Top5
	Ac-SDQEAKPSTEDLGDKK	2	895.4262	-0.27	45.45	$4.21 \times 10^5$	Full range Top5
	Ac-pSDQEAKPSTEDLGDKK	3	623.9420	0.05	46.94	$6.37 \times 10^5$	Full range Top5
	Ac-SDQEAKPSTEDLGDKK	3	597.2865	-0.07	48.02	$2.40 \times 10^6$	Full range Top5
	Ac-pSDQEAKPSTEDLGDKK	2	935.4093	0.05	29.78	$2.19 \times 10^6$	SIM CID/HCD
(from His-SUMO-2)	Ac-pSDQEAKPSTEDLGDKKEGEYIK	3	863.7250	0.43	24.76	$1.46 \times 10^6$	Full range Top5
	Ac-SDQEAKPSTEDLGDKKEGEYIK	3	837.0695	0.31	41.87	$1.44 \times 10^6$	Full range Top5
Yeast Smt3	pSDSEVNQEAK	2	593.7348	-0.47	38.73	$8.85 \times 10^4$	Full range Top5
	Ac-pSDSEVNQEAKPEVKPEVK	3	712.3350	-0.5	34.52	$2.13 \times 10^5$	Full range Top5
	Ac-pSDSEVNQEAKPEVKPEVK	2	1067.9988	-0.52	59.79	$2.14 \times 10^5$	Full range Top5
<i>Drosophila</i> Smt3	Ac-pSDEKKGGGETEHINLK	2	903.9171	-0.77	28.19	$7.88 \times 10^5$	
	Ac-pSDEKKGGGETEHINLK	3	602.9472	-1.76	21.36	$2.71 \times 10^6$	

NaVO<sub>4</sub>, 5 mM NaF, complete protease inhibitor mixture (Roche)). The cell lysate was centrifuged for 1 h at 60 000g at 4 °C and the supernatant was collected.

The YAL6B yeast strain<sup>13</sup> was grown in YPD liquid medium to log-phase (OD<sub>600</sub> 0.8), harvested by centrifugation for 5 min at 4000g, washed with cold water and frozen in liquid nitrogen. The cells were resuspended in lysis buffer (50 mM HEPES, pH 6.8, 150 mM KOAc, 2 mM MgOAc, 5 mM β-glycerophosphate, 1 mM NaVO<sub>4</sub>, 5 mM NaF, complete protease inhibitor mixture (Roche)) and lysed under nondenaturing conditions using glass beads. The yeast lysate was centrifuged at 10 000g for 10 min at 4 °C and the supernatant was collected.

**Electrophoresis, Coomassie Staining, and Protein Digestion.** Proteins were separated by one-dimensional SDS-PAGE using Novex 4–12 % precast gels and MES SDS running buffer (Invitrogen) according to the manufacturer's instructions. The gel was stained with Coomassie blue using Colloidal Blue Staining Kit (Invitrogen). The protein bands in the approximate molecular weight range 5–20 kDa (Supplementary Figure 1) were excised from the gel and subjected to in-gel digestion with trypsin (Promega), essentially as described in Shevchenko et al.<sup>14</sup> The resulting peptide mixture from HeLa cells was cleaned, desalted, concentrated and enriched using self-made StageT-ips.<sup>15</sup> Phosphopeptides from the yeast mixture were enriched by titanium dioxide (TiO<sub>2</sub>) chromatography, as described previously.<sup>16</sup>

**Mass Spectrometry.** LC-MS/MS analysis was performed by a linear ion trap-orbitrap hybrid mass spectrometer (LTQ-Orbitrap, Thermo Fisher Scientific) equipped with a nano-electrospray ion source (Proxeon Biosystems) and coupled with an Agilent 1200 nano-HPLC system (Agilent Technologies). The digested peptides were separated on an in-house packed 75 μm reversed-phase C18 column.<sup>17</sup> Survey scan acquisition was performed in the orbitrap (*R* = 60 000 ion accumulation to a target value of 1 000 000). The five most intense ions were fragmented and acquired in the LTQ using collisionally induced dissociation (CID) ("Full range Top5" method).<sup>18</sup> In the second measurement, the phosphorylated and N-terminally acetylated peptide of SUMO-1 was analyzed in the selected-ion monitoring (SIM) mode with a scanning range of *m/z* 932.4–938.4 (*R* = 60 000 ion accumulation to a target value of 100 000). The two most intense ions were isolated and fragmented in the LTQ using CID and in the C-trap by higher energy CID (HCD). The acquisition of the resulting MS/MS spectra was performed in the orbitrap analyzer (*R* = 30 000, ion accumulation to a target value of 75 000) ("SIM CID/HCD" method). Multistage activa-

tion was enable on all CID MS/MS spectra and the lock mass option was enabled in all measurements to further increase mass accuracy, as described previously.<sup>18</sup>

**Data Processing and Analysis.** Raw data files were analyzed with the quantitative proteomics software MaxQuant (version 1.0.9.9),<sup>19</sup> which in this case was used for peak list generation, identification and filtering, and Mascot search engine (version 2.2.0; Mascot Science). MS/MS spectra were searched against the human International Protein Index (IPI) database (Version 3.37),<sup>20</sup> the *Drosophila* proteome database (FlyBase 5.4) or a yeast *Saccharomyces cerevisiae* ORF protein database. All three databases contained the forward and reversed sequences and were supplemented with the most commonly observed contaminants. For database search, enzyme specificity was set to trypsin allowing up to three missed cleavages. Carbamidomethylation (C) was set as fixed modification, while Oxidation (M), N-Acetyl (Protein) and phospho (STY) were searched as variable modifications. Initial maximum mass deviation was set to 7 ppm and the fragment ion mass tolerance to 0.5 Da.

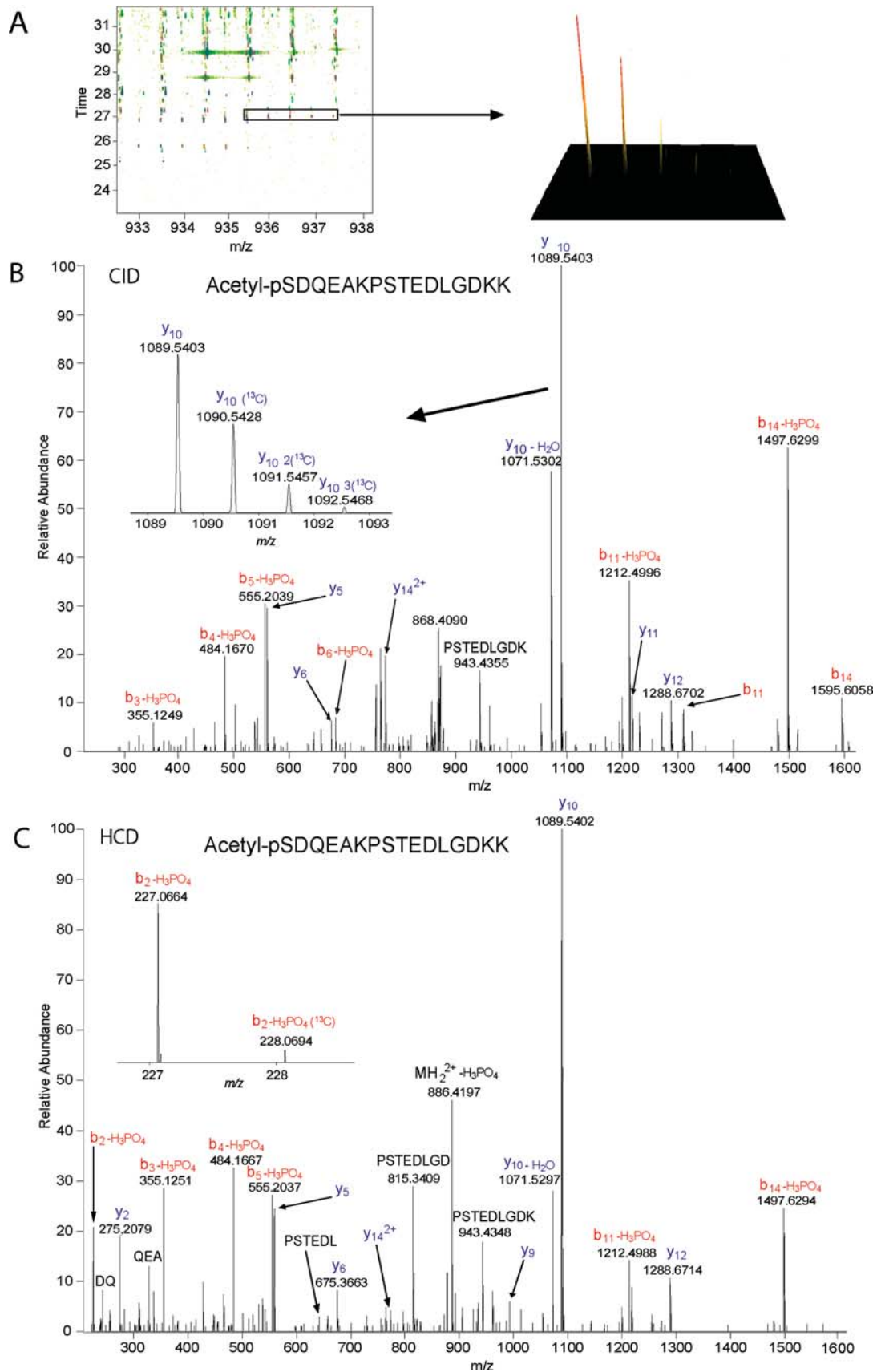
The Viewer tool of MaxQuant was used for a 2D and 3D visualization of LC-MS runs. The signal intensities are color-coded, with white and black representing the lowest intensities and green and red the highest intensities, respectively, in the 2D and 3D view.

Multiple sequence alignment was performed using the ClustalX program.<sup>21</sup>

## Results and Discussion

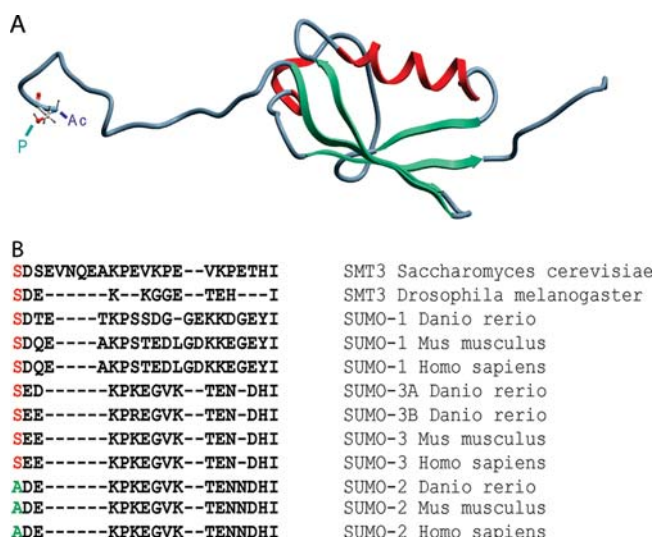
**Phosphorylation of Endogenous Human SUMO-1 *in Vivo*.** Recent mass spectrometric identification of sumoylated and ubiquitinated SUMO peptides<sup>10–12</sup> has induced us to investigate the possibility that SUMO may be targeted by an additional major regulatory PTM, phosphorylation.

SUMO is usually prepared for MS analysis from stable cell lines expressing a tagged form of the protein, such as the His<sub>6</sub>-SUMO-2 HeLa cell line.<sup>10</sup> These tagged forms have the advantage of good purification yield, but also the drawback that tagged SUMO expressed at nonendogenous levels may behave differently from its endogenous counterpart. To avoid this and still obtain sufficient material for analysis, we decided to lyse cells under nondenaturing conditions. Under these conditions, cellular SUMO is rapidly deconjugated from target proteins due to the strong activity of the desumoylation machinery.<sup>22</sup> Therefore, previously conjugated SUMO is enriched in the region of the gel whose molecular weight corresponds to that



**Figure 1.** Serine 2 of endogenous human SUMO-1 is phosphorylated *in vivo*. HeLa total cell lysate was size-fractionated by SDS-PAGE. (A) Enlarged 2D view of the MS data from the in gel-digested band is shown and the precursor ion isotopic distribution is represented three-dimensionally. Precursor ion was measured in the orbitrap mass spectrometer ( $m/z$  935.4093 (2+); mass deviation, 0.05 ppm), and the peptide was fragmented by CID in the LTQ (B) and by HCD in the C-trap (C). In both cases, the fragment ions were analyzed in the orbitrap analyzer. Insets, magnifications of the most intense fragment ion (B) and of the low mass region (C).

## Phosphorylation of SUMO-1



**Figure 2.** Position of the N-acetylated and phosphorylated residue and its evolutionary conservation. (A) SUMO-1 structure (1A5R)<sup>6</sup> was visualized using ICM Browser Pro (Molsoft, Redmond, WA; version 3.5-1m). (B) Sequence alignment of the N-terminal arm of SUMO family.

of free SUMO. We lysed HeLa cells, and after separation by SDS-PAGE, excised and in-gel digested the band in the approximate range 5–20 kDa (Supplementary Figure 1). Dephosphorylation and protein degradation were prevented by adding phosphatase and protease inhibitors. The resulting peptides were analyzed by LC-MS/MS on a hybrid linear ion trap–orbitrap mass spectrometer (LTQ-Orbitrap). We employed the standard method used in our laboratory, which consists of a high resolution scan in the orbitrap concurrent with up to five low resolution MS/MS scans in the linear ion trap (“Full range Top5” method, see Materials and Methods).

We found that the N-terminal peptide always lacked the initial methionine and that it was always N-acetylated in HeLa cells. This has also very recently been observed by others.<sup>23</sup> Unexpectedly, however, this peptide occurred in two forms; nonphosphorylated or phosphorylated at the N-terminal serine (Ser2). The charge state of the nonphosphopeptide was predominantly triply charged, whereas the phosphopeptide was predominantly doubly charged, presumably because of the negatively charged phosphogroup (Table 1). To obtain a rough measure of the stoichiometry of phosphorylation, we integrated the ion current of the two peptide forms. Often, phosphopeptides are assumed to ionize less well than nonphosphopeptides, although this is not necessarily true.<sup>24</sup> Even assuming equal ionization efficiency, our data is consistent with 35–40 % of Ser2 being phosphorylated. Thus, our data indicates that a large proportion of SUMO is phosphorylated *in vivo*.

Because SUMO is the object of intense biological interest and because no phosphorylation site has yet been described, we next wanted to confirm the identity of the phosphopeptide beyond doubt. The identified phosphopeptides of SUMO-1 (Table 1, Supplementary Figures 2–4) had very low mass deviations from the corresponding calculated values (low ppm range or below). However, their MS/MS spectra had low resolution and low mass accuracy (for a direct comparison of resolutions between orbitrap and LTQ see Matic et al.<sup>10</sup>). Unambiguous identification of this phosphopeptide was achieved using the capability of the LTQ-Orbitrap mass spectrometer to obtain very high mass accuracy at both MS and

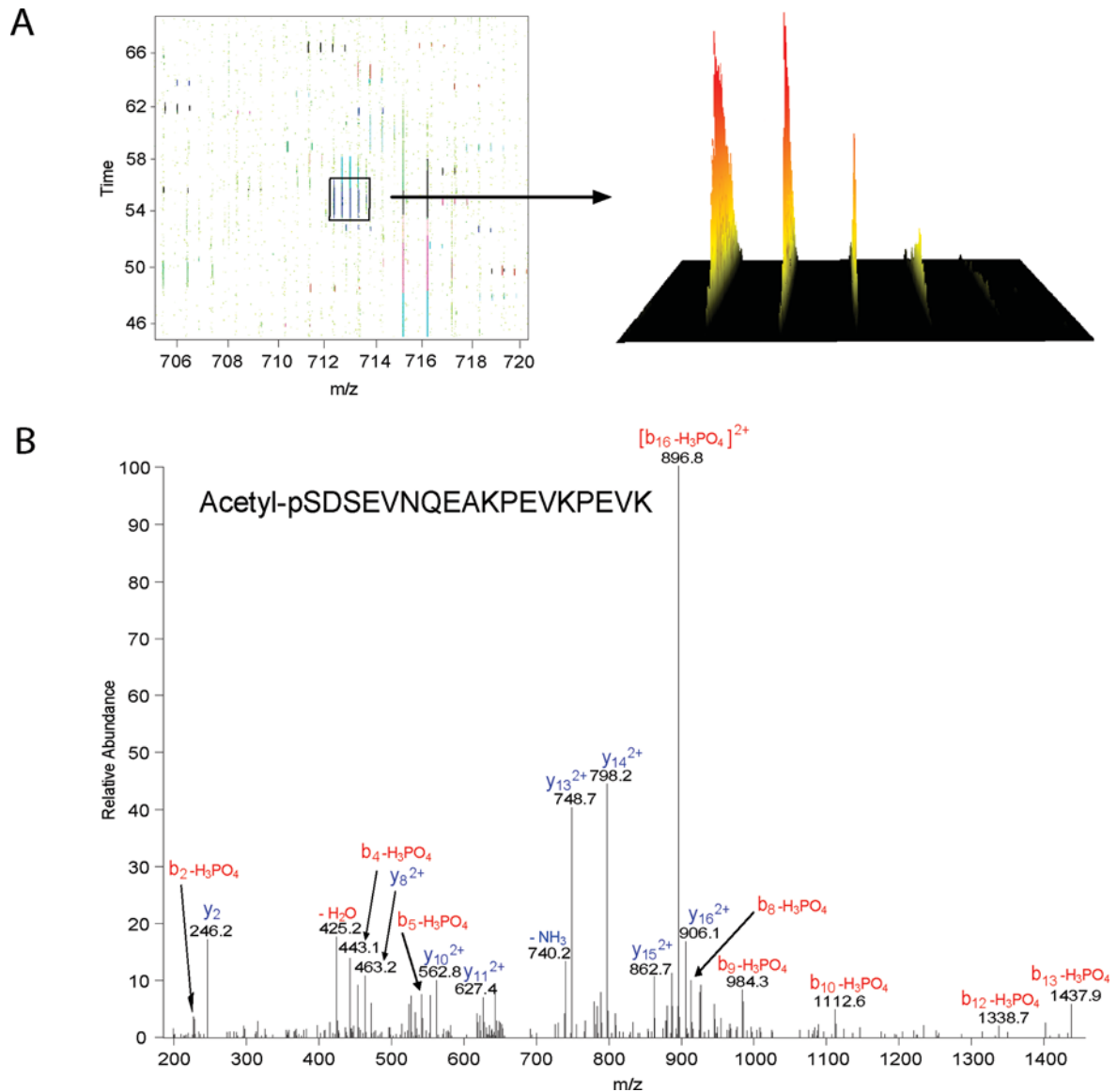
MS/MS level.<sup>18</sup> On the basis of the information from the standard, “Full range Top5” method, we devised a targeted peptide-specific method, similarly to the approach used previously for the identification of the SUMO-1/SUMO-2 branched peptide.<sup>10</sup> A narrow mass range of 6 Th around the  $m/z$  value of the precursor ion of interest was monitored in the survey scan (“SIM CID/HCD” method). The two most abundant peaks were isolated and fragmented both in the LTQ by CID and in the C-trap by higher energy dissociation (HCD) (Figure 1). In the HCD method, the C-trap, which is normally used for storage of ions prior to injection in the orbitrap, is used to produce highly efficient quadrupole-like fragmentation.<sup>25</sup> The fragment ions were acquired in the orbitrap producing high-resolution and high-accuracy MS/MS spectra. Note the similarity of MS/MS profiles in Figure 1, panels B and C. These high resolution spectra verified the identity of the b and y ions, crucial for localizing the modification, where b-ions were observed with loss of phosphoric acid ( $H_3PO_4$ ). In particular, the  $b_3$  to  $b_6$  ion series in Figure 1A and the  $b_2$  ion in Figure 1C, all with loss of phosphoric acid, as well as the nonmodified  $y_{10}$  and  $y_{12}$  ions localize the phosphorylation to the very N-terminus of the peptide. The fragmentation pathway obtained by HCD fragmentation in the C-trap also leads to a pronounced internal fragmentation (Figure 1C), giving rise to the fragment ions in the low mass part of the spectrum that were absent in CID fragmentation (Figure 1B). These internal fragments were not phosphorylated. Low resolution fragmentation spectra of the peptide with and without missed cleavage further confirmed the assignments (Table 1 and Supplementary Figures 2-4). In addition, we also found Ser2 phosphorylation of SUMO-1 in our previous global study of phosphorylation dynamics during EGF signaling in HeLa cells.<sup>16</sup> Thus, by multiple lines of mass spectrometric evidence, we unambiguously confirmed that the endogenous human SUMO-1 is phosphorylated in HeLa cells.

**Phosphorylation of Endogenous Human SUMO-1 in His<sub>6</sub>-SUMO-2 Conjugates.** We next asked if SUMO-1 could be phosphorylated when conjugated to target proteins. For this purpose, we interrogated the raw spectra from the His<sub>6</sub>-SUMO-2 fraction, in which we had previously detected mixed SUMO-branched peptides.<sup>10</sup> That fraction did not contain any free SUMO-1 as determined by Western blotting. We reanalyzed the data with phosphorylation as a variable modification and detected the phosphorylation on SUMO-1 (Supplementary Figure 8). This shows that SUMO-1 conjugated to target proteins (likely SUMO-2 in this case) can be phosphorylated.

Identically to the in-gel digested fraction, Ser2 was always N-terminally acetylated. The nonphosphorylated and phosphorylated forms of the peptide have approximately equal integrated ion current (Table 1), which confirms that a large fraction of endogenous SUMO-1 is phosphorylated.

**Evolutionary Conservation of SUMO Phosphorylation.** The identified peptides indicate that the N-terminal tail of SUMO-1 undergoes three co- or post-translational modifications: removal of the starting methionine residue, N-terminal acetylation and phosphorylation of the first amino acid (Ser2) of the mature SUMO-1. Since Ser2 is in a highly flexible and accessible part of SUMO-1, this phosphoserine is a good candidate for the regulation of protein binding (Figure 2A).

To investigate evolutionary conservation of the Ser2, we aligned the amino acid sequences of SUMOs of five different eukaryotic species (Figure 2B). The alignment shows that Ser2 of human SUMO-1 is conserved in the single SUMO forms present in *S. cerevisiae* and *D. melanogaster* (both



**Figure 3.** Serine 2 of endogenous yeast Smt3 is phosphorylated *in vivo*. Yeast total cell lysate was separated by SDS-PAGE, the band was digested in gel, and the phosphopeptides were enriched by  $TiO_2$  chromatography. The precursor ion, shown in a 2D and 3D representation (A), was analyzed in the orbitrap mass spectrometer ( $m/z$  712.3350 (3+); mass deviation,  $-0.58$  ppm), and subsequently fragmented by CID in the LTQ. The resulting fragmentation spectra were acquired in the LTQ analyzer (B).

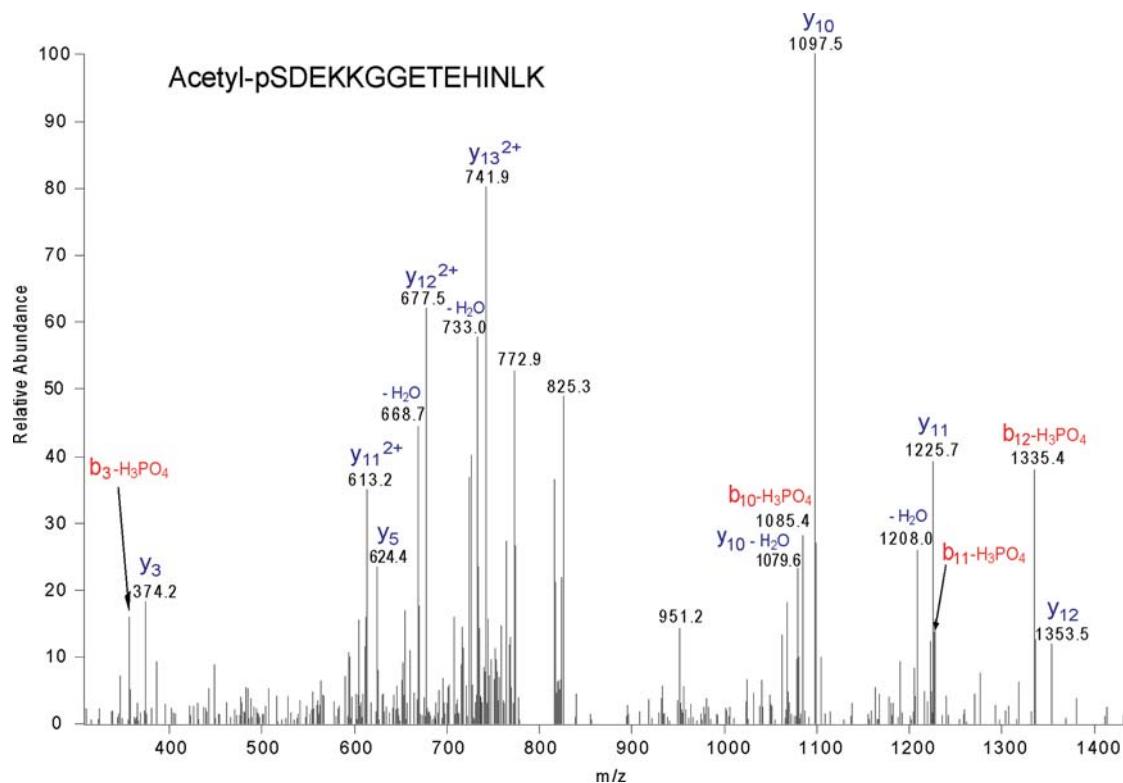
called Smt3). In vertebrates, which have three SUMO paralogs, position 2 is serine in SUMO-1 and SUMO-3, whereas it is alanine in SUMO-2.

It is interesting to note that the amino acid at position 2 is one of only three amino acids that are different between human SUMO-2 and SUMO-3. SUMO-2/3 are so far functionally indistinguishable, but their sequences are quite different from SUMO-1. Thus, SUMO-1 and SUMO-3, but not SUMO-2, can in principle be phosphorylated at their N-terminus.

Despite different strategies, including purification of nuclei, in-solution and in-gel digestion followed by  $TiO_2$  enrichment and the targeted mass spectrometric approach, in this study, we were not able to detect the SUMO-3 phosphoserine-containing peptide. Digestion with trypsin could produce a peptide which is too small to be detected by MS (Figure 3B). However, the phosphopeptide was not identified even after digestion with endoprotease Asp-N, which should produce a longer and more easily detectable peptide. Our failure may be

due to the low stoichiometry of the modification and technical limitations of our approach and does not prove that SUMO-3 cannot be phosphorylated.

**Phosphorylation of Endogenous *S. cerevisiae* and *D. melanogaster* Smt3 *in Vivo*.** As mentioned above, Ser2 is evolutionary conserved in eukaryotic species. To test whether N-terminal acetylation and phosphorylation of Ser2 are also conserved, we enriched yeast Smt3 as described for mammalian cells above and digested the proteins in gel with trypsin. Phosphopeptides were enriched by  $TiO_2$  chromatography and analyzed with the LTQ-Orbitrap using the standard acquisition method. The identified peptide of interest lacks the first methionine and Ser2 is N-acetylated and phosphorylated (Figure 3, Table 1 and Supplementary Figures 5 and 6), just as in human SUMO-1. The presence of the  $b_2$  ion with loss of phosphoric acid and of the unmodified double charged  $y_{16}$  ion (Figure 3B) excludes the possibility that the phosphorylation is localized on the nonconserved Ser4. As in the human case, the N-terminal



**Figure 4.** Serine 2 of endogenous *Drosophila* Smt3 is phosphorylated *in vivo*. Proteins were separated by SDS-PAGE and digested in gel. Phosphopeptides were enriched using TiO<sub>2</sub> beads. The precursor ion was analyzed in the orbitrap mass spectrometer ( $m/z$  903.9171 (2+); mass deviation,  $-0.77$  ppm), and subsequently fragmented and acquired in the LTO.

peptide of yeast Smt3 was always lacking the initial methionine, but contrary to what we observed in HeLa cells, we also detected a peptide in which Ser2 is phosphorylated, but not N-acetylated (Table 1, Supplementary Figure 5).

We next interrogated data from an ongoing *Drosophila* phosphoproteome project (Hilger et al., manuscript in preparation). Proteins extracted from Schneider SL2 cells were in-gel digested and phosphopeptides were enriched by TiO<sub>2</sub> chromatography. As shown in Figure 4, Table 1 and Supplementary Figure 7, the N-terminal peptide is acetylated and phosphorylated in this species as well.

## Conclusions

The ubiquitin and ubiquitin like proteins are unique in that they are both post-translational modifiers but can also themselves be modified. Thus, they have a “modifying” mode, but also a “modified” mode. This latter property has only been explored in depth for ubiquitin and SUMO chain formation.<sup>3,9,10</sup> Here, we have shown that a member of the Ubl family, SUMO-1, is the target of phosphorylation. We used state-of-the-art high-resolution mass spectrometry at the MS and MS/MS levels to establish the presence of this modification beyond doubt. Two factors argue that this modification may be of functional significance rather than ‘background phosphorylation’. First, the stoichiometry of the modification is high—the signal of the phosphopeptide was almost as large as that of the unphosphorylated peptide in HeLa cells. It appears unlikely that a nonfunctional site, especially a highly conserved site potentially important for protein–protein interaction, would be phosphorylated at such high levels. Secondly, we found this modification over an evolutionary distance of at least one billion years

and in organisms as different as a unicellular eukaryote, an insect and human.

Interestingly, a SUMO polymerization site resides within the amino-terminal arm.<sup>10</sup> Here, we have established that two PTMs not belonging to the ubl/ubiquitin family, namely, phosphorylation and N-terminal acetylation, contribute to the SUMO “modified” mode. The unique feature of the SUMO proteins compared to other UbIs is the long and very flexible N-terminal protrusion. It has been suggested that this unstructured tail of the protein might provide an additional surface for protein–protein interaction.<sup>6</sup> In this perspective, phosphorylation of this extreme N-terminal serine could modulate the ability of proteins to interact with SUMO.

Interestingly, similar amino-terminal tail domains are present in histones and are essential for chromatin dynamics. Modifications of histone tails regulate gene expression by mediating the recruitment of proteins to chromatin.<sup>26</sup> The N-terminal arms of both SUMOs and histones protrude from the protein core, contain many charged amino acids and different PTMs sites. Analogously to SUMO, the first serine of histones H2A and H4 are N-acetylated and phosphorylated.<sup>27</sup> Thus, one can speculate that the N-terminal extensions of SUMO may recruit interaction partners to sumoylated proteins or SUMO itself in a similar way that histones amino terminal tails recruit proteins to chromatin.

SUMO-2 and SUMO-3 have yet to be functionally differentiated. All three residues different between these two paralogs are localized on the N-terminal protrusion and serine is conserved from SUMO-1 to SUMO-3, but not SUMO-2. The slight difference in the amino acid sequences between SUMO-2 and SUMO-3 tails and the hypothetical SUMO-3 specific phosphorylation could lead to a different recruitment of

interacting proteins. If regulated Ser2 phosphorylation on SUMO-3 can be detected, this would be the first functional difference between SUMO-2 and SUMO-3. Thus, the functional meaning of the phosphorylation on SUMO Ser2 and the investigation of interaction partners merits further studies.

**Abbreviations:** PTM, post-translational modification; UbLs, ubiquitin-like proteins; SUMO, small ubiquitin-like modifier; E1, SUMO-activating enzyme, E2, SUMO-protein carrier protein; E3, SUMO ligase; LC-MS, liquid chromatography-mass spectrometry; MS/MS, tandem mass spectrometry; SIM, selected ion monitoring; CID, collision-induced dissociation; HCD, higher-energy C-trap dissociation; LTQ, linear quadrupole ion trap; ppm, parts per million.

**Acknowledgment.** We thank Florian Fröhlich and Michael Rehman for advice, helpful discussion and help with multiple sequence analysis. We thank other members of our department for fruitful discussions. This work was supported by The European Community (RUBICON, VI Framework).

**Supporting Information Available:** Supplementary Figure 1, gel region around free SUMO-1 or Smt3 (approximately 5–20 kDa) was digested in gel and analyzed by LC-MS/MS; Supplementary Figure 2, MS/MS fragmentation spectrum of the human SUMO-1 peptide Acetyl-pSDQEAKPSTEDLGDK. Precursor ion mass was measured in the orbitrap analyzer ( $m/z$  871.3618 (2+); mass deviation,  $-0.06$  ppm) and the peptide was fragmented by CID and acquired in the LTQ mass spectrometer; Supplementary Figure 3, MS/MS fragmentation spectrum of the human SUMO-1 peptide Acetyl-pSDQEAKPSTEDLGDKK. Precursor ion mass was measured in the orbitrap analyzer ( $m/z$  935.4093 (2+); mass deviation,  $-0.19$  ppm) and the peptide was fragmented by CID and acquired in the LTQ mass spectrometer; Supplementary Figure 4, MS/MS fragmentation spectrum of the human SUMO-1 peptide Acetyl-pSDQEAKPSTEDLGDKK. Precursor ion mass was measured in the orbitrap analyzer ( $m/z$  623.9420 (3+); mass deviation,  $0.05$  ppm) and the peptide was fragmented by CID and acquired in the LTQ mass spectrometer; Supplementary Figure 5, MS/MS fragmentation spectrum of the yeast Smt3 peptide pSDSEVNQEAK. Precursor ion mass was measured in the orbitrap analyzer ( $m/z$  593.7348 (2+); mass deviation,  $-0.47$  ppm) and the peptide was fragmented by CID and acquired in the LTQ mass spectrometer; Supplementary Figure 6, MS/MS fragmentation spectrum of the yeast Smt3 peptide Acetyl-pSDSEVNQEAKPEVKPEVK. Precursor ion mass was measured in the orbitrap analyzer ( $m/z$  1067.9988 (2+); mass deviation,  $-0.52$  ppm) and the peptide was fragmented by CID and acquired in the LTQ mass spectrometer; Supplementary Figure 7, MS/MS fragmentation spectrum of the *Drosophila* Smt3 peptide Acetyl-pSDEKKGGETEHINLK. Precursor ion mass was measured in the orbitrap analyzer ( $m/z$  602.9472 (3+); mass deviation,  $-1.76$  ppm) and the MS/MS spectra were acquired in the LTQ mass spectrometer; Supplementary Figure 8, His<sub>6</sub>-SUMO-2 MS/MS data from Matic et al.<sup>10</sup> were analyzed with MaxQuant. MS/MS fragmentation spectrum of the human SUMO-1 peptide Acetyl-pSDQEAKPSTEDLGDKKEGEYIK. Precursor ion mass was measured in the orbitrap analyzer ( $m/z$  863.7250 (2+); mass deviation,  $0.44$  ppm) and the peptide was fragmented by CID and acquired in the LTQ mass spectrometer. Note that the multistage activation was not enabled and that the main

fragmentation ion is the neutral loss of the precursor ion. The enlarged left and right regions of the MS/MS spectrum show the presence of other fragment ions. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Krishna, R. G.; Wold, F. Post-translational modification of proteins. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1993**, *67*, 265–98.
- (2) Seet, B. T.; Dikic, I.; Zhou, M. M.; Pawson, T. Reading protein modifications with interaction domains. *Nat. Rev. Mol. Cell. Biol.* **2006**, *7* (7), 473–83.
- (3) Welchman, R. L.; Gordon, C.; Mayer, R. J. Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nat. Rev. Mol. Cell. Biol.* **2005**, *6* (8), 599–609.
- (4) Hay, R. T. SUMO: a history of modification. *Mol. Cell* **2005**, *18* (1), 1–12.
- (5) Johnson, E. S. Protein modification by SUMO. *Annu. Rev. Biochem.* **2004**, *73*, 355–82.
- (6) Bayer, P.; Arndt, A.; Metzger, S.; Mahajan, R.; Melchior, F.; Jaenicke, R.; Becker, J. Structure determination of the small ubiquitin-related modifier SUMO-1. *J. Mol. Biol.* **1998**, *280* (2), 275–86.
- (7) Bossis, G.; Melchior, F. SUMO: regulating the regulator. *Cell Div.* **2006**, *1*, 13.
- (8) Pickart, C. M.; Fushman, D. Polyubiquitin chains: polymeric protein signals. *Curr. Opin. Chem. Biol.* **2004**, *8* (6), 610–6.
- (9) Tatham, M. H.; Jaffray, E.; Vaughan, O. A.; Desterro, J. M.; Botting, C. H.; Naismith, J. H.; Hay, R. T. Polymeric chains of SUMO-2 and SUMO-3 are conjugated to protein substrates by SAE1/SAE2 and Ubc9. *J. Biol. Chem.* **2001**, *276* (38), 35368–74.
- (10) Matic, I.; van Hagen, M.; Schimmel, J.; Macek, B.; Ogg, S. C.; Tatham, M. H.; Hay, R. T.; Lamond, A. I.; Mann, M.; Vertegaal, A. C. In vivo identification of human small ubiquitin-like modifier polymerization sites by high accuracy mass spectrometry and an in vitro to in vivo strategy. *Mol. Cell. Proteomics* **2008**, *7* (1), 132–44.
- (11) Tatham, M. H.; Geoffroy, M.-C.; Shen, L.; Plechanovova, A.; Hattersley, N.; Jaffray, E. G.; Palvimo, J. J.; Hay, R. T. RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. *Nat. Cell Biol.* **2008**, *10* (5), 538–46.
- (12) Schimmel, J.; Larsen, K. M.; Matic, I.; van Hagen, M.; Cox, J.; Mann, M.; Andersen, J. S.; Vertegaal, A. C. The ubiquitin-proteasome system is a key component of the SUMO-2/3 cycle. *Mol. Cell. Proteomics* **2008**,
- (13) Gruhler, A.; Olsen, J. V.; Mohammed, S.; Mortensen, P.; Faergeman, N. J.; Mann, M.; Jensen, O. N. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol. Cell. Proteomics* **2005**, *4* (3), 310–27.
- (14) Shevchenko, A.; Tomas, H.; Havlis, J.; Olsen, J. V.; Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **2006**, *1* (6), 2856–60.
- (15) Rappsilber, J.; Mann, M.; Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2007**, *2* (8), 1896–906.
- (16) Olsen, J. V.; Blagoev, B.; Gnäd, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **2006**, *127* (3), 635–48.
- (17) Olsen, J. V.; Ong, S. E.; Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* **2004**, *3* (6), 608–14.
- (18) Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **2005**, *4* (12), 2010–21.
- (19) Cox, J.; Mann, M. Is proteomics the new genomics. *Cell* **2007**, *130* (3), 395–8.
- (20) Kersey, P. J.; Duarte, J.; Williams, A.; Karavidopoulou, Y.; Birney, E.; Apweiler, R. The International Protein Index: an integrated database for proteomics experiments. *Proteomics* **2004**, *4* (7), 1985–8.
- (21) Thompson, J. D.; Gibson, T. J.; Plewniak, F.; Jeanmougin, F.; Higgins, D. G. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **1997**, *25* (24), 4876–82.
- (22) Mahajan, R.; Delphin, C.; Guan, T.; Gerace, L.; Melchior, F. A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell* **1997**, *88* (1), 97–107.
- (23) Lallemand-Breitenbach, V.; Jeanne, M.; Benhenda, S.; Nasr, R.; Lei, M.; Peres, L.; Zhou, J.; Zhu, J.; Raught, B.; de The, H. Arsenic

## Phosphorylation of SUMO-1

- degrades PML or PML-RAR[alpha] through a SUMO-triggered RNF4/ubiquitin-mediated pathway. *Nat. Cell Biol.* **2008**, *10* (5), 547–55.
- (24) Steen, H.; Jebanathirajah, J. A.; Rush, J.; Morrice, N.; Kirschner, M. W. Phosphorylation analysis by mass spectrometry: myths, facts, and the consequences for qualitative and quantitative measurements. *Mol. Cell. Proteomics* **2006**, *5* (1), 172–81.
- (25) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* **2007**, *4* (9), 709–12.

## research articles

- (26) Iizuka, M.; Smith, M. M. Functional consequences of histone modifications. *Curr. Opin. Genet. Dev.* **2003**, *13* (2), 154–60.
- (27) Barber, C. M.; Turner, F. B.; Wang, Y.; Hagstrom, K.; Taverna, S. D.; Mollah, S.; Ueberheide, B.; Meyer, B. J.; Hunt, D. F.; Cheung, P.; Allis, C. D. The enhancement of histone H4 and H2A serine 1 phosphorylation during mitosis and S-phase is evolutionarily conserved. *Chromosoma* **2004**, *112* (7), 360–71.

PR800368M



## **7.5 Publication:**

### **Triple SILAC to determine stimulus specific interactions in the Wnt pathway**

**Maximiliane Hilger** and Matthias Mann

In revision at **Mol Cell Proteomics**

This manuscript is currently under revision at Molecular Cellular Proteomics and the authors are Matthias Mann and me. This work results from my final thesis project.



# Triple SILAC to determine stimulus specific interactions in the Wnt pathway

Maximiliane Hilger§ and Matthias Mann§

Many important regulatory functions are performed by dynamic multi-protein complexes that adapt their composition and activity in response to different stimuli. Here we employ quantitative affinity purification coupled with mass spectrometry to efficiently separate background from specific interactors but add an additional quantitative dimension to explicitly characterize stimulus dependent interactions. This is accomplished by stable isotope labeling by amino acids in cell culture in a triple-labeling format, in which pull-downs with bait, with bait and stimulus and without bait are quantified against each other. As baits we use full-length proteins fused to the green fluorescent protein and expressed under endogenous control. We applied this technology to Wnt signaling, which is important in development, tissue homeostasis and in cancer, and investigated interactions of the key components APC, Axin-1, DVL2 and CtBP2 with differential pathway activation. Our screens identify many known Wnt complex components and link novel candidates to Wnt signaling, including FAM83B and Girdin, which we found as interactors to multiple Wnt pathway players. APC and Axin-1 share constitutive interactors but also several bin-

ders whose binding was modulated by stimulation with the ligand Wnt3a. The core destruction complex itself, which regulates beta-Catenin stability as the key step in canonical Wnt signaling, remained essentially unchanged. Girdin binds to DVL2 independent of stimulation but to Axin-1 and APC in a stimulus dependent manner. Triple SILAC interaction analysis is a powerful tool for AP-MS based interaction studies that determines specific interactors of a protein of interest and at the same time resolves stimulus specific interaction dynamics.

The study of protein-protein interactions is of pivotal importance because most biological functions are mediated by protein complexes. In contrast to most other techniques, affinity purification combined with mass spectrometry (AP-MS)<sup>1</sup> is unbiased in that it does not require knowledge of potential interaction partners and it can be used for systems-wide analysis of protein-protein interaction networks (1). AP-MS has often been performed with the goal of a high degree of purification of protein complexes using tandem affinity purification (TAP) tagging approaches (2). However, this requires

---

§ From the Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Am Klopferspitz 18, D-82152 - Martinsried, Germany  
To whom correspondence may be addressed: Matthias Mann, Ph.: 49-89-8578-2557; Fax: 49-89-8578-2219; E-mail: mmann@biochem.mpg.de

---

<sup>1</sup> The abbreviations used are: AP-MS, affinity purification coupled with mass spectrometry; BAC, bacterial artificial chromosome; FDR, false discovery rate; GFP, green fluorescent protein; GO, Gene Ontology; LTQ, linear trap quadrupole; QUBIC, quantitative BAC interactomics; SILAC, stable isotope labeling by amino acids in cell culture; TAP, tandem affinity purification

large amounts of starting material because of the two purification steps. Furthermore, the stringent washing conditions involved in TAP purifications often lead to loss of weakly bound protein complex members. Quantitative mass spectrometry can overcome these limitations by distinguishing specific interactors from unspecific background binders by the ratios of proteins in bait versus control pull-downs (3-5). This allows single-step low stringent purification and high confidence interaction mapping including weak interactors.

There are many different formats for AP-MS and for quantitative AP-MS (6-11). Recently, our laboratory has established an integrated workflow for quantitative AP MS using bacterial artificial chromosomes (BACs) containing the gene of interest fused to the green fluorescent protein (GFP), which leads to expression of the full-length, GFP-tagged proteins from their endogenous promoters (12). This system, termed QUBIC for QUantitative BAC InteraCtomics, has several advantages. Most importantly the bait protein is expressed close to endogenous levels because the entire gene encoding the bait protein, including up- and downstream regulatory elements, is stably integrated into the genome of the cell (13-16). As tagged transcripts and proteins are processed by the cell equally to the endogenous counterpart, different splice isoforms can be expressed and proteins are post-translationally modified in the correct manner. Furthermore, cell lines expressing tagged versions of very large proteins can be created. In contrast to APs of the endogenous proteins, the QUBIC strategy does not rely on the availability of highly specific and immunoprecipitating antibodies for each protein of interest.

Most protein complexes, especially those with regulatory functions, are dynamic structures that form or change their composition and activity in response to cellular perturbations (17). Stimulation-dependent changes in protein conformation,

subcellular localization or modification determine the interaction properties of the different complex members. Quantitative AP MS using stable isotope labeling with amino acids in cell culture (SILAC) (18, 19) in a double-labeling format is frequently employed for the characterization of protein interactions. SILAC with three isotope states has previously mainly been used to study the time dimension of the proteome (20-22) but has also enabled differentiation of isoform specific interactors (6) and the change in composition of RNA polymerase upon inhibition of transcription (23).

Here we wished to establish and characterize a general method for characterizing constitutive and stimulation dependent dynamic interaction partners of regulatory protein complexes. We combined the QUBIC approach with triple SILAC labeling to differentiate background binders from specific binders and, in the same experiment, constitutive interactors from those that associate with a complex in a stimulus dependent manner. We applied this method to the analysis of complexes in the Wnt signaling pathway and investigated differential complex formation dependent on stimulation of cells with the Wnt ligand Wnt3a.

The canonical Wnt pathway regulates cell fate, proliferation and self-renewal of adult stem and progenitor cells during the entire lifespan of metazoan organisms (24-28). Aberrant regulation of this pathway leads to different diseases, most prominently sporadic colon cancer. The key step in canonical Wnt signaling is the regulation of  $\beta$ -Catenin. In the absence of Wnt ligands  $\beta$ -Catenin levels are low as a result of its continuous phosphorylation by the destruction complex, which triggers ubiquitylation and subsequent proteasomal degradation. Core components of the destruction complex are APC (Adenomatous Polyposis Coli) and Axin-1, which both function as scaffolds, and the kinases glycogen synthase kinase 3 $\beta$  (GSK-3 $\beta$ )

and casein kinase I- $\alpha$  (CKI- $\alpha$ ). Upon Wnt ligand binding to the receptors Frizzled and LRP5/6, the destruction complex function is attenuated, at least in part through relocalization to the plasma membrane and interactions with Dishevelled (DVL) (24-28). Levels of  $\beta$ -Catenin then accumulate in the cytoplasm and  $\beta$ -Catenin translocates to the nucleus where it binds to TCF/LEF transcription factors and coactivates transcription of target genes.

Because of its central importance, the Wnt pathway is intensively studied and new pathway players that may be potential therapeutic targets are still found using a variety of approaches (29-31). Although canonical Wnt signaling has been investigated in depth, the exact mechanism by which the destruction complex is inhibited and  $\beta$ -Catenin is stabilized is still not fully understood. Our Wnt pathway interactome study identifies potential novel Wnt pathway members and sheds light on the dynamics of the complexes involved.

## EXPERIMENTAL PROCEDURES

*Cell culture* - HeLa cells stably expressing BACs with human GFP-tagged APC, Axin-1, DVL2 or CtBP2 were grown at 37°C and 5% CO<sub>2</sub> for at least five passages in SILAC media (Dulbecco's modified eagle medium minus L-arginine and L-lysine (Invitrogen) with 10% dialyzed fetal bovine serum (Invitrogen) and 100 U/mL penicillin/streptomycin (Invitrogen)) containing 73  $\mu$ g/mL L arginine (Arg0) and 49  $\mu$ g/mL L lysine (Lys0) (light), 73  $\mu$ g/mL 13C614N4-L arginine (Arg6) (Euriso-top) and 49  $\mu$ g/mL 4,4,5,5-D4-L-lysine (Lys4) (medium) or 73  $\mu$ g/mL 13C615N4-L arginine (Arg10) and 49  $\mu$ g/mL 13C615N2-L Lysine (Lys8) (heavy) each supplemented with 400  $\mu$ g/mL geneticin (Invitrogen). The untransfected control HeLa Kyoto cells were only light or heavy SILAC labeled. Cells were expanded to four, 80% confluent 15 cm dishes per affinity purification and per SILAC label (in total 12 dishes for one triple SILAC experiment). The cell lines were generated by the

BAC recombineering technology (13, 14) and used as transgenic cell pools.

*Wnt stimulation and cell harvest* - In the 'forward' experiment heavy labeled transgenic cell lines were stimulated for two hours with 200 ng/mL recombinant mouse Wnt3a (RD Systems), dissolved in carrier solution (0.1% BSA in PBS). The corresponding medium labeled transgenic cells and light labeled untransfected control cells were incubated with the carrier solution for two hours. In the 'reverse' experiment the labels in the previous 'heavy' and 'light' conditions were interchanged, whereas the medium labeled condition was unchanged. Subsequently cells were trypsinized, pelleted, resuspended in PBS and counted. Equal cell numbers of each SILAC condition were separately pelleted, snap frozen and stored at -80°C.

*Preparation of cell extract*-Cell pellets were thawed on ice and resuspended in 2 mL ice-cold lysis buffer (basic buffer (150 mM NaCl, 50 mM Tris-HCl (pH 7.5), 5% glycerol), 1% IGPAL-CA-630 (Sigma), 1 mM MgCl<sub>2</sub>, 1% Benzomase (Merck) and 1x EDTA-free complete protease inhibitors (Roche). After incubation for 40 min on a rotation wheel at 4°C, lysates were centrifuged at 4,000xg for 15 min at 4°C. Supernatants were subjected to affinity purification.

*Affinity purification* - Each cleared SILAC extract was incubated separately with 100  $\mu$ L  $\mu$ MACS mouse monoclonal anti-GFP antibody coupled magnetic microbeads (Miltenyi Biotech) for 15 min. One  $\mu$ Column (Miltenyi Biotech) per SILAC extract was equilibrated with 250  $\mu$ L basic buffer containing 1% IGPAL-CA-630 (Sigma) using a hand magnet (Miltenyi Biotech). After incubation with the beads, lysates were applied to the columns. Subsequently columns were rinsed four times with 1 mL basic buffer containing 0.05% IGPAL-CA-630 (Sigma). For unspecific protein elution 25  $\mu$ L of preheated (95°C) SDS gel loading buffer (50 mM Tris HCl (pH 8), 50 mM DTT, 1% SDS, 0.005% bromophenol blue, 10% glycerol) were added and incubated for 5 min. Eluates were collected by adding additional 30  $\mu$ L preheated SDS gel loading buffer to each column. Corresponding eluates of the triple SILAC experiment and 30  $\mu$ L NuPAGE LDS sample buffer (Invitrogen) were combined.

*Protein digestion* - Combined eluates were separated by 1D-SDS PAGE (4-12 % Novex mini-

gel) (Invitrogen) and visualized by colloidal Coomassie staining (Invitrogen). Proteins were separated in three adjacent lanes that were subsequently cut into 8 slices. All gel slices were subjected to in-gel digestion with trypsin (Promega) (32, 33). Resulting tryptic peptides were extracted with 30 % ACN in 3 % TFA, concentrated until full evaporation of organic solvent and further concentrated and desalted on reversed phase C18 StageTips (34, 35). Shortly prior to high resolution liquid chromatography tandem mass spectrometry (LC MS/MS) analysis peptides were eluted twice from the StageTips with 20  $\mu$ L buffer B (80% ACN in 0.5% acetic acid) solution into a 96 sample well plate (Abgene), concentrated in a SpeedVac concentrator until removal of the organic solvent and reconstituted with buffer A\* (2% ACN in 0.1 % TFA).

*LC-MS/MS analysis* - Eluted peptides were analyzed by a nanoflow HPLC (Proxeon Biosystems) coupled on-line via a nano-electrospray ion source (Proxeon Biosystems) to a linear trap quadrupole (LTQ)-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific). Peptide mixtures were loaded with an IntelliFlow of maximal 500nl/min onto a C18 reversed phase column (15 cm long, 75  $\mu$ m inner diameter, packed in-house with ReproSil-Pur C18 AQ 3  $\mu$ m resin (Dr. Maisch)) in buffer A (0.5% acetic acid). Peptides were eluted with a multi-segment linear gradient of 5–60% buffer B (80% ACN and 0.5% acetic acid) at a constant flow rate of 250 nl/min over 107 min. Mass spectra were acquired in the positive ion mode applying a data-dependent automatic switch between survey scan and tandem mass spectra (MS/MS) acquisition. A 'top 10' method was applied that acquires one Orbitrap survey scan in the mass range of m/z 300 1650 followed by MS/MS of the ten most intense ions in the LTQ. The target value in the LTQ-Orbitrap was 1,000,000 for survey scan at a resolution of 60,000 at m/z 400. Fragmentation in the LTQ was performed by collision-induced dissociation with a target value of 5,000 ions. The ion selection threshold was 500 counts. Selected sequenced ions were dynamically excluded for 90 seconds.

*Data analysis*-Raw mass spectrometric data were analyzed with the MaxQuant software (version 1.0.14.9) (36, 37). A false discovery rate (FDR) of 0.01 for proteins and peptides and a

minimum peptide length of 6 amino acids were required. A time-dependent mass recalibration algorithm was used instead of lock masses for recalibration to improve the mass accuracy of precursor ions. MS/MS spectra were searched by Mascot (version 2.2.04, Matrix Science) against the IPI human data base (version 3.62) (containing 83,947 entries) combined with 262 common contaminants and concatenated with the reversed versions of all sequences. For the Mascot search, trypsin allowing for cleavage N-terminal to proline was chosen as enzyme specificity. Cysteine carbamido-methylation was selected as a fixed modification, while protein N-terminal acetylation and methionine oxidation were selected as variable modifications. MaxQuant was used for scoring of the peptides for identification. It also determined the SILAC state of peptides by the mass differences between SILAC peptide pairs and this information was used to perform searches with fixed Arg6 and Lys4 or Arg10 and Lys8 modifications as appropriate. Maximally two missed cleavages and three labeled amino acids were allowed. Initial mass deviation of precursor ion was up to 7 ppm, mass deviation for fragment ions was 0.5 units on the m/z scale. Protein identification required two peptides one of which had to be unique to the protein group. Quantification in MaxQuant was performed as described (36). The 'Match between runs' option was selected, which enabled the transfer of identifications between the MS analysis of the same and the adjacent gel slices of all replicates and their quantification across the replicates. The 'Requantify' option was enabled, which in effect integrates noise levels for undetected SILAC partners in order to estimate a lower bound on the SILAC ratio. Data analysis plots were either performed in the MaxQuant environment (Perseus) or in the R environment (38). The label swap of the control cell line to light SILAC label was additionally used to filter out non-assigned contaminants such as rare keratins not contained in the contaminant database (Supplemental FIG. 3). Gene Ontology (GO) analysis was performed with AMIGO (39). Cytoscape (40, 41) was used to visualize APC and Axin-1 interactome overlap. All protein group files, containing the information of all protein pull-downs, are uploaded to TRANCHE (see instructions at the end of the manuscript).

Fluorescence microscopy - HeLa cells stably expressing BACs with human GFP-tagged FAM73A were grown in 35 mm glass bottom dishes (MatTek). After staining cells with MitoTracker Red CM-H2XROS (Molecular Probes, Invitrogen), cells were imaged with a spinning-disk confocal microscope (TiLL iMIC CSU22; Andor) using a back-illuminated EM charge-coupled device camera (iXonEM 897; Andor) and a 60x 1.4 NA oil immersion objective (Olympus). 16-bit images were collected using Image iQ (version 1.9; Andor) in the linear range of the camera. They were deconvoluted with Huygens Software and cropped with ImageJ (<http://rsbweb.nih.gov/ij/>).

## RESULTS

*Triple SILAC enables detection of constitutive and dynamic interactions* - We established QUBIC (12) in a triple-encoding SILAC format to allow analysis of both constitutive and dynamic interactions of bait proteins that belong to diverse levels of canonical Wnt signaling under differential pathway activation. These encompassed APC and Axin-1 scaffold components of the destruction complex, DVL2, a mediator of the Wnt signal from the membrane to the destruction complex, as well as CtBP, a coregulator of Wnt target gene transcription (24-28).

The cells expressing GFP-tagged protein were light (L) and medium (M) SILAC labeled, while the control cell line without BAC transgene was heavy (H) SILAC labeled (FIG. 1A). The light labeled transgenic cells were stimulated with Wnt3a for two hours. To prevent light to heavy or medium to heavy exchange of specific interaction partners during the immunopurification procedure, (42, 43) GFP pull-downs were performed separately for each SILAC condition and eluates were mixed afterwards. For in-depth interactome characterization, we reduced the sample complexity by one-dimensional gel separation into eight slices. Eluates were characterized at high sensitivity on the state-

of-the-art, high-resolution LTQ-Orbitrap Velos instrument (44). Detected peptides can be classified according to their SILAC triplet peak patterns (FIG. 1B). For unspecific background binders to either beads or GFP-antibody, this pattern shows no change between the three states ('one to one to one'). A specific interactor with the GFP-bait will have peptide ratios between the pull-down with the non-stimulated cell population and the untransfected control cell population (M/H ratio in this case) and/or between the stimulated cell population and the control (L/H ratio). A stimulus dependent interactor has a significant ratio for the peptide intensity of the GFP-bait pull-down from the stimulated cell population compared to the GFP-bait pull-down from the non-stimulated cell population (L/M ratio).

Via a two-dimensional plot proteins can be grouped into constitutive interactors and dynamic interactors according to the ratios of GFP-bait to control cells (here M/H) and GFP-bait with stimulus to GFP-bait without stimulus (L/M). Figure 1C illustrates this principle in cartoon form, with outliers in the positive x-direction (M/H) representing specific binding to the bait protein. This dimension contains the information of a standard SILAC interaction experiment. Plotting the ratio between stimulated and non-stimulated cell populations (L/M) on the y-axis adds the stimulus dependent dimension. Outliers in the positive y-direction bind more strongly upon Wnt stimulus, whereas outliers in the negative direction selectively dissociate from the complex upon stimulation. The three main protein classes are therefore proteins that bind specifically to the bait but not in a signal dependent manner (blue oval in FIG. 1C), proteins that increase their binding upon stimulation (green oval) and proteins that decrease their binding (red oval).

While two of the three possible ratios of the triplets are in principle sufficient to represent the constitutive and stimulus

dependent aspects of interaction, in practice all three dimensions are often needed. This is because the third ratio can be more accurately determined directly rather than estimated from the other two. Furthermore, proteins that only bind specifically in the presence of the stimulus are not optimally represented in the graph. They are only separated from background binding proteins in one dimension (positive outlier on the y-axis). When plotting stimulus dependent binding to control versus stimulus dependent binding to non-stimulus dependent binding, these proteins are readily separated from the origin in both x- and y-directions.

Heat maps turned out to be a very valuable additional visualization method, combining all the information from triple SILAC pull-downs into a single picture. This was particularly true for integrating the data from reverse labeling experiments (see below). To generate these heat maps, we placed all pull-down ratios between the triplet states on the horizontal axis and performed one-dimensional hierarchical clustering of the multiple ratios of each quantified prey protein in the vertical dimension.

We typically detected about 1,200 proteins per pull-down experiment of which about 1,100 proteins were quantified with at least two ratio counts. These large numbers reflect the single step purification procedure and our low stringency washing conditions. However, the quantitative information encoded in the SILAC ratios immediately excluded the vast majority of these proteins from consideration (typically more than 95% of all quantified proteins).

For each bait, experiments were repeated after swapping the SILAC labels between the control and the stimulated cell populations. This introduced an additional dimension of specificity and provided a minimum of two biological replicates. Because we required at least two peptides and two ratio counts, there was a minimum of four data points for

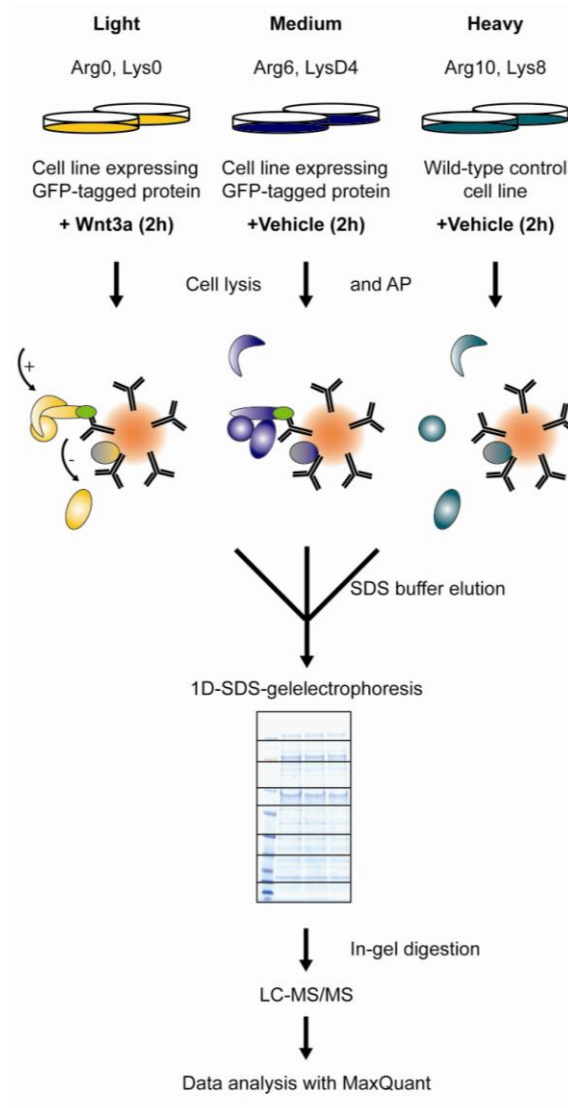
quantitation per protein. 'Forward' and 'reverse' experiments together took 1.5 days of measurement time.

*APC is an interaction scaffold that is altered by Wnt3a stimulation* - Adenomatous polyposis coli (APC) is a large (~310kDa) tumor suppressor protein that is mutated in most sporadic colorectal cancers in early tumorigenesis. Apart from its role in Wnt signaling as a member of the  $\beta$ -Catenin destruction complex, APC is involved in various other cellular processes such as cell migration, cell division, transcriptional regulation and DNA repair (45). Consequently, it has been reported to be localized in many different compartments of the cell, including the nucleus, mitochondria, mitotic spindle, centrosome, microtubules and the plasma membrane. Determination of interactors would help to further elucidate these diverse APC functions. Previous AP-MS of this protein has provided important information (29, 30). However, systematic study of its interactions has been challenging due to its large size (necessitating the use of cDNA fragments instead of full-length cDNA) and to the difficulty in distinguishing specific from non-specific proteins in the absence of quantitative methods.

We analyzed the Wnt3a stimulation dependent and independent APC interactome with the triple SILAC strategy described above and show the results of one experiment in FIG. 2A. More than 1,000 quantified proteins are plotted according to their interaction specificity (x-axis) against their stimulus specificity (y-axis). Unassigned contaminants are not labeled and are therefore readily apparent by their pattern after the label-swapping experiment and were removed (Supplemental FIG. 3). A large majority of proteins cluster around the origin, indicating that they bound equally well in the presence or absence of the bait and stimulus. We extracted the fold-change distribution of the background binders (Supplemental FIG. 1)



A



B

Unspecific background binders

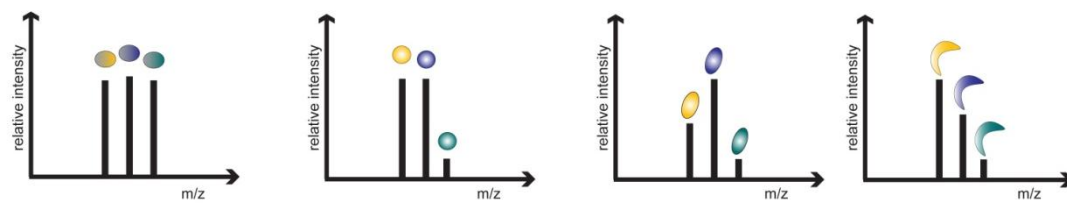
Bait protein and specific interactors

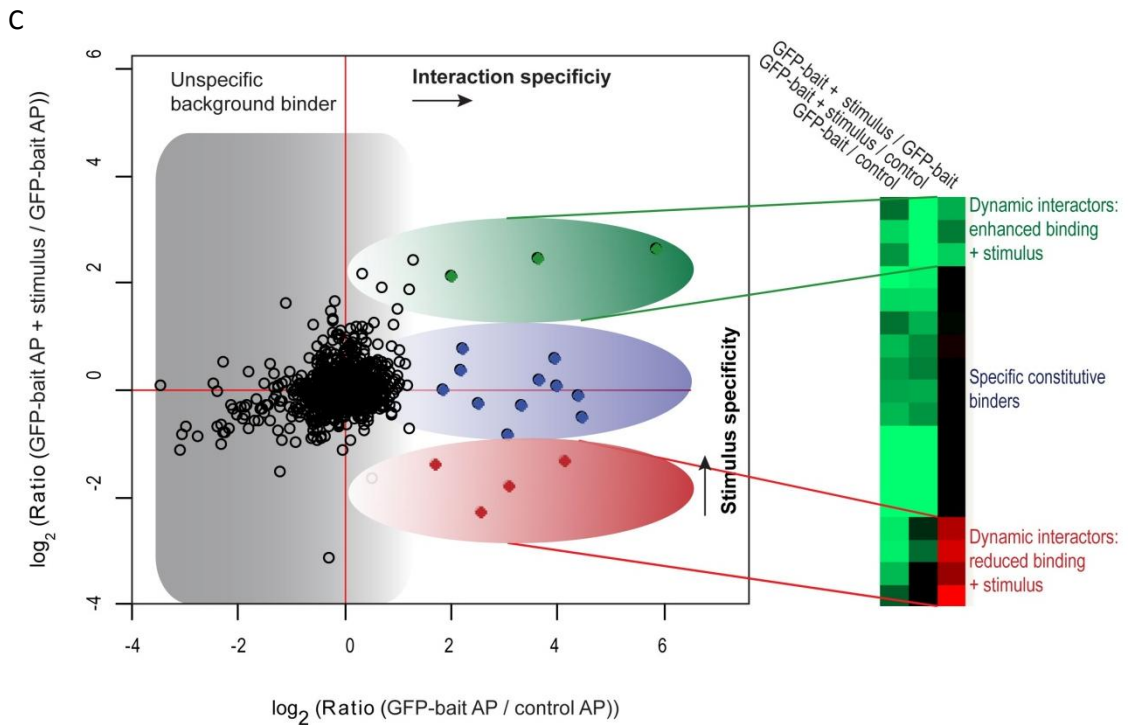
Constitutive interactions

Stimulus dependent interactions

Reduced binding

Enhanced binding





**FIG.1 Analysis of interaction dynamics by QUBIC triple SILAC based quantitative mass spectrometry.** (A) Experimental workflow for triple SILAC pull-downs to determine Wnt3a dependent interaction dynamics. The cell line expressing the GFP-tagged protein of interest is light and medium SILAC labeled, the untransfected wild-type control cell line is heavy SILAC labeled. Cells are lysed after two hour treatment with Wnt3a (200 ng/mL) or vehicle solution, respectively. GFP-pull-downs are performed separately for each SILAC state. Eluates are combined, separated on a one-dimensional gel into eight slices and in-gel digested. Resulting peptide mixtures are analyzed by high resolution LC-MS/MS on an LTQ-Orbitrap Velos. SILAC ratios are automatically quantified by MaxQuant. (B) SILAC peptide triplets representing peak profiles characteristic of background, constitutive and dynamic binders. The complete elution profile of the isotope cluster of the peptide is shown (m/z scale in x-direction, elution time in y-direction and MS signal in z-direction). (C) Data analysis plot of the ratio representing interaction specificity versus the ratio representing stimulus specificity of the interaction. Filled dots represent significant interactors and of these, constitutive interactors are depicted in blue. Dynamic interactors with enhanced binding to the bait protein are shown in red and those with reduced binding to the bait protein in green.

and determined significant outliers with box plot statistics (Supplemental FIG. 2). Many proteins are clearly separated from this background in the x-direction (specific interactors) but not in the stimulus dependent dimension. These proteins cluster around the x-axis and are colored in blue. Eight proteins are in the upper right quadrant, indicating that they bound specifically to APC and that this binding was increased upon Wnt stimulation (colored in

green). Conversely, there were nine proteins whose binding to APC decreased upon Wnt stimulation (colored in red, lower right quadrant). Interestingly, several proteins showed no specific binding to APC without stimulus at all but were recruited upon Wnt stimulation. As explained above, these proteins are more easily visualized when using the ratio of binding to GFP-APC with Wnt stimulation against control as the x-axis (FIG. 2B). In that plot APC binders that are

recruited in a stimulus dependent manner are located in the upper right quadrant (colored in green).

In total we performed four biological replicates, two of these with switched heavy and light labels (Experimental Procedures). To statistically assess the biological reproducibility of the APC interactions, we employed a one sample t-test and separately plotted the median of all protein ratios from the four replicates in volcano plots for the cases without and with Wnt stimulation (FIG. 3A and 3B). This revealed 28 reproducible APC interactors ( $p$  value  $< 0.1$  combined with a minimum ratio of 4). Next, we created a heat map of the median ratios of all proteins (Supplemental FIG. 5). The 28 reproducible APC interactors clustered together in two subgroups (FIG. 4). We also visualized the  $p$ -values determined from the t-tests as a heat map (FIG. 4). Together, these two heat maps conveniently combine the information obtained from the replicate triple pull-down experiments.

Our APC interactome includes well-known binders such as  $\beta$ -Catenin and the transcription regulator CtBP2. It covers proteins with GO cellular component annotation of all described APC localizations (Supplemental Table 1). For example, the novel APC interactor Cep170 localizes to the centrosome (46) and another novel interactor, the kinesin family member KIF2A, localizes to microtubules (47). The novel protein FAM73A had no known compartmental localization. Microscopy of a GFP-BAC line of this protein showed co-staining with mitochondrial outer membranes (Supplemental FIG. 4).

Significant APC binders also include Axin-1, CKI- $\alpha$  and Wtx, the known binding partners of APC in the cytoplasmic  $\beta$ -Catenin destruction complex. This complex is usually thought to partly disassemble upon Wnt stimulation, although evidence against this has also been reported (48). Interestingly, in our triple SILAC experiments, we did not

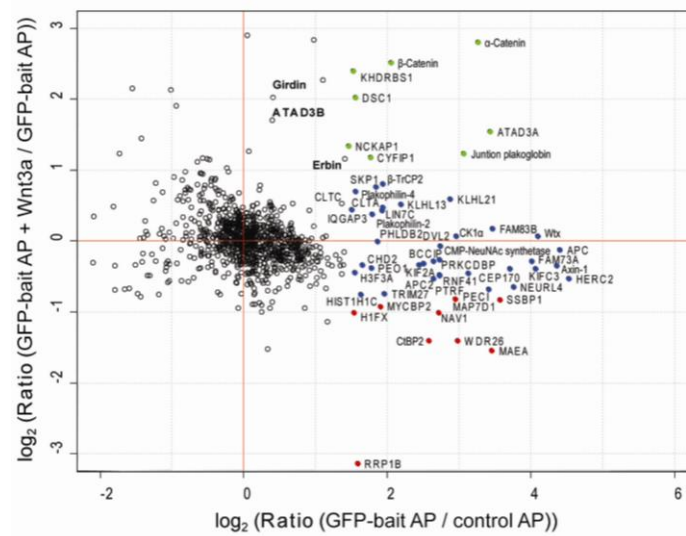
detect dynamic APC interaction changes for members of this complex upon Wnt pathway activation with Wnt3a, at least not after two hours of stimulation.

While the majority of specific interactors showed stimulus independent binding to APC, we also identified dynamic interaction changes upon Wnt3a activation (FIG. 2 and 4) such as the enhanced binding of APC to  $\beta$ -Catenin. Since the major mechanism of Wnt activation is stabilization of  $\beta$ -Catenin leading to its accumulation, this could simply be the result of more available  $\beta$ -Catenin. We likewise observed increased binding of  $\alpha$ -Catenin to APC.  $\alpha$ -Catenin is reported to indirectly associate with APC via  $\beta$ -Catenin (49), these three proteins contact each other at microtubule ends (50). Cytosolic complexes of  $\alpha$ - and  $\beta$ -Catenin have also been described (51). Thus, dynamic  $\alpha$ -Catenin binding to APC is most likely due to its association with increased levels of  $\beta$ -Catenin. Furthermore, APC binds  $\beta$ -Catenin not only within the destruction complex but also in the nucleus to enhance  $\beta$ -Catenin nuclear export in the non-Wnt activated cell (52).

Additional dynamic interactors of APC with increased binding upon Wnt3a stimulation included ATAD3A and ATAD3B. These paralogs have been reported to be localized in mitochondria, but they expose a cytosolic AAA domain (53, 54). ERBB2-interacting protein (Erbin) is a novel interactor, which potentially links APC and ERBB2 signaling. Furthermore, there is evidence that Erbin binds to  $\beta$ -Catenin and negatively regulates Wnt induced gene expression (55, 56).

We identified Girdin as a novel APC interactor that exclusively binds upon Wnt3a activation (compare FIG. 2 A and B). Girdin is a known substrate and regulator of Akt signaling (57). Furthermore, Girdin is a paralog of Daple, which has been reported to interact with the central Wnt signal mediator DVL, through its Gly-Cys-Val C-terminal motif

A



B

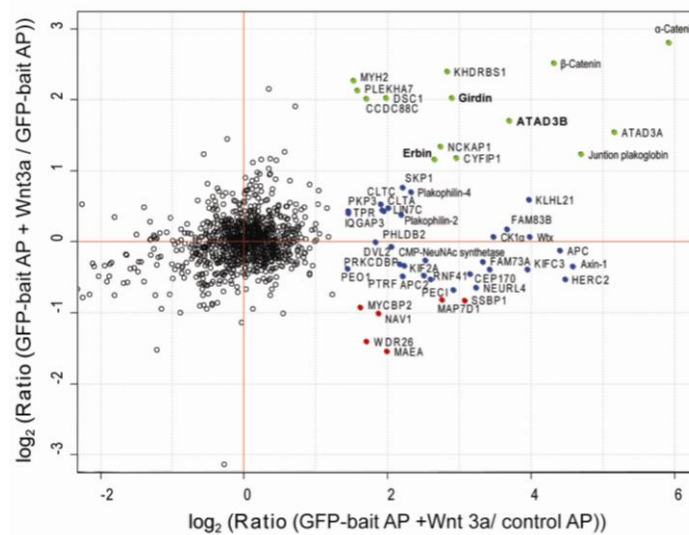
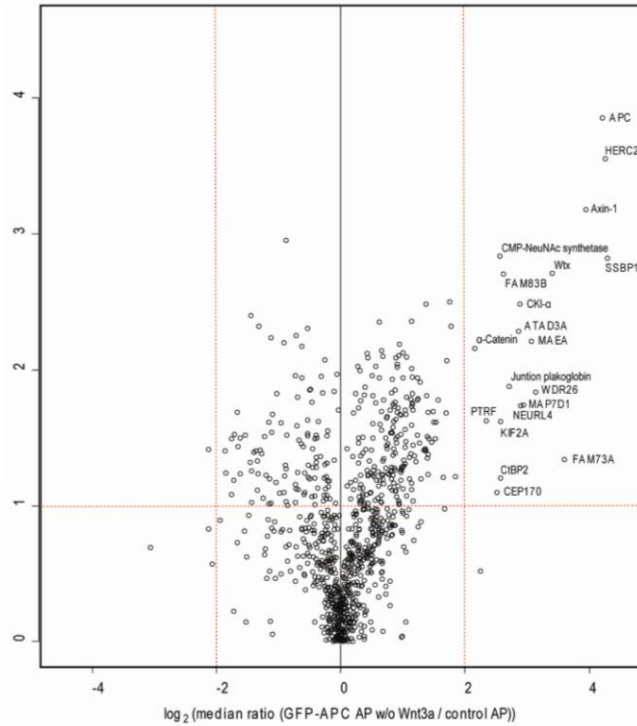


FIG. 2. **Dynamic APC interactome.** (A) Results from a triple SILAC pull-down as described in FIG. 1 using Wnt3a as the stimulus and GFP-APC as bait protein, plotted as explained in FIG. 1C. Annotated filled circles represent specific interactors determined by box plot statistics of the fold-change distribution of unstimulated pull-down against control. Specific dynamic interactors with enhanced binding to APC upon Wnt stimulation are depicted in green; the ones with decreased binding are depicted in red. Significance thresholds for dynamic changes were obtained from a box plot of fold-change distribution of stimulated pull-down against unstimulated pull-down. Constitutive, specific interactors are shown in blue. (B) Same experiment as in (A) but with the fold-change distribution of stimulated pull-down against control on the x-axis. In this plot dynamic interactors move to the upper right hand quadrant as can be seen for the proteins shown in bold.

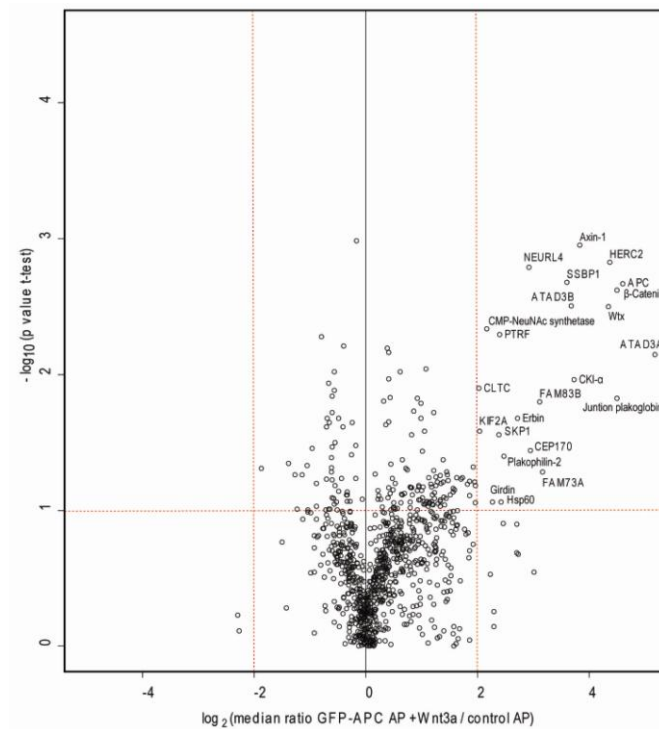
A

**FIG. 3. Volcano plot to determine reproducible APC interactors.**

(A) Log<sub>2</sub> ratios of the median of four pull-downs of GFP-APC against control (x-axis) are plotted versus  $-\log_{10}$  of the p-values derived from a t-test. Proteins with a minimum four-fold change combined with a p value smaller than 0.1 are considered significant (red lines). (B) Same as (A) but for simulated pull-downs against controls.



B



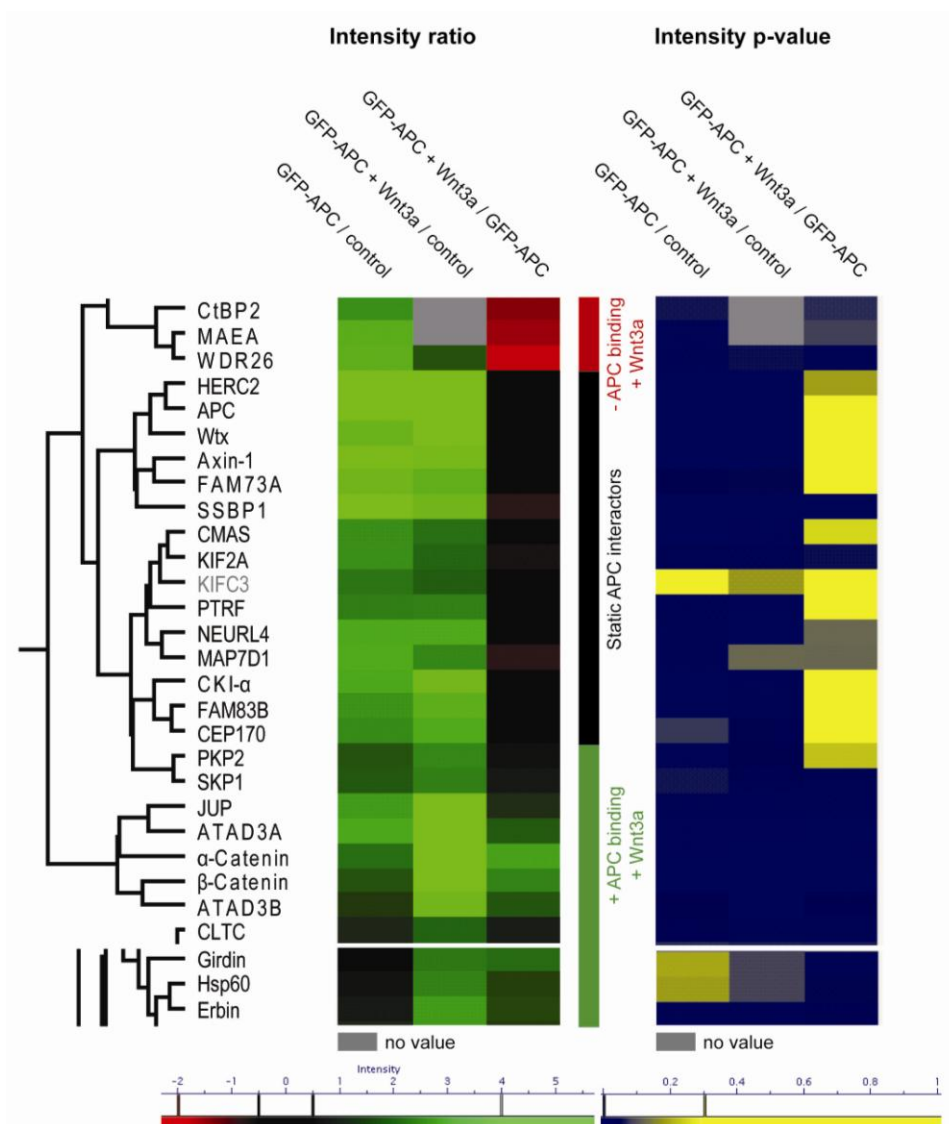


FIG. 4. **Dynamic APC interactome visualized by one-way hierarchical clustering.** The three ratios of the triple SILAC pull-down (median of four experiments) are used to cluster the reproducible APC interactors (determined in FIG. 3) by one-way hierarchical clustering. A green color value signifies specific binding to APC without Wnt stimulation (first column) or with Wnt stimulation (second column). The third column depicts the SILAC ratio of simulated against unstimulated bait pull-down. In this column a green color value represents enhanced binding to APC upon Wnt activation and a red color represents reduced binding. Constitutive interactors have no significant ratio and therefore appear in black. Additionally, t-test results for the ratio reproducibility (FIG. 3) were visualized in another panel after the clustering process. Proteins with reproducible ratios are in blue and those below the threshold in yellow. All ratio intensities are shown in log scale using the indicated color code.

(57). Interestingly, however, Girdin does not have this motif and therefore at least it must bind DVL in a different manner.

In the nucleus APC competes with the transcription factor TCF for  $\beta$ -Catenin and the APC-  $\beta$ -Catenin complex is then thought to bind CtBP2 (58). The APC  $\beta$ -Catenin CtBP2 complex reduces the pool of  $\beta$ -Catenin that can bind to TCF factors and thereby represses Wnt-dependent gene expression. The transcriptional regulator CtBP2 displays the exact opposite APC binding dynamics to Girdin. Upon Wnt3a activation, this protein is released from its association with APC. Colorectal cell lines with truncated APC have diminished binding of APC to CtBP2 (58), which therefore contributes to increased expression of Wnt target genes. Our observation that the APC-CtBP2 interaction is lost upon Wnt activation demonstrates that the truncation of APC is mechanistically equivalent to stimulation by the Wnt ligand in abolishing CtBP2 binding to the C-terminal part of APC.

Our APC interactome also contained WDR26 and MAEA, whose binding was diminished upon pathway activation. These proteins had previously been found in an Axin-1 interaction screen (29). Because APC and Axin-1 each have important roles in the destruction complex, binders to both proteins are more likely to also have Wnt related functions. This motivated us to investigate if more APC interactors might be linked to Wnt signaling in the same way. We therefore performed dynamic interaction screens for other Wnt signaling components with the aim of integrating their interactomes.

*The dynamic Axin-1 interactome reveals shared APC and Axin-1 interactors linking them to Wnt signaling* - We determined dynamic Axin-1 interactors by triple SILAC pull-downs with and without two hour Wnt3a activation (FIG. 1A). Experiments were done in biological duplicates with switched SILAC labeling (Supplemental Table 1 and

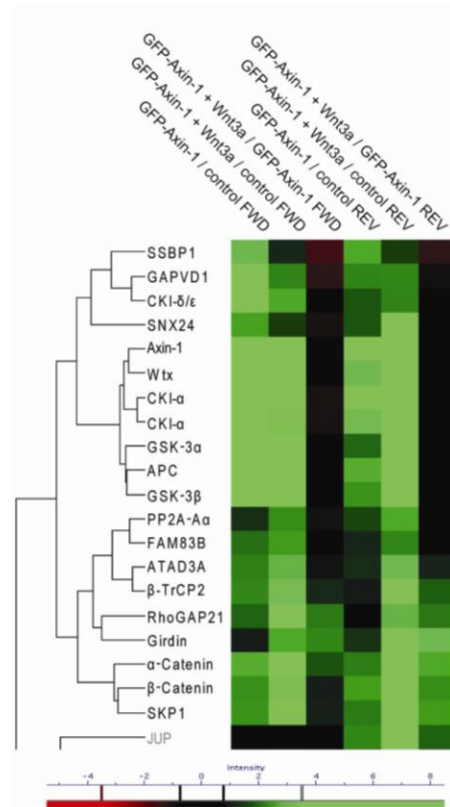
Supplemental FIG. 7). We required a minimum ratio of four for significant interactors, which was determined from box plot statistics. Results are summarized in FIG. 5A in the heat map format. In total we identified 18 specific Axin-1 interactors that were present in both duplicates. We extensively cover  $\beta$ -Catenin destruction complex component members such as APC, CKI- $\alpha$ , Wtx, GSK-3 $\beta$  and PP2A, all of which turned out to bind constitutively to Axin-1, in agreement with a recent report (48). Additionally, we found moderately increased Axin-1 interaction of SKP1 and  $\beta$ -TrCP2. Both of these proteins are members of the ubiquitin ligase complex that targets  $\beta$ -Catenin for proteasomal degradation (59). CDK1 $\delta$  is known to phosphorylate and activate DVL after Wnt3a activation (60, 61) and our data show that its binding to Axin-1 does not depend on Wnt activation. Rho GTPase activating protein 21 (RhoGAP21) was reported as a  $\beta$ -Catenin interactor on the basis of TAP pull-downs (29) and as  $\alpha$ -Catenin interactor that is required for  $\alpha$ -Catenin recruitment to adherens junctions (62). We identify this protein as a dynamic interactor to Axin-1, whose binding is markedly enhanced by Wnt activation. Further supporting its role in Wnt signaling is the observation that RhoGAP21 is also a dynamic interactor of APC (Supplemental Table 2). However, because of our stringent identification criteria it only appears in the final Axin-1 and not the APC interactome results.

Comparing the Axin-1 and APC interactomes revealed ten shared interactors (FIG. 6). Remarkably, the interaction dynamics of each shared component with either of these proteins was highly similar, which suggests that they interact with an APC-Axin-1 complex. Among the shared interactors are the known destruction complex members CKI- $\alpha$  and Wtx as well as the ubiquitin ligase component SKP1. We detect enhanced binding of  $\beta$ -Catenin to APC

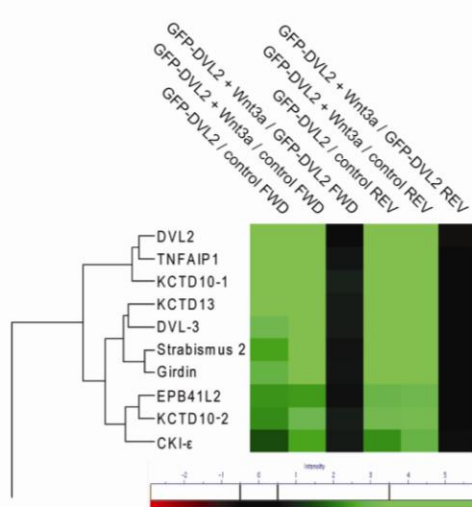
FIG. 5. **Dynamic Axin-1 (A), CtBP2 (B) and DVL2 (C) interactomes.**

One-way hierarchical clustering of SILAC ratios from biological duplicates of triple SILAC experiments performed with label switch (FWD and REV experiment). The part of the heat map with significant ratio intensities is depicted. For color coding see Fig. 4.

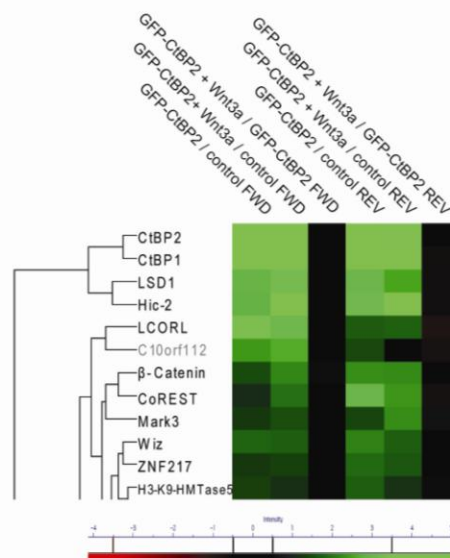
A



B



C



and Axin-1, which as noted above may partly reflect  $\beta$ -Catenin accumulation upon stimulation. Moreover, a complex of these three proteins has been reported to localize

at the membrane after Wnt3a activation (63). The interaction of Axin-1 with  $\alpha$ -Catenin was greatly enhanced by Wnt3a stimulation, similar to its binding to APC. Intriguingly



Girdin is also a highly significant dynamic interactor of Axin-1 and APC upon Wnt3a activation. The novel protein FAM83B interacts constitutively with both Axin-1 and APC. Since we have additionally observed interaction of this protein with GSK-3 $\beta$  in separate QUBIC experiments (Supplemental FIG. 6), it may be a novel member of the destruction complex and is very likely involved in Wnt signaling.

*The DVL2 interactome reveals constitutive binding of Girdin independent of Wnt activation* – Disheveled (DVL) proteins are important Wnt signal mediators that counteract destruction complex action upon Wnt3a stimulation. The exact mechanism by which the Wnt signal is transduced - including the complex formation at the membrane, which involves phosphorylated DVL - is still not fully understood (26). Furthermore, DVL proteins integrate different branches of Wnt signaling including the planar cell polarity (PCP) pathway. Triple SILAC DVL2 pull-downs covered known DVL2 binder such as DVL3 and the positive Wnt regulator CKI- $\epsilon$  (64). The negative Wnt PCP regulator Vang-like protein 1/Strabismus 2 (65, 66), was also a significant interactor of DVL2. None of the ten identified interactions were modulated by the Wnt signal (FIG. 5B, Supplemental FIG. 8). Among the newly discovered interactors, we found the three members of the BTB/POZ domain-containing protein family KCTD10/KCTD13/TNFAIP1 (67). We detect Girdin as a novel and specific interactor for DVL2. In contrast to its dynamically increased binding to APC and Axin-1 upon Wnt3a stimulation, Girdin binds constitutively to DVL2.

*CtBP2 binds to  $\beta$  catenin in a non-stimulus dependent manner* - We determined the dynamic CtBP2 interactome to characterize a potentially dynamic nuclear regulator of Wnt signaling. We detected 12 specific interactors of CtBP2 but none of them with significant stimulus dependent changes (FIG. 5C,

Supplemental FIG. 8). Among these was  $\beta$  catenin, which is known to interact with CtBP2 as well as with APC to repress transcription of Wnt target genes (see also above). Upon Wnt stimulation  $\beta$  catenin interacts with CtBP2 and TCF to contribute to transcriptional activation (68). Our observation that the  $\beta$  catenin-CtBP2 interaction is not dynamically regulated by Wnt, agrees with suggestions that these two proteins function in both repression (58) and activation of gene expression (68). The majority of specific CtBP2 interactors were members of complexes that deposit activating or remove repressive histone methylation marks (69, 70). This concurs with known CtBP2 functions and interactions, which are broad and not only specific to Wnt signaling (71).

#### DISCUSSION AND OUTLOOK

Here we have described a three-state quantitative proteomics approach to study the dynamics of protein-protein interactions. We used the SILAC technology because of its simplicity and accuracy of quantification when coupled to a high resolution mass spectrometric readout. Proteins were expressed as GFP fusions from bacterial artificial chromosomes that had been integrated into the host cell genome, ensuring close to endogenous expression levels, correct modification state and compartmentalization of the bait proteins. This is especially important for studying a signaling pathway such as Wnt, in which the regulation of protein amount is critical and in which the signal propagates through different cellular locations. The approach described here is generic as BAC-GFP cell lines can be produced in a streamlined procedure (16). We employed one-dimensional gel separation using somewhat more material (four 15 cm dishes per condition) than in single run analyses (12). Analysis of the results was

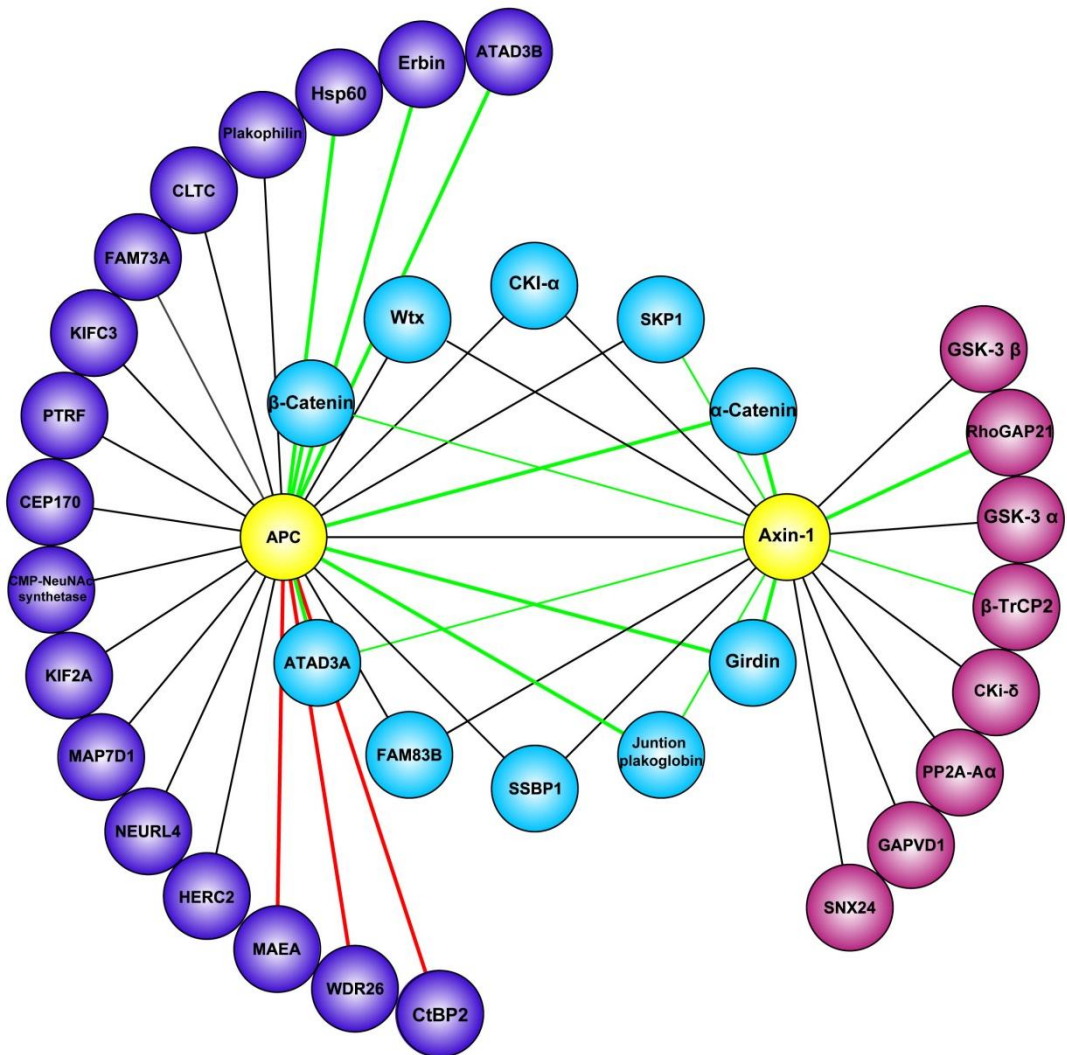


Fig. 6. **Overlap of APC and Axin-1 interactomes.** Protein-protein interactions were drawn in Cytoscape, after importing the pull-down data from FIG. 4 and FIG. 5A. Baits are depicted in yellow, shared APC and Axin-1 interactors in blue and unique interactors for APC and Axin-1 in purple and pink, respectively. Lines represent detected interactions. Green lines indicate enhanced interaction upon Wnt3a activation while red lines indicate reduced interactions upon Wnt3a activation. Line width reflects the SILAC ratio intensity for the dynamic interactors.

more complex than double-labeling SILAC because three states are compared. However, these analysis steps have now been incorporated into the freely available MaxQuant environment or as R-scripts. Consequently, dynamic analysis of interaction partners is relatively streamlined and it can now be used routinely for pathways of interest or as a follow up on initial high-throughput protein interaction screens.

There are several obvious extensions of the workflow described here. For example, dynamic interaction measurements can be repeated at different time points to investigate the changing composition of signaling complexes over time. In this experiment the three SILAC states can each represent different time points because the specificity of the binders has already been established.

Here we have applied the QUBIC triple SILAC dynamic interaction screen to the challenging case of Wnt signaling. We performed pull-downs on central members of the pathway from the destruction complex and from other different pathway levels. The SILAC ratios efficiently filtered out non-specific binders, reducing an initial set of about 1,000 identified proteins to a relatively small number (10 to 50). These proteins contained many positive controls that were either known interaction partners or that already had some other connection to Wnt signaling. Among the novel interaction partners we focused on those shared by APC and Axin-1, which are most likely to be functional members of the Wnt pathway. Interestingly, almost all of these interaction partners turned out to have similar interaction dynamics, consistent with a role in a shared complex with APC and Axin-1. Other dynamic interaction partners of APC or Axin-1 are also good candidates for functional roles in this pathway by virtue of their Wnt-dependent interaction modulation.

One example is Girdin (CCDC88A), which we separately found as a novel interactor of

APC, Axin 1 and Disheveled 2 (DVL2). Intriguingly, interaction with APC and Axin-1 are contingent on Wnt-stimulation, whereas interaction with DVL2 is not. This raises the possibility that Girdin and DVL2 are in a pre-formed complex, which may then recruit destruction complex members to the membrane receptors upon Wnt pathway stimulation. In this context, we observed that interaction of APC and Axin-1 with other destruction complex members did not change upon Wnt3a activation. This finding sheds some light on the unresolved mechanism of destruction complex dynamics at the plasma membrane. In concordance with recent observations (48), our data is consistent with a potential translocation of a relatively intact destruction complex, at least after two hours of Wnt stimulation.

The majority of colon cancers have a truncated form of APC. This interferes with APC's role in the destruction complex. Additionally, truncated APC loses binding to CtBP2, diminishing its repressor function (58). Our data show that CtBP2 binding to full length APC is entirely Wnt-dependent. This implies that colon cancer cells could also lose this binding by another manipulation of the Wnt signal cascade, different from APC truncation.

In conclusion, we have described a streamlined interaction screen, which accurately discriminates constitutive from dynamic, signal dependent interactions. In contrast to targeted techniques, such as western blotting, it can both discover and characterize such dynamic interactors in the same experiment. Interaction proteomics is now clearly ready to add a dynamic dimension to protein interaction studies.

*Acknowledgments-* We thank Bianca Splettstoesser for technical help and Markus Räsche and Nina C. Hubner for critical reading of the manuscript and helpful discussions. Juergen Cox advised on the data analysis. Natalie Kraemer from the group of Organelle Architecture and

Dynamics helped with the microscopy. We also thank Ina Poser and Anthony A. Hyman from the Max Planck Institute of Molecular Cell Biology and Genetics for kindly providing the GFP-tagged cell lines. This work was supported by Munich Center for Integrated Protein Science (CIPSM) and by European Commission's 7th Framework Program PROteomics SPECification in Time and Space (PROSPECTS, HEALTH-F4-2008-021648).

## REFERENCES

1. Gingras, A. C., Gstaiger, M., Raught, B., and Aebersold, R. (2007) Analysis of protein complexes using mass spectrometry. *Nature reviews* 8, 645-654.
2. Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nature biotechnology* 17, 1030-1032.
3. Ranish, J. A., Yi, E. C., Leslie, D. M., Purvine, S. O., Goodlett, D. R., Eng, J., and Aebersold, R. (2003) The study of macromolecular complexes by quantitative proteomics. *Nature genetics* 33, 349-355.
4. Blagoev, B., Kratchmarova, I., Ong, S. E., Nielsen, M., Foster, L. J., and Mann, M. (2003) A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nature biotechnology* 21, 315-318.
5. Vermeulen, M., Hubner, N. C., and Mann, M. (2008) High confidence determination of specific protein-protein interactions using quantitative mass spectrometry. *Current opinion in biotechnology* 19, 331-337.
6. Trinkle-Mulcahy, L., Andersen, J., Lam, Y. W., Moorhead, G., Mann, M., and Lamond, A. I. (2006) Repo-Man recruits PP1 gamma to chromatin and is essential for cell viability. *The Journal of cell biology* 172, 679-692.
7. Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dumpelfeld, B., Edlmann, A., Heurtier, M. A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A. M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J. M., Kuster, B., Bork, P., Russell, R. B., and Superti-Furga, G. (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440, 631-636.
8. Kocher, T., and Superti-Furga, G. (2007) Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nature methods* 4, 807-815.
9. Sowa, M. E., Bennett, E. J., Gygi, S. P., and Harper, J. W. (2009) Defining the human deubiquitinating enzyme interaction landscape. *Cell* 138, 389-403.
10. Wepf, A., Glatter, T., Schmidt, A., Aebersold, R., and Gstaiger, M. (2009) Quantitative interaction proteomics using mass spectrometry. *Nature methods* 6, 203-205.
11. Mak, A. B., Ni, Z., Hewel, J. A., Chen, G. I., Zhong, G., Karamboulas, K., Blakely, K., Smiley, S., Marcon, E., Roudeva, D., Li, J., Olsen, J. B., Punna, T., Isserlin, R., Chetyrkin, S., Gingras, A. C., Emili, A., Greenblatt, J., and Moffat, J. (2010) A lentiviral-based functional proteomics approach identifies chromatin remodelling complexes important for the induction of pluripotency *Mol Cell Proteomics* In press.
12. Hubner, N. C., Bird, A. W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A., Mann, M. (2010) Quantitative proteomics combined with BAC TransgeneOmics reveals in-vivo protein interactions. *JCB* in press.
13. Zhang, Y., Buchholz, F., Muylers, J. P., and Stewart, A. F. (1998) A new logic for DNA engineering using recombination in *Escherichia coli*. *Nature genetics* 20, 123-128.
14. Kittler, R., Pelletier, L., Ma, C., Poser, I., Fischer, S., Hyman, A. A., and Buchholz, F. (2005) RNA interference rescue by bacterial artificial chromosome transgenesis in mammalian tissue culture cells. *Proceedings of the National Academy of Sciences of the United States of America* 102, 2396-2401.
15. Sarov, M., Schneider, S., Pozniakovski, A., Roguev, A., Ernst, S., Zhang, Y., Hyman, A. A., and Stewart, A. F. (2006) A recombineering pipeline for functional genomics applied to *Caenorhabditis elegans*. *Nature methods* 3, 839-844.
16. Poser, I., Sarov, M., Hutchins, J. R., Heriche, J. K., Toyoda, Y., Pozniakovski, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A. W., Pelletier, L., Kittler, R., Hua, S., Naumann, R., Augsburg, M., Sykora, M. M., Hofemeister, H., Zhang, Y., Nasmyth, K., White, K. P., Dietzel, S., Mechtler, K., Durbin, R., Stewart, A. F., Peters, J.

- M., Buchholz, F., and Hyman, A. A. (2008) BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nature methods* 5, 409-415.
17. Charbonnier, S., Gallego, O., and Gavin, A. C. (2008) The social network of a cell: recent advances in interactome mapping. *Biotechnology annual review* 14, 1-28.
18. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 1, 376-386.
19. Mann, M. (2006) Functional and quantitative proteomics using SILAC. *Nature reviews* 7, 952-958.
20. Blagoev, B., Ong, S. E., Kratchmarova, I., and Mann, M. (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nature biotechnology* 22, 1139-1145.
21. Andersen, J. S., Lam, Y. W., Leung, A. K., Ong, S. E., Lyon, C. E., Lamond, A. I., and Mann, M. (2005) Nucleolar proteome dynamics. *Nature* 433, 77-83.
22. Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127, 635-648.
23. Boulon, S., Ahmad, Y., Trinkle-Mulcahy, L., Verheggen, C., Cogley, A., Gregor, P., Bertrand, E., Whitehorn, M., and Lamond, A. I. (2009) Establishment of a protein frequency library and its application in the reliable identification of specific protein interaction partners. *Mol Cell Proteomics*.
24. Clevers, H. (2006) Wnt/beta-catenin signaling in development and disease. *Cell* 127, 469-480.
25. Klaus, A., and Birchmeier, W. (2008) Wnt signalling and its impact on development and cancer. *Nat Rev Cancer* 8, 387-398.
26. Angers, S., and Moon, R. T. (2009) Proximal events in Wnt signal transduction. *Nature reviews* 10, 468-477.
27. MacDonald, B. T., Tamai, K., and He, X. (2009) Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Developmental cell* 17, 9-26.
28. van Amerongen, R., and Nusse, R. (2009) Towards an integrated view of Wnt signaling in development. *Development (Cambridge, England)* 136, 3205-3214.
29. Major, M. B., Camp, N. D., Berndt, J. D., Yi, X., Goldenberg, S. J., Hubbert, C., Biechele, T. L., Gingras, A. C., Zheng, N., Maccoss, M. J., Angers, S., and Moon, R. T. (2007) Wilms tumor suppressor WTX negatively regulates WNT/beta-catenin signaling. *Science (New York, N.Y)* 316, 1043-1046.
30. Major, M. B., Roberts, B. S., Berndt, J. D., Marine, S., Anastas, J., Chung, N., Ferrer, M., Yi, X., Stoick-Cooper, C. L., von Haller, P. D., Kategaya, L., Chien, A., Angers, S., MacCoss, M., Cleary, M. A., Arthur, W. T., and Moon, R. T. (2008) New regulators of Wnt/beta-catenin signaling revealed by integrative molecular screening. *Science signaling* 1, ra12.
31. Sue Ng, S., Mahmoudi, T., Li, V. S., Hatzis, P., Boersema, P. J., Mohammed, S., Heck, A. J., and Clevers, H. (2010) MAP3K1 functionally interacts with Axin1 in the canonical Wnt signalling pathway. *Biological chemistry* 391, 171-180.
32. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Analytical chemistry* 68, 850-858.
33. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., and Mann, M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature protocols* 1, 2856-2860.
34. Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Analytical chemistry* 75, 663-670.
35. Rappsilber, J., Mann, M., and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature protocols* 2, 1896-1906.
36. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* 26, 1367-1372.

37. Cox, J., Matic, I., Hilger, M., Nagaraj, N., Selbach, M., Olsen, J. V., and Mann, M. (2009) A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nature protocols* 4, 698-705.
38. Team, R. D. C. (2008) R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria.
39. Carbon, S., Ireland, A., Mungall, C. J., Shu, S., Marshall, B., and Lewis, S. (2009) AmiGO: online access to ontology and annotation data. *Bioinformatics (Oxford, England)* 25, 288-289.
40. Cline, M. S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B., Hanspers, K., Isserlin, R., Kelley, R., Killcoyne, S., Lotia, S., Maere, S., Morris, J., Ono, K., Pavlovic, V., Pico, A. R., Vailaya, A., Wang, P. L., Adler, A., Conklin, B. R., Hood, L., Kuiper, M., Sander, C., Schmulevich, I., Schwikowski, B., Warner, G. J., Ideker, T., and Bader, G. D. (2007) Integration of biological networks and gene expression data using Cytoscape. *Nature protocols* 2, 2366-2382.
41. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 13, 2498-2504.
42. Mousson, F., Kolkman, A., Pijnappel, W. W., Timmers, H. T., and Heck, A. J. (2008) Quantitative proteomics reveals regulation of dynamic components within TATA-binding protein (TBP) transcription complexes. *Mol Cell Proteomics* 7, 845-852.
43. Wang, X., and Huang, L. (2008) Identifying dynamic interactors of protein complexes by quantitative mass spectrometry. *Mol Cell Proteomics* 7, 46-57.
44. Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E. R., Senko, M., Makarov, A., Mann, M., and Horning, S. (2009) A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol Cell Proteomics* 8, 2759-2769.
45. Brocardo, M., and Henderson, B. R. (2008) APC shuttling to the membrane, nucleus and beyond. *Trends in cell biology* 18, 587-596.
46. Guarguaglini, G., Duncan, P. I., Stierhof, Y. D., Holmstrom, T., Duensing, S., and Nigg, E. A. (2005) The forkhead-associated domain protein Cep170 interacts with Polo-like kinase 1 and serves as a marker for mature centrioles. *Molecular biology of the cell* 16, 1095-1107.
47. Ganem, N. J., and Compton, D. A. (2004) The Kif1 kinesin Kif2a is required for bipolar spindle assembly through a functional relationship with MCAK. *The Journal of cell biology* 166, 473-478.
48. Yokoyama, N., Yin, D., and Malbon, C. C. (2007) Abundance, complexation, and trafficking of Wnt/beta-catenin signaling elements in response to Wnt3a. *Journal of molecular signaling* 2, 11.
49. Rubinfeld, B., Souza, B., Albert, I., Munemitsu, S., and Polakis, P. (1995) The APC protein and E-cadherin form similar but independent complexes with alpha-catenin, beta-catenin, and plakoglobin. *The Journal of biological chemistry* 270, 5549-5555.
50. Harris, T. J., and Peifer, M. (2005) Decisions, decisions: beta-catenin chooses between adhesion and transcription. *Trends in cell biology* 15, 234-237.
51. Bienz, M. (2005) beta-Catenin: a pivot between cell adhesion and Wnt signalling. *Curr Biol* 15, R64-67.
52. Henderson, B. R. (2000) Nuclear-cytoplasmic shuttling of APC regulates beta-catenin subcellular localization and turnover. *Nature cell biology* 2, 653-660.
53. Bogenhagen, D. F., Rousseau, D., and Burke, S. (2008) The layered structure of human mitochondrial DNA nucleoids. *The Journal of biological chemistry* 283, 3665-3675.
54. He, J., Mao, C. C., Reyes, A., Sembongi, H., Di Re, M., Granycome, C., Clippingdale, A. B., Fearnley, I. M., Harbour, M., Robinson, A. J., Reichelt, S., Spelbrink, J. N., Walker, J. E., and Holt, I. J. (2007) The AAA+ protein ATAD3 has displacement loop binding properties and is involved in mitochondrial nucleoid organization. *The Journal of cell biology* 176, 141-146.
55. Ress, A., and Moelling, K. (2008) The PDZ protein erbin modulates beta-catenin-dependent transcription. *Eur Surg Res* 41, 284-289.
56. Ress, A., and Moelling, K. (2006) Interaction partners of the PDZ domain of erbin. *Protein and peptide letters* 13, 877-881.

57. Enomoto, A., Ping, J., and Takahashi, M. (2006) Girdin, a novel actin-binding protein, and its family of proteins possess versatile functions in the Akt and Wnt signaling pathways. *Annals of the New York Academy of Sciences* 1086, 169-184.
58. Hamada, F., and Bienz, M. (2004) The APC tumor suppressor binds to C-terminal binding protein to divert nuclear beta-catenin from TCF. *Developmental cell* 7, 677-685.
59. Liu, C., Kato, Y., Zhang, Z., Do, V. M., Yankner, B. A., and He, X. (1999) beta-Trcp couples beta-catenin phosphorylation-degradation and regulates *Xenopus* axis formation. *Proceedings of the National Academy of Sciences of the United States of America* 96, 6273-6278.
60. Bryja, V., Schulte, G., and Arenas, E. (2007) Wnt-3a utilizes a novel low dose and rapid pathway that does not require casein kinase 1-mediated phosphorylation of Dvl to activate beta-catenin. *Cellular signalling* 19, 610-616.
61. Peters, J. M., McKay, R. M., McKay, J. P., and Graff, J. M. (1999) Casein kinase I transduces Wnt signals. *Nature* 401, 345-350.
62. Sousa, S., Cabanes, D., Archambaud, C., Colland, F., Lemichez, E., Popoff, M., Boisson-Dupuis, S., Gouin, E., Lecuit, M., Legrain, P., and Cossart, P. (2005) ARHGAP10 is necessary for alpha-catenin recruitment at adherens junctions and for *Listeria* invasion. *Nature cell biology* 7, 954-960.
63. Hendriksen, J., Jansen, M., Brown, C. M., van der Velde, H., van Ham, M., Galjart, N., Offerhaus, G. J., Fagotto, F., and Fornerod, M. (2008) Plasma membrane recruitment of dephosphorylated beta-catenin upon activation of the Wnt pathway. *Journal of cell science* 121, 1793-1802.
64. Sakanaka, C., Leong, P., Xu, L., Harrison, S. D., and Williams, L. T. (1999) Casein kinase Iepsilon in the wnt pathway: regulation of beta-catenin function. *Proceedings of the National Academy of Sciences of the United States of America* 96, 12548-12552.
65. Park, M., and Moon, R. T. (2002) The planar cell-polarity gene *stbm* regulates cell behaviour and cell fate in vertebrate embryos. *Nature cell biology* 4, 20-25.
66. Angers, S., Thorpe, C. J., Biechele, T. L., Goldenberg, S. J., Zheng, N., MacCoss, M. J., and Moon, R. T. (2006) The KLHL12-Cullin-3 ubiquitin ligase negatively regulates the Wnt-beta-catenin pathway by targeting Dishevelled for degradation. *Nature cell biology* 8, 348-357.
67. Van Bogaert, P., Azizieh, R., Desir, J., Aeby, A., De Meirleir, L., Laes, J. F., Christiaens, F., and Abramowicz, M. J. (2007) Mutation of a potassium channel-related gene in progressive myoclonic epilepsy. *Annals of neurology* 61, 579-586.
68. Fang, M., Li, J., Blauwkamp, T., Bhambhani, C., Campbell, N., and Cadigan, K. M. (2006) C-terminal-binding protein directly activates and represses Wnt transcriptional targets in *Drosophila*. *The EMBO journal* 25, 2735-2745.
69. Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstone, J. R., Cole, P. A., Casero, R. A., and Shi, Y. (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 119, 941-953.
70. Shi, Y., Sawada, J., Sui, G., Affar el, B., Whetstone, J. R., Lan, F., Ogawa, H., Luke, M. P., Nakatani, Y., and Shi, Y. (2003) Coordinated histone modifications mediated by a CtBP corepressor complex. *Nature* 422, 735-738.
71. Chinnadurai, G. (2002) CtBP, an unconventional transcriptional corepressor in development and oncogenesis. *Mol Cell* 9, 213-224.

Supplemental data are provided upon request: hilger@biochem.mpg.de





## 7.6 Curriculum Vitae

# Maximiliane Hilger

---

10 April 1982  
German

Franz-Senn-Straße 28  
81377 Munich  
Germany

Tel. +49 (0) 89 32 60 81 52  
hilger@biochem.mpg.de

## EDUCATION

---

### PhD studies

Since 10/2006      **International Max Planck Research School for Molecular and Cellular Life Sciences (IMPRS)**  
**Max Planck Institute of Biochemistry**  
Martinsried, Germany

### University

10/2001 – 09/2006      **University of Bielefeld**  
Bielefeld, Germany  
Diplom (MSc equivalent) in Molecular Biotechnology 2006  
Vordiplom (BSc equivalent) in Molecular Biotechnology 2004

09/2003 – 01/2004      **Dublin City University**  
Dublin, Ireland  
ERASMUS scholar

### School

1992 – 2001      **Märkisches Gymnasium Schwelm (High School)**  
Schwelm, Germany  
Abitur (equivalent to 'A' level) 2001

## RESEARCH EXPERIENCE

---

Since 10/2006	<b>PhD research</b>	
	Supervisor	Prof. Dr. Matthias Mann Max Planck Institute of Biochemistry Martinsried, Germany Department of Proteomics and Signal Transduction
	Project	Unraveling signal transduction networks by high-resolution and quantitative mass spectrometry-based proteomics
01/2006 – 09/2006	<b>Diplom project</b>	
	Supervisor	PD Dr. Oliver Müller (since 2008 Prof. at FH Kaiserslautern) Max Planck Institute of Molecular Physiology Dortmund, Germany Group of Tumour Genetics Department of Structural Biology (PI: Prof. Dr. Alfred Wittinghofer)
	Project	Regulation of the target genes <i>c-jun</i> and <i>fra-1</i> by the Wnt pathway and their downregulation by T7-siRNA

## INTERNSHIPS

---

02/2004	<b>Dublin City University</b> Fermentation technology
09/2002	<b>MPI of Molecular Physiology in Dortmund</b> Molecular biology
07/2001 – 08/2001	<b>Bitop (Gesellschaft für biotechnische Optimierung mbH)</b> Microbiology and molecular biology
07/2000	<b>Hospital Klinikum Wuppertal GmbH</b> Oncology

## EXTRACURRICULAR ACTIVITIES

---

2008 – 2009	PhD representative of the department
2009	Organizer of PhD symposium <interact> in Munich, responsible for posters and speakers
2004 – 2006	Formation of a branch office of the “biotechnologische Studenteninitiative (btS)” in Bielefeld and participation as second board of management

---

## PUBLICATIONS

---

**Hilger, M.**, Bonaldi, T., Gnad, F., and Mann, M. (2009) Systems-wide analysis of a phosphatase knock-down by quantitative proteomics and phosphoproteomics. **Mol Cell Proteomics** 8, 1908-1920.

Cox, J., Matic, I., **Hilger, M.**, Nagaraj, N., Selbach, M., Olsen, J. V., and Mann, M. (2009) A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. **Nature protocols** 4, 698-705.

Matic, I., Macek, B., **Hilger, M.**, Walther, T. C., and Mann, M. (2008) Phosphorylation of SUMO-1 occurs in vivo and is conserved through evolution. **Journal of proteome research** 7, 4050-4057.

---

## ATTENDED CONFERENCES

---

**Signal Transduction and Disease  
Trinational Fall Meeting of the Biochemical Societies**

Aachen, Germany, September 2009

Poster presentation

**Wnt Meeting**

Washington DC, USA, June 2009

Talk

**HUPO 7th Annual World Congress**

Amsterdam, Netherlands, August 2008

Talk

**Phosphorylation, Signaling & Disease**

Cold Spring Harbor, NY, USA, May 2007

Poster presentation