

# **Phosphoproteome studies of human mitotic spindle proteins**

Dissertation  
zur Erlangung des Doktorgrades der Naturwissenschaften  
(Dr. rer. nat.)  
der Fakultät für Biologie  
der Ludwig-Maximilians-Universität München

Vorgelegt von  
Bin Wang  
München 2009

Dissertation eingereicht am 17.12.2009  
Datum der mündlichen Prüfung: 04.05.2010

Erstgutachter: Prof. Dr. Erich A. Nigg  
Zweitgutachter: Prof. Dr. Angelika Boettger

**Ehrenwörtliche Erklärung und Erklärung über frühere Promotionsversuche**

Hiermit erkläre ich, dass ich, Bin Wang, die vorliegende Dissertation selbständig und ohne unerlaubte Hilfe angefertigt habe. Sämtliche Experimente wurden von mir selbst durchgeführt, soweit nicht explizit auf Dritte verwiesen wird. Ich habe weder anderweitig versucht, eine Dissertation einzureichen oder eine Doktorprüfung durchzuführen, noch habe ich diese Dissertation oder Teile derselben einer anderen Prüfungskommission vorgelegt.

Bin Wang

München, den 16.Dec.2009



Table of contents .....	
<b>Acknowledgement.....</b>	<b>1</b>
<b>Summary .....</b>	<b>3</b>
<b>1 Introduction .....</b>	<b>5</b>
1.1 Cell cycle .....	5
1.2 Mitosis.....	5
1.2.1 The mitotic spindle .....	7
1.2.2 Phosphorylation in mitotic progression .....	8
1.3 Polo-like kinase 1 (Plk1) .....	9
1.4 Mass spectrometry based proteomics .....	9
1.4.1 General workflow of MS-based proteomics.....	9
1.4.2 MS instrumentation .....	11
1.4.3 MS-based Quantitative Proteomics.....	16
1.4.4 Phosphoproteomics .....	20
<b>2 Aims of this study .....</b>	<b>22</b>
<b>3 Results .....</b>	<b>23</b>
3.1 Quantitative identification of Polo-like kinase 1 (Plk1) specific phosphorylation sites on human mitotic spindle proteins .....	23
3.1.1 Introduction .....	23
3.1.2 Experimental design for the comparison of the spindle phosphoproteome between cells in the presence or absence of Plk1 .....	25
3.1.3 Plk1-dependent human mitotic spindle proteome and phosphoproteome .....	29
3.1.4 Plk1-dependent quantitative human mitotic spindle phosphoproteome .....	30
3.1.5 Plk1 spindle substrates identification .....	35
3.1.6 Validation of Plk1-dependent phosphorylation sites on spindle proteins through a candidate-based <i>in vitro</i> kinase screening .....	35
3.1.7 Reconsideration of the Plk1 consensus motif .....	38
3.1.8 Plk1-dependent localization of spindle components .....	39
3.1.9 Integration of Aurora A and Plk1 activity .....	41
3.2 Phosphoproteome analysis of human mitotic spindle proteins by using low- specificity protease elastase .....	44
3.2.1 Introduction .....	44
3.2.2 Experimental outline.....	46
3.2.3 Adjustment of the Elastase Concentration .....	46
3.2.4 MS Mass Accuracy .....	46

3.2.5	Elastase cleavage specificity .....	48
3.2.6	Elastase cleavage scores.....	50
3.2.7	Comparison of phosphorylation site identification using trypsin and elastase...	51
3.2.8	“Custom database” search strategy .....	53
3.2.9	Complementary phosphorylated peptide coverages using trypsin and elastase	54
3.2.10	Testing for undersampling effects .....	56
<b>4</b>	<b>Discussion .....</b>	<b>58</b>
<b>5</b>	<b>Material and Methods.....</b>	<b>63</b>
5.1	Quantitative identification of Polo-like kinase 1 (Plk1) specific phosphorylation sites on human mitotic spindle protein.....	63
5.1.1	Materials .....	63
5.1.2	SILAC media .....	63
5.1.3	Cell culture and spindle isolation .....	63
5.1.4	<i>In vitro</i> kinase assays .....	64
5.1.5	Gel Electrophoresis and In-Gel Digestion .....	64
5.1.6	Enrichment of Phosphorylated Peptides and Desalting .....	64
5.1.7	Nano LC-MS/MS .....	65
5.1.8	Database searching and Data Filtering .....	66
5.2	Phosphoproteome analysis of human mitotic spindle proteins by using low-specificity protease elastase .....	67
5.2.1	Materials .....	67
5.2.2	Cell Culture and Mitotic Spindle preparation .....	67
5.2.3	Gel Electrophoresis and In-Gel Digestion .....	67
5.2.4	Phosphorylated Peptide Enrichment and Sample Desalting .....	68
5.2.5	Nano LC-MS/MS .....	68
5.2.6	Data Processing and Analysis.....	69
<b>6</b>	<b>Abbreviations.....</b>	<b>70</b>
<b>7</b>	<b>References .....</b>	<b>73</b>
<b>8</b>	<b>Appendix .....</b>	<b>80</b>
	<b>Curriculum Vitae .....</b>	<b>89</b>

---

## Acknowledgement

Firstly, I would like to give my thanks to Prof. Dr. Erich A Nigg for providing me the opportunity to pursue my doctoral degree in his laboratory. I thank Erich for all the invaluable and helpful advice he has given me. I also thank Erich and Elena Nigg for the kind invitation to their family at the beginning of my stay in Germany. I appreciate that warm welcome.

I am thankful to Prof. Dr. Xinmiao Liang, who was the supervisor of my Masters studies. I thank Prof. Liang for giving me the chance of doing my postgraduate studies in his laboratory. I am especially grateful to Prof. Liang for encouraging me to do my Ph.D. at MPI and all the help he has given me in the past years.

I would like to express my deepest gratitude to Dr. Roman Körner, who is my supervisor of my Ph.D. studies. I thank Roman for everything he taught me at the early period of my Ph.D. with great patience, and also the guidance and all the helpful ideas for the projects I worked on. I appreciate the precious friendship with him in private time, and I will never forget the help he offered me for my personal life in Munich. I am very lucky to work with him and have him as my supervisor.

I am grateful to Ms. Alison Dalfovo for her great secretary assistance. I thank Alison for all the help and kind suggestions she gave me in the past years.

I am thankful to Dr. Anna Santamaria for the warm help at the early days of my stay in Munich and for her great job for the collaborating Plk1 project in this thesis. I have learned so much from our collaboration.

I want to thank Dr. Rainer Malik for his contribution in the elastase project. I also want to thank Kalyan Dulla for all the discussions and help. I thank Dr. Rene Lenobel for his kind help at the first year of my studies in the lab.

My special thanks to Albert Ries for his countless help and advices.

I thank my friends Eunice, Anja W., Stefan L. and Shin for all the great time I spent with.

I want to thank all the Nigg department mates for creating such a stimulating and friendly atmosphere in the lab. I have enjoyed and benefited a lot from this unforgettable time.

I want to thank Dr. Spielmann, Ms. Panglung, Mr. Fang and Mr. Shi for running the MPG-CAS Ph.D. promotion program.

---

I thank Prof. Dr. Matthias Mann and Dr. Jürgen Cox for early access to the MaxQuant software package.

I thank Xiaofang Yang for being with me and the wonderful time spent together.

Lastly, I would like to thank my parent Qijiang Wang and Shuxian Li for their endless love and continues support.



---

## Summary

Mitosis is the process during which a eukaryotic cell segregates the chromosomes into two identical daughter cells. The chromosomes are physically separated to the opposite poles of the cell by the mitotic spindle, which is a highly dynamic microtubule (MT)-based macromolecular machine. Spindle assembly begins early in mitosis and is completed when the bipolar attachment of microtubules to kinetochore (KT) pairs is achieved. Many mitotic spindle associated proteins are regulated by phosphorylation and dephosphorylation. A comprehensive phosphorylation site mapping of mitotic spindle proteins would help us to gain a better understanding of its functions and regulation in mitotic progression.

Mass spectrometry (MS) has become the most powerful and robust tool for protein identification and the study of their post-translational modifications. The development of MS hardware and data processing software made large scale quantitative proteomics and phosphoproteomics studies feasible.

In the project of quantitative identification of Polo-like kinase 1 (Plk1) specific phosphorylation sites on human mitotic spindle proteins, we present the first quantitative study of the Polo-like kinase 1 (Plk1) dependent phosphoproteome of the human mitotic spindle. Mammalian Plk1 controls centrosome maturation, spindle assembly and microtubule attachment to kinetochores. However, despite the essential and diverse functions of Plk1, only a limited number of substrates have been identified. In our study, we combine SILAC with Plk1 depletion or inactivation, followed by spindle isolation and mass spectrometry to compare the phosphoproteomes of mitotic spindles isolated from cells lacking Plk1 activity to spindles from cells with fully active kinase. We report the down-regulation of more than 700 phosphorylation sites upon Plk1 inactivation, including 360 sites on known spindle proteins, of which more than 100 were validated as direct Plk1 targets using *in vitro* phosphorylation of synthetic peptide arrays. This approach identifies many novel Plk1 substrates and suggests a broader phosphorylation consensus motif for this kinase than previously recognized. Furthermore, we analyze in a quantitative manner the consequences of Plk1 inactivation on the localization of its substrates to the spindle. We report the mis-localization of certain centrosomal and kinetochore proteins upon Plk1 inhibition and further analyze the kinetochore protein CENP-F as an example. Finally, we identify several Aurora A-dependent phosphorylation events that depend on Plk1, thus highlighting the role of Plk1 in the

---

regulation of Aurora A activity. Collectively, our data provides a rich source of information on Plk1-dependent phosphorylation, Plk1 docking to substrates, the influence of phosphorylation on protein localization, and the functional interaction between Plk1 and Aurora A on the early mitotic spindle.

In the project of phosphoproteome analysis of human mitotic spindle proteins by using the low-specificity protease elastase, we evaluated the use of elastase in a phosphoproteome study of human mitotic spindle associated proteins. Despite the usefulness and obvious advantages of trypsin in bottom-up proteomics experiments, it still has certain limitations in sequence coverage and cleavage efficiency. As an alternative to specific enzymes, low-specificity proteases have attracted our attention. Elastase became the enzyme of choice to supplement tryptic digestions, mainly because it generates medium sized peptides in the mass range between 500 Da and 1500 Da which are optimal for detection with current LC-MS hardware. We took advantage of the high mass accuracy of Orbitrap mass spectrometers and optimized the database search specificity by analyzing both elastase cleavage preferences and employing a dedicated two-step database search strategy. Through this approach, we have approximately doubled the number of detectable phosphorylation sites from elastase digested samples. Remarkably, phosphorylation sites detected by trypsin and elastase were highly complementary with an overlap of less than 10%. In total, we identified 1068 phosphorylation sites using trypsin and 467 phosphorylation sites using elastase. Approximately 30% of the phosphorylation sites were exclusively identified after digestion by elastase, demonstrating the value of this enzyme for phosphoproteome studies.

---

# **1 Introduction**

## **1.1 Cell cycle**

The cell cycle is an ordered series of events that take place in a cell and leads to its division and replication. Newly formed daughter cells can themselves grow and divide. In bacteria and yeast, each round of cell division is sufficient to produce a complete new organism. In higher eukaryotes, multiple and complex rounds of cell divisions are required to produce functional organism and to replace dead cells. The most fundamental characteristic of cell cycle through all living organism is to pass on exact genetic information to newly formed daughter cells. To accomplish this, each cell first duplicates its genetic material and then equally segregates it between two daughter cells to ensure each cell retains the exact same content of genetic information. In eukaryotes, this process is divided into four sequential phases: G1 (Gap phase1), S (Synthesis phase), G2 (Gap phase2) and M (Mitosis and cytokinesis). DNA synthesis and replication occur in S phase. In M phase, cell growth and protein production are greatly reduced, and duplicated chromosomes segregate equally to the daughter cells. Between S and M phases, there are two gap phases called G1 and G2. During G1 (after M phase and before S phase), cells grow in size, produce RNA and synthesize proteins. In G2 (after S phase and before M phase), cells continue to grow and prepare for entry into mitosis.

## **1.2 Mitosis**

Mitosis is relatively brief but a key event during the cell cycle. In M phase, equal segregation of genetic and cytoplasmic material takes place. Mitosis has five stages: prophase, prometaphase, metaphase, anaphase and telophase (figure 1). In prophase, the interphase chromatin condenses into chromosomes, which each comprises two chromatids held together by protein complex cohesion. The centrosomes, which have been duplicated during S phase, start moving to opposite poles and increase the nucleation of highly dynamic microtubules (MTs) to build mitotic spindles <sup>1,2</sup>. During prometaphase, the nuclear envelope breaks down. The microtubules emanating from opposite spindle poles are captured by kinetochores which are situated on the centromeres of both sister chromatids. Attached chromosomes are then moved to the equator of the cell in a process called chromosome congression. In metaphase,

all attached chromosomes align on an equatorial plate known as the metaphase plate in the middle of the cell. Once all chromosomes are properly aligned, the spindle assembly checkpoint, which is a surveillance mechanism that is active in prometaphase to prevent inaccurate separation of the sister chromatids, is satisfied. This leads to anaphase promoting complex mediated ubiquitin-dependent degradation of the separase inhibitor securin, resulting in the removal of sister chromatid cohesion and anaphase onset. In anaphase A, the sister chromatids separate and are pulled to the opposite poles by shortening of kinetochore microtubules. In anaphase B, the centrosomes move towards the cell cortex accelerating chromosome separation. Mitosis ends with telophase, in which the nuclear envelope reforms, the chromosomes arrive at the poles of spindles and start decondensing.

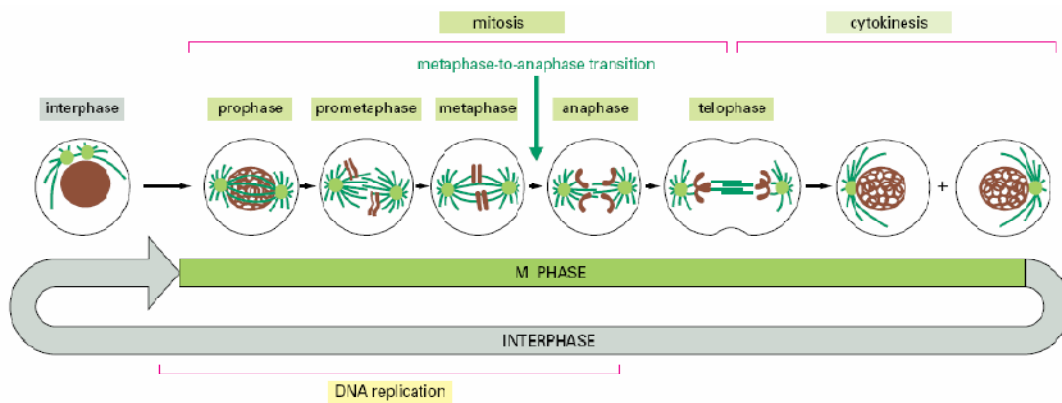


Figure 1. Schematic representation of different stages of mitosis and cytokinesis. The colours shown here are brown for DNA, light green for centrosomes and dark green for MTs. Image adapted from Alberts et al, Molecular Biology of the Cell, fourth edition, 2002.

Mitosis is followed by cytokinesis. In cytokinesis, the contraction of an actomyosin-based contractile ring (formed during late anaphase) takes place at the site of the spindle midzone, leading to furrow ingression. Cell division is completed by abscission and two daughter cells are formed, each with one nucleus and one centrosome<sup>3</sup>.

---

### 1.2.1 The mitotic spindle

Mitotic spindles pull the sister chromatids apart and move a complete set of chromosomes to each pole of the cell, where they are packaged into daughter nuclei. This ensures the faithful segregation of the genetic material during mitosis.

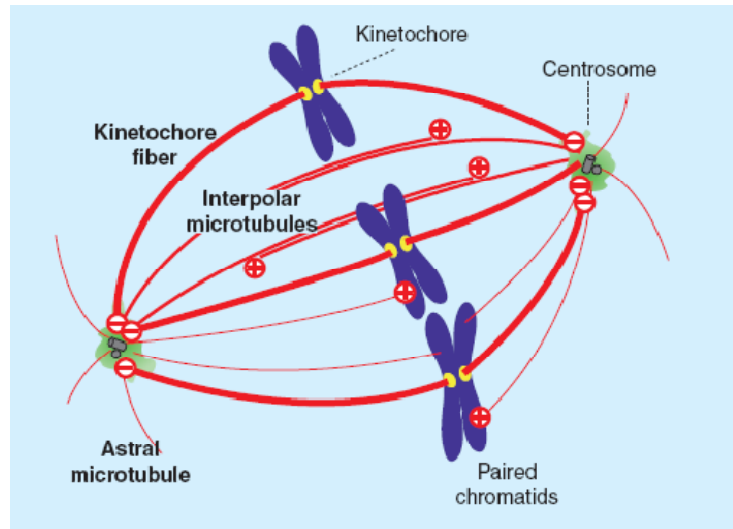


Figure 2. The features of the metaphase mitotic spindle. With their minus ends tethered at the spindle poles, microtubules extend either to the kinetochores of paired chromatids (kinetochore fibers), to the central spindle where they form an overlapping antiparallel array (inter polar microtubules), or away from the spindle towards the cell cortex (astral microtubules). Image adapted from Gadde S et al, Current Biology, 2004.

The microtubule array is biopolar, with minus ends tethered at the poles and plus ends attached to the chromosomes (figure 2). In most vertebrate cells, the spindle poles are the microtubule organizing centers, they are formed by the centrosomes. The mitotic spindle is a highly dynamic microtubule-based structure. Microtubules continuously grow and shrink, this behavior is regulated by motor proteins which can travel along microtubules, and many other microtubule associated proteins (MAPs) which have important roles in the assembly and stability of the microtubule array and the movement of chromosomes on the spindle.

### 1.2.2 Phosphorylation in mitotic progression

Mitotic progression is regulated by reversible protein phosphorylation (figure 3) and irreversible protein degradation. These two mechanisms are interdependent since the proteolytic machinery is regulated by phosphorylation and many mitotic kinases are down-regulated by protein degradation. Cyclin-dependent kinase 1 (Cdk1), Polo-like kinase 1 (Plk1) and Aurora kinases are the most prominent and well characterized mitotic regulatory kinases. Cdk1 protein level is constant throughout the cell cycle. Its catalytic subunit has to bind to a regulatory subunit called cyclin (Cyclin A or B) to become enzymatically active<sup>4, 5</sup>.

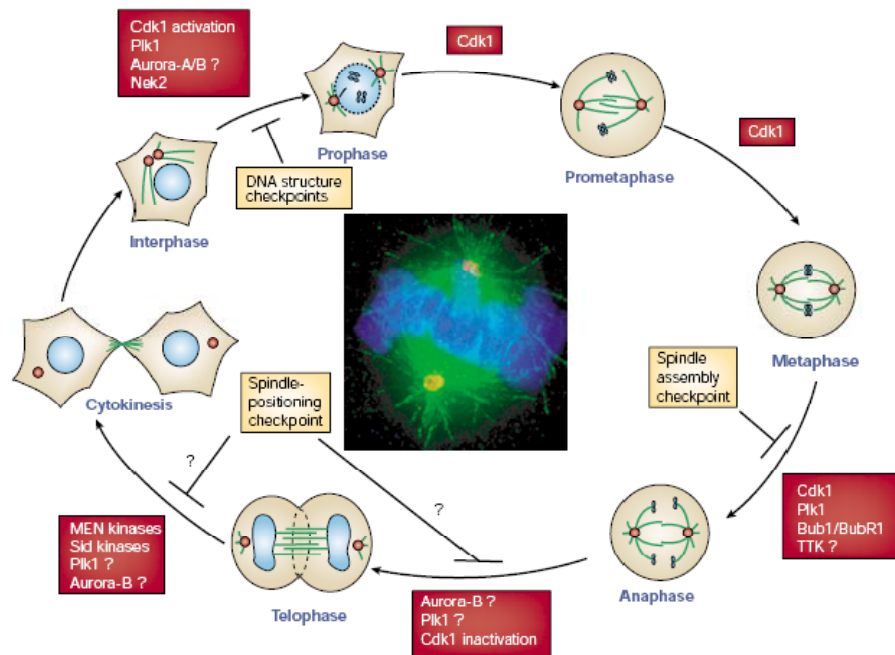


Figure 3. Kinases play important roles at different mitotic stages. Image adapted from Nigg E A, Nature Reviews, Molecular Cell Biology, 2001

Activated Cdk1 has many phosphorylation substrates in all mitotic stages, including proteins important for nuclear envelope breakdown, chromatin condensation, spindle assembly and centrosome separation. Furthermore, Cdk1/Cyclin B complexes are involved in the regulation of the anaphase-promoting complex/cyclosome (APC/C) which triggers the degradation of several mitotic regulators such as securin and Cyclin B itself<sup>6, 7</sup>.

---

Aurora kinases are also highly active during mitosis. Aurora A activity is closely associated with the centrosomes and spindle assembly from prophase to telophase<sup>8</sup>. It controls the timely mitotic entry by modulating nuclear envelope breakdown<sup>9</sup> and assists in centrosome maturation. Aurora B localizes to kinetochores and central spindles at different mitosis stages<sup>10</sup>. The kinase activity of Aurora B is required for spindle checkpoint signaling, central spindle formation and cytokinesis<sup>11</sup>.

### **1.3 Polo-like kinase 1 (Plk1)**

Polo-like kinase 1 is well-characterized and has been reported to be involved in all phases of mitosis<sup>12-14</sup>. It is highly conserved from yeast to human<sup>14</sup>. Plk1 N-terminal harbors the kinase catalytic domain whereas the C-terminal features a polo-box domain (PBD). The PBD is required for Plk1 substrate targeting and subcellular localization. It recognizes and binds to phosphorylated docking proteins, which results in a conformational change and liberates the kinase domain of Plk1. Activated Plk1 can then phosphorylate either the docking proteins or downstream targets<sup>15</sup>. The docking proteins are phosphorylated by the so-called priming kinases. Cdk1/Cyclin B, Calmodulin dependent kinase II<sup>16</sup>, MAP kinase Erk2<sup>17</sup> and Plk1 itself<sup>18</sup> are the most prominent priming kinases.

The localization of Plk1 is highly dynamic throughout mitosis. It accumulates at centrosomes during early mitosis to regulate centrosome maturation and separation<sup>14, 19</sup>. It also binds to the kinetochores to assure their correct attachments to the mitotic spindle<sup>20</sup>. Plk1 is further involved in the dissociation of chromosome cohesion through the removal of cohesin, which holds the two sister chromatids together<sup>21, 22</sup>. In addition, Plk1 has been shown to regulate the function of the anaphase-promoting complex/cyclosome (APC/C), which is essential for mitotic exit<sup>7, 23</sup>. At metaphase-anaphase transition, Plk1 relocates to the spindle midzone and plays a key role in cytokinesis<sup>14</sup>.

## **1.4 Mass spectrometry based proteomics**

### **1.4.1 General workflow of MS-based proteomics**

Modern proteomic approaches include mass spectrometry, protein microarray technologies, large scale two-hybrid analysis, high-throughput protein production and crystallization<sup>24</sup>.

Mass spectrometry measures the mass-to-charge ratios ( $m/z$ ) of ionized molecules in the gasphase and can be used to determine the elemental composition provided that the mass accuracy of the instrument is sufficiently high. In proteomics, the measured  $m/z$  ratios and fragmentation patterns reflect the amino acids composition and modifications of analyzed peptides or proteins. In the past decade, mass spectrometry has been developed further successfully and become a popular technique in a wide variety of biological studies, especially for protein identification, the determination of protein post-translational modifications, and the study of protein-protein interactions.

There are two general approaches in MS: “bottom-up” and “top-down”. In the “bottom-up” strategy, proteins are enzymatically digested into peptides using proteases. The collection of peptide products is then analyzed by the mass spectrometers. In the “top-down” strategy, intact proteins are directly ionized and analyzed by mass spectrometry.

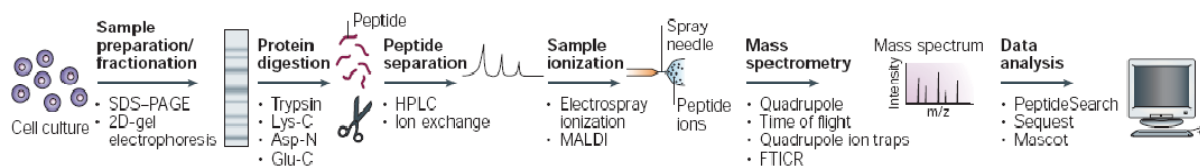


Figure 4. General workflow of bottom-up MS-based proteomics experiment. Illustration adapted from Steen H et al., Nature Reviews on Molecular Cell Biology, 2004.

Due to the advantages in detection sensitivity, sample preparation, ionization and fragmentation efficiency, “bottom-up” is by far the most widespread and successful method in MS-based proteomics. A general workflow is depicted in figure 4. The proteins to be analyzed are isolated or purified from cells or tissues. The purification of proteins often includes gel electrophoresis as the last step. The proteins are then separated in the gel according to their molecular weight or isoelectric point. The obtained gel lane is cut into several slices, and then in-gel digestion of proteins is performed. Various proteases or chemicals are available for this step. For the most popular enzyme trypsin, most peptides consist of 6-20 amino acids, which is optimal for detection and identification by mass spectrometry. Prior to MS analysis, the peptide mixtures normally undergo single or multiple dimensions of on-line or off-line separation. Peptides are then ionized, their  $m/z$  values determined by mass spectrometry, and further structural information generated by low energy



---

excitation (MS/MS fragmentation). Finally, obtained mass spectra are searched against protein databases via computer algorithms, returned search results identify peptide sequences and post-translational modifications, thus allowing the identification of proteins and the characterization of their modifications<sup>25, 26</sup>.

In the “top-down” approach, intact proteins are directly ionized and measured in mass spectrometers without previous proteolytic digestion. MS/MS fragmentation is also performed by various activation methods. As an alternative approach, “top-down” obtains better sequence coverage for protein identifications, and it can in principle provide complete information on post-translational modifications. However, the application is usually limited to small proteins with molecular weight up to 30 kDa and the analytical sensitivity is low compared to the bottom-up strategy. Furthermore, insufficient automated hardware and software tools within the field make high throughput and routine laboratory use very difficult<sup>27, 28</sup>.

#### **1.4.2 MS instrumentation**

A mass spectrometer measures the mass to charge ratio in the gas phase of ionized analytes. A complete MS system consists of an ion source, a mass analyzer and a detector. Typically, the peptide or protein ions are generated in the ion source by one of two soft ionization methods: electrospray ionization (ESI)<sup>29</sup> and matrix-assisted laser desorption/ionization (MALDI)<sup>30</sup>.

In the field of proteomics, time-of-flight (TOF), ion trap, quadrupole, Fourier transform ion cyclotron resonance (FTICR)-MS and Orbitrap<sup>31</sup> analysers are the most frequently used instruments<sup>25</sup>.

##### **1.4.2.1 Electrospray ionization (ESI)**

Dole and colleagues introduced the idea of using electrospray dispersion to produce gas phase ions from solutions in 1968. Fenn and coworkers applied this technique to biomolecules and developed the modern day ESI<sup>29</sup>. Further improvements were made by Mann and coworkers in the development of nanoelectrospray ionization (nanoESI), which is compatible with nano-flow rates for minute amount of samples thus strongly increasing the detection sensitivity<sup>32</sup>.

Electrospray generates gas phase ions with high efficiency and is extremely gentle, enabling the study of a wide range of polar biomolecules by mass spectrometry<sup>33</sup>. Since large biomolecules become usually multiply charged upon ESI, these high-mass ions can still be detected by instruments with a relatively low  $m/z$  detection range (frequently 50-2000 Da). Furthermore, since samples are applied in solution, ESI is compatible with many types of separation techniques such as liquid chromatography and capillary electrophoresis.

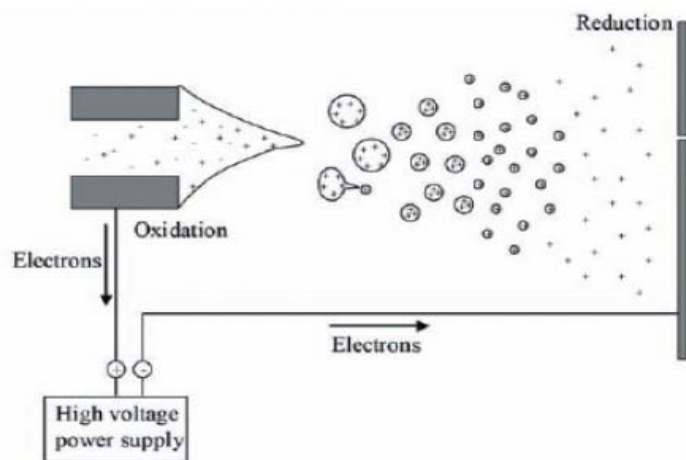


Figure 5. Schematic presentation of the electrospray ionization process. From Nielsen et al<sup>34</sup>.

In the classical ESI process, the sample dissolved in a polar solvent such as methanol/water flows through a fine capillary or a needle with a thin, sharp, conductive tip to which positive or negative high voltage of several kilovolts is applied. The capillary is placed towards the ESI source, of which the orifice is warmed up to around 150-200 degrees to aid desolvation. The surface of the emerging liquid gets charged under such a high voltage field and is being attracted toward the counter-charged electrode, whereas the surface tension acts in an opposite direction. As a result of these two forces, a pointed liquid cone forms at the tip of the capillary if the electric field is sufficiently strong. If the strength of the electric field increases further, a thin, charged liquid jet emerges from the cone, which quickly subdivides into smaller charged droplets to minimize its surface area<sup>29</sup>. The droplets quickly shrink by evaporation and become unstable as the charge density on the surface increases. This leads the droplets to blow apart into a cloud of tiny, highly charged droplets which move towards the orifice (figure 5). As the droplets move, more and more solvent evaporates and eventually

---

charged solute molecules are liberated into the gas phase of which the  $m/z$  ratio can be determined by the mass analyzer<sup>35</sup>.

#### 1.4.2.2 Linear quadrupole ion trap (LTQ)

The linear quadrupole ion trap can be used in various applications such as protein identification and quantitative differential expression analysis, biomarker studies and PTM identification, metabolite identification and quantification, drug screening etc. The basic construction of a commercial LTQ contains two pairs of orthogonal hyperbolic rods, each segmented into three sections, with the length of the central section is about three times larger than the two end sections. There are two small slits along the rods at the Z axis to allow ions to be scanned out of the trap. Two detectors are placed symmetrically towards the slits to record the radially ejected ions. Direct current (DC) voltages are applied to three sections allowing trapping of the ions along the axis in the central part (figure 6).

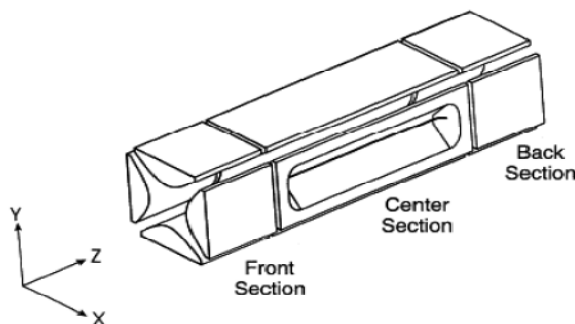


Figure 6. Basic structure of the two-dimensional linear ion trap. Image adapted from Schwartz J C et al., Journal of the American Society for Mass Spectrometry, 2002.

In a linear ion trap, injected ions are confined radially ( $x$ - and  $y$ - directions) by a two dimensional radio frequency (RF) field, and axially by stopping potentials applied to the end electrodes. The two pairs of opposing rods receive the same voltages, but neighboring rods have opposite voltages but with the same amplitude. Along the  $z$ - axis, different DC voltages are applied to the three sections to generate a deep electric potential well in the central section for ion trapping in the axial direction. In addition, a supplemental alternating current (AC) voltage across the  $x$  rods is applied to assist ion isolation, activation and ejection<sup>36</sup>.

---

Once in the trap, the ions are cooled by collision with an inert gas and fly along the  $z$ -axis between the end electrodes, while simultaneously oscillating in the  $xy$  plane owing to the application of the RF potential on the rods<sup>37</sup>. Importantly, the ion oscillation frequencies in the radial and axial directions depend on the mass-to-charge ratios, so ions can be specifically fragmented or ejected by applying the corresponding RF resonance frequencies to the trap. The analyzer is scanned by applying RF at a fixed frequency but with variable amplitudes. For analyzing trapped ions, the mass-selective instability principle is employed where the primary RF voltage increases at a constant rate. Under the two-dimensional electric field the ions are pushed and pulled in the  $x$ - and  $y$ -directions as they travel along the  $z$ -axis, this process filters out ions of all  $m/z$  values except for those of interest. The ramping of the RF potential amplitude causes all ions to eventually exceed the resonance ejection limit and be ejected through the exit slits in the  $x$ -rods to the detectors placed right beyond the slots, to provide maximum sensitivity<sup>36, 38</sup>. In this way, ions with specific mass to charge ratio can be isolated and ejected from the trap.

The LTQ can perform MS analysis alone or be connected with other analyzers as a hybrid mass spectrometer. The mass accuracy of LTQ can be up to 50 ppm and resolution close to 1000 ( $m/z$  400, at an ion target value of 5000)<sup>39</sup>. Compared to a conventional 3D ion trap, LTQ has an increased ion storage capacity (15 times higher), improved ion trapping efficiency (up to 70%), high detection efficiency (up to 100%) and better scan rate (3 times faster)<sup>36</sup>.

#### **1.4.2.3 Orbitrap**

The Orbitrap is the most recent type of ion trap mass analyzer. It uses an electrostatic field to trap ions instead of a RF or magnet field. The Orbitrap consists of an outer barrel-like electrode and a coaxial inner spindle-like electrode (figure 7). The outer electrode is split at the middle allowing ions to be injected into the trap. The two electrodes generate a static, quadro-logarithmic electrostatic field. The ions enter the electric field between the two electrodes in a tangential way, and they are trapped because of their electrostatic attraction to the inner electrode is balanced by centrifugal forces. The ions cycle around in circles and also move back and forward along the axis.

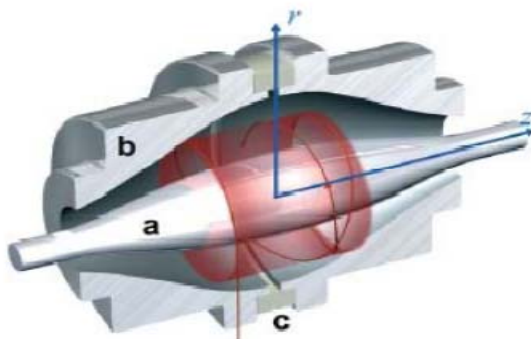


Figure 7. A cut-away model of the orbitrap mass analyzer. From Scigelova M. et al., Proteomics, 2006.

The frequency of the oscillations of the ions is proportional to the square root of the mass-to-charge ratio but is completely independent of the energy and the spatial spread of the ions<sup>40</sup>. By sensing the ion oscillation frequencies using Fourier transformation, the trap can be used as a mass analyzer. The Orbitrap can provide very high mass accuracy (1-2 ppm) and mass resolution (up to 200,000)<sup>41, 42</sup>.

#### 1.4.2.4 LTQ-Orbitrap Hybrid Mass Spectrometer

The LTQ-Orbitrap is widely used in MS-based proteomics studies. Produced by Thermo Fisher, the LTQ-Orbitrap combines the advantages of fast scanning rate and high sensitivity of the LTQ and the advantages of high mass accuracy, high resolution and high dynamic range of the Orbitrap<sup>43</sup>. MS related experiments in this thesis were all performed on a LTQ-Orbitrap.

The Linear ion trap is placed in front of the Orbitrap (figure 8), which detects MS and MS<sup>n</sup> with very high sensitivity but relatively low resolution and mass accuracy. Ions injected from the ESI source at first scanned in the LTQ to determine the ion current within the mass range of interest. This allows storing a user-defined number of ions<sup>43</sup>. Accumulated ions can be transferred into an RF-only quadrupole called C-trap. The C-trap accumulates and stores the ions. The additional storage improves the analytical capabilities considerably<sup>31</sup>. After leaving the C-trap, the ions are accelerated to high kinetic energies and pass through a small entrance aperture and enter the Orbitrap tangentially<sup>42</sup>. Ions are captured in the Orbitrap by rapidly increasing the electric field. Signals from each of the Orbitrap outer electrodes are amplified by a differential amplifier and transformed into a frequency spectrum by fast

---

Fourier transformation. The frequency spectrum is converted into a mass spectrum and processed with Xcalibur software<sup>43</sup>.

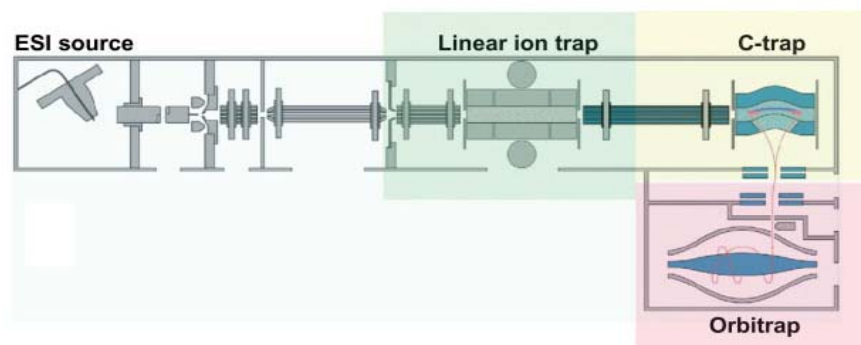


Figure 8. A schematic of the LTQ Orbitrap. From Scigelova M et al., Proteomics, 2006.

### 1.4.3 MS-based Quantitative Proteomics

Many biological processes are triggered and regulated not only by turning protein expression on and off. Protein expression levels and their dynamic changes are also essential for various cellular activities. Therefore, it is not always sufficient to only identify proteins to decipher different biological phenomena. Precise quantification of protein expression and their modifications are of great demand for biological research.

Traditionally, quantitative proteomics utilizes dyes, fluorophores or radioactivity. In MS-based proteomics quantification, established methods can be classified as metabolic labeling, chemical and enzymatic labeling, label free strategies and absolute quantitation<sup>44, 45</sup>. General MS-based quantification workflows are illustrated in figure 9.

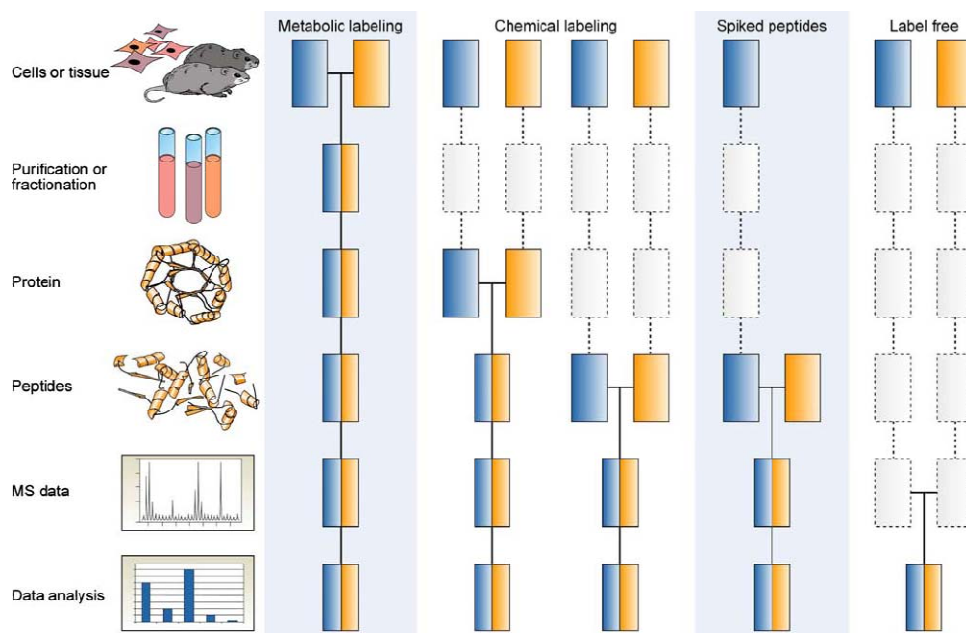


Figure 9. General MS-based quantitative proteomics workflows. Boxes in blue and yellow represent two experimental conditions. Horizontal lines indicate when samples are combined. Dashed lines indicate points at which experimental variation and thus quantification errors can occur. Adapted from Bantscheff M et al., *Analytical and bioanalytical chemistry*, 2007.

Metabolic labeling introduces a stable isotope signature into proteins at the early stages of protein expression. Labeled proteins with stable isotopes thus have shifted  $m/z$  values in mass spectra compared to their natural, non-labeled counterparts but are otherwise identical in all respects. Metabolic labeling allows mixing of labeled/unlabeled samples at the level of intact cells, which avoids accumulating systematic errors and inaccurate quantitation results. Stable isotope labeling by amino acid in cell culture (SILAC) is currently the most widely used method in metabolic labeling (detailed description in 1.4.3.1).

Labeling of proteins or peptides after biosynthesis is often performed by in-vitro enzymatic or chemical labeling. Enzymatic labeling incorporates  $^{18}\text{O}$  molecules onto the C-termini of peptides during protein digestion. This results for most proteases in a 4 Da mass difference between labeled and unlabeled peptides<sup>46, 47</sup>. Chemical labeling can be performed both at intact protein or peptide level. Most often used methods include isotope-coded affinity

---

tags (ICAT)<sup>48</sup>, isotope tags for relative and absolute quantification (iTRAQ)<sup>49, 50</sup>, isotope-coded protein label (ICPL)<sup>51</sup> and HysTag<sup>52</sup>.

Label free quantification is an emerging method: it compares the mass spectrometric signal intensity of peptides precursor ions belonging to a protein or the numbers of fragment spectra identifying a given protein between two biological samples<sup>45</sup>. The method demands very precise and accurate performance of sample-preparation and analytical systems<sup>53</sup>. Absolute quantitation can be achieved by spiking known amount of a standard reference into samples to be analyzed. Several methods are available for different applications, such as culture-derived isotope tags (CDITs)<sup>54</sup> and absolute quantitation (AQUA)<sup>55</sup>.

#### **1.4.3.1 Stable Isotope labeling by amino acid in cell culture (SILAC)**

Using the SILAC method, based on labeling differences, up to three biological conditions can be directly compared in a single experiment. One population is fed with growth medium containing normal amino acids whereas the other one or two are fed with media supplemented with amino acids labeled with certain heavy isotopes. SILAC in combination with tryptic digestion often uses labeled arginine and lysine since trypsin predominantly cleaves peptide chains at the carboxyl side of lysine and arginine, thus allowing the relative quantitation of virtually all detected peptides. The most commonly used labeled forms are  $^{13}\text{C}_6\text{-Arg}$ ,  $^{13}\text{C}_6^{15}\text{N}_4\text{-Arg}$ ,  $^2\text{H}_4\text{-Lys}$  and  $^{13}\text{C}_6^{15}\text{N}_2\text{-Lys}$ .<sup>44, 56, 57</sup>

Taking  $^{13}\text{C}_6^{15}\text{N}_2\text{-Lys}$  and  $^{13}\text{C}_6^{15}\text{N}_4\text{-Arg}$  labeling as an example, the light form of cells are grown with normal amino acids whereas the heavy labeled cells are fed with  $^{13}\text{C}_6^{15}\text{N}_2\text{-Lys}$  and  $^{13}\text{C}_6^{15}\text{N}_4\text{-Arg}$  containing medium (figure 10). This results in 8 Da and 10 Da mass shifts for each lysine and arginine in the detected peptides, respectively. Light and heavy peptides co-elute from liquid chromatography column (heavier peptides containing deuterium atoms elute earlier than their unlabeled counterparts<sup>58, 59</sup>). The quantitation is done by comparing the extracted ion current (XIC) of two peptides, which is proportional to the peptide abundance (figure 11).

In a cell-line dependent manner, the SILAC accuracy can be compromised by metabolic conversion of arginine to proline. This results in the formation of multiple satellite peak clusters for all proline containing peptides in the labeled states and thus hampers accurate quantitation<sup>60, 61</sup>. Solutions include empirically reducing the amount of arginine used in the



labeling media, mathematical corrections for all proline-containing peptides<sup>62</sup>, providing extra proline in media<sup>63</sup>, or internal correction strategies<sup>61</sup>.

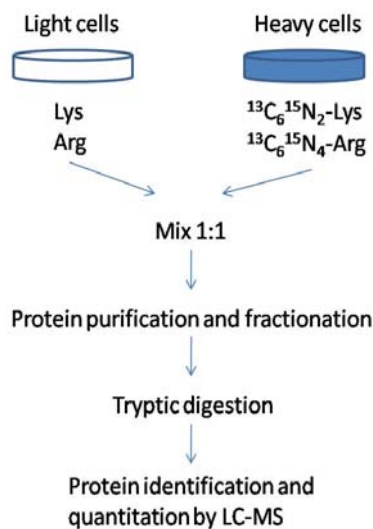


Figure 10. Outline of a double SILAC experimental workflow

SILAC has proven to be a powerful quantitative method in many biological research fields, such as cell signaling<sup>64</sup>, protein-protein interaction<sup>65</sup>, protein phosphorylation<sup>66</sup> and gene function analyses<sup>67</sup>. The limitation of SILAC is that it is not feasible for *in vivo* tissue sample applications of large mammals.

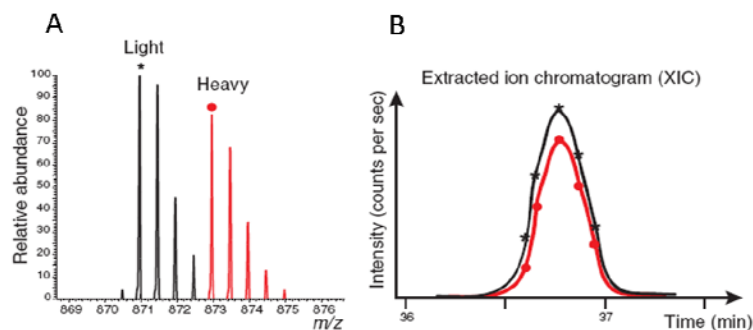


Figure 11. (A) MS spectra of isotope clusters of a SILAC peptide pair. (B) Extracted ion chromatogram of co-eluted SILAC peptide pair from HPLC column. Adapted from Ong S E et al., Nature Chemical Biology, 2005

---

#### 1.4.4 Phosphoproteomics

Reversible phosphorylation of proteins is among the most widespread post-translational modifications in eukaryotic cells. It regulates many biological processes such as signal transduction, cell division, gene expression, molecular recognition. It has been estimated that about 30% of the proteome is affected by phosphorylation at a given time point <sup>64, 68, 69</sup>. Therefore, in-depth studies on protein phosphorylation are of great importance to help us to better understand diverse biological phenomena. However, complete phosphorylation profiling still remains a challenge due to its frequently low stoichiometry. Often only a few percent of the entire protein amount is present in the phosphorylated form <sup>70</sup>.

Traditionally, radioactive <sup>32</sup>P-labeling coupled with 2D gel-electrophoresis and Edman sequencing was widely used to characterize phosphorylation on proteins. Western blotting using phospho-antibodies also has high sensitivity and specificity for detecting specific phosphorylation sites.

Nowadays, mass spectrometry has become the most powerful and popular method in the field of phosphoproteomics. Generally, target proteins are in-solution or in-gel digested before phosphopeptides are isolated from highly complex peptide mixtures and subjected to mass spectrometry analysis. A comprehensive phosphorylation profiling should include the identification of phosphorylated peptides, and the exact localization of the phosphorylated amino acids within the peptide sequences.

The ability of performing phosphoproteomics is highly dependent on effectively isolating phosphorylated peptides from non-phosphorylated species. There are many developed enrichment strategies such as TiO<sub>2</sub>, strong cation exchange chromatography (SCX), immobilized metal affinity chromatography (IMAC), ZrO<sub>2</sub> <sup>71</sup> etc. Projects described in this thesis are based on TiO<sub>2</sub> and SCX as phosphorylated peptide enrichment methods.

##### 1.4.4.1 TiO<sub>2</sub> phosphorylated peptide enrichment

Titanium dioxide (TiO<sub>2</sub>) beads have specific affinity to organic phosphates in solution. Additionally, their high mechanical, chemical and thermal stability facilitates their use to selectively enrich phosphorylated peptides from complex mixtures. Very high enrichment efficiencies (> 90%) in low-complexity samples have been reported <sup>72</sup>. In large-scale phosphoproteomics, unspecific binding of non-phosphorylated peptides and acidic peptides to

---

TiO<sub>2</sub> beads become severe. Loading peptide mixture to TiO<sub>2</sub> in organic solvent with strong acid and “non-phosphorylated peptide excluder” as a modifier can significantly increase enrichment selectivity. The “non-phosphorylated peptide excluders” should be able to compete with the binding of non-phosphorylated peptides to TiO<sub>2</sub> and 2,5-dihydroxy benzoic acid (DHB), phthalic acid, glycolic acid, and lactic acid have proven to be efficient additives<sup>73, 74</sup>.

#### **1.4.4.2 SCX phosphorylated peptide enrichment**

SCX often uses silica-based materials with bonded coating of hydrophilic, anionic polymer to selectively separate phosphorylated peptides from non-phosphorylated ones based on their charges. Between pH 2.7-3.0, which is used in the SCX loading buffer, most tryptic peptides carry at least two positive charges<sup>75</sup>. Because of added negatively charged phosphate groups, phosphorylated peptides will have reduced numbers of positive net charges, and therefore exhibit weak or no binding to the SCX separation column, while non-phosphorylated peptides bind usually stronger. However, the interfering effect from acidic peptides has not yet been overcome in this strategy. SCX is a robust phosphorylated peptide enrichment method but with modest enrichment efficiency<sup>75, 76</sup>.

---

## **2 Aims of this study**

The mitotic spindle is essential for accurate chromosome segregation during cell division. Reversible protein phosphorylation is a key regulatory mechanism involved in the formation and regulation of spindles. Insight into the phosphorylation network of spindle proteins is of great interest for us to gain a better understanding of the regulation and function of the mitotic spindle.

The Plk1 substrate identification part of this thesis aims at identifying the specific phosphorylation substrates of a key mitosis regulator, Polo-like kinase 1, on human mitotic spindles, and to explore the role of this kinase in the phosphorylation network of the spindle associated proteins.

The evaluation of the use of elastase project aims at combining the use of the specific enzyme trypsin and the low specificity enzyme elastase to achieve a better sequence coverage for the phosphoproteomics study of human mitotic spindles.

---

## 3 Results

### 3.1 Quantitative identification of Polo-like kinase 1 (Plk1) specific phosphorylation sites on human mitotic spindle proteins

#### 3.1.1 Introduction

In mitosis, a eukaryotic cell separates the chromosomes into two identical sets in two daughter cells. A series of tasks such as mitotic entry, spindle assembly, chromosome segregation, and cytokinesis, must be carefully coordinated to ensure the error-free distribution of chromosomes into the newly forming daughter cells. The physical separation of the chromosomes to opposite poles of the cell is driven by the mitotic spindle, a highly dynamic microtubule (MT)-based, protein-associated macromolecular machine. Spindle assembly begins early in mitosis and is completed when bipolar attachment of microtubules to kinetochore pairs occurs<sup>77, 78</sup>. Polo-like kinase 1 (Plk1), a serine/threonine-specific kinase first identified in *Drosophila*<sup>79</sup>, is a key regulator of this essential mitotic process<sup>13, 14</sup>.

Plk1 has diverse functions and its localization during mitosis is dynamic. Plk1 first associates with centrosomes in prophase, then associates with the spindle poles and kinetochores (KTs) in prometaphase and metaphase, is recruited to the central spindle in anaphase, and finally accumulates at the midbody during telophase. Proteomics studies using oriented peptide libraries suggest that the two polo-box domains (PBDs) of Plk1 are crucial for its localization to cellular structures<sup>15</sup>. These domains bind to specific amino acid sequence motives which were phosphorylated by priming kinases or self-primed by Plk1, thus providing an efficient mechanism to regulate localization and substrate selectivity in time and space.

Plk1 has fundamental roles in early mitosis in centrosome maturation, kinetochore-microtubule attachments and spindle assembly, but to date only a limited number of targeted proteins and phosphorylation sites on these substrates have been conclusively identified<sup>12, 80</sup>. Moreover, the current understanding of the functions of these site-specific Plk1 phosphorylations is far from being complete, in part due to low abundance of the substrates, technical difficulties in determining *in vivo* phosphorylation sites, and the requirement of Plk1 localization for substrate recognition in some of the cases<sup>81</sup>.

---

In recent years, mass spectrometry (MS) has become a key technology for mapping protein phosphorylation sites. The optimization of methods for enriching and detecting phosphorylated peptides have shown great promise for large scale studies<sup>82</sup>. Previously, we explored the human mitotic spindle by MS and successfully identified a large number of novel spindle proteins and phosphorylation sites<sup>83, 84</sup>. The development of quantitative methods to monitor *in vivo* phosphorylation changes in complex samples<sup>64, 85</sup>, represents a unique opportunity to address the role of individual kinases in spindle function.

In this work, we performed quantitative analysis for human mitotic spindle proteins using SILAC in combination with a phosphopeptide enrichment approach. Aimed at improving our current experimental coverage of Plk1 substrates and gaining insight into the direct and indirect effects of Plk1, we have analyzed how reduced Plk1 activity affects the phosphoproteome of purified human mitotic spindles. We report the combination of the selective inhibition of Plk1 using either an inducible Plk1 depletion shRNA cell line or the small molecule inhibitor TAL<sup>86</sup>. The experimental strategy was applied to analyze Plk1 phosphorylation on spindle proteins at an early mitosis stage. Furthermore, we set out to validate the phosphorylation sites found to be down-regulated in the absence of Plk1 activity using *in vitro* phosphorylation assays on peptide spots arrays.

Our approach identified novel Plk1 substrates and validated more than 100 *in vivo* direct Plk1 phosphorylation sites. By probing a large set of candidate Plk1 peptides, we were able to broaden the consensus phosphorylation motif of this kinase. Moreover, we showed that the Plk1-dependent localization of some of the identified protein substrates to the centrosomes and kinetochores. In addition, we could confirm that the Aurora A signaling pathway is downstream of Plk1.

I emphasize that major biological experiments (cell culture, spindle isolation, IF, western blotting etc.) in the Plk1 substrate identification project of this thesis were performed by Dr. Anna Santamaria (MPIB). The results are shown here in order to confer biological relevance to the proteomics analysis that was carried out by myself. The *in vitro* kinase assay experiments were performed by Dr. Sabine Elowe (MPIB).

### 3.1.2 Experimental design for the comparison of the spindle phosphoproteome between cells in the presence or absence of Plk1

To gain insight of the phosphoproteomes of human mitotic spindles from Plk1-inhibited cells with spindles from control cells with fully active Plk1, we developed a quantitative phosphoproteomics strategy that combines SILAC with selective enrichment of spindle-associated proteins by isolating human mitotic spindles (figure 12).

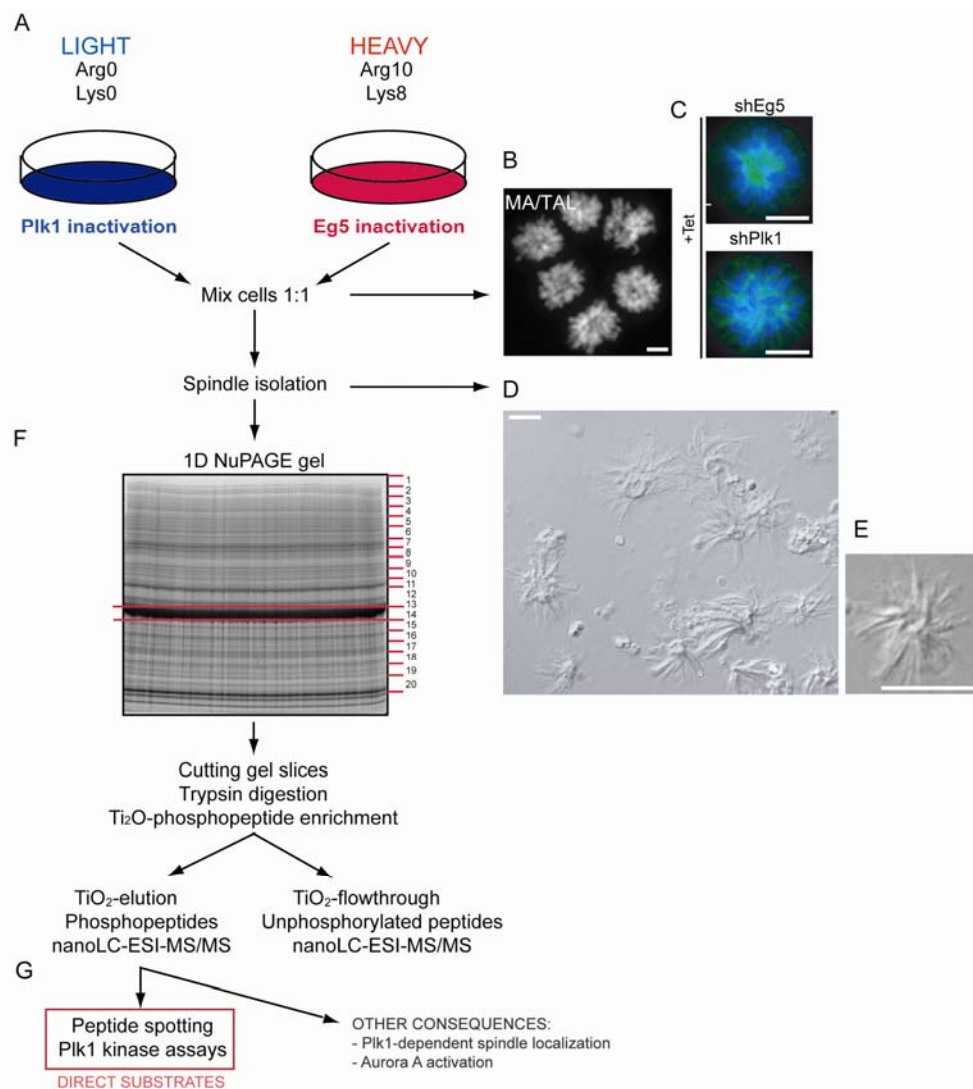


Figure 12. Schematic of the experimental design for the SILAC experiments. (A) HeLa S3 cells grown in the presence of normal or stable isotope-labeled arginine and lysine were synchronized by Tet induction and MA/TAL treatment, respectively. (B, C) DNA and/or spindle morphology of MA and TAL-treated cells (DAPI staining is shown) and shEg5 and shPlk1-induced cells ( $\alpha$ -Tubulin is shown in green and DNA was visualized using DAPI (blue)) prior to mixture and spindle isolation. (D, E) DIC pictures of isolated mitotic spindles. Scale bar, 10  $\mu$ m. (F) Analytical strategy to map phosphorylation sites in spindle proteins. (G) Validation of direct Plk1 phosphorylation sites by peptide spotting and *in vitro* Plk1 kinase assays.

To synchronize control cells to the same mitotic stage and importantly, obtain similar microtubule arrays as in Plk1-inactivated cells, we interfered with the function of the kinesin motor protein Eg5. Similar to Plk1 inactivation<sup>87</sup>, Eg5 inhibition results in a prometaphase arrest with a monoastral microtubule array surrounded by a ring of chromosomes<sup>88</sup>. Importantly, Plk1 localization and kinase activity were retained in Eg5-inactivated cells, supporting the use of Eg5 as control (figure 13)<sup>86</sup>.

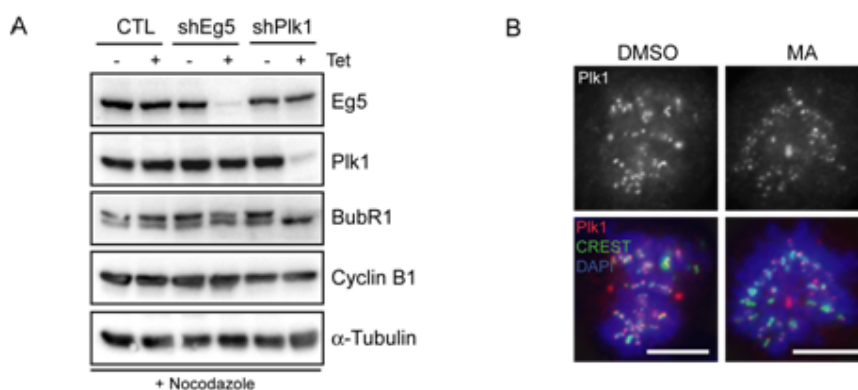


Figure 13. (A) Lysates from CTL, shEg5 and shPlk1 cells treated for 36 h with Tet or left for 36 h in Tet-free medium, treated for the last 12 h with nocodazole, to obtain cells in the same mitotic stage. Membranes were probed for Eg5, Plk1, BubR1, Cyclin B1 and  $\alpha$ -Tubulin, as loading control. (B) DMSO, MA-treated cells were fixed and stained for the indicated antibodies. Plk1 is shown in red and CREST in green. DNA was visualized using DAPI (blue). Scale bar is 10  $\mu$ m (experiment performed by Dr. Anna Santamaria).



---

We employed two different experimental strategies. First, we interfered with Plk1 function by lowering Plk1 protein levels. To do this systematically, we generated tetracycline (Tet) inducible stable HeLa S3 cell lines for shRNA-mediated knockdown of Plk1 and Eg5, as control (shPlk1 and shEg5 cell line, respectively). After shRNA induction, Eg5 and Plk1 protein levels were efficiently reduced when compared to non-induced or control cells (a stable cell line generated with an empty vector driving non specific shRNA expression; CTL cells) (figure 13A and 14A). Consistently, depletion of Plk1 but not Eg5 led to loss of BubR1 hyperphosphorylation (figure 13A), which was shown to be Plk1 dependent <sup>89-91</sup>. Flow cytometry (FACS) analysis of Tet-induced shEg5 or shPlk1 cells revealed an accumulation of cells with a 4N DNA content, indicative of a G2/M arrest (data not shown). Concomitantly, a striking increase in the mitotic index could be observed, in agreement with the predominant phenotype seen upon interference with Plk1 or Eg5 <sup>87,88</sup>.

Analysis of the protein levels of Plk1 and Eg5 upon Tet induction by MS in the appropriate shRNA cell line provided an independent assessment for depletion efficiency in our shRNA cell lines (figure 14B and C).

As a second complementary strategy, a chemical biology based approach using the specific small molecule Plk1 inhibitor TAL <sup>86</sup> was employed to directly interfere with the enzymatic activity (rather than protein levels) of Plk1. Monastrol (MA), a small molecule inhibitor of Eg5 <sup>92</sup>, was used to treat control cells (figure 12B and 13B).

Two independent experiments were performed for each Plk1 inactivation strategy which served as biological replicates.

To enable accurate quantitation of the changes of phosphorylated peptide abundances by MS, cells were labeled by growing them in medium containing either normal arginine and lysine (Arg0/Lys0) or the heavy isotopic variants <sup>13</sup>C<sub>6</sub>-<sup>15</sup>N<sub>4</sub>-arginine and <sup>13</sup>C<sub>6</sub>-<sup>15</sup>N<sub>2</sub>-lysine (Arg10/Lys8) <sup>44</sup>. In three of the four experiments shPlk1 or TAL-treated cells were grown in SILAC medium supplemented with light amino acids, whereas shEg5 or MA-treated cells were grown in SILAC medium supplemented with heavy amino acids, whereas an additional experiment was performed under reverse labeling condition (experiment Inhibitor-1).

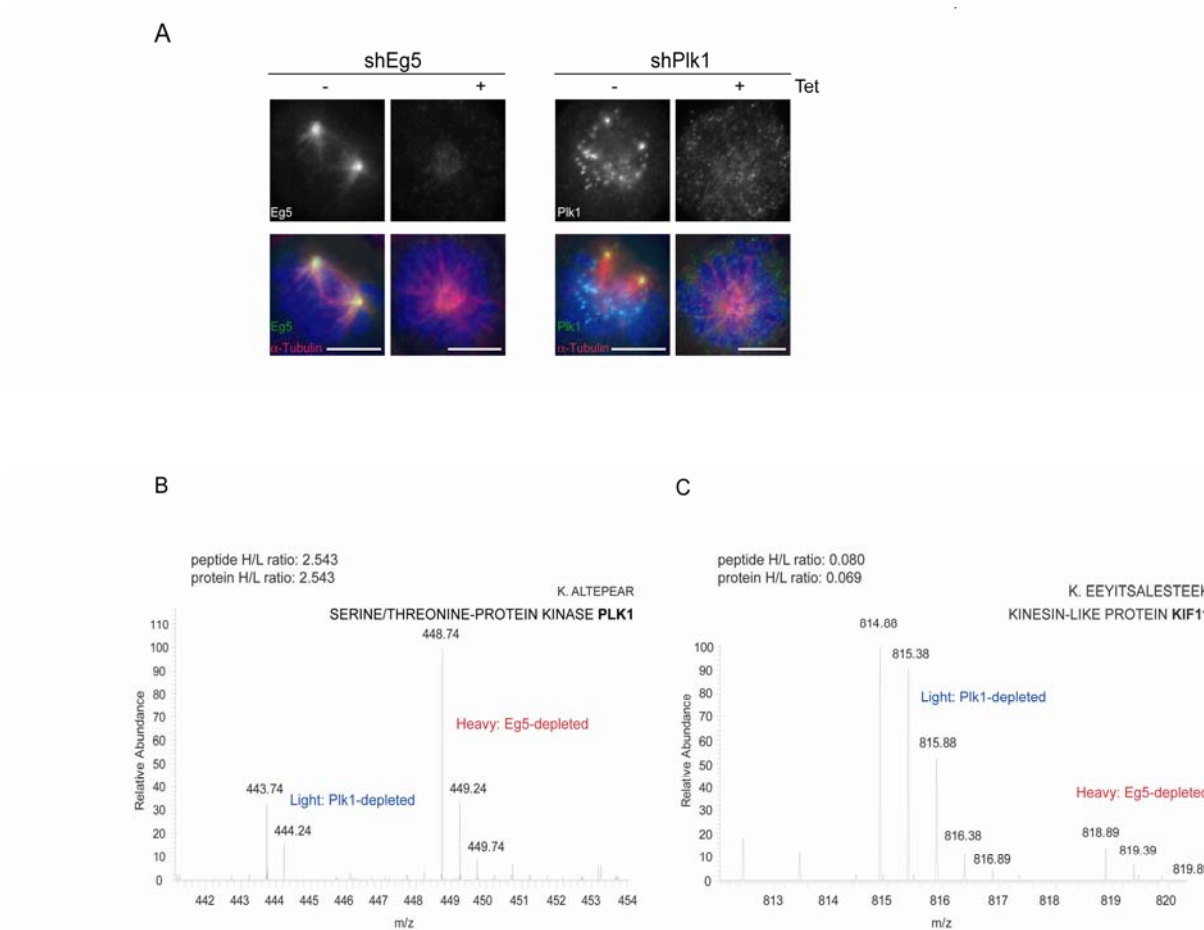


Figure 14. (A) shEg5 and shPlk1 cells treated for 36 h with Tet or left for 36 h in Tet-free medium were fixed and stained for the indicated antibodies. Eg5 and Plk1 are shown in green and  $\alpha$ -Tubulin in red. DNA was visualized using DAPI (blue). Scale bar is 10  $\mu$ m. (experiment performed by Dr. Anna Santamaria, MPIB) (B) Mass spectra representing a SILAC peak-pair for Plk1 and Eg5, respectively, from shEg5/shPlk1-induced cells.

Taxol-stabilized mitotic spindles (containing KT and centrosomes) were purified essentially as previously described<sup>93</sup>. To ensure minimal variation in sample handling and processing, equal amounts of Plk1 and Eg5-depleted/inhibited cells were combined prior to spindle isolation (figure 12A-C). Examination of the purified spindles by differential interference contrast (DIC) light microscopy revealed predominantly monopolar spindles (Figure 12D, E), as expected, largely free of other cellular structures.

---

Purified spindles were subsequently solubilized and separated by one-dimensional SDS-PAGE gradient gel electrophoresis. The gel was then cut into 16-20 slices, covering all spindle protein mass ranges (Figure 12F). Spindle Proteins were digested in-gel with trypsin and the resulting peptide fractions were subjected to phosphorylated peptide enrichment on TiO<sub>2</sub> beads prior to analysis of both TiO<sub>2</sub>-elution and TiO<sub>2</sub>-flow-through fractions (phosphorylated and unphosphorylated peptides, respectively), using a nanoLC-LTQ-Orbitrap system.

### **3.1.3 Plk1-dependent human mitotic spindle proteome and phosphoproteome**

From 226 LC-MS/MS runs for proteome and phosphoproteome quantification, we identified 3395 unique proteins at a false discovery rate (FDR) of 2%. About 60% of the proteome (1996 proteins) was phosphorylated, and more than 90% (3108) of identified proteins had quantification profiles (i.e. could be quantified with SILAC). The phosphorylation output file of MaxQuant contained 9544 phosphopeptides. However, this list still contains phosphopeptides with low identification Mascot score and duplicates in the form of phosphorylation-isoforms, for which the exact site of phosphorylation within the peptide sequence could not be determined.

After applying a Mascot score cut-off  $\geq 12$  for all assigned phosphorylated peptides and a phosphorylation site localization probability score cut-off  $\geq 0.75$  (class I sites)<sup>64</sup>, 4420 unique phosphorylation sites with high confidence were identified at a false-discovery rate (FDR) of 0.89% from 4 different experiments (3526 from inhibitor experiments and 3212 from shRNA experiments). Distributions of phosphoserine (79.8%), phosphothreonine (19.6%) and phosphotyrosine (0.6%) sites (localization probability  $> 0.75$ <sup>64</sup>), resembled the ones reported in an independent quantitative analysis of the human spindle phosphoproteome at distinct mitotic stages<sup>94</sup>.

To control for fluctuations in levels of individual proteins between the different experimental conditions, all measured phosphorylated peptide ratios were normalized against the levels of the protein in question.

Collectively, these analyses led to a final number of 3894 unique phosphorylation sites with SILAC and corresponding protein ratios, including 3092 and 2808 phosphorylation sites from experiments performed with small molecule inhibitors and the shRNA cell lines. Key

---

features of the identified phosphoproteome are summarized in figure 15.

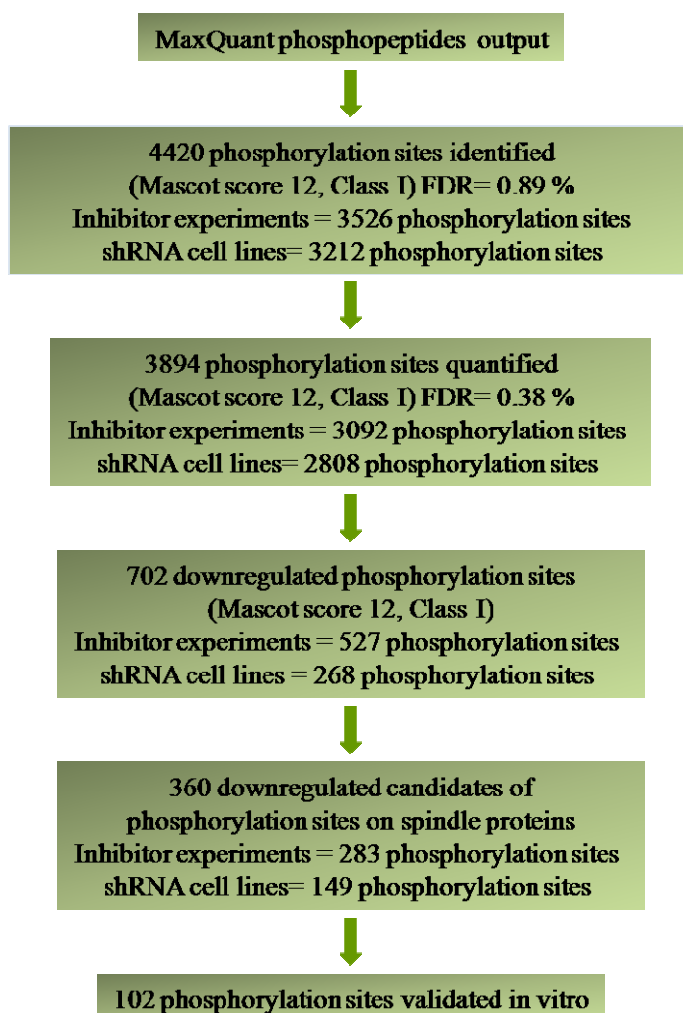


Figure 15. A phosphoproteome data processing work-flow for the detection of Plk1-dependent phosphorylation sites in isolated mitotic spindles.

### **3.1.4 Plk1-dependent quantitative human mitotic spindle phosphoproteome**

To describe changes in phosphorylation upon Plk1 inactivation, an independent small scale pilot experiment was performed to determine a threshold for significant phosphorylation regulation. TAL was applied to SILAC light and heavy-labeled HeLa S3 cells, respectively.

---

Mitotic cells were combined 1:1 and the sample analyzed by MS (1612 phosphorylation sites were identified with high confidence). The error rates for normalized ratios with significance threshold  $< 2/3$  and  $> 3/2$  were 4.4% and 1.9%, respectively.

We applied this threshold to our Plk1 experiments. In three out of four experiments, the Plk1 inhibited samples was light labeled and ratios  $> 3/2$  were considered as significant, and in the fourth experiment, Plk1 inhibited sample was heavy labeled and ratios  $< 2/3$  were considered as significant. Thus, the average quantitation error in our combined dataset can be estimated as  $(3 \times 1.9\% + 4.4\%) / 4$ , resulting in 2.5%. The selected quantitative change above 50% as significant for regulated phosphorylation events is also in line with other recent SILAC based studies<sup>67, 95-97</sup>. Furthermore, since our experimental strategy includes a downstream *in vitro* kinase assay for the verification of potential Plk1 direct targets, we consider this error rate as a good compromise between experimental sensitivity and stringency.

Based on the chosen ratios of  $> 1.5$  ( $3/2$ ) and  $< 0.66$  ( $2/3$ ) as thresholds for considering any observed change as significant, the automated quantification was accomplished by MaxQuant software. The ratios of each ion pair with different SILAC labeling were normalized by corresponding protein expression levels which were quantified only with unphosphorylated peptides.

A total of 702 phosphorylation sites were identified as down-regulated in at least one of the four experiments, 527 sites after TAL treatment and 268 sites after Plk1 depletion by shRNA (figure 15), with an overlap of 93 sites (figure 16B).

Between replicate experiments, a few phosphorylation sites with inconsistent trends (sites down-regulated in one experiment but up-regulated in the corresponding biological replicate) were excluded from further analysis. For the remaining sites (the vast majority), the overlaps of identified phosphorylation sites between two strategies and biological replicate experiments are listed in figure 16A.

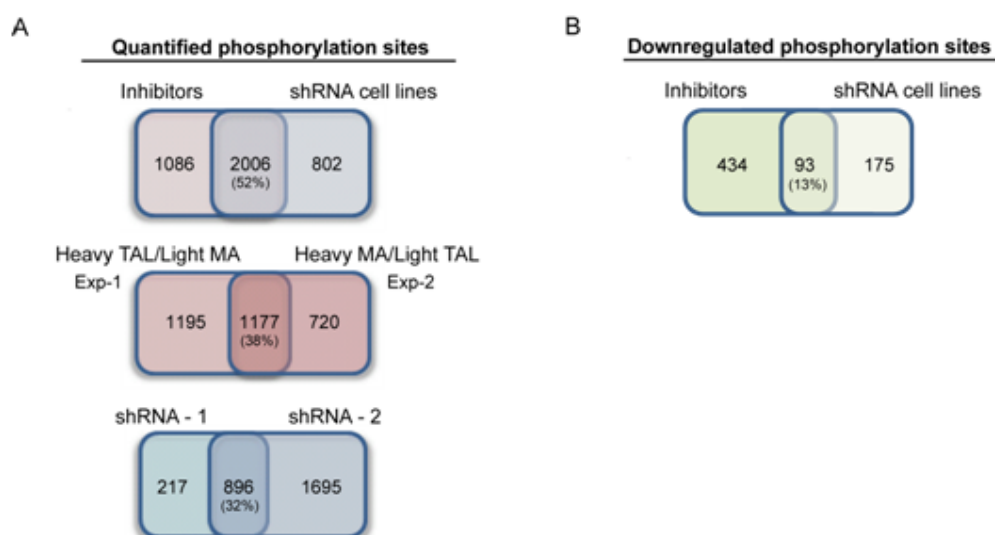


Figure 16. Comparative assessment of phosphoproteomics data. (A) Venn diagrams illustrating overlap of the phosphorylation sites identified in this study. Overlap between the two different strategies (upper diagram) and the two biological replicates from inhibitor or shRNA experiments (middle and lower diagram, respectively) are shown. (B) Overlap of the down-regulated phosphorylation sites identified from two approaches.

As an example for the quantification of SILAC pairs of phosphorylated peptides, the MS spectra of the kinesin KIF4A identified in experiments performed with both the shRNA cell lines and small molecule inhibitors are shown in figure 17.

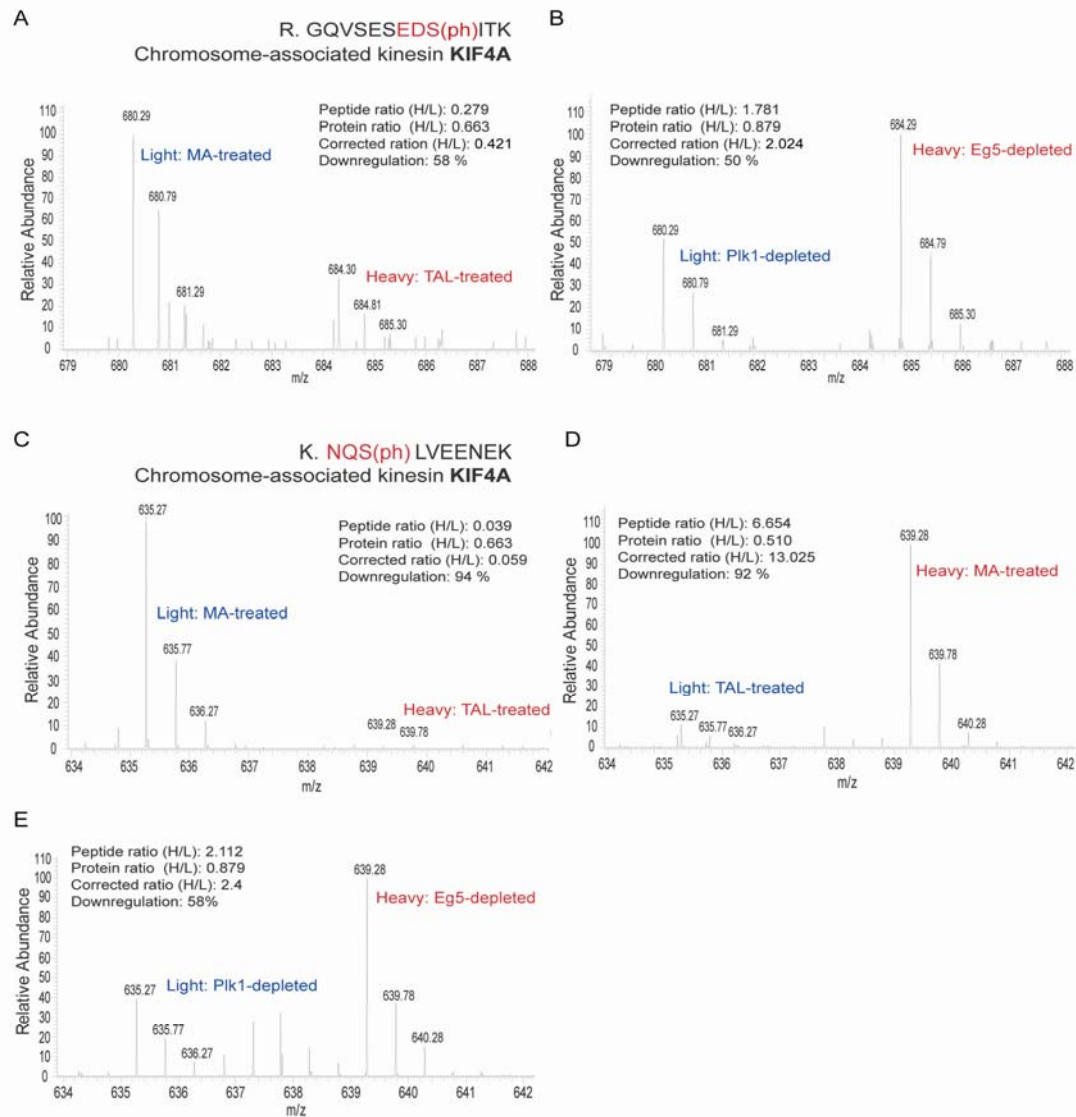


Figure 17. SILAC Kif4A phosphorylated peptides. Four independent experiments were carried out with different SILAC conditions. Due to the SILAC strategy for metabolic labeling, arginine- and/or lysine-containing peptides co-eluted in LC-MS as doublets with a defined mass difference. The ion signals allowed accurate measurement of relative peptide abundance. Representative MS spectra of a KIF4A phosphorylated peptides following the classical Plk1 consensus motif quantified on spindles isolated from MA/TAL-treated cells (A) and shEg5/shPlk1-induced cells (B). Measured and normalized peptide ratios are shown. (C-E) Representative MS spectra of a KIF4A phosphorylated peptide containing an N in position -2, quantified on spindles isolated from MA/TAL-treated cells (C and D) or

shEg5/shPlk1-induced cells (E). Reciprocal SILAC labeling in experiments shown in C and D resulted in inversed peptide ratios in the MS spectra (crossover experiment). MaxQuant assigned the same phosphorylation site within the peptide with high confidence (localization probability score 1). Measured and normalized peptide ratios are shown.

To address the reproducibility of phosphorylation site identification and quantification, half of the purified spindle protein sample from one of the experiments, performed with the Tet-inducible stable cell lines, was separated on a distinct 1D gel and subsequently analyzed as a technical replicate. A good reproducibility ( $r^2$ -value: 0.84) was observed (figure 18), in line with other recent phosphoproteome studies<sup>85,98</sup>.

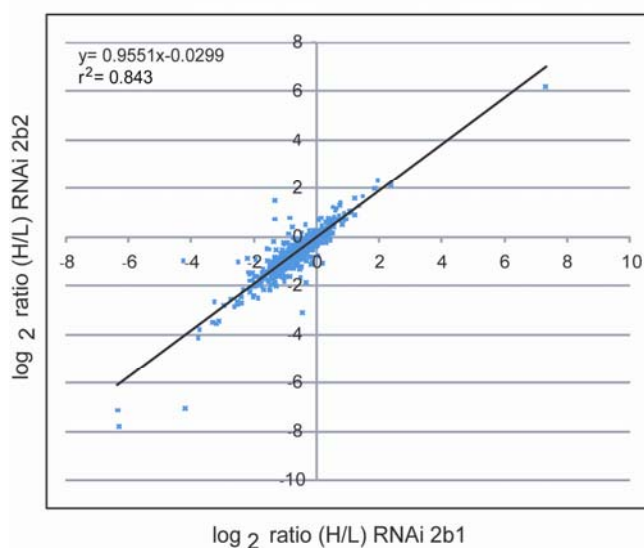


Figure 18. Reproducibility of phosphorylated peptides SILAC abundance measurements. Enriched spindle proteins from 20 gel slices of one RNAi experiment were analyzed twice in LC-MS/MS, datasets named as RNAi 2b1 and RNAi 2b2. For SILAC phosphorylated peptides identified in both datasets, the log transformed abundance ratios (H/L) from RNAi 2b1 were plotted against that from RNAi 2b2 (1236 unique phosphorylation sites).



### 3.1.5 Plk1 spindle substrates identification

There were 360 candidates of down-regulated phosphorylation sites quantified on spindle proteins upon Plk1 inactivation, either by TAL-treatment or stable shPlk1-mediated depletion (figure 15 and appendix, note: 68 sites were included regardless of low localization probability threshold for further *in vitro* kinase assays). Importantly, the majority (63%) of those down regulated phosphorylation had not been reported earlier.

In agreement with essential functions of Plk1 at the spindle poles, centrosomal and MT-associated proteins (MAPs), as well as motor proteins, were identified upon Plk1 inactivation (figure 19).

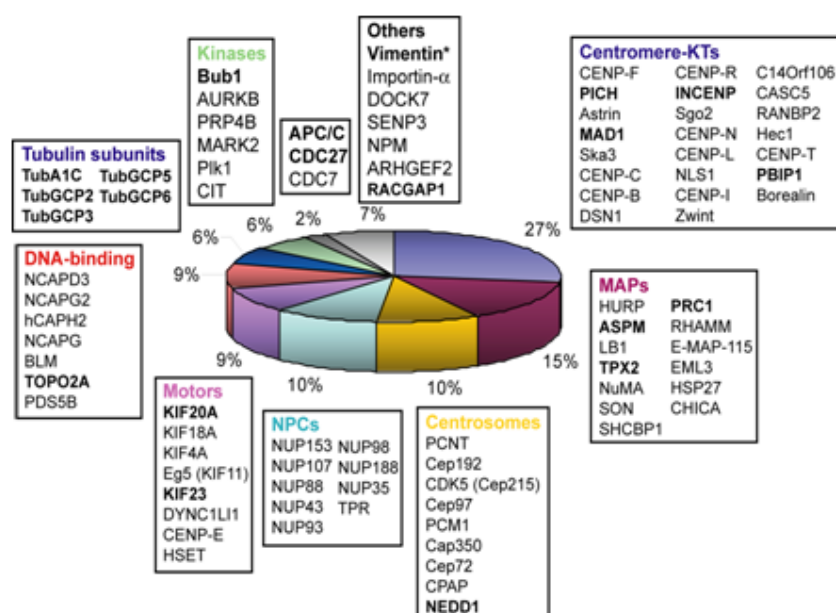


Figure 19. Plk1-dependent substrates on the mitotic spindle. The pie chart depicts proteins showing down-regulated phosphorylation sites upon Plk1 inactivation. They are classified into categories based on either function or localization. Known Plk1 substrates are shown in bold.

### 3.1.6 Validation of Plk1-dependent phosphorylation sites on spindle proteins through a candidate-based *in vitro* kinase screening

As a consequence of reduced Plk1 activity, phosphorylation sites on direct Plk1 substrates are expected to be down-regulated in Plk1-depleted or TAL-treated cells. However, considering that kinases and phosphatases often form complex signaling networks, the pool of

regulated phosphorylation sites is also expected to include indirect targets of Plk1 activity, notably proteins that are targets of kinases or phosphatases that act downstream of Plk1.

To distinguish direct from indirect substrates of Plk1 activity, we used a peptide spotting approach to validate potential direct Plk1 phosphorylation sites (experiments performed by Dr. Sabine Elowe, MPIB). Using peptides synthesized directly on cellulose membranes as substrates, we performed *in vitro* Plk1 kinase assays with recombinant Plk1 purified from insect cells (figure 12G). All down-regulated phosphorylation sites on spindle components were arrayed on the membrane as 12-mer peptides, with the potential phosphoacceptor amino acid kept constant at position 7 in either its serine/threonine/tyrosine native form or changed to alanine to verify signal specificity. Incubation of duplicate peptide arrays (generated simultaneously) with either kinase and [<sup>32</sup>P]-ATP or [<sup>32</sup>P]-ATP alone, followed by extensive washing, resulted in clear and specific signals on membranes phosphorylated by Plk1, whereas virtually no signal was detected on control membranes incubated with [<sup>32</sup>P]-ATP alone.

In total, 102 down-regulated sites (figure 15 and appendix) were verified as direct *in vitro* targets of Plk1 phosphorylation using this approach (table 1).

Table 1. Plk1 phosphorylation sites on known spindle proteins		
Proteins	Acc. Number	Phosphorylation sites
ARHGEF2	Q92974	S737
<b>ASPM</b>	Q8IZT6	T178, S267, S270, S280, S355, T356, S1825, S3426
AURKB	Q96GD4	Y239
BLM	P54132	S26*, S28*
<b>BUB1</b>	O43683	S661
CASC5	Q8NG31	S1013, S1808
<b>CDC27</b>	P30260	<b>S441</b>
CENPB	P07199	S306, S307
CENPC1	Q03188	S96, S104, S110, T231*, T516, S250, T734
CENPE	Q02224	S611, S112
CENPF	P49454	S242, S838, S1248, S1324, S1750, S2512, S2513
CENPI	Q92674	S709, S710
CENPJ (CPAP)	Q9HC77	S556*
CENPL	Q8N0S6	S53
CENPN	Q96H22	S226, S235, S282
CENPT	Q96BT3	S45
<b>CENPU (PBIP1)</b>	Q71F23	S194

C14Orf106	Q6P0N0	S135, S191, S192
Cep97	Q8IW35	S308
Cep170	Q5SW79	S880*
Cep192	Q8TEP8	S1502
Cep215	Q96SN8	S613, S1102
Cep350	Q5VT06	S2689
DGL7 (HURP)	Q86T11	S777
DSN1	Q9H410	S123
<b>ERCC6L (PICH)</b>	Q2NKX8	S774, S790
<b>INCENP</b>	Q9NQS7	S72, S330
KIF18A	Q8NI77	S681*
<b>KIF20A</b>	O95235	Y558, S635
<b>KIF23</b>	Q02241	S867*, S889
KIFC1 (HSET)	Q9BW19	S33
KIF4A	O95239	S815, S951
<b>MAD1L1</b>	Q9Y6D9	S8*, S490*
NCAPD3	P42695	T430, S508
NCAPG2	Q86XI2	S30, T1114
NCAPH	Q15003	T98
<b>NPM</b>	P06748	S198
NuMA	Q14980	T1818
Nup93	Q8N1F7	S72
Nup98	P52948	S591
Nup107	P57740	S4, S57
Nup153	P49790	S343
Nup188	Q5SRE5	S1523
PCM1	Q15154	S110
PCNT	O95613	S554, S813, S815, T1690, S2594, T3325, S3326
<b>PRC1</b>	O43663	S554
<b>RACGAP1</b>	Q9H0H5	S164, <b>S170, S214</b>
SGOL2	O562F6	S436, S1151*
SON	P18583	S154, T2022
SPAG5 (ASTRIN)	Q96R06	S401
<b>TOP2A</b>	P11388	S285
TPR	P12270	S1185
TPX2	Q9ULW0	S356, T361, S654
<b>TUBA1C</b>	Q9BQE3	S48, T51
<b>TUBGCP5</b>	Q96RT8	S182
Vimentin	P08670	S83, S409, T458, S459
ZWINT	O95229	S84*

\* sites found in pilot experiments and in this study but excluded from the appendix due to Mascot score < 12 or contradictory ratio between inhibitor or cell lines experiments. Known Plk1 substrates and phosphorylation sites are marked in bold.

---

### 3.1.7 Reconsideration of the Plk1 consensus motif

Analysis of the *in vitro* verified Plk1 sites using the Motif-X software tool (established in collaboration with Dr. Rainer Malik, MPIB), revealed many peptides that conform to the general Plk1 kinase consensus motif D/E-X-S/T-Φ (where X is any amino acid and Φ corresponds to a hydrophobic amino acid) (appendix)<sup>81</sup>. Interestingly, the percentage of positive peptides containing a glutamine (Q) in -2 position was comparable to the one observed on peptides that conformed to the classical Plk1 consensus motif (59% and 56%, respectively) (appendix). Furthermore, an asparagine (N) in -2 position was also observed in a large number (58%) of the *in vitro* validated Plk1 phosphorylation sites, and we note that an N at the -2 position has previously been reported in a number of *in vivo* substrates of Plk1 or its yeast homolog Cdc5p, including the APC subunit Cdc27<sup>149</sup>, the cohesion subunit Rec8<sup>150</sup>, the cytokinesis effector kinase Rock2<sup>151</sup> and the centralspindlin complex component MgcRacGAP<sup>152</sup>. Furthermore, N, E and D at the -2 position occurred with similar frequency in our set of validated sites, suggesting that an N at the -2 position is prevalent in the target sequence of a broad range of *in vivo* Plk1 substrates.

To further explore the optimal Plk1 consensus motif, an *in vitro* kinase assay was performed to screen for Plk1 phosphorylation on the CENP-F peptide containing S1324 (experiment performed by Dr. Sabine Elowe, data not shown). Originally, the peptide contained an N at the -2 position. Each residue from positions -4 to +2 was substituted with every other natural amino acid, and the ability of Plk1 to phosphorylate the resulting peptides was analyzed in an *in vitro* phosphorylation assay. The results clearly demonstrate a strong preference of Plk1 for particular residues N-terminal to the phosphoacceptor amino acid; instead, virtually all amino acid substitutions were well tolerated at the +1 and +2 position, although there was also a clear preference for hydrophobic residues. A notable exception was proline at the +1 position, which was not tolerated. This strongly suggests that S/T-P motifs are not substrates for Plk1. Furthermore, we note the equal preference for N or E at the -2 position. Significantly, a leucine in position -3 was prominent in many validated phosphorylated peptides and substitution of this residue in the CENP-F peptide screen was poorly accepted.

---

Collectively, these observations are in agreement with both our MS data and the results of the peptide spotting screen. The data presented above suggest that the consensus sequence for Plk1 should be broadened to L( $\Phi$ )-E/N/D(Q)-X-S/T-( $\Phi$ ).

### **3.1.8 Plk1-dependent localization of spindle components**

In accordance with the requirement for Plk1 activity to properly localize some of its substrates on the mitotic spindle<sup>13</sup>, we considered it likely that the localization of some of its downstream targets would be altered as a consequence of Plk1 inhibition. To analyze this possibility in a systematic and quantitative manner, we prepared total lysates from cells grown in SILAC medium treated with either MA or TAL, and compared the SILAC ratios to the ones quantified in experiments with isolated spindles.

Proteins showing a reduced or increased association with the spindle apparatus are displayed in figure 20A (note that only proteins with SILAC ratios identified in both spindle and total lysate experiments are shown). Strikingly, upon Plk1 inactivation many proteins were depleted from the spindle and many contained down-regulated phosphorylation sites (figure 19 and appendix), arguing that Plk1 activity is required for their spindle localization.

Our MS analysis demonstrates that levels of the  $\gamma$ -Tubulin complex subunits and the centrosomal proteins pericentrin, Cep192, and Cep215 (PCM proteins implicated in centrosomal  $\gamma$ -Tubulin attachment<sup>99</sup>), were also highly reduced upon Plk1 inhibition. Importantly, our study revealed novel Plk1-dependent phosphorylation events on these proteins that might contribute to their role in centrosome maturation and spindle assembly.

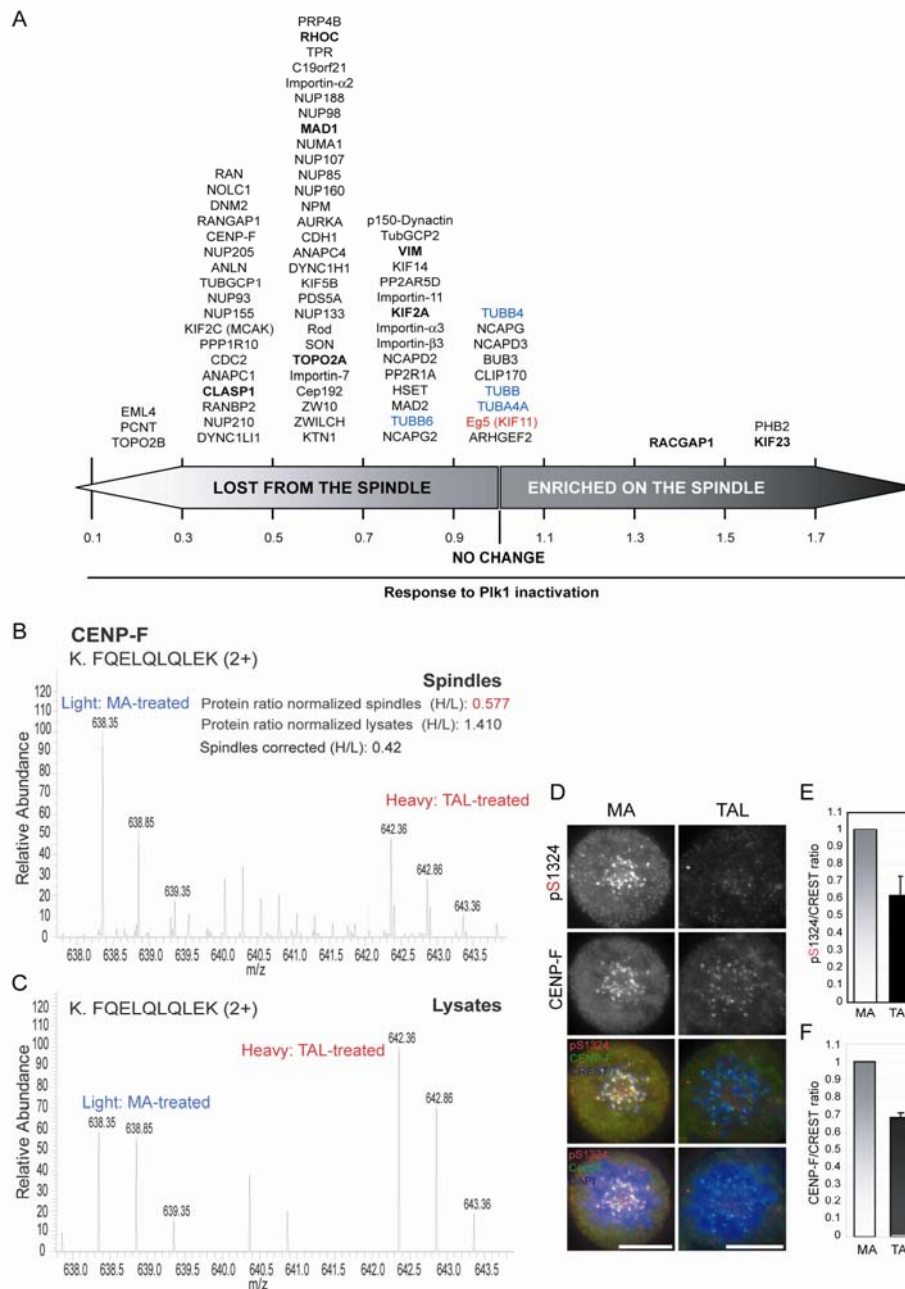


Figure 20. Plk1 dependent localization of spindle components. (A) Proteins with altered spindle levels upon Plk1 inactivation are shown. Known Plk1 substrates are shown in bold. Proteins were subcategorized according to their relative spindle amount from 0 to 1.7. Ratio = 1 indicates no change on spindle levels relative to total lysate amounts. Tubulin subunits used to normalize all ratios are depicted in blue, whereas the control protein Eg5 (KIF11) with a ratio close to 1 is shown in

---

red. (B, C) Representative MS spectra of a CENP-F phosphorylated peptide quantified on spindles (B) and total lysates (C) of MA/TAL-treated cells. Measured, corrected and normalized peptide ratios are shown. (D) HeLa S3 cells treated with MA or TAL were fixed and immunostained for pS1324 CENP-F (red), CENP-F (green) and CREST (far red, depicted in blue). DNA was visualized using DAPI (blue). Scale bar, 10  $\mu$ m. (E, F) Fluorescence intensity of pS1324 on CENP-F and CENP-F, respectively, was quantified at KTJs in MA or TAL-treated cells. Intensities were plotted as a CENP-F or pS1324 versus CREST ratio (10 cells/>10 KTJs per cell, n= 3 independent experiments). Columns and error bars show average and standard deviation of 3 independent experiments. (experiments for figures D-F were performed by Dr. Anna Santamaria, MPIB)

An intriguing case was CENP-F, a large coiled-coil protein that transiently localizes to outer KTJs, and preferentially to KTJs of unaligned chromosomes<sup>100</sup>. Although it was previously shown that CENP-F exhibits slower electrophoretic mobility upon mitotic entry<sup>101</sup>, very little is known about CENP-F phosphorylation. We identified 19 serine and threonine phosphorylation sites on CENP-F that were suppressed by more than 2-fold upon Plk1 inhibition, 8 of which we could validate *in vitro* (table 1 and appendix). Collectively, these results identify CENP-F as a novel Plk1 substrate.

Our MS data further suggest that the levels of CENP-F on the spindle apparatus were reduced upon Plk1 inactivation when compared to the corresponding levels in total lysates (figure 20A-C). Immunofluorescence analysis also revealed a significant reduction of total CENP-F and CENP-F S1324 phosphorylation signal at KTJs upon TAL treatment (figure 20D-F). These data demonstrate that CENP-F localization is directly regulated by Plk1 phosphorylation.

### **3.1.9 Integration of Aurora A and Plk1 activity**

Aurora A localization to centrosomes has also recently been shown to depend on Plk1 activity<sup>86, 102, 103</sup>. In agreement with this, our MS data shows that Aurora A levels on the isolated spindle were reduced by 50% upon Plk1 inhibition when normalized against the H/L ratio for Aurora A in total lysates (figure 21A and B).

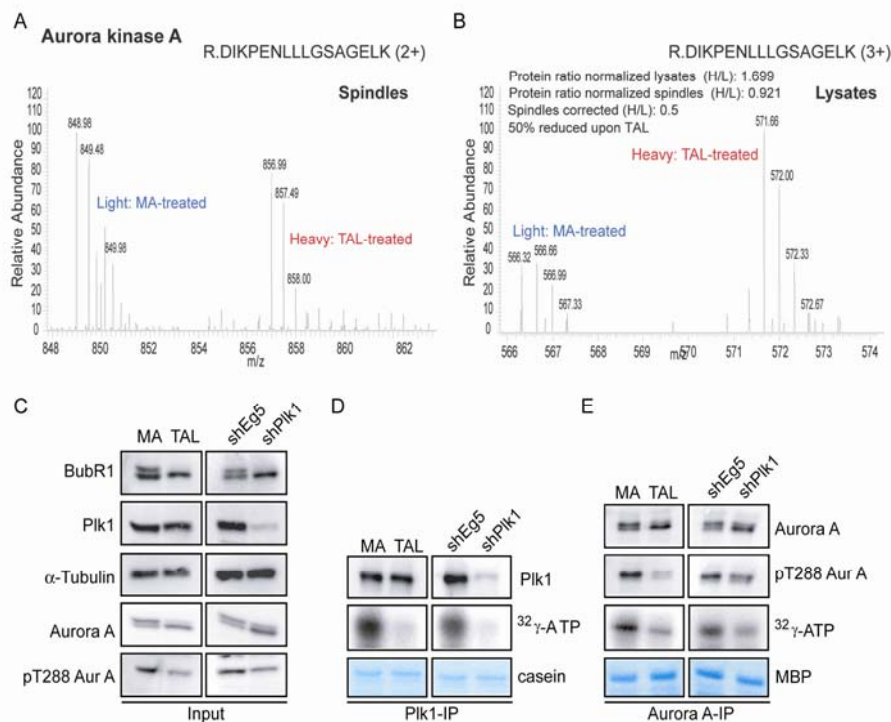


Figure 21. Reduced Aurora A activity as consequence of Plk1 inactivation. (A, B) Representative MS spectra of an Aurora A phosphorylated peptide quantified from spindles (A) and total lysates (B) of MA/TAL-treated cells. Corrected and normalized peptide ratios are shown. (C-E) Plk1 and Aurora A *in vitro* kinase assays. Kinases were immunoprecipitated from MA/TAL treated cells and shEg5/shPlk1-induced cells. Casein and MBP were used as model substrates (D and E, respectively). Protein phosphorylation is visualized by autoradiography [ $^{32}$ P]. Total cell lysates (Input, C) and immunoprecipitates (D and E) were probed by Western blotting for the indicated proteins (experiments for figures C-E were performed by Dr. Anna Santamaria, MPIB)

It has also recently been postulated that Plk1 may contribute to efficient Aurora A activation<sup>104, 105</sup>. Indeed, when the two kinases were immunoprecipitated from either shPlk1-depleted/TAL-treated cells or shEg5-depleted/MA-treated cells (for control) and then assayed *in vitro* on appropriate substrates (casein and myelin basic protein (MBP), respectively), both kinases showed strongly reduced activity (figure 21C-E). While the reduction in activity was expected for Plk1, the significant reduction in activity observed in parallel for Aurora A strongly suggests that inhibition of Plk1 also reduces Aurora A activity.



---

In further support of the above conclusion we emphasize that several of the phosphorylation sites identified as down-regulated after Plk1 depletion/inhibition by MS analysis conform to the Aurora A consensus motif<sup>106, 107</sup>. Interestingly, many of these sites were present in proteins likely to be Aurora A substrates, including known Aurora A interactors such as TPX2<sup>108</sup> or centrosomal proteins and MAPs involved in centrosome maturation and spindle assembly.

Interestingly, we also identified substrates common to both Plk1 and Aurora A, such as pericentrin, the  $\gamma$ -Tubulin subunit GCP2, and the centrosomal protein Cep215, suggesting that these two kinases cooperate to control key early mitotic functions through co-regulation of several important substrates. Collectively, these observations support the idea that Aurora A is a downstream effector of Plk1 activity<sup>105</sup>.

---

## 3.2 Phosphoproteome analysis of human mitotic spindle proteins by using low-specificity protease elastase

### 3.2.1 Introduction

Proteomics aims at the large scale characterization of the protein contents of complex biological systems<sup>109, 110</sup>. As a result of improved mass accuracy, resolution, sensitivity, and speed, mass spectrometry has become a robust technique for proteome analysis<sup>111-113</sup>. It allows not only the generation of comprehensive protein inventories, but also the large scale mapping of post translational modifications, such as protein phosphorylation, which has emerged as a new sub-discipline in the field of proteomics<sup>114-116</sup>. Reversible protein phosphorylation is a key regulatory mechanism involved in almost all cellular activities<sup>69, 117, 118</sup> and thousands of phosphorylation sites have been mapped by combining mass spectrometry with dedicated enrichment strategies<sup>72, 73, 119-121</sup>. Despite this progress, however, the complete characterization of all phosphorylation sites is still challenging.

High throughput protein identification is based on the mapping of the best matches between experimental tandem mass spectrometry (MS/MS) spectra and theoretical fragmentation patterns calculated from known protein sequences<sup>122, 123</sup>. Even though intact proteins have been successfully analyzed by MS/MS<sup>124, 125</sup>, this technique is far more sensitive when applied to peptides in the range of 5 to 25 amino acids. Therefore, proteins are digested by proteases prior to MS/MS analysis, and trypsin (cleaving at the C-termini of lysine and arginine) is generally the enzyme of choice due to its high cleavage specificity<sup>111</sup>.

Detection and fragmentation efficiencies of peptides depend strongly on their size and sequence<sup>112</sup>. This bias of MS/MS towards a subset of peptides is not a serious obstacle for confident protein identification since only a few good peptide-spectrum-matches (PSMs) are sufficient to accomplish this task. However, complete sequence coverage of proteins is mandatory for a comprehensive mapping of post-translational modifications. Since this goal can rarely be achieved using only one enzyme, analysis of samples after proteolysis with different enzymes is a common strategy to increase protein sequence coverage<sup>126, 127</sup>. In particular, highly specific proteases, such as Glu-C (cleaving at the C-terminus of glutamic acid) and Lys-C (cleaving at the C-terminus of lysine), have been used in phosphoproteome studies<sup>128</sup>. However, Lys-C frequently produces very large peptides, which are often less

---

suitable for collision induced dissociation (CID) MS/MS analysis, and the cleavage specificity of Lys-C partially overlaps with that of trypsin. Similarly, Glu-C digestion generates mostly large peptides and a dramatic decrease in the number of detected peptides has been observed when this enzyme was applied to complex samples <sup>128</sup>.

As an alternative to specific enzymes, low-specificity proteases such as elastase, pepsin, thermolysin and proteinase K have also been applied to increase the sequence coverage of selected proteins <sup>127, 129-131</sup>. Amongst these enzymes, elastase has attracted particular attention since it generates mainly medium sized peptides in the mass range between 500 Da and 1500 Da <sup>129</sup>. The disadvantage of low-specificity proteases, however, is a strong increase in search time and a strong decrease in search specificity. Whereas spectra of tryptic peptides only need to be matched against all tryptic database peptides in the measured mass window, the search space increases more than 100 fold if no enzyme is specified <sup>111</sup>. Thus, a search without enzyme specificity corresponds to a search against a much larger database resulting in decreased detection sensitivity <sup>132</sup> and higher false discovery rate (FDR). Therefore, low-specificity enzymes have mostly been used to increase the sequence coverage of well defined samples of modest complexity <sup>127, 129-131</sup>. Even though low-specificity proteases have not yet been used successfully in large-scale phosphoproteome studies, their application may profit strongly from the recent development of highly accurate and sensitive Orbitrap mass spectrometers <sup>31</sup>. As a result of mass accuracies in the low ppm range <sup>133</sup>, the search specificity increases strongly since only peptides in a very narrow mass window have to be matched against the corresponding MS/MS spectra.

The present study was aimed at evaluating to what degree the disadvantages of low-specificity enzymes may be compensated for by the high mass accuracy of the Orbitrap. Specifically, we have analyzed the performance of the low-specificity protease elastase, in comparison with trypsin, in a large-scale phosphoproteome analysis of mitotic spindle preparations from human cells <sup>84</sup>. Mitotic spindles are essential for accurate chromosome segregation during cell division <sup>83, 134-137</sup>. Many spindle proteins are regulated by kinases and phosphatases and, therefore, a comprehensive mapping of phosphorylation sites in these proteins would clearly be important <sup>138, 139</sup>. The results of our present analysis demonstrate that elastase is highly complementary to trypsin and efficient in increasing the coverage of phosphorylation sites. With an optimized MS mass error window, an improved database

---

search strategy, and the introduction of a novel elastase cleavage specificity factor, we have successfully limited the FDR in a large scale dataset. (All optimization of data analysis was carried out in collaboration with Dr. Rainer Malik, MPIB).

### **3.2.2 Experimental outline**

Elastase was compared to trypsin for the identification of phosphorylation sites from a complex mixture of proteins associated with the human mitotic spindle<sup>83, 84</sup>. To this end, purified spindles were separated on two 1D gels and gel slices digested with elastase or trypsin, respectively. Subsequently, TiO<sub>2</sub> enriched phosphorylated peptides and the flow-through fractions from the TiO<sub>2</sub> columns (un-phosphorylated peptides) were analyzed by LC-MS/MS on an Orbitrap mass spectrometer.

### **3.2.3 Adjustment of the Elastase Concentration**

To adjust the elastase concentration for proteolysis, we digested gel-separated samples of Bovine Serum Albumin (BSA) at elastase concentrations of 3.75 ng/μl, 7.5 ng/μl, 15 ng/μl, 30 ng/μl, and 60ng/μl for 16-hours. We found that both the sequence coverage and the background of peptides derived from elastase auto-proteolysis increased with increasing enzyme concentration (data not shown). A concentration of 15 ng/μl was chosen as the best trade-off between sequence coverage and elastase auto-proteolysis peaks and hence was used for all subsequent experiments.

### **3.2.4 MS Mass Accuracy**

The high mass accuracy of the LTQ-Orbitrap was expected to be crucial to partially compensate for the lower specificity of elastase cleavage. To evaluate the dependencies of the number of true positive identifications and false discovery rates (FDRs) from the employed MS mass error settings, trypsin and elastase derived datasets (18 gel slices, each) were first searched with MASCOT against SwissProt human (merged with a reversed decoy database) at a mass error of 7 ppm. Results were then combined using MaxQuant and subsequently filtered at mass accuracy stringencies ranging from 1 ppm to 7 ppm. The FDR was initially reduced by applying a MASCOT cut-off score of 20. Next, we calculated (table 2) and plotted (figure 22) the ratios between the hits in the forward database and the FDR to identify the best compromise between search sensitivity (number of identifications) and accuracy (low FDR). From figure 22, it becomes obvious that the ratios between the number of identifications and

---

the FDR peak at a mass accuracy setting of 2 ppm, so this value was employed for all subsequent analyses.

---

**Table 2.** Numbers of phosphorylated peptide identifications in the forward database and FDRs for different mass accuracy settings. Search results of LC-MS/MS spectra with MASCOT identification scores >20 of trypsin and elastase digested samples were combined.

---

mass error window	peptides FDR	no. of reverse hits	no. of forward hits	no. of forward hits/FDR*100
±1ppm	1.0%	18	3633	3633.0
±2ppm	1.0%	20	4171	4171.0
±3ppm	1.1%	23	4281	3891.8
±4ppm	1.2%	27	4313	3594.2
±5ppm	1.2%	27	4327	3605.8
±6ppm	1.3%	28	4335	3334.6
±7ppm	1.3%	28	4345	3342.3

---

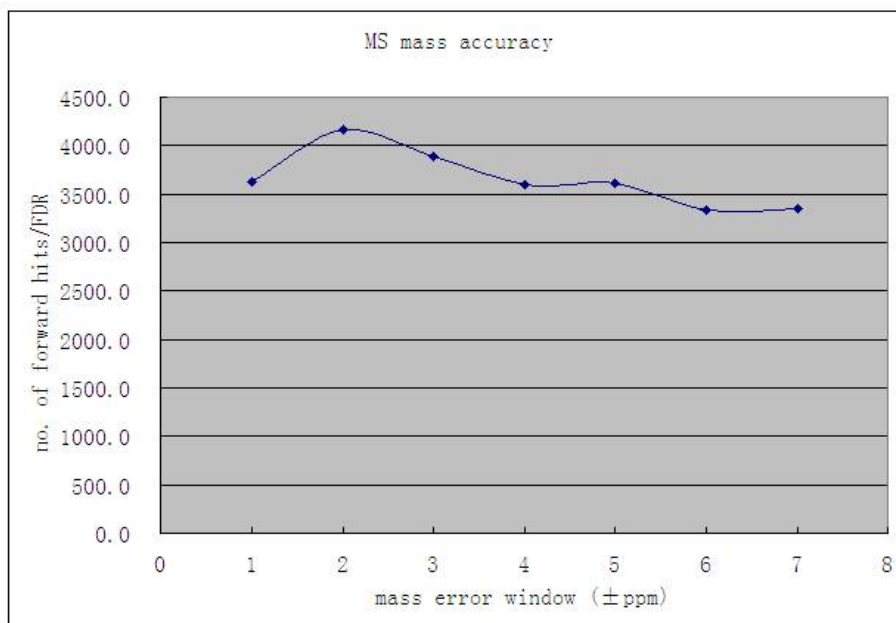


Figure 22. Ratios between the numbers of identifications in the forward database and the FDRs for different mass accuracy settings. Search results of LC-MS/MS spectra with MASCOT identification scores >20 of trypsin and elastase digested samples were combined.

### 3.2.5 Elastase cleavage specificity

It has been reported previously that elastase cleaves predominantly C-terminal of the amino acids Gly, Val, Leu, Ala, and Ile, and to a lesser extent after Phe, Pro, Glu, and Arg<sup>131, 140</sup>. These results, however, were derived from only one substrate protein (elastin), and the conditions used for digestion did not exactly match the ones employed in our study. Therefore, we determined the cleavage frequencies of elastase C-terminal to the 20 common amino acids based on 931 high-scoring peptides (MASCOT score >25) from the elastase 1 dataset (table 3). The cleavage probabilities were calculated by dividing the numbers of observed cleavages C-terminal to each the amino acid by the number of the occurrences of the same amino acid in the sequences of the detected peptides. Cleavage probabilities above 10% were observed (in decreasing order) for Ala, Val, Ile, Thr, Ser, Leu, and Met, which is only in partial agreement with the studies cited above. This discrepancy, however, may be explained

---

by the different source of enzyme, different digestion conditions (i.e. in-gel versus in-solution), and the much broader range of substrates in our analysis.

---

**Table 3. Cleavage Frequencies of Elastase at the C- Termini of 20 Standard Amino Acids<sup>a</sup>**

amino acid	cleavages observed	total number of residues	cleavage probability
A	357	820	43.5%
V	314	861	36.5%
I	241	694	34.7%
T	190	627	30.3%
S	197	919	21.4%
L	267	1367	19.5%
M	31	198	15.7%
H	14	154	9.1%
G	85	969	8.8%
N	28	435	6.4%
C	2	33	6.1%
R	22	378	5.8%
Q	33	599	5.5%
Y	10	204	4.9%
K	13	466	2.8%
F	7	521	1.3%
D	10	1056	0.9%
E	9	1427	0.6%
P	3	783	0.4%
W	0	50	0%

---

---

<sup>a</sup> From dataset elastase 1, SwissProt human search, 931 elastase derived peptides from TiO<sub>2</sub> elution and TiO<sub>2</sub> flow-through samples. FDR 1%, MS mass accuracy  $\pm 2$  ppm, MS/MS mass accuracy  $\pm 0.3$  Da.

---

### 3.2.6 Elastase cleavage scores

\*Established in collaboration with Dr. Rainer Malik (MPIB)

Even though proteolysis by elastase is much less specific than proteolysis by trypsin, the determined cleavage probabilities may be used to increase the specificity of peptide identifications in protein databases. Unlike for trypsin, however, the strict cleavage rules applied in MASCOT searches and the upper limit of nine missed cleavage sites are not appropriate for elastase, since cleavages probabilities are more broadly distributed over the 20 amino acids (table 3). Therefore, we decided to search the spectra of elastase derived peptides with the “no enzyme” setting and then filter the identified peptides in a second step, using the determined cleavage probabilities to reduce the number of false positive identifications. To this end, we defined external cleavage scores for identified peptides in the forward and decoy databases (at an FDR of 5%). These scores were calculated as the product of the elastase cleavage probability of the amino acid at the C-terminus of each peptide multiplied by the cleavage probability of the C-terminal amino acid of the preceding peptide (the amino acid N-terminal to the peptide’s N-terminus):

$$External\_Score = cleav.prob.(N) * cleav.prob.(C)$$

where  $cleav.prob(N)$  represents the cleavage probability of the C-terminal amino acid of the preceding peptide (the amino acid N-terminal to the peptide’s N-terminus) and  $cleav.prob(C)$  stands for the cleavage probability of the amino acid at the peptide’s C-terminus (table 3).

Next, we compared the medians of the external cleavage scores for the identifications in the forward versus the decoy database. As expected, the median external score among decoy database hits (0.0112) was considerably lower (7.6 fold) than that among forward database hits (0.0850), suggesting a substantial discriminative power of the external cleavage scores.

Peptides composed of many amino acids with high elastase cleavage probabilities likely represent false positive identifications, since proteolytic cleavages within these peptides



---

would be expected. To evaluate if this expectation holds true, we calculated an internal cleavage score for each identified peptide in the forward and decoy databases (at a FDR of 5%), defined as the product of 1 minus the cleavage probabilities of all but the C-terminal amino acid:

$$Internal\_Score = \prod_1^N (1 - cleav.prob.(N))$$

where N represents all amino acids within the peptide sequence except the C-terminal amino acid and cleav.prob. stands for the elastase cleavage probabilities reported in table 3.

As for the external cleavage score, we calculated the median internal scores for the hits in the decoy database (score: 0.145) and the forward database (0.191). Even though the discriminative power of the internal cleavage score (factor 1.3) is much lower than that of the external score (factor 7.6) we decided to use this factor as an additional means to increase search specificity.

Instead of consecutively filtering the lists of forward and decoy hits using appropriate cutoff values of MASCOT scores, external scores, and internal scores, we calculated a combined score composed of these three factors by using Discriminant Function Analysis (DFA)<sup>141</sup>, named as discriminant score (D-score).

The centroids of the discriminant score distributions of hits in the forward and decoy databases were calculated as + 0.076 and -0.563, respectively, underlining the discriminative power of the D-score. Next, we determined that filtering the elastase dataset using a discriminant score cutoff of  $D = -1.25$  resulted in the desired FDR of 1%. Therefore, we used this cutoff value for all subsequent data analyses.

### **3.2.7 Comparison of phosphorylation site identification using trypsin and elastase**

After having optimized the database search strategy for elastase derived peptides, we aimed to evaluate to what degree digestion by elastase could complement digestion by trypsin for phosphorylation site identification. In experiment 1, equal amounts of spindle proteins were processed in parallel with trypsin and elastase, respectively. MS spectra were searched against SwissProt human (including a reversed decoy database) using MASCOT. Whereas

---

trypsin derived samples were searched using trypsin specificity settings, the “no enzyme” option was used for elastase derived spectra. Next, the FDR was reduced by filtering of the peptide lists with the optimized  $\pm 2$  ppm mass accuracy criterion, and, for the elastase dataset, the discriminant score. Furthermore, in order to ensure the reliability of phosphorylation site identifications and to further decrease the FDR, a cutoff of the localization probability score (as provided by MaxQuant) of 75% was set, corresponding to the class of phosphorylation sites with highest confidence in localization <sup>64</sup>. This resulted in the identification of 703 phosphorylation sites using trypsin versus 185 phosphorylation sites from the elastase sample (figure 23). Although, the number of unique phosphorylation sites found with elastase was much lower than that from the tryptic sample, there was an overlap of only 4.2%, suggesting that trypsin and elastase are highly complementary in large scale phosphoproteome studies.

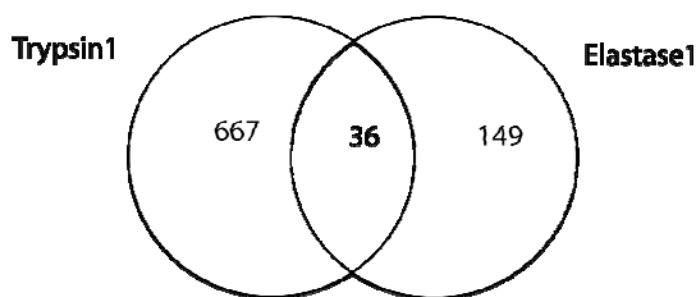


Figure 23. Detected Phosphorylation sites using trypsin or elastase for database searches against SwissProt human.

To evaluate whether the lower number of detected phosphorylation sites using elastase digestion was related to the larger search space (“no enzyme” search setting), we also searched the trypsin derived dataset against SwissProt human using the “no enzyme” option. Compared to the more specific search, the number of identified peptides decreased greatly (40% reduction, data not shown), suggesting that the larger search space is indeed a major factor in explaining the lower number of identified peptides using elastase. Therefore, we next aimed at decreasing the search space for the elastase datasets without compromising the confidence of peptide identifications.

---

### 3.2.8 “Custom database” search strategy

It has been reported that database size significantly affects the sensitivity of database search algorithms<sup>142</sup>. Since large databases include more protein/peptide sequences, the chance of random assignments also increases. On the other hand, a database should be sufficiently large to contain all detectable proteins present in the sample since spectra may be matched to the false second-best-hit if the correct hit is not present in the database.

Here we introduce a two-step search strategy to increase the search sensitivity while maintaining high stringency for phosphorylation site identification. In the first step, the MS spectra of trypsin and elastase digested TiO<sub>2</sub> elution and flow-through samples were searched against SwissProt human (including the decoy database) with a relaxed protein/peptide FDR < 5%. After removing the decoy hits, a “custom database” containing 1030 identified proteins was built and concatenated with a decoy version of this novel database. To ensure that the “custom database” indeed covers the detectable proteins in our samples, the numbers of unassigned spectra in four representative elastase LC-MS/MS runs against SwissProt human, and the “custom database”, were compared. The number of unassigned spectra decreased slightly (by 86 unassigned spectra for the four LC-MS/MS runs) for the “custom database” search strategy whereas the opposite trend would be expected if the custom database would not largely cover all detectable proteins in the sample.

The elastase and trypsin datasets were then searched against this “custom database” using identical filtering criteria as for the SwissProt search. Employing this new minimal database strategy, the number of identified phosphorylation sites using elastase increased by 85% to 343 sites (figure 24), while the number of phosphorylation sites found with trypsin increased only moderately by 13% (795 phosphorylation sites, figure 24). Importantly, the overlap between the datasets is still low (6.8%), presumably reflecting high complementarity of the elastase and trypsin cleavage specificities.

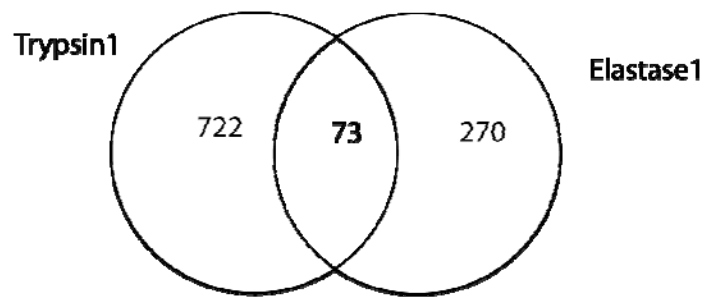


Figure 24. Detected Phosphorylation sites using trypsin or elastase for database searches against a “custom database”.

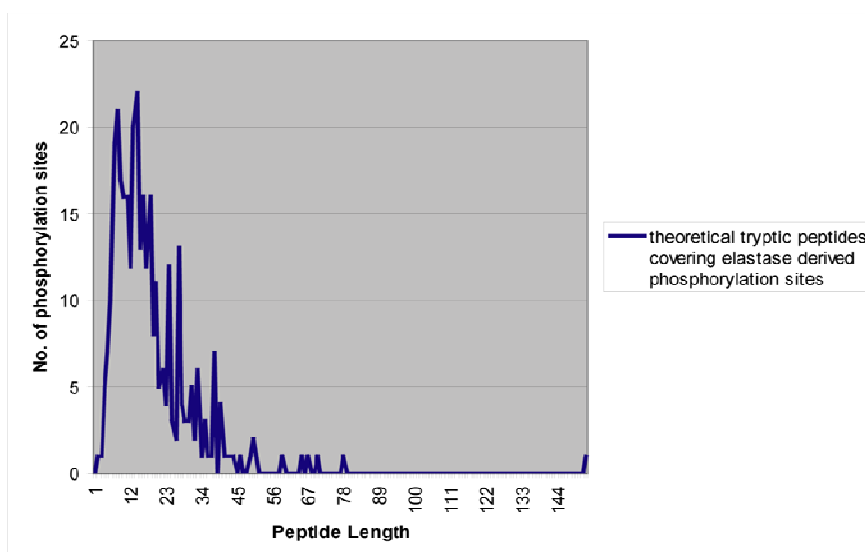
### 3.2.9 Complementary phosphorylated peptide coverages using trypsin and elastase

\*Established in collaboration with Dr. Rainer Malik (MPIB)

To rationalize the highly complementary coverage of phosphorylated peptides detected by trypsin and elastase we analyzed whether the theoretical size of tryptic peptides covering elastase derived phosphorylation sites would be outside the optimal sensitivity range for our experimental strategy. Figure 25 shows the length-distributions of phosphorylated peptides observed after tryptic digestions, elastase digestions, and the theoretical tryptic peptides covering elastase derived phosphorylation sites. Interestingly, the size distribution of the theoretical tryptic peptides (mean: 18.32 amino acids, standard deviation: 13.64 amino acids) is much broader than the observed distributions of experimentally determined tryptic (mean 11.24 amino acids, standard deviation: 2.80 amino acids) and elastase (mean 11.56 amino acids, standard deviation: 2.52 amino acids) derived phosphorylated peptides (figure 25B).

Importantly, many of the tryptic peptides phosphopeptides would not have been detected using trypsin since the resulting peptide sizes are not optimal. The most pronounced example regarding this observation is the phosphorylation site S486 (phosphorylated peptide GGS(ph)FGAYGAQEEAQCPT) of the major centromere autoantigen B (CENPB\_HUMAN) which was detected after elastase digestion. The corresponding tryptic peptide would have a molecular weight of 17 kDa (153 AA), thus it would be very unlikely to observe such a peptide in a large-scale experiment after trypsin digestion.

(A)



(B)

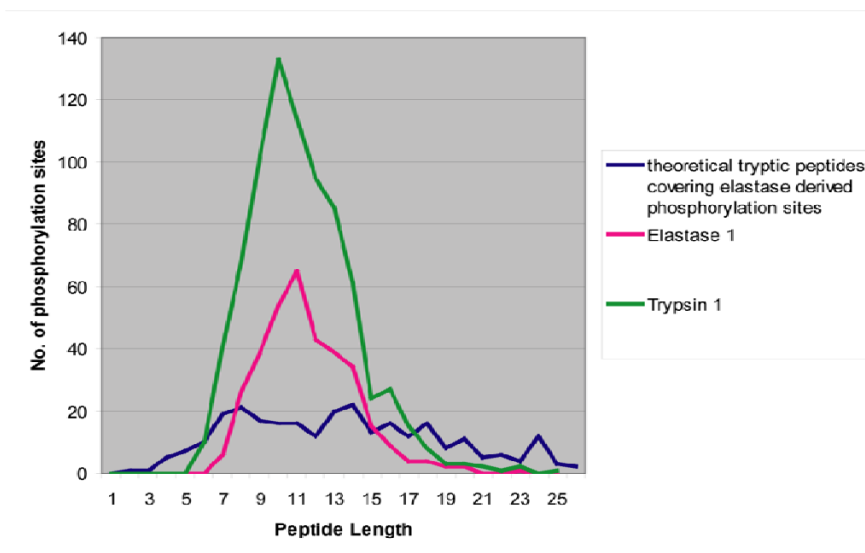


Figure 25. (A) Size distribution of theoretical tryptic peptides covering elastase derived phosphorylation sites. (B) Size distributions in the range between 1 and 25 amino acids of experimentally identified tryptic phosphorylated peptides (green), elastase derived phosphorylated

---

peptides (red), and theoretical tryptic peptides covering elastase derived phosphorylation sites (blue).

### **3.2.10 Testing for undersampling effects**

Finally, we aimed to confirm that the observed low overlap between the trypsin and elastase datasets was indeed a consequence of enzyme complementarity and not a sole undersampling effect due to high sample complexity<sup>143</sup>. To this end, in a second experiment, we analyzed the elastase and trypsin fractions twice (details see materials and methods) and compared the observed overlaps of identified phosphorylation sites from the four datasets (trypsin 2a, trypsin 2b, elastase 2a, and elastase 2b). Figures 26A and 26B show that the overlaps from the repeated trypsin and elastase analyses are about 50% each, whilst the overlap between trypsin and elastase derived datasets is clearly lower (8.1%, figure 26C). Therefore, it can be concluded that the observed low overlap between the phosphorylation sites identified using either trypsin and elastase indeed reflects the complementary cleavage properties of these enzymes.

Combining the results of both experiments, we have identified 1068 phosphorylation sites from tryptic peptides and 467 phosphorylation sites using elastase. The overlap of those two datasets is only 9.9% (139 sites), demonstrating the value of using elastase in addition to trypsin to increase phosphorylation site coverage in complex samples.

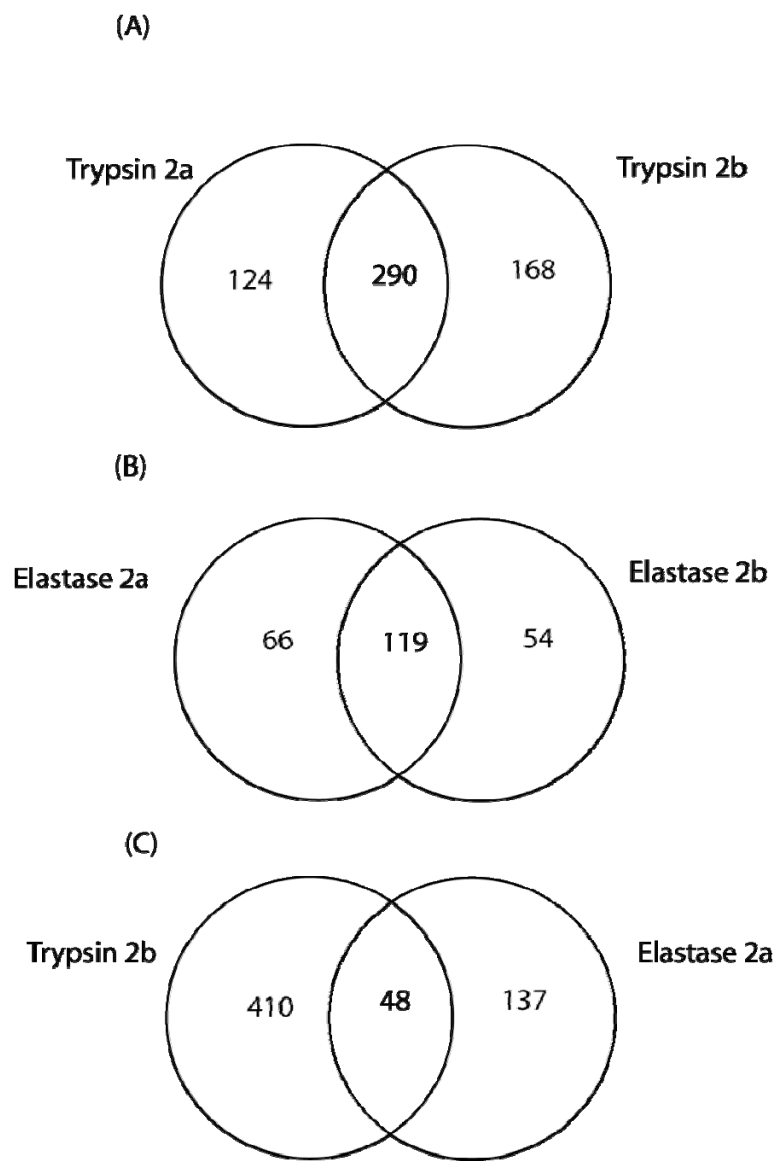


Figure 26. Detected phosphorylation sites for repeated experiments using trypsin or elastase for database searches against the “custom database”.

---

## 4 Discussion

The Plk1 dependent spindle phosphoproteome project demonstrates the use of mass spectrometry based proteomics techniques to generate insight into important biological signaling networks. Using rigorous filtering criteria, we identified in total 4420 unique phosphorylation sites from inhibitor and shRNA experiments (with a false-positive rate (FDR) of 0.89 % (see Results)). Of these, 1943 phosphorylation sites (more than 40%) have not been annotated in Uniprot (version 15.11), and to the best of our knowledge, have not been previously reported in the literature. The large number of confidently identified phosphorylation sites demonstrates the high efficiency of the employed proteomics strategy. Especially, enrichment of phosphorylated peptides by titanium affinity beads combined with highly sensitive LC-MS/MS techniques can be considered as key techniques to reproducibly improve phosphoproteome coverage.

The phosphoproteomes identified with high confidence from inhibitor and shRNA approaches comprise similar numbers of phosphorylation sites, 3526 and 3212, respectively. However, from 702 down-regulated phosphorylation sites, 527 were observed after TAL treatment and only 268 sites after Plk1 depletion by shRNA. Chemical inhibition of Plk1 resulted in a larger number of identified down-regulated phosphorylation sites and stronger down-regulation at individual sites than Plk1 depletion, likely reflecting both incomplete depletion of Plk1 in the shRNA cell line and the potent inhibition of Plk1 by TAL <sup>86</sup>. Nevertheless, although the number of down-regulated sites identified upon Plk1 depletion was lower, more than two thirds of these sites did not overlap with those down-regulated upon TAL treatment (figure 16B), supporting the use of complementary strategies to interfere with Plk1 for better coverage of significantly regulated phosphorylation site identification.

There were 304 human proteins identified with Plk1 dependent phosphorylation sites. 35% of these proteins are related to mitotic spindle function <sup>83, 84</sup>, demonstrating a substantial enrichment for spindle components in our samples.

Among the spindle proteins containing down-regulated phosphorylated sites, those associated with the centromere or KT structure were overrepresented, consistent with an important role for Plk1 in KT function (figure 19). This included the centromeric protein PBIP1, the protein kinase Bub1 and the subunit of the chromosomal passenger complex, INCENP, all known to contribute to KT localization of Plk1 during mitosis <sup>144-146</sup>.



---

Interestingly, members of the nuclear pore complex (NUPs), previously shown to associate with KTs at the onset of mitosis<sup>147</sup>, also showed down-regulated sites after Plk1 inhibition (Figure 19). Furthermore, APC subunits, other mitotic kinases, tubulin subunits and proteins functioning in chromosome structure, appeared also to be regulated by Plk1. Our data set contained known Plk1 substrates, such as Cdc27, INCENP, Kif20A, and TOPOIIA<sup>13, 148</sup>, which validate our experimental strategy.

The analysis of the sequence motifs of confirmed Plk1 phosphorylation sites (see Results) revealed that, in addition to the previously reported D/E-X-S/T-Φ motif, many Plk1 phosphorylation sites conformed to slightly different sequence features, i.e. a N at the -2 position was observed frequently. Detailed alignment analyses and additional peptide spotting experiments suggested the broadened motif L(Φ)-E/N/D(Q)-X-S/T-(Φ) as a Plk1 substrate consensus, thus refining our current knowledge on Plk1 activity.

Interestingly, we found that many Plk1 substrates showed reduced association with the spindle, when Plk1 levels were reduced by shRNA or when Plk1 activity was inhibited by TAL. This observation suggests that phosphorylation by Plk1 is required for the correct localization of many spindle components, thus providing important insight into the functions of Plk1 phosphorylation.

In summary, in the first part of this thesis we have identified 360 unique Plk1-dependent phosphorylation sites on mitotic spindle proteins, including novel substrates, illustrating the complexity of the Plk1-dependent signaling network. Over 100 sites were validated by *in vitro* peptide arrays. Importantly, in line with the distinct localization of Plk1 at the centrosomes, spindle and KTs in early mitosis, and its relevant roles in centrosome maturation, spindle assembly and KT-MT stability, the isolation of purified spindles allowed us to enrich for relevant Plk1 targets and phosphorylation sites on the mitotic spindle. Our comprehensive study of phosphoregulation by Plk1 vastly extends the inventories of direct Plk1 substrates, Plk1-dependent phosphorylation sites on the mitotic spindle apparatus, and downstream effectors of Plk1 in mammalian cells. These data should thus serve as a valuable resource for future research in the field. The strategies developed here for the analysis of the Plk1-dependent proteome and phosphoproteome on the mitotic spindle can in principle be applied to other protein kinases, provided that effective shRNA duplexes and/or specific

---

inhibitors are available. These strategies thus respond to a growing contemporary need for assigning experimental phosphoproteome data sets to specific individual kinases.

The results of the Plk1 project demonstrate that thousands of phosphorylation sites can be identified by mass-spectrometry based phosphoproteome analysis in a robust fashion, thus allowing for comparative SILAC based analyses of considerable biological interest. However, many important phosphorylation sites may still remain undetected due to bias of the employed enrichment strategy towards certain sequence related features. Especially, the specificity of trypsin, the protease employed for in-gel digestion of gel-separated proteins, may favor the detection of phosphorylation sites within tryptic peptides of optimal sizes and physiochemical properties. The detection and sequencing sensitivity of most mass spectrometers is optimal for peptides of 6-20 amino acid lengths so longer or shorter peptides may frequently remain undetected. Trypsin is almost exclusively used in proteomics experiments since it works highly efficient and because its strict cleavage specificity strongly increases the database search efficiency (only tryptic peptides have to be considered).

However, complete sequence coverage of proteins can unfortunately not be achieved employing a single protease. Therefore, in the second part of this thesis we decided to evaluate the use of alternative enzymes to complement trypsin based proteome analyses. Low specific enzymes should be well suited to reach higher sequence coverage since overlapping peptides are frequently generated. In particular, elastase generates mainly medium sized peptides in the mass range between 500 Da and 1500 Da and has the potential advantage of being a good complement to trypsin due to its cleavage low-specificity. To evaluate elastase for phosphoproteome studies, equal amounts of spindle proteins were processed in parallel with trypsin and elastase for phosphorylation sites identification. Obtained MS spectra were searched against the same database whereas trypsin search was under trypsin specificity setting and elastase search was under “no enzyme” setting. Using this strategy, we identified 703 phosphorylation sites identified using trypsin versus 185 phosphorylation sites using elastase (see figure 23). The reduced number of identified phosphorylation sites after elastase digestion compared to trypsin, can be explained by several factors: First, trypsin derived peptides are especially suitable for MS/MS based identification due to protonated amino acids (Arg, Lys) at their C-termini, often resulting in pronounced y-ion series, and the low frequency of internal Arg and Lys residues, known to adversely affect fragmentation patterns.

---

Second, the detection sensitivity after elastase digestion might be hampered by the possible spread of the signal to multiple overlapping peptides. Third, as mentioned in the Results section, the vastly increased search space as a result of the “no enzyme” setting may adversely affect the sensitivity of the MASCOT search engine. As shown in the Results section, the database size has indeed a strong influence on the number of identified phosphorylation sites when searches without enzyme specificity are conducted.

The much higher increase of phosphorylation site identifications for elastase compared to trypsin employing the “custom database” search strategy can be explained by the much larger effect on the search space when using the “no enzyme” setting. Whilst the sensitivity of the MASCOT search algorithm was sufficient to detect a high proportion of PSMs using tryptic search specificity, this sensitivity suffered apparently strongly from the larger search space for the elastase data. Thus, the result for the elastase dataset profits strongly from the more specific search against the “custom database” whereas most correct PSMs of the tryptic dataset were already detected during the SwissProt search. Importantly, the estimated FDRs for these datasets are still low (0.3% for trypsin, 1.2% for elastase), even though it must be noted that these numbers cannot be directly compared to the SwissProt FDRs as FDRs generally depend on search space, even for identical MS/MS spectra. Importantly, however, the search space of the elastase search against the custom database (1030 sequences) is still about seven times larger compared to the trypsin dataset search against SwissProt human (14475 sequences), since the “no enzyme” setting for the elastase search increases the search space by approximately 100 times<sup>111</sup>. Thus, we can conclude that the proposed search strategy is statistically sound, provided that the “custom database” represents a good coverage of the detectable proteins in the samples, as shown in the Results section.

We have demonstrated that the employed search strategy provides an efficient means to increase the number of detectable peptides using low-specificity enzymes. It is conceptually similar to the second-pass search strategy<sup>153</sup>, in which highly specific search criteria are used for the first search, and the coverage of peptides for the identified proteins is increased in the second pass search using more relaxed search criteria (i.e. more variable modifications, wider mass error windows). Even though in our search strategy the “second-pass search” is run with identical search criteria, we also take advantage of the strongly reduced search space to identify more phosphorylation sites.

---

By employing the two-step database search method, the overlap of protein phosphorylation site identification between trypsin and elastase was still low (6.8%), which reflects high complementarity of the elastase and trypsin cleavage specificities. This point was further confirmed by the length-distributions of phosphorylated peptides observed after tryptic and elastase digestions (figure 25). By comparing the length-distributions of tryptic, elastase and the theoretical tryptic peptides covering elastase derived phosphorylation sites, it can be concluded that many of the phosphorylation sites discovered after elastase digestion would be either too small or too big to be detected with high sensitivity after tryptic digestion. In addition to peptide lengths, other sequence related physicochemical properties<sup>112</sup> may further contribute to the low overlap between trypsin and elastase derived phosphorylation sites.

In summary, in the second part of this thesis we have investigated the use of elastase in comparison to trypsin for the phosphoproteome analysis of samples enriched for human mitotic spindle proteins. We found that trypsin was clearly more efficient than elastase in generating detectable peptides for phosphorylation site identification. However, by determining the cleavage preferences of elastase and optimizing the database search strategy, we show that it is possible to robustly improve phosphorylation site identification from low specificity protease digests.

Importantly, the cleavage specificities of trypsin and elastase were found to be highly complementary, providing about 30% additional phosphorylation sites using elastase. Therefore, as already established for studies focused on specific proteins, we propose to use parallel digestions with various specific and unspecific enzymes to approach towards complete sequence coverage in large-scale proteomics studies. To this end, the presented improved search strategy may also be applied to other low-specificity proteases, such as pepsin, proteinase K, or thermolysin.

---

## **5 Material and Methods**

### **5.1 Quantitative identification of Polo-like kinase 1 (Plk1) specific phosphorylation sites on human mitotic spindle protein**

#### **5.1.1 Materials**

Glycine, trifluoroacetic acid (25%), acetic acid, formic acid (98-100%), acetonitrile, methanol and ammonia solution (25%) were obtained from Merck. 2-Propanol was from Carl Roth. Lactic acid (~90%) was ordered from Fluka. NuPAGE Bis-Tris gels (1.0mm thick, 12 lanes per gel) and NuPAGE lithium dodecyl sulfate (LDS, 4 ×) sample buffer were from Invitrogen. DL-Dithiothreitol (DTT, >99%) and iodoacetamide were obtained from Sigma. Proteomics grade trypsin was purchased from Roche (Mannheim, Germany). Titansphere (10 µm) was from GL Sciences Inc. (Japan) and GELoader tips were purchased from Eppendorf. The 3M Empore C8 and C18 disks were obtained from 3M Bioanalytical Technologies (St. Paul, MN).

#### **5.1.2 SILAC media**

DMEM high glucose medium deficient in amino acids arginine and lysine was supplemented with 5% dialysed FCS, penicillin-streptomycin (100 U/mL and 100 µg/mL, respectively), and either unlabelled Arginine.HCl and Lysine.HCl (SILAC light) or L-arginine-U-13C6-15N4.HCl and L-lysine-U-13C6-15N2.HCl (Cambridge Isotope Laboratories) (SILAC heavy) at concentrations of 42 µg/mL (Arginine) and 72 µg/mL (Lysine).

#### **5.1.3 Cell culture and spindle isolation**

Cells were grown at 37 °C in a humidified incubator with a 5 % CO<sub>2</sub> atmosphere. Cells were adapted to the appropriate SILAC medium for at least 6 passages to achieve complete incorporation of the isotopically labeled amino acids. For large-scale mitotic spindle isolation, each population of HeLa S3 labeled cells was propagated to 5 culture triple-flasks, with a total surface of 500 cm<sup>2</sup>. Cells were first pre-synchronized with thymidine (2 mM) for 20 h, then washed twice with PBS, and released from the thymidine block into TAL or Monastrol-containing medium (1µM or 150 µM, respectively), or by induction with 1 µg/ml tetracycline

---

for 36 or 48 h (Plk1 or Eg5 shRNA respectively), in case of the inducible cell lines. Taxol-stabilized mitotic spindles (including kinetochores and centrosomes) were isolated essentially as described previously<sup>93</sup> Briefly, taxol (5 µg/mL) was added to pre-lysis and lysis buffers to stabilize spindles. Cells were then treated with DNases and latrunculin B to partially remove chromosomes and to depolymerize the actin cytoskeleton, respectively. Intermediate filaments were subsequently depolymerized in low-ionic-strength buffer and the mitotic spindles collected by centrifugation. The purity of isolated spindles was evaluated by DIC light microscopy. (performed by Anna Santamaria)

#### **5.1.4 *In vitro* kinase assays**

Plk1 and Aurora A activity was assayed in kinase buffer (50mM Tris-HCl pH7.5, 10 mM MgCl<sub>2</sub>, 1 mM DTT, 100 uM NaF, 10uM Sodium Vanadate) and (20 mM Hepes pH 7.4, 150 mM KCl, 5 mM MnCl<sub>2</sub>, 5 mM NaF and 1mM DTT), respectively, using His-tagged Plk1 and Aurora A, respectively. Casein and MBP were used as model substrates. Reactions were mixed, incubated for 30 min at 30 °C and stopped by the addition of Laemmli sample buffer. Proteins were resolved by SDS-PAGE and gels were dried before being visualized by autoradiography. (performed by Sabine Elowe)

#### **5.1.5 Gel Electrophoresis and In-Gel Digestion**

Enriched spindle proteins were separated by SDS-PAGE and in-gel<sup>154</sup> digested with trypsin. In each experiment, about 400 µg spindle proteins were loaded to one NuPAGE Bis-Tris gel and run for 50 min using a 200-V/100-mA program. Gels were stained with a 1:1 mixture of 0.2% Coomassie blue in MeOH and 20% acidic acid for 40 min, and destained using a mixture of 30% MeOH and 10% acetic acid for 15 min, 10% acetic acid for 1 hour, and 2% acetic acid over night at 4°C. Gels were cut into 18 slices each according to the separation of spindle proteins. After reduction and alkylation, proteins were digested by 15 ng/µl trypsin for 16h. Digested peptides were extracted with 30% AcN/5% formic acid and dried.

#### **5.1.6 Enrichment of Phosphorylated Peptides and Desalting**

Phosphorylated peptides were selectively enriched by titanium dioxide beads with lactic acid as a modifier<sup>74</sup>. A piece of C8 material was plugged at the constricted end of GELoader

---

tip and about 3 mg titanium dioxide beads was transferred to each of the micro-columns. The micro-columns were washed with 40  $\mu$ l of 0.3  $\mu$ g/ $\mu$ l lactic acid in a mixture of 80% ACN and 0.2% TFA. Digested peptides were re-dissolved in 0.3  $\mu$ g/ $\mu$ l lactic acid in a mixture of 80% ACN and 2% TFA, and applied to micro-columns with slow flow rate allowing phosphorylated peptides to bind to the titanium dioxide beads. The micro-columns were subsequently washed with 40  $\mu$ l of 0.3  $\mu$ g/ $\mu$ l lactic acid in a mixture of 80 % ACN and 0.2% TFA and 40  $\mu$ l of a mixture of 80% ACN and 0.2% TFA. Phosphorylated peptides were eluted with 40  $\mu$ l of 0.6% NH<sub>4</sub>OH and 40  $\mu$ l of a mixture of 80% ACN and 0.2% TFA. The eluates were dried and re-dissolved with 0.5% formic acid for LC-MS/MS analysis.

The flow-through fractions were desalted with C18 reversed-phase material prior to LC-MS/MS for protein expression level measurement. Briefly, a piece of C18 material was plugged into a GELoader tip, washed with 20  $\mu$ l of 2-propanol and 20  $\mu$ l of 5% formic acid. The dried flow-through fractions were re-dissolved with 20  $\mu$ l 5% formic acid and applied to the C18 micro-columns. The columns were then washed with 20  $\mu$ l of 5% formic acid and subsequently peptides were eluted with 2 $\times$ 20  $\mu$ l of 50% methanol/2% formic acid. The eluates were dried and re-dissolved with 0.5% formic acid for LC-MS/MS.

### **5.1.7 Nano LC-MS/MS**

The nano LC-MS/MS analysis was performed with a nanoACQUITY ultra performance liquid chromatography (UPLC) system (Waters, UK) connected to a hybrid linear ion trap/Orbitrap tandem mass spectrometer (Thermo Electron, Germany). Dissolved peptides were loaded at 500 nl/min into a pulled and fused silica capillary with an inner diameter of 75  $\mu$ m and a tip of 8  $\mu$ m (New Objective, USA) packed to a length of 12 cm with reversed-phase ReproSil-Pur C18-AQ 3  $\mu$ m resin (Maisch, Germany), and eluted at 200 nl/min by a stepwise 180 min gradient of 0-100% between buffer A (0.2% formic acid in water) and buffer B (0.2% formic acid in acetonitrile).

The LTQ mass spectrometer was operated in a data-dependent MS/MS mode. Survey full scan MS spectra (from m/z 300 to 2000) were acquired in the FT-Orbitrap with a resolution of 60 000 at m/z 400. A maximum of five peptides were sequentially isolated for fragmentation in the linear ion trap using collision induced dissociation (CID). The lock mass option was enabled to improve mass accuracy as described<sup>133</sup>. All spectra were acquired with Xcalibur software.

---

### 5.1.8 Database searching and Data Filtering

MS spectral were searched via MASCOT <sup>123</sup> (version 2.2.0, Matrix Science, London, U.K.) search engine against a composite Human International Protein Index sequence database (IPI, Version 3.48) which was concatenated with its reversed complement <sup>155</sup>. Peak lists were generated using MaxQuant <sup>156</sup> (version 1.0.12.5), which also performs SILAC quantitation, FDR determination, peptide/protein grouping, phosphorylation sites localization probability scoring, and data filtration based on MASCOT search results. Maximum allowed monoisotopic precursor mass error was set to  $\pm 7$  ppm, whereas an accuracy of  $\pm 0.5$  Da was used for MS/MS peaks. Enzyme specificity was set to trypsin and allowing cleavages N-terminally to proline. Carbamidomethylation was set as fixed modification. Oxidation, protein N-terminal acetylation, Arg10, Lys8 and phosphorylation (STY) were considered as variable modifications. A maximum of three labeled amino acids, two missed tryptic cleavages and a minimum of six amino acids peptide length were allowed. A minimum of one unique peptide was required for protein identification. Estimated peptide/protein FDR were calculated as (number of hits in the reversed database / number of hits in the forward database)  $\times 100\%$  <sup>157</sup>. Maximum 5% peptide FDR and 2% protein FDR were allowed for the first-step peptide/protein identification. MASCOT ion score  $\geq 12$  and phosphorylation sites localization probability  $\geq 75\%$  were used for data filtering.

Automated quantitation was accomplished by MaxQuant software. The ratios of each ion pair with different SILAC labeling were normalized by corresponding protein expression levels which were quantitated only with unphosphorylated peptides. Normalized phosphorylated peptide ratios show either an increase higher than 3/2-fold or a decrease lower than 2/3-fold in relative abundance under different SILAC labeling conditions, were considered as significantly down-regulated upon the depletion/inhibition of Plk1 <sup>95</sup>.



---

## **5.2 Phosphoproteome analysis of human mitotic spindle proteins by using low-specificity protease elastase**

### **5.2.1 Materials**

Elastase (from porcine pancreas) and proteomics grade trypsin were purchased from Roche (Mannheim, Germany), bovine serum albumin (>99%), DL-Dithiothreitol (DTT, >99%) and iodoacetamide from Sigma, NuPAGE Bis-Tris gels (1.0mm thick, 12 lanes per gel) and NuPAGE lithium dodecyl sulfate (LDS, 4 ×) sample buffer from Invitrogen, glycine, acetonitrile, trifluoroacetic acid (25%), acetic acid, methanol and ammonia solution (25%) from Merck, and lactic acid (~90%) from Fluka. Titanspheres (10 µm) were obtained from GL Sciences Inc. (Japan), GELoader tips from Eppendorf, and the 3M Empore C8 and C18 disks from 3M Bioanalytical Technologies (St. Paul, MN).

### **5.2.2 Cell Culture and Mitotic Spindle preparation**

Human mitotic spindles were isolated and enriched from 9×10<sup>7</sup> HeLa S3 cells as described in detail<sup>93</sup>, except that thymidine (5 mM for 21 hours) was used to synchronize cells in S-phase prior to the mitotic nocodazole block. In short, the synchronized mitotic cells were released after shake-off, and taxol was added to stabilize microtubules when most cells had reached metaphase. After lysis, DNase was added to degrade chromosomes, whereas intermediated filaments were depolymerized in low ionic strength buffer. Finally, mitotic spindles were collected by centrifugation.

### **5.2.3 Gel Electrophoresis and In-Gel Digestion**

In both experiments, isolated spindle proteins were split into two equal fractions, separated in parallel by 1D gel electrophoresis, and in-gel digested by trypsin or elastase, respectively. In experiment 1, for each gel 200 µg spindle proteins were suspended in 241 µl 0.1 M glycine buffer and mixed with 80 µl 4 × LDS sample buffer and 16 µl of 0.5 M DTT and heated for 10 min at 75°C. Spindle proteins were then separated on NuPAGE Bis-Tris gels for 50 min at 200-V/100mA. Gels were stained using a 1:1 mixture of 0.2% Coomassie blue in 50% MeOH and 20% acetic acid for 40 min, and destained using a mixture of 30% MeOH and 10% acetic acid for 15 min, 10% acetic acid for 1 hour, and 2% acetic acid over night at 4°C. Gels were cut into 18 slices each. Spindle proteins were then in-gel digested by

---

trypsin<sup>154</sup> or elastase at concentrations of 15 ng/μl for 16 hours. After digestion, peptides were extracted and dried in a vacuum centrifuge. These samples were named trypsin 1 and elastase 1. In experiment 2, 350 micrograms spindle proteins were applied to each gel. The samples were processed as in experiment 1, except that they were split prior to LC-MS analysis for a replicate analysis. These samples were named trypsin 2a, trypsin 2b, elastase 2a and elastase 2b, respectively.

#### **5.2.4 Phosphorylated Peptide Enrichment and Sample Desalting**

Titanium dioxide beads were used to selectively enrich for phosphorylated peptides with lactic acid as a modifier<sup>74</sup>. About 3 mg of titanium dioxide beads were transferred to a GELoader tip (plugged at the constricted end by a small piece of C8 material) and washed with 40 μl of 0.3 μg/μl lactic acid in a mixture of 80% ACN and 0.2% TFA. Digested peptides were dissolved in 20 μl of 0.3 μg/μl lactic acid in a mixture of 80% ACN and 2% TFA, and applied to micro-columns allowing phosphorylated peptides to bind to the titanium dioxide phase. Micro-columns were then washed with 40 μl of 0.3 μg/μl lactic acid in a mixture of 80 % ACN and 0.2% TFA and 40 μl of a mixture of 80% ACN and 0.2% TFA. Finally, phosphorylated peptides were eluted slowly with 40 μl of 0.6% NH<sub>4</sub>OH and subsequently 40 μl of a mixture of 80% ACN and 0.2% TFA. Eluates and flow-through fractions were desalted with C18 reversed-phase material prior to LC-MS/MS analysis essentially as described<sup>158</sup>.

#### **5.2.5 Nano LC-MS/MS**

The digested peptide mixtures were chromatographically separated with a nanoACQUITY ultra performance liquid chromatography (UPLC) system (Waters, UK) connected to a hybrid linear ion trap/Orbitrap tandem mass spectrometer (Thermo Electron, Germany). Peptides dissolved in 0.5% formic acid were loaded into a 14 cm pulled and fused silica capillary with an inner diameter of 75 μm and a tip of 8 μm (New Objective, USA) packed with reversed-phase ReproSil-Pur C18-AQ 3 μm resin (Maisch, Germany). Peptides were separated and eluted by a stepwise 180 min gradient of 0-100% between buffer A (0.2% formic acid in water) and buffer B (0.2% formic acid in acetonitrile). The mass spectrometer was operated in data-dependent MS/MS mode. Survey full scan MS spectra (from m/z 300 to 2000) were acquired in the FT-Orbitrap with a resolution of 60 000 at m/z 400. A maximum

---

of five peptides were sequentially isolated for fragmentation in the linear ion trap using collision induced dissociation. The Orbitrap lock mass feature was applied to improve mass accuracy as described<sup>133</sup>.

### **5.2.6 Data Processing and Analysis**

Peak lists were generated using MaxQuant (version 1.0.9.4)<sup>156</sup>, which also features spectra re-calibration based on high scoring peptides, phosphorylation site scoring within the sequences of identified phosphorylated peptides, false-discovery rate determination, and data filtering options after import of MASCOT result files. MASCOT (version 2.2.0, Matrix Science, London, U.K.)<sup>123</sup> was used for peptide and protein identifications. For initial searches, the precursor mass tolerance was set to  $\pm 7$  ppm, whereas an accuracy of  $\pm 0.3$  Da was used for MS/MS spectra. The MS mass error window was subsequently optimized during data analysis (see below). Carbamidomethylation was set as fixed modification, whereas oxidation, acetylation (protein N-terminal), and phosphorylation (STY) were considered as variable modifications. Trypsin/p was set as enzyme for tryptic peptides, and “no enzyme” was set for elastase derived peptides. Searches were performed against SwissProt human (release 54.4) and “custom databases” (see Results), which were concatenated with a reverse database<sup>155, 157</sup>. Peptide/protein false-discovery rates (FDR) were calculated as (number of hits in the reversed database / number of hits in the forward database)  $\times 100\%$ <sup>157</sup>.

---

## 6 Abbreviations

AC	Alternating current
ACN	Acetonitrile
APC/C	Anaphase-promoting complex/cyclosome
AQUA	Absolute quantitation
BSA	Bovine Serum Albumin
CID	collision induced dissociation
CDITs	Culture-derived isotope tags
Cdk1	Cyclin-dependent kinase 1
DC	Direct current
DHB	2,5-dihydroxy benzoic acid
DFA	Discriminant Function Analysis
DTT	DL-Dithiothreitol
ESI	Electrospray ionization
FACS	Flow cytometry
FDR	False discovery rate
FTICR	Fourier transform ion cyclotron resonance
ICAT	Isotope-coded affinity tags

---

ICPL	Isotope-coded protein label
IF	Immunofluorescence
IMAC	Immobilized Metal Affinity Chromatography
iTRAQ	Isotope tags for relative and absolute quantification
KTs	Kinetochores
LC	Liquid chromatography
LDS	Lithium dodecyl sulfate
LTQ	Linear quadrupole ion trap
m/z	Mass to charge ratio
MA	Monastrol
MALDI	Matrix-assisted laser desorption/ionization
MAPs	Microtubule associated proteins
MBP	Myelin basic protein
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MTs	Microtubules
NUPs	Nuclear pore complex
PBDs	Polo-box domains

---

Plk1	Polo-like kinase 1
PSMs	Peptide-spectrum-matches
RF	Radio frequency
SCX	Strong Cation exchange chromatography
SILAC	Stable isotope labeling by amino acid in cell culture
Tet	Tetracycline
TFA	Trifluoroacetic acid
TOF	Time-of-flight
UPLC	Ultra performance liquid chromatography
XIC	Extracted ion current

---

## 7 References

- (1) Doxsey, S. *Nat Rev Mol Cell Biol* **2001**, 2, 688-698.
- (2) Meraldi, P.; Nigg, E. A. *FEBS Lett* **2002**, 521, 9-13.
- (3) Pines, J.; Rieder, C. L. *Nat Cell Biol* **2001**, 3, E3-6.
- (4) Hunt, T. *Semin Cell Biol* **1991**, 2, 213-222.
- (5) Nigg, E. A. *Bioessays* **1995**, 17, 471-480.
- (6) Morgan, D. O. *Annu Rev Cell Dev Biol* **1997**, 13, 261-291.
- (7) Peters, J. M. *Nat Rev Mol Cell Biol* **2006**, 7, 644-656.
- (8) Berdnik, D.; Knoblich, J. A. *Curr Biol* **2002**, 12, 640-647.
- (9) Hachet, V.; Canard, C.; Gonczy, P. *Dev Cell* **2007**, 12, 531-541.
- (10) Carmena, M.; Earnshaw, W. C. *Nat Rev Mol Cell Biol* **2003**, 4, 842-854.
- (11) Giet, R.; Glover, D. M. *J Cell Biol* **2001**, 152, 669-682.
- (12) Archambault, V.; Glover, D. M. *Nat Rev Mol Cell Biol* **2009**, 10, 265-275.
- (13) Petronczki, M.; Lenart, P.; Peters, J. M. *Dev Cell* **2008**, 14, 646-659.
- (14) Barr, F. A.; Sillje, H. H.; Nigg, E. A. *Nat Rev Mol Cell Biol* **2004**, 5, 429-440.
- (15) Elia, A. E.; Cantley, L. C.; Yaffe, M. B. *Science* **2003**, 299, 1228-1231.
- (16) Rauh, N. R.; Schmidt, A.; Bormann, J.; Nigg, E. A.; Mayer, T. U. *Nature* **2005**, 437, 1048-1052.
- (17) Fabbro, M.; Zhou, B. B.; Takahashi, M.; Sarcevic, B.; Lal, P.; Graham, M. E.; Gabrielli, B. G.; Robinson, P. J.; Nigg, E. A.; Ono, Y.; Khanna, K. K. *Dev Cell* **2005**, 9, 477-488.
- (18) Neef, R.; Preisinger, C.; Sutcliffe, J.; Kopajtich, R.; Nigg, E. A.; Mayer, T. U.; Barr, F. A. *J Cell Biol* **2003**, 162, 863-875.
- (19) Glover, D. M. *Oncogene* **2005**, 24, 230-237.
- (20) Rape, M. *Curr Biol* **2007**, 17, R506-508.
- (21) Sumara, I.; Vorlaufer, E.; Stukenberg, P. T.; Kelm, O.; Redemann, N.; Nigg, E. A.; Peters, J. M. *Mol Cell* **2002**, 9, 515-525.
- (22) Uhlmann, F. *Exp Cell Res* **2004**, 296, 80-85.
- (23) Moshe, Y.; Boulaire, J.; Pagano, M.; Hershko, A. *Proc Natl Acad Sci U S A* **2004**, 101, 7937-7942.
- (24) Cravatt, B. F.; Simon, G. M.; Yates, J. R., 3rd *Nature* **2007**, 450, 991-1000.
- (25) Aebersold, R.; Mann, M. *Nature* **2003**, 422, 198-207.
- (26) Steen, H.; Mann, M. *Nat Rev Mol Cell Biol* **2004**, 5, 699-711.
- (27) Waanders, L. F.; Hanke, S.; Mann, M. *J Am Soc Mass Spectrom* **2007**, 18, 2058-2064.
- (28) Siuti, N.; Kelleher, N. L. *Nat Methods* **2007**, 4, 817-821.
- (29) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, 246, 64-71.
- (30) Karas, M.; Hillenkamp, F. *Anal Chem* **1988**, 60, 2299-2301.
- (31) Scigelova, M.; Makarov, A. *Proteomics* **2006**, 6 Suppl 2, 16-21.
- (32) Wilm, M.; Mann, M. *Anal Chem* **1996**, 68, 1-8.

- 
- (33) Katta, V.; Chowdhury, S. K.; Chait, B. T. *Anal Chem* **1991**, *63*, 174-178.
- (34) Nielsen, M. L. *Ph.D. thesis* **2007**.
- (35) Mora, J. F.; Van Berkel, G. J.; Enke, C. G.; Cole, R. B.; Martinez-Sanchez, M.; Fenn, J. B. *J Mass Spectrom* **2000**, *35*, 939-952.
- (36) Schwartz, J. C.; Senko, M. W.; Syka, J. E. *J Am Soc Mass Spectrom* **2002**, *13*, 659-669.
- (37) Hoffmann, E. d. **2007**.
- (38) Watson, J. T. **2008**.
- (39) Gorshkov, M. V.; Zubarev, R. A. *Rapid Commun Mass Spectrom* **2005**, *19*, 3755-3758.
- (40) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R. *J Mass Spectrom* **2005**, *40*, 430-443.
- (41) Makarov, A.; Denisov, E.; Lange, O.; Horning, S. *J Am Soc Mass Spectrom* **2006**, *17*, 977-982.
- (42) Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S. *Anal Chem* **2006**, *78*, 2113-2120.
- (43) Yates, J. R.; Cociorva, D.; Liao, L.; Zabrouskov, V. *Anal Chem* **2006**, *78*, 493-500.
- (44) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Mol Cell Proteomics* **2002**, *1*, 376-386.
- (45) Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B. *Anal Bioanal Chem* **2007**, *389*, 1017-1031.
- (46) Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. *Anal Chem* **2001**, *73*, 2836-2842.
- (47) Reynolds, K. J.; Yao, X.; Fenselau, C. *J Proteome Res* **2002**, *1*, 27-33.
- (48) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat Biotechnol* **1999**, *17*, 994-999.
- (49) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlett-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. *Mol Cell Proteomics* **2004**, *3*, 1154-1169.
- (50) Pierce, A.; Unwin, R. D.; Evans, C. A.; Griffiths, S.; Carney, L.; Zhang, L.; Jaworska, E.; Lee, C. F.; Blinco, D.; Okoniewski, M. J.; Miller, C. J.; Bitton, D. A.; Spooner, E.; Whetton, A. D. *Mol Cell Proteomics* **2008**, *7*, 853-863.
- (51) Schmidt, A.; Kellermann, J.; Lottspeich, F. *Proteomics* **2005**, *5*, 4-15.
- (52) Olsen, J. V.; Andersen, J. R.; Nielsen, P. A.; Nielsen, M. L.; Figeys, D.; Mann, M.; Wisniewski, J. R. *Mol Cell Proteomics* **2004**, *3*, 82-92.
- (53) Wisniewski, J. R. *Arch Pathol Lab Med* **2008**, *132*, 1566-1569.
- (54) Ishihama, Y.; Sato, T.; Tabata, T.; Miyamoto, N.; Sagane, K.; Nagasu, T.; Oda, Y. *Nat Biotechnol* **2005**, *23*, 617-621.
- (55) Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P. *Proc Natl Acad Sci U S A* **2003**, *100*, 6940-6945.
- (56) Ong, S. E.; Mann, M. *Nat Chem Biol* **2005**, *1*, 252-262.



- 
- (57) Mann, M. *Nat Rev Mol Cell Biol* **2006**, 7, 952-958.
- (58) Ong, S. E.; Foster, L. J.; Mann, M. *Methods* **2003**, 29, 124-130.
- (59) Boersema, P. J.; Raijmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A. J. *Nat Protoc* **2009**, 4, 484-494.
- (60) Ong, S. E.; Kratchmarova, I.; Mann, M. *J Proteome Res* **2003**, 2, 173-181.
- (61) Van Hoof, D.; Pinkse, M. W.; Oostwaard, D. W.; Mummery, C. L.; Heck, A. J.; Krijgsveld, J. *Nat Methods* **2007**, 4, 677-678.
- (62) Gruhler, A.; Olsen, J. V.; Mohammed, S.; Mortensen, P.; Faergeman, N. J.; Mann, M.; Jensen, O. N. *Mol Cell Proteomics* **2005**, 4, 310-327.
- (63) Bendall, S. C.; Hughes, C.; Stewart, M. H.; Doble, B.; Bhatia, M.; Lajoie, G. A. *Mol Cell Proteomics* **2008**, 7, 1587-1597.
- (64) Olsen, J. V.; Blagoev, B.; Gnäd, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. *Cell* **2006**, 127, 635-648.
- (65) Wang, X.; Huang, L. *Mol Cell Proteomics* **2008**, 7, 46-57.
- (66) Pan, C.; Gnäd, F.; Olsen, J. V.; Mann, M. *Proteomics* **2008**, 8, 4534-4546.
- (67) Bonaldi, T.; Straub, T.; Cox, J.; Kumar, C.; Becker, P. B.; Mann, M. *Mol Cell* **2008**, 31, 762-772.
- (68) Hubbard, M. J.; Cohen, P. *Trends Biochem Sci* **1993**, 18, 172-177.
- (69) Hunter, T. *Cell* **2000**, 100, 113-127.
- (70) Reinders, J.; Sickmann, A. *Proteomics* **2005**, 5, 4052-4061.
- (71) Thingholm, T. E.; Jensen, O. N.; Larsen, M. R. *Proteomics* **2009**, 9, 1451-1468.
- (72) Pinkse, M. W.; Uitto, P. M.; Hilhorst, M. J.; Ooms, B.; Heck, A. J. *Anal Chem* **2004**, 76, 3935-3943.
- (73) Larsen, M. R.; Thingholm, T. E.; Jensen, O. N.; Roepstorff, P.; Jorgensen, T. J. *Mol Cell Proteomics* **2005**, 4, 873-886.
- (74) Jensen, S. S.; Larsen, M. R. *Rapid Commun Mass Spectrom* **2007**, 21, 3635-3645.
- (75) Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. *Proc Natl Acad Sci U S A* **2004**, 101, 12130-12135.
- (76) Ballif, B. A.; Villen, J.; Beausoleil, S. A.; Schwartz, D.; Gygi, S. P. *Mol Cell Proteomics* **2004**, 3, 1093-1101.
- (77) Cheeseman, I. M.; Desai, A. *Nat Rev Mol Cell Biol* **2008**, 9, 33-46.
- (78) Musacchio, A.; Salmon, E. D. *Nat Rev Mol Cell Biol* **2007**, 8, 379-393.
- (79) Llamazares, S.; Moreira, A.; Tavares, A.; Girdham, C.; Spruce, B. A.; Gonzalez, C.; Karess, R. E.; Glover, D. M.; Sunkel, C. E. *Genes Dev* **1991**, 5, 2153-2165.
- (80) Plyte, S.; Musacchio, A. *Curr Biol* **2007**, 17, R280-283.
- (81) Nakajima, H.; Toyoshima-Morimoto, F.; Taniguchi, E.; Nishida, E. *J Biol Chem* **2003**, 278, 25277-25280.
- (82) Schreiber, T. B.; Mausbacher, N.; Breitkopf, S. B.; Grundner-Culemann, K.; Daub, H. *Proteomics* **2008**, 8, 4416-4432.
- (83) Sauer, G.; Korner, R.; Hanisch, A.; Ries, A.; Nigg, E. A.; Sillje, H. H. *Mol Cell Proteomics* **2005**, 4, 35-43.

- 
- (84) Nousiainen, M.; Sillje, H. H.; Sauer, G.; Nigg, E. A.; Korner, R. *Proc Natl Acad Sci U S A* **2006**, *103*, 5391-5396.
- (85) Dephoure, N.; Zhou, C.; Villen, J.; Beausoleil, S. A.; Bakalarski, C. E.; Elledge, S. J.; Gygi, S. P. *Proc Natl Acad Sci U S A* **2008**, *105*, 10762-10767.
- (86) Santamaria, A.; Neef, R.; Eberspacher, U.; Eis, K.; Husemann, M.; Mumberg, D.; Prechtl, S.; Schulze, V.; Siemeister, G.; Wortmann, L.; Barr, F. A.; Nigg, E. A. *Mol Biol Cell* **2007**, *18*, 4024-4036.
- (87) Lane, H. A.; Nigg, E. A. *J Cell Biol* **1996**, *135*, 1701-1713.
- (88) Blangy, A.; Lane, H. A.; d'Herin, P.; Harper, M.; Kress, M.; Nigg, E. A. *Cell* **1995**, *83*, 1159-1169.
- (89) Matsumura, S.; Toyoshima, F.; Nishida, E. *J Biol Chem* **2007**, *282*, 15217-15227.
- (90) Lenart, P.; Petronczki, M.; Steegmaier, M.; Di Fiore, B.; Lipp, J. J.; Hoffmann, M.; Rettig, W. J.; Kraut, N.; Peters, J. M. *Curr Biol* **2007**, *17*, 304-315.
- (91) Elowe, S.; Hummer, S.; Uldschmid, A.; Li, X.; Nigg, E. A. *Genes Dev* **2007**, *21*, 2205-2219.
- (92) Mayer, T. U.; Kapoor, T. M.; Haggarty, S. J.; King, R. W.; Schreiber, S. L.; Mitchison, T. J. *Science* **1999**, *286*, 971-974.
- (93) Sillje, H. H.; Nigg, E. A. *Methods* **2006**, *38*, 25-28.
- (94) Malik, R.; Lenobel, R.; Santamaria, A.; Ries, A.; Nigg, E. A.; Korner, R. *J Proteome Res* **2009**.
- (95) Bose, R.; Molina, H.; Patterson, A. S.; Bitok, J. K.; Periaswamy, B.; Bader, J. S.; Pandey, A.; Cole, P. A. *Proc Natl Acad Sci U S A* **2006**, *103*, 9773-9778.
- (96) Kolkman, A.; Daran-Lapujade, P.; Fullaondo, A.; Olsthoorn, M. M.; Pronk, J. T.; Slijper, M.; Heck, A. J. *Mol Syst Biol* **2006**, *2*, 2006 0026.
- (97) Tang, L. Y.; Deng, N.; Wang, L. S.; Dai, J.; Wang, Z. L.; Jiang, X. S.; Li, S. J.; Li, L.; Sheng, Q. H.; Wu, D. Q.; Li, L.; Zeng, R. *Mol Cell Proteomics* **2007**, *6*, 1952-1967.
- (98) Nishizuka, S.; Charboneau, L.; Young, L.; Major, S.; Reinhold, W. C.; Waltham, M.; Kouros-Mehr, H.; Bussey, K. J.; Lee, J. K.; Espina, V.; Munson, P. J.; Petricoin, E., 3rd; Liotta, L. A.; Weinstein, J. N. *Proc Natl Acad Sci U S A* **2003**, *100*, 14229-14234.
- (99) Haren, L.; Stearns, T.; Luders, J. *PLoS One* **2009**, *4*, e5976.
- (100) Varis, A.; Salmela, A. L.; Kallio, M. J. *Chromosoma* **2006**, *115*, 288-295.
- (101) Zhu, X.; Chang, K. H.; He, D.; Mancini, M. A.; Brinkley, W. R.; Lee, W. H. *J Biol Chem* **1995**, *270*, 19545-19550.
- (102) De Luca, M.; Lavia, P.; Guarguaglini, G. *Cell Cycle* **2006**, *5*, 296-303.
- (103) Hanisch, A.; Wehner, A.; Nigg, E. A.; Sillje, H. H. *Mol Biol Cell* **2006**, *17*, 448-459.
- (104) Eckerdt, F.; Pascreau, G.; Phistry, M.; Lewellyn, A. L.; Depaoli-Roach, A. A.; Maller, J. L. *Cell Cycle* **2009**, *8*.
- (105) Chan, E. H.; Santamaria, A.; Sillje, H. H.; Nigg, E. A. *Chromosoma* **2008**, *117*, 457-469.
- (106) Ferrari, S.; Marin, O.; Pagano, M. A.; Meggio, F.; Hess, D.; El-Shemerly, M.; Krystyniak, A.; Pinna, L. A. *Biochem J* **2005**, *390*, 293-302.

- 
- (107) Miller, M. L.; Jensen, L. J.; Diella, F.; Jorgensen, C.; Tinti, M.; Li, L.; Hsiung, M.; Parker, S. A.; Bordeaux, J.; Sicheritz-Ponten, T.; Olhovsky, M.; Pasculescu, A.; Alexander, J.; Knapp, S.; Blom, N.; Bork, P.; Li, S.; Cesareni, G.; Pawson, T.; Turk, B. E.; Yaffe, M. B.; Brunak, S.; Linding, R. *Sci Signal* **2008**, *1*, ra2.
- (108) Kufer, T. A.; Sillje, H. H.; Korner, R.; Gruss, O. J.; Meraldi, P.; Nigg, E. A. *J Cell Biol* **2002**, *158*, 617-623.
- (109) Wilkins, M. R. W.; K. L.; Appel, R. D.; Hochstrasser, D. F. (eds.) *Proteome Research: New Frontiers in Functional Genomics*; Springer: Berlin, 1997.
- (110) Williams, K. L. *Electrophoresis* **1999**, *20*, 678-688.
- (111) Olsen, J. V.; Ong, S. E.; Mann, M. *Mol Cell Proteomics* **2004**, *3*, 608-614.
- (112) Mallick, P.; Schirle, M.; Chen, S. S.; Flory, M. R.; Lee, H.; Martin, D.; Ranish, J.; Raught, B.; Schmitt, R.; Werner, T.; Kuster, B.; Aebersold, R. *Nat Biotechnol* **2007**, *25*, 125-131.
- (113) Peng, J.; Gygi, S. P. *J Mass Spectrom* **2001**, *36*, 1083-1091.
- (114) McLachlin, D. T.; Chait, B. T. *Curr Opin Chem Biol* **2001**, *5*, 591-602.
- (115) Mann, M.; Ong, S. E.; Gronborg, M.; Steen, H.; Jensen, O. N.; Pandey, A. *Trends Biotechnol* **2002**, *20*, 261-268.
- (116) Mann, M.; Jensen, O. N. *Nat Biotechnol* **2003**, *21*, 255-261.
- (117) Hunter, T. *Cell* **1995**, *80*, 225-236.
- (118) Cohen, P. *Eur J Biochem* **2001**, *268*, 5001-5010.
- (119) Ficarro, S. B.; McClelland, M. L.; Stukenberg, P. T.; Burke, D. J.; Ross, M. M.; Shabanowitz, J.; Hunt, D. F.; White, F. M. *Nat Biotechnol* **2002**, *20*, 301-305.
- (120) Posewitz, M. C.; Tempst, P. *Anal Chem* **1999**, *71*, 2883-2892.
- (121) Bodenmiller, B.; Mueller, L. N.; Mueller, M.; Domon, B.; Aebersold, R. *Nat Methods* **2007**, *4*, 231-237.
- (122) Yates, J. R., 3rd; Eng, J. K.; McCormack, A. L.; Schieltz, D. *Anal Chem* **1995**, *67*, 1426-1436.
- (123) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551-3567.
- (124) Mortz, E.; O'Connor, P. B.; Roepstorff, P.; Kelleher, N. L.; Wood, T. D.; McLafferty, F. W.; Mann, M. *Proc Natl Acad Sci U S A* **1996**, *93*, 8264-8267.
- (125) Kelleher, N. L. *Anal Chem* **2004**, *76*, 197A-203A.
- (126) Choudhary, G.; Wu, S. L.; Shieh, P.; Hancock, W. S. *J Proteome Res* **2003**, *2*, 59-67.
- (127) Schlosser, A.; Vanselow, J. T.; Kramer, A. *Anal Chem* **2005**, *77*, 5243-5250.
- (128) Molina, H.; Horn, D. M.; Tang, N.; Mathivanan, S.; Pandey, A. *Proc Natl Acad Sci U S A* **2007**, *104*, 2199-2204.
- (129) Schlosser, A.; Pipkorn, R.; Bossemeyer, D.; Lehmann, W. D. *Anal Chem* **2001**, *73*, 170-176.
- (130) Schlosser, A.; Bodem, J.; Bossemeyer, D.; Grummt, I.; Lehmann, W. D. *Proteomics* **2002**, *2*, 911-918.
- (131) Getie, M.; Schmelzer, C. E.; Neubert, R. H. *Proteins* **2005**, *61*, 649-657.

- 
- (132) Kapp, E. A.; Schutz, F.; Connolly, L. M.; Chakel, J. A.; Meza, J. E.; Miller, C. A.; Fenyo, D.; Eng, J. K.; Adkins, J. N.; Omenn, G. S.; Simpson, R. J. *Proteomics* **2005**, *5*, 3475-3490.
- (133) Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. *Mol Cell Proteomics* **2005**, *4*, 2010-2021.
- (134) Wittmann, T.; Hyman, A.; Desai, A. *Nat Cell Biol* **2001**, *3*, E28-34.
- (135) Gadde, S.; Heald, R. *Curr Biol* **2004**, *14*, R797-805.
- (136) Downing, K. H.; Nogales, E. *Curr Opin Struct Biol* **1998**, *8*, 785-791.
- (137) Nogales, E. *Annu Rev Biochem* **2000**, *69*, 277-302.
- (138) Trinkle-Mulcahy, L.; Lamond, A. I. *Curr Opin Cell Biol* **2006**, *18*, 623-631.
- (139) Nigg, E. A. *Nat Rev Mol Cell Biol* **2001**, *2*, 21-32.
- (140) Mecham, R. P.; Broekelmann, T. J.; Fliszar, C. J.; Shapiro, S. D.; Welgus, H. G.; Senior, R. M. *J Biol Chem* **1997**, *272*, 18071-18076.
- (141) Barbara G. Tabachnick, L. S. F. *Discriminant function analysis. using multivariate statistics*, 4 ed.; Allyn & Bacon: Boston, 2001.
- (142) Resing, K. A.; Meyer-Arendt, K.; Mendoza, A. M.; Aveline-Wolf, L. D.; Jonscher, K. R.; Pierce, K. G.; Old, W. M.; Cheung, H. T.; Russell, S.; Wattawa, J. L.; Goehle, G. R.; Knight, R. D.; Ahn, N. G. *Anal Chem* **2004**, *76*, 3556-3568.
- (143) Garza, S.; Moini, M. *Anal Chem* **2006**, *78*, 7309-7316.
- (144) Qi, W.; Tang, Z.; Yu, H. *Mol Biol Cell* **2006**, *17*, 3705-3716.
- (145) Kang, Y. H.; Park, J. E.; Yu, L. R.; Soung, N. K.; Yun, S. M.; Bang, J. K.; Seong, Y. S.; Yu, H.; Garfield, S.; Veenstra, T. D.; Lee, K. S. *Mol Cell* **2006**, *24*, 409-422.
- (146) Goto, H.; Kiyono, T.; Tomono, Y.; Kawajiri, A.; Urano, T.; Furukawa, K.; Nigg, E. A.; Inagaki, M. *Nat Cell Biol* **2006**, *8*, 180-187.
- (147) Guttinger, S.; Laurell, E.; Kutay, U. *Nat Rev Mol Cell Biol* **2009**, *10*, 178-191.
- (148) Li, H.; Wang, Y.; Liu, X. *J Biol Chem* **2008**, *283*, 6209-6221.
- (149) Kraft, C.; Herzog, F.; Gieffers, C.; Mechtler, K.; Hagting, A.; Pines, J.; Peters, J. M. *Embo J* **2003**, *22*, 6598-6609.
- (150) Brar, G. A.; Kiburz, B. M.; Zhang, Y.; Kim, J. E.; White, F.; Amon, A. *Nature* **2006**, *441*, 532-536.
- (151) Lowery, D. M.; Clauser, K. R.; Hjerrild, M.; Lim, D.; Alexander, J.; Kishi, K.; Ong, S. E.; Gammeltoft, S.; Carr, S. A.; Yaffe, M. B. *Embo J* **2007**, *26*, 2262-2273.
- (152) Burkard, M. E.; Maciejowski, J.; Rodriguez-Bravo, V.; Repka, M.; Lowery, D. M.; Clauser, K. R.; Zhang, C.; Shokat, K. M.; Carr, S. A.; Yaffe, M. B.; Jallepalli, P. V. *PLoS Biol* **2009**, *7*, e1000111.
- (153) Jensen, O. N.; Podtelejnikov, A. V.; Mann, M. *Anal Chem* **1997**, *69*, 4741-4750.
- (154) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. *Anal Chem* **1996**, *68*, 850-858.
- (155) Elias, J. E.; Gygi, S. P. *Nat Methods* **2007**, *4*, 207-214.
- (156) Graumann, J.; Hubner, N. C.; Kim, J. B.; Ko, K.; Moser, M.; Kumar, C.; Cox, J.; Scholer, H.; Mann, M. *Mol Cell Proteomics* **2008**, *7*, 672-683.
- (157) Choi, H.; Nesvizhskii, A. I. *J Proteome Res* **2008**, *7*, 47-50.

---

(158) Rappsilber, J.; Ishihama, Y.; Mann, M. *Anal Chem* **2003**, 75, 663-670.

## 8 Appendix

Part II: Key features of identified down-regulated phosphorylation sites on spindle proteins upon Plk1 inactivation.

Gene Names	Position	Local. Prob.	Spotting result	Modified Sequence
ASPM	1825	1	positive	_QQS(ph)IAALK_
ASPM	480	0.507887	negative	_FSAVQDISS(ph)HSHNK_
ASPM	277	0.526316	NC	_AVT(ph)ETSFNSVNVNGQR_
ASPM	283	0.648605	negative	_AVTETSFNS(ph)VNVNGQR_
ASPM	479	0.844812	negative	_FSAVQDIS(ph)SHSHNK_
ASPM	355	0.854537	positive	_DNSQPVHLES(ph)TIAHEIYQK_
ASPM	356	0.889526	positive	_DNSQPVHLEST(ph)IAHEIYQK_
ASPM	279	0.975622	negative	_AVTET(ph)SFNSVNVNGQR_
ASPM	263	0.99979	negative	_ELLNVS(ph)ANVSK_
ASPM	267	0.999981	positive	_ELLNVSANVS(ph)K_
ASPM	280	0.999992	positive	_AVTETS(ph)FNSVNVNGQR_
ASPM	463	0.999995	negative	_SNYY(ph)FIK_
ASPM	270	1	positive	_VS(ph)FNEK_
ASPM	3426	0.314908	positive	_NS(ph)SISIPFIPETPVR_
ASPM	3443	0.985525	negative	_TRIVS(ph)R_
ASPM	3425	0.990031	negative	_NS(ph)SISIPFIPETPVR_
ASPM	178	0.999243	positive	_T(ph)FSVSQK_
BLM	336	0.333333	negative	_EDVLS(ph)TSKDLLSKPEK_
BLM	337	0.483386	negative	_KEDVLST(ph)SK_
BLM	338	0.483386	negative	_KEDVLST(ph)SK_
BLM	1083	0.954642	negative	_DVTDDVKS(ph)IVR_
CDCA8	154	0.974926	NC	_TQS(ph)IQGK_
CDCA8	152	0.980595	negative	_RT(ph)QSIQGK_
CDCA8	169	0.999799	negative	_ANT(ph)VTPAVGR_
CDCA8	180	0.999981	positive	_LEVS(ph)M(ox)VKPTPGLTPR_
CDCA8	171	0.999983	negative	_ANTVT(ph)PAVGR_
CDCA8	106	1	negative	_LTAEAIQT(ph)PLK_
CDK5RAP2	1100	0.495431	negative	_VSVM(ox)GTDQS(ph)ESINTSNETEYLK_
CDK5RAP2	1102	0.495431	positive	_VSVM(ox)GTDQS(ph)ESINTSNETEYLK_
CDK5RAP2	1244	0.888278	negative	_DLSPPRYDS(ph)LVQSQR_
CDK5RAP2	255	0.94	negative	_SQM(ox)ACPDENV(ph)SGELR_
CDK5RAP2	706	0.999999	negative	_LASTEQTELLAS(ph)K_
CDK5RAP2	613	1	positive	_TLEEIS(ph)EIR_
Q9H930 (Uniprot)	180	1	negative	_QGEVPES(ph)PEAR_
NEDD1	332	0.971055	negative	_S(ph)VNVNAASGGVQNSGIVR_
NEDD1	384	0.999997	negative	_S(ph)INTDTLSK_

NEDD1	339	0.999999	negative	_SVNVNAAS(ph)GGVQNSGIVR_
NEDD1	581	1	negative	_IADS(ph)IGNNR_
DLG7	827	0.483843	NC	_AGLEVLGSSDPTT(ph)SASR_
DLG7	828	0.989155	negative	_AGLEVLGSSDPTTS(ph)ASR_
DLG7	787	0.999611	negative	_IS(ph)QSELFDNK_
DLG7	774	0.999903	negative	_NTAS(ph)QNSILEEGETK_
DLG7	777	1	positive	_NTASQNS(ph)ILEEGETK_
CDC27	440	0.499972	NC	_LDS(ph)SIIEGK_
CDC27	444	0.824847	NC	_LDSSIIS(ph)EGK_
CDC27	441	0.996426	positive	_LDSS(ph)IIEGK_
CDCA7L	117	0.999896	negative	_ASLVS(ph)EEEEDEEEDKATPR_
CDCA7L	114	1	negative	_AS(ph)LVS(ph)EEEEDEEEDK_
CENPC1	104	0.478864	positive	_EAS(ph)LQFVVEPSEATNR_
CENPC1	515	0.857115	NC	_LVPEEVT(ph)TVTK_
CENPC1	732	0.922007	negative	_LVLPS(ph)NTPNVR_
CENPC1	516	0.997091	positive	_LVPEEVTST(ph)VTK_
CENPC1	518	0.999636	NC	_LVPEEVTSTVT(ph)K_
CENPC1	734	0.999998	positive	_LVLPSNT(ph)PNVR_
CENPC1	439	1	negative	_PAEEQLDVGQS(ph)K_
CENPC1	96	1	positive	_EAS(ph)LQFVVEPSEATNR_
CENPC1	250	1	positive	_IRDS(ph)EYEIQR_
CENPC1	110	1	positive	_S(ph)VQAHEVHQK_
CENPC1	316	1	negative	_S(ph)WITIPR_
CENPF	1774	0.443244	negative	_NESCDISKEHT(ph)SETTER_
CENPF	1639	0.5	NC	_KET(ph)PSAPAK_
CENPF	3177	0.729498	negative	_SPT(ph)DSPREGLR_
CENPF	3175	0.829851	negative	_SVPVNNLPERS(ph)PTDSPR_
CENPF	1777	0.853823	negative	_EHTSET(ph)TERTPK_
CENPF	2512	0.895738	positive	_ISELEIINS(ph)SFENILQEK_
CENPF	1248	0.923143	positive	_GDLETS(ph)NLQDM(ox)QSQEISGLK_
CENPF	1748	0.928854	negative	_LQLQGLDLSS(ph)R_
CENPF	838	0.948684	positive	_VDSLEFS(ph)LESQK_
CENPF	1766	0.971172	negative	_NES(ph)CDISK_
CENPF	242	0.97434	positive	_DFS(ph)ASYFSGEQEVTPSR_
CENPF	1637	0.981015	negative	_KET(ph)PSAPAK_
CENPF	1747	0.996815	NC	_LQLQGLDLS(ph)SR_
CENPF	773	0.998991	negative	_SKDAS(ph)LVTNEDHQR_
CENPF	2513	0.999571	positive	_ISELEIINSS(ph)FENILQEK_
CENPF	876	0.999995	negative	_AEQM(ox)HQS(ph)FVAETSQR_
CENPF	1750	1	positive	_S(ph)LLGIDTEDAIQGR_
CENPF	1324	1	positive	_YELVTELNDS(ph)R_
CENPF	1988	1	positive	_VNDS(ph)WKER_
CENPI	709	0.5	positive	_TVNVVS(ph)SIR_

CENPI	710	0.5	NC	_TVNVS(ph)SIR_
CENPI	705	1	negative	_T(ph)VNVSSIR_
CENPJ	892	1	negative	_LNIS(ph)QDQPPGDNAR_
CENPL	53	1	positive	_KIPQCS(ph)QLQEDVDPQK_
CENPT	397	0.999776	negative	_AAS(ph)PESASSTPESLQAR_
CENPT	45	0.932011	positive	_RALLET(ph)ASPR_
CENPT	10	0.997079	negative	_(ac)ADHNPDSDS(ph)TPR_
CENPT	47	0.999999	negative	_ALLETAS(ph)PR_
MLF1IP	194	0.999999	positive	_TGPLSAQPS(ph)VEK_
CENPE	612	0.482225	positive	_M(ox)DLS(ph)YSLESIEDPK_
CENPE	611	0.914385	positive	_M(ox)DLS(ph)YSLESIEDPK_
CEP192	1484	1	negative	_AS(ph)LESTSDLGASGK_
CEP192	1502	1	positive	_HGGNVS(ph)LDVLPVK_
CEP192	285	0.838187	negative	_T(ph)CSIDNKLQDVGNDK_
CEP192	287	0.984293	negative	_TCS(ph)IDNKLQDVGNDK_
CEP72	387	0.928274	negative	_TLSQPEAS(ph)ETEEQR_
CEP97	308	0.967084	positive	_QLM(ox)NQS(ph)QNEELSPLVPVETR_
CEP350	2689	0.999107	positive	_VSIAAEDDTLDNT(ph)FSEEEK_
KIF4A	953	0.5	NC	_QLEES(ph)VSEK_
KIF4A	951	0.999675	negative	_QLEES(ph)VSEK_
KIF4A	810	0.99974	negative	_GQVS(ph)ESEDSTK_
KIF4A	815	0.999801	positive	_GQVSESEDS(ph)ITK_
KIF4B	1038	1	negative	_FLEQS(ph)MDIEDLK_
KIF4B	394	1	positive	_NQS(ph)LVEENEK_
CCDC86	110	0.886644	negative	_QPEYS(ph)PESPR_
CCDC86	113	0.989544	negative	_QPEYSPE(ph)PR_
CCDC86	58	1	negative	_AGLGS(ph)PERPPK_
NCAPH	16	0.451243	negative	_M(ox)GPPGALPATM(ox)NNS(ph)SSETR_
NCAPH	98	0.886582	positive	_SIDISAT(ph)IPK_
NCAPH	17	0.903612	negative	_GPPGALPATM(ox)NNS(ph)ETR_
NCAPH	96	0.939989	negative	_SIDIS(ph)ATIPK_
NCAPH	201	0.954973	negative	_DAPSLEEVEGHVADGS(ph)ATEM(ox)GTTK_
NCAPH	203	0.976778	negative	_DAPSLEEVEGHVADGSAT(ph)EMGTTK_
NCAPH	15	0.993995	NC	_GPPGALPATM(ox)NNS(ph)SSETR_
NCAPH	92	0.995883	negative	_S(ph)IDISATIPK_
NCAPH	233	1	negative	_TIEQNINNLNVS(ph)EADR_
NCAPG	390	0.999999	negative	_GDFS(ph)YIGNLMTK_
NCAPD3	430	0.920321	positive	_EVDNT(ph)LSLEHQK_
NCAPD3	325	0.974873	negative	_APLAVTS(ph)QVINCR_
NCAPD3	508	0.99979	positive	_NSS(ph)AFSYQR_
NCAPD3	1250	0.999997	negative	_QLAS(ph)ELEYDM(ox)K_
NCAPG2	1114	0.993942	positive	_T(ph)FM(ox)EITLEEDSIER_
NCAPG2	1119	0.999923	negative	_TFM(ox)EIT(ph)LEEDSIER_



NCAPG2	30	0.999969	positive	_EAS(ph)DPFSLNELLDELSR_
NCAPG2	605	1	NC	_ENVV(ph)VLDK_
DYNC1LI1	516	0.871028	negative	_KPVTVSPTTPTS(ph)PTEGEAS_
CKAP2	533	0.99993	positive	_EVS(ph)IEDTGVDVDPEK_
DOCK7	907	0.460531	negative	_SLSNSNPDISGT(ph)PTSPDDEVR_
DOCK7	909	0.460531	negative	_SLSNSNPDISGT(ph)PTSPDDEVR_
TOP2A	282	0.5	positive	_DKLDET(ph)GNSLK_
TOP2A	285	0.5	negative	_DKLDET(ph)GNSLK_
TOP2A	1458	1	NC	_SVVS(ph)DLEADDVK_
RANBP2	2280	0.995942	NC	_SALSPSKS(ph)PAK_
RANBP2	2293	0.996196	negative	_LNQSGTSVGT(ph)DEESDVTQEEER_
RANBP2	2450	0.999573	NC	_DSLIT(ph)PHVSR_
RANBP2	2454	0.999936	negative	_DSLITPHVS(ph)R_
TOPORS	98	0.999991	negative	_LQQTVPADAS(ph)PDSK_
EML3	208	1	negative	_S(ph)NYNLEGISVK_
EPPK1	2508	1	negative	_AEAEAGS(ph)PRPDPR_
ERCC6L	805	0.499844	negative	_GT(ph)GSADSIATLPK_
ERCC6L	790	0.5	positive	_QDLS(ph)SIK_
ERCC6L	791	0.934394	NC	_QDLSS(ph)IK_
ERCC6L	807	0.998481	negative	_GTGS(ph)ADSIATLPK_
ERCC6L	14	1	negative	_FPEAEALS(ph)PEQAAHYLR_
ERCC6L	774	1	positive	_M(ox)AS(ph)VVIDDLPK_
ERCC6L	995	1	negative	_AGFVHS(ph)K_
FANCD2	592	0.991802	negative	_SES(ph)PSLTQER_
FNBP4	432	0.997156	negative	_ALEEGDGSVSGSS(ph)PR_
TUBGCP2	873	1	negative	_LERLS(ph)AER_
TUBGCP2	899	0.886792	negative	_LRVSLGT(ph)R_
TUBGCP2	896	0.999999	negative	_VS(ph)LGTR_
TUBGCP5	182	1	positive	_EDS(ph)GIQVDR_
TUBGCP6	1068	0.999993	NC	_VGENVS(ph)DVAPTQPR_
TUBGCP6	1230	1	negative	_VGENVS(ph)DVAPIR_
HSPB1	199	0.997277	negative	_AQLGGPEAAKS(ph)DETAAK_
HSPB1	65	1	negative	_PLPPAAIES(ph)PAVAAPAYSR_
HSPB1	15	1	negative	_GPS(ph)WDPFR_
HMMR	65	1	NC	_DTTLPAS(ph)AR_
HMMR	704	1	negative	_ENFALKT(ph)PLK_
KPNA2	490	0.999998	negative	_ASLS(ph)LIEK_
INCENP	509	0.49832	negative	_NQM(ox)LM(ox)TPT(ph)SAPR_
INCENP	420	0.769604	negative	_PAAS(ph)SPETPSAGQQEA_
INCENP	915	0.848169	negative	_VPSS(ph)LAYSLKKH_
INCENP	294	0.908119	negative	_VLAPILPDNFSTPT(ph)GSR_
INCENP	510	0.924064	negative	_NQM(ox)LM(ox)TPTS(ph)APR_
INCENP	94	0.972951	negative	_SS(ph)QLSSR_

INCENP	93	0.985121	negative	_S(ph)SQLSSR_
INCENP	914	0.98594	negative	_VPS(ph)SLAYSLK_
INCENP	507	0.995107	negative	_NQMLMT(ph)PTSAPR_
INCENP	323	0.999999	NC	_YS(ph)LVAK_
INCENP	72	1	positive	_RIS(ph)YVQDENR_
INCENP	524	1	negative	_RNT(ph)PLRM(ox)DPK_
INCENP	350	1	negative	_TAEPPAAS(ph)GR_
INCENP	330	1	positive	_QES(ph)VVR_
KIF11	931	0.999928	NC	_S(ph)YLYPSTLVR_
KIF18A	695	0.995156	negative	_SPPSQSVQLNDS(ph)LSK_
KIF20A	48	0.46704	negative	_NLLSDCSVVST(ph)SLEDK_
KIF20A	558	0.5	positive	_ADTGLDDDIENEADIS(ph)M(ox)YGK_
KIF20A	47	0.811131	negative	_NLLSDCSVVS(ph)TSLEDK_
KIF20A	635	0.853842	positive	_ESLTS(ph)FYQEEIQR_
KIF20A	662	0.957139	negative	_QQS(ph)VAHQQSGSELALR_
KIF20A	670	0.963194	negative	_QQSVAHQQSGS(ph)ELALR_
KIF20A	668	0.964931	negative	_QQSVAHQQS(ph)GSELALR_
KIF20A	49	0.999532	negative	_NLLSDCSVVSTS(ph)LEDK_
KIF20A	41	0.999761	negative	_NLLS(ph)DCS(ph)VVSTSLEDK_
KIF20A	44	0.999924	negative	_NLLSDCS(ph)VVS(ph)TSLEDK_
KIF20A	244	1	negative	_S(ph)VYIESR_
KIF20A	825	1	NC	_LQGQVS(ph)AKK_
KIF20A	556	1	negative	_ADTGLDDDIENEADIS(ph)MYGK_
KIF20A	109	1	negative	_DSFALKS(ph)NER_
KIFC1	33	0.499915	positive	_APSQLPLS(ph)GSR_
KIFC1	31	0.99985	NC	_APSQLPLS(ph)GSR_
NDC80	49	0.916435	NC	_LSINKPT(ph)SER_
NDC80	50	0.996448	negative	_LSINKPTS(ph)ER_
NDC80	44	0.998157	NC	_LS(ph)INKPTSER_
NDC80	55	1	NC	_KVS(ph)LFGK_
DSN1	38	0.423053	negative	_T(ph)SASLEM(ox)NQGVSEER_
DSN1	39	0.425809	negative	_TS(ph)ASLEMNQGVSEER_
DSN1	41	0.90672	negative	_TSAS(ph)LEM(ox)NQGVSEER_
DSN1	125	0.999799	negative	_SIS(ph)VDLAESK_
DSN1	123	0.999999	negative	_S(ph)ISVDLAESK_
DSN1	331	1	negative	_SMQQLDPS(ph)PAR_
NSL1	22	0.493728	negative	_ELAAGT(ph)ESQALVSATPR_
NSL1	24	0.902749	negative	_ELAAGTES(ph)QALVSATPR_
CENPB	306	0.5	positive	_LAAQSLDT(ph)SGLR_
CENPB	307	0.995467	positive	_LAAQSLDTS(ph)GLR_
MAP7D1	113	0.302259	NC	_RSS(ph)QPSPTAVPASDSPPTKQEVK_
MAP7D1	118	0.302259	negative	_RSS(ph)QPSPTAVPASDSPPTKQEVK_
MAP7D1	123	0.302259	negative	_RSS(ph)QPSPTAVPASDSPPTKQEVK_

MAP7D1	316	0.95963	NC	_SRSAVT(ph)LPR_
MAP7D1	452	0.999976	negative	_ARLS(ph)ASTASELSPK_
BUB1	306	0.5	negative	_LHQVVET(ph)SHEDLPASQER_
BUB1	314	0.634657	negative	_LHQVVETSHEDLPAS(ph)QERSEVNPAR_
BUB1	307	0.812046	negative	_KLHQVVETS(ph)HEDLPASQER_
BUB1	661	1	positive	_FS(ph)PIQEKS(ph)PK_
MYBBP1A	1163	0.999989	NC	_EIPSATQS(ph)PISK_
Em:U62317.2	493	1	negative	_FVQETELS(ph)QR_
Em:U62317.2	96	0.997212	negative	_QLSS(ph)VQEDR_
NUMA1	1788	0.903678	negative	_SQAPLES(ph)SLDSLGDVFLDSGR_
NUMA1	1818	0.905349	positive	_TTQIINIT(ph)M(ox)TK_
NUP107	45	0.883423	negative	_VLLQASQDENFGNT(ph)TPR_
NUP107	57	0.8837	positive	_TPS(ph)SFRQPFTPTSR_
NUP107	4	1	positive	_S(ph)GFGEISSPVIR_
NUP153	520	0.383793	negative	_VQM(ox)TSPS(ph)STGSPM(ox)FK_
NUP153	518	0.418952	negative	_VQM(ox)TS(ph)PSSTGSPM(ox)FK_
NUP153	515	0.829829	negative	_VQM(ox)T(ph)SPSSTGSPM(ox)FK_
NUP153	522	0.978026	NC	_VQM(ox)TSPSSTGS(ph)PM(ox)FK_
NUP153	516	0.988075	negative	_VQM(ox)TS(ph)PSSTGSPM(ox)FK_
NUP153	343	1	positive	_S(ph)GIDITDFQAK_
NUP88	540	1	positive	_S(ph)VANPAFLK_
NUP93	75	0.999965	negative	_GLDISHIS(ph)QR_
NUP93	72	1	positive	_GLDIS(ph)HIS(ph)QR_
NUP98	980	0.893409	negative	_AS(ph)LLTDEEDVDM(ox)ALDQR_
NUP98	983	0.970659	negative	_ASLLT(ph)DEEDVDM(ox)ALDQR_
NUP98	591	0.998516	positive	_NLNNS(ph)NLFSPVNR_
NUP98	606	0.999997	negative	_DSENLAS(ph)PSEYPENGER_
NUP98	595	1	negative	_NLNNSNLFS(ph)PVNR_
NPM1	198	1	positive	_KGQES(ph)FK_
NUP188	1709	0.991639	negative	_GAPSS(ph)PATGVLPSQGK_
NUP188	1523	0.999703	negative	_VQRPPS(ph)AASAAPSSSK_
NUP43	361	0.743713	negative	_SS(ph)LFHQGGR_
NUP35	73	1	negative	_SPLLAGGS(ph)PPQPVVPAHK_
TPR	1185	0.99272	positive	_EGVQGPLNVS(ph)LSEEGK_
PCNT	1690	0.499995	positive	_LDMQNS(ph)QTAVSLR_
PCNT	3325	0.817391	positive	_LGGVLPDST(ph)SK_
PCNT	2479	0.926108	negative	_LSGS(ph)DLGGHSSLLER_
PCNT	815	0.975834	positive	_SLT(ph)EQQGR_
PCNT	813	0.982922	positive	_S(ph)LTEQQGR_
PCNT	1688	0.998154	positive	_LDM(ox)QNS(ph)QTAVSLR_
PCNT	2594	0.999999	positive	_TLSEEKANS(ph)VQK_
PCNT	1653	1	negative	_RES(ph)EVLDLKEQLEK_
PCNT	2963	1	negative	_RAAGS(ph)DADHLREQQR_

PCNT	2044	1	negative	_NEMRLS(ph)LEDGGK_
PCNT	2860	1	negative	_ELRCS(ph)LER_
PCNT	2433	1	negative	_KEDEIQDIS(ph)LHGGK_
PCNT	682	1	negative	_VQLS(ph)LLQTEK_
PCNT	201	0.998933	negative	_GIFTIS(ph)DHPAEQR_
PCM1	861	0.999321	negative	_QGLAETAS(ph)PVAVSLR_
PCM1	110	1	positive	_INFS(ph)DLDQR_
PLEC1	4389	0.90212	negative	_SSSVGS(ph)SSSYPIPAVSR_
CASC5	1805	0.465659	negative	_S(ph)ANSVLIKNSLR_
CASC5	479	0.909009	negative	_TIYS(ph)GEENM(ox)DITK_
CASC5	1013	0.978665	positive	_LVANDS(ph)QLTPLEEWSNNR_
CASC5	1808	0.999998	positive	_SANS(ph)VLIK_
FAM83B	334	0.969072	negative	_LDS(ph)SYFK_
FAM83D	368	0.863728	NC	_LARLS(ph)STPR_
PRC1	593	0.479591	negative	_DPSLSDSS(ph)TVGLQR_
PRC1	592	0.935751	positive	_DPSLSDSS(ph)TVGLQR_
DKFZp434J046	1147	0.451398	negative	_AGTGY(ph)ASPDRTTHVLAAGK_
DKFZp434J046	1145	0.60411	negative	_AGT(ph)GYASPDRTTHVLAAGK_
DKFZp434J046	975	0.901806	negative	_KEVEAGPGDQQGDSY(ph)LR_
DKFZp434J046	1149	0.999998	negative	_AGTGYS(ph)PDR_
DKFZp434J046	974	1	positive	_KEVEAGPGDQQGDS(ph)YLR_
Q86Y91				
(Uniprot)	443	1	negative	_ALQEES(ph)LGM(ox)EAQVER_
RACGAP1	169	0.5	NC	_TDESLDWDS(ph)SLVK_
RACGAP1	164	0.999731	positive	_TDES(ph)LDWDSSLVK_
RACGAP1	170	0.999836	positive	_TDESLDWDS(ph)LVK_
RACGAP1	214	1	positive	_SIGSAVDQGNES(ph)IVAK_
RGPD8	1534	1	NC	_AVVS(ph)PPK_
RGPD8	21	0.999697	negative	_YVASVLGLTPS(ph)PR_
ARHGEF2	737	0.722985	positive	_S(ph)ESLESPPGER_
ARHGEF2	739	0.998405	negative	_SES(ph)LESPR_
SENP3	57	0.821678	NC	_SGGGFGPDGSGTT(ph)VPAR_
AURKB	61	0.93013	negative	_VM(ox)ENS(ph)SGTPDILTR_
AURKB	62	0.994939	negative	_VM(ox)ENSS(ph)GTPDILTR_
AURKB	19	0.997002	negative	_QTAPS(ph)GLSTLPQR_
AURKB	16	0.997888	negative	_QT(ph)APSGSTLPQR_
AURKB	35	0.999987	negative	_KEPVT(ph)PSALVLM(ox)SR_
AURKB	7	0.999999	negative	_(ac)AQKENS(ph)YPWPYGR_
AURKB	45	1	negative	_S(ph)NVQPTAAPGQK_
AURKB	236	0.333333	negative	_KT(ph)M(ox)CGTLDYLPPEM(ox)IEGR_
AURKB	239	0.333333	positive	_KT(ph)M(ox)CGTLDYLPPEM(ox)IEGR_
AURKB	232	0.999944	negative	_T(ph)M(ox)CGTLDYLPPEM(ox)IEGR_
MARK2	569	0.989246	negative	_VPVAS(ph)PSAHNISSGGAPDR_

PLK1	326	1	negative	_FS(ph)IAPSSLDPSNR_
PRPF4B	277	0.999999	NC	_KKS(ph)PIINESR_
SHCBP1	5	0.999369	negative	_(ac)ADGS(ph)LTGGGLEAAAM(ox)APER_
SGOL2	436	0.740113	positive	_TERS(ph)DVLDGKR_
SGOL2	1150	0.841756	NC	_SLSEIHSPNIQDS(ph)SFDSVR_
SGOL2	421	0.921569	negative	_NSS(ph)DVDIGEKIENR_
SGOL2	433	0.992403	negative	_T(ph)ERSDVLDGK_
PDS5B	1359	0.858497	negative	_AES(ph)PESSAISTQSTPQK_
SON	97	0.838414	negative	_CVSVQT(ph)DPTDEIPTTK_
SON	2020	0.988273	NC	_S(ph)RTPLR_
SON	160	0.997173	NC	_FDS(ph)EPSAVALLPTR_
SON	94	0.999614	negative	_CVS(ph)VQTDPTDEIPTK_
SON	152	0.999878	NC	_SHDDGNIDLES(ph)DSFLK_
SON	2022	1	NC	_S(ph)RT(ph)PLRR_
SON	154	1	positive	_SHDDGNIDLESDS(ph)FLK_
SPAG5	400	0.393666	negative	_STNT(ph)SQTGLVGTK_
SPAG5	50	0.398686	negative	_SPACSSLT(ph)PSLCK_
SPAG5	114	0.651142	negative	_T(ph)SEEAVDPLGNM(ox)VK_
SPAG5	115	0.74228	negative	_TS(ph)EEAVDPLGNM(ox)VK_
SPAG5	403	0.930181	negative	_STNTSQT(ph)GLVGTK_
SPAG5	401	0.987596	positive	_STNTS(ph)QTGLVGTK_
SPAG5	159	0.997926	negative	_LDTM(ox)AETNSIS(ph)LNGPLR_
SPAG5	157	0.999591	negative	_LDTM(ox)AETNS(ph)ISLNGPLR_
SSFA2	737	0.579884	positive	_S(ph)QSLPTTLLS(ph)PVR_
SSFA2	1161	0.896832	negative	_ASVALTPTAPS(ph)R_
SSFA2	739	0.97224	positive	_SQS(ph)LPTTLLSPVR_
SSFA2	746	0.999919	negative	_SQS(ph)LPTTLLS(ph)PVR_
SSFA2	668	1	negative	_SLAS(ph)IEAK_
TPX2	358	0.652646	NC	_KDDINLLPSKS(ph)SVTK_
TPX2	361	0.695135	positive	_KDDINLLPSKSSVT(ph)K_
TPX2	359	0.794102	NC	_KDDINLLPSKSS(ph)VTK_
TPX2	356	0.934409	positive	_DDINLLPS(ph)KSSVTK_
TPX2	80	0.499997	negative	_KANLQQAIVT(ph)PLKPVDNTYYK_
TPX2	646	0.82756	negative	_S(ph)VAEGLSGSLVQEPFLATEK_
TPX2	652	0.99448	negative	_SVAEGLS(ph)GSLVQEPFLATEK_
TPX2	654	0.999767	positive	_SVAEGLSGS(ph)LVQEPFLATEK_
TPX2	174	0.999781	negative	_KPEEEGS(ph)AHQDTAEK_
TPX2	72	1	negative	_ANLQQAIVT(ph)PLKPVDNTYYK_
TPX2	369	1	negative	_DPQT(ph)PVLQTK_
TPX2	486	1	negative	_S(ph)PAFALKNR_
TPX2	113	1	negative	_KT(ph)PAQPQR_
TUBA1C	51	0.407613	negative	_TIGGGDDS(ph)FNTFFSETGAGK_
TUBA1C	48	0.999533	positive	_TIGGGDDS(ph)FNTFFSETGAGK_

---

TUBA1C	41	0.999999	positive	_T(ph)IGGGDDSFNTFFSETGAGK_
C14orf106	134	0.446311	NC	_DKQEQRNS(ph)SLLEPQK_
C14orf106	191	0.49996	negative	_ASVQGVPLES(ph)SNNDIFLPVK_
C14orf106	192	0.49996	positive	_ASVQGVPLES(ph)SNNDIFLPVK_
C14orf106	135	0.999675	positive	_NSS(ph)LLEPQK_
C13orf3	319	0.493847	negative	_TNSS(ph)SNDLEVEDR_
C13orf3	318	0.932538	negative	_TNSS(ph)SNDLEVEDR_
C13orf3	317	0.977196	negative	_TNS(ph)SSNDLEVEDR_
C18orf21	158	0.999805	NC	_GKS(ph)PASVFR_
CENPN	226	1	positive	_S(ph)LGLDINMDSR_
CENPN	282	1	positive	_SGLNGS(ph)ILAER_
CENPN	235	1	positive	_SLGLDINM(ox)DS(ph)R_
CIT	433	0.997708	negative	_SES(ph)VVSGLDSPAK_
ITGB3BP	112	0.495291	negative	_LNHPS(ph)LTESK_
ITGB3BP	110	0.999986	negative	_LNHPS(ph)LTESK_
KIF23	897	0.548666	positive	_GGGQSVQFTDIET(ph)LKQESPNGSR_
KIF23	889	0.999994	positive	_GGGQS(ph)VQFTDIETLK_
KIF23	450	1	NC	_AICGLT(ph)PGRR_
MAP7	200	0.997359	negative	_RLS(ph)SSSATLLNSPDR_
VIM	458	0.988223	positive	_DGQVINET(ph)SQHHDDLE_
VIM	459	0.993847	positive	_DGQVINETS(ph)QHHDDLE_
VIM	83	0.999979	positive	_LLQDS(ph)VDFSLADAINTEFK_
VIM	409	1	positive	_LLEGEES(ph)R_
VIM	299	1	negative	_FADLS(ph)EAANR_

NC= non-conclusive sites. Positive also in the membrane incubated with ATP alone or showing the same signal intensity in the A version of the peptide.

---

## Curriculum Vitae

First Name        Bin  
Last Name        Wang  
Date of Birth      10 July 1982  
Telephone        + (49) / (0) 176 6838 6046  
Email             bwang@biochem.mpg.de  
Address           Brunellenweg 28, 80689, Munich



## Education

2006- present    Ph.D. studies

Ludwig-Maximilians-Universität München

Max-Planck Institute of Biochemistry, Martinsried, Germany

Thesis: Phosphoproteome studies of human mitotic spindle proteins

Supervisors: Dr. Roman Körner and Prof. Erich A. Nigg

2005 – 2006     Master's studies

Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China

Project: Purification and identification of bio-active molecules from traditional Chinese herbs

Supervisor: Prof. Xinmiao Liang

2004 – 2005     Master's studies

University of Science and Technology of China, Hefei, China

Supervisor: Prof. Xinmiao Liang

2000 – 2004     Bachelor's studies

Central South University, Changsha, China

Thesis: Fingerprint analysis of volatile components from tobacco flavor (in collaboration with Changde tobacco co., Ltd., Changde, China)

---

## Languages

- Chinese: Native
- English: Fluent (written/spoken)
- German: Basic level

## Computer software

MS: Xcalibur, Masslynx, FLEXControl, FlexAnalysis

LC: nanoLC ACQUITY, ChemStation, Empower

MS data: Mascot, MSQuant, MaxQuant, GPMW

Microsoft office: Word, Excel, PowerPoint

Adobe: Adobe Acrobat, Adobe Illustrator

## Publications

\* denotes equal contribution

- Wang B\*, Malik R\*, Nigg EA and Körner R. Evaluation of the Low-Specificity Protease Elastase for Large-Scale Phosphoproteome Analysis. Analytical Chemistry, 2008, 80 (24), p9526–9533
- Santamaria A\*, Wang B\*, Elowe S\*, Malik R, Zhang F, Silljé H HW, Körner R and Nigg EA. Identification of Plk1 specific phosphorylation sites on spindle proteins. (submitted)
- Gaitanos TN, Santamaria A\*, Jeyaprakash AA\*, Wang B, Conti E and Nigg EA. Stable kinetochore–microtubule interactions depend on the Ska complex and its new component Ska3/C13Orf3. The EMBO journal, 2009, 28 (10), p1442-1452



---

## Major Honors/Scholarships

- 2006 Max-Planck Society & Chinese Academy of Sciences Joint Ph.D. Promotion Scholarship
- 2004 Qualification for Postgraduate Study at Chinese Academy of Sciences with exam-exemption
- 2004 Graduate of Hunan province with Distinction
- 2004 Graduate of Central South University with Distinction
- 2001 - 2002 Top Grade Scholarship of Central South University
- 2001 - 2002 Outstanding Scholarship from Hainan Airlines of China

## References

Dr. Roman Körner  
Group Leader  
Dept. Cellular Biochemistry  
Max-Planck Institute of Biochemistry  
Martinsried, Germany  
Email: rkoerner@biochem.mpg.de

Prof. Erich A. Nigg  
Director  
Biozentrum  
University of Basel  
Switzerland  
Email: erich.nigg@unibas.ch

Prof. Xinmiao Liang  
Director  
Dalian Institute of Chemical Physics  
Dalian, China  
Email: liangxm@dicp.ac.cn