Phosphoproteomic analysis of
*Arabidopsis thaliana* ribosomes

Hanna Klang Årstrand
She's got a secret garden
Where everything you want
Where everything you need
Will always stay
A million miles away

- Bruce Springsteen
Abstract

Ribosomes serve as the site of protein synthesis in all living cells. Ribosomes were discovered in 1955 by George E. Palade when he was studying the endoplasmic reticulum which is covered by ribosomes. He received the Nobel Prize in Physiology or Medicine in 1974 for this discovery. Ribosomes are large protein and rRNA complexes which are made up from one small and one large subunit that work together to translate mRNA into a protein chain. Eukaryotic translation is mainly controlled during the initiation, which involves protein phosphorylation. In plants there is a general increase of protein synthesis during the day in order to synthesize proteins needed for photosynthesis. Phosphorylation can alter protein function and localization and is reversibly added and removed by kinases and phosphatases, respectively.

The aim of the studies in this thesis was to elucidate the phosphorylation status of ribosomal proteins in the *Arabidopsis thaliana* 80S ribosome. I have focused on comparing ribosomal protein phosphorylation between different conditions and subcellular locations, namely day/night conditions and cytosol/nucleus location.

By using Fe$^{3+}$IMAC to enrich phosphorylated peptides from cytosolic ribosomes followed by mass spectrometric analysis eight serine residues in six ribosomal proteins were found to be phosphorylated. Among these was a novel phosphorylation site in 40S ribosomal protein S6 at Serine 231. By using quantification with stable isotope labeling and mass spectrometry this phosphorylated residue and three other ribosomal phosphopeptides were found to have increased phosphorylation levels during day as compared to night ranging from 2 to 4 times. This phosphorylation increase can in turn effect the modulation of the diurnal protein synthesis in *Arabidopsis thaliana*.

Ribosome biogenesis involves shuttling of proteins and ribosomal subunits between the cell nucleus and cytoplasm. By purifying ribosomal proteins from these two cellular compartments and enriching for phosphopeptides using TiO$_2$ affinity chromatography combined with mass spectrometry I was able to analyze their phosphorylation status. This method identified 13 phosphopeptides derived from 11 ribosomal proteins as well as phosphopeptides from two ribosomal associated proteins. 40S ribosomal protein S2-3 was found phosphorylated only in the cytoplasmic samples while 60S ribosomal protein L13-1 and the two ribosomal associated proteins were found only in the nuclear enriched samples.
**Populärvetenskaplig sammanfattning**

DNA är livets kod och det är ribosomerna som är de maskiner som är de centrala delarna i proteinsynthesen vilka tillverkar proteiner utifrån DNA-koden. Denna process kallas för translation vilket syftar på att ribosomer översätter den genetiska koden till proteiner. Ribosomer består dels av ribosomalt RNA men också av ett stort antal proteiner som tillsammans bildar en stor och en liten enhet som sätts ihop till en fullständig ribosom. Eukaryoter har en 80S ribosom som består av en 40S enhet (den lilla) och en 60S enhet (den större).

Ett sätt för cellen att ändra ett proteins egenskaper och funktion är genom s.k. post-translationella modifieringar och av dessa är fosforylering en av de viktigaste. Proteinfosforylering förekommer vid initieringssteget som sker i början av i proteinsynthesen som ett sätt att reglera mängden protein som tillverkas.


I avhandlingens andra del har jag studerat ribosomer som har renats fram dels från cellkärnor och dels från cytoplasman. Avsikten var att studera skillnader i
List of papers

Paper I

Turkina MV, Klang Årstrand H, Vener AV. (2011)

Paper II

Klang Årstrand H, Vener AV. (2012)
Phosphorylation of Ribosomal Proteins in Cytoplasm and Nucleus of Arabidopsis thaliana. Manuscript.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CID</td>
<td>collision-induced dissociation</td>
</tr>
<tr>
<td>CKI</td>
<td>casein kinase I</td>
</tr>
<tr>
<td>CKII</td>
<td>casein kinase II</td>
</tr>
<tr>
<td>ESI</td>
<td>electrospray ionization</td>
</tr>
<tr>
<td>ETD</td>
<td>electron transfer dissociation</td>
</tr>
<tr>
<td>HPLC</td>
<td>high-performance liquid chromatography</td>
</tr>
<tr>
<td>IMAC</td>
<td>immobilized metal affinity chromatography</td>
</tr>
<tr>
<td>L</td>
<td>large ribosome subunit</td>
</tr>
<tr>
<td>LC</td>
<td>liquid chromatography</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>MS/MS</td>
<td>tandem mass spectrometry</td>
</tr>
<tr>
<td>m/z</td>
<td>mass over charge ratio</td>
</tr>
<tr>
<td>NLS</td>
<td>nuclear localization sequence</td>
</tr>
<tr>
<td>NPC</td>
<td>nuclear pore complex</td>
</tr>
<tr>
<td>PTM</td>
<td>post-translational modification</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>s</td>
<td>phosphorylated serine residue</td>
</tr>
<tr>
<td>S</td>
<td>small ribosome subunit, or the Svedberg unit</td>
</tr>
<tr>
<td>t</td>
<td>phosphorylated threonine residue</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
</tbody>
</table>
# Table of contents

ABSTRACT ................................................................................................................................. V
POPULÄRVETENSKAPLIG SAMMANFATTNING .................................................................. VII
LIST OF PAPERS ....................................................................................................................... IX
ABBREVIATIONS ....................................................................................................................... X
TABLE OF CONTENTS ............................................................................................................. XI

1. INTRODUCTION .................................................................................................................... 1
   1.1 The ribosome .................................................................................................................. 1
      1.1.1 Ribosome function ................................................................................................. 1
      1.1.2 Ribosome protein composition .............................................................................. 3
      1.1.3 The P-proteins ......................................................................................................... 3
   1.2 The nucleus ..................................................................................................................... 4
   1.3 Ribosome phosphorylation ............................................................................................ 5
   1.4 Increased protein synthesis in light ............................................................................... 6

2. AIMS ....................................................................................................................................... 7

3. METHODOLOGIES ................................................................................................................. 9
   3.1 Arabidopsis thaliana as a model organism ..................................................................... 9
   3.2 Plant growth and protein preparation .......................................................................... 9
      3.2.1 Isolation of ribosomes by ultracentrifugation ...................................................... 9
   3.3 Proteomics and protein phosphorylation ..................................................................... 10
      3.3.1 Enrichment and quantitation of phosphorylated peptides ............................... 11
   3.4 Mass spectrometry ....................................................................................................... 12

4. PRESENT INVESTIGATION ................................................................................................ 15
   4.1 Characterization of phosphoproteins from day and night (Paper I) ....................... 15
   4.2 Cytoplasm and nucleus (Paper II) .............................................................................. 18
      4.2.1 Motif analysis ......................................................................................................... 20

5. CONCLUSIONS ..................................................................................................................... 23

6. ACKNOWLEDGEMENTS ..................................................................................................... 24

REFERENCES ........................................................................................................................... 26
1. Introduction

1.1 The ribosome

Protein synthesis is an essential process in all living cells and it is performed by huge protein and RNA complexes called ribosomes. Ribosomes have a fundamental role in all living cells and the overall structure and function of the ribosome is conserved between prokaryotes and eukaryotes. The eukaryotic ribosome has two subunits, a small 40S subunit and a large 60S subunit, that together form an 80S ribosome (Figure 1). S stands for the Svedberg unit which is a sedimentation coefficient in centrifugation and not a measurement of size.

Figure 1. The 40S small subunit and 60S large subunit form the 80S ribosome.

The 40S subunit contains one strand of ribosomal RNA (rRNA) (18S rRNA) and 32 proteins while the 60S subunit contains three rRNA strands (5S rRNA, 28S rRNA, 5.8S rRNA) and 48 proteins. The plant 80S ribosome is slightly smaller than its equivalent in rat and yeast because of the smaller mass of the large subunit. This is due to differences in the 28S rRNA and heterogeneity in the protein composition (Cammarano et al. 1972; Schnare et al. 1996; Chang et al. 2005; Ben-Shem et al. 2011).

1.1.1 Ribosome function

The function of the ribosome is to translate mRNA into protein by catalyzing the peptidyl transferase reaction of protein synthesis. The small subunit uses mRNA as a template and reads each codon (3 nucleotides) of mRNA to pair it with the correct amino
acid which is provided by a tRNA molecule. The tRNA contains an anticodon which is complementary to the mRNA codon. The catalytic activity is carried out by the rRNA in the large subunit. The ribosome contains three binding sites: the aminoacyl binding site (A-site), the peptidyl binding site (P-site) and an exit site (E-site). The A-site binds a tRNA molecule that carries the next amino acid in the protein sequence. The P-site is where the peptidyl bond is formed between the new amino acid and the peptide being synthesized. The E-site binds empty tRNA and discharges it from the ribosome (Figure 2). When the A-site binds to a stop codon a release factor can bind and the ribosome and mRNA complex is disassembled. The protein is released from the ribosome and transported to the Golgi apparatus where it is completed and subsequently released (Schmeing and Ramakrishnan 2009; Klinge et al. 2012).

Figure 2. Translation. The ribosome reads the mRNA sequence and uses it as template in order to synthesize proteins. The amino acids are delivered by tRNA molecules which enter the ribosome and bind to the mRNA codon.
1.1.2 Ribosome protein composition

When it comes to the protein content of the Arabidopsis ribosome most of the protein gene families are expressed and could be present in the ribosome at different time points in the development (Schmid et al. 2005). An analysis of expressed sequence tags identified 249 ribosomal protein genes that encode 32 small subunit (S) proteins and 48 large subunit (L) proteins, in total 80 ribosomal proteins (Barakat et al. 2001). The genes are organized into multigene families with two and up to seven members with an average of three family members. Many families have members with very high sequence conservation from 65% to 100%. An *in silico* analysis of theoretical gene specific peptides concluded that it therefore is difficult, and in some cases impossible, to find gene-specific matches for all family members. This is because the identified tryptic peptides can be derived from several or all family members when analyzing the ribosomal protein content using mass spectrometry. 10 families have no gene-specific peptides at all: S18, S29, S30, L11, L21, L23, L36a, L38, L40 and L41. Another 8 families have such high level of sequence identity that it is unlikely that any gene-specific peptides would be found: S4, S13, S17, S26, S27, L15, L35a, L37 and L29 (Carroll et al. 2008; Byrne 2009). The presence of so many multigene families indicates that the protein composition of the ribosome could be dynamic and affect the translation regulation process in plants.

Ribosomal proteins are present mainly at the ribosome surface where they can interact with other components in the cell such as regulatory proteins (Brodersen and Nissen 2005). In fact, the core of the ribosome where the catalytic activity takes place is composed almost only of rRNA. Thus it has been speculated that proteins were added later for fine tuning of the ribosome machinery and that the ribosome was built initially from only rRNA.

1.1.3 The P-proteins

The acidic phosphoribosomal proteins (P-proteins) form a lateral stalk structure in the 60S ribosomal subunit which is present in the peptidyl transferase region in both prokaryotes and eukaryotes. The P-proteins are phosphoproteins that have unique N-terminal and central regions followed by a section of acidic residues and a conserved C-
terminal phosphorylated region. The P-stalk is needed for efficient elongation during translation and enhances aminoacyl-tRNA binding, binding of elongation and release factors, stimulating GTPase activity of elongation factor-2, releasing of empty tRNA and the mRNA movement (Bailey-Serres et al. 1997). In eukaryotes the P-stalk is formed of five acidic P-proteins, one P0, two P1, and two P2 that in turn corresponds to the prokaryotic homologues L10 and L7/L12 (Liljas 1991; Montoya-Garcia et al. 2002). Arabidopsis contain one additional P-protein called P3 which is found in mono- and dicotyledonous plants (Bailey-Serres et al. 1997; Szick et al. 1998).

1.2 The nucleus

The nucleus is a complex membrane enclosed organelle present in eukaryotic cells that contains most of the genetic material of the cell and is thus involved in DNA replication and gene expression. The nucleolus is a nuclear subcompartment that is not surrounded by a membrane in which the assembly of ribosomal subunits takes place before they are exported out to the cytoplasm. It is broken down and reformed in each mitotic cell cycle in order to respond to the need for synthesis of new ribosomes (Lamond and Earnshaw 1998; Hernandez-Verdun 2011). Nearly 60 ribosomal proteins present in the Arabidopsis nucleolus were previously identified (Pendle et al. 2005). Ribosome biogenesis in the nucleus needs import of ribosomal proteins from the cytoplasm where they are transcribed. Proteins are imported into the nucleus through the nuclear pore complex (NPC) and ribosomal subunits are transported out into the cytoplasm and assembled into complete ribosomes (Gorlich and Kutay 1999; Guttler and Gorlich 2011). A nuclear localization sequence (NLS) present in the protein is required in order for it to be imported into the nucleus. The classic NLS motif consists of a short sequence of positively charged arginines and lysines but there are also other motifs which facilitate nuclear import (Merkle 2010). The NLS motif is recognized by nuclear transport receptors (NTRs) which bind the protein, shuttle it across the nuclear envelope and release it on the other side. Furthermore, that a protein contains a NLS does not mean that it has to be localized in the nucleus since the NLS can be masked by molecular interactions or the protein can be anchored to static components (Jones et al. 2009). Proteins that do not contain a NLS signal may also piggy-back together with other proteins during transport into the nucleus (Steidl et al. 2004).
1.3 Ribosome phosphorylation

Post-translational modifications (PTMs) are chemical modifications of a protein after it has been translated. PTMs are important for regulating protein function and thereby other processes such as catalytic activity, localization, protein stability and interaction with partners (Kersten et al. 2009). Known ribosomal protein PTMs include methionine removal, methionine oxidation, methylation, acetylation, ubiquitination and phosphorylation and of these phosphorylation has been subject to several large-scale studies (Chang et al. 2005; Carroll et al. 2008; Reiland et al. 2009). Protein phosphorylation is one of the most important, most prevalent and most studied PTMs. In eukaryotes protein synthesis is mainly controlled at the initiation stage and this involves several cases of reversible protein phosphorylation (Jackson et al. 2010). The most studied phosphorylated ribosomal proteins are 40S ribosomal protein S6 and the acidic ribosomal P-proteins P0, P1, P2 and P3 (Bailey-Serres et al. 1997; Szick-Miranda and Bailey-Serres 2001; Montoya-Garcia et al. 2002; Ruivinsky and Meyuhas 2006; Henriques et al. 2010; Rosner et al. 2010; Hutchinson et al. 2011).

40S ribosomal protein S6 is the most known and studied ribosomal phosphoprotein and it is needed for assembly of ribosomal subunits. It is localized in the mRNA binding site of the ribosome and 40S ribosomal subunit. S6 phosphorylation has been shown to be directly associated with regulating cell size and S6 heterozygosity leads to early embryonal lethality in mice (Ruivinsky and Meyuhas 2006). S6 is phosphorylated C-terminally in response to mitogenic stimuli in both animals and plants but there is no significant sequence similarity between them (Williams et al. 2003; Rosner et al. 2010). Arabidopsis have two S6 isoforms, S6-1 and S6-2, which have been found to be phosphorylated at several C-terminal residues. A large-scale study identified Ser-127, Ser-237, Ser-240, Ser-247, Thr-249 in S6-1 and Ser-237, Ser-240 in S6-2 (Reiland et al. 2009) and the Ser-240 phosphorylation was also characterized in a study using Arabidopsis cell culture (Carroll et al. 2008).

In Arabidopsis S6 is phosphorylated by the 40S ribosomal protein S6 kinase (S6K) which is a member of the AGC family of serine/threonine kinases (Bögre et al. 2003). Arabidopsis have two S6K genes, S6K1 and S6K2, which have high sequence similarity. S6K is downstream regulated by the target of rapamycin (TOR) pathway. S6


phosphorylation has been shown to be predominantly nuclear and could be involved in the localization of S6 between nucleus and cytoplasm (Rosner et al. 2010).

All four plant acidic P-proteins (P0, P1, P2 and P3) are also phosphorylated in their C-terminal region. (Chang et al. 2005; Carroll et al. 2008; Reiland et al. 2009). Two members of the 60S ribosomal protein L13 family have been identified as phosphoproteins: L13-1 was found to be C-terminally phosphorylated in (Carroll et al. 2008) and L13-3 was indicated as a phosphoprotein in (Reiland et al. 2009) but in the latter study the position of the phosphorylated residue was not determined.

1.4 Increased protein synthesis in light

During the day photosynthetic organisms are in need of a functional photosynthetic machinery. Plants have circadian networks that regulate their biological processes to meet their photosynthetic requirements (Thines and Harmon 2010; McWatters and Devlin 2011). Diurnal regulation increases translation rates for many proteins and metabolic enzymes during the day phase which also affect the capacity for cellular growth. The amount of ribosomes present in polysomes and the occupancy of mRNA transcripts increase in the day compared to the night which suggests a need for increased protein synthesis (Piques et al. 2009).
2. AIMS

The general aim of this project was to investigate differences in the phosphorylation status of ribosomal proteins in *Arabidopsis thaliana*. The focus of each paper was as follows:

**Paper 1**
To investigate and quantify differences in phosphorylation of ribosomal proteins between leaves harvested during day or night.

**Paper 2**
To identify phosphorylation of the ribosomal proteins in two different subcellular compartments and examine any variations between them.
3. Methodologies

3.1 Arabidopsis thaliana as a model organism

*Arabidopsis thaliana* is a small flowering plant in the mustard family. Its complete genome was sequenced in 2000 (The Arabidopsis Genome Initiative 2000) and it has since become a widely used model organism in plant biology, genetics and cell biology. *Arabidopsis* 157 megabase pair genome is organized into five chromosomes and is one of the smallest genomes among plants. *Arabidopsis* has several advantages as a model organism; it has a small genome, short six weeks generation time and it produces large amount of seeds that can be easily collected and grown (Meinke et al. 1998). By using Agrobacterium-meditated transformation (Clough and Bent 1998) it is possible to insert T-DNA at a gene of interest and study the characteristics of the resulting mutant that lacks the protein as compared to wild type plants. Today, T-DNA insertion mutants for nearly all *Arabidopsis* genes are commercially available via international seed stock centers. This has been extensively used to gain knowledge about the function of many gene products in recent years.

3.2 Plant growth and protein preparation

For research purposes *Arabidopsis thaliana* plants can be grown either on soil or in a hydroponic system (Norén et al. 2004). Hydroponics is a soil free and clean system in which plants are grown in water with added nutrients and continuous aeration. In this work *Arabidopsis thaliana* wild type plants were grown hydroponically at 20-22 °C with 65-70% relative humidity. The photoperiod was 8 h light and 16 h dark at a light flux of 150 µmol of photons m⁻²s⁻¹. Plants from day samples were harvested 4 h after the light period started and night samples were harvested 2 h before the dark period ended.

3.2.1 Isolation of ribosomes by ultracentrifugation

Ribosomes can be purified by ultracentrifugation through a dense sucrose cushion to protect the ribosomes and to separate them from other cytoplasmic contaminants. Denser components such as ribosomes will pellet through the cushion while less dense components such as plasma membranes remain in the upper fraction. During isolation
of intact ribosomes it is important to manage the Mg\(^{2+} \) (~30 mM) and K\(^+ \) (~200 mM) concentrations since too little Mg\(^{2+} \) and too much K\(^+ \) will cause the ribosomes to dissociate into subunits (Chua et al. 1973; Carroll et al. 2008) because the metal ions are needed for charge neutralization in the ribosome.

In Paper I cytoplasmic ribosomes were isolated using ultracentrifugation (Williams et al. 2003). In Paper II ribosomes were prepared as in paper I and also from isolated nuclei (Folta and Kaufman 2000; 2006).

### 3.3 Proteomics and protein phosphorylation

Proteomics is defined as the large-scale study of proteins, i.e. analysis of the whole protein content of the cell which is also known as the proteome. It is particularly challenging to study the proteome due to its complexity; not only can proteins be post-translationally modified by many different combinations of PTMs, the protein content also varies between cell types and with different time points in the same cell. This is because genes are expressed differently in different cell types during the varying stages of development. Stresses and signals that are transmitted from outside or inside the cell also cause the proteome to differ between cells or whole biological systems. In the past mRNA analysis was used to predict the proteome but since mRNA is not always translated into protein this method is not reliable and no longer used.

Phosphorylation is one of the most often occurring and most important PTMs in eukaryotes and it is has a key role in regulatory mechanisms that are important in many signaling pathways. About 30% of all proteins in a cell are thought to be phosphorylated at any given time and plant genomes contain more than 1000 protein kinases (Pawson and Scott 2005; Steen et al. 2006; de la Fuente van Bentem et al. 2008). Protein phosphorylation is a reversible process that requires enzymes that add and remove phosphate groups, called kinases and phosphatases, respectively. When studying protein phosphorylation it is not only important to identify phosphoproteins but also to map and quantify phosphorylation sites to elucidate their roles in cell signaling networks.
Analysis of the entire set of phosphorylated proteins in a cell is difficult to achieve. The main obstacle in modern proteomics is the huge variation of the dynamic range of proteins in a cell which makes mass spectrometric identification and quantification challenging due to the inherent properties of the equipment (Aebersold 2003; Mann and Jensen 2003; Cox and Mann 2007). Mass spectrometers have a limited dynamic range and are only able to sequence the most abundant peptide during each sequence event. This in combination with the fact that each peptide sequencing cycle takes a certain amount of time, depending on the hardware, which limits the number of sequenced peptides means that low abundant peptides will never be sequenced. By coupling nano-LC to the mass spectrometer the complexity of the sample is decreased while the separation and sensitivity is enhanced but the problem of lacking protein coverage still remains (Mann et al. 2002). In summary there are five reasons as to why it is difficult to study phosphoproteomics: 1. Low phosphorylation stoichiometry. 2. Phosphorylation sites on a protein can vary and none, some or all of them can be phosphorylated at any given time. 3. Low phosphoprotein abundance means that enrichment methods have to be used. 4. The dynamic range of the assays used has limits and therefore major phosphosites are detected but not the minor phosphosites. 5. Phosphoproteins are dephosphorylated by phosphatases during sample preparation. Phosphatase inhibitors may be needed.

3.3.1 Enrichment and quantitation of phosphorylated peptides

It is generally required to enrich the phosphorylated peptides before analysis by mass spectrometry due to their low abundance in comparison with non-phosphorylated peptides which results in suppression effects. By using an enrichment procedure the peptide complexity is reduced and the resulting peptide mixture is easier to analyze. One often used method to enrich phosphopeptides is immobilized metal affinity chromatography (IMAC) (Andersson and Porath 1986). Negatively charged phosphate groups interact with positively charged metal ions, immobilized on a matrix, and are selectively enriched on the IMAC column. Several metal ions such as Ga$^{3+}$ and Fe$^{3+}$ can be employed and of these Fe$^{3+}$ has been shown to be successful in enrichment of phosphopeptides in Arabidopsis chloroplasts (Vener et al. 2001). However, the metal ions will also bind to and enrich non-specific peptides with many carboxyl containing
residues which are negatively charged, i.e. aspartic and glutamic acid. To overcome this inherent disadvantage of IMAC a method of methyl esterification has been developed that converts carboxyl residues to methyl esters (Ficarro et al. 2002; Vainonen et al. 2005). The esterification step provides the possibility to introduce differential labeling with stable isotopes in order to quantify protein phosphorylation in two different samples as in Paper I (Lemeille et al. 2010). Using this approach it is possible to determine phosphorylation ratios of samples from different categories such as mutants versus wild type, environmental conditions or subcellular location.

A more recent affinity chromatography method employs metal oxide resin containing TiO$_2$ which is specific for enrichment of phosphopeptides (Pinkse et al. 2004; Larsen et al. 2005). TiO$_2$ is more stable than the silica-based IMAC resin and when coupled with lactic acid in the loading buffer it is highly specific for binding of phosphopeptides rather than acidic peptides (Sugiyama et al. 2007; Dunn et al. 2010).

3.4 Mass spectrometry

Mass spectrometry has within the last decade become the golden standard in proteomic protein identification and characterization in large-scale studies. With mass spectrometry it is possible to determine the position of PTMs and thus exactly which amino acid that is phosphorylated. The Nobel Prize in chemistry 2002 was awarded to two ionization techniques that are used in mass spectrometry in order to charge peptides: electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI). Half the Nobel Prize was jointly awarded to the developers of these methods Koichi Tanaka and John B. Fenn (Tanaka et al. 1988; Fenn et al. 1989). The peptide sample is inserted into a LC-system where the peptides are separated according to hydrophobicity and eluted with a gradient into the mass spectrometer. The LC column contains reverse phase C18 material. Its hydrophobic alkyl chains interact with the analyte molecules and thus retain more hydrophobic molecules. The introduction of nano-LC to mass spectrometry has dramatically increased sensitivity as well as peptide separation (Jensen 2004). Nano-LC also helps to decrease ion suppression by minimizing coelution of peptides (Schmidt et al. 2003). The main steps in the proteomic approach used in this work are shown in Figure 3.
Figure 3. Proteomic workflow for preparation and analyses of ribosomal proteins used in this work. After endoproteinase digestion the peptides are treated in four different ways. 1. Direct analysis on LC-MS/MS to identify what proteins are present in the sample. 2. Phosphopeptide enrichment by IMAC followed by LC-MS/MS. 3. Labeling by differential stable isotopes followed by IMAC enrichment and LC-MS/MS. 4. Phosphopeptide enrichment using TiO$_2$ affinity chromatography followed by LC-MS/MS. Peptides are loaded on a LC-column, separated, eluted and vaporized into the mass spectrometer where they are separated according to their mass-to-charge (m/z) ratio. The mass spectrometer selects peptide ions which are fragmented and their fragment ion spectra are submitted to data analysis software and are searched and matched against known peptide sequences in a database.
Tandem mass spectrometry (MS/MS) is used to fragment peptides and determine their sequence. Peptide fragments are acquired from collision activation of charged precursor ions. Two fragmentation methods were used in this work: collision-induced dissociation (CID) and electron transfer dissociation (ETD). CID is the most common method for peptide fragmentation used in MS/MS. With CID the main fragmentation pathway for phosphopeptide ions is loss of the phosphate group (neutral loss) and water since this is the most energetically favored cleavage (Gruhler et al. 2005). ETD mainly yield fragment ions originating from peptide backbone cleavage with preservation of the side chains such as the phosphate group which allows better detection of PTMs with ETD as compared to CID (Syka et al. 2004). ETD favors multiply charged ions and works best for ions with charge states above 3+ and has been shown to detect more phosphopeptides than CID (Jufvas et al. 2011).
4. Present investigation

4.1 Characterization of phosphoproteins from day and night (Paper I)

In Paper I the aim was to determine if there are any changes in phosphorylation status of ribosomal proteins during day or night, either qualitative or quantitative changes. In order to examine qualitative changes we used the proteomic program Scaffold. This program helps to show a visual comparison of the identified proteins in complex mass spectrometric proteomic experiments. It also makes it easy to compare multiple samples and sample categories as we have in this study with day and night samples. Ribosomes were purified using a sucrose cushion. rRNA was removed, the ribosomal proteins were precipitated using acetone and cleaved into peptides with trypsin before being subjected to mass spectrometric characterization.

Analysis of peptides identified with mass spectrometry did not reveal any qualitative changes in ribosomal protein composition. Instead, the preparations from day and night had nearly identical protein content. 95 proteins from the 80S ribosome were identified and of them 72 proteins were identified with unique peptides. Accordingly 23 proteins were not identified with unique peptides and can therefore only be attributed to a gene family and not to a specific gene product.

Aliquots from day and night ribosomal preparations were enriched for phosphopeptides with Fe₃⁺-IMAC. The enriched samples were subjected to nanoLC-MS/MS with alternating CID and ETD fragmentation. Six ribosomal proteins were found to be phosphorylated: S2-3, S6-1, S6-2, P0-2, P1 and L29-1. All six phosphoproteins were present in both day and night samples so there was no qualitative difference in phosphoprotein composition in response to an increased need for protein synthesis during the day. In total eight phosphorylated residues were identified in these six proteins of which one was a previously unknown phosphorylation on Ser-231 in the 40S ribosomal protein S6. In the CID spectrum of this peptidethere is significant neutral loss of phosphoric acid (98 Da) from the parent ion at m/z 322.5 as indicated in the spectrum in Figure 4A. However, the ETD spectrum shows intact phosphorylated residues without loss of phosphoric acid (Figure 4B). Phosphorylation identification confidence is increased when combining CID and ETD fragmentation (Molina et al.)
The three mapped phosphorylated residues in the C-terminus of 40S ribosomal protein S6, Ser-231, Ser-237 and Ser-240, are highly conserved in higher plants which suggest a physiological role for the three identified C-terminal phosphorylations.

Figure 4. Identification of 40S ribosomal protein S6 phosphorylated peptide SEs\textsuperscript{231}LAK. A: CID spectrum where methylated doubly charged SEs\textsuperscript{231}LAK at m/z 371.6 is indicated. The b and y fragment ions are marked both in the spectrum and in the peptide sequence. B: ETD spectrum with the same peptide, methylated doubly charged SEs\textsuperscript{231}LAK at m/z 371.6. The corresponding c and z fragment ions are marked in the spectrum. The mass increments of 167 which corresponds an intact phosphorylated serine residue between z3 and z4 fragment ions and c2 and c3 fragment ions are shown with double pointed arrows.

These results prompted us to investigate if there was a quantitative difference in the phosphorylation stoichiometry of the six identified phosphopeptides in response to the diurnal cycle. Quantitative mass spectrometric analysis with differential stable isotope labeling revealed a statistically significant increase in phosphorylation during the day for four serine residues: 40S ribosomal protein S6 at Ser-231, 40S ribosomal protein S6-1 and S6-2 at Ser-240 and 60S ribosomal protein L29-1 at Ser-58 (Table 1). There was a 2 to 4 times increase in the phosphorylation of three serine residues in the 40S ribosomal protein S6. Additionally there was also a small increase in phosphorylation of a serine in 60S ribosomal protein L29-1. 60S ribosomal protein P1 also showed a non significant phosphorylation increase during the day (Table 1).

Quantification analysis of the 40S ribosomal protein S6 phosphorylated peptide SEs\textsuperscript{231}LAK, which was found for the first time in this work, is shown in Figure 5. When labeled with h\textsuperscript{3}-methanol m/z was 742.3 and when labeled with d\textsuperscript{3}-methanol m/z was
In Figure 5A the solid line shows the intensity of the day sample and the dashed line is the intensity of the night sample. It is the ratio of these intensities that are listed in Table 1. Figure 5B shows the reciprocal labeling control experiment where the day sample is marked with a dashed line and the night sample with a solid line. In both A and B the day samples have twice as high intensities as the night samples. The phosphopeptides from S6-1 and S6-2 also show a 2 fold and 4 fold increase in the day, respectively. The differential labeling experiment for 60S acidic protein P1 and 60S ribosomal protein L29-1 are shown in Paper I. The other identified phosphopeptides, doubly phosphorylated 40S ribosomal protein S6-1 and S6-2 and 60S ribosomal protein P0-2 and 40S ribosomal protein S2-3, did not yield reliable quantitative data due to low signal to noise ratios in the spectrums.

**Figure 5.** Comparative analysis of 40S ribosomal protein S6 phosphopeptides in day and night samples. Pictured are extracted ion chromatograms of a quantitative analysis of 40S ribosomal protein S6 phosphopeptides. The chromatogram shows three phosphopeptides labeled with light or heavy isotopes. The peptide sequences are shown above the corresponding peak. A: day sample labeled with heavy d3-methanol (solid line) and night sample labeled with light d3-methanol (dashed line). B: Reciprocal labeling of the same samples as control. Day sample labeled with light d3-methanol (dashed line) and night sample labeled with heavy d3-methanol (solid line).
Table 1. Relative quantification of ribosomal phosphopeptides from day and night. Phosphorylated serine and threonine residues are marked using lower case s and t in the sequences and the superscript numbers correspond to the amino acid positions. * and ** represents statistically significance at p<0.05 and p<0.01, respectively.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Peptide sequence</th>
<th>Ratio day/night</th>
</tr>
</thead>
<tbody>
<tr>
<td>S6</td>
<td>RSEs231LAK</td>
<td>2.2±0.2**</td>
</tr>
<tr>
<td>S6-1</td>
<td>SRLs249SAAAKPSVTA</td>
<td>4.2±0.4**</td>
</tr>
<tr>
<td>S6-2</td>
<td>SRLs249SAPAKPVA</td>
<td>1.8±0.2**</td>
</tr>
<tr>
<td>L29-1</td>
<td>AGENAs59AE</td>
<td>1.6±0.3*</td>
</tr>
<tr>
<td>P1</td>
<td>KDEP/AEEx103DGDLGFLFD</td>
<td>1.3±0.3</td>
</tr>
</tbody>
</table>

The higher level of phosphorylation during the day can to some degree be derived from the higher general level of day time protein synthesis as discussed in chapter 1.4. However, analysis of spectral counts from all identified proteins in day and night did not reveal higher levels in day as compared to night samples (Paper I, supplemental Table S2). From these results we can speculate that there is a diurnal regulation of 40S ribosomal protein S6-1 and S6-2 and 60S ribosomal protein L29-1 but the mechanism behind this regulation is still unknown.

4.2 Cytoplasm and nucleus (Paper II)

In Paper II we sought to answer the question if there are differences in ribosomal protein phosphorylation between subcellular compartments. The shuttling of ribosomal proteins and subunits between the cell nucleus and cytoplasm during ribosome biogenesis prompted us to investigate these two compartments. Cytoplasm located ribosomal proteins were prepared as in Paper I with ultracentrifugation and acetone precipitation. Nuclear located ribosomal proteins were prepared by first making a nuclear enriched sample by using hexylene glycol in the buffer. This compound stabilizes nuclei and prevents their lysis (Peterson et al. 1997). After treatment with trypsin phosphopeptides from cytoplasmic and nuclei samples were enriched using TiO₂ and analyzed with nanoLC-MS/MS with alternating CID and ETD fragmentation. Using this approach we were able to map in total 13 phosphopeptides from 11 ribosomal proteins and two phosphopeptides from non ribosomal proteins (NAC1 and NAC3), see Table 2. S2-3 phosphorylation was only present in cytoplasmic samples. L13-1, NAC1 and NAC3 phosphopeptides were only present in nuclear enriched samples. S6-1, S6-2,
L13-3, L29-1, P0-2, P0-3, P1 and P3 phosphopeptides were present in both cytoplasmic and nuclear enriched samples.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Peptide sequence</th>
<th>Nucleus</th>
<th>Cytoplasm</th>
</tr>
</thead>
<tbody>
<tr>
<td>S2-3</td>
<td>ALs^{273}TSKDPVVEDQA</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>S6</td>
<td>RSes^{231}LAK</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>S6-1</td>
<td>SRLs^{248}SAAAKPSVTA</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>s^{237}RLs^{248}SAAAKPSVTA</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>S6-2</td>
<td>SRLs^{248}SAPAKPVA-A</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>L13-1</td>
<td>AGDS^{138}PEELANATQVQGDYLPVR</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>L13-3</td>
<td>AGDS^{138}PEELANATQVQGDYMPIASVK</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>L29-1</td>
<td>AGENAS^{159}sEE</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>P0-2</td>
<td>EE^{305}sDEEDYGGDFGLFDVEE</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>VEEKEEs^{305}sDEEDYGGDFGLFDVEE</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>P0-3</td>
<td>KEEs^{308}sDEEDYEGGFGLFDVEE</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>P1</td>
<td>KDEPAEs^{103}E-DDLGFLFD</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>DEPAEs^{103}E-DDLGFLFD</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>P2</td>
<td>EE^{101}sDDMGFSFLFE</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>KEKEEs^{101}sDDMGFSFLFE</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>P3</td>
<td>KKKEEs^{102}E-EEE-GDFGLFD</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>NAC1</td>
<td>IDLDKPEVEDDNDDEDs^{30}DDDKDDDEADGLDGEAGGR</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>NAC3</td>
<td>IDLDKPEVEDDNDDEDs^{30}EDDEAEHGHDGEAGGR</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

In Paper I cytoplasmic ribosomes were enriched using Fe$^{3+}$IMAC while we used TiO$_2$ in Paper II. The eight phosphorylated residues that were mapped in Paper I were also found in Paper II. In addition to these residues TiO$_2$ also enriched phosphopeptides from 60S ribosomal protein L13-1 and L13-3, 60S acidic protein P0-3, 60S acidic protein P2, 60S acidic protein P3 and two non ribosomal phosphoproteins NAC1 and NAC3. A study designed to compare Fe$^{3+}$IMAC and TiO$_2$ concluded that they are largely complementary enrichment methods with 35% overlap in identified phosphopeptides (Bodenmiller et al. 2007).
The non ribosomal proteins NAC1 and NAC3 (nascent polypeptide-associated complex subunit alpha-like proteins 1 and 3) were only found phosphorylated in the nuclear enriched samples. The phosphorylation of these proteins can thus be involved in their subcellular localization and/or function which is suggested to be promotion of appropriate targeting of ribosome nascent polypeptide complexes in the cytosol. These proteins of the NAC complex has previously been shown to interact with the large ribosome subunit (Giavalisco et al. 2005)

4.2.1 Motif analysis

We used the software tool Motif-X to find overrepresented sequence patterns among the 14 identified phosphopeptides (Schwartz and Gygi 2005). This resulted in five significant motifs (p=0.04) of which three patterns were in agreement with the general motif for CK2 mediated phosphorylation (SXXE/D and variations of SXE/D and S/D). For further analysis the phosphopeptide sequences were analyzed using NetPhosK, an online tool that predicts which kinase that is phosphorylating the submitted amino acid sequences (Blom et al. 2004). The predicted result from NetPhosK was that P0-2, P0-3, P2, P3, L13-1, L13-3, L29-1, NAC1 and NAC3 are phosphorylated by CK2, P1 by CK1 and S6-1, S6-2 and S2-3 by PKC as seen in Figure 6.

CK2 requires presence of acidic amino acids surrounding the target residue (Panasyuk et al. 2006). This is in accordance with the extracted P1 motif and CK2 is the most likely candidate for P1 phosphorylation. PKC is predicted to phosphorylate S6-1, S6-2 and S2-3 but Arabidopsis lacks obvious orthologs to this kinase. Instead other kinase families of the kinase AGC group are preserved and expanded and the kinases that phosphorylate 40S ribosomal protein S6 (S6K1 and S6K2) belong to this group (Henriques et al. 2010; Hutchinson et al. 2011).
Figure 6. Alignment of phosphopeptide sequences and their predicted kinases grouped according to NetPhosK prediction. TAIR accession numbers, protein names and number of the last amino acid in the corresponding protein sequence are listed. Symbols denote identical (*), strongly similar (:), and weakly similar (.) residues. Phosphorylated amino acid residues are boxed.

<table>
<thead>
<tr>
<th>Accession</th>
<th>Name</th>
<th>Sequence</th>
<th>NetPhosK Score</th>
<th>Last Amino Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>At1g01100</td>
<td>RLA11_ARATH</td>
<td>MGAVGAGG-GGAPVAAAFAGGAAFAEKKDEFAKEDOGDGLGLFD</td>
<td>112</td>
<td></td>
</tr>
<tr>
<td>At3g12390</td>
<td>NACA1_ARATH</td>
<td>MTTEKK11AAKLEEQKIQIKQIKQKEVPEVDGQGNEEEDDGDODIDDEAGLDGFAQGKSKQS</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>At5g13850</td>
<td>NACA3_ARATH</td>
<td>MTARQKVRLAKSLREEQQIQIQKQKEVPEVDGQGNEEEDDGDODIDDEAGLDGFAQGKSKQS</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>At3g49010</td>
<td>RL11_ARATH</td>
<td>KLVIFPRARVKAGDFFELANATVQGDDILFIVXKIPMLELVKTSEMGSFPAQK</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td>At5g23900</td>
<td>RL133_ARATH</td>
<td>KLVVPRRHSRKAGDFEEVLANATVQGDDMFASVKAAMELVKTLALKAYQDKI</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td>At3g06700</td>
<td>RL291_ARATH</td>
<td>-----------</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>At3g09200</td>
<td>RLA02_ARATH</td>
<td>AVSADAGGGGAPAAA-</td>
<td>320</td>
<td></td>
</tr>
<tr>
<td>At3g11250</td>
<td>RLA03_ARATH</td>
<td>AVSADAGGGGAGQAGAAAXVEEExQDIDEYDDGDFGLP</td>
<td>323</td>
<td></td>
</tr>
<tr>
<td>At4g27710</td>
<td>RLA21_ARATH</td>
<td>SATVRDGGGGAAPAAPAEKXEXXXEFPPQXGDOQFSLFF--</td>
<td>115</td>
<td></td>
</tr>
<tr>
<td>At4g25890</td>
<td>RLA31_ARATH</td>
<td>GASSPQDGAGEAAAFKXEXXXEFPPQXGDOQFSLFQ--</td>
<td>119</td>
<td></td>
</tr>
</tbody>
</table>

** Figure 6. Alignment of phosphopeptide sequences and their predicted kinases grouped according to NetPhosK prediction. TAIR accession numbers, protein names and number of the last amino acid in the corresponding protein sequence are listed. Symbols denote identical (*), strongly similar (:), and weakly similar (.) residues. Phosphorylated amino acid residues are boxed. **
5. Conclusions

This work has focused on the study of protein phosphorylation on ribosomes derived from either different circadian cycle phases or different subcellular compartments. In summary the most important conclusions from this work are:

Eight phosphorylated amino acid residues corresponding to six ribosomal proteins were identified in ribosomes from different phases of the circadian cycle. Four of these peptides were found to have significant elevated levels of phosphorylation during the day as compared to the night (Paper I). Among these is a previously unknown phosphorylation of 40S ribosomal protein S6 at Ser-231.

After phosphopeptide enrichment of peptides from the cell nucleus and cytoplasm I identified 11 phosphoproteins and two ribosome associated proteins. One ribosomal protein was only found phosphorylated in the cytoplasm and one was only found phosphorylated in the nucleus. Additionally, the two ribosome associated proteins were also only phosphorylated in the nucleus (Paper II).
6. Acknowledgements

I would like to thank:

My supervisor Professor Alexander Vener for giving me the opportunity to work in your lab. Your truly great scientific knowledge and advice have been invaluable during these years.

Min biträdehandledare professor Cornelia Spetea Wiklund för alla givande samtal och vetenskapliga diskussioner genom åren.

Min biträdehandledare professor Ingemar Rundquist som kom och gav stöd och råd när det som bäst behövdes.

Annelie Lindström: tack för all hjälp, det hade inte fixat sig utan dig.

Gruppen:

Maria Turkina: min medförfattare och masspektrometerexpert. Så mycket som vi har avhandlat genom åren, från antikviteter till fosforyleringar, du har kunskap inom allt.

Björn Ingelsson: Boppe ploppe, bästa rumskompisen även när du var lite kort. Alla dessa knäppa, roliga och ibland allvarliga samtal kommer jag att minnas länge, länge.

Rikard Fristedt: Rille lille. Det är tyst här när ingen längre sjunger sina sjungar i kylrummet. Nu har du i alla fall det viktigaste; receptet på bananmuffins. En klippa i labbet som efterlämnade ett stort tomrum.

Åsa Jufvas: min kompis i mushandsproblematiken som har stått ut med snack om proteomik, stickning och lite till. Tack för all hjälp!

Lan Yin: Thanks for the jokes, chats and all the fun during the years. Time will tell if we meet again in Göteborg!

Thomas Kieselbach för all tålmodig hjälp med bioinformatiken.

De studenter som har livat upp i labbet genom åren: Jacob Kuruvilla, Pontus Granath, Daniel, Pavaan och Henier. Thanks for all the fun we’ve had.

Problemlösarna **Ulf Hannestad, Håkan Wiktander och Camilla Höglund** som kan fixa allt.

Alla på plan 13 och 12 för samtal i och runt fikarummet och labbet: **Cilla, Lotta, Emelie, Eva, Anders, Kerstin, Anita S, Anita L, Birgitta, Vivian, Maria, Sara, Christine, Viviana, Susana, Ia, Cecilia, Elin, Siri, Anna G, Anna L-M, Narges, Meenu, Simon, Marie, Martin, Tobias, Sebastian** och **Chamilly**.

**Berit** och **Leif** för all härlig tid på Pensionat Paradiset.

**Sandra, Kaj, My och Lia** som har stöttat in i det sista för att jag skulle kunna färdigställa den här boken.

**Mamma** och **pappa** som alltid har stöttat mig när det gäller "skolan". **Henrik** och **Marcela** som aldrig fått nått svar på frågan; Vad är det du gör egentligen? Nu kan ni läsa boken och få svaret(?)

Bästa **Christoffer** för allt stöd och mitt lilla hjärta **Ludvig**.
References


