A sneak preview of the daily life of *S. elongatus* by MS-based proteomics

Ana C. L. Guerreiro

Dedicated to my family and partner

The research presented in this thesis was performed in the Biomolecular Mass Spectrometry and Proteomics group at the UIPS at Utrecht University.

Printed by: Proefschriftmaken.nl || Uitgeverij BOXPress

ISBN: 978-94-6295-428-1

A sneak preview of the daily life of *S. elongatus* by MS-based proteomics

Sneak preview van het dagelijks leven van *S. elongatus* door MS-gebaseerde proteomics

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op maandag 11 januari 2016 des middags te 4.15 uur

door

Ana Catarina Leite Guerreiro

geboren op 14 september 1986 te Lissabon, Portugal

Promotor: Prof. dr. A. J. R. Heck

Copromotor: Dr. A. F. M. Altelaar

Table of Contents

Chapter 1

Introduction

Introduction to Mass spectrometry based proteomics

Proteins are molecules essential for life in general and are the facilitators of all cellular processes. Their composition is based on polymers of amino acids with specific sequences, as determined by the translation of the mRNA, which is sequentially determined by gene transcription. Therefore, proteins represent the final product of our genes, reflecting its abnormalities and making their study increasingly crucial.

A number of cellular states, both in health and disease, are associated with specific changes at the protein level. To some extent, the proteins present in a cell are much more dynamic when compared to the genome and the transcriptome. This is due to the fact that their expression is dependent on the needs of the cell and its environment; additionally, their chemical modification state (post-translational modifications – PTMs) determines interactions with other molecules in the cell, localization and activation/deactivation of certain cellular processes, including degradation. The proteome is then defined as the complete set of proteins expressed by the cell at a certain time and in certain conditions/states. Consequently, proteomics is the field of science dedicated to the study of the proteome, through its identification, characterization and quantification (1, 2); it complements and completes the information coming from the Genomics and Transcriptomics fields.

In order to study the proteome there are a number of techniques that can be used. Mass spectrometry, however, is usually the common denominator among a variety of approaches. In proteomics, there are several aspects of the experimental design and types of instruments that have to be considered depending on the question being addressed (3–5). These different aspects will be addressed further in this introduction.

1. Mass spectrometry based Proteomics

Mass spectrometry (MS), in general, allows the measurement of the analytes mass, consequently allowing their identification. Recent advances allowed the MS application to proteomics, due to the possibility of analyzing bigger molecules, such as peptides or even proteins.

There are different strategies that can be adopted for proteomics studies associated with mass spectrometric analysis. The analysis of proteins can be approached at the peptide level, as in bottom-up proteomics (6), or at the protein level, as in top-down proteomics (7). There is also the middle down

approach (8), which relies on long sequence peptides. The most common approach implemented is the bottom-up, since it's more straightforward.
The bottom-up experiments generally rely on several common steps of sample preparation prior to the MS analysis itself, which also include fractionation techniques. The next steps include MS data analysis, for the protein identification and quantification, through employment of bioinformatics tools (Figure 1).

1.1. Sample preparation

To identify proteins and study the proteome using mass spectrometry in a bottom-up approach, there are several sample preparation steps that have to be followed to obtain the peptides. The first step consists of extracting/isolating the desired proteins from the sample matrix, which can be composed of simple synthetic protein mixtures, cultured cells, organelle extracts, or tissue. In general, it's necessary to disrupt the cell membrane/wall, solubilize and isolate the proteins. This can be achieved using the right buffer conditions and performing sonication, followed by centrifugation. This process mainly allows recovery of the cytosolic proteins, however to reach proteins from specific organelles and from the membrane more efficiently, more steps or other techniques are required. To preserve protein modifications and also to prevent premature protein degradation by endogenous proteases, certain enzyme inhibitors are also added to the buffer solution.
Following the protein extraction, a protein digestion has to be performed to obtain peptides. This is achieved after denaturation, disruption of the protein's disulfide bonds and alkylation of the resulting reduced ends, which prevents re-folding of the proteins.

To allow efficient identification of the peptide sequences, specific proteases must be chosen. The most commonly used enzyme is trypsin, which cleaves peptide bonds at the carboxyl side of lysine and arginine amino acids (9). To increase proteome digestion efficiency, it is common to use a sequential combination of proteases (10, 11). Lys-C or Lys-N are examples of enzymes that have been combined with trypsin to complement global proteome digestion (12, 13).

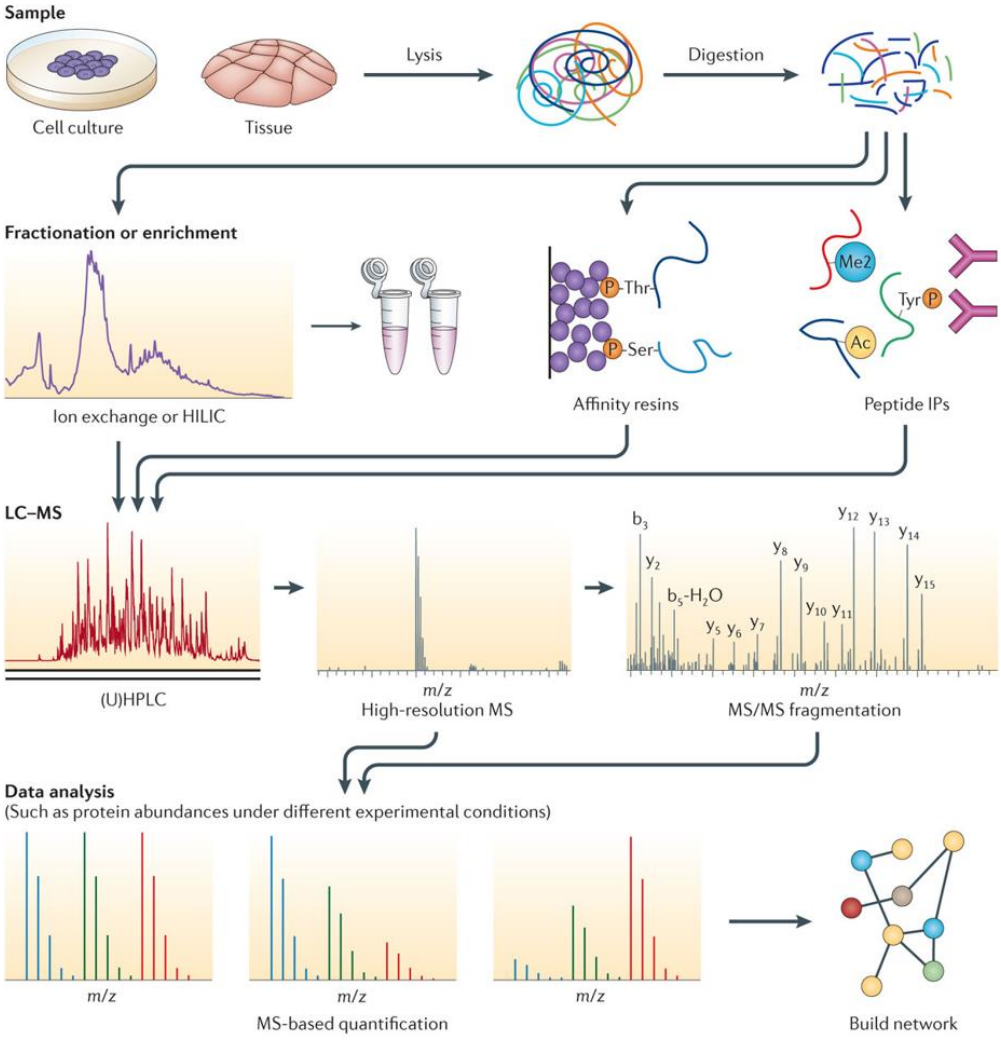


Figure 1 – General workflow of bottom-up proteomics experiments. Adopted from Altelaar *et al*. (5).

If the peptide mixture is to be analyzed directly by the mass spectrometer, it's often necessary to make sure that there are no significant contaminants in the sample and that it is suitable for MS analysis. For this purpose small scale chromatographic columns are used, such as reversed phase C18 columns.

1.2. Fractionation and multidimensionality

In the bottom-up shotgun proteomics strategy several thousands of peptides are generated through the digestion of the proteins present in whole cell lysates. This high sample complexity prevents a comprehensive MS analysis, since the mass spectrometer is limited by a characteristic sequencing speed. To decrease sample complexity and consequently make these samples more suitable for MS analysis, it is typical to separate the peptides with different physicochemical properties or enrich one set of peptides of interest. To achieve this it is common to employ liquid chromatography (LC). For highly complex samples, a multidimensional chromatography approach is usually employed, consisting of a combination of several complementary/orthogonal chromatographic methods.

One of the most frequent choices for the first fractionation step is strong cation exchange (SCX). This ion exchange technique uses a negative stationary phase to capture positive charges of the peptides, thus separating peptides according to its charge (14–17). The low resolving power of this technique is then overcome by combining it to reversed phase liquid chromatography (RP-LC). The latter is the ideal and most common separation technique used to be coupled directly to the mass spectrometer. RP-LC uses an apolar stationary phase to separate the peptides based on their hydrophobicity (18). It's highly orthogonal to SCX and it has a greater resolving power (19). Moreover, the buffers used for this chromatographic separation are compatible with MS, allowing direct infusion of the sample in the mass spectrometer, through electrospray ionization (ESI) (20). In the work described in this thesis a RP-LC was used as an 'online' method and the SCX method was coupled in an 'offline' manner. In an offline approach the sample is first separated by SCX, the fractions are collected and subsequently injected individually into the RP-LC coupled to the mass spectrometer. This set-up carries no limitation to the SCX column size and buffers used, allowing for larger sample amounts to be analyzed and non-MS-compatible SCX buffers, which can be eliminated prior to MS analysis. Additionally, an online two-dimensional approach was introduced, a so-called multidimensional protein identification technology (MudPIT). Here, the SCX and RP columns are sequentially combined in one silica capillary which is in turn coupled to the MS in an online manner (21, 22). The online method is advantageous in the sense that it's an automated technique and therefore the analysis time is reduced. However, only a limited amount of sample can be loaded. On the other hand, since the samples have to be desalted before the analysis, an

RP-SCX-RP column should be used and that carries some drawbacks, such as the reduced column lifetime and the risk of sample loss due to an error during this extended analysis. The introduction of an ultrahigh-pressure liquid chromatography (UHPLC) together with SCX (UHP-MudPIT) approach was thought to be beneficial in this case and was already tested (23). Contrary to the frequently used high-pressure liquid chromatography (HPLC), the UHPLC operates at pressures up to 1300 bar, making the analysis time faster. In addition, raising the temperature of the column will enable more controlled and stable operating pressure, leading to even better results (24).

Recently, another combination for multidimensional separation, based on RP-RP-LC, emerged with great promise. Although, the same type of separation is being applied consecutively, the simple change of the mobile phase pH can impact the orthogonality of this two-dimensional separation strategy. In fact, changing the mobile phase pH creates more impact than the change of the stationary phase type (25–28). A greater orthogonality can be achieved when the pH difference between the two dimensions is increased, i.e., a first dimension with a pH of 10 and the second with a pH of 2.6 (29); and by using a concatenation strategy of pooling samples from different parts of the gradient (30). The high-pH RP-LC used in an offline manner, has been proven to offer higher resolving power and higher orthogonality with the low-pH RP-LC when compared with SCX-RP-LC strategies, increasing the peptide identifications in the MS analysis as well as decreasing sample loss/processing (30, 31). On the other hand, this approach has its limitations, such as the incompatibility of the mobile phases, which make it difficult to employ it in an online manner; and the damaging effect of the high pH on the silica-based stationary phases. The second issue can be addressed by using a more resistant stationary phase (27, 29).

Another successful example of a first dimension to combine with RP-LC is the hydrophilic interaction liquid chromatography (HILIC), which uses a sufficiently polar stationary phase and a low aqueous mobile phase (5-20% water in acetonitrile), creating a water-rich liquid layer around the stationary phase (32). This technique has shown high orthogonality when combined with RP-LC (29, 33), since it retains peptides with higher hydrophilicity. Nevertheless, the elution order in HILIC is not completely opposite to the one of RP-LC, since it depends on several physicochemical properties. The exact separation mechanism is still not completely understood, but it was determined that the peptide elution depends on the type and concentration of salts and pH, as well as the mobile and stationary phase used (34, 35). The latter can include underivatized (36, 37) or derivatized silica (neutral (38), ionic

(32, 39, 40) or zwitterionic (41–43)). Using HILIC as a first dimension instead of SCX provides higher separation resolution (43). An online set-up of HILIC-RP is quite difficult, since a high concentration of acetonitrile is employed in HILIC and this is a strong eluent for RP. To circumvent this problem an 'in line' set-up was created (43), which is an offline method with limited sample handling. Here, the sample eluting from the HILIC column is collected in a well plate containing an acidified water solvent, which can be injected directly into a RP-LC-MSMS system.

The separation step can also be performed before digestion, where the sample is less complex. Different techniques are also available for separation/fractionation at the protein level. Among the most commonly used techniques in proteomics are gel based electrophoresis, as well as liquid chromatography techniques, such as size exclusion. The electrophoresis technique makes use of the proteins' amphoteric properties, which allow them to adopt a global positive or negative charge depending on the sample pH, making them susceptible to electric fields. They can be separated according to their molecular weight, using sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE), or according to their isoelectric point (pI), using isoelectric focusing separation (IEF). SDS-PAGE uses SDS to denature and even the charge-density ratio of all proteins in the sample; and the polyacrylamide gel forms a matrix with uniform pore size, allowing the differential migration and separation of the different proteins, where proteins with higher molecular weight are increasingly retained compared to low molecular weight proteins (44, 45). The separated proteins can subsequently be visualized in the gel through different staining strategies (Coomassie brilliant blue, or Silver staining). In IEF, charged proteins are loaded on a gel surface containing a pH gradient, enabled by so-called Immobilines™ buffers. Through the imposition of an electric field, proteins migrate until they reach the pH corresponding to their pI, which leads to a halt in migration (46, 47). These two techniques are also commonly used in combination (two-dimensional electrophoresis – 2-DE), where proteins go through IEF first and then SDS-PAGE, increasing the resolving power of separation (48, 49). Gel-based methods are inexpensive, powerful, simple and easy to use, nonetheless they have limited resolving power, poor recovery, significant manual operation, poor reproducibility and limitations in molecular weight and pI ranges. As an alternative we have many LC techniques such as reversed-phase and ion-exchange, already mentioned above, and also size-exclusion chromatography (SEC). The latter separates proteins according to their molecular weight, due to a porous stationary phase. Contrary to the SDS-

PAGE method, proteins with higher molecular weight migrate in between the stationary phase-beads, eluting before the lower molecular weight proteins, which in turn pass through the pores of the beads as well as in between them, delaying their elution. Both chromatographic techniques were recently used on an innovative way, since they were applied to native proteins as a fractionation step, facilitating the separation and discrimination of protein complex interactions (50, 51).

1.3. Enrichment techniques for modified peptides

Despite the existing pre-fractionation options to tackle sample complexity issues before MS analysis, a dynamic range issue could still be present. An example of this is associated with the analysis of the low abundant post translational modified (PTM) proteins. In fact, phosphorylated proteins are a minority when compared to the unmodified proteins, and this difference is even more pronounced at the peptide level. In this case an enrichment method is necessary to eliminate the undesired/unmodified peptides and concentrate the peptides of interest. Here, chromatography also plays an important role, however certain immunoprecipitation methods can be applied. These techniques can be performed before, after or instead of the pre-fractionation step.

The most popular enrichment strategies used in phosphoproteomics are based in metal ion affinity chromatography, which essentially comprises a stationary phase with positive metal ions to capture the negatively charged tryptic phosphopeptides. The most common approaches are the immobilized metal ion affinity chromatography (IMAC) and the metal oxide affinity chromatography (MOAC). In the first, metal ions such Fe^{3+}, Al^{3+} and Ga^{3+} are quelated to the stationary phase (52–54); and in the second, metal oxides such as TiO_2 and ZrO_2 are quelated instead (55, 56). The TiO_2 technique has been demonstrated as being more specific than IMAC, since there is a bidentate type of bond between the beads and the phosphopeptides (57). Both methods have lower specificity for peptides containing multiple basic residues and the latter a decrease of retention of peptides with multiple acidic residues. This can be partly solved by the O-methyl esterification of the carboxylic groups in acidic peptides prior to the enrichment (58). For highly complex mixtures it was suggested the use of both techniques, as a so-called sequential elution from IMAC (SIMAC), since it was demonstrated a certain complementarity in terms of the types of phosphopeptides identified (59). It

was recently added to this list another method from a new generation of IMAC materials, the so-called Ti^{4+}-IMAC, which demonstrates phosphopeptide affinity similar to the TiO_2. This method has shown higher specificity and reproducibility then the above-mentioned methods, being able to capture the phosphopeptides with multiple basic residues (60, 61). Despite its popularity, the aforementioned methods can be irreproducible and time-consuming, considering that it's frequently applied after fractionation, employing manual steps in the protocol and normally self-constructed micro-columns. A recent alternative method, which revisited the Fe-IMAC method in a HPLC column set-up, showed promise (62). This method employs a commercially available column, which can be used for offline phosphopeptide enrichment from digested lysates, being more selective, comprehensive and reproducible than the most popular methods.

The pre-fractionation techniques presented in the previous section can also be employed as an additional step in the PTM enrichment protocol, separating the modified peptides from the unmodified ones. SCX has been routinely employed in the enrichment of phosphorylated peptides and it can also be used for N-terminal acetylated peptides. In the low-pH SCX, only the basic residues and phosphorylated residues contribute to the net charge of the tryptic peptides, reducing it in comparison with their unmodified equivalent. Thus, it's possible to separate the unmodified from the modified peptides, as well as the multiply phosphorylated from the singly phosphorylated peptides (63). However, the singly phosphorylated peptides with more than one basic residue still co-elute with the unmodified peptides. For the enrichment of these phosphopeptides, a recent strategy was employed, which combines two tandem SCX separations, first with a pH of 3 and second with a pH of 1 (64). The second dimension of this approach is able to separate the now neutral phosphorylated peptides from the other peptides, which remain the same at both pH levels. Other ion exchange techniques have also been suggested as alternatives, such as strong anion exchange (SAX) and weak anion exchange (WAX), which inversely to SCX contain positively charged stationary phase to capture negatively charged peptides (65, 66). HILIC appears again as an alternative method to SCX, for the PTM enrichment of peptides with glycosylation (33, 67), phosphorylation (68, 69) and N-acetylation (33). For the phosphopeptide enrichment it was shown that this technique by itself is not sufficient for a specific enrichment, being favorable the addition of other enrichment steps based on metal ion affinity, such as IMAC (68). A similar principle is applicable to the analysis of glycosylated peptides, where other chromatography approaches are used in combination with HILIC (70).

Zwiterionic stationary phase (ZIC)-HILIC has been proposed as an alternative separation method for N-acetylated tryptic peptides (33), although it cannot be used for lysine acetylated peptides, since trypsin cannot cleave acetylated residues.

A good alternative to the enrichment and separation methods described above is the immunoaffinity enrichment, which can be used for analysis of phosphorylated proteins and is also very popular for isolation of acetylated (71, 72) and methylated proteins (73, 74). In phosphoproteomics, this technique relies on antibodies against specific phosphorylated amino acid residues or peptide sequence motifs containing phosphorylated residues such as serine, threonine and tyrosine (75–79). Unfortunately, for this method several milligrams of sample material are needed for an efficient enrichment, contrary to the metal ion chromatography techniques, and an additional enrichment step is also advisable (80).

1.4. Mass spectrometry instrumentation

Mass spectrometry analysis essentially consists of two steps: the ionization of the analyte, i.e. the peptide in this case, which is achieved in the ion source; and the storage and detection of the peptide mass, which is achieved by the analyzer and detector. In tandem mass spectrometry (MS/MS) there is a middle step during which the peptide precursor ions are fragmented, producing the so called daughter ions, which are ultimately detected. The most common MS instruments and its constituents will be described next, as well as the fragmentation techniques performed in tandem MS.

1.4.1. Ion source

The ion source, as the name indicates, is responsible for producing and delivering the ions to the mass analyzer. The peptides need to be ionized and brought to the gas phase using mainly one of two soft-ionization techniques: matrix-assisted laser desorption/ionization (MALDI); or electrospray ionization (ESI).

The ESI technique was used to perform the experiments in this work and nowadays is the most commonly applied. Here the ions are generated from a liquid solution. Firstly, a high voltage (1-3kV) is applied between the conductive capillary tip containing the solution and the entrance of the mass spectrometer. This creates an electric field, and consequently, charged droplets at the tip. Secondly, the solvent starts to evaporate, reducing the

droplet size until the charge density is too high so that the droplet splits. This point of fissure, when the Coulombic repulsion is equal to the tension in the surface of the droplet, is called Rayleigh limit (81) and generates increasingly smaller charged droplets across several cycles, until the gas phase ions are formed (Figure 2a). This whole process can also be accelerated by the use of a heating source and a nebulizer gas. Eventually, multiply charged ions are generated, depending on the length of the amino acid sequence and also on the type of amino acids. MALDI was not used in this work, but it can be used

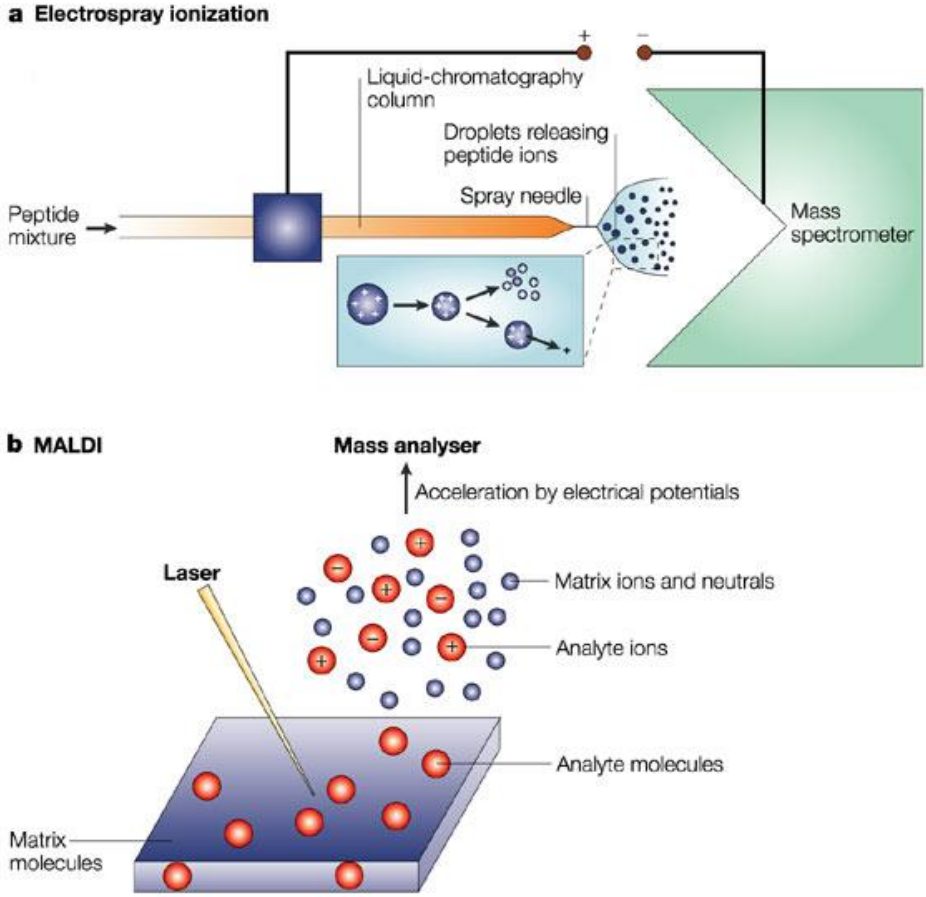


Figure 2 – Soft ionization techniques commonly used for peptide ionization. a) matrix-assisted laser desorption/ionization (MALDI) uses laser irradiation to generate singly charged peptides from a solid state solution matrix; b) electrospray ionization (ESI) uses high voltage and temperature to form increasingly small charged droplets until the gas phase peptide ions are formed. Adopted from Steen *et al*. (3).

as an alternative to ESI, in which singly charged peptides are generated from a solid state solution, through the irradiation of a laser (82) (Figure 2b). In addition to being highly compatible with liquid chromatography, ESI also has the advantage of maximizing the analysis time of the MS when compared with MALDI, since in the latter the ionization occurs in pulses. In this sense, ESI is mostly used for high-throughput experiments and MALDI for imaging mass spectrometry.

1.4.2. Mass analyzers

The mass analyzers are ultimately involved in the measurement of the ions' masses. In specific terms, they can store, separate and guide the ions generated in the ion source. Several types of mass analyzers can be found for this purpose, which operate according to different principles. The mass analyzers used during the preparation of this work include: quadrupole (Q), ion trap (IT), time-of-flight (TOF) and the orbitrap. Nowadays, in shot-gun proteomics experiments it is common to use a combination of several of these mass analyzers instead of just one at a time, making up so called hybrid instruments. Below, each mass analyzer and the hybrid MS instruments used in this work are going to be discussed.

Quadrupole

As the name indicates, the quadrupole is composed of four circular (ideally hyperbolic) metal rods, which are perfectly aligned in a parallel fashion (83, 84). The application of a combination of direct current (DC) and radio frequency (RF) voltages to the rods results in a radial motion of the ions across this tunnel-like structure. These voltages are applied in a dynamic manner, however, since opposite rods are connected, the same potential is applied to these, while opposite potentials are applied to adjacent rods instead. The dynamic nature of the electric field of this analyzer allows the preservation of specific ions in stable trajectories, while other particles or unwanted ions have unstable trajectories. The latter are eliminated through ejection into the vacuum or a collision with the rods, being subsequently discharged and unable to be transmitted further. This discrimination is dependent on the m/z in the sense that different combinations of electric fields lead to the stabilization of the trajectory of ions with certain m/z, which ultimately pass through the quadrupole in a corkscrew-like movement. Apart from the four rod arrangement, there are also hexapole (six rods) or octapole (eight rods) arrangements (Figure 3).

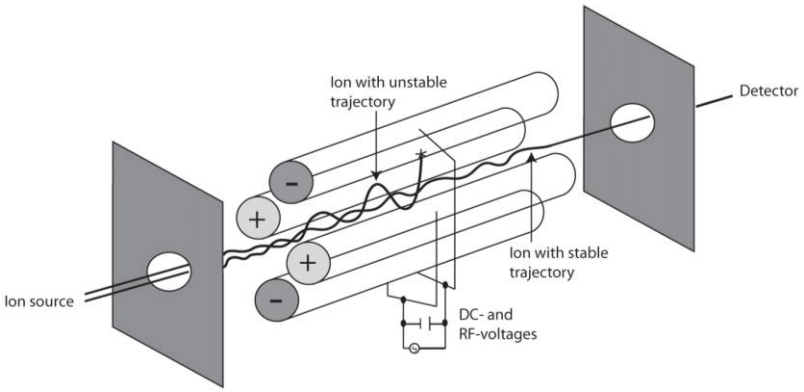


Figure 3 – Schematic representation of a quadrupole analyzer.

The quadrupole is commonly used as a mass filter. This is achieved through manipulation of the voltages applied to the rods, stabilizing the trajectory of ions with a particular m/z value. On the other hand, the filter can also be set to be wider and scan all m/z values, being used for ion guiding purposes. Lastly, the quadrupole can also be used as a reaction chamber for peptide fragmentation purposes, which will be discussed later in this chapter.

Linear Ion trap

The linear ion trap (LIT) or 2D-trap is essentially composed of a quadrupole with one electrode at each end (85, 86). These two electrodes, which can be two plates or two small quadrupoles, are kept at a static electric potential. This allows trapping of the ions in two dimensions, i.e., along an axis. In the LIT, the ions are not selectively entering the quadrupole, but they are selectively ejected in a radial or axial manner, according to their m/z (Figure 4).

In this type of analyzers a space-charge effect occurs, meaning that the charged ions repel each other when confined in a small space, leading to unstable ion trajectories (87). This effect can be partially corrected with an inert gas, such as helium, which is used to cool down the ions with an excess of kinetic energy. In this way, higher storage volumes can be achieved in the LIT, ultimately leading to an increase in dynamic range and, consequently, an increase in sensitivity. The use of helium in this analyzer also adds the possibility of performing peptide fragmentation. On the other hand, the space charge-effect can also be minimized through the regulation of the ion population inside this instrument. This can be achieved by monitoring the ion production with a pre-scan and an automatic adjustment of the ion

accumulation time in the LIT, employing the so-called automatic gain control (AGC) (88).

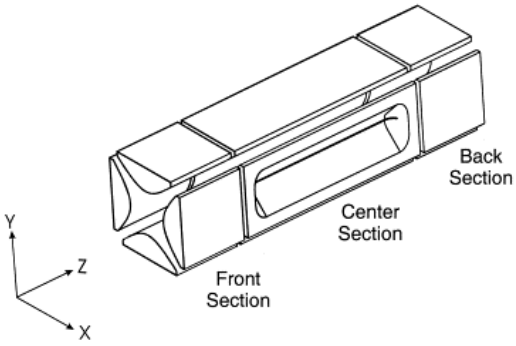


Figure 4 – Illustration of a two-dimensional linear ion trap. Adopted from Schwartz *et al*. (89).

Time-of-flight

The time-of-flight (TOF) analyzer is a free drift tube where ions travel without the application of an electric field (90). Ideally, the packets of ions are initially static and in the same position, after which they are accelerated by means of an electric field. The velocity at which the different ions "fly" across the tube is dependent on the initial potential applied to them and inversely proportional to their m/z. Consequently, the different ions hit the detector at variable times due to their different m/z. In order to correct for the possibility of different initial kinetic energies, a reflectron is commonly used. This is essentially an electrostatic ion mirror which allows the prolongation of the "flight" time (91).
This mass analyzer is able to achieve high-resolution (>20,000 FMHW) and it is one of the fastest scanning instruments, producing a full scan in microseconds.

Orbitrap

The Orbitrap is the most recent mass analyzer to be developed. However, it rapidly became one of the most popular instruments in the proteomics field.
This instrument is composed of an inner and an outer electrode, respectively shaped as a spindle and as a barrel (92). The ions are trapped by an electrostatic field in an orbital motion around the inner electrode, in the same way as described for the Kingdom trap. This movement also occurs along the axis of the inner electrode, where the ions oscillate back and forward. The frequency of these oscillations is inversely proportional to the m/z. Thus, ions with larger m/z have a lower frequency of oscillation.
The popularity of this instrument is partially related to the achievable high-resolution (>100,000 FMHW) and mass accuracy (<2 ppm). A comparable

resolution and even higher mass accuracy (sub-ppm) can be reproduced by another instrument, the Fourier transform-ion cyclotron resonance (FT-ICR) mass analyzer (93). However, its lower scanning speed and a demanding and costly maintenance adds to a choice towards the Orbitrap.

Hybrid instruments

As mentioned before, the mass analyzers described above are typically implemented as components of hybrid MS instruments, rather than as individual instruments. This MS analyzers are indeed complementary and can lead to higher rates of protein identification. In this work, three types of hybrid instruments were used: the LTQ-Orbitrap Velos; the Q-Exactive (QqOrbitrap) and the Triple TOF (QqTOF).

The LTQ-Orbitrap Velos has a combination of multipoles, quadrupoles, linear ion traps and the orbitrap (94) (Figure 5). The quadrupole and multipoles are used as ion guides and a collision cell. Two types of detectors are present in this instrument, one capable of fast scanning and high sensitivity and another capable of high accuracy. The first type is essentially composed of two ion traps positioned sequentially, one with higher gas pressure, for capture and fragmentation of the ions; while another has low gas pressure, which improves ion scanning. This dual pressure conformation allows for faster detection and higher sensitivity (95). Another C-shaped trap, the C-trap, is used to capture the ions and inject them into the second type of detector (Orbitrap), for the high-resolution ion detection. This hybrid instrument is very versatile allowing parallel detection of ions with the ion trap and orbitrap. Moreover, this instrument can perform the so called data-dependent decision tree methods. This type of method allows the selection of the best fragmentation technique according to each ion type (96, 97).

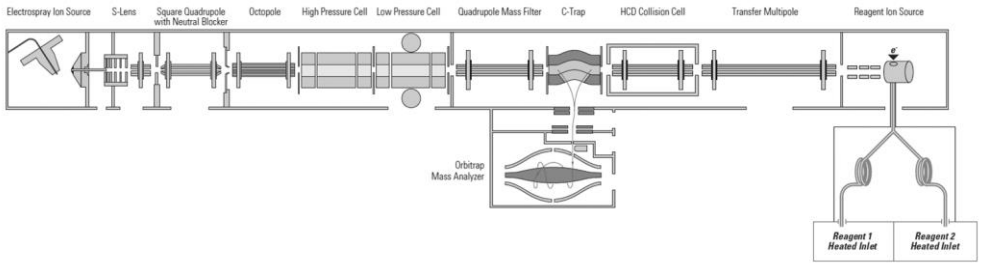


Figure 5 – A hybrid orbitrap instrument essentially composed of a linear ion trap and orbitrap analyzer (LTQ-Orbitrap Velos).

The Q-Exactive is essentially composed of multipoles, quadrupoles, a C-trap and the orbitrap (98) (Figure 6). The main difference between this instrument and the LTQ-Orbitrap Velos, is the fact that the dual ion-trap has been substituted by a quadrupole, which is used here as a mass filter. This quadrupole allows even faster ion detection than the previous instrument, with almost simultaneous ion isolation and fragmentation. Additionally, it can perform next generation targeted analysis, namely parallel reaction monitoring (PRM), which will be discussed in the Targeted proteomics section. On the other hand, this instrument is not as versatile as the LTQ-Orbitrap Velos.

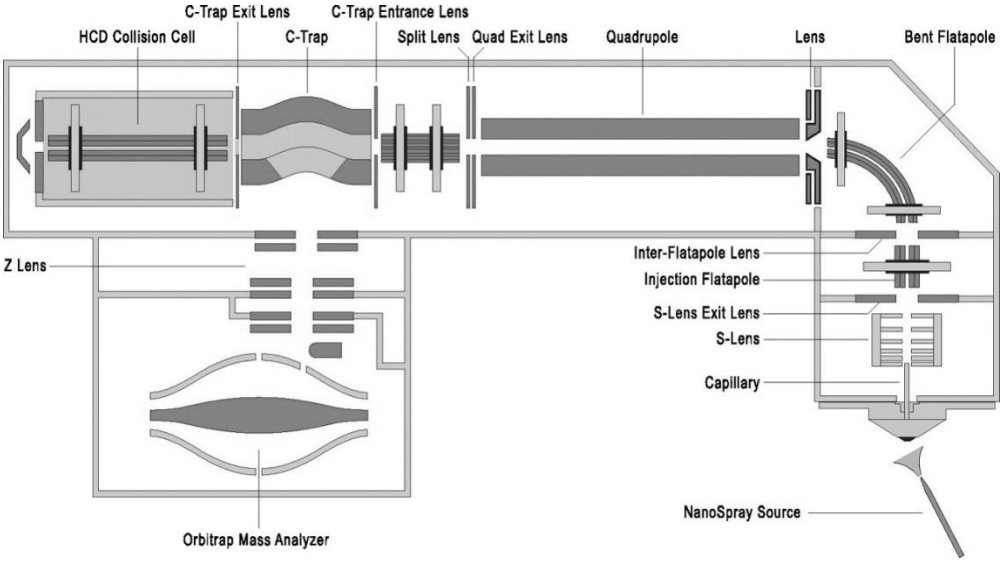


Figure 6 – A hybrid orbitrap instrument essentially composed of a quadrupole and orbitrap analyzers (Q Exactive). Adopted from Michalski *et al.* (98).

The TripleTOF is essentially composed of three quadrupoles and a TOF analyzer (99) (Figure 7). The quadrupoles act as mass filters, being able to selectively transmit ions to the TOF analyzer. As mentioned before, the TOF analyzer allows for fast detection and, consequently, this is one of the fastest scanning instruments available, leading to high protein identification rates. This instrument can be used for both high-resolution discovery and targeted proteomics approaches. This is due to the triple quadrupole conformation, similar to the common QqQ instruments used for targeted experiments. In the Triple-TOF case, the ions are guided through Q0, followed by selection in Q1 and fragmentation in Q2, which precedes the TOF analyzer, used for the high

resolution measurements. Another interesting feature of this instrument is the ability to perform data independent acquisition (DIA) or sequential windowed acquisition of all theoretical MS (SWATH™) experiments (100). Theoretically, in this approach, all ions across the full MS spectrum are selected and fragmented, which can lead to the identification of every ion detected by the MS instrument.

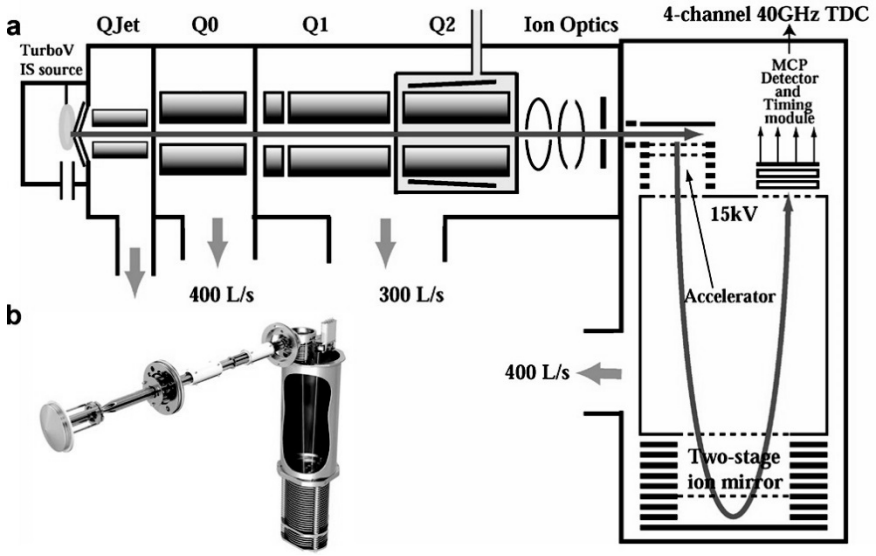


Figure 7 – Quadrupole-Time of flight instrument or TripleTOF 5600. a) schematic representation of the TripleTOF 5600; b) image of the instrument. Adopted from Andrews *et al.* (99).

1.4.3. Peptide fragmentation techniques

Peptide fragmentation is performed in tandem MS instruments' collision cell. As mentioned before, the measurement of both the precursor and fragment ions allowed in these instruments is necessary to achieve confident identification of the peptides and its corresponding proteins. This is due to the fact that a measurement of the m/z of the peptide is performed, as well as the determination of its sequence, through the MS/MS or fragmentation spectrum. There are essentially two types of fragmentation that are frequently used in MS proteomics approaches: collision-induced dissociation (CID) and electron transfer dissociation (ETD).

CID is the most common technique used for peptide ion fragmentation (101). Here, the protonated peptides are subjected to collisions with neutral gas molecules, which can be helium, nitrogen or xenon and are present in the collision cell. These collisions lead to a conversion of the kinetic energy of the peptide ion into vibrational energy, which ultimately spreads across the molecule and disrupts the peptide bonds. Usually CID can be performed in an ion-trap or in a quadrupole-type collision cell (102). The first option is a so-called tandem-in-time type of fragmentation and occurs through slow activation of the ions, which lead to the collisions and dissociation. This type of fragmentation requires a few eV of energy and occurs in a millisecond time scale. In the quadrupole, a tandem-in-space fragmentation occurs. This is a beam-type fragmentation and the ions are accelerated to the collision cell filled with the gas molecules where they suffer multiple collisions. The energy applied in this case is higher and occurs in a faster time-scale. This type of fragmentation is particularly useful for quantitative techniques that rely on isobaric tags, which will be discussed in the following section of this chapter.
These two types of fragmentation of the amide bond produce ions with the N-termini of the peptide (b ions) and with C-termini (y ions) (Figure 8). Nonetheless, there could be additional fragment ions resulting from these fragmentations, such as internal fragments, including immonium ions, and/or neutral losses of ammonia, water or phosphate groups, in the case of phosphorylated residues.

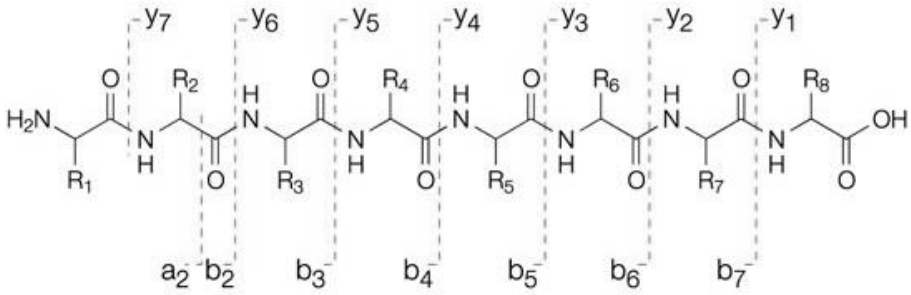


Figure 8 – Nomenclature for the fragment ions as proposed by Roepstorff and Fohlman. N-terminal fragment ions are named a/b/c and C-terminal fragment ions are named x/y/z, according to the cleavage site.

ETD is considered to be a complementary technique to CID, since it was proven to be more suited for fragmentation of multiply charged peptides (>2+)

(103). It is also very popular for the analysis of modified peptides with posttranslational modifications (PTMs). Here, a radical anion molecule transfers its electrons to the protonated peptide, leading to the fragmentation of the N-Cα bond of the amino acids. This type of fragmentation also produces ions with N-termini of the peptide (c ions) and C-termini (z ions) (Figure 8). The suitability of this technique for peptide PTM analysis, relies on the fact that these modifications remain intact during the fragmentation process, enabling the determination of the modification position in the sequence.
The peptide sequence information lies in these MS/MS spectra and it is obtained based on the difference between the fragment ions, which ultimately points to the unique mass of 19 out of 20 amino acids (excluding leucine and isoleucine).
As mentioned before, in modern instruments like the LTQ-Orbitrap Velos, one can use different methods of fragmentation for the analysis of one sample. For example, one can choose to do CID for 2+ charged ions in the ion trap and ETD for ions with charge >3+ and with phosphorylated peptides in the ion trap.

1.5. Data analysis

Peptide identification and subsequent protein identification is frequently achieved through the use of different algorithms available, such as Mascot (104), Sequest (105) and Andromeda (integrated in MaxQuant) (106, 107). These search engines perform a database search of the MS/MS spectra obtained. The process starts with the generation of a theoretical peptide database containing their masses, through an *in silico* digestion of the proteome specified. This digestion is determined by a specific protease with a certain cleavage specificity. The precursors acquired are subsequently matched to the theoretical peptides generated, within a specified mass tolerance. Next, a list of theoretical MS/MS spectra is generated for each matched precursor, according to expected known fragmentation rules. The experimental MS/MS spectra generated is also compared to this new list and a score is given to each peptide-to-spectrum-match (PSM). The score varies with the different algorithms, however it depends essentially on the number and type of fragment ions matched. Unfortunately, when dealing with large databases, there is a high risk of random and incorrect peptide identification matches, despite the use of high-resolution instrumentation. Usually, to account for these false-positives, a supplementary list of spectra containing reversed, randomized or scrambled sequences is generated from the selected

database. This so-called decoy database is used for a second search, where every match to the experimental spectrum is considered a false positive. A false discovery rate (FDR) can then be calculated by dividing the decoy PSM by the total PSMs (108). Usually, an FDR of 1% is accepted within the proteomics community, nevertheless to increase the stringency, additional thresholds to the peptide score and/or sequence length can be applied.

2. Quantitative proteomics

Nowadays, many of the bottom-up proteomics experiments are also quantitative. Several questions can be answered through the quantitative analysis of proteomes in different cell states. However, this type of experiments can be influenced by several mass spectrometer dependent factors, such as solubility, or ionizability, which will influence the peptide's intensity. Therefore, several methods have been developed to enable a more accurate MS based quantitation, and these include label-free and label-based methods. These two generally fall into the relative quantification category. An absolute quantification approach can be achieved with the use of internal standards in combination with a targeted method.

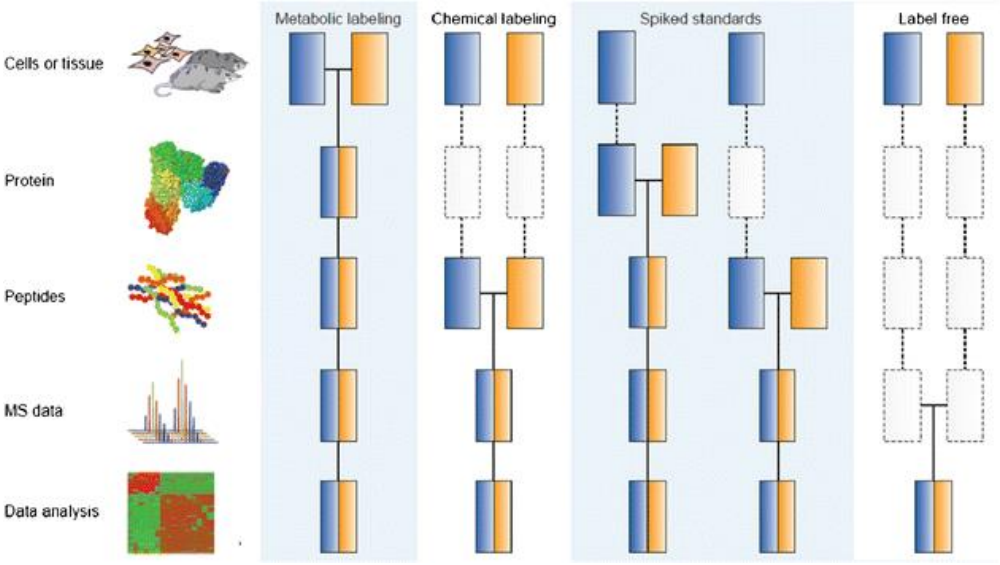


Figure 9 – Quantitative mass spectrometry based proteomics different workflows. Different strategies allow for experimental variation in different stages of the workflow. The colors blue and orange represent different experimental conditions to be compared. The dashed lines indicate presence of experimental variability. Adapted from Bantscheff *et al*. (109).

2.1. Label-free quantification

Label-free quantification can be seen as the most simple and straightforward quantification technique, since it's based on the direct comparison of peptide intensity measurements between several samples without having to add extra sample preparation steps (Figure 9). This method suffers in terms of accuracy, as it's influenced by the instrument's capabilities and peptide properties, making peptide sequencing essentially stochastic. The main approaches to this method rely on precursor ion intensities (110) or spectral counting of the peptides measured (111).
The first is based on the calculation of the XICs for each peptide across the LC-MSMS run. To ensure better precision, this method requires an alignment/normalization of the chromatographic runs. This can be achieved through the use of internal standards and specific algorithms/software. On the other hand, the spectral counting method is based on the assumption that the number of acquired spectra matching to a peptide correlates to its abundance.

2.2. Labeled quantification

Label based techniques are essentially applied through the incorporation of stable isotopes into the peptides from different samples. When measuring these samples by MS, equal peptides from each sample will be distinguishable through the different masses given by the labels. With this type of method there are additional steps required; however this also ensures higher accuracy and precision when compared to the label-free methods. The introduction of these labels can be done at different experimental stages, either at the protein level or peptide level.

2.2.1. Metabolic labeling

In metabolic labeling heavy isotopes are introduced in growing cells media, being consequently incorporated into proteins or peptides (Figure 9). This can be applied not only to cultured cells but also to complex organisms, such as plants or animals. The most used isotopes are ^{15}N and ^{13}C.

A very successful example of this technique is the Stable Isotope Labeling by Amino acids in Cell culture (SILAC) (112). In this case, the amino acids containing heavy isotopes are added to the media, substituting its natural counterparts. Then the cells go through its normal growth and division cycles until approximately 98% of its peptides contain heavy isotopes. It's important

to choose well which amino acids to use. Firstly, it's required that the cells are auxotrophic for the amino acids chosen. Secondly, it´s necessary to ensure that all peptides contain the heavy isotope, therefore the heavy amino acids chosen have to be present in all peptides. This means that, if the protease of choice is trypsin, the amino acids being incorporated should be the heavy Lysine and Arginine, for example.
This method is advantageous in the sense that it enables a decrease in technical variations and an increase in accuracy, since it's introduced before other sample preparation steps. On the other hand, it can be quite costly and is also time consuming. Most importantly, this technique cannot be applied to every organism or its tissues, or primary cells and it's limited to the comparison of 2 to 3 different samples. Recently, a so-called super-SILAC approach was applied, which enables accurate protein quantification in human tissue, through the combination of the tissue being studied and several corresponding SILAC labeled cell lines, which serve as internal standards (113).

2.2.2. Chemical labeling

Chemical labeling represents a good alternative for metabolic labeling, since it can be applied to every biological sample containing peptides or proteins (Figure 9). The labeling occurs through the reaction of favorable peptide or protein sites with reagents containing isotopes. In this method there can be used isotopic tags or isobaric tags. The first produces distinguishable peptides at the MS1 level and in the second the peptides can be distinguished at the MS2 level.
An efficient isotopic tag technique is the stable isotope dimethyl labeling (114, 115). The labeling occurs in the amino group at the N-terminal of peptides and, in the case of lysines, at its ε-position as well. This technique makes use of formaldehyde and cyanoborohydride with different combinations of ^{13}C isotope and deuterium. The result is a multiplexed experiment of up to 3 samples with a 4 Da difference in their mass spectrum. The most appealing aspects of this technique reside in the fact that it can be applied to every type and amount of sample at a low cost. However, the use of deuterated molecules imposes some disadvantages. The main concern is the retention time shift when using LC based separation techniques, which leads to less accurate quantification. Despite the fact that the labels are incorporated at a later stage in the proteomics workflow, it was concluded in a recent study that

this doesn't affect greatly the accuracy of this approach, when carefully controlled (116).
A very popular alternative to these methods is the use of isobaric tags. The most known representatives are the isobaric tag for relative and absolute quantification (iTRAQ) (117) and the tandem mass tags (TMT) (118). The iTRAQ labels can be used to compare up to 8 different samples (8-plex) and the TMT labels can be used to compare up to 10 different samples (10-plex). Each different label incorporated into the peptides has the following common composition: a reactive group, a reporter and a balancer/normalizer group (Figure 10). The amine reactive group enables the binding of the label to the N-termini and lysine side chains. The reporter group has different masses for each corresponding label, differing 1 Da from each other. This is achieved due to the different composition of ^{13}C and ^{15}N present in these molecules. The balancer/normalizer group also contains different masses in the same way as the reporter group, except for the fact that it´s present to counter-balance the reporter's masses and maintain the total mass of the tag equal for all the different labels. The combination of these different components allows peptides labeled with these tags to be indistinguishable at the MS1 level, which doesn't increase the MS scan complexity. The quantification is normally enabled after CID fragmentation (quadrupole), when the reporter ion is released (Figure 10). Thus, the reporter ions are visible at the low mass range in the MS/MS spectra, with their relative intensities corresponding to the abundance of the peptides of interest.

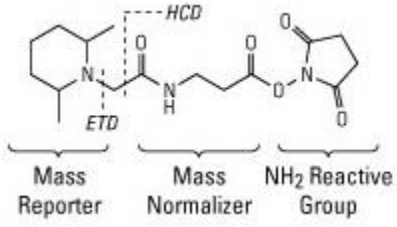


Figure 10 – Tandem mass tag representation as an example for an isobaric label.

The isobaric tags have a noticeable advantage in terms of sample multiplexing, making it a very appealing technique for high-throughput type analysis and time-course experiments, such as the one presented in this thesis. On the other hand, the detection of the low m/z ions cannot be easily performed in an ion trap, except with the low-throughput pulsed Q collision induced dissociation (PQD), therefore a more suitable fragmentation technique has to be applied. A more important issue is the potential existence of ions with similar masses at the MS1 level, which can be co-isolated for

fragmentation. This decreases the quantification accuracy, since we have reporter ion interference from different peptides in the same MS/MS spectrum, leading to a sum of the intensities from the different peptides present. In this case, it is beneficiary to add separation steps prior to MS analysis to decrease the possibility of finding those interfering peptides together. It's also viable to apply more restricted isolation windows. Since these only offer partial solutions, it was suggested a MS3 based methods to overcome this problem in a more significant way (119). The aim is to decrease the intensity of the background peptides through double isolation. Here, the peptides can be fragmented by ion-trap CID, followed by the re-isolation of the top N fragments for CID fragmentation in the quadrupole to obtain the reporter ions, which are analyzed in the Orbitrap. In turn, the fragments generated by ion-trap CID are used for peptide identification. However, this approach leads to decreased quantification rate, due to slower duty cycle speeds and lower sensitivity.

2.3. Absolute quantification

The determination of the absolute amount of a protein can be approached in different ways. The most common approach is based on the use reference peptides containing stable isotopes (internal standards). The labeled internal standards are analogs of the peptides of interest and have known concentrations. Once spiked in the sample, the intensity of both the peptide of interest and the internal standard can be compared, and an estimation of the true amount of the peptide of interest can be obtained. The standard can be introduced in the peptide form, just before the LC-MSMS analysis (referred to as AQUA) (120) or in the form of protein, before sample digestion (QconCAT – concatenated proteotypic peptides (121); PSAQ – full length proteins (122)). In general, these methods improve quantification accuracy, especially for the QconCAT and PSAQ approaches, and eliminate the labeling step, being very appealing for biomarker related research. However, obtaining the right standards is time-consuming and expensive. Additionally, this method has also limited multiplexing.

A faster way of getting an estimate of the absolute protein amount is by applying a label-free approach, through precursor peak area results. In this case, it is common to calculate the average of the 3 most intense precursor ions – the top 3 method (123, 124); or calculate the sum of all precursor ions, normalized by the number of theoretical peptides – iBAQ (intensity based absolute quantification) (125).

2.4. Targeted proteomics

When a few peptides or proteins are considered for quantification experiments it's common to apply a more targeted technique, instead of the shot-gun approaches. A selected reaction monitoring (SRM) approach can be applied to obtain a relative or absolute amount of the analyte (126). The purpose here is to reduce the interference of background ions, leading to highly accurate and reproducible quantification results, even with low abundant proteins. The triple quadrupoles (QqQ) are the classical instruments of choice, in order to perform SRM. Here, the intended precursor ions are selected in Q1, fragmented in Q2 and the fragment ions are further selected and detected in Q3 (Figure 11).

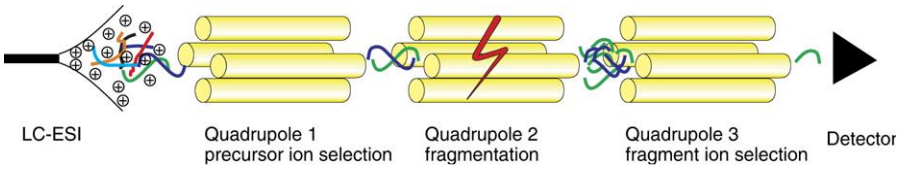


Figure 11 – Representation of a triple quadrupole. Adopted from Lange *et al.* (127).

In the targeted SRM method a series of steps are usually taken before obtaining the final results. First, the proteins and peptides of interest are chosen, sometimes as consequence of previous shot-gun experiments. Second, with the application of certain software, such as Skyline (128, 129), pairs of precursor ion-fragment ion (peptide transition) are determined. A method is then created with specific parameters (m/z, retention time), for the selection and measurement of the peptide's transitions, and it's tested in the instrument. Finally, after obtaining the right parameters and the optimal peptide transitions for the desired proteins, the quantification experiment can be performed. The AQUA approach can be used in conjunction with this method, in order to obtain more sensitive and precise results, at the testing and the final experiment stages.

Nowadays, a new generation of targeted methods can be performed in quadrupole time-of-flight (QqTOF – TripleTOF™) (Figure 7) and quadrupole orbitrap (QqOrbi – Q Exactive™) (Figure 6) hybrid instruments, as mentioned before. The first can perform the high-resolution multiple reaction monitoring (MRM^{HR}) and MS/MS^{ALL} SWATH acquisition (or DIA) (100) and the second performs the PRM acquisition (130, 131) (similar to MRM^{HR}) and also DIA.

The PRM and MRM^{HR} methods rely on the high resolution and mass accuracy of the correspondent mass detectors, as well as the fast and sensitive MS2 scanning. These methods use orbitrap or TOF MS for the measurement of all transitions (fragment ions) obtained from Q1 selection and Q2 fragmentation, contrarily to the classic QqQ. After the acquisition, quantification is performed on the extracted ion chromatograms (XICs). This has the advantage of a better confirmation of the peptide of interest, since more transitions are observed; and a shorter pre-analysis time, since there is no need for prior knowledge of the transitions. The latter is even better addressed with the use of DIA methods, making this method a high-throughput target approach with possible retrospective data analysis. On the other hand, the dynamic range, accuracy and precision might be limited for targeted analysis of complex samples, when compared to PRM and MRM^{HR} methods.

3. Proteome-based research

Solely quantifying genes or transcripts does not provide the complete information about the function of those genes. This can only be observed through the quantitative analysis of proteins, in relation with their functional analysis, i.e., analysis of PTMs or cellular localization. A global or large-scale analysis of the proteome allows the study of multiple cellular processes simultaneously, as well as possible crosstalk between them. That said, proteomes can paint a picture of the state and abnormalities in our cells through a systems biology view at the molecular level. However, proteome analysis is challenging, since the level of complexity increases from gene to transcript, and from transcript to the protein level, due to alternative splicing or (post-) translational regulation. In the case of human proteome analysis we could be dealing with at least 100 000 protein isoforms, making proteome studies of higher organisms extremely demanding. An alternative for such challenging approaches, in this case complex proteome studies, is to investigate selected molecular mechanisms in simpler organisms from earlier forms of life, such as bacteria or other unicellular organisms such as *Saccharomyces cerevisiae* (Yeast). These model organisms can be very advantageous for systems biology approaches, since they are easier to obtain/culture and manipulate while preserving molecular features similar to ours, giving us clues for more complex organisms. Indeed, microorganisms such as *Escherichia coli* (*E. coli*) and Yeast are frequently used for probing particular protein families, such as the heat shock proteins (132); for the study of protein networks and interactions (133, 134), including for the mapping of

the human interactomes (135) and for the purpose of drug targets discovery (136); as well as probing post-translational modifications crosstalk (137) and kinase and phosphatase global interactions (138). A similar approach was taken for the work described in this thesis, by choosing a cyanobacterial model.

Introduction to Cyanobacterial circadian rhythm

1. Circadian rhythm definition

The 24h cyclic Earth rotation imposes cyclic behavior and physiological changes in most living organisms. Most of the organisms have adjusted their daily life to this succeeding day and night environmental changes, taking advantage of these phases. The way they adapted to these changes may vary, however higher organisms, such as Plants and Animals, and lower organisms, such as Bacteria and Archaea, have developed a so called circadian clock, which allows the scheduling of different cellular processes in different times of the day. This endogenous biological clock is composed of molecular oscillators which provide a self-sustained mechanism with a ~24h period. In general, these oscillators comprise positive and negative components that form feedback loops, which can receive environmental input in a direct or indirect way, allowing entrainment of the clock with the Earth light/dark cycles (139, 140). The positive loop components activate the transcription of so-called 'clock genes', which encode negative components, which are in turn responsible for the inhibition of the positive components. Consequently, this inhibition will decrease the transcription of the clock genes. At the same time, a phosphorylation-induced degradation of the clock genes will lead to the reactivation of the positive components, restarting the cycle once again. Additionally, certain negative components can also activate the expression of positive components in order to ensure the stability of the feedback loop. In this manner they can coordinate rhythmic outputs, which regulate circadian gene expression and cellular processes (141–143). Moreover, rhythmic transcription was proven to be essential for oscillator function in several organisms (144, 145), while in others like cyanobacteria, it was shown that protein abundances of the clock genes are dispensable (146). Alternatively, there is a core oscillator, or pacemaker, which operates independently and is ultimately responsible for the entrainment of the mechanism, driving the rhythmic output through coordination of other oscillators or by itself. Ultimately, certain circadian mechanism are quite conserved despite the fact that in unicellular organisms there is rhythm within cellular processes and in multicellular organism there is a coordinated rhythmicity among different cell types/tissues.

2. *S. elongatus* circadian clock

The cyanobacterium was one of the first unicellular organisms found to have a circadian rhythm, contradicting the hypothesis that fast diving and single-compartmented organisms couldn't have an endogenous pacemaker (147–151). Cyanobacteria are among the oldest and most abundant organisms on Earth, forming a highly diverse group of prokaryotes. They can inhabit freezing, dry or tropical regions and water, soil or rock surfaces, being also tolerant to a large range of salt and temperature conditions. Their morphology is also variable, since they exist as unicellular ovoid or rod shaped organisms, but also as multicellular filaments comprised of different cell types (152, 153). Some species of these photosynthetic cyanobacteria have to regulate incompatible processes, such as photosynthesis and nitrogen fixation (154, 155), which might explain the necessity for such simple organisms to have evolved towards circadian regulation systems. Among these bacteria the fresh-water *Synechococcus elongatus* PCC7942 (*S. elongatus*) stood out as a circadian rhythm model for having the right genetic tools available, the full structure of the core clock proteins and for being the only organism from which the core clock mechanism can be reproduced *in vitro*.

As a simple organism, *S. elongatus* also possesses the simplest core clock mechanism known to date, comprising a set of three proteins, named KaiA, KaiB, and KaiC (from the Japanese *kaiten* for "turning of the heavens"), which are unique to cyanobacteria species (156, 157). KaiA forms homodimers, KaiB can exist in a monomeric, homodimeric or a homotetrameric state (158) and KaiC forms a homohexamer, with a double doughnut shape (159–161). The mechanism of the core clock depends on the auto-phosphorylation/dephosphorylation of KaiC, which occurs in a period of approximately 24h. First, KaiA binds to KaiC and consequently KaiC becomes sequentially hyperphosphorylated at the T432 and S431. Next, KaiB binds to KaiA-C, which leads to a conformational change of KaiC, enabling a stable KaiA-B-C complex and KaiC hypophosphorylation, first of the T432 and then the S431. Finally, the hypophosphorylated KaiC releases KaiA and KaiB (Figure 12). The cyclic events of the core oscillator can be stably reproduced *in vitro* with the addition of ATP/Mg^{2+}, lasting for at least 10 days without any further external cues (162).

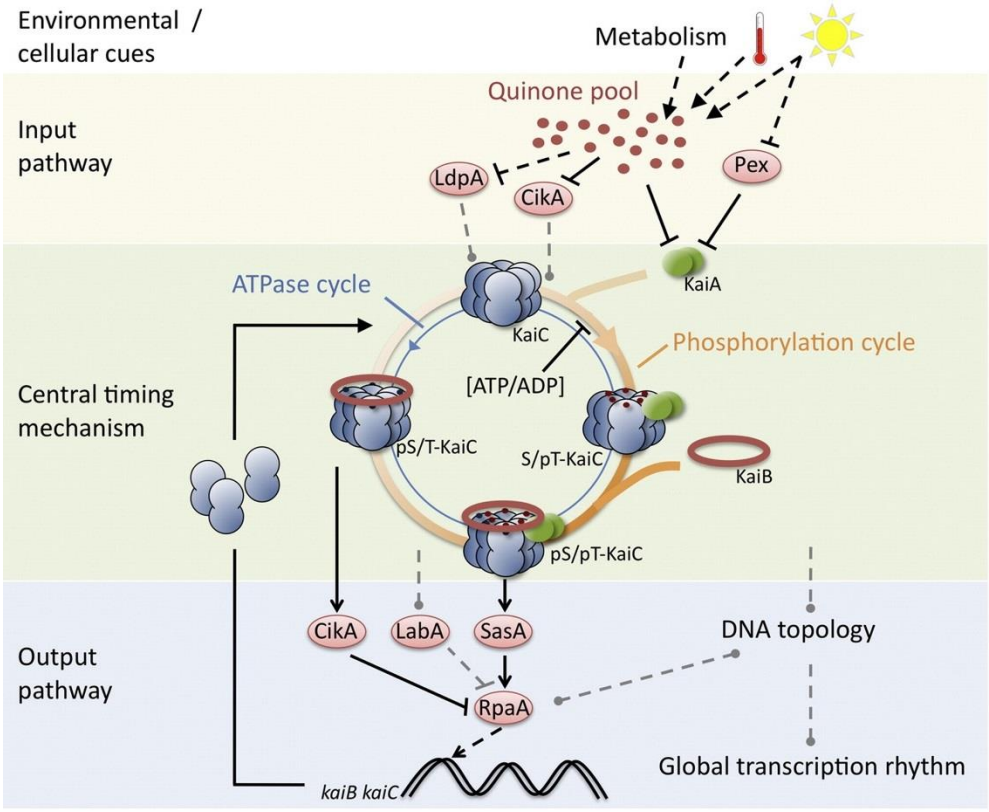


Figure 12 – Schematic representation of the molecular mechanisms comprising the *Synechococcus elongatus* PCC7942 circadian clock. Adopted from Axmann *et al*. (163).

The evidence of a posttranslational oscillator (PTO), which was stably reconstituted *in vitro*, proved that this is the core mechanism driving the circadian rhythm in *S. elongatus*, and contradicted the assumption that all circadian oscillators are dependent on a transcription/translation feedback loop (TTFL) (162). Moreover, it was proven that transcription and translation were also independent of KaiC kinase activity (164), indicating that the TTFL provides input to the core PTO as well. This tightly intertwined system is implicated in the regulation of gene expression (165), metabolism (141), and cell division (151) (Figure 12). However, the exact connection between the central clock and its outputs is still not clear.

Alternative models were proposed for the regulation of rhythmic gene expression, which could ultimately work in unison (166). The oscilloid model

refers to transcription regulation through DNA compaction/decompaction cycles (167) and supercoiling changes (168). The transcription factor model, as the name indicates, refers to transcription factor regulation of the gene expression, including the kaiBC genes. Several output factors were discovered in this case, such as SasA, RpaA, LabA and CikA. The histidine kinase SasA (*Synechococcus* adaptive sensor) was shown to interact with KaiC and with the putative transcription factor RpaA (regulator of phycobilisome-associated) (169, 170). SasA autophosphorylates upon interaction with KaiC and consequently phosphorylates RpaA, which in turn leads to the activation of the kaiBC gene expression (171). This pathway was proposed to be active only during the day (172). On the other hand, CikA (circadian input kinase) is regulated by KaiBC (173) and in a similar way as LabA (low-amplitude and bright) acts as an RpaA inactivator (174), consequently repressing kaiBC gene expression.

There are also input pathways which allow entrainment of the central clock to the environmental time. So far, a mechanism was found which senses redox state changes, including the histidine kinase CikA (circadian input kinase) and LdpA (light dependent period); and a mechanism that is regulated through light changes, which includes Pex (period extender). CikA has a double function, being involved in the output and input pathways. In the latter, CikA is destabilized when binding to quinone (175), similarly to what was shown for KaiA (176). The changes in the quinone pool are directly connected to photosynthesis and consequently, to light changes. CikA was also shown to interact with additional factors (NhtA, PrkE, IrcA, CdpA) that might be part of this input pathway (177). LdpA also senses quinone pool changes in an indirect way (178). It's still not known how CikA and LdpA transduce its signals to the central clock, but it is speculated that could be through interaction with KaiA. Pex expression is suppressed by light and functions as a KaiA gene expression repressor during the night (179). Moreover, it was recently shown that KaiC can sense energy changes through variations of the ATP/ADP ratio (180). The mechanisms behind both metabolic products mentioned (ATP and quinone) seem to act together to reset the PTO according to light changes (181). An elevated ATPase activity of KaiC was suggested to provide a checkpoint that allows cell division (182), additionally it might implicate the CikA and CdpA factors (177, 182). This observation was consequence of the fact that it was shown that the timing of the cell cycle doesn't perturb the one of the circadian rhythm, since the clock of a daughter cell is in phase with the one from the mother cell (150). In addition to the light/dark (LD) entrainment

there is usually a temperature regulation of the core circadian rhythm. In the case of *S. elongatus* only a few *in vitro* studies demonstrated that the temperature influenced the rhythm's phase (183, 184).

3. Cyanobacteria circadian rhythm studies

Cyanobacteria are known to play an important role in the carbon and nitrogen cycles, and they largely contribute to the biomass production of the Earth, including the oxygen we breathe. The interest in cyanobacteria has risen among the scientific community for these and other reasons, including the potential for industrial applications, such as in bio-fuel production, agriculture, drug-discovery and biopharmaceutical precursors. In order to understand the molecular mechanisms of these organisms several postgenomics studies were performed, including transcriptomics and proteomics studies.

Following the publication of the full sequenced genome of several cyanobacteria species several transcriptomics studies based in DNA microarray technology were performed. The different experiments addressed gene expression changes upon high light intensities; phosphate, nitrogen and sulphur deprivation; irradiation of UV-B and white light; addition of electron transport inhibitors (redox response); salt stress; light to dark transition; carbon sequestering; cold stress; osmotic stress; iron-deficiency versus iron-reconstituted cells; adverse environmental conditions (photosynthetic pathway stress response); magnesium concentration regulation (185). A number of microarray circadian studies were also performed. Here, the cyanobacterium is usually grown in constant light (LL) conditions to eliminate the external influence of light changes. Such studies indicated a circadian abundance variation of 10% to 80% of the transcripts, on species such as *Synechocystis* sp. PCC6803, *Crocosphaera watsonii* and *Prochlorococcus MED4*. For *S. elongatus* in particular, the microarray analysis demonstrated a frequency of rhythmic abundant transcripts of 30% to 60% (186, 187). This result shows a certain discrepancy with a previous bioluminescence study effectuated using the luciferase protein as a reporter and observing rhythms in the luminescence from the cells, which revealed that almost all promotors are regulated in a circadian manner (188). On the other hand, it was also shown that not all genes with circadian expression regulation are associated with processes which present circadian rhythm (in LL), but only a daily rhythm (in LD) (189).

Studies involving protein level regulation are not as common and are not yet as comprehensive as transcriptomics studies. Early protein studies in cyanobacteria focused in the study of specific enzymes and their corresponding functions, showing only isolated responses. But, a more systemic view of the cell phenotype is important for understanding cellular responses. Initial proteomics studies focused mainly on the analysis of the plasma membrane and thylakoid proteins, regulation of microcystin production, the effects of UV-B stimulation, sat stress, cell division or differentiation, light induction and photosynthetic process activity within several highly active photosynthetic mutants. These studies were only based on small numbers of proteins, being identified around 105 (190) or 234 proteins (191) at most in *Synechocystis* sp. PCC 6803, through the use of 2D PAGE and N-terminal sequencing techniques. Initial high-throughput shot-gun proteomics approaches using 2D-LC or 3D-LC (SEC and SCX-RP-LC) methods and ESI-QqTOF mass spectrometry techniques, took a step forward enabling the identification of around 1500 proteins in *Nostoc punctiforme* ATCC 29133 (192). On the other hand, quantitative studies based on the use isobaric tags allowed quantification of up to 702 proteins in the same species (193). Further advances in proteomics techniques led to an increase of proteome coverage and the attempt to perform a global profiling of cyanobacterial proteins. Recently, such studies were performed for several cyanobacteria species, such as *Synechocystis* (194), *Anabaena* (195), *Cyanothece* (196), and *Prochlorococcus* (197), the last two being focused on diurnal rhythms of the proteome. Since transcriptome studies were previously conducted in many different cyanobacteria species, there was also an interest to combine those with the new proteome data to infer on transcript *versus* protein synthesis regulation. In *Cyanothece*, LL and light/dark (LD) experiments were performed, using metabolic labeling (SILAC) (198, 199) or spectral counts for protein quantification. In the case of the latter, up to 3616 proteins (68.2% of the predicted genome) were quantified across a combination of several different growth conditions and fractionation methods, on soluble or membrane cell compartments, revealing that no more than 20% of those (5% of the predicted genome) possess circadian rhythms in abundance. Through the conjunction of previously obtained mRNA data, it was possible to infer that the role of post-translational processes can be quite significant for the variances in protein abundance. In the *Prochlorococcus* work, only LL conditions were tested and a metabolic labeling strategy was used. Here, only the genes for which both the protein and mRNA showed circadian rhythm expression were considered, corresponding to 312 genes.

In this study it was determined that there are differences in amplitude and in the phase of the abundance profiles between proteins and transcripts. Similarly to the *Cyanothece* study, it was also observed a variance in the timing of protein expression compared to mRNA. Besides the interesting findings concerning the circadian regulation at the protein and mRNA levels, these studies also highlight the discrepancy between these two levels of expression and the fact that one cannot always rely solely on one of them to make conclusions. A dynamic proteome analysis of *S. elongatus* was only recently performed and will be described in Chapter 2.
The studies of the cyanobacterial circadian clock are useful for understanding these organisms themselves. Nevertheless, they also shed a light into the origin of circadian rhythms and therefore might teach us something about our own circadian clocks and the mechanisms involved.

Outline

The work described in the coming chapters comprises the proteome analysis of the cyanobacteria *S. elongatus* in light and dark conditions, using a variety of quantitative proteomics strategies.

In chapter 2, high-resolution MS is combined with the TMT 6-plex labeling technique to probe cyclic protein abundance variations across 82% of the *S. elongatus* proteome. This 48-hour time-series study revealed that, in contrast to abundant occurrences of cyclic and even circadian rhythms at the transcript level (30-60%), at the protein level these variations are much less pronounced, with only 5% of the quantified proteome showing significant cyclic variations.

In the next chapter, the focus moved towards the analysis of protein interactions and complexes. A novel approach, combining native size exclusion chromatography and mass spectrometry, was used to uncover differences in protein assemblies between light and dark states. Several proteins involved in higher order complexes, which didn't show diurnal adaptation in protein abundance in the previous chapter, were found to have association dynamics. Some examples are the photosynthetic and ribosomal proteins, which showed differences in abundances as well as variable assemblies.

In chapter 4, a first look at the phosphoproteome of *S. elongatus* is described. The combination of MS, Ti^{4+}-IMAC phosphopeptide enrichment and TMT 6-

plex labeling was explored to uncover clues for regulation of molecular rhythms by protein phosphorylation. This approach revealed new phosphorylation sites on proteins associated to diverse molecular pathways.

Finally, an outlook on the use of quantitative mass spectrometric techniques is given in the fifth chapter. Recent trends and future perspectives on whole proteome analysis are discussed, with emphasis on its application to the study of cyanobacteria and circadian research. This outlook describes improvements in data collection and interpretation, which, in combination with the study of different levels of gene expression, can stimulate meaningful investigations in this fascinating research area.

References

1. Abbott, A. (1999) A post-genomic challenge: learning to read patterns of protein synthesis. Nature 402, 715–20
2. Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J. C., Yan, J. X., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L., and Hochstrasser, D. F. (1996) From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology. (N. Y).* 14, 61–65
3. Steen, H., and Mann, M. (2004) The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* 5, 699–711
4. Yates, J. R., Ruse, C. I., and Nakorchevsky, A. (2009) Proteomics by mass spectrometry: approaches, advances, and applications. *Annu. Rev. Biomed. Eng.* 11, 49–79
5. Altelaar, a F. M., Munoz, J., and Heck, A. J. R. (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* 14, 35–48
6. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* 422, 198–207
7. Tran, J. C., Zamdborg, L., Ahlf, D. R., Lee, J. E., Catherman, A. D., Durbin, K. R., Tipton, J. D., Vellaichamy, A., Kellie, J. F., Li, M., Wu, C., Sweet, S. M. M., Early, B. P., Siuti, N., LeDuc, R. D., Compton, P. D., Thomas, P. M., and Kelleher, N. L. (2011) Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* 480, 254–8
8. Young, N. L., DiMaggio, P. A., Plazas-Mayorca, M. D., Baliban, R. C., Floudas, C. A., and Garcia, B. A. (2009) High throughput characterization of combinatorial histone codes. *Mol. Cell. Proteomics* 8, 2266–2284
9. Olsen, J. V, Ong, S.-E., and Mann, M. (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* 3, 608–614
10. Bian, Y., Ye, M., Song, C., Cheng, K., Wang, C., Wei, X., Zhu, J., Chen, R., Wang, F., and Zou, H. (2012) Improve the coverage for the analysis of

phosphoproteome of HeLa cells by a tandem digestion approach. *J. Proteome Res.* 11, 2828–2837

11. Wiśniewski, J. R., and Mann, M. (2012) Consecutive proteolytic digestion in an enzyme reactor increases depth of proteomic and phosphoproteomic analysis. *Anal. Chem.* 84, 2631–2637
12. Glatter, T., Ludwig, C., Ahrné, E., Aebersold, R., Heck, A. J. R., and Schmidt, A. (2012) Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *J. Proteome Res.* 11, 5145–56
13. Gauci, S., Helbig, A. O., Slijper, M., Krijgsveld, J., Heck, A. J. R., and Mohammed, S. (2009) Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. *Anal. Chem.* 81, 4493–501
14. Isobe, T., Takayasu, T., Takai, N., and Okuyama, T. (1982) High-performance liquid chromatography of peptides on a macroreticular cation-exchange resin: Application to peptide mapping of Bence-Jones proteins. *Anal. Biochem.* 122, 417–425
15. Cachia, P. J., van Eyk, J., Chong, P. C. S., Taneja, A., and Hodges, R. S. (1983) Separation of basic peptides by cation-exchange high-performance liquid chromatography. *J. Chromatogr. A* 266, 651–659
16. Mant, C. T., and Hodges, R. S. (1985) Separation of peptides by strong cation-exchange high-performance liquid chromatography. *J. Chromatogr.* 327, 147–55
17. Alpert, A. J., and Andrews, P. C. (1988) Cation-exchange chromatography of peptides on poly(2-sulfoethyl aspartamide)-silica. *J. Chromatogr.* 443, 85–96
18. Gruber, K. A., Stein, S., Brink, L., Radhakrishnan, A., and Udenfriend, S. (1976) Fluorometric assay of vasopressin and oxytocin: a general approach to the assay of peptides in tissues. *Proc. Natl. Acad. Sci. U. S. A.* 73, 1314–8
19. Shen, Y., Zhang, R., Moore, R. J., Kim, J., Metz, T. O., Hixson, K. K., Zhao, R., Livesay, E. A., Udseth, H. R., and Smith, R. D. (2005) Automated 20 kpsi RPLC-MS and MS/MS with chromatographic peak capacities of 1000-1500 and capabilities in proteomics and metabolomics. *Anal. Chem.* 77, 3090–3100
20. Guo, D., Mant, C. T., Taneja, A. K., Parker, J. M. R., and Rodges, R. S. (1986) Prediction of peptide retention time in reversed-phase high-performance liquid chromatography I. Determination of retention coefficients of amino acid residues of model synthetic peptides. *J. Chromatogr. A* 359, 499–518
21. Washburn, M. P., Wolters, D., and Yates, J. R. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* 19, 242–247
22. Wolters, D. A., Washburn, M. P., and Yates, J. R. (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* 73, 5683–5690
23. Motoyama, A., Venable, J. D., Ruse, C. I., and Yates, J. R. (2006) Automated ultra-high-pressure multidimensional protein identification technology (UHP-MudPIT) for improved peptide identification of proteomic samples. *Anal. Chem.* 78, 5109–5118

24. Fränzel, B., and Wolters, D. A. (2011) Advanced MudPIT as a next step toward high proteome coverage. *Proteomics* 11, 3651–3656
25. Gilar, M., Olivova, P., Daly, A. E., and Gebler, J. C. (2005) Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J. Sep. Sci.* 28, 1694–1703
26. Delmotte, N., Lasaosa, M., Tholey, A., Heinzle, E., and Huber, C. G. (2007) Two-dimensional reversed-phase x ion-pair reversed-phase HPLC: An alternative approach to high-resolution peptide separation for shotgun proteome analysis. *J. Proteome Res.* 6, 4363–4373
27. Toll, H., Oberacher, H., Swart, R., and Huber, C. G. (2005) Separation, detection, and identification of peptides by ion-pair reversed-phase high-performance liquid chromatography-electrospray ionization mass spectrometry at high and low pH. *J. Chromatogr. A* 1079, 274–286
28. Manadas, B., English, J. A., Wynne, K. J., Cotter, D. R., and Dunn, M. J. (2009) Comparative analysis of OFFGel, strong cation exchange with pH gradient, and RP at high pH for first-dimensional separation of peptides from a membrane-enriched protein fraction. *Proteomics* 9, 5194–5198
29. Gilar, M., Olivova, P., Daly, A. E., and Gebler, J. C. (2005) Orthogonality of separation in two-dimensional liquid chromatography. *Anal. Chem.* 77, 6426–6434
30. Song, C., Ye, M., Han, G., Jiang, X., Wang, F., Yu, Z., Chen, R., and Zou, H. (2010) Reversed-phase-reversed-phase liquid chromatography approach with high orthogonality for multidimensional separation of phosphopeptides. *Anal. Chem.* 82, 53–56
31. Wang, Y., Yang, F., Gritsenko, M. A., Wang, Y., Clauss, T., Liu, T., Shen, Y., Monroe, M. E., Lopez-Ferrer, D., Reno, T., Moore, R. J., Klemke, R. L., Camp, D. G., and Smith, R. D. (2011) Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* 11, 2019–2026
32. Alpert, A. J. (1990) Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J. Chromatogr.* 499, 177–196
33. Boersema, P. J., Divecha, N., Heck, A. J. R., and Mohammed, S. (2007) Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. *J. Proteome Res.* 6, 937–946
34. Guo, Y., and Gaiki, S. (2005) Retention behavior of small polar compounds on polar stationary phases in hydrophilic interaction chromatography. *J. Chromatogr. A* 1074, 71–80
35. Guo, Y., and Gaiki, S. (2011) Retention and selectivity of stationary phases for hydrophilic interaction chromatography. *J. Chromatogr. A* 1218, 5920–5938
36. Jandera, P. (2011) Stationary and mobile phases in hydrophilic interaction chromatography: A review. *Anal. Chim. Acta* 692, 1–25
37. Jandera, P. (2008) Stationary phases for hydrophilic interaction chromatography, their characterization and implementation into multidimensional chromatography concepts. *J. Sep. Sci.* 31, 1421–1437

38. Yoshida, T. (1998) Calculation of peptide retention coefficients in normal-phase liquid chromatography. *J. Chromatogr. A* 808, 105–112
39. Lindner, H., Sarg, B., Meraner, C., and Helliger, W. (1996) in *Journal of Chromatography A*, pp 137–144.
40. Alpert, A. J. (2008) Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. *Anal. Chem.* 80, 62–76
41. Intoh, A., Kurisaki, A., Fukuda, H., and Asashima, M. (2009) Separation with zwitterionic hydrophilic interaction liquid chromatography improves protein identification by matrix-assisted laser desorption/ionization-based proteomic analysis. *Biomed. Chromatogr.* 23, 607–14
42. Jiang, W., and Irgum, K. (2002) Tentacle-type zwitterionic stationary phase prepared by surface-initiated graft polymerization of 3-[N,N-dimethyl-N-(methacryloyloxyethyl)-ammonium] propanesulfonate through peroxide groups tethered on porous silica. *Anal. Chem.* 74, 4682–4687
43. Di Palma, S., Boersema, P. J., Heck, A. J. R., and Mohammed, S. (2011) Zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC and ZIC-cHILIC) provide high resolution separation and increase sensitivity in proteome analysis. *Anal. Chem.* 83, 3440–7
44. Scherl, A., Couté, Y., Déon, C., Callé, A., Kindbeiter, K., Sanchez, J.-C., Greco, A., Hochstrasser, D., and Diaz, J.-J. (2002) Functional proteomic analysis of human nucleolus. *Mol. Biol. Cell* 13, 4100–4109
45. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685
46. Bossi, A., Righetti, P. G., and Chiari, M. Immobilized pH gradients: new pK values of acrylamido buffers in poly(N-acryloylaminoethoxyethanol) matrices. *Electrophoresis* 15, 1112–7
47. Bjellqvist, B., Ek, K., Righetti, P. G., Gianazza, E., Görg, A., Westermeier, R., and Postel, W. (1982) Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *J. Biochem. Biophys. Methods* 6, 317–339
48. Kenrick, K. G., and Margolis, J. (1970) Isoelectric focusing and gradient gel electrophoresis: A two-dimensional technique. *Anal. Biochem.* 33, 204–207
49. Tonella, L., Walsh, B. J., Sanchez, J. C., Ou, K., Wilkins, M. R., Tyler, M., Frutiger, S., Gooley, A. A., Pescaru, I., Appel, R. D., Yan, J. X., Bairoch, A., Hoogland, C., Morch, F. S., Hughes, G. J., Williams, K. L., and Hochstrasser, D. F. (1998) '98 Escherichia coli SWISS-2DPAGE database update. *Electrophoresis* 19, 1960–1971
50. Kristensen, A. R., Gsponer, J., and Foster, L. J. (2012) A high-throughput approach for measuring temporal changes in the interactome. *Nat. Methods* 9, 907–9
51. Havugimana, P. C., Hart, G. T., Nepusz, T., Yang, H., Turinsky, A. L., Li, Z., Wang, P. I., Boutz, D. R., Fong, V., Phanse, S., Babu, M., Craig, S. A., Hu, P., Wan, C., Vlasblom, J., Dar, V.-N., Bezginov, A., Clark, G. W., Wu, G. C., Wodak, S. J., Tillier, E. R. M., Paccanaro, A., Marcotte, E. M., and Emili, A. (2012) A census of human soluble protein complexes. *Cell* 150, 1068–81

52. Villén, J., and Gygi, S. P. (2008) The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat. Protoc.* 3, 1630–8
53. Li, X., Gerber, S. A., Rudner, A. D., Beausoleil, S. A., Haas, W., Villén, J., Elias, J. E., and Gygi, S. P. (2007) Large-scale phosphorylation analysis of alpha-factor-arrested Saccharomyces cerevisiae. *J. Proteome Res.* 6, 1190–7
54. Posewitz, M. C., and Tempst, P. (1999) Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal. Chem.* 71, 2883–92
55. Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., and Jørgensen, T. J. D. (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell. Proteomics* 4, 873–886
56. Kweon, H. K., and Håkansson, K. (2006) Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. *Anal. Chem.* 78, 1743–1749
57. Pinkse, M. W. H., Uitto, P. M., Hilhorst, M. J., Ooms, B., and Heck, A. J. R. (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal. Chem.* 76, 3935–43
58. Ficarro, S. B., McCleland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F., and White, F. M. (2002) Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat. Biotechnol.* 20, 301–305
59. Thingholm, T. E., Jensen, O. N., Robinson, P. J., and Larsen, M. R. (2008) SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. *Mol. Cell. Proteomics* 7, 661–671
60. Zhou, H., Ye, M., Dong, J., Han, G., Jiang, X., Wu, R., and Zou, H. (2008) Specific phosphopeptide enrichment with immobilized titanium ion affinity chromatography adsorbent for phosphoproteome analysis. *J. Proteome Res.* 7, 3957–67
61. Zhou, H., Ye, M., Dong, J., Corradini, E., Cristobal, A., Heck, A. J. R., Zou, H., and Mohammed, S. (2013) Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. *Nat. Protoc.* 8, 461–80
62. Ruprecht, B., Koch, H., Medard, G., Mundt, M., Kuster, B., and Lemeer, S. (2015) Comprehensive and reproducible phosphopeptide enrichment using iron immobilized metal ion affinity chromatography (Fe-IMAC) columns. *Mol. Cell. Proteomics* 14, 205–15
63. Gauci, S., Helbig, A. O., Slijper, M., Krijgsveld, J., Heck, A. J. R., and Mohammed, S. (2009) Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. *Anal. Chem.* 81, 4493–4501

64. Hennrich, M. L., Van Den Toorn, H. W. P., Groenewold, V., Heck, A. J. R., and Mohammed, S. (2012) Ultra acidic strong cation exchange enabling the efficient enrichment of basic phosphopeptides. *Anal. Chem.* 84, 1804–1808
65. Han, G., Ye, M., Zhou, H., Jiang, X., Feng, S., Jiang, X., Tian, R., Wan, D., Zou, H., and Gu, J. (2008) Large-scale phosphoproteome analysis of human liver tissue by enrichment and fractionation of phosphopeptides with strong anion exchange chromatography. *Proteomics* 8, 1346–61
66. Hennrich, M. L., Groenewold, V., Kops, G. J. P. L., Heck, A. J. R., and Mohammed, S. (2011) Improving depth in phosphoproteomics by using a strong cation exchange-weak anion exchange-reversed phase multidimensional separation approach. *Anal. Chem.* 83, 7137–7143
67. Palmisano, G., Lendal, S. E., Engholm-Keller, K., Leth-Larsen, R., Parker, B. L., and Larsen, M. R. (2010) Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. *Nat. Protoc.* 5, 1974–1982
68. McNulty, D. E., and Annan, R. S. (2008) Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. *Mol. Cell. Proteomics* 7, 971–980
69. Wu, C. J., Chen, Y. W., Tai, J. H., and Chen, S. H. (2011) Quantitative phosphoproteomics studies using stable isotope dimethyl labeling coupled with IMAC-HILIC-nanoLC-MS/MS for estrogen-induced transcriptional regulation. *J. Proteome Res.* 10, 1088–1097
70. Scott, N. E., Parker, B. L., Connolly, A. M., Paulech, J., Edwards, A. V. G., Crossett, B., Falconer, L., Kolarich, D., Djordjevic, S. P., Højrup, P., Packer, N. H., Larsen, M. R., and Cordwell, S. J. (2011) Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of Campylobact. *Mol. Cell. Proteomics* 10, M000031–MCP201
71. Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V, and Mann, M. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325, 834–840
72. Zhang, K., Tian, S., and Fan, E. (2013) Protein lysine acetylation analysis: current MS-based proteomic technologies. *Analyst* 138, 1628–36
73. Guo, A., Gu, H., Zhou, J., Mulhern, D., Wang, Y., Lee, K. a, Yang, V., Aguiar, M., Kornhauser, J., Jia, X., Ren, J., Beausoleil, S. a, Silva, J. C., Vemulapalli, V., Bedford, M. T., and Comb, M. J. (2014) Immunoaffinity enrichment and mass spectrometry analysis of protein methylation. *Mol. Cell. Proteomics* 13, 372–87
74. Biggar, K. K., and Li, S. S.-C. (2015) Non-histone protein methylation as a regulator of cellular signalling and function. *Nat. Rev. Mol. Cell Biol.* 16, 5–17
75. Grønborg, M., Kristiansen, T. Z., Stensballe, A., Andersen, J. S., Ohara, O., Mann, M., Jensen, O. N., and Pandey, A. (2002) A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a

novel protein, Frigg, as a protein kinase A substrate. *Mol. Cell. Proteomics* 1, 517–27
76. Zhang, H., Zha, X., Tan, Y., Hornbeck, P. V, Mastrangelo, A. J., Alessi, D. R., Polakiewicz, R. D., and Comb, M. J. (2002) Phosphoprotein analysis using antibodies broadly reactive against phosphorylated motifs. *J. Biol. Chem.* 277, 39379–87
77. Moritz, A., Li, Y., Guo, A., Villen, J., Wang, Y., MacNeill, J., Kornhauser, J., Sprott, K., Zhou, J., Possemato, A., Ren, J. M., Hornbeck, P., Cantley, L. C., Gygi, S. P., Rush, J., and Comb, M. J. (2010) Akt-RSK-S6 Kinase Signaling Networks Activated by Oncogenic Receptor Tyrosine Kinases. *Sci. Signal.* 3, ra64–ra64
78. Rush, J., Moritz, A., Lee, K. A., Guo, A., Goss, V. L., Spek, E. J., Zhang, H., Zha, X.-M., Polakiewicz, R. D., and Comb, M. J. (2005) Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat. Biotechnol.* 23, 94–101
79. Boersema, P. J., Foong, L. Y., Ding, V. M. Y., Lemeer, S., van Breukelen, B., Philp, R., Boekhorst, J., Snel, B., den Hertog, J., Choo, A. B. H., and Heck, A. J. R. (2010) In-depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immunoaffinity purification and stable isotope dimethyl labeling. *Mol. Cell. Proteomics* 9, 84–99
80. Kettenbach, A. N., and Gerber, S. A. (2011) Rapid and reproducible single-stage phosphopeptide enrichment of complex peptide mixtures: Application to general and phosphotyrosine-specific phosphoproteomics experiments. *Anal. Chem.* 83, 7635–7644
81. Gomez, A., and Tang, K. (1994) Charge and fission of droplets in electrostatic sprays. *Phys. Fluids* 6, 404–414
82. Cohen, S. L., and Chait, B. T. (1996) Influence of matrix solution conditions on the MALDI-MS analysis of peptides and proteins. *Anal. Chem.* 68, 31–37
83. W., P., and H., S. (1953) Ein neues Massenspektrometer ohne Magnetfeld. *Z. Naturforschg* 8a, 448
84. Hoffmann, E. de, and Stroobant, V. (2007) *Mass spectrometry : principles and applications* (WILEY, West Sussex). 3rd Ed.
85. Hager, J. W., and Le Blanc, J. C. Y. (2003) High-performance liquid chromatography–tandem mass spectrometry with a new quadrupole/linear ion trap instrument. *J. Chromatogr. A* 1020, 3–9
86. Douglas, D. J., Frank, A. J., and Mao, D. (2005) Linear ion traps in mass spectrometry. *Mass Spectrom. Rev.* 24, 1–29
87. Schwartz, J. C., Senko, M. W., and Syka, J. E. P. (2002) A two-dimensional quadrupole ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.* 13, 659–669
88. Page, J. S., Bogdanov, B., Vilkov, A. N., Prior, D. C., Buschbach, M. A., Tang, K., and Smith, R. D. (2005) Automatic gain control in mass spectrometry using a jet disrupter electrode in an electrodynamic ion funnel. *J. Am. Soc. Mass Spectrom.* 16, 244–253
89. Schwartz, J. C., and Senko, M. W. (2002) F OCUS : Q UADRUPOLE I ON T RAPS A Two-Dimensional Quadrupole Ion Trap Mass Spectrometer. 0305,

90. Stephens, W. E. (1946) A pulsed mass spectrometer with time dispersion. *Phys. Rev.* 69, 691–691
91. Grix, R., Kutscher, R., Li, G., Grüner, U., Wollnik, H., and Matsuda, H. (1988) A time-of-flight mass analyzer with high resolving power. *Rapid Commun. Mass Spectrom.* 2, 83–85
92. Makarov, A. (2000) Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Anal. Chem.* 72, 1156–1162
93. Marshall, A. G., and Hendrickson, C. L. (2002) Fourier transform ion cyclotron resonance detection: Principles and experimental configurations. *Int. J. Mass Spectrom.* 215, 59–75
94. Olsen, J. V, Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E. R., Senko, M., Makarov, A., Mann, M., and Horning, S. (2009) A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteomics* 8, 2759–69
95. Olsen, J. V, Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E. R., Senko, M., Makarov, A., Mann, M., and Horning, S. (2009) A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteomics* 8, 2759–69
96. Swaney, D. L., McAlister, G. C., and Coon, J. J. (2008) Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat. Methods* 5, 959–964
97. Frese, C. K., Altelaar, a F. M., Hennrich, M. L., Nolting, D., Zeller, M., Griep-Raming, J., Heck, A. J. R., and Mohammed, S. (2011) Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J. Proteome Res.* 10, 2377–88
98. Michalski, A., Damoc, E., Hauschild, J.-P., Lange, O., Wieghaus, A., Makarov, A., Nagaraj, N., Cox, J., Mann, M., and Horning, S. (2011) Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol. Cell. Proteomics* 10, M111.011015
99. Andrews, G. L., Simons, B. L., Young, J. B., Hawkridge, A. M., and Muddiman, D. C. (2011) Performance Characteristics of a New Hybrid Quadrupole (TripleTOF 5600). *Anal. Chem.*, 5442–5446
100. Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012) Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* 11, O111.016717
101. McLuckey, S. A. (1992) Principles of collisional activation in analytical mass spectrometry. *J. Am. Soc. Mass Spectrom.* 3, 599–614
102. Sleno, L., and Volmer, D. A. (2004) Ion activation methods for tandem mass spectrometry. *J. Mass Spectrom.* 39, 1091–1112
103. Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 101, 9528–33

104. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551–67
105. Eng, J. K., McCormack, A. L., and Yates, J. R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* 5, 976–89
106. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–72
107. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011) Andromeda: A peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 10, 1794–1805
108. Elias, J. E., and Gygi, S. P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* 4, 207–214
109. Bantscheff, M., Lemeer, S., Savitski, M. M., and Kuster, B. (2012) Quantitative mass spectrometry in proteomics: Critical review update from 2007 to the present. *Anal. Bioanal. Chem.* 404, 939–965
110. Chelius, D., and Bondarenko, P. V. (2002) Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J. Proteome Res.* 1, 317–323
111. Liu, H., Sadygov, R. G., and Yates, J. R. (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* 76, 4193–4201
112. Ong, S.-E. (2002) Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteomics* 1, 376–386
113. Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R., and Mann, M. (2010) Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat. Methods* 7, 383–5
114. Hsu, J. L., Huang, S. Y., Chow, N. H., and Chen, S. H. (2003) Stable-Isotope Dimethyl Labeling for Quantitative Proteomics. *Anal. Chem.* 75, 6843–6852
115. Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S., and Heck, A. J. R. (2009) Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protoc.* 4, 484–94
116. Altelaar, A. F. M., Frese, C. K., Preisinger, C., Hennrich, M. L., Schram, A. W., Timmers, H. T. M., Heck, A. J. R., and Mohammed, S. (2012) Benchmarking stable isotope labeling based quantitative proteomics. *J. Proteomics*,
117. 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., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* 3, 1154–69
118. Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., and Hamon, C. (2003) Tandem mass tags: A novel

quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* 75, 1895–1904

119. Ting, L., Rad, R., Gygi, S. P., and Haas, W. (2011) MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat. Methods* 8, 937–40
120. Gerber, S. a, Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U. S. A.* 100, 6940–5
121. Pratt, J. M., Simpson, D. M., Doherty, M. K., Rivers, J., Gaskell, S. J., and Beynon, R. J. (2006) Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nat. Protoc.* 1, 1029–1043
122. Brun, V., Dupuis, A., Adrait, A., Marcellin, M., Thomas, D., Court, M., Vandenesch, F., and Garin, J. (2007) Isotope-labeled protein standards: toward absolute quantitative proteomics. *Mol. Cell. Proteomics* 6, 2139–2149
123. Grossmann, J., Roschitzki, B., Panse, C., Fortes, C., Barkow-Oesterreicher, S., Rutishauser, D., and Schlapbach, R. (2010) Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods. *J. Proteomics* 73, 1740–1746
124. Silva, J. C., Gorenstein, M. V, Li, G.-Z., Vissers, J. P. C., and Geromanos, S. J. (2006) Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol. Cell. Proteomics* 5, 144–156
125. Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011) Global quantification of mammalian gene expression control. *Nature* 473, 337–342
126. Gallien, S., Duriez, E., and Domon, B. (2011) Selected reaction monitoring applied to proteomics. *J. Mass Spectrom.* 46, 298–312
127. Lange, V., Picotti, P., Domon, B., and Aebersold, R. (2008) Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* 4, 222
128. Prakash, A., Tomazela, D. M., Frewen, B., Maclean, B., Merrihew, G., Peterman, S., and Maccoss, M. J. (2009) Expediting the Development of Targeted SRM Assays : Using Data from Shotgun Proteomics to Automate Method Development research articles. 2733–2739
129. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26, 966–8
130. Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S., and Coon, J. J. (2012) Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol. Cell. Proteomics*,
131. Gallien, S., Duriez, E., Crone, C., Kellmann, M., Moehring, T., and Domon, B. (2012) Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Mol. Cell. Proteomics* 11, 1709–23
132. Lemaux, P. G., Herendeen, S. L., Bloch, P. L., and Neidhardt, F. C. (1978) Transient rates of synthesis of individual polypeptides in E. coli following temperature shifts. *Cell* 13, 427–434

133. Butland, G., Peregrín-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., Davey, M., Parkinson, J., Greenblatt, J., and Emili, A. (2005) Interaction network containing conserved and essential protein complexes in Escherichia coli. *Nature* 433, 531–7
134. Hu, J. C., Kornacker, M. G., and Hochschild, A. (2000) Escherichia coli one- and two-hybrid systems for the analysis and identification of protein-protein interactions. *Methods* 20, 80–94
135. Rolland, T., Taşan, M., Charloteaux, B., Pevzner, S. J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R., Kamburov, A., Ghiassian, S. D., Yang, X., Ghamsari, L., Balcha, D., Begg, B. E., Braun, P., Brehme, M., Broly, M. P., Carvunis, A.-R., Convery-Zupan, D., Corominas, R., Coulombe-Huntington, J., Dann, E., Dreze, M., Dricot, A., Fan, C., Franzosa, E., Gebreab, F., Gutierrez, B. J., Hardy, M. F., Jin, M., Kang, S., Kiros, R., Lin, G. N., Luck, K., MacWilliams, A., Menche, J., Murray, R. R., Palagi, A., Poulin, M. M., Rambout, X., Rasla, J., Reichert, P., Romero, V., Ruyssinck, E., Sahalie, J. M., Scholz, A., Shah, A. A., Sharma, A., Shen, Y., Spirohn, K., Tam, S., Tejeda, A. O., Trigg, S. A., Twizere, J.-C., Vega, K., Walsh, J., Cusick, M. E., Xia, Y., Barabási, A.-L., Iakoucheva, L. M., Aloy, P., De Las Rivas, J., Tavernier, J., Calderwood, M. A., Hill, D. E., Hao, T., Roth, F. P., and Vidal, M. (2014) A Proteome-Scale Map of the Human Interactome Network. *Cell* 159, 1212–1226
136. Becker, F., Murthi, K., Smith, C., Come, J., Costa-Roldán, N., Kaufmann, C., Hanke, U., Degenhart, C., Baumann, S., Wallner, W., Huber, A., Dedier, S., Dill, S., Kinsman, D., Hediger, M., Bockovich, N., Meier-Ewert, S., Kluge, A. F., and Kley, N. (2004) A three-hybrid approach to scanning the proteome for targets of small molecule kinase inhibitors. *Chem. Biol.* 11, 211–23
137. Swaney, D. L., Beltrao, P., Starita, L., Guo, A., Rush, J., Fields, S., Krogan, N. J., and Villén, J. (2013) Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. *Nat. Methods* 10, 676–82
138. Breitkreutz, A., Choi, H., Sharom, J. R., Boucher, L., Neduva, V., Larsen, B., Lin, Z.-Y., Breitkreutz, B.-J., Stark, C., Liu, G., Ahn, J., Dewar-Darch, D., Reguly, T., Tang, X., Almeida, R., Qin, Z. S., Pawson, T., Gingras, A.-C., Nesvizhskii, A. I., and Tyers, M. (2010) A global protein kinase and phosphatase interaction network in yeast. *Science* 328, 1043–6
139. Reppert, S. M., and Weaver, D. R. (2002) Coordination of circadian timing in mammals. *Nature* 418, 935–941
140. Young, M. W., and Kay, S. A. (2001) Time zones: a comparative genetics of circadian clocks. *Nat. Rev. Genet.* 2, 702–715
141. Hellwegera, F. L. (2010) Resonating circadian clocks enhance fitness in cyanobacteria in silico. *Ecol. Modell.* 221, 1620–1629
142. Qin, X., Byrne, M., Xu, Y., Mori, T., and Johnson, C. H. (2010) Coupling of a core post-translational pacemaker to a slave transcription/translation feedback loop in a circadian system. *PLoS Biol.* 8, e1000394
143. Yang, Q., Pando, B. F., Dong, G., Golden, S. S., and van Oudenaarden, A. (2010) Circadian gating of the cell cycle revealed in single cyanobacterial cells. *Science* 327, 1522–1526

144. Hardin, P. E. (2004) Transcription regulation within the circadian clock: the E-box and beyond. *J. Biol. Rhythms* 19, 348–360
145. Ditty, J. L., Williams, S. B., and Golden, S. S. (2003) A cyanobacterial circadian timing mechanism. *Annu. Rev. Genet.* 37, 513–543
146. Tomita, J., Nakajima, M., Kondo, T., and Iwasaki, H. (2005) No transcription-translation feedback in circadian rhythm of KaiC phosphorylation. *Science* 307, 251–254
147. Pittendrigh, C. S. (1993) Temporal organization: reflections of a Darwinian clock-watcher. *Annu. Rev. Physiol.* 55, 16–54
148. Johnson, C. H. (2010) Circadian clocks and cell division: what's the pacemaker? *Cell Cycle* 9, 3864–73
149. Kondo, T., Mori, T., Lebedeva, N. V, Aoki, S., Ishiura, M., and Golden, S. S. (1997) Circadian rhythms in rapidly dividing cyanobacteria. *Science* 275, 224–7
150. Mihalcescu, I., Hsing, W., and Leibler, S. (2004) Resilient circadian oscillator revealed in individual cyanobacteria. *Nature* 430, 81–5
151. Mori, T., Binder, B., and Johnson, C. H. (1996) Circadian gating of cell division in cyanobacteria growing with average doubling times of less than 24 hours. *Proc. Natl. Acad. Sci. U. S. A.* 93, 10183–8
152. Green, J. W., Knoll, A. H., and Swett, K. (1989) Microfossils from silicified stromatolitic carbonates of the Upper Proterozoic Limestone-Dolomite "Series", central East Greenland. *Geol. Mag.* 126, 567
153. Shi, L., Carmichael, W. W., and Miller, I. (1995) Immuno-gold localization of hepatotoxins in cyanobacterial cells. *Arch. Microbiol.* 163, 7–15
154. Millineaux, P. M., Gallon, J. R., and Chaplin, A. E. (1981) Acetylene reduction (nitrogen fixation) by cyanobacteria grown under alternating light-dark cycles. *FEMS Microbiol. Lett.* 10, 245–247
155. Gallon, J. R., LaRue, T. A., and Kurz, W. G. W. (1974) Photosynthesis and nitrogenase activity in the blue-green alga Gloeocapsa. *Can. J. Microbiol.* 20, 1633–1637
156. Ishiura, M. (1998) Expression of a Gene Cluster kaiABC as a Circadian Feedback Process in Cyanobacteria. *Science (80-.).* 281, 1519–1523
157. Loza-Correa, M., Gomez-Valero, L., and Buchrieser, C. (2010) Circadian clock proteins in prokaryotes: hidden rhythms? *Front. Microbiol.* 1, 130
158. Snijder, J., Burnley, R. J., Wiegard, A., Melquiond, A. S. J., Bonvin, A. M. J. J., Axmann, I. M., and Heck, A. J. R. (2014) Insight into cyanobacterial circadian timing from structural details of the KaiB-KaiC interaction. *Proc. Natl. Acad. Sci. U. S. A.* 111, 1379–84
159. Hayashi, F., Suzuki, H., Iwase, R., Uzumaki, T., Miyake, A., Shen, J.-R., Imada, K., Furukawa, Y., Yonekura, K., Namba, K., and Ishiura, M. (2003) ATP-induced hexameric ring structure of the cyanobacterial circadian clock protein KaiC. *Genes to Cells* 8, 287–296
160. Mori, T., Saveliev, S. V, Xu, Y., Stafford, W. F., Cox, M. M., Inman, R. B., and Johnson, C. H. (2002) Circadian clock protein KaiC forms ATP-dependent hexameric rings and binds DNA. *Proc. Natl. Acad. Sci. U. S. A.* 99, 17203–8

161. Pattanayek, R., Wang, J., Mori, T., Xu, Y., Johnson, C. H., and Egli, M. (2004) Visualizing a circadian clock protein: crystal structure of KaiC and functional insights. *Mol. Cell* 15, 375–88
162. Nakajima, M., Imai, K., Ito, H., Nishiwaki, T., Murayama, Y., Iwasaki, H., Oyama, T., and Kondo, T. (2005) Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro. *Science* 308, 414–5
163. Axmann, I. M., Hertel, S., Wiegard, A., Dörrich, A. K., and Wilde, A. (2014) Diversity of KaiC-based timing systems in marine Cyanobacteria. *Mar. Genomics* 14, 3–16
164. Kitayama, Y., Nishiwaki, T., Terauchi, K., and Kondo, T. (2008) Dual KaiC-based oscillations constitute the circadian system of cyanobacteria. *Genes Dev.* 22, 1513–1521
165. Qin, X., Byrne, M., Xu, Y., Mori, T., and Johnson, C. H. (2010) Coupling of a core post-translational pacemaker to a slave transcription/translation feedback loop in a circadian system. *PLoS Biol.* 8, e1000394
166. Chabot, J. R., Pedraza, J. M., Luitel, P., and van Oudenaarden, A. (2007) Stochastic gene expression out-of-steady-state in the cyanobacterial circadian clock. *Nature* 450, 1249–1252
167. Smith, R. M., and Williams, S. B. (2006) Circadian rhythms in gene transcription imparted by chromosome compaction in the cyanobacterium Synechococcus elongatus. *Proc. Natl. Acad. Sci. U. S. A.* 103, 8564–8569
168. Woelfle, M. A., Xu, Y., Qin, X., and Johnson, C. H. (2007) Circadian rhythms of superhelical status of DNA in cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* 104, 18819–24
169. Iwasaki, H., Williams, S. B., Kitayama, Y., Ishiura, M., Golden, S. S., and Kondo, T. (2000) A KaiC-Interacting Sensory Histidine Kinase, SasA, Necessary to Sustain Robust Circadian Oscillation in Cyanobacteria. *Cell* 101, 223–233
170. Takai, N., Nakajima, M., Oyama, T., Kito, R., Sugita, C., Sugita, M., Kondo, T., and Iwasaki, H. (2006) A KaiC-associating SasA-RpaA two-component regulatory system as a major circadian timing mediator in cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* 103, 12109–14
171. Hanaoka, M., Takai, N., Hosokawa, N., Fujiwara, M., Akimoto, Y., Kobori, N., Iwasaki, H., Kondo, T., and Tanaka, K. (2012) RpaB, another response regulator operating circadian clock-dependent transcriptional regulation in Synechococcus elongatus PCC 7942. *J. Biol. Chem.* 287, 26321–7
172. Taniguchi, Y., Takai, N., Katayama, M., Kondo, T., and Oyama, T. (2010) Three major output pathways from the KaiABC-based oscillator cooperate to generate robust circadian kaiBC expression in cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* 107, 3263–8
173. Gutu, A., and O'Shea, E. K. (2013) Two antagonistic clock-regulated histidine kinases time the activation of circadian gene expression. *Mol. Cell* 50, 288–94
174. Taniguchi, Y., Katayama, M., Ito, R., Takai, N., Kondo, T., and Oyama, T. (2007) labA : a novel gene required for negative feedback regulation of the cyanobacterial circadian clock protein KaiC. 60–70

175. Ivleva, N. B., Gao, T., LiWang, A. C., and Golden, S. S. (2006) Quinone sensing by the circadian input kinase of the cyanobacterial circadian clock. *Proc. Natl. Acad. Sci. U. S. A.* 103, 17468–73
176. Wood, T. L., Bridwell-Rabb, J., Kim, Y.-I., Gao, T., Chang, Y.-G., LiWang, A., Barondeau, D. P., and Golden, S. S. (2010) The KaiA protein of the cyanobacterial circadian oscillator is modulated by a redox-active cofactor. *Proc. Natl. Acad. Sci. U. S. A.* 107, 5804–9
177. Mackey, S. R., Choi, J.-S., Kitayama, Y., Iwasaki, H., Dong, G., and Golden, S. S. (2008) Proteins found in a CikA interaction assay link the circadian clock, metabolism, and cell division in Synechococcus elongatus. *J. Bacteriol.* 190, 3738–46
178. Ivleva, N. B., Bramlett, M. R., Lindahl, P. a, and Golden, S. S. (2005) LdpA: a component of the circadian clock senses redox state of the cell. *EMBO J.* 24, 1202–10
179. Kutsuna, S., Kondo, T., Ikegami, H., Uzumaki, T., Katayama, M., and Ishiura, M. (2007) The Circadian Clock-Related Gene pex Regulates a Negative cis Element in the kaiA Promoter Region. *J. Bacteriol.* 189, 7690–7696
180. Rust, M. J., Golden, S. S., and O'Shea, E. K. (2011) Light-driven changes in energy metabolism directly entrain the cyanobacterial circadian oscillator. *Science* 331, 220–3
181. Kim, Y.-I., Vinyard, D. J., Ananyev, G. M., Dismukes, G. C., and Golden, S. S. (2012) Oxidized quinones signal onset of darkness directly to the cyanobacterial circadian oscillator. *Proc. Natl. Acad. Sci. U. S. A.* 109, 17765–9
182. Dong, G., Yang, Q., Wang, Q., Kim, Y.-I., Wood, T. L., Osteryoung, K. W., Oudenaarden, A. van, and Golden, S. S. (2010) Elevated ATPase Activity of KaiC Applies a Circadian Checkpoint on Cell Division in Synechococcus elongatus. *Cell* 140, 529–539
183. Mori, T., Williams, D. R., Byrne, M. O., Qin, X., Egli, M., Mchaourab, H. S., Stewart, P. L., and Johnson, C. H. (2007) Elucidating the ticking of an in vitro circadian clockwork. *PLoS Biol.* 5, 841–853
184. Yoshida, T., Murayama, Y., Ito, H., Kageyama, H., and Kondo, T. (2009) Nonparametric entrainment of the in vitro circadian phosphorylation rhythm of cyanobacterial KaiC by temperature cycle. *Proc. Natl. Acad. Sci. U. S. A.* 106, 1648–1653
185. Burja, A. M., Dhamwichukorn, S., and Wright, P. C. (2003) Cyanobacterial postgenomic research and systems biology. *Trends Biotechnol.* 21, 504–11
186. Ito, H., Mutsuda, M., Murayama, Y., Tomita, J., Hosokawa, N., Terauchi, K., Sugita, C., Sugita, M., Kondo, T., and Iwasaki, H. (2009) Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in Synechococcus elongatus. *Proc. Natl. Acad. Sci. U. S. A.* 106, 14168–73
187. Vijayan, V., Zuzow, R., and O'Shea, E. K. (2009) Oscillations in supercoiling drive circadian gene expression in cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* 106, 22564–8

188. Liu, Y., Tsinoremas, N. F., Johnson, C. H., Lebedeva, N. V, Golden, S. S., Ishiura, M., and Kondo, T. (1995) Circadian orchestration of gene expression in cyanobacteria. *Genes Dev.* 9, 1469–1478
189. Yen, U.-C., Huang, T.-C., and Yen, T.-C. (2004) Observation of the circadian photosynthetic rhythm in cyanobacteria with a dissolved-oxygen meter. *Plant Sci.* 166, 949–952
190. Simon, W. J., Hall, J. J., Suzuki, I., Murata, N., and Slabas, A. R. (2002) Proteomic study of the soluble proteins from the unicellular cyanobacterium Synechocystis sp. PCC6803 using automated matrix-assisted laser desorption/ionization-time of flight peptide mass fingerprinting. *Proteomics* 2, 1735–1742
191. Sazuka, T., Yamaguchi, M., and Ohara, O. (1999) in *Electrophoresis*, pp 2160–2171.
192. Anderson, D. C., Campbell, E. L., and Meeks, J. C. (2006) A soluble 3D LC/MS/MS proteome of the filamentous cyanobacterium Nostoc punctiforme. *J. Proteome Res.* 5, 3096–3104
193. Ow, S. Y., Nolrel, J., Cardona, T., Taton, A., Lindblad, P., Stensjö, K., and Wright, P. C. (2009) Quantitative overview of N 2 fixation in nostoc Punctiforme ATCC 29133 through cellular enrichments and iTRAQ shotgun proteomics. *J. Proteome Res.* 8, 187–198
194. Wegener, K. M., Singh, A. K., Jacobs, J. M., Elvitigala, T., Welsh, E. a, Keren, N., Gritsenko, M. a, Ghosh, B. K., Camp, D. G., Smith, R. D., and Pakrasi, H. B. (2010) Global proteomics reveal an atypical strategy for carbon/nitrogen assimilation by a cyanobacterium under diverse environmental perturbations. *Mol. Cell. Proteomics* 9, 2678–89
195. Barrios-Llerena, M. E., Chong, P. K., Gan, C. S., Snijders, A. P. L., Reardon, K. F., and Wright, P. C. (2006) Shotgun proteomics of cyanobacteria--applications of experimental and data-mining techniques. *Brief. Funct. Genomic. Proteomic.* 5, 121–32
196. Stöckel, J., Jacobs, J. M., Elvitigala, T. R., Liberton, M., Welsh, E. a, Polpitiya, A. D., Gritsenko, M. a, Nicora, C. D., Koppenaal, D. W., Smith, R. D., and Pakrasi, H. B. (2011) Diurnal rhythms result in significant changes in the cellular protein complement in the cyanobacterium Cyanothece 51142. *PLoS One* 6, e16680
197. Waldbauer, J. R., Rodrigue, S., Coleman, M. L., and Chisholm, S. W. (2012) Transcriptome and proteome dynamics of a light-dark synchronized bacterial cell cycle. *PLoS One* 7, e43432
198. Aryal, U. K., Stöckel, J., Krovvidi, R. K., Gritsenko, M. a, Monroe, M. E., Moore, R. J., Koppenaal, D. W., Smith, R. D., Pakrasi, H. B., and Jacobs, J. M. (2011) Dynamic proteomic profiling of a unicellular cyanobacterium Cyanothece ATCC51142 across light-dark diurnal cycles. *BMC Syst. Biol.* 5, 194
199. Aryal, U. K., Stöckel, J., Welsh, E. a, Gritsenko, M. a, Nicora, C. D., Koppenaal, D. W., Smith, R. D., Pakrasi, H. B., and Jacobs, J. M. (2012) Dynamic proteome analysis of Cyanothece sp. ATCC 51142 under constant light. *J. Proteome Res.* 11, 609–19

Chapter 2

Daily rhythms in the cyanobacterium *Synechococcus elongatus* probed by high-resolution mass spectrometry based proteomics reveals a small-defined set of cyclic proteins

Ana C.L. Guerreiro[1,2], Marco Benevento[1,2], Robert Lehmann[3], Bas van Breukelen[1,2], Harm Post[1,2], Piero Giansanti[1,2], A.F. Maarten Altelaar[1,2] , Ilka M. Axmann[3,4] and Albert J.R. Heck[1,2]

Published in Molecular and Cellular proteomics, 2014

[1] Biomolecular Mass Spectrometry and Proteomics Group, Utrecht Institute for Pharmaceutical Sciences and Bijvoet Centre for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands
[2] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands
[3]Institute for Theoretical Biology (ITB), Humboldt-Universitaet zu Berlin, Invalidenstrasse 43, D-10115 Berlin, Germany
[4]Institute for Synthetic Microbiology, Heinrich-Heine-Universitaet Duesseldorf, Universitaetsstrasse 1, D-40225 Duesseldorf, Germany

Summary

Circadian rhythms are self-sustained and adjustable cycles, typically entrained with light/dark and/or temperature cycles. These rhythms are present in animals, plants, fungi and several bacteria. The central mechanism behind these 'pacemakers' and the connection to the circadian regulated pathways are still poorly understood. The circadian rhythm of the cyanobacterium *Synechococcus elongatus* PCC 7942 (*S. elongatus*) is highly robust and controlled by only three proteins, named KaiA, KaiB and KaiC. This central clock system has been extensively studied functionally, structurally and can be reconstituted *in vitro*. These characteristics together with the relatively small genome (2.7 Mbp) of *S. elongatus*, make it an ideal model system for the study of circadian rhythms.

Different approaches have already been used to reveal the influence of the central *S. elongatus* clock on rhythmic gene expression, rhythmic mRNA abundances, rhythmic DNA topology changes and cell division. However, a global analysis of its proteome dynamics has not been reported yet.

To uncover the variation in protein abundances during 48 hours under light and dark cycles (12:12-hours), we used quantitative proteomics, with TMT 6-plex isobaric labeling. We queried the *S. elongatus* proteome at ten different time points spanning a single 24 hour period, leading to 20 time points over the full 48 hour period.

Employing multi-dimensional separation and high-resolution mass spectrometry, we were able to find evidence for a total of 82% of the *S. elongatus* proteome. Among the 1537 proteins quantified over the time-course of the experiment, only 77 of them underwent significant cyclic variations. Interestingly, our data provides evidence for in- and out-of-phase correlation between mRNA and protein levels for a set of specific genes and proteins. Since a range of cyclic proteins are functionally not well annotated, this work provides a resource for further studies to explore the role of these proteins in the cyanobacterial circadian rhythm. The data is available via ProteomeXchange with identifier PXD000510.

Introduction

Circadian clocks or rhythms are widely observed in many species and throughout the different kingdoms, from plants and animals to fungi and bacteria. Circadian rhythms are defined as self-sustained and adjustable

cycles, typically entrained by light/dark and/or temperature cycles, occurring within 24 hour periods (1).

Circadian rhythms were originally associated with more complex organisms, based on the notion that simple and fast dividing organisms, such as bacteria, would not need a robust cyclic mechanism, governing cellular processes (2, 3). More recently, cyclic rhythms have been shown to exist also in evolutionary very old systems like cyanobacteria, which possess a very robust circadian machinery (pacemaker), regulating gene expression, metabolism and even cell cycle (4, 5).

The cyanobacterium *S. elongatus* in particular has become a very useful model for studying circadian rhythms. Besides having a small genome (2.7 Mbp) (6, 7), this cyanobacterium also features a very simple and extremely robust core clock mechanism controlling its circadian rhythm. The core circadian system consists of just three proteins (KaiA, KaiB, KaiC). KaiC undergoes a phosphorylation/dephosphorylation cycle, where KaiA is the promoter of its hyperphosphorylated state and KaiB opposes KaiA's action, promoting KaiC's hypophosphorylated state. Next to this phosphorylation, the dynamic co-assembly of larger KaiABC complexes plays a controlling role in this circadian system (8). This relatively simple circadian pacemaker provides such a robust system that can even be reproduced *in vitro*, only requiring the three Kai proteins and the addition of ATP/Mg^{2+} (9).

The connection of the central circadian mechanism (or the core posttranslational oscillator (PTO)) to its output is less well defined. However, it has been shown to encompass a transcription/translation feedback loop (TTFL), where the KaiABC genes and their products participate in positive and negative autoregulatory feedback loops (10, 11), similarly to what is observed in higher organisms (12). Different studies connect the KaiABC central clock with control over transcriptional activity through different effectors such as the histidine kinase SasA (*synechococcus* adaptative sensor A), the putative transcription factor RpaA (regulator of phycobilisome-associated) (13, 14), the RpaB regulator (15), LabA (low amplitude and bright A) or the sensor histidine kinase CikA (circadian input kinase A) (16, 17) and the input factors Pex (period extender - PadR family transcriptional regulator) and LdpA (light-dependent period A). Others have proposed an oscilloid model, in which gene expression is influenced by control over DNA topology (18, 19). These observations link the central clock to the global gene expression, including the expression of the *kaiABC* genes themselves (10).

Several bioluminescence studies have indicated a cyclic promoter activity of 100% (20), while microarray studies report a 30-60% rate of cyclic mRNA abundances in *S. elongatus* (21, 22). Despite the fact that these experiments were performed under different conditions and experimental setups, this discrepancy may be also caused by post-transcriptional regulation, and therefore protein abundance might likewise differ from observed mRNA abundance patterns.

Besides extensive analysis of gene expression and mRNA levels in cyanobacteria, a limited number of proteome analyses have been performed. To our knowledge, a global proteome analysis of *S. elongatus* has not been published so far, although proteome analyses have been reported for some related species such as *Synechocystis* (23), *Cyanothece* (24–26), *Prochlorococcus* (27), and *Anabaena* (28).

To enable the global analysis of the *S. elongatus* proteome dynamics, we used high-resolution quantitative MS-based proteomics. To uncover the variation in protein abundances during 48 hours, under light and dark cycles (LD) (12:12-hours), we employed quantitative shot-gun proteomics, using TMT 6-plex isobaric labeling. These isobaric tags provide a sensitive labeling method for the analysis of several different experimental conditions (up to 6 at a time). We queried the *S. elongatus* proteome at ten different time points spanning a single 24 hours period, resulting in 20 time points over the full 48 hours period.

With this approach we were able to detect the abundances of proteins covering 82% of the *S. elongatus* genome, within which we observed significant abundance changes of 544 proteins. Among these proteins, 77 proteins showed well-defined cyclic abundance profiles. The comparison of our results to previously published mRNA abundance profiles yields a significantly lower degree of cyclic expression, pointing to the importance of post-transcriptional and/or post-translational regulatory mechanisms. Moreover, our data provides novel insights into the phasing of cyclic protein abundances relative to corresponding mRNA levels, as we observe that several protein abundance levels cycle, albeit out-of-phase, with their corresponding cycling mRNA levels.

Experimental Procedures

1. Cyanobacteria cell culture

The wild-type strain of *Synechococcus elongatus* PCC 7942 was routinely grown photoautotrophically in BG11-medium (29) at 30°C under continuous illumination with white light of 80 µmol of photons/m^2s (Versatile Environmental Test Chamber, SANYO) and a continuous stream of air. Cell concentrations were measured by determining the optical densities of the culture at 750 nm (OD750) (SPECORD®200 PLUS, Analytik Jena). The culture was kept in log growth phase (up to an OD750 of 1.0) by dilution up to a specific volume and transferred to 12:12-hours light/dark cycle (LD) for 3 days. Synchronized culture was finally diluted to an OD750 of approximately 0.4 one day before the sampling started. At certain time points (Fig. 1A), which varied from one to three hour intervals, 40 ml of the culture was centrifuged at 15,000 x g for 10 min and the supernatant was removed. Cell pellet was resuspended in 1 ml BG11-medium, centrifuged again for 5 min. Supernatant was removed and pellet washed with 1 ml PBS buffer before last round of centrifugation. Cell pellets were frozen in liquid nitrogen prior to storage at -20°C.

2. Sample preparation

Cyanobacteria pellets from samples of 20 time-points were lysed in 8 M Urea, 50 mM triethyl ammonium bicarbonate (TEAB), containing 1 tablet EDTA-free protease inhibitor cocktail (Sigma) and 1 tablet PhosSTOP phosphatase inhibitor cocktail (Roche). After three sonication cycles at 4 ^{0}C, total cell lysates were obtained through centrifugation at 14,000 rpm for 30 min, at 4 ^{0}C. The supernatant was recovered and protein concentration was determined with the Bradford method (BioRad). The proteins were subjected to reduction and alkylation of the cysteine residues, using 200 mM dithiotreitol (DTT) (Sigma) and 200 mM iodoacetamide (Sigma). The proteins were first digested with Lys-C (Roche Diagnostics, Ingelheim, Germany), at an enzyme:protein ratio of 1:75, for 4 hours at 37 ^{0}C, followed by 4 times dilution of the samples with 50 mM TEAB and digestion with trypsin (Roche Diagnostics, Ingelheim, Germany), at an enzyme:protein ratio of 1:100, overnight at 37 ^{0}C. 40 µg from each sample and from a mixture of all samples were desalted using 1cc Sep Pack C18 columns (Waters) and dried *in vacuo*.

3. Peptide labeling

Peptides were labeled with tandem mass tags (TMT) using the TMT 6plex labeling kit (Pierce). Two separate experiments, each one of them consisting of two TMT 6plex labeling experiments were performed in total. Each experiment consisted of the use of 5 tags, one for each time-point, and the 6th tag for the mixture of all time-points (internal standard). The manufacturer's protocol was followed, with a few adjustments. After desalting, 40 µg of peptides per channel were dissolved in 100 µl 200 mM TEAB. The TMT labeling reagents were dissolved in 40 µL acetonitrile (ACN) (Biosolve) per vial and added to the samples in two steps, to maximize the labeling efficiency. First, 10 µl of reagent solution was added to the sample, after 5 min, the other 10 µl were added and the reaction was incubated for 1 hour at room temperature. In the quenching step, 4 µL of 5% hydroxylamine were added. After 15 min, the six channels were mixed in a 1:1 ratio and stored at −20 °C.

4. Strong Cation eXchange (SCX) fractionation

Peptides were fractionated using strong-cation exchange (SCX) as described previously (30). In short, four SCXs were performed using a ZorbaxBioSCX-Series II column (0.8 mm×50 mm, 3.5 µm). Solvent A consisted of 0.05% formic acid in 20% ACN, while solvent B consisted of 0.05% formic acid, 0.5 M NaCl in 20% ACN. The following gradient was used: 0–0.01 min (0–2% B); 0.01–8.01 min (2–3% B); 8.01–14.01 min (3–8% B); 14.01–28 min (8–20% B); 28–38 min (20–40% B); 38–48 min (40–90%B); 48–54 min (90% B); 54–60 min (0% B). The fractions were collected from minute 50 until minute 90. The resulting 50 fractions were dried *in vacuo*, resuspended in 10% formic acid and stored at −20 °C.
For the SWATH (Sequential Windowed Acquisition of all Theoretical spectra) approach, unlabeled samples from a 24h time-series were pooled together and fractionated by SCX, to create a spectral library.

5. Liquid Chromatography-tandem Mass Spectrometry (LC-MS/MS)

Peptide fractions were analyzed on an Orbitrap Velos (Thermo Fisher Scientific, Bremen) that was coupled to an Agilent 1200 HPLC system (Agilent Technologies, Waldbronn). Both the trap (20 mm×100 µm I.D.) and the analytical column (35 cm×50 µm I.D.) were packed in-house using Reprosil-pur C18 3 µm (Dr. Maisch). Peptides were trapped at 5 µL/min in 100%

solvent A (0.1 M acetic acid (Merck)). Elution was achieved with the solvent B (0.1 M acetic acid in 80% acetonitrile) at 100 nL/min. The 180 minutes gradient used was as follows: 0-10 min, 100% solvent A; 10.1-115 min, 10–24% solvent B; 115-160 min, 24–50% solvent B; 160-163 min, 50-100% solvent B; 163-164 min, 100% solvent B; 164-165 min, 0-100% solvent A; 165-180 min, 100% solvent A. Nanospray was achieved using a coated fused silica emitter (New Objective, Cambridge, MA) (o.d. 360 µm; i.d. 20 µm, tip i.d. 10 µm) biased to 1.7 kV. The mass spectrometer was operated in the data dependent acquisition mode. A MS2 method was used with a FT survey scan from 350 to 1500 m/z (resolution 30,000; AGC target 5E5). The 10 most intense peaks were subjected to HCD fragmentation (resolution 7500; AGC target 3E4, NCE 45%, max. injection time 500 ms, dynamic exclusion 60 s). Predictive AGC was enabled.

For the SWATH approach, samples were analyzed by a TripleTOF 5600 fitted with a Nanospray III source (AB SCIEX, Concord, ON) coupled to an Agilent 1290 Infinity UHPLC (Ultra-High Pressure Liquid Chromatography) system (Agilent Technologies, Waldbronn, DE). Briefly, for the spectral library data, the mass spectrometer was operated in data-dependent (DDA) mode to obtain MS/MS spectra for the 20 most abundant parent ions following each survey MS1 scan. Additional data sets were recorded as triplicates in data-independent mode using SWATH MS2 acquisitions, essentially as described in Gillet *et al*. (31). In summary, a window of 26 m/z (containing 1 m/z for the window overlap) is passed in 32 incremental steps over the full mass range 350–1250 m/z.

6. Database search and validation

Raw data was converted to .mgf file type with Proteome Discoverer (version 1.3, Thermo). Mascot (version 2.3.02, Matrix science) was used to search the MS/MS data against the *S. elongatus* PCC 7942 Uniprot database (version 4-2010) including a list of common contaminants and concatenated with the reversed versions of all sequences (5,826 sequences). Trypsin was chosen as cleavage specificity allowing two missed cleavages. Carbamidomethylation (C) was set as a fixed modification. The variable modifications used were oxidation (M), TMT6plex (K) and TMT6plex (N-term). The database searches were performed using a peptide tolerance of 50 ppm and a fragment mass tolerance of 0.05 Da (HCD). The 50-ppm mass window was chosen to allow random assignment of false positives that were later removed by filtering using the instruments actual mass accuracy (10 ppm). A

.dat file of each of the four experiments was exported from Mascot and filtered with Rockerbox (32) to an FDR of 1% using the concatenated database decoy method. Quantification was performed with the [R] package Isobar (33). A minimum peptide score of 20 and a minimum protein score of 40 were used as identifications thresholds. The peak intensities obtained were corrected for isotope impurity of the TMT labels and normalized with the median peak intensity. For all experiments, a minimum of one unique peptide was considered for protein quantification.

Raw data obtained with the TripleTOF 5600 was searched with ProteinPilot™, using the Paragon™ search engine (version 4.3, AB SCIEX). MS/MS data was searched against the same *S. elongatus* PCC 7942 Uniprot database. Trypsin was chosen as cleavage specificity. For the modifications, Cys alkylation was set to iodoacetamide and ID focus was set to biological modifications (*i.e.* phosphorylations, amidations, semitryptic fragments, etc.), as special factors, urea denaturation was selected. The database search was performed with a thorough effort, with a detected protein threshold (unused protscore (confidence)) set to achieve 99% confidence. The FDR analysis option was selected.

Datasets from SWATH MS2 acquisitions were processed using the full scan MS/MS filtering module for data-independent acquisition within Skyline 1.3 (34). The .group file obtained from ProteinPilot was converted to .xml using the group2xml.exe script (version 4.3.0.1456, AB SCIEX), to create the spectral library (.blib). The top 6 peptides and fragment ions were extracted from SWATH MS2 acquisitions within Skyline using a fragment ion resolution setting of 15,000. Peak areas were normalized with total area sums (TAS) method.

7. Protein copy numbers calculations

The sum of the number of peptide-spectrum matches (PSMs) obtained in all the experiments was normalized by the molecular weight of each corresponding protein and then divided by the total abundance factor calculated for all the identified proteins. This relative abundance factor was multiplied by the total amount of protein material used in all the experiments and divided by the protein molecular weights, leading to the protein copy number values. To determine the copy numbers of each protein in each cyanobacterium cell, we then divided this value by the number of cells used in the experiment.

8. Significance analysis and clustering

Quantitative data, containing protein intensities from 20 time-points, as obtained from the experiments described before, was analyzed as follows. The table containing protein names and respective intensities in all time points was loaded into [R] (version 3.0.0) (35) as a datamatrix (eset) object. Subsequently all rows, containing proteins with their intensities over the 20 time-points were filtered and processed in [R] using the following criteria: i) No missing data (intensities) in any of the time-points per protein was allowed. Proteins containing missing data were removed from the matrix; ii) All data in the matrix was scaled (Z transformed); iii) All proteins that have an Inter Quartile Range (IQR) variation less than 1 (e.g. those proteins that show no real change in any of the time-points) were removed; iv) The data was exported to a tab-delimited file for subsequent analysis. The filtered protein data was loaded into the Multi Experiment Viewer software (MeV v. 4.8.1) (36, 37). Based on the figure of merit, a method to determine the optimal number of k-means clusters, we have chosen to perform k-means clustering with 6 clusters. All clusters with their containing proteins were exported and used for further analysis.

9. Cyclic profile validation

The proteins considered significant and observed across all different time-points were subjected to a Pearson correlation analysis between the first 24 hour ratios and second 24 hour ratios. For this purpose the statistical software IBM® SPSS® Statistics Data Editor was used. For a protein profile to be considered cyclic, a significant correlation with a p-value<0.05 was required. Consequently, a heatmap was produced by hierarchical clustering of the protein abundance profiles, using Pearson correlation and complete linkage. The data was Z-score transformed for better visualization. A visual inspection of the profiles was performed to confirm their cyclic nature. The same procedure was applied for the analysis of the mRNA profiles from Ito et al. (21). The “raw” data (GSE14225) was extracted from GEO (Gene Expression Omnibus) with [R] (35) and an average of the two replicates was performed.

Results

1. Global proteome analysis of the cyanobacterium *S. elongatus*

Here we present the first in-depth quantitative proteome analysis of the cyanobacterium *S. elongatus*, over a 48 hours timespan under light/dark (LD) conditions. This was accomplished by sampling 20 independent time points with 1 to 3 hour intervals. The sampling scheme consists of one hour intervals at the LD and DL transitions, whereas two and three hour intervals cover the respective day and night phases (Fig. 1A). The resulting 20 samples were quantified using the isobaric tandem mass tag (TMT) labeling strategy and four 6-plex quantitation experiments with one pooled sample as internal control. After cell lysis, protein digestion and TMT labeling, the peptides were fractionated by strong cation exchange (SCX), this was done in order to decrease the complexity of the proteome, and at the same time reduce precursor ion interference upon MS/MS analysis, which is inherent to isobaric quantification. This yielded an average of 20 fractions per experiment and approximately 80 LC-MSMS runs (Fig. 1A). The subsequent search analysis was performed using Mascot in combination with Isobar (33). Over 45,092 unique peptides (used for quantitation) and 2,179 proteins were identified, thus covering 82% of the predicted *S. elongatus* proteome (2,657 proteins). Out of the identified proteins, 82% were represented by two or more peptides. The presented work comprises one of the most complete coverage of a proteome reported to date. The concentrations of each detected protein were evaluated using the well-established method based on spectral counts (38–40), as described in the experimental procedures. In this experiment, we covered a dynamic range in protein abundance of over 5 orders of magnitude (Fig. 1B). The most abundant proteins found are mainly involved in photosynthesis, a crucial part of the energetic metabolism. However, proteins involved in this pathway are also spread throughout the whole dynamic range (Fig. 1B and Suppl. Fig. 1A). Notwithstanding the high proteome coverage we obtain for *S. elongatus* we observed that 768 (~35%!) of the identified proteins are still annotated in the database as functionally uncharacterized. These proteins are distributed over the entire protein abundance range, even in the top 50 of abundant proteins. These not-annotated proteins might be implicated in all kind of processes, both in specialized functions as well as ‘housekeeping’ (Suppl. Fig. 1B).

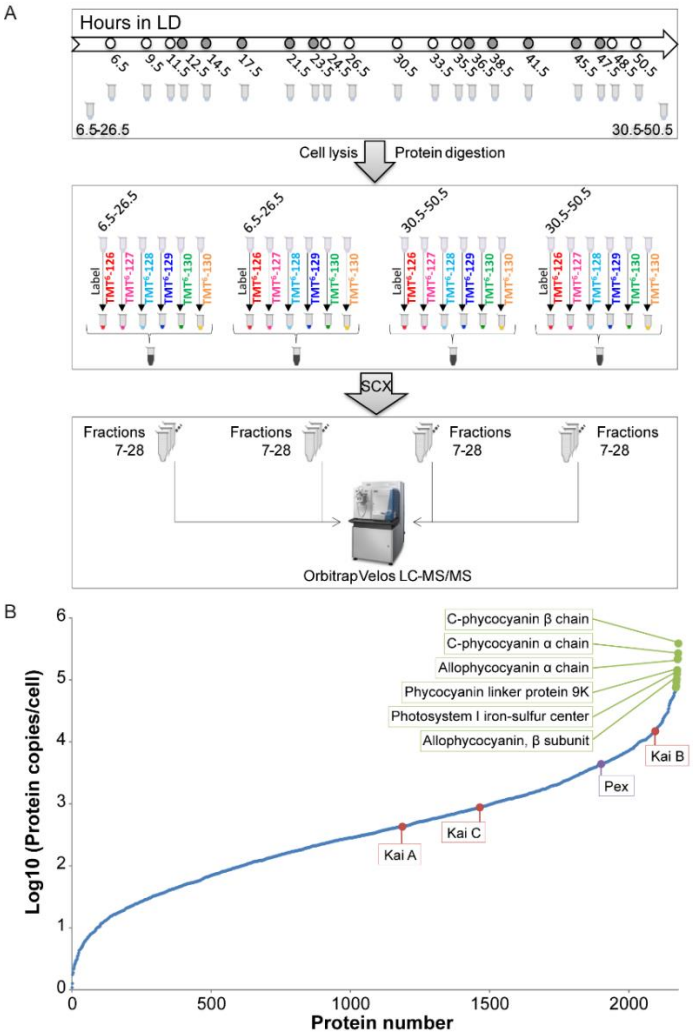


Figure 1 – High-resolution proteomics analysis of the S. elongatus proteome. **A)** *S. elongatus* cells were synchronized and cultured in light/dark (12:12-hours) conditions. Samples were collected across a 48-hour period in 20 independent time points with 1 to 3 hour intervals. Peptides were labeled with TMT 6-plex isobaric tags, were the 6th channel was used to label a mixture of all the time-points. Before LC-MSMS analysis with an Orbitrap Velos, the four resulting sample mixtures were fractionated by SCX. **B)** Logarithmic protein copy numbers of all the 2,179 proteins identified, as calculated using spectral counts (described in the Experimental Procedures). Their abundances cover a dynamic range of 5 orders of magnitude. The most abundant proteins belong to the photosynthetic pathway (highlighted in green) and some proteins relevant to the circadian rhythm mechanism have medium to high abundances (central clock mechanism proteins - KaiA, KaiB and KaiC - are highlighted in red and the transcriptional regulator, PadR family (Pex) is highlighted in purple).

2. Abundance profiles of the Kai proteins and their putative input and output channels

We first investigated the abundance profiles of some well-studied proteins in *S. elongatus*, and compared these profiles to previous published work (Fig. 2). The main engine driving the circadian clock in *S. elongatus* is composed of the three proteins KaiA, KaiB and KaiC. In accordance with a previous study (41), KaiA did not display a cyclic abundance profile (Fig.2A). The profiles of KaiB and KaiC also do not show obvious cyclic abundances and moreover seem to show opposite abundance profiles in our data (Fig. 2A). Cyclic abundances for KaiB and KaiC have been reported under continuous light conditions (LL) (41) and synchronization in their total cellular abundances. However, in contrast, Qin et al.(10) reported that KaiC is not cyclic under LD conditions, which is in close agreement with our results.

Next we assessed proteins, which are linked to the circadian rhythm regulation or propagation and for which abundance profiles have been published. Since current studies on cyanobacterial circadian rhythms consider transcript abundance levels rather than protein abundance, the number of available datasets is limited. One protein that fulfills these criteria is pex, a PadR family transcriptional regulator which is essential for circadian rhythm in *S. elongatus*. Pex is responsible for the elongation and delay of the circadian rhythm in *S. elongatus* by the negative regulation of KaiA expression (42, 43). Takai et al. (42) demonstrated that the Pex protein exhibits cyclic abundance in LD conditions with its maximum during the subjective night. This western-blot based analysis agrees with the presented proteomics dataset (Fig. 2B). Another component of the circadian clock output pathways is the KaiC expression inhibitor LabA. While LabA was found to exhibit cyclic mRNA abundance in LL conditions (21), its protein abundance is also found to oscillate (not shown). However, most of the other reported components (e.g. RpaA, RpaB, SasA, CikA, LdpA) were not found to have cyclic abundances at the protein level as revealed by our analysis.

In order to provide further confidence in the quantitative TMT data, additionally an independent SWATH analysis was performed on a small set of selected proteins, namely 6 cyclic, 3 non-significant and non-cyclic and 3 significant but not-cyclic proteins (Suppl. Fig. 4), according to our TMT data analysis. Comparing the protein expression profiles obtained by SWATH with the latter, we observe a very good agreement.

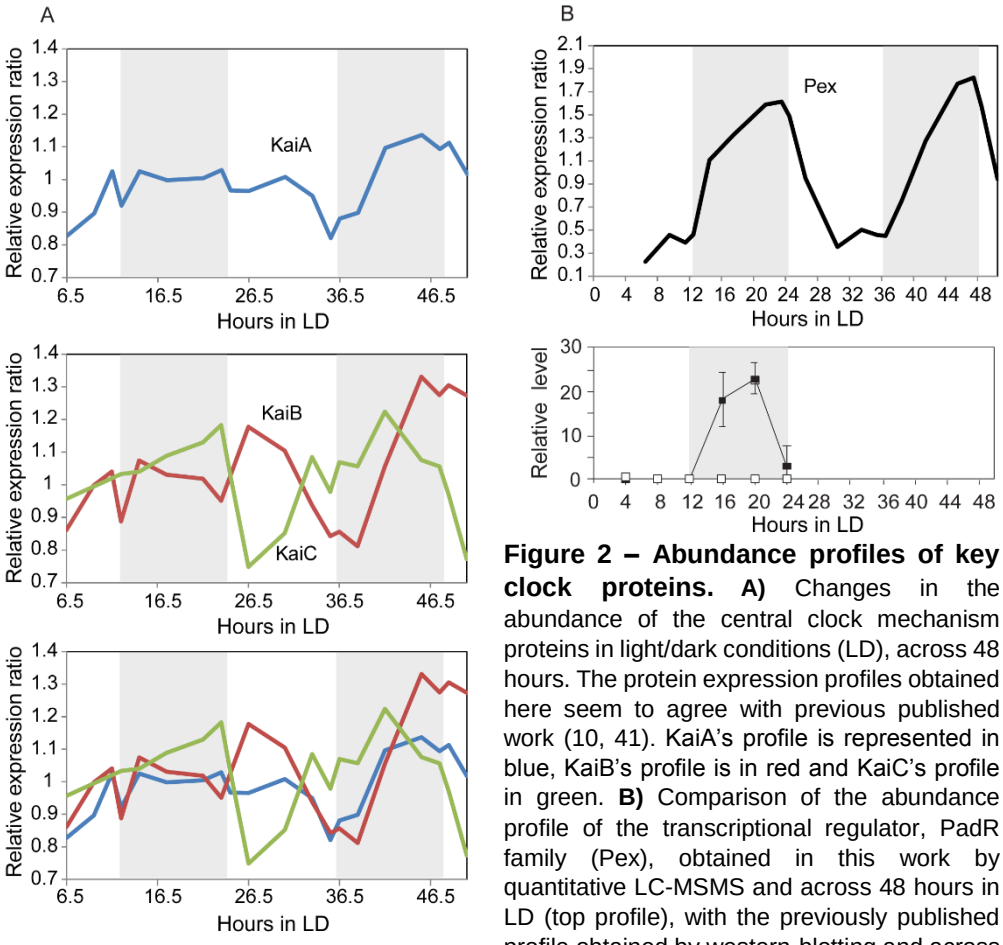


Figure 2 – Abundance profiles of key clock proteins. A) Changes in the abundance of the central clock mechanism proteins in light/dark conditions (LD), across 48 hours. The protein expression profiles obtained here seem to agree with previous published work (10, 41). KaiA's profile is represented in blue, KaiB's profile is in red and KaiC's profile in green. **B)** Comparison of the abundance profile of the transcriptional regulator, PadR family (Pex), obtained in this work by quantitative LC-MSMS and across 48 hours in LD (top profile), with the previously published profile obtained by western-blotting and across 24 hours in LD (bottom profile, adapted from (42)). Both protein abundance profiles show an increase in abundance during the dark period. Grey areas represent the dark period and the white areas represent the light period.

3. Significance analysis

Next, we conducted a global analysis of the abundance behavior of all proteins quantified in our analysis. Through this approach we sought to find profiles that could be linked to circadian rhythm. We filtered our dataset, taking first only proteins that had quantitative information throughout the whole 48 hours experiment (1,537 proteins). Next we used inter quartile range analysis to find significant variations in the 48-hour profiles. Ratios were normalized and

transformed to a z-score, followed by removal of the proteins with an inter quartile range (IQR) variation less than 1. From this evaluation we found 544 proteins to be significantly regulated in at least one out of 20 time points. Included in this set were the known clock pathway proteins KaiB, Pex, CikA, RpaA and LabA.

In our data analysis, we also observed ratio compression, which can be associated with the precursor ion interferences inherent to isobaric quantification. Nevertheless, this underestimation of the fold change is of no influence on our analysis as here we determine whether the ratio is significantly different from the background, which is similarly compressed. The resulting significant protein abundance profiles were clustered with K-means to facilitate the global analysis and the search for novel cyclic abundance profiles. This was done by the use of the MultiExperiment Viewer software (MeV). The 544 protein abundance profiles were divided into 6 clusters, as optimized with the figure of merit (FOM) (Fig.3A). Clusters 1 and 2 show anti-correlated profiles and more pronounced differential abundance over time than the remaining clusters. Clusters 4 and 5 also exhibit anti-correlated cyclic profiles, but with less pronounced changes in abundance over time. Cluster 3 and 6 show proteins with specific trends, which have a noticeable relative abundance. However, they do not show cyclic profiles nor show correlation with the other clusters or with each other.

4. Functional analysis of proteins co-occurring within clusters

To investigate whether any of the clusters was associated with a specific pathway, protein interaction network, localization or function we used the Database for annotation, visualization and integrated discovery (DAVID) to look for functional enrichments (44, 45). Although this analysis revealed no significant enrichment of specific functions in the clusters, we made some interesting observations using the Kyoto encyclopedia of genes and genomes (KEGG) for pathway analysis. For cluster 1 we found 6 proteins to be associated with the two-component system and penicillin and cephalosporin biosynthesis, associated with signal transduction and the synthesis of secondary metabolites, respectively. An interesting example of the first pathway is the protein RpaA which is known to be a transcription factor involved in genome wide circadian gene expression and controlled by the kinase SasA, which in turn is controlled by KaiC, indicating a possible link with circadian control. Cluster 2 has a lower number of associated proteins but most of them display a quite prominent cyclic profile. Since no specific

functional association can be made with cluster 2 one could envision that the variety of pathways controlled by the circadian clock is reflected by the variety of proteins present in this cluster. Indeed we find proteins involved in transcription, biotin metabolism and response to stress. Cluster 3 contains proteins that are part of the amino acid, carbohydrate and nucleotide metabolisms, as well as DNA replication and repair and protein export. In cluster 4 the photosynthetic proteins appear most prominent, in cluster 5 ribosomal proteins and in Cluster 6 proteins that are related to gene expression and translation are present. These results show that functionally related proteins do not necessarily associate with specific trends of protein abundance, however, some trends may be associated with pathway components or protein complexes.

5. Protein profiles analysis in selected networks

Next, we sought to analyze selected pathways and/or protein complexes in more detail with the protein profiles [R] package. This package looks for the significance between similar abundance profiles, independently of their intensities (46). We started with the photosynthetic pathway and its complexes, also well characterized in cyanobacteria. Here we could observe groups of proteins with very similar abundance profiles. One example is the phycobilisome complex, of which the components capture energy from light and transfer it to the photosystem II through chlorophyll A. These proteins have significantly similar profiles (p-value<<0.001) (Fig. 3B and Table 1), up regulated in the light period, although not completely cyclic. Within this group of proteins we find an even more pronounced profile, namely the phycobilisome core-membrane linker polypeptide and the phycobilisome rod-core linker polypeptide showing an increase in abundance during the dark period (p-value=0.01) (Fig. 3B and Table 1). The two components serve as stabilizers at the core of the phycobilisome complex in collaboration with other proteins, therefore they are not directly involved in the light capturing process.

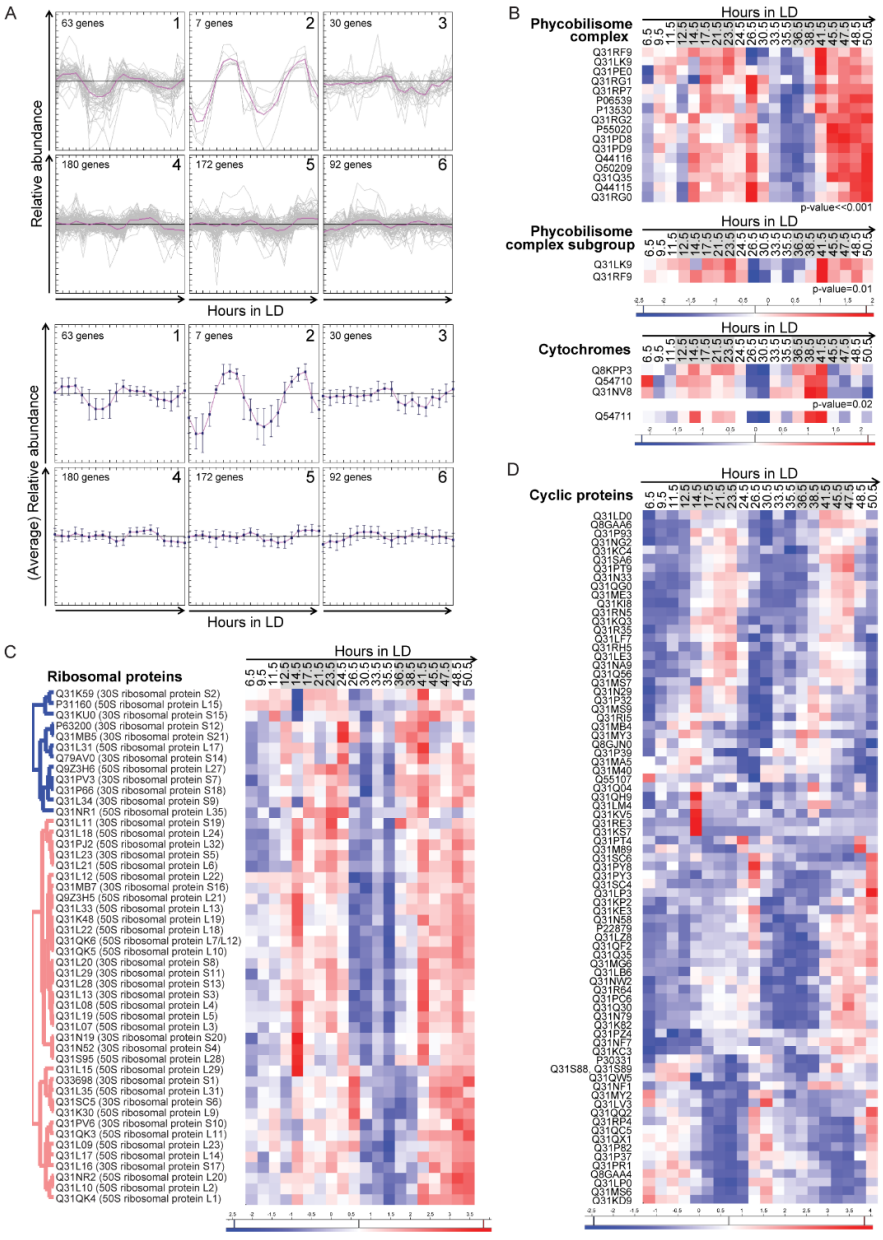


Figure 3 – Clustering analysis on the significantly changing protein abundance profiles. A) K-means clustering analysis on the 544 protein abundance profiles found to be significantly changing across the 48 hour LD experiment, as described in the Experimental Procedures section. At the top all the different protein profiles are visible in each corresponding cluster (the pink line represents the average profile of each cluster); at the bottom, only the averaged abundance profiles and their variance bars are visible for each cluster. **B)** Hierarchical

clustered heatmaps of three sets of proteins from the photosynthetic pathway. Each interest group of proteins has significantly similar abundance changes (p-value<0.05) across 48 hours, as determined by the Protein Profiles [R] package analysis. Red represents high abundance and blue low abundance of the proteins. All proteins are represented by their corresponding Uniprot accession number. **C)** Hierarchical clustered heatmap of ribosomal proteins. The type of analysis and representation is the same as the one mentioned in B). **D)** Hierarchical clustered heatmap of proteins with cyclic abundance profiles. These protein profiles were first analyzed with respect to their significance, as mentioned above, and then a Pearson correlation was performed, comparing the first 24 hours of the profiles with the second 24 hours, as explained in the Experimental Procedures section. The representation and clustering procedure is as the one mentioned above. The grey areas represent the dark period. The color scale is shown below each heatmap.

Another example is the cytochrome complex b6/f, which is responsible for the transfer of electrons from photosystem II to photosystem I. Although we were not able to cover all of its components, we could observe very similar protein abundance profiles, specifically between the apocytochrome f and the cytochrome b6-f complex subunit 4. These two combined with the cytochrome b559, which is part of the photosystem II, also show a similar trend (p-value=0.02) (Fig. 3B and Table 1). All three of them have higher abundance during the dark period. Another b6/f component, the cytochrome b6, exhibits a similar behavior despite not having a significantly similar profile to the ones above (Fig. 3B and Table 1).

Next, we investigated different proteins from the photosystems I and II and the ATP synthase complex. For the photosystem I we found 4 proteins with significantly similar profiles (p-value= 0.02). These show an increase of abundance during the dark period. In photosystem II we found 7 proteins with a significantly similar nocturnal trend (p-value<<0.001), and 2 proteins with similarly significant trends, with higher abundance during both the dark and light phases (p-value=0.004). For the ATP synthase complex, we found that most of the components we identified have similar profiles (p-value=0.03). However, during the light period they differ slightly. Finally, for the ribosomal proteins all the proteins from the small subunit showed significantly similar profiles (p-value=0.01), as did all proteins from the large subunit (p-value=10^{-4}) (Fig. 3C). Some components of the two subunits also shared significant profiles (p-value<<0.001). Altogether, these observations show that many of the observed protein complexes or particular components of these complexes share abundance trends hinting at the specific regulation of different subsets of proteins.

Table 1 – Similarity in abundance profiles in related proteins. Three sets of proteins from the photosynthetic pathway, with significantly similar abundance changes (p-value<0.05) across 48h, as determined by the Protein Profiles [R] package analysis.

Accession	Protein name
Phycobilisome complex (p-value<<0.001)	
Q31RF9	Phycobilisome core-membrane linker polypeptide
Q31LK9	Phycobilisome rod-core linker polypeptide
Q31PE0	Phycobilisome rod linker polypeptide
Q31RG1	Allophycocyanin, beta subunit
Q31RP7	Putative uncharacterized protein
P06539	C-phycocyanin beta chain
P13530	C-phycocyanin alpha chain
Q31RG2	Phycobilisome 7.8 kDa linker polypeptide, allophycocyanin-associated core
P55020	Plastocyanin
Q31PD8	Phycocyanin linker protein 9K
Q31PD9	Phycobilisome rod linker polypeptide
Q44116	Phycobilisome maturation protein
O50209	Allophycocyanin, beta subunit
Q31Q35	Allophycocyanin alpha chain-like
Q44115	Phycocyanobilin lyase subunit alpha
Q31RG0	Allophycocyanin alpha chain
Phycobilisome complex subgroup (p-value=0.01)	
Q31LK9	Phycobilisome rod-core linker polypeptide
Q31RF9	Phycobilisome core-membrane linker polypeptide
Cytochromes (p-value=0.02)	
Q8KPP3	Cytochrome b559 subunit alpha
Q54710	Cytochrome b6-f complex subunit 4
Q31NV8	Apocytochrome f
Q54711	Cytochrome b6

6. Focusing on proteins displaying cyclic abundances

The cluster analysis gave an indication of at least 2 clusters with possible cyclic trends and potential circadian associated protein abundances (Fig. 3A). Cluster 2 shows little variability between each trend, while in cluster 1, both amplitude and variability are higher. To make a more precise assessment of the cyclic proteins, we performed a Pearson correlation and decided on the cyclic properties of the profiles based on the significant linear correlation between the data points from the first day and the second, corresponding to a p-value lower than 0.05, using the IBM SPSS Statistics software.

The Pearson correlation analysis and a visual confirmation of the profiles resulted in the stringent definition of 77 well-defined cyclic abundance profiles. This group included several functionally uncharacterized/not well annotated proteins. To facilitate the visualization of these profiles we created a heatmap using [R] (Fig. 3D and Table 2). This heatmap reveals 39 proteins featuring an abundance peak during the night. In contrast, only 29 proteins exhibit a maximum abundance during the light phase. We also found proteins, which have certain ambiguities concerning the peak phase as well as the oscillation period (24 hours or 12 hours). Despite this, we considered those cases to correspond to cyclic proteins. The group of 29 proteins with higher abundance during the light period are reported to be involved in a broad variety of processes including secretion, cell wall/membrane biogenesis, general stress, transcription, translation, protein turnover and folding, signal transduction, circadian rhythm, amino acid metabolism, photosynthesis, with a small predominance of the cell wall/membrane biogenesis. The group of proteins with their abundance peaking in the dark phase contains a variety of proteins involved in cell/ wall biogenesis, chaperoning, signal transduction, transcription, translation, defense mechanisms, DNA repair, different metabolisms such as coenzyme and vitamin, carbohydrate, amino acid, lipid and energy production related, including photosynthesis (light capture related). Notably, the majority of the proteins with higher abundance during the night are involved in transcription.

From these results one might suppose that the cyanobacterium has increased metabolism and maintenance during the night cycle as compared to the day. Also, it is interesting to see that not all photosynthetic proteins show cyclic profiles in our analysis and those that do, even demonstrate a possible diminished photosynthetic capacity during the day. The proteins ambiguous on their phase are part of secondary metabolite biosynthesis, transport and catabolism, inorganic ion transport, signal transduction and transcription.
Interestingly, most of the protein abundance profiles do not have a sinusoidal peak shape. In fact, most of the profiles with maxima during the light period seem to have a more steep increase in abundance followed by a slow decrease in abundance. Moreover, profiles with maxima during the dark period display exactly the opposite behavior. Consistently, the DL transition is

Table 2 – Proteins with cyclic abundance profiles. These proteins profiles were first analyzed with respect to their significance, with IQR (interquartile range) method and then a Pearson correlation was performed, comparing the first 24h of the profiles with the second 24 hours.

Accession	Protein Name
Q31LD0	ATPase
Q8GAA6	Queuine tRNA-ribosyltransferase
Q31P93	Phosphomethylpyrimidine synthase
Q31NG2	Metal dependent phosphohydrolase
Q31KC4	Putative uncharacterized protein
Q31SA6	Aminotransferase
Q31PT9	Putative uncharacterized protein
Q31N33	Putative uncharacterized protein
Q31QG0	Transcriptional regulator, PadR family
Q31ME3	RNA polymerase sigma factor rpoD2
Q31KI8	Heat shock protein Hsp20
Q31RN5	Putative uncharacterized protein
Q31KQ3	Putative uncharacterized protein
Q31R35	Putative uncharacterized protein
Q31LF7	Elongation factor EF-G
Q31RH5	UPF0102 protein Synpcc7942_0312
Q31LE3	Diguanylate cyclase with GAF sensor
Q31NA9	Putative uncharacterized protein
Q31Q56	Phosphoenolpyruvate synthase
Q31MS7	Pyridine nucleotide transhydrogenase alpha subunit
Q31N29	Group3 RNA polymerase sigma factor SigF
Q31P32	Putative uncharacterized protein
Q31MS9	NAD(P) transhydrogenase subunit beta
Q31RI5	Phospholipase D/Transphosphatidylase
Q31MB4	Phosphate starvation-induced protein
Q31MY3	Putative uncharacterized protein
Q8GJN0	Light-independent protochlorophyllide reductase subunit B
Q31P39	Putative uncharacterized protein
Q31MA5	Group3 RNA polymerase sigma factor SigF
Q31M40	RNA polymerase sigma factor
Q55107	Bicarbonate transport ATP-binding protein CmpC

Table 2 – Proteins with cyclic abundance profiles (continuation).

Accession	Protein Name
Q31Q04	Putative uncharacterized protein
Q31QH9	Putative uncharacterized protein
Q31LM4	Putative uncharacterized protein
Q31KV5	2-hydroxy-6-oxohepta-24-dienoate hydrolase
Q31RE3	Probable peptidase
Q31KS7	UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
Q31PT4	Putative uncharacterized protein
Q31M89	Putative uncharacterized protein
Q31SC6	Putative uncharacterized protein
Q31PY8	Putative uncharacterized protein
Q31PY3	Response regulator receiver domain protein
Q31SC4	Putative uncharacterized protein
Q31LP3	Putative uncharacterized protein
Q31KP2	Putative uncharacterized protein
Q31KE3	Methionine aminopeptidase
Q31N58	Imidazole glycerol phosphate synthase subunit HisH
P22879	60 kDa chaperonin
Q31LZ8	LabA
Q31QF2	60 kDa chaperonin 1
Q31Q35	Allophycocyanin alpha chain-like
Q31MG6	Carotene isomerase
Q31LB6	Anthranilate phosphoribosyltransferase
Q31NW2	DNA repair protein RadC
Q31R64	Putative uncharacterized protein
Q31PC6	Putative uncharacterized protein
Q31Q30	Putative uncharacterized protein
Q31N79	Putative uncharacterized protein
Q31K82	Putative uncharacterized protein
Q31PZ4	Phosphoesterase PHP-like
Q31NF7	Diguanylate cyclase/phosphodiesterase
Q31KC3	Two component transcriptional regulator, winged helix family

Table 2 – Proteins with cyclic abundance profiles (continuation).

Accession	Protein Name
P30331	Metallothionein
Q31S88, Q31S89	Pilin polypeptide PilA-like
Q31QW5	Putative uncharacterized protein
Q31NF1	Carbonate dehydratase
Q31MY2	RNA polymerase sigma factor
Q31LV3	Putative uncharacterized protein
Q31QQ2	PDZ/DHR/GLGF
Q31RP4	Possible high light inducible polypeptide HliC
Q31QC5	C-terminal processing peptidase-2. Serine peptidase. MEROPS family S41A
Q31QX1	C-terminal processing peptidase-2. Serine peptidase. MEROPS family S41A
Q31P82	TPR repeat
Q31P37	Rare lipoprotein A
Q31PR1	Outer envelope membrane protein
Q8GAA4	Putative uncharacterized protein sek0026
Q31LP0	RNA-binding region RNP-1
Q31MS6	Putative uncharacterized protein
Q31KD9	General secretion pathway protein D

marked by a steeper response than the LD transition, even though sampling and light/dark transition were kept the same.

Among the proteins with abundance profiles over the complete period, we also found 4 proteins with abundance cycles shorter than 24 hours. In fact it has been shown by Westermark and Herzel (47) that components of the circadian clock can generate 12 hour rhythms in gene expression of animals. They demonstrated that 12 hour genes have alternating peak heights, which is consistent with our observation at the protein level of pilin polypeptide PilA-like (signal transduction), carbonate dehydratase (nitrogen metabolism), 2-hydroxy-6-oxohepta-24-dienoate hydrolase (no specific function) and a putative uncharacterized protein. Recent reports describe rhythms shorter than 24 hours called the ultradian rhythms, which has also been reported to occur in related cyanobacteria such as *Cyanothece* (48, 49) and *Prochlorococcus* (50).

From all proteins with significant changing abundances, we classified 14% to experience cyclic abundances. The percentage of proteins showing cyclic abundances is low, when compared to data reported on the transcript abundance. Transcript data for *S. elongatus* revealed that 30-60% of the mRNA levels exhibited cyclic profiles (21, 51). These somewhat conflicting results could indicate the existence of post-transcriptional and/or post-translational mechanisms regulating the protein abundances, as further discussed in the following section. However, it must be noted that differences in experimental conditions as well as the experimental and theoretical methods can influence the observed mRNA and protein abundances. The difference in methodology and the stringency of our data analysis might therefore account for some of the described differences.

7. In- and out-of-phase correlation between mRNA and protein levels of cyclic genes/proteins

To compare circadian mRNA abundance profiles with the cyclic protein profiles we used the published results from Ito et al. (21), on the quantitative profiles of *S. elongatus* transcripts during 48 hours under continuous light conditions. We took this dataset from Ito et al. as we observed that the majority of mRNA transcripts measured by them could also be quantified in our study at the protein level and the taken time-points matched best with our experimental set-up (Fig 4A). Of the identified transcripts, 800 were reported to be cyclic. However, when we perform a Pearson correlation analysis on the reported transcript profiles, we find 1057 transcripts to be cyclic. In a direct comparison of these cyclic mRNA versus our 77 cyclic protein abundances, we find an overlap of 37 genes with the reported data by Ito et al. and an overlap of 46 genes with the data analyzed by Pearson correlation, uncovering an extra 13 cyclic genes in common with our cyclic proteins (Fig 4B and Suppl. Fig. 2). Moreover, an additional 27 proteins found cyclic in our study did not show cyclic behavior at the mRNA level. These results indicate a large discrepancy between the abundance at the transcript and protein level and clearly indicate post-transcriptional regulation.

Next, we compared the cyclic protein profiles that overlapped with the cyclic mRNA data (46 proteins) in more detail. Here, we observed examples of proteins involved in several biological functions with a small prevalence for DNA repair, transcription and metabolism related proteins. The 31 cases where only the protein abundances are cyclic, equally showed diverse protein functions with a small prevalence for signal transduction, transcription and

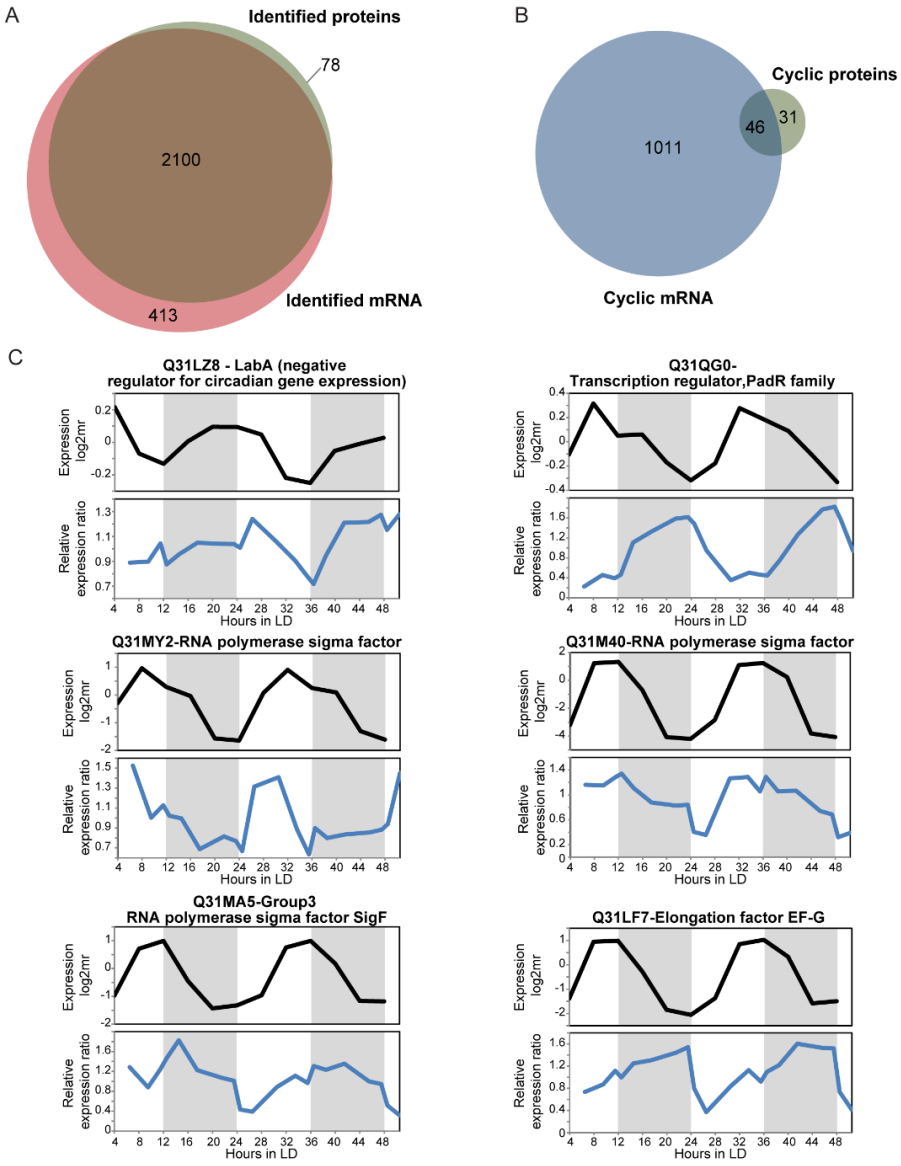


Figure 4 – Comparison of the proteome data with the mRNA data. A) Comparison of the number of proteins identified in this work, by quantitative LC-MSMS, with the number of transcripts identified in Ito et al. work (21), by microarray. **B)** Comparison of the number of proteins with cyclic abundance profiles with the number of transcripts with cyclic abundance profiles, identified by Ito et al., but reanalyzed using Pearson correlation analysis, as described in the Experimental Procedures section. **C)** Examples on the comparison of the abundance profiles of the proteins (blue lines) and transcripts (black lines) considered cyclic, showing in-phase or out-of-phase correlations between protein and transcript data. The dark period is represented by the grey area.

folding. Visual analysis of the peak phase of the mRNA profiles showed some ambiguity, similar to that observed for the cyclic analysis at the protein level. We observed two transcripts with 12-hour cycles and abundance peaks within the subjective switch between day and night.
Interestingly, we noticed that the delays of the peak times between mRNA and proteins sometimes are not the same within the 48 hours. We performed a comparison of the peak times between our cyclic abundance profiles and the cyclic transcript profiles from Ito et al., which revealed delays between zero to 22 hours, hinting at in phase and out of phase correlations between these two levels (Fig 4C and Suppl. Fig. 3). For 8 genes out of 46, there was a clear in phase correlation between protein and mRNA profiles, and for 33 genes there was a clear out of phase correlation, 15 being completely anti-phase. Among the genes with delays shorter than 12 hours we mainly have proteins implicated in transcription, DNA repair and metabolism. Genes with a longer delay are involved in cell wall biogenesis, transcription, general stress response and coenzyme metabolism. Proteins related to transcription are observed with differing time delays from the transcript data (e.g. transcription factor PadR family and RNA polymerase *rpoD5* versus RNA polymerase sigma factors *rpoD5*/SigC and SigF2), indicating alternative means to post transcriptional control of protein expression.
Our observations show a complex regulation of protein expression, which for several proteins occurs at the post-transcriptional level. For many proteins the observed cyclic behavior of mRNA abundance does not translate into the same behavior at the protein level. Moreover, for those gene products that do display cyclic behavior at both mRNA and protein level, a large portion do not display similar abundance profiles but instead show different levels of delay. These different delay times of actual translation of the mRNA into protein might be attributed to the necessity of the bacteria be able to quickly express certain proteins upon external signals.

Discussion

Here we report one of the most complete proteomes to date, identifying evidence for 82% of the predicted genes at the protein level. Although the genome of the cyanobacterium *S. elongatus* is not as extensive as mammalian systems, this percentage is still impressive, also when compared to the 60-70% reported for other smaller genomes such as yeast (52).
We monitored the relative protein abundance of hundreds of proteins over two light/dark (LD) cycles with a total of 20 samples. We focused our analysis on

the proteins showing a cycling abundance profile. 77 proteins were found to exhibit cycling abundance with a 24-hour period and quite distinct behavior. Firstly, the majority of cycling proteins were more abundant in the night, which is unexpected for photosynthetic bacteria. We also observed many proteins within a single protein complex or pathway revealing similar abundance profiles, albeit not always cycling. Interestingly, a range of proteins was found to exhibit 12-hour oscillations in abundance (e.g. Carbonate dehydratase, Pilin polypeptide PilA-like, 2-hydroxy-6-oxohepta-24-dienoate hydrolase). Most remarkable was the observation of several cycling proteins, whose cycle was clearly out of phase with the cycling behavior observed for the corresponding transcript. This data represents some of the clearest evidence for post-transcriptional regulation of many of the processes involved in the circadian rhythm and the clear need to analyze these processes at the protein level.

Although a few of the proteins we observe to be cyclic have been linked to the circadian rhythm in *S. elongatus,* most of these proteins have not yet been linked and originate from a variety of biological processes. Therefore, we believe our dataset provides several new starting-points for further investigation of proteins and pathways involved or regulated by the circadian rhythm in cyanobacteria.

Supporting information

Supporting Figures and Supporting tables can be found online at http://www.mcponline.org/content/13/8/2042/suppl/DC1.

Acknowledgments

We would like to acknowledge Anne Rediger, for her contribution in the cell culturing and sample collection, and Stefanie Hertel, for the helpful comments during manuscript preparation. This work was supported by The Netherlands Proteomics Centre, and the Netherlands Organization for Scientific Research (NWO) with the VIDI grant for AFMA (723.012.102). Additionally, the PRIME-XS Project, Grant Agreement 262067, funded by the European Union Seventh Framework Program, is kindly acknowledged for financial support. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (53) with the dataset identifier PXD000510.

References

1. Johnson, C. H., Stewart, P. L., and Egli, M. (2011) The cyanobacterial circadian system: from biophysics to bioevolution. *Annu. Rev. Biophys.* 40, 143–67
2. Johnson, C. H., Golden, S. S., Ishiura, M., and Kondo, T. (1996) MicroReview Circadian clocks in prokaryotes. *Mol. Microbiol.* 21, 5–11
3. Pittendrigh, C. S. (1993) Temporal organization: reflections of a Darwinian clock-watcher. *Annu. Rev. Physiol.* 55, 16–54
4. Dong, G., Yang, Q., Wang, Q., Kim, Y.-I., Wood, T. L., Osteryoung, K. W., Oudenaarden, A. van, and Golden, S. S. (2010) Elevated ATPase Activity of KaiC Applies a Circadian Checkpoint on Cell Division in Synechococcus elongatus. *Cell* 140, 529–539
5. Kondo, T., Strayer, C. a, Kulkarni, R. D., Taylor, W., Ishiura, M., Golden, S. S., and Johnson, C. H. (1993) Circadian rhythms in prokaryotes: luciferase as a reporter of circadian gene expression in cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* 90, 5672–6
6. Andersson, C. R., Tsinoremas, N. F., Shelton, J., Lebedeva, N. V., Yarrow, J., Min, H., and Golden, S. S. (2000) ScienceDirect.com - Methods in Enzymology - Application of bioluminescence to the study of circadian rhythms in cyanobacteria. *Methods Enzymol.* 305, 527–542
7. Holtman, C. K., Chen, Y., Sandoval, P., Gonzales, A., Nalty, M. S., Thomas, T. L., Youderian, P., and Golden, S. S. (2005) High-throughput functional analysis of the Synechococcus elongatus PCC 7942 genome. *DNA Res.* 12, 103–15
8. Brettschneider, C., Rose, R. J., Hertel, S., Axmann, I. M., Heck, A. J. R., and Kollmann, M. (2010) A sequestration feedback determines dynamics and temperature entrainment of the KaiABC circadian clock. *Mol. Syst. Biol.* 6, 389
9. Nakajima, M., Imai, K., Ito, H., Nishiwaki, T., Murayama, Y., Iwasaki, H., Oyama, T., and Kondo, T. (2005) Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro. *Science* 308, 414–5
10. Qin, X., Byrne, M., Xu, Y., Mori, T., and Johnson, C. H. (2010) Coupling of a core post-translational pacemaker to a slave transcription/translation feedback loop in a circadian system. *PLoS Biol.* 8, e1000394
11. Zwicker, D., Lubensky, D. K., and ten Wolde, P. R. (2010) Robust circadian clocks from coupled protein-modification and transcription-translation cycles. *Proc. Natl. Acad. Sci. U. S. A.* 107, 22540–5
12. Hardin, P. E., Hall, J. C., and Rosbash, M. (1990) Feedback of the Drosophila period gene product on circadian cycling of its messenger RNA levels. *Nature* 343, 536–40
13. Takai, N., Nakajima, M., Oyama, T., Kito, R., Sugita, C., Sugita, M., Kondo, T., and Iwasaki, H. (2006) A KaiC-associating SasA-RpaA two-component regulatory system as a major circadian timing mediator in cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* 103, 12109–14

14. Iwasaki, H., Williams, S. B., Kitayama, Y., Ishiura, M., Golden, S. S., and Kondo, T. (2000) A KaiC-Interacting Sensory Histidine Kinase, SasA, Necessary to Sustain Robust Circadian Oscillation in Cyanobacteria. *Cell* 101, 223–233
15. Hanaoka, M., Takai, N., Hosokawa, N., Fujiwara, M., Akimoto, Y., Kobori, N., Iwasaki, H., Kondo, T., and Tanaka, K. (2012) RpaB, another response regulator operating circadian clock-dependent transcriptional regulation in Synechococcus elongatus PCC 7942. *J. Biol. Chem.* 287, 26321–7
16. Taniguchi, Y., Takai, N., Katayama, M., Kondo, T., and Oyama, T. (2010) Three major output pathways from the KaiABC-based oscillator cooperate to generate robust circadian kaiBC expression in cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* 107, 3263–8
17. Schmitz, O., Katayama, M., Williams, S. B., Kondo, T., and Golden, S. S. (2000) CikA, a Bacteriophytochrome That Resets the Cyanobacterial Circadian Clock. *Sci.* 289 , 765–768
18. Mori, T., and Johnson, C. H. (2001) Circadian programming in cyanobacteria. *Semin. Cell Dev. Biol.* 12, 271–278
19. Woelfle, M. A., Xu, Y., Qin, X., and Johnson, C. H. (2007) Circadian rhythms of superhelical status of DNA in cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* 104, 18819–24
20. Liu, Y., Tsinoremas, N. F., Johnson, C. H., Lebedeva, N. V, Golden, S. S., Ishiura, M., and Kondo, T. (1995) Circadian orchestration of gene expression in cyanobacteria. *Genes Dev.* 9, 1469–1478
21. Ito, H., Mutsuda, M., Murayama, Y., Tomita, J., Hosokawa, N., Terauchi, K., Sugita, C., Sugita, M., Kondo, T., and Iwasaki, H. (2009) Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in Synechococcus elongatus. *Proc. Natl. Acad. Sci. U. S. A.* 106, 14168–73
22. Vijayan, V., Jain, I. H., and O'Shea, E. K. (2011) A high resolution map of a cyanobacterial transcriptome. *Genome Biol.* 12, R47
23. Wegener, K. M., Singh, A. K., Jacobs, J. M., Elvitigala, T., Welsh, E. a, Keren, N., Gritsenko, M. a, Ghosh, B. K., Camp, D. G., Smith, R. D., and Pakrasi, H. B. (2010) Global proteomics reveal an atypical strategy for carbon/nitrogen assimilation by a cyanobacterium under diverse environmental perturbations. *Mol. Cell. Proteomics* 9, 2678–89
24. Aryal, U. K., Stöckel, J., Welsh, E. a, Gritsenko, M. a, Nicora, C. D., Koppenaal, D. W., Smith, R. D., Pakrasi, H. B., and Jacobs, J. M. (2012) Dynamic proteome analysis of Cyanothece sp. ATCC 51142 under constant light. *J. Proteome Res.* 11, 609–19
25. Aryal, U. K., Stöckel, J., Krovvidi, R. K., Gritsenko, M. a, Monroe, M. E., Moore, R. J., Koppenaal, D. W., Smith, R. D., Pakrasi, H. B., and Jacobs, J. M. (2011) Dynamic proteomic profiling of a unicellular cyanobacterium Cyanothece ATCC51142 across light-dark diurnal cycles. *BMC Syst. Biol.* 5, 194
26. Stöckel, J., Jacobs, J. M., Elvitigala, T. R., Liberton, M., Welsh, E. a, Polpitiya, A. D., Gritsenko, M. a, Nicora, C. D., Koppenaal, D. W., Smith, R. D., and Pakrasi, H. B. (2011) Diurnal rhythms result in significant changes in the

cellular protein complement in the cyanobacterium Cyanothece 51142. *PLoS One* 6, e16680
27. Waldbauer, J. R., Rodrigue, S., Coleman, M. L., and Chisholm, S. W. (2012) Transcriptome and proteome dynamics of a light-dark synchronized bacterial cell cycle. *PLoS One* 7, e43432
28. Barrios-Llerena, M. E., Chong, P. K., Gan, C. S., Snijders, A. P. L., Reardon, K. F., and Wright, P. C. (2006) Shotgun proteomics of cyanobacteria--applications of experimental and data-mining techniques. *Brief. Funct. Genomic. Proteomic.* 5, 121–32
29. Rippka, R., Deruelles, J., Waterbury, J., Herdman, M., and Stanier, R. (1979) Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J. Gen. Microbiol.* 110, 1–61
30. De Graaf, E. L., Altelaar, a F. M., van Breukelen, B., Mohammed, S., and Heck, A. J. R. (2011) Improving SRM assay development: a global comparison between triple quadrupole, ion trap, and higher energy CID peptide fragmentation spectra. *J. Proteome Res.* 10, 4334–41
31. Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012) Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* 11, O111.016717
32. Van den Toorn, H. W. P., Muñoz, J., Mohammed, S., Raijmakers, R., Heck, A. J. R., and van Breukelen, B. (2010) RockerBox: Analysis and Filtering of Massive Proteomics Search Results. *J. Proteome Res.* 10, 1420–1424
33. Breitwieser, F. P., Müller, A., Dayon, L., Köcher, T., Hainard, A., Pichler, P., Schmidt-Erfurth, U., Superti-Furga, G., Sanchez, J.-C., Mechtler, K., Bennett, K. L., and Colinge, J. (2011) General statistical modeling of data from protein relative expression isobaric tags. *J. Proteome Res.* 10, 2758–66
34. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26, 966–8
35. R Development Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Au
36. Saeed, A. I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V., and Quackenbush, J. (2003) TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34, 374–8
37. Saeed, A., Bhagabati, N., Braisted, J., Liang, W., Sharov, V., Howe, E., Li, J., Thiagarajan, M., White, J., and Quackenbush, J. (2006) TM4 microarray software suite. *Methods Enzymol.* 411, 134–93
38. Peng, M., Taouatas, N., Cappadona, S., van Breukelen, B., Mohammed, S., Scholten, A., and Heck, A. J. R. (2012) Protease bias in absolute protein quantitation. *Nat Meth* 9, 524–525
39. Ringrose, J. H., van den Toorn, H. W. P., Eitel, M., Post, H., Neerincx, P., Schierwater, B., Maarten Altelaar, A. F., and Heck, A. J. R. (2013) Deep

proteome profiling of Trichoplax adhaerens reveals remarkable features at the origin of metazoan multicellularity. *Nat Commun* 4, 1408
40. Malmström, J., Beck, M., Schmidt, A., Lange, V., Deutsch, E. W., and Aebersold, R. (2009) Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. *Nature* 460, 762–5
41. Kitayama, Y., Iwasaki, H., Nishiwaki, T., and Kondo, T. (2003) KaiB functions as an attenuator of KaiC phosphorylation in the cyanobacterial circadian clock system. *EMBO J.* 22, 2127–34
42. Takai, N., Ikeuchi, S., Manabe, K., and Kutsuna, S. (2006) Expression of the circadian clock-related gene pex in cyanobacteria increases in darkness and is required to delay the clock. *J. Biol. Rhythms* 21, 235–44
43. Kutsuna, S., Kondo, T., Ikegami, H., Uzumaki, T., Katayama, M., and Ishiura, M. (2007) The Circadian Clock-Related Gene pex Regulates a Negative cis Element in the kaiA Promoter Region. *J. Bacteriol.* 189, 7690–7696
44. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2008) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57
45. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13
46. Hansson, J., Rafiee, M. R., Reiland, S., Polo, J. M., Gehring, J., Okawa, S., Huber, W., Hochedlinger, K., and Krijgsveld, J. (2012) Highly coordinated proteome dynamics during reprogramming of somatic cells to pluripotency. *Cell Rep.* 2, 1579–92
47. Westermark, P., and Herzel, H. (2013) Mechanism for 12 Hr Rhythm Generation by the Circadian Clock. *Cell Rep.* 3, 1228–1238
48. Cerveny, J., Sinetova, M. a, Valledor, L., Sherman, L. a, and Nedbal, L. (2013) Ultradian metabolic rhythm in the diazotrophic cyanobacterium Cyanothece sp. ATCC 51142. *Proc. Natl. Acad. Sci. U. S. A.* 110, 13210–5
49. Elvitigala, T., Stöckel, J., Ghosh, B. K., and Pakrasi, H. B. (2009) Effect of continuous light on diurnal rhythms in Cyanothece sp. ATCC 51142. *BMC Genomics* 10, 226
50. Shalapyonok, A., Olson, R. J., and Shalapyonok, L. S. (1998) Ultradian Growth inProchlorococcus spp. *Appl. Environ. Microbiol.* 64, 1066–1069
51. Vijayan, V., Zuzow, R., and O'Shea, E. K. (2009) Oscillations in supercoiling drive circadian gene expression in cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* 106, 22564–8
52. Gunaratne, J., Schmidt, A., Quandt, A., Neo, S. P., Saraç, O. S., Gracia, T., Loguercio, S., Ahrné, E., Xia, R. L. H., Tan, K. H., Lössner, C., Bähler, J., Beyer, A., Blackstock, W., and Aebersold, R. (2013) Extensive mass spectrometry-based analysis of the fission yeast proteome: the Schizosaccharomyces pombe PeptideAtlas. *Mol. Cell. Proteomics* 12, 1741–51
53. Vizcaino, J., Vizcaíno, J. A., Côté, R. G., Csordas, A., Dianes, J. a, Fabregat, A., Foster, J. M., Griss, J., Alpi, E., Birim, M., Contell, J., O'Kelly, G., Schoenegger, A., Ovelleiro, D., Pérez-Riverol, Y., Reisinger, F., Ríos, D.,

Wang, R., and Hermjakob, H. (2013) The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* 41, D1063–D1069

Chapter 3

Monitoring light/dark association dynamics of multiprotein complexes in cyanobacteria using size exclusion chromatography-based proteomics

Ana C. L. Guerreiro[1,2], Renske Penning[1,2], Linsey Raaijmakers[1,2], Ilka M. Axman[3], Albert J. R. Heck[1,2] and A. F. Maarten Altelaar[1,2]

Manuscript under revision

[1]Biomolecular Mass Spectrometry and Proteomics, Utrecht Institute for Pharmaceutical Sciences and Bijvoet Centre for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands;
[2]Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, the Netherlands;
[3]Institute for Synthetic Microbiology, Heinrich Heine University Düsseldorf, Universitaetsstrasse 1, D-40225 Düsseldorf, Germany.

Abstract

Diurnal rhythms are recurring 24h patterns such as light/dark cycles that affect many natural environmental and biological processes. The cyanobacterium *Synechococcus elongatus* PCC 7942 (*S. elongatus*) produces its energy through photosynthesis and therefore its internal molecular machinery is strongly influenced by these diurnal rhythms. Moreover, it has one of the simplest, self-sustained, circadian rhythms, which has been extensively studied functionally and structurally and can be reconstituted *in vitro*. These characteristics together with the relatively small genome (2.7 Mbp) of *S. elongatus*, make it an ideal model system for the study of diurnal and circadian rhythms. Although it has been shown that the expression of many gene transcripts fluctuates in phase with the circadian rhythm, this was less pronounced at the protein level. With so little fluctuations in protein abundances we hypothesized that the diurnal adaptation may occur at the level of higher organization of protein complexes. Therefore, we probed the abundance and constituency of protein complexes present in cyanobacteria at two extreme time points, i.e. at the height of the day, and the depth of the night. Following several well-known complexes such as the RNA polymerase, the ribosome and complexes involved in photosynthesis, we observe for the first time that these complexes change not only in abundance but also in constituency, with associated proteins being either present or absent at these two different time-points. Therefore, we conclude that the dynamic assembly of protein complexes is indeed also a key-player in the processes governing the diurnal rhythm.

Introduction

The succession of day and night periods imposes drastic changes in all living organisms, during which also different molecular processes adapt. These may include different regulatory mechanisms and cellular pathways, such as the photosynthetic pathway, which occurs primarily during the day when sunlight is available. These different processes are usually entrained with the day and night cycles either through diurnal rhythms or through so-called circadian rhythms. Diurnal rhythms are biological cycles dependent on external cues, such as light and dark variations. On the other hand, circadian rhythms are adjustable to those cues, but they are also self-sustained biological cycles of 24 hours. Both types of rhythms are widely present in all kingdoms of life,

including in lower complexity organisms such as cyanobacteria[1]. The circadian rhythms in particular, have non-universal molecular mechanisms governing them, differing from species to species, and they are known to be rather complex, involving many genes in higher eukaryotic systems. Cyanobacteria, such as the *Synechococcus elongatus* PCC 7942 (*S. elongatus*), also exhibit a self-sustained biological clock, which is from a molecular perspective much more simple. Therefore, the cyanobacterial model, with its rather small genome (2.7 Mbp) [2,3], has been widely adopted to investigate circadian regulation [2,4–6]. The central core of this very robust pacemaker is composed of only three proteins, namely KaiA, KaiB and KaiC. In short, the mechanism involves the cyclic phosphorylation/ dephosphorylation of KaiC, by KaiA or KaiB, respectively, which also dynamically regulates the assembly of these three proteins into larger protein complexes [6,7].

The connection between the central circadian oscillator and its output to the rest of the cell is not completely known. Although it has been shown that the expression of many gene transcripts heavily fluctuates in phase with the circadian rhythm[5,8], our recent global proteomics investigation revealed that the diurnal fluctuations seemed to be less pronounced at the protein level[4]. It has been established that the central clock regulates gene expression through several effectors, such as the histidine kinase SasA (*Synechococcus* adaptive sensor A), the master regulator RpaA (regulator of phycobilisome-associated A/two component transcriptional regulator)[9–11], the RpaB regulator[12],LabA (low amplitude and bright A) or the sensor histidine kinase CikA (circadian input kinase A)[13,14]. The central clock is in turn entrained by the input factors Pex (PadR family transcriptional regulator) and LdpA (light-dependent period A).The regulatory functions mentioned above, and many others, depend on protein-protein interactions (PPIs) [15] and/or protein assemblies. Taking this into account and the fact that so few fluctuations in protein abundances were observed earlier[4], we argued that the diurnal adaptation might occur at the level of dynamic changes in protein assembly. Therefore, here we set out to analyze the relative abundance and constituencies of protein assemblies in cyanobacteria at the maximum of the light and dark phases.

The most common approaches for the study of protein interactions and assemblies consist of either yeast two-hybrid screens[16] or affinity purification coupled to mass-spectrometry (AP-MS) [17,18]. In the latter, the protein of interest is used as a "bait" to isolate and "pull down" its interacting proteins. AP-MS has been widely used and nowadays allows the comparison of different complexes at the same time[19,20], with increasingly high-

throughput[21,22]. It has been applied for instance to map full interactomes, also in the bacteria *E. coli*[23]. As a minor drawback of the approach, it has been shown that this technique, by using N- or C-terminal tags, can affect protein interactions and functions[24,25], besides being limited by the availability of tagged-constructs or antibodies, restraining the study of the dynamic PPIs across many conditions. Ultimately, as a targeted approach, AP-MS can only identify complexes of baits chosen *a priori*, disregarding the dynamics of unexpected PPIs. Recently, an alternative method was introduced coupling size exclusion chromatography (SEC) and high-resolution MS, which tackles some of the disadvantages of AP-MS. SEC is a widely used classical biochemistry technique in which proteins in solution are separated according to their shape/size. Under native conditions, this technique can be used to separate protein oligomers and or complexes based on their size and can be used as a pre-fractionation step, prior to mass spectrometry based proteomics analysis. This strategy has already been applied to chloroplast samples[26], as well as human cell lysates[27,28].

To our knowledge, a global analysis of the PPIs of *S. elongatus* has not yet been performed, and there is even more limited information on how such complexes would adapt to day (light) and night (dark) conditions. *A priori* it may be hypothesized that biological processes such as photosynthesis, which is regulated by large proteins assemblies, display night and day differences, possibly reflected by the recruitment or disassembly of different accessory proteins. Similarly, essential cellular processes, such as transcription and protein synthesis, previously associated with circadian regulation, may display comparable differences. To study adaptation of the *S. elongatus* PPIs to light and dark conditions, we used an MS centered approach conjugated with the pre-fractionation of protein assemblies by SEC. To compare the light and dark (LD) variations in protein complexes, we used *S. elongatus* cultures grown in LD cycles (12:12h) and collected samples from two opposite time-points separated by 12 hours: one from the light phase and one from the dark phase. The protein complexes obtained were separated according to their size by SEC, and identified and quantified through a high-resolution liquid chromatography (LC)-MS/MS approach, using standard protein complexes as internal standards. Using biological duplicates we were able to identify protein complexes from 100 kDa up to several megadalton, and analyze differences in their composure under light and dark conditions. Our data on several different protein assemblies include confirmations of previous observations, but also provides evidence for the putative role of new protein constituents in

playing specific roles at either the night or day phase of the cyanobacteria daily cycle.

Methods

1. Cyanobacteria cell culture

The wild-type strain of *Synechococcus elongatus PCC 7942* was grown as previously described[29] photoautotrophically in BG11-medium at 30°C under continuous illumination with white light of 80 µmol of photons/m^2s (Versatile Environmental Test Chamber, SANYO) and a continuous stream of air. Cell concentrations were measured by determining the optical densities of the culture at 750 nm (OD750) (SPECORD®200 PLUS, Analytik Jena). The culture was kept in log growth phase (up to an OD750 of 1.0) by dilution up to a specific volume and transferred to 12:12-hours light/dark cycle (LD) for 3 days. Synchronized culture was finally diluted to an OD750 of approximately 0.4 one day before the sampling started. Two time points with 12 hours intervals were sampled, for each of the two replicates. 40 ml of the culture was centrifuged at 15,000g for 10 min and the supernatant was removed. Cell pellet was resuspended in 1 ml BG11-medium, centrifuged again for 5 min. Supernatant was removed and pellet washed with 1 ml PBS buffer before last round of centrifugation. Cell pellets were frozen in liquid nitrogen prior to storage at -20°C.

2. Cell lysis

Cyanobacteria pellets were lysed in 50 mM KCl and 50 mM NaCOO (pH=7.2), containing 1 tablet EDTA-free protease inhibitor cocktail (Sigma) and 1 tablet PhosSTOP phosphatase inhibitor cocktail (Roche). After three mild sonication cycles at 4 ºC, total cell lysates were obtained through centrifugation at 14,000 rpm for 20 min, at 4 ºC. The supernatant was recovered and protein concentrations were determined with the Bradford method (BioRad). A Vivaspin500 100kMWCO filter (Sartorius Stedim Biotech GmbH, Goettingen, DE) was used to pre-select protein complexes larger than 100 kDa, by centrifugation at 15000 RCF at 4 ºC.

3. SEC (Size-Exclusion Chromatography)

The filtered cyanobacterial cell lysates were mixed with a mixture of protein and protein complex standards (Sigma-Alderich, St Louis, US), containing: thyroglobulin (3.3 μM, 669 kDa, complex), apoferritin (3.1 μM, 443 kDa, complex), beta-amylase (5.5 μM, 200 kDa, complex), alcohol dehydrogenase (9.1 μM, 150 kDa, complex) and carbonic anhydrase (28 μM, 29 kDa). For SEC the samples were injected (100 µl) onto a BioSep SEC S4000 column (Phenomenex) with a fractionation range of 15-1500kDa, supported by an Akta Basic HPLC system (GE Healthcare). Equilibration and elution were performed with 50 mM KCl, 50 mM NaCOO (pH=7.2). The flow rate was 450 µl/min and 54 fractions were collected at 260 µl/min, into Eppendorf tubes.

4. In solution digestion

The protein complexes collected in the SEC fractions were denatured using Rapigest (Waters) surfactant and subjected to 99 ºC incubation for 5 min. Next, reduction and alkylation of the cysteine residues was performed using 200 mM dithiotreitol (DTT) (Sigma) and 200 mM iodoacetamide (Sigma), respectively. The proteins were first digested with Lys-C (Roche Diagnostics, Ingelheim, Germany), at an enzyme:protein ratio of 1:75, for 4 hours at 37 ºC, followed by digestion with trypsin (Roche Diagnostics, Ingelheim, Germany), at an enzyme:protein ratio of 1:100, overnight at 37 ºC. Samples were desalted using an Oasis µElution plate (Waters) and dried *in vacuo*. The bovine serum albumin (Sigma) spiked-in later and used for normalization was digested separately, under similar conditions, with the exception of the surfactant being substituted by a solution of 8 M urea and 50 mM ammonium bicarbonate.

5. Liquid Chromatography-tandem Mass Spectrometry (LC-MS/MS)

LC MS/MS data was obtained using a Q Exactive Orbitrap (Thermo Fisher Scientific, Bremen) that was coupled to an Agilent 1290 Infinity UHPLC (Ultra-High Pressure Liquid Chromatography) system (Agilent Technologies, Waldbronn, DE). Both the trap (20 mm×100 µm I.D.) and the analytical column (35 cm×50 µm I.D.) were packed in-house using Poroshell C18 2.7 µm(Agilent Technologies, Waldbronn, DE). Peptides were trapped at 5µL/min in 100% solvent A (0.1 M acetic acid (Merck)). Elution was achieved with the solvent B (0.1 M acetic acid in 80% acetonitrile) at 100 nL/min. The 60 minutes gradient

used was as follows: 0-10 min, 100% solvent A; 10.1-45 min, 13–41% solvent B; 45-48 min, 41–100% solvent B; 48-49 min, 100% solvent B; 49-50 min, 0-100% solvent A; 50-60 min, 100% solvent A. The autosampler was programed to inject 2 fmol of tryptic digest of bovine serum albumin prior to sample injection. Nanospray was achieved using a coated fused silica emitter (New Objective, Cambridge, MA) (o.d. 360 µm; i.d. 20 µm, tip i.d. 10 µm) biased to 1.7 kV. The mass spectrometer was operated in the data dependent acquisition mode. A MS2 method was used with a FT survey scan from 350 to 1500 m/z (resolution 35,000; AGC target 3E6). The 10 most intense peaks were subjected to HCD fragmentation (resolution 17500; AGC target 5E4, NCE 25%, max. injection time 120 ms, dynamic exclusion 10 s). Predictive AGC was enabled.

6. Database search and validation

Raw data was analyzed by using Proteome Discoverer (version 1.3, Thermo), with the Mascot search engine (version 2.3.02, Matrix science). MS/MS data was searched against the *S. elongatus* PCC 7942 Uniprot database (version 4-2010) including a list of common contaminants and concatenated with the reversed versions of all sequences (5,826 sequences). The sequences from the SEC standard proteins and protein complexes (alcohol dehydrogenase, beta-amylase, apo-ferritin, carbonic anhydrase and thyroglobulin) and the bovine serum albumin were also added to this database. Trypsin was chosen as cleavage specificity allowing two missed cleavages. Carbamidomethylation (C) was set as a fixed modification. The variable modification used was methionine oxidation. The database searches were performed using a peptide tolerance of 50 ppm and a fragment mass tolerance of 0.05 Da (HCD). The 50-ppm mass window was chosen to allow random assignment of false positives that were later removed by filtering using the instruments actual mass accuracy (10 ppm). The identifications thresholds were set as a minimum peptide score of 20, a peptide rank of 1, a peptide length between 6 and 23 amino acids and a high peptide confidence. The peak area of each protein was exported from Proteome discoverer for further processing.

7. Data processing and analysis

The peak area of the spiked-in BSA internal standard was used for normalization. Each fraction was normalized according to the average peak

area of BSA, calculated across all fractions. This was followed by a similar normalization per experiment. The peak areas from the two biological replicate experiments were averaged and proteins identified with less than two peptides were excluded. Finally, since the SEC standard proteins were added with the same quantity to each replicate, we also normalized the peak areas to give a standard constant peak area ratio of 1 for the protein standards between light and dark.

To identify protein complexes that display regulated features between the night and day phase, we compared the averaged profiles of light and dark using an in-house developed analysis and visualization tool. Perseus software (http://www.perseus-framework.org/) was used for the clustering analysis and replicate correlation calculations. For all the hierarchical clustering analysis made, we first performed a Z-score normalization on the peak areas from light and dark, simultaneously, this allows better correlation of protein profiles, although one can only compare abundances between light and dark conditions. Secondly, we performed an imputation of the missing values by substituting them with zero; finally, we performed the unsupervised clustering of the rows using Pearson correlation, with k-means pre-processing and a complete linkage. The approximate molecular weight corresponding to each fraction was calculated considering the theoretical molecular weight of the SEC standard proteins and protein complexes and their elution profiles. A logarithmic function was fitted to this distribution to extrapolate the molecular weight across all fractions.

Results and Discussion

In many organisms, biological processes such as transcription, translation and photosynthesis follow circadian patterns and are regulated by large proteins assemblies. Cyanobacteria represent some of the oldest organisms that exhibit a very simple but effective circadian rhythm. Here, we set out to provide a global quantitative label-free analysis of native protein complexes originating from the cyanobacteria *Synechococcus elongatus* PCC7942, under light and dark (LD) conditions, using a combination of size exclusion chromatography (SEC) and high-resolution mass spectrometry (Fig. 1A). To evaluate LD variations in protein complexes, we selected two “extreme” time-points separated by 12 hours across a 24-h LD cycle, specifically one in the middle of the light phase and one in the middle of the dark phase. By using mild cell lysis conditions, we were able to preserve protein complexes in their

native state and subsequently separate them according to their complex/protein size, using SEC. The SEC chromatograms obtained were reproducible between light and dark samples and corresponding biological replicates, as shown by their overlap (Fig.1B). To normalize the proteins' peak areas, a SEC standard composed of five protein complexes (alcohol dehydrogenase, beta-amylase, apo-ferritin, carbonic anhydrase and thyroglobulin) was spiked into the samples before SEC and a bovine serum albumin (BSA) digest standard before LC-MS analysis (Fig. 1A).

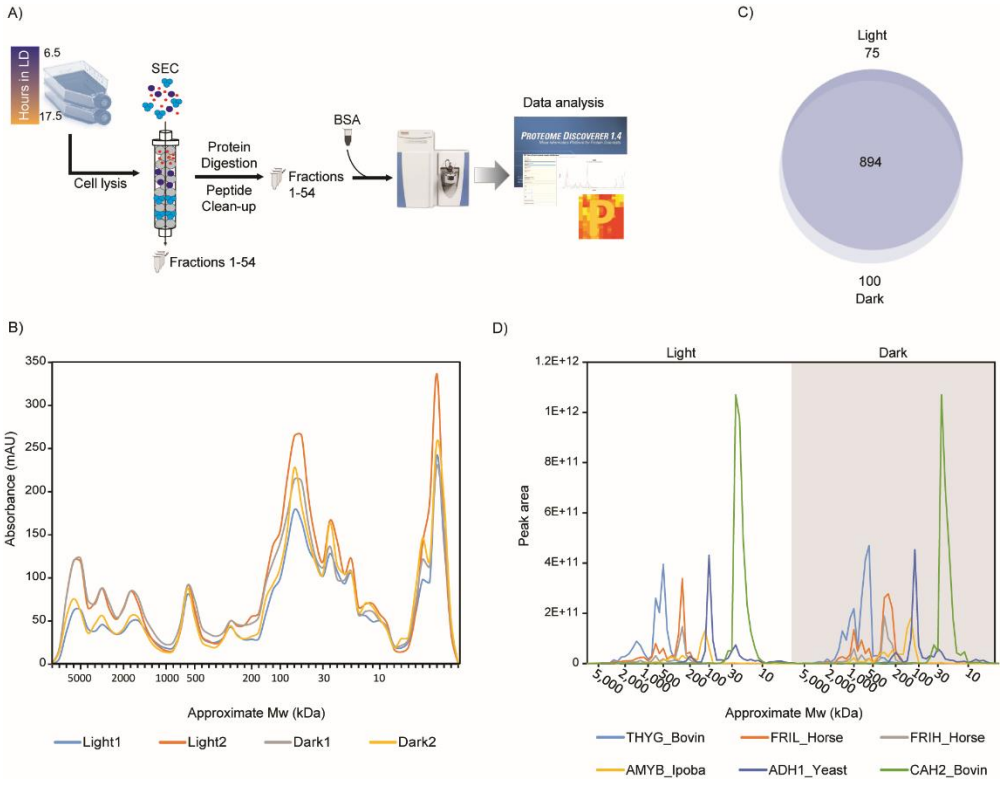


Figure 1 – Native size exclusion chromatography of S. elongatus protein complexes. A) *S. elongatus* cells were synchronized and cultured in LD (12:12h) conditions. Samples were collected from two time-points separated by 12 hours across a 24-h LD cycle. Mild cell lysis conditions enabled native protein complex preservation, followed by separation according to their complex/protein size using SEC. B) SEC chromatogram reproducibility across light and dark samples and corresponding replicates. C) Overlap of protein identifications between light and dark conditions. D) SEC standards elution profiles for light and dark conditions, after replicate normalization and averaging.

Each SEC chromatogram was fractionated in 54 fractions that were individually analyzed by high-resolution nanoLC-MS/MS following the spike-in of the tryptic BSA. In total, 2x54 fractions for the light condition and 2x54 fractions for the dark condition were analyzed, leading to 216 nanoLC-MS/MS experiments performed. In the data analysis, only proteins identified with two or more peptides were considered, resulting in confident identification of 1061 proteins across the light and dark phases, with an overlap of ~84% (Fig. 1C). To measure the reproducibility, a Pearson correlation was calculated for each fraction comparing the light and dark replicates, which showed high correlation with an average of 0.8. Next, the normalized peak areas were averaged between the two replicates of each experiment, to allow quantitative comparisons, resulting in very similar peak areas of the spiked-in five SEC internal protein complex standards for the light and dark experiments (Fig. 1D), indicating correct normalization.

1. Global clustering analysis

To compare the SEC elution profiles of each protein in the light and in the dark phase, we performed an unsupervised hierarchical clustering analysis taking both LD profiles in parallel (Fig. 2), using Perseus software (http://www.perseus-framework.org/). Generally similar elution profiles between the light and dark cycle were obtained, albeit that several distinct differences between the two conditions were observed. The main differences are relative abundance variation and to a lesser extent spread of the corresponding elution peaks. In some specific cases proteins were found eluting in certain SEC fractions only in either the light or dark phase, as discussed further below. Here we will focus on proteins identified eluting in the high molecular weight range, as shown in Figure 2, which was particularly enriched for 70S ribosome and many proteins involved in photosynthesis. This is as expected since the native molecular weight of these complexes is in the mega-Da (MDa) range. Also the RNA polymerase complex could be observed quite distinctively, appearing isolated in one relatively narrow SEC cluster. These complexes will be examined and described in more detail in the next sections. Of note, many other proteins were observed to elute at high Mw ranges. However, many proteins in cyanobacteria are only weakly functionally annotated and even less is known about their quaternary structure and/or contribution/association into larger protein complexes. Illustratively, ~30% of all proteins identified in our analysis are annotated as functionally uncharacterized.

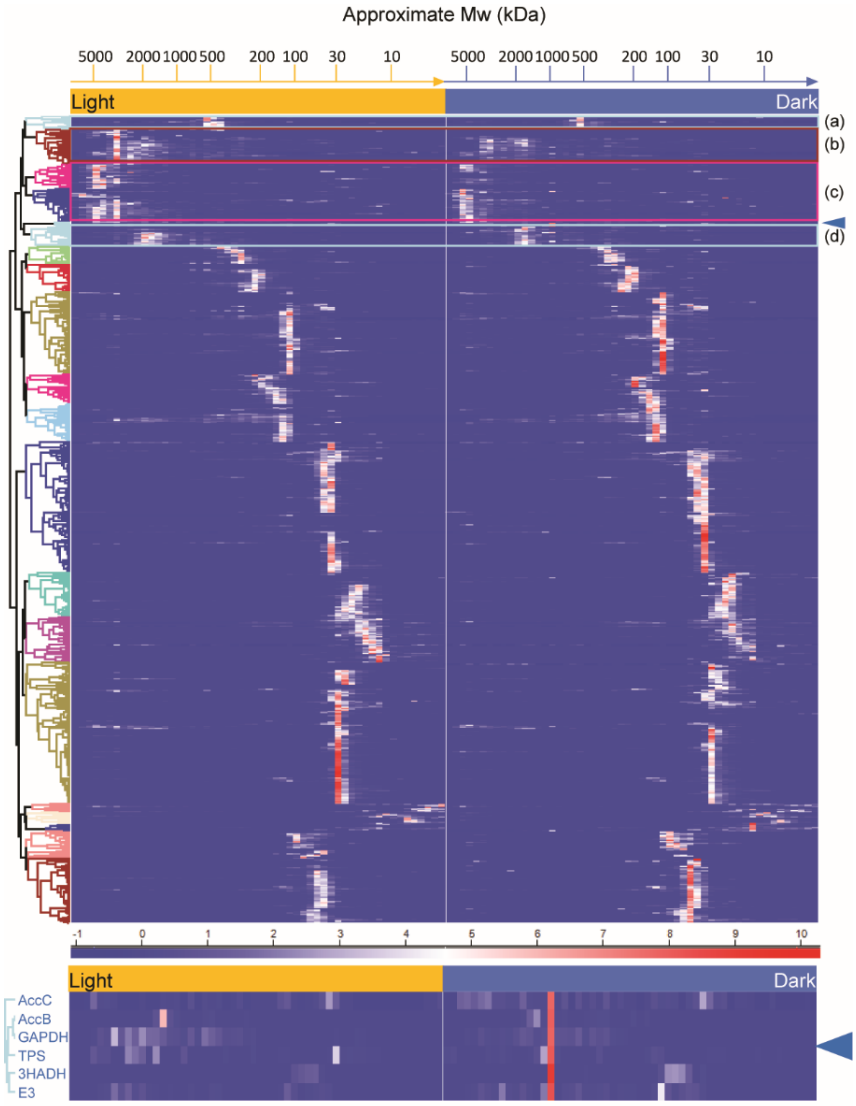


Figure 2 – Hierarchical clustering of all identified proteins. Depicted heat map of unsupervised hierarchical clustering taking the standardized (Z-score) LD profiles in parallel, using Perseus software, as described in the methods section. Red corresponds to higher protein abundance and blue to lower abundance. Cluster enrichment in the MDa range: a) RNA polymerase; b) 50S ribosomal proteins; c) photosynthetic proteins; d) 30S ribosomal proteins; light-blue arrow (zoom-in below): LD dynamic cluster.

Our experiments may be the first evidence of their presences and/or association within larger assemblies in cyanobacteria. However, as little is known about these proteins we did choose to focus on the better annotated protein assemblies. To minimize elution biases, we performed separate cluster analysis for these MDa complexes, containing the complex constituents and known or highly probable interactors (based on common biological processes).

2. RNA polymerase

In the global clustering analysis we observed a clear cluster enriched for the RNA polymerase (RNAP) complex (Fig. 2 (a) and Suppl. Fig. 1), eluting around ~500 kDa. Like in *E. coli*, in cyanobacteria this complex is typically constituted of two alpha subunits, a beta and a beta' subunit and an omega subunit, all identified in our analysis. The cyanobacterial RNAP has also a gamma subunit which originates from the beta' subunit of E. coli[30]. This cyanobacterial core-complex should have a theoretical Mw of ~415 kDa. For transcription, this core RNA polymerase enzyme complex recruits a variety of sigma factors enabling transcription of specific DNA template strands. If we zoom into the RNAP protein core complex profiles, we first see a substantial higher abundance of all core-components during the light phase (Fig 3A). In our earlier cyanobacterial circadian proteome study, where we reanalyzed transcript data from Ito et al.[5], we did observe that certain components of the RNAP (alpha, beta' and SigA1) indeed exhibit circadian rhythms in their transcripts, but not in their protein abundances[4]. Moreover, here the components of this core-complex are observed in several distinct fractions, also depending on the light or dark conditions. This likely indicates that other accessory proteins co-assemble to this complex depending on the LD phases. To investigate the nature of these possible interactors of the RNAP, a hierarchical clustering was performed only for the RNA polymerase and transcriptional related proteins (Suppl. Fig. 1). The targeted clustering analysis, reveals that the RNAP core complex elutes together with one of the expected sigma factors SigA1 (light and dark) and the diguanylate cyclase/phosphodiesterase with PAS/PAC and GAF sensor(S) (DGC/PDE) (only light). As described above, the sigma factor assists RNAP in binding to the correct location on the DNA – the -10 region within the promoter - for transcription initiation[31]. The phosphodiesterase is related to transcription regulation, albeit primarily through its phosphorelay sensor kinase activity. This protein was only detected in the light phase and it is known that the PAS

domains are associated with monitoring of light changes[32]. Therefore, this potential co-assembly might be related to differential gene expression control.

In the targeted clustering analysis we also detected the circadian clock protein kinase KaiC (light and dark), as a transcriptional regulator. KaiC elutes across different fractions, including together with the RNAP complex, SigA1 and DGC/PDE. Since interactions between the central clock proteins, transcription factors[9] and sensor proteins[33,34] have been suggested, our data make it plausible that KaiC interacts with SigA1 and DGC/PDE. Notably, in our analysis KaiC also co-elutes at much higher Mw (~4 MDa) with several DNA binding proteins, related to DNA replication and repair.

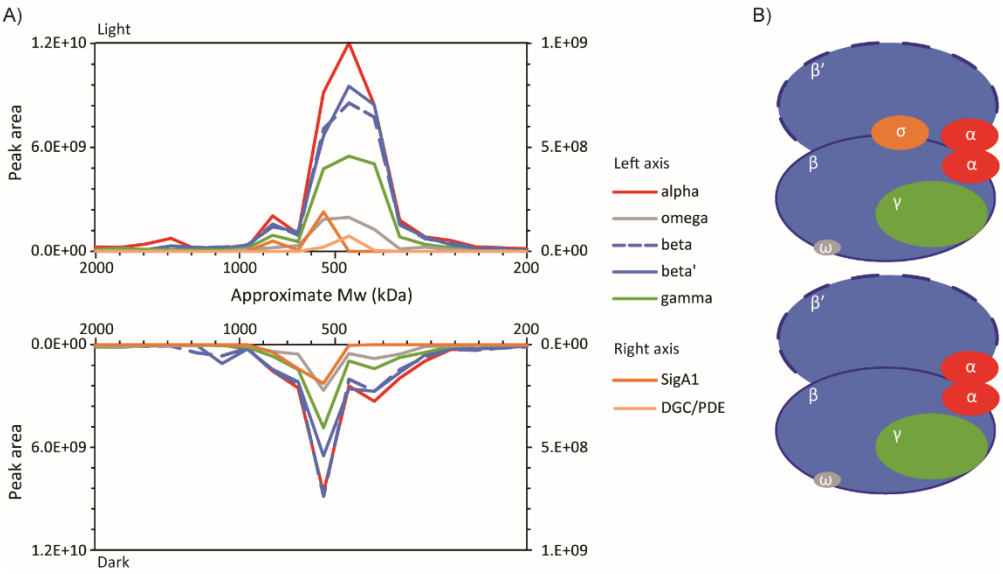


Figure 3 – RNA polymerase complex dynamics. A) SEC protein LD profiles of RNA polymerase subunits and co-eluting proteins. Essentially two types of interactions were detected; RNA polymerase with and without its sigma factor (SigA1). In the light phase RNA polymerase additionally co-elutes with the Diguanylate cyclase/Phosphodiesterase with PAS/PAC and GAF sensors (DGC/PDE). A different intensity scale is represented on the right y-axis of the plots, for amplification of the SigA1 and DGC/PDE profiles. B) Simplistic representation of both complexes.

In summary, considering the predicted Mw and the co-elution patterns, we observe the interaction of the RNAP core-complex with SigA1 and the DGC/PDE dimer, during the light phase (Fig. 3A). There is also an interaction of the RNAP complex and SigA1 with the KaiC hexamer (Suppl. Fig. 1). During

the dark phase the RNAP core-complex can be seen alone, interacting with SigA1 (Fig. 3A-B) and with KaiC (Suppl. Fig. 1). Our data clearly highlight the cyclic behavior of these assemblies.

3. Ribosomal complex

The bacterial ribosome (70S) is composed of two subunits, the small (30S) and the large (50S). In this study, we were able to identify all proteins of the ribosomal complex, where 20 proteins from the 30S subunit and 27 from the 50S subunit passed our stringent filtering criteria. The S20 protein, from the 30S subunit, and 4 proteins from the 50S subunit (L16, L32, L34 and L36) were identified with only one unique peptide. In the global clustering analysis, most of the ribosomal proteins co-eluted, being enriched in two different clusters (Fig. 2 (b) and (d)).To have a closer look at the ribosome and its potential interactors, we performed a clustering analysis on the small and large ribosomal subunits and proteins that are known to bind to the ribosome, such as those involved in the regulatory steps of translation, co-translational modifications and ribosome biogenesis (Fig. 4 and Suppl. Fig. 2). In the resulting heatmap, we can clearly see a separation of the small 30S subunit and the large 50S subunit. The small subunit eluted at ~1.5-2 MDa and the large subunit at ~3.5 MDa. The total molecular weight for the ribosome is 2.5 MDa, meaning that in our SEC analysis they have an apparent higher Mw. This has been observed before, where it was found that the rRNAs belonging to these subunits were not detached in the course of the analysis[26]. Potentially we capture these ribosomal complexes in different stages of maturation or even as polysomes.

Looking closer into the 50S cluster, we observe the co-elution of two initiation factors (IF-2 and IF-3) (blue arrows Fig. 4A), where IF2 is most abundant when co-eluting with the 50S subunit and IF3 with the 30S subunit. It is noticeable that IF-3 co-elutes with the peak of the 30S subunit at a higher abundance during the dark phase and in similar abundance with 50S during the light phase, which is in agreement with the translation initiation complex being initially constituted of the 30S subunit and IF-3[35]. On the other hand, this can also indicate a dynamic relation between the initiation complex formation and the ribosome formation, the latter being more frequent during the day. If we look at the 30S cluster, we find also the RNA methyltransferase TrmH (brown arrow Fig. 4A and B, suppl. Fig. 2).

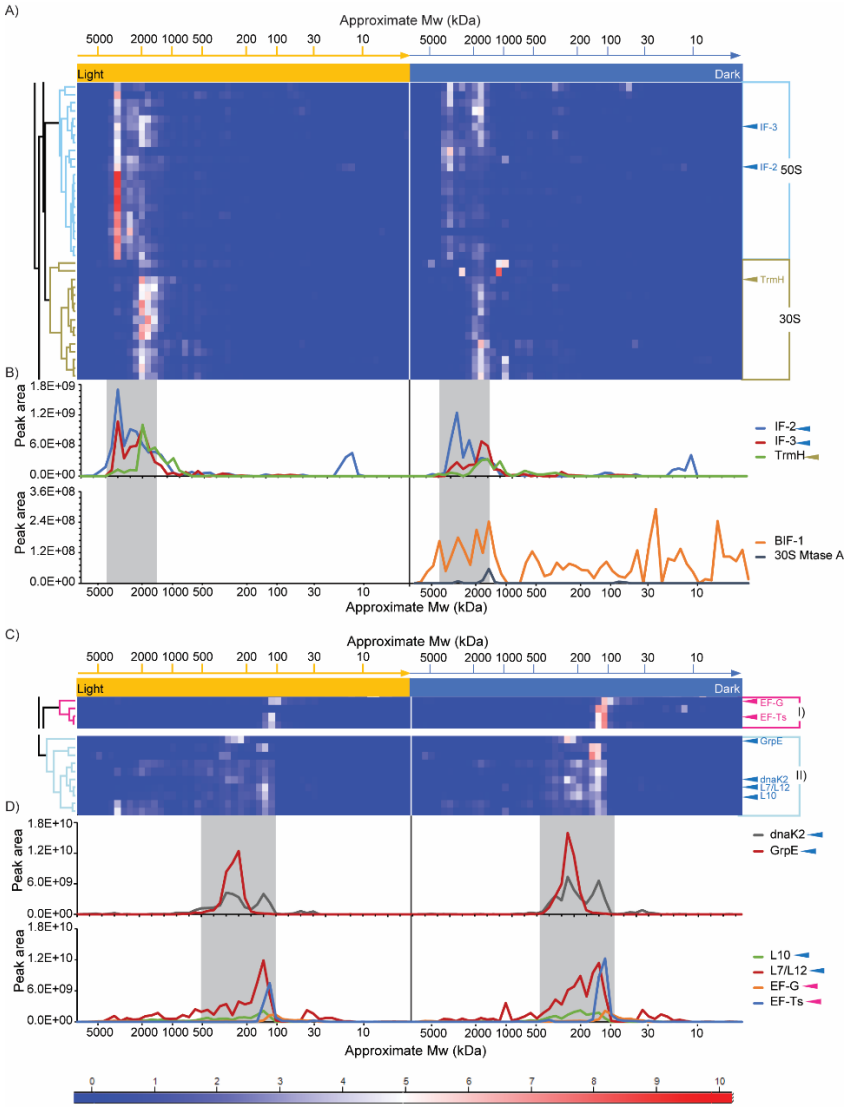


Figure 4 – Dynamic association of ribosomal and translation related proteins. Detailed heatmap of the unsupervised hierarchical clustering, the standardized (Z-score) LD profiles of the ribosomal and translation related proteins in parallel, as described in the methods section. Red corresponds to higher protein abundance and blue to lower abundance. Grey areas emphasize co-elution in the profiles. A) 50S and 30S enriched clusters and protein elution profiles of co-eluting translation factors. B) L7/12 stalk cluster and co-eluting translation factors and chaperones, with individual elution profiles below for clarification purposes.

This methyltransferase is involved in the biogenesis of the ribosome, more specifically in the modification and maturation of the rRNA. This important factor was previously observed to have a possible interaction with the 30S particle[26].

There are other proteins co-eluting with the two main subunits, yet are not included in the same cluster due to small profile differences. These co-eluting proteins include the bacterial translation initiation factor 1 (BIF-1) and the rRNA small subunit methyltransferase A (30S Mtase A) (Fig. 4B, bottom graph). The fact that they mainly co-elute with the 30S subunit in the dark phase is causing the difference in the profiles. For now there is no obvious explanation for this observation, however it is known that the methyltransferase exerts its function in the biogenesis of the 30S particle; and BIF-1, as an initiation factor, is primarily associated to the small subunit[35]. Also eluting with the ribosome, we observe the ribosome maturation factor RimP, involved in rRNA processing (Suppl. Fig. 2).
The DnaK/Hsp70 and GrpE (chaperone and co-chaperone) do co-elute with several ribosomal proteins, namely in the L10-L7/L12 stalk, at a lower mass (Fig. 4C and D). The same occurs, for the elongation factor EF-G, EF-Ts and the sigma54 modulation protein/SSU ribosomal proteinS30P (SSU S30P) (Suppl. Fig. 2).This is expected, as these proteins are supposed to bind to that specific ribosomal region. Moreover, it was suggested that in Eukaryotic cells L12 orthologs exist in an unbound state and L12 is part of a regulatory recruitment mechanism of translational factors[36]. Despite the fact that in *E. coli* this was stated to be unlikely[37], our results suggest such behavior in cyanobacteria.

Comparing the light and dark conditions, we see very similar trends (Fig. 4 and Suppl. Fig. 2). However, in the light phase, the peak areas of the ribosomal subunits are more intense. These observations seem to show that the effective and functional ribosome or ribosomes are more prominent during the light phase, showing more dynamic interchanges between its different states. In our previous global proteomics study the ribosomal proteins showed cyclic 48h profiles at the transcript level (re-analyzed data from Ito *et al*.[5]) and cyclic differences between light and dark at the proteome level[4]. The now observed pattern occurs in an opposite manner than the one obtained by global proteomics, which might be attributed to the fact that during the dark phase the majority of the proteins are in a free form and during the light phase most of the proteins form larger assemblies.

4. Complexes involved in photosynthesis

Another prominent set of two clusters in the global cluster analysis is enriched for the photosynthetic and thylakoid membrane-associated proteins (Fig. 2 c)). The photosynthetic pathway is essentially composed of six major complexes: phycobilisome, photosystem I (PSI) and II (PSII), cytochrome b6f, NAD(P)H-quinone oxireductase complex (NDH-1)) and ATP synthase. In the clustering analysis, we included the 52 proteins that we were able to identify, out of the 91 proteins that make up these complexes (Supp. Fig. 3). Below, we will mainly discuss the phycobilisome, the PSI and PSII complexes.

The photosynthetic pathway is essentially composed of six major complexes: phycobilisome, photosystem I (PSI) and II (PSII), cytochrome b6f, NAD(P)H-quinone oxireductase complex (NDH-1)) and ATP synthase. In the clustering analysis, we included the 52 proteins that we were able to identify, out of the 91 proteins that make up these complexes (Supp. Fig. 3). Below, we will mainly discuss the phycobilisome, the PSI and PSII complexes.

Looking at the photosystems separately, we see that the corresponding core of each complex clusters together (Fig. 5). In the case of photosystem I (Fig. 5A), the proteins eluting at high mass include the P700 chlorophyll *a* apo-protein A1 (PsaA) and A2 (PsaB) (core complex of PSI monomer), the reaction center subunits III (PsaF) and XI (PsaL) (at the interface between PSI monomers) and the iron-sulfur center protein (PsaC) and the reaction center subunit IV (PsaE) (cytoplasm side). The NDH-1 complex co-elutes with PSI at high Mw, which agrees with their involvement in the cyclic electron flow process during photosynthesis[38]. For the photosystem II (Fig. 5B (I)), among the proteins eluting at higher mass, we observe the CP47 chlorophyll apo-protein (CP47), the 44 kDa reaction center protein (CP43) and the D2 protein (D2) (core complex of PSII monomer), the manganese-stabilizing polypeptide (PsbO) and lipoprotein Psb27 (Psb27) (luminal side). Additionally, the cytochrome b6-f complex iron-sulfur subunit (Cyt b6-f) was found to co-elute at high Mw (Suppl. Fig. 3A). This complex is responsible for electron transfer between the PSI and PSII and the cyclic electron transfer[39]. In this range, the PSII subunits elute in different fractions leading to the possibility of different states of the photosystem. It is interesting to see that, in those different states, Psb27 and PsbO do not always co-elute. This is in line with a previous report[40], which shows that these two proteins bind in the same place. During the PSII *de novo* assembly, Psb27 binds at an earlier stage than PsbO.

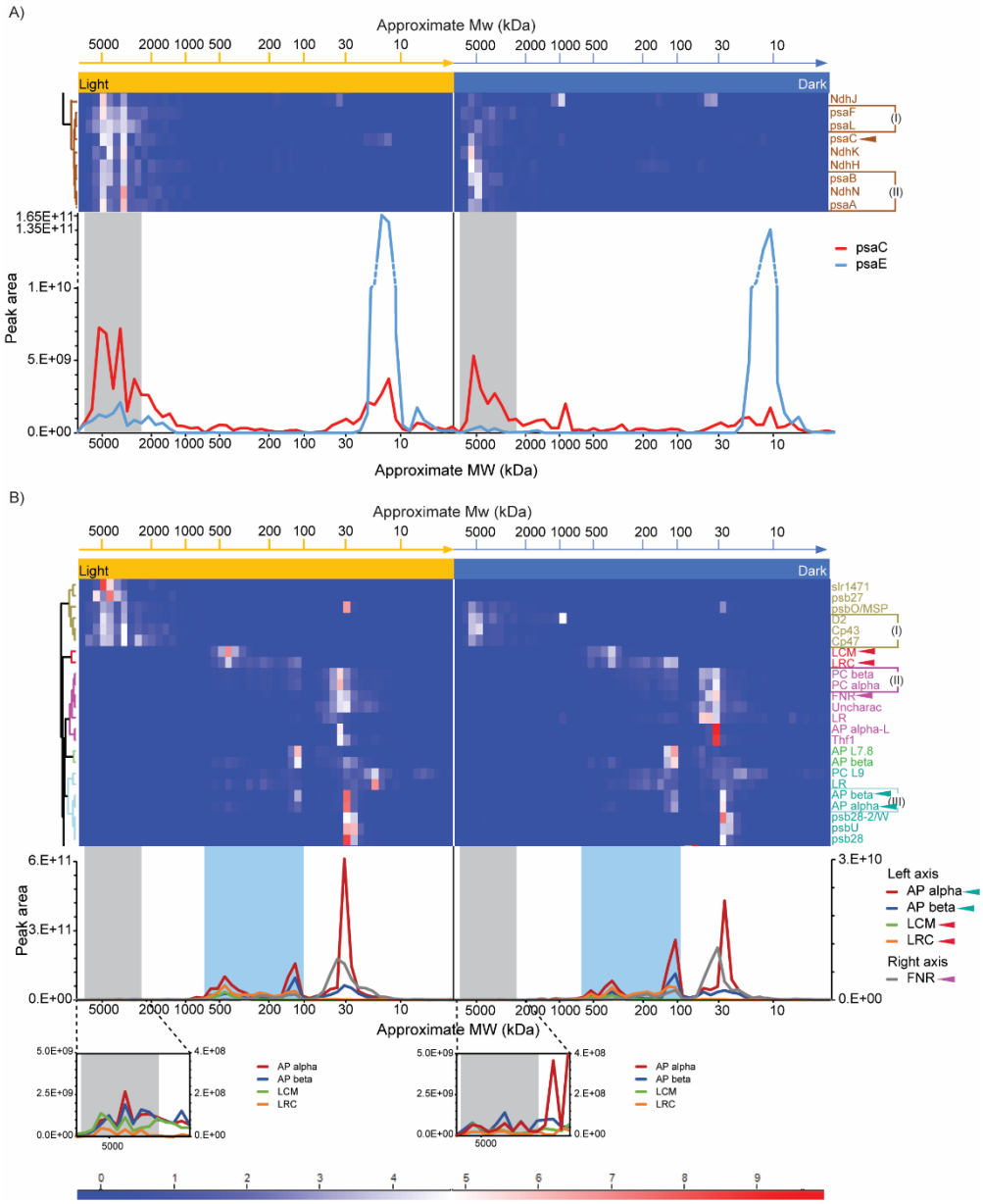


Figure 5 – Dynamics of the proteins from the photosynthetic pathway. Detailed heatmap of the unsupervised hierarchical clustering, performed with the standardized (Z-score) LD profiles of the photosynthetic proteins in parallel, as described in the methods section. Red corresponds to higher protein abundance and blue to lower abundance. Grey and blue areas emphasize co-elution in the profiles. A) Cluster of the proteins from the photosystem I (PSI) and NAD(P)H-quinone oxireductase complexes (NDH-1); highlighted SEC LD profiles: I)

Photosystem I interfaces and II) Photosystem I core (psaB and psaA). PsaE and PsaC profiles shown below demonstrate their elution at the MDa range. Dashed line indicates a break in the scale. B) Heatmap of proteins from the photosystem II (PSII) and phycobilisome complexes; highlighted SEC LD profiles: I) Photosystem II core proteins, II) Phycobilisome rod proteins, III) Phycobilisome core proteins. Profiles below demonstrate the co-elution of L_{CM}, L_{RC}, AP alpha and beta at the MDa range (zoomed in); and the co-elution of FNR (right axis) with the core proteins (AP).

When comparing the light and dark conditions, the complexes show more dynamic interactions and higher intensity during the light phase. This observation can be easily explained by the photosynthetic function of these complexes during the light phase, while they are apparently actively down regulated during the dark phase. Similar to the observations made for the ribosome data, most proteins belonging to the phycobilisome system displayed a cyclic pattern between light and dark in our previous global proteomics study[4], but in an opposite manner, being higher during the night phase. An explanation similar to the ribosomal complex can be applied, in which protein complexes are disassembled during the night and proteins are more abundantly present as single entities.

The main proteins of both PSI and PSII, namely the core proteins, the luminal side from PSI and the interfaces from PSI, co-eluted in the high mass range. They do not interact physically, but they are usually found in close proximity, in order to ensure the electron flow. The fact that they elute together at such high MW (4 MDa) could be explained by an indirect interaction between the PSI and PSII, through the ferredoxin-NADP+ oxidoreductase (FNR) and the phycobilisome (discussed below), which was suggested before[41,42].

Next, we looked more specifically at the phycobilisome complex (Fig. 5B). The phycobilisome acts as an antenna, harvesting the photon-energy. It is composed of a core complex with allophycocyanins (AP) and its linkers; and several rod complexes with phycocyanins (PC) and its linkers. We were able to identify all of its components as well as their assembly factors (Suppl. Fig. 3). In the heatmap in Figure 5B, we can see a clear differentiation between these sub-complexes, namely the rod (II) and core (III) proteins. We also noticed specific interactions within these sub-complexes. In the rod complex, we see a clear co-elution with FNR which is associated with PSI (Fig. 5B, light blue area). FNR is in turn also eluting at high Mw with a lower abundance. This coincides with our previous observation for an interaction of PSI with PSII, which could have a very transient nature and is mediated by the

interaction of FNR with the phycobilisome complex and PSI, and the interaction of the phycobilisome with PSII. Additionally, we see the core-membrane linker (L_{CM}) eluting in the 4MDa range (Fig. 5B, zoom in), demonstrating the interaction of phycobilisome and PSI further. This is in line with the recent identification of a megacomplex (MCL), consisting of phycobilisome-PSII-PSI, with L_{CM} as the central binding place for PSI[43]. The L_{CM} itself is more prominent in light conditions and mostly at high mass, co-eluting with the AP alpha, AP beta and the rod-core linker (L_{RC}), which stabilizes the connection between the rods and the core, facilitating energy transfer (Fig. 5B, grey and blue areas); but also with PSII proteins, which are the ultimate energy receivers (Fig. 5B, heatmap).These interactions indicate an active antenna complex (phycobilisome-PSII-PSI) during the light phase. The rest of the linkers located in the rod and core structures, have a more variable distribution across the fractions. Since we also observe greater variability of Mw within the rod proteins, we can speculate that these are related to the dynamic characteristics of the rods, which have been shown to vary in number of single units according to the light changes[41].

5. Light versus dark dynamics

Our global clustering analysis revealed several dynamic protein complex elution profiles. One case of extreme difference between the light and dark phase concerns a group of proteins that do not co-elute during the light phase but during the dark phase align around one MDa (Fig. 2 blue arrow and Fig 6A). In this cluster, we find 3-hydroxyacid dehydrogenase (3HADH), dihydroplipoyl dehydrogenase (E3), thiamine-phosphate synthase (TPS), acetyl-coenzyme A (acetyl-CoA) carboxylase carboxyl transferase subunit alpha/biotin carboxylase (AccC), biotin carboxyl carrier protein (AccB) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH). An immediate connection between these proteins is not easily made, since in the search tool for the retrieval of interacting genes/proteins (STRING)[44], the only known interaction with medium confidence is the one between AccB and AccC. These two proteins are part of the acetyl-CoA carboxylase complex hexamer (ACC), from which we also were able to identify the subunit AccD. If we look closer into these proteins' profiles (Fig. 6B), we see that the AccB elutes alone in the light phase, but it elutes together with AccC in the dark phase. However, the interaction between these two does not make up for the total mass corresponding to the fraction where they elute, unless they oligomerize (~1 MDa).

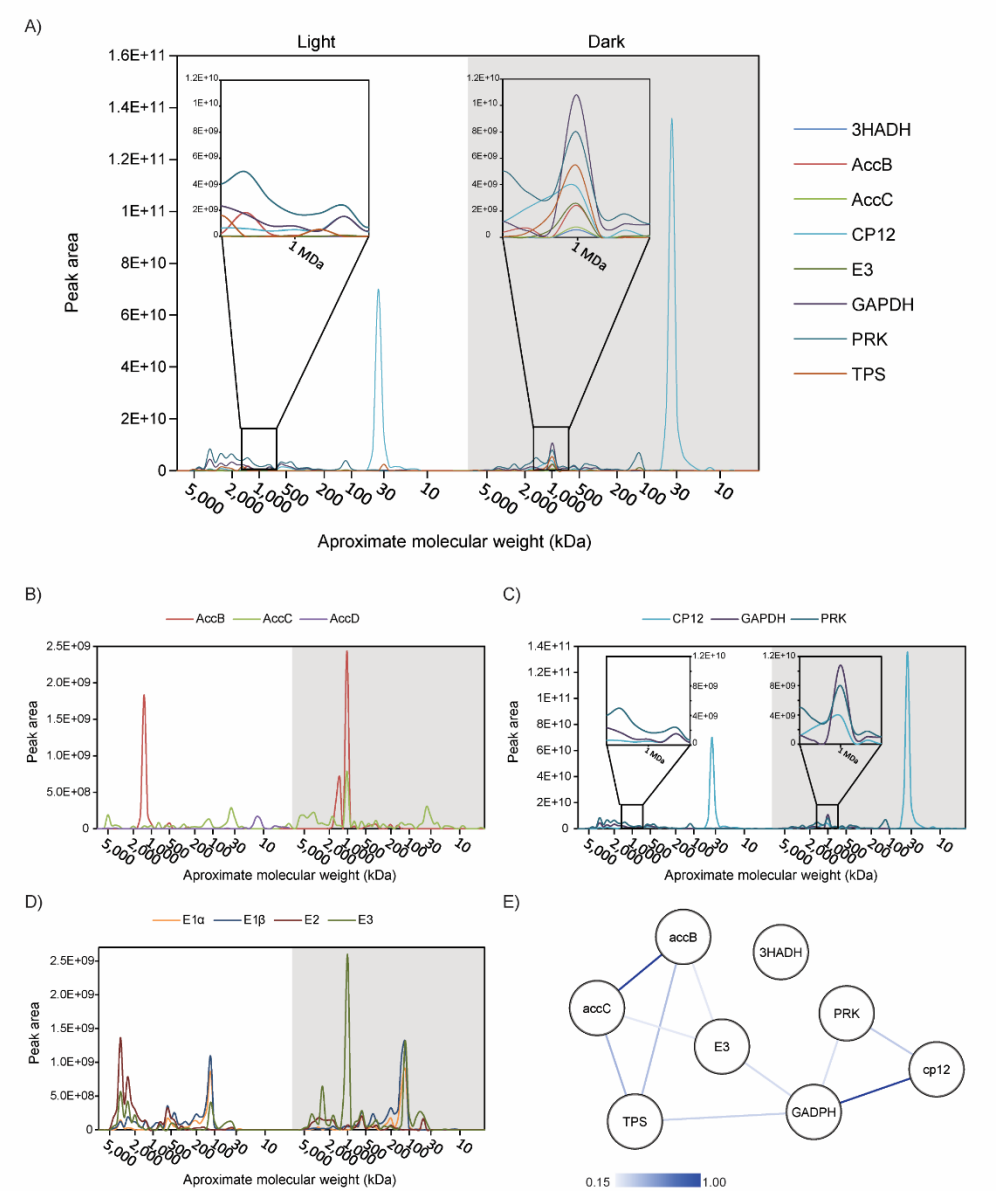


Figure 6 – Cluster with LD dynamics. A) SEC LD profiles of proteins co-eluting at 1MDa, in dark conditions, namely 3-hidroxyacid dehydrogenase (3HADH), dihydroplipoyl dehydrogenase (E3), thiamine-phosphate synthase (TPS), acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha/biotin carboxylase (AccC), biotin carboxyl carrier protein (AccB), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), phosphoribulokinase (Prk) and the linker CP12. A zoom in of the 1MDa range is shown for clarity. B) SEC LD profiles of the acetyl-CoA carboxylase complex proteins identified (AccB, AccC, AccD). C) SEC LD

profiles of the carbon fixation complex GAPDH, Prk and CP12. A zoom in of the 1MDa range is shown for clarity. D) SEC LD profiles of the pyruvate dehydrogenase complex proteins (E1α, E1β, E2, and E3). E) Interaction network of the proteins co-eluting at the 1MDa range. Interactions based on gene neighborhood, co-occurrence, co-expression, experimental evidence of interaction, pathway interaction and text mining, retrieved from STRING[44] and analyzed with Cytoscape[50]

If we look at other proteins in this cluster we can make similar observations. For the GAPDH, we could identify known interactors in our dataset: the phosphoribulokinase (Prk) and the linker CP12[45–47]. Prk and GAPDH are keys enzymes in carbon fixation, through the Calvin cycle, which occurs with light. As expected, the GAPDH-CP12-Prk complex co-elutes more prominently in the dark phase, since their association at night corresponds to their inactive state (Fig. 6C). In the case of the E3 protein, which is part of different multi-enzyme α-keto acid dehydrogenases, we were also able to find known interactors, which make up the pyruvate dehydrogenase (PDC): the pyruvate dehydrogenase E1 component subunit alpha (E1α), the pyruvate/2-oxoglutarate dehydrogenase complex dehydrogenase (E1β) component (beta subunit) and the pyruvate dehydrogenase dihydrolipoamide acetyltransferase component (E2)[48,49]. These proteins form a complex of ~4MDa, which is visible in the SEC profile (Fig. 6D). This megadalton PDC appears to exist mainly during the light phase. However, during the dark phase, only the E3 component elutes at 1MDa, aligning with the rest of the cluster in figure 6A. By extending the STRING network to all these different components observed eluting at 1MDa, we now find a possible interaction between the PDC and the ACC (Fig. 6E). The PDC is involved in the formation of Acetyl-CoA from pyruvate, which leads to synthesis of citric acid, in the citric acid cycle. ACC on the other hand is activated by citric acid and uses acetyl-CoA in fatty acid metabolism. These common aspects can culminate in an effective interaction related to their regulation.

Conclusion

In summary, we report the first global analysis of macromolecular assemblies in the cyanobacterium *S. elongatus* across day (light) and night (dark) conditions. Our SEC and MS based proteomics approach enabled the identification of megadalton protein assemblies in cyanobacteria, including the ribosomal and the photosynthetic complexes. This approach, which was implemented with higher separation resolution than the previous study in chloroplasts[26], allowed us the use of hierarchical clustering of the SEC protein

elution profiles to better infer particular PPIs. This methodology allowed us to infer dynamic interactions for several protein complexes such as for the sigma factor SigA1 and the phosphodiesterase sensor in the RNA polymerase complex. For the megadalton complexes, involved in transcription and photosynthesis, we observe an intensity decrease during the dark period, which is expected especially for the photosynthetic pathway. These complexes also seem to have more assembly variety during the light phase. We also observe an unexpected association of (members of) complexes that are involved in glycolysis, pyruvate metabolism and carbon fixation, which specifically cluster together in the dark phase, hinting at a common role or regulation. The differences observed between the light and dark conditions in these complexes indicate a cyclic regulation of essential cellular processes, which have been connected to circadian rhythm regulation before[51–53]. In conclusion, our work demonstrates the advantages of using SEC coupled to high resolution LC-MS based proteomics in the study of PPI dynamics and provides novel insights in PPIs, opening opportunities to new investigations at a more targeted level.

Supplementary information

Supplementary Figures can be found in Chapter 7.

Acknowledgements

This work was partly supported by the PRIME-XS project (grant agreement number 262067) funded by the European Union 7th Framework Program. This work is part of the project Proteins At Work, a program of the Netherlands Proteomics Centre financed by the Netherlands Organization for Scientific Research (NWO) as part of the National Roadmap Large-scale Research Facilities of the Netherlands (project number 184.032.201). A.F.M.A. acknowledges additional support by the Netherlands Organization for Scientific Research (NWO) through a VIDI grant (723.012.102).

References

(1) Johnson, C. H.; Stewart, P. L.; Egli, M. The cyanobacterial circadian system: from biophysics to bioevolution. *Annu. Rev. Biophys.* **2011**, *40*, 143–167.

(2) Holtman, C. K.; Chen, Y.; Sandoval, P.; Gonzales, A.; Nalty, M. S.; Thomas, T. L.; Youderian, P.; Golden, S. S. High-throughput functional analysis of the Synechococcus elongatus PCC 7942 genome. *DNA Res.* **2005**, *12*, 103–115.
(3) Andersson, C. R.; Tsinoremas, N. F.; Shelton, J.; Lebedeva, N. V.; Yarrow, J.; Min, H.; Golden, S. S. ScienceDirect.com - Methods in Enzymology - Application of bioluminescence to the study of circadian rhythms in cyanobacteria. *Methods Enzymol.* **2000**, *305*, 527–542.
(4) Guerreiro, A. C. L.; Benevento, M.; Lehmann, R.; van Breukelen, B.; Post, H.; Giansanti, P.; Maarten Altelaar, a F.; Axmann, I. M.; Heck, A. J. R. Daily rhythms in the cyanobacterium synechococcus elongatus probed by high-resolution mass spectrometry-based proteomics reveals a small defined set of cyclic proteins. *Mol. Cell. Proteomics* **2014**, *13*, 2042–2055.
(5) Ito, H.; Mutsuda, M.; Murayama, Y.; Tomita, J.; Hosokawa, N.; Terauchi, K.; Sugita, C.; Sugita, M.; Kondo, T.; Iwasaki, H. Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in Synechococcus elongatus. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 14168–14173.
(6) Brettschneider, C.; Rose, R. J.; Hertel, S.; Axmann, I. M.; Heck, A. J. R.; Kollmann, M. A sequestration feedback determines dynamics and temperature entrainment of the KaiABC circadian clock. *Mol. Syst. Biol.* **2010**, *6*, 389.
(7) Nakajima, M.; Imai, K.; Ito, H.; Nishiwaki, T.; Murayama, Y.; Iwasaki, H.; Oyama, T.; Kondo, T. Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro. *Science* **2005**, *308*, 414–415.
(8) Vijayan, V.; Jain, I. H.; O'Shea, E. K. A high resolution map of a cyanobacterial transcriptome. *Genome Biol.* **2011**, *12*, R47.
(9) Takai, N.; Nakajima, M.; Oyama, T.; Kito, R.; Sugita, C.; Sugita, M.; Kondo, T.; Iwasaki, H. A KaiC-associating SasA-RpaA two-component regulatory system as a major circadian timing mediator in cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 12109–12114.
(10) Iwasaki, H.; Williams, S. B.; Kitayama, Y.; Ishiura, M.; Golden, S. S.; Kondo, T. A KaiC-Interacting Sensory Histidine Kinase, SasA, Necessary to Sustain Robust Circadian Oscillation in Cyanobacteria. *Cell* **2000**, *101*, 223–233.
(11) Markson, J. S.; Piechura, J. R.; Puszynska, A. M.; O'Shea, E. K. Circadian control of global gene expression by the cyanobacterial master regulator RpaA. *Cell* **2013**, *155*, 1396–1408.
(12) Hanaoka, M.; Takai, N.; Hosokawa, N.; Fujiwara, M.; Akimoto, Y.; Kobori, N.; Iwasaki, H.; Kondo, T.; Tanaka, K. RpaB, another response regulator operating circadian clock-dependent transcriptional regulation in Synechococcus elongatus PCC 7942. *J. Biol. Chem.* **2012**, *287*, 26321–26327.
(13) Taniguchi, Y.; Takai, N.; Katayama, M.; Kondo, T.; Oyama, T. Three major output pathways from the KaiABC-based oscillator cooperate to generate robust circadian kaiBC expression in cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 3263–3268.

(14) Schmitz, O. CikA, a Bacteriophytochrome That Resets the Cyanobacterial Circadian Clock. *Science (80-.).* **2000**, *289*, 765–768.
(15) Mackey, S. R.; Choi, J.-S.; Kitayama, Y.; Iwasaki, H.; Dong, G.; Golden, S. S. Proteins found in a CikA interaction assay link the circadian clock, metabolism, and cell division in Synechococcus elongatus. *J. Bacteriol.* **2008**, *190*, 3738–3746.
(16) Rolland, T.; Taşan, M.; Charloteaux, B.; Pevzner, S. J.; Zhong, Q.; Sahni, N.; Yi, S.; Lemmens, I.; Fontanillo, C.; Mosca, R.; et al. A Proteome-Scale Map of the Human Interactome Network. *Cell* **2014**, *159*, 1212–1226.
(17) Dunham, W. H.; Mullin, M.; Gingras, A.-C. Affinity-purification coupled to mass spectrometry: basic principles and strategies. *Proteomics* **2012**, *12*, 1576–1590.
(18) Altelaar, a F. M.; Munoz, J.; Heck, A. J. R. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* **2013**, *14*, 35–48.
(19) Collins, M. O.; Choudhary, J. S. Mapping multiprotein complexes by affinity purification and mass spectrometry. *Curr. Opin. Biotechnol.* **2008**, *19*, 324–330.
(20) Gingras, A.-C.; Gstaiger, M.; Raught, B.; Aebersold, R. Analysis of protein complexes using mass spectrometry. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 645–654.
(21) Binai, N. A.; Marino, F.; Soendergaard, P.; Bache, N.; Mohammed, S.; Heck, A. J. R. Rapid analyses of proteomes and interactomes using an integrated solid-phase extraction-liquid chromatography-MS/MS system. *J. Proteome Res.* **2015**, *14*, 977–985.
(22) Hosp, F.; Scheltema, R. A.; Eberl, C.; Kulak, N. A.; Keilhauer, E. C.; Mayr, K.; Mann, M. A double-barrel LC-MS/MS system to quantify 96 interactomes per day. *Mol. Cell. Proteomics* **2015**.
(23) Butland, G.; Peregrín-Alvarez, J. M.; Li, J.; Yang, W.; Yang, X.; Canadien, V.; Starostine, A.; Richards, D.; Beattie, B.; Krogan, N.; et al. Interaction network containing conserved and essential protein complexes in Escherichia coli. *Nature* **2005**, *433*, 531–537.
(24) Goel, A.; Colcher, D.; Koo, J. S.; Booth, B. J.; Pavlinkova, G.; Batra, S. K. Relative position of the hexahistidine tag effects binding properties of a tumor-associated single-chain Fv construct. *Biochim. Biophys. Acta* **2000**, *1523*, 13–20.
(25) Rumlová, M.; Benedíková, J.; Cubínková, R.; Pichová, I.; Ruml, T. Comparison of Classical and Affinity Purification Techniques of Mason-Pfizer Monkey Virus Capsid Protein: The Alteration of the Product by an Affinity Tag. *Protein Expr. Purif.* **2001**, *23*, 75–83.
(26) Olinares, P. D. B.; Ponnala, L.; van Wijk, K. J. Megadalton complexes in the chloroplast stroma of Arabidopsis thaliana characterized by size exclusion chromatography, mass spectrometry, and hierarchical clustering. *Mol. Cell. Proteomics* **2010**, *9*, 1594–1615.

(27) Kristensen, A. R.; Gsponer, J.; Foster, L. J. A high-throughput approach for measuring temporal changes in the interactome. *Nat. Methods* **2012**, *9*, 907–909.
(28) Kirkwood, K. J.; Ahmad, Y.; Larance, M.; Lamond, A. I. Characterization of native protein complexes and protein isoform variation using size-fractionation-based quantitative proteomics. *Mol. Cell. Proteomics* **2013**, *12*, 3851–3873.
(29) Rippka, R.; Deruelles, J.; Waterbury, J.; Herdman, M.; Stanier, R. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J. Gen. Microbiol.* **1979**, *110*, 1–61.
(30) Gunnelius, L.; Hakkila, K.; Kurkela, J.; Wada, H.; Tyystjärvi, E.; Tyystjärvi, T. The omega subunit of the RNA polymerase core directs transcription efficiency in cyanobacteria. *Nucleic Acids Res.* **2014**, *42*, 4606–4614.
(31) Lee, D. J.; Minchin, S. D.; Busby, S. J. W. Activating transcription in bacteria. *Annu. Rev. Microbiol.* **2012**, *66*, 125–152.
(32) Taylor, B. L.; Zhulin, I. B. PAS Domains: Internal Sensors of Oxygen, Redox Potential, and Light. *Microbiol. Mol. Biol. Rev.* **1999**, *63*, 479–506.
(33) Ivleva, N. B.; Bramlett, M. R.; Lindahl, P. a; Golden, S. S. LdpA: a component of the circadian clock senses redox state of the cell. *EMBO J.* **2005**, *24*, 1202–1210.
(34) Ivleva, N. B.; Gao, T.; LiWang, A. C.; Golden, S. S. Quinone sensing by the circadian input kinase of the cyanobacterial circadian clock. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 17468–17473.
(35) Schmeing, T. M.; Ramakrishnan, V. What recent ribosome structures have revealed about the mechanism of translation. *Nature* **2009**, *461*, 1234–1242.
(36) Gonzalo, P.; Reboud, J.-P. The puzzling lateral flexible stalk of the ribosome. *Biol. Cell* **2003**, *95*, 179–193.
(37) Diaconu, M.; Kothe, U.; Schlünzen, F.; Fischer, N.; Harms, J. M.; Tonevitsky, A. G.; Stark, H.; Rodnina, M. V; Wahl, M. C. Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell* **2005**, *121*, 991–1004.
(38) Battchikova, N.; Aro, E.-M. Cyanobacterial NDH-1 complexes: multiplicity in function and subunit composition. *Physiol. Plant.* **2007**, *131*, 22–32.
(39) Cramer, W. A.; Zhang, H.; Yan, J.; Kurisu, G.; Smith, J. L. Transmembrane traffic in the cytochrome b6f complex. *Annu. Rev. Biochem.* **2006**, *75*, 769–790.
(40) Nickelsen, J.; Rengstl, B. Photosystem II assembly: from cyanobacteria to plants. *Annu. Rev. Plant Biol.* **2013**, *64*, 609–635.
(41) MacColl, R. Cyanobacterial phycobilisomes. *J. Struct. Biol.* **1998**, *124*, 311–334.
(42) Schluchter, W. M.; Bryant, D. A. Molecular characterization of ferredoxin-NADP+ oxidoreductase in cyanobacteria: cloning and sequence of the petH gene of Synechococcus sp. PCC 7002 and studies on the gene product. *Biochemistry* **1992**, *31*, 3092–3102.

(43) Liu, H.; Zhang, H.; Niedzwiedzki, D. M.; Prado, M.; He, G.; Gross, M. L.; Blankenship, R. E. Phycobilisomes supply excitations to both photosystems in a megacomplex in cyanobacteria. *Science* **2013**, *342*, 1104–1107.
(44) Franceschini, A.; Szklarczyk, D.; Frankild, S.; Kuhn, M.; Simonovic, M.; Roth, A.; Lin, J.; Minguez, P.; Bork, P.; von Mering, C.; et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **2013**, *41*, D808–D815.
(45) Gontero, B.; Avilan, L. Creating order out of disorder: structural imprint of GAPDH on CP12. *Structure* **2011**, *19*, 1728–1729.
(46) Kitatani, T.; Nakamura, Y.; Wada, K.; Kinoshita, T.; Tamoi, M.; Shigeoka, S.; Tada, T. Structure of apo-glyceraldehyde-3-phosphate dehydrogenase from Synechococcus PCC7942. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **2006**, *62*, 727–730.
(47) Matsumura, H.; Kai, A.; Maeda, T.; Tamoi, M.; Satoh, A.; Tamura, H.; Hirose, M.; Ogawa, T.; Kizu, N.; Wadano, A.; et al. Structure basis for the regulation of glyceraldehyde-3-phosphate dehydrogenase activity via the intrinsically disordered protein CP12. *Structure* **2011**, *19*, 1846–1854.
(48) Neveling, U.; Bringer-Meyer, S.; Sahm, H. Gene and subunit organization of bacterial pyruvate dehydrogenase complexes. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1998**, *1385*, 367–372.
(49) Engels, A.; Kahmann, U.; Ruppel, H. G.; Pistorius, E. K. Isolation, partial characterization and localization of a dihydrolipoamide dehydrogenase from the cyanobacterium Synechocystis PCC 6803. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1997**, *1340*, 33–44.
(50) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504.
(51) Johnson, C. H.; Egli, M. Metabolic compensation and circadian resilience in prokaryotic cyanobacteria. *Annu. Rev. Biochem.* **2014**, *83*, 221–247.
(52) Qin, X.; Byrne, M.; Xu, Y.; Mori, T.; Johnson, C. H. Coupling of a core post-translational pacemaker to a slave transcription/translation feedback loop in a circadian system. *PLoS Biol.* **2010**, *8*, e1000394.
(53) Pattanayak, G.; Rust, M. J. The cyanobacterial clock and metabolism. *Curr. Opin. Microbiol.* **2014**, *18*, 90–95.

Chapter 4

A first look at *S. elongatus* phosphorylation dynamics through mass spectrometry based proteomics

Ana C.L. Guerreiro[1,2], Marco Benevento[1,2], Harm Post[1,2], Ilka M. Axmann[3], A.F. Maarten Altelaar[1,2] and Albert J.R. Heck[1,2]

[1]Biomolecular Mass Spectrometry and Proteomics Group, Utrecht Institute for Pharmaceutical Sciences and Bijvoet Centre for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands
[2]Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands
[3]Institute for Synthetic Microbiology, Heinrich Heine University Düsseldorf, Universitaetsstrasse 1, D-40225 Düsseldorf, Germany.

Abstract

Protein post-translational modifications (PTMs) are essential for signal transduction and ultimately for cellular function. Phosphorylation in particular is a very specific and common signaling PTM, known to promote enzyme activation or deactivation and even conformational changes in the protein. Signaling mechanisms involving Serine/Threonine/Tyrosine phosphorylation were until recently only associated to eukaryotes, and no phosphorylation or only Histidine phosphorylation was associated to prokaryotes. Nowadays, due to advances in mass spectrometry based proteomics technologies, there are several studies showing Ser/Thr/Tyr phosphorylation in bacteria. In the particular group of cyanobacteria, where a phosphorylation-based clock mechanism subsists, this takes particular importance. Several regulatory connections were already made to the central mechanism, through transcription factor associations. However, considering the cyclic regulation of protein abundances observed previously, which connect the clock to numerous metabolic processes, it is likely that phosphorylation signaling plays part in the regulation process as well. Here we describe a first step towards the global analysis of protein phosphorylation dynamics, to uncover cyclic/circadian variations. Despite our method not being the most suitable, we were able to identify known phosphoproteins and infer on possible phosphorylation dependent regulation in new proteins.

Introduction

Protein post-translational modifications are essential for signal transduction and ultimately for cellular function. They are responsible for the molecular response upon sensing internal and external signals via transfer of information down signaling cascades. Phosphorylation in particular is a very specific and common signaling post-translational modification (PTM), which occurs on serine (Ser), threonine (Thr) and tyrosine (Tyr) amino acid residues (1), and less frequently on histidine (His), arginine (Arg) and lysine (Lys) residues (2). This reaction is known to promote enzyme activation or deactivation and even conformational changes in the protein, which can mediate activity regulation or protein-protein interactions (3).

Signaling mechanisms involving Ser/Thr/Tyr phosphorylation were thought to only occur in highly complex organisms such as eukaryotes, until 1979 when Garnak and Reeves (4) identified the first phosphoprotein in prokaryotes.

Additionally, histidine phosphorylation has been primarily associated with bacterial signal transduction, being involved in two-component sensory pathways. Nowadays, due to advances in mass spectrometry based proteomics technologies, Ser/Thr/Tyr phosphorylation was shown to occur in prokaryotes (5–8). In cyanobacteria, a large group of photoautotrophic prokaryotes, phosphorylation is particularly important, since it is also part of the central clock mechanism regulating circadian rhythms in cellular function. The *S. elongatus'* clock mechanism in particular is based on the near-24h phosphorylation oscillation cycle of the protein KaiC. This phosphorylation/dephosphorylation is regulated by the consecutive association/dissociation of proteins KaiA and KaiB. This mechanism can be reproduced in vitro with the three clock protein and ATP/Mg^{2+}, due the auto-phosphorylation capability of KaiC. The processes downstream and upstream to the clock are not fully characterized, however several protein effectors and input factors were already connected to it. Among them are the histidine kinase SasA (*Synechococcus* adaptive sensor), the master regulator RpaA (regulator of phycobilisome-associated A/two component transcriptional regulator) (9–11), the RpaB regulator (12), LabA (low amplitude and bright A) or the sensor histidine kinase CikA (circadian input kinase A) (13, 14); and the input factors Pex (PadR family transcriptional regulator) and LdpA (light-dependent period A). Our previous observations relate cyclic protein abundance oscillations (15) and protein assembly LD variations (Chapter 3) with different cellular processes, such as transcription, carbohydrate metabolism and photosynthesis. Therefore, we argued that additional downstream and upstream pathways might be connected to the clock through phosphorylation signaling and established a proteomics approach to investigate this.

Initial studies in cyanobacterial protein phosphorylation have demonstrated its function in sensing illumination changes and nutrient supply. These studies employed a more targeted approach focusing on the signal transduction protein P-II, involved in the nitrogen signaling pathway (16), on phycobiliproteins and phycobilisome linker proteins (17, 18), or on the circadian clock protein KaiC (19). Recently, global phosphoproteomics studies were performed in species such as *Synechococcus sp* PCC7002 (20) and *Synechocystis* (21, 22), including qualitative or quantitative approaches. In the first, 245 phosphoproteins were identified using a gel-free qualitative approach with TiO_2 phosphopeptide enrichment; while in the second, a dimethyl labeling quantitative approach was used together with the TiO_2,

enabling the differential quantification of 148 phosphorylation events in response to nitrogen starvation.

A phosphoproteomics study of this dimension has not yet been performed for the cyanobacteria *S. elongatus*, as well as a global study of the cyclic changes in protein phosphorylation as consequence of circadian regulation. This could contribute to the investigation of a possible connection of the central clock, which is governed by cyclic phosphorylation of KaiC, to different regulatory pathways. Here, we took the first step towards the study of protein phosphorylation dynamics in light/dark (LD) conditions, to ultimately uncover potential cyclic/circadian variations, using *Synechococcus elongatus* PCC 7942 as a cyanobacterial model. We implemented a mass spectrometry based approach and tested the combination of known quantitative, fractionation and enrichment proteomics strategies, namely 6-plex tandem mass tags (TMT), strong cation exchange (SCX) and Ti^{4+}-IMAC, respectively. We were able to identify known and new phosphorylated proteins and sites of several abundant proteins. However, the combination of these proteomics approaches proved to be less promising to convey the phosphorylation dynamics of this organism.

Results and Discussion

This report comprises a preliminary assessment for the study of the cyclic/circadian dynamics of protein phosphorylation in cyanobacteria. The combination of strong cation exchange (SCX) with Ti^{4+}-IMAC were the methods of choice for phosphopeptide fractionation and enrichment, constituting a known successful association (23). TMT 6-plex was chosen as a quantitative method due to its multichannel advantage (Figure 1a). The SCX approach enabled the separation of the peptides with different charges, where singly charged and neutral peptides, corresponding to singly or doubly phosphorylated peptides, respectively, elute in the beginning of the chromatogram (Figure 1b).

In order to reduce subsequent sample handling and increase reproducibility, the SCX fractions with lower chromatographic intensity, and therefore with lower peptide abundance, were combined upon peptide desalting. The amount of Ti^{4+}-IMAC material used in the enrichment step was also determined by the estimated amount of peptides in the corresponding fractions.

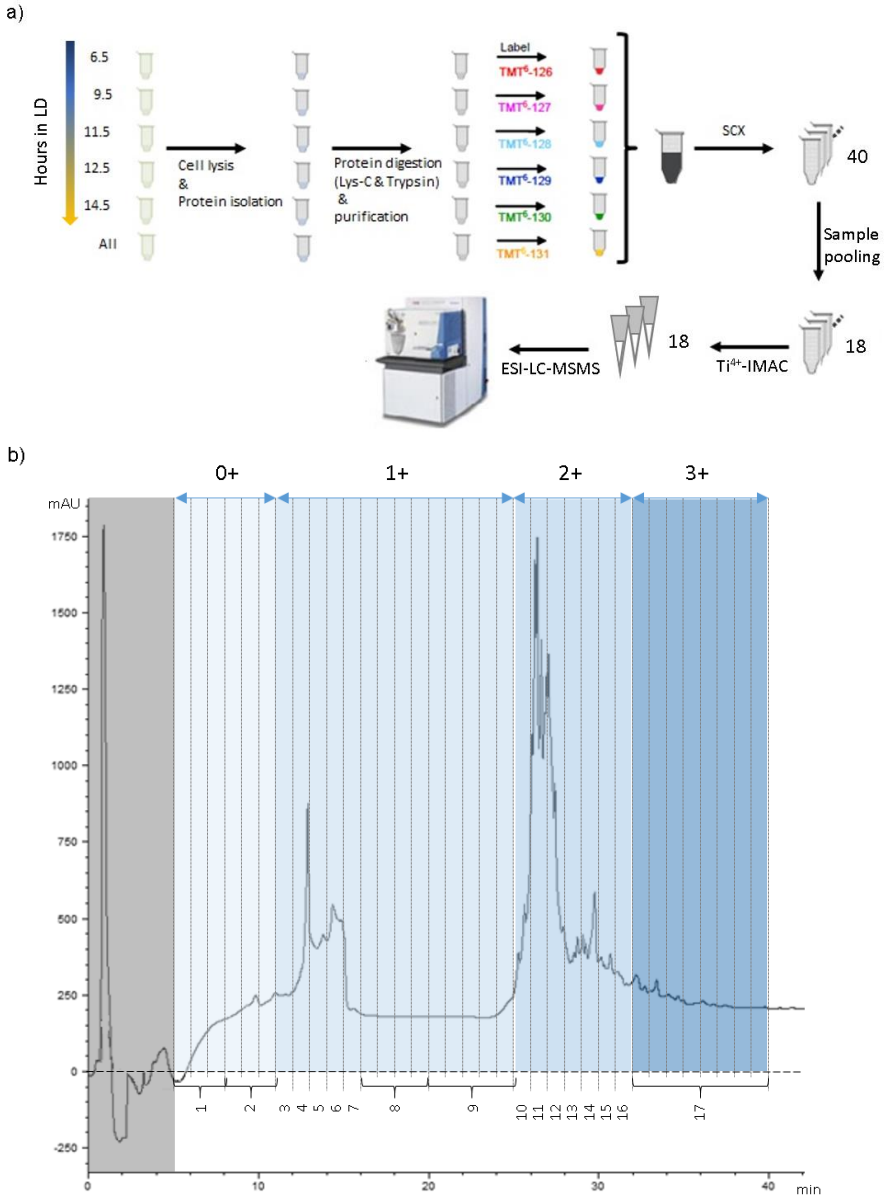


Figure 1 - Experimental workflow and SCX chromatogram. a) Samples were collected from *S. elongatus* cultures synchronized to LD cycles in five time-points. Following the protein isolation and digestion, the different samples were labeled with TMT 6-plex and mixed prior to SCX fractionation. Some of the fractions collected were pooled and subjected to Ti^{4+}-IMAC enrichment of phosphopeptides. Samples were analyzed

in a LTQ-Orbitrap Velos, with HCD activation. b) Peptide elution profile concerning the SCX separation, with the representation of each charge elution area and the pooling scheme of the fractions. Gray area represents the void.

With this approach we were able to identify 10 phosphoproteins and 15 phosphopeptides with high confidence, which include some of the most abundant cyanobacteria proteins and previous identified phosphoproteins (Table 1). We also identified 10 other phosphoproteins, including putative and uncharacterized proteins, which contained either only one peptide-to-spectrum match (PSM) per phosphopeptide or uncertain phosphosite assignments, making the identifications less reliable. These results represent lower identifications than expected, nonetheless we investigated the phosphorylation events identified, which we describe bellow. Still, to assess possible reasons to justify the low number of identifications, we performed a survey of the kinases presently characterized in *S. elongatus*, within the Uniprot database, to estimate on protein phosphorylation. The survey indicated the presence of 28 protein kinases, including histidine, serine, threonine and tyrosine (STY) kinases. Considering this, it can be assumed that the TMT labeling and Ti^{4+}-IMAC enrichment are not compatible for the study of phosphopeptide dynamics, or there were excessive steps employed in this enrichment protocol, leading to sample loss. In fact, recent studies combined these two techniques but in a reversed order, where labeling was performed after the enrichment, which appeared to be successful (24). Adding to this, we obtained similar or slightly better results in a preliminary measurement with no quantitative information, before the enrichment step. Since this test was performed on a different instrument, it cannot be used as a final proof of the previous assumption. Nevertheless, in this analysis we obtained a different set of phosphorylated proteins and peptides, which in its majority were represented by only one PSM. These proteins are involved in amino acid, pyruvate, peptidoglycan, porphyrin and chlorophyll metabolism, photosynthesis and homologous recombination. Only three peptides and their corresponding uncharacterized proteins were identified with more than two PSMs, namely BioY protein, ABC-transporter membrane fusion protein and a putative protein.

Among the known phosphoproteins, we identified and quantified the intensively studied nitrogen regulatory protein P-II (Figure 3). This signal transduction protein was shown to be involved in the cellular response to

nitrogen supply, where P-II is phosphorylated at Ser49 upon nitrogen starvation (16). Here, we see a differential phosphorylation on Ser49, identified and quantified on three independent phosphopeptides, which shows dependence on illumination. In the dark there is a slow decrease in protein phosphorylation, as perceived by the reduction of the phosphopeptides abundances; the phosphorylation increases again in the light phase. The protein abundance was determined from the same analysis as given by the quantitative software (Proteome discoverer), having a similar trend to the abundance determined in the global proteomics study described in chapter 2 (15). We observe that the protein level decreases less and its increase, which starts still in the dark phase, does not correlate initially with the phosphopeptides abundance. This light dependent change in phosphorylation was already suggested, although it was considered to be an indirect response, since it would most probably be due to residual nitrogen assimilation and the arrest of CO_2 fixation in darkness (16).

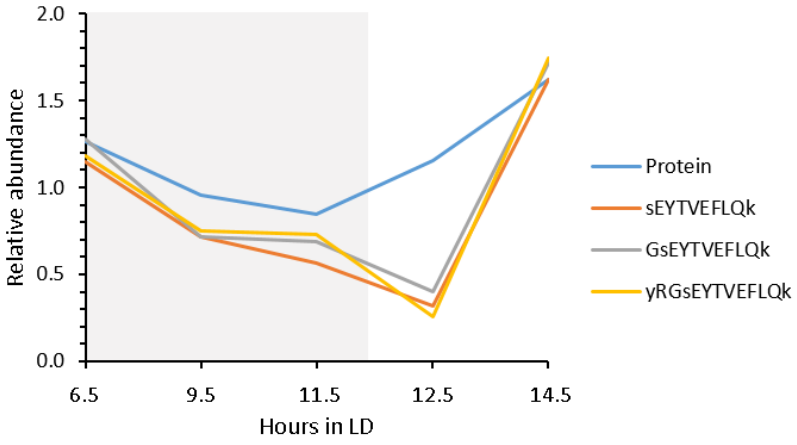


Figure 3 – Nitrogen regulatory protein P-II abundance and phosphorylation profile. Abundance of the protein without phosphorylation is represented in blue and the remaining colors correspond to the three phosphorylated peptides identified. Serine residue in lower case represents the phosphorylation site Ser49.

In addition, we identified phosphorylated proteins involved in photosynthesis (antenna and photosystem I), carbohydrate metabolism and pentose phosphate pathway (6-phosphogluconolactonase), purine and pyrimidine metabolism (nucleoside diphosphate kinase), which in some cases have been

identified as phosphoproteins before (Table 1). These proteins didn't show dynamic phosphorylation in our study, thus they seem to be independent of environmental light related changes. Phosphorylation of the photosynthetic proteins was previously observed, first in plants (25, 26) and recently in cyanobacteria (20). Here, we observed phosphorylation of Ser and Thr residues in antenna proteins (phycobilisomes), which are responsible for photo-energy absorption; and in the photosystem I, which performs the final stage of photosynthesis. The phosphorylation of phycobilisome proteins was considered to play a role in the regulation of energy distribution between the two photosystems in plants, through conformational changes of the antenna proteins and their binding to the different photosystems (27, 28). The phosphorylation of the phycobilisome linker in particular is related to the regulation of phycobilisome assembly/disassembly and remodeling/turnover in cyanobacteria (18, 29). These proteins were also identified in our previous global proteomics studies. In our quantitative daily rhythm work (described in chapter 2) these proteins were not found to have a significant abundance in the 48h time-course experiment, nonetheless we observed a change in abundance between light and dark conditions. This was also reflected in the abundance variations of protein assemblies in the work described in chapter 3, which might connect abundance and complex formation regulation to phosphorylation signaling.

Enzymes from other metabolic pathways were also found phosphorylated, namely 6-phosphogluconolactonase (Lactonase), from the pentose phosphate pathway, and Thioredoxin-1, which is involved in the regulation of redox homeostasis. The first was phosphorylated at Thr45 which was not observed before. However, this protein is acetylated in *E. coli* (30) and another enzyme from the pentose phosphate pathway (6-phosphogluconate dehydrogenase) is phosphorylated in the *Synechocystis* sp. PCC 6803 (21), which can imply a possible activity regulation associated to this modification. The thioredoxin is associated to oxidative stress response and is also responsible for the activation of several Calvin cycle enzymes (31). In the latter, thioredoxin is usually activated through its reduction, which is mediated by ferredoxin, with no phosphorylation involved. Consequently, its phosphorylation at Ser2 may play a part in other type of reaction or pathway. In Yeast, one of the thioredoxins isoforms was also found to be phosphorylated, however so far no function was attributed (32).

Table 1 – Identified phosphoproteins and corresponding phosphopeptides. * Found in cyanobacteria.

Accessions	Protein name	Phosphosite	Site Probability (%)	PSMs
Q31RG1	Allophycocyanin, beta subunit*	S21	100	4
		S139	100	1
Q31RG0	Allophycocyanin alpha chain*	S46*	100	1
		S46*	100	1
		S18	100	1
		Y78/80	50/50	1
Q31QV8	6-phosphogluconolactonase*	T45	100	11
Q31QV2	Photosystem I iron-sulfur center	T74	99.1	5
Q31QF8	RNA-binding region RNP-1	S88	100	1
Q31PZ5	Putative uncharacterized protein	-	20	5
Q31PS2	Putative uncharacterized protein	S5/T7	50	3
Q31PI7	Photosystem I reaction center subunit II*	S114	100	1
Q31PD9	Phycobilisome rod linker polypeptide*	S283	98.9	3
Q31PD8	Phycocyanin linker protein 9K	S5	100	1
Q31NL5	DEAD/DEAH box helicase-like	S306	100	2
Q31NA5	Putative uncharacterized protein	S68	99.3	1
Q31M25	Putative uncharacterized protein	-	33.3	4
Q31LS0	Transcriptional regulator AbrB	S25	100	1
P50590	Nucleoside diphosphate kinase*	Y49*	100	3
		T91	100	1
P33171	Elongation factor Tu	T252	100	1
P13530	C-phycocyanin alpha chain*	S147	100	2
		S37	100	2
		S15	100	1
P12243	Thioredoxin-1*	S2	100	3
P0A3F4	Nitrogen regulatory protein P-II*	S49*	100	1
		S49*	100	1
		S49*	100	1
		S49*	99.5	4
P06539	C-phycocyanin beta chain*	S144	100.0	1
		S50/51	50.0	2

The nucleoside diphosphate kinase was also phosphorylated in our data. This protein is a major component in the synthesis of nucleoside triphosphates, other than ATP, and it auto-phosphorylates during an intermediate step. In *E. coli* this transient phosphorylation occurs at a histidine residue, however here the phosphorylated residue was at Tyr49, which is in agreement with a recent study in *Synechocyctis* sp. PCC 6803 (21). This phosphosite could also be associated with an intermediary state or with a regulation of this protein's activity. The proteins mentioned above didn't show significant protein abundance oscillations in our global quantitative proteomics study (described in chapter 2) (15). Nevertheless, their phosphorylation might instead be related to their direct function and not to abundance regulation.

We didn't observe histidine phosphorylation, which should be quite abundant in bacteria. This type of protein phosphorylation is particularly related to two-component signaling pathways, whose constituent proteins we also weren't able to identify. However, it was shown that histidine phosphorylation identification is not compatible with standard proteomics workflow conditions, since this type of modification is very labile in acidic conditions (33).

The clock protein KaiC was also not present in our phosphoprotein identification list. This is due to the fact that the peptide containing the supposed phosphorylated residues assumes a very long stretch when using trypsin for the digestion protocol. An alternative would be the use of another enzyme, such as AspN, which cleaves at the N-terminal side of aspartic and cysteic acid residues, enabling the size reduction of the phosphorylated peptide.

This preliminary study used well known proteomics techniques to uncover the phosphoprotein dynamics in *S. elongatus*, in different illumination conditions. However, the current combination Ti^{4+}-IMAC enrichment after TMT 6-plex labeling turned out to be not compatible. Nevertheless, the results showed some promise, since we were able to identify, among the most abundant proteins, known phosphoproteins and infer on possible phosphorylation dependent regulation in new proteins. This is an encouragement for the continuation of phosphoproteomics studies in *S. elongatus*, to uncover daily or circadian rhythms.

Methods

1. Cyanobacteria cell culture

The wild-type strain of *Synechococcus elongatus* PCC 7942 was routinely grown photoautotrophically in BG11-medium (34) at 30°C under continuous illumination with white light of 80 µmol of photons/m^2s (Versatile Environmental Test Chamber, SANYO) and a continuous stream of air. Cell concentrations were measured by determining the optical densities of the culture at 750 nm (OD750) (SPECORD®200 PLUS, Analytik Jena). The culture was kept in log growth phase (up to an OD750 of 1.0) by dilution up to a specific volume and transferred to 12:12-hours light/dark cycle (LD) for 3 days. Synchronized culture was finally diluted to an OD750 of approximately 0.4 one day before the sampling started. At certain time points (Figure 1a), which varied from one to three hour intervals, 40 ml of the culture was centrifuged at 15,000 x g for 10 min and the supernatant was removed. Cell pellet was resuspended in 1 ml BG11-medium, centrifuged again for 5 min. Supernatant was removed and pellet washed with 1 ml PBS buffer before last round of centrifugation. Cell pellets were frozen in liquid nitrogen prior to storage at -20°C.

2. Sample preparation

Cyanobacteria pellets from samples of 5 time-points were lysed in 8 M Urea, 50 mM triethyl ammonium bicarbonate (TEAB), containing 1 tablet EDTA-free protease inhibitor cocktail (Sigma) and 1 tablet PhosSTOP phosphatase inhibitor cocktail (Roche). After three sonication cycles at 4 ^{0}C, total cell lysates were obtained through centrifugation at 14,000 rpm for 30 min, at 4 ^{0}C. The supernatant was recovered and protein concentration was determined with the Bradford method (BioRad). The proteins were subjected to reduction and alkylation of the cysteine residues, using 200 mM dithiotreitol (DTT) (Sigma) and 200 mM iodoacetamide (Sigma). The proteins were first digested with Lys-C (Roche Diagnostics, Ingelheim, Germany), at an enzyme:protein ratio of 1:75, for 4 hours at 37 ^{0}C, followed by 4 times dilution of the samples with 50 mM TEAB and digestion with trypsin (Roche Diagnostics, Ingelheim, Germany), at an enzyme:protein ratio of 1:100, overnight at 37 ^{0}C. 100 µg from each sample and from a mixture of all samples were desalted using 1cc Sep Pack C18 columns (Waters) and dried in vacuo.

3. Peptide labeling

Peptides were labeled with tandem mass tags (TMT) using the TMT 6plex labeling kit (Pierce). The labeling consisted of the use of 5 tags, one for each time-point, and the 6th tag for the mixture of all time-points (internal standard). The manufacturer's protocol was followed, with a few adjustments. After desalting, 100 µg of peptides per channel were dissolved in 100 µl 200 mM TEAB. The TMT labeling reagents were dissolved in 40 µL acetonitrile (ACN) (Biosolve) per vial and added to the samples in two steps, to maximize the labeling efficiency. First, 20 µl of reagent solution was added to the sample, after 5 min, the other 20 µl were added and the reaction was incubated for 1 hour at room temperature. In the quenching step, 4 µL of 5% hydroxylamine were added. After 15 min, the six channels were mixed in a 1:1 ratio and stored at −20 °C.

4. Strong Cation eXchange (SCX) fractionation

Peptides were fractionated using strong-cation exchange (SCX) as described previously (23). In short, four SCXs were performed using a Polysulfoethyl A column (200 x 2.1mm) (PolyLC). Solvent A consisted of 5 mM KH_2PO_4, 0.05% formic acid and 30% acetonitrile (pH 2.7), while solvent B consisted of 5 mM KH_2PO_4, 0.05% formic acid, 30% acetonitrile and 350 mM KCl (pH 2.7). The following gradient was used: 0–10 min (100% A); 10–15 min (0–26% B); 15–40 min (26-35% B); 40–45 min (35–60% B); 45–49 min (60–100% B); 49–57 min (100%B); 57–65 min (100% A). The fractions were collected every minute for 40 minutes. The resulting 40 fractions were dried *in vacuo*, resuspended in 10% formic acid and stored at −20 °C.

5. Phosphopeptide enrichment

Prior to enrichment, the fractions obtained from SCX fractionation were desalted and pooled together to obtain 18 samples with similar amount of peptides. Ti^{4+}-IMAC material was prepared and used as previously described previously (35). Briefly, the Ti^{4+}-IMAC beads (400 or 500 µg of beads/200 µl pipette tip) were loaded onto GELoader tips (Eppendorf, Hamburg, Germany) using a C8 plug. To reduce variations during the enrichment process, the spin tip enrichment was performed in-parallel. The Ti^{4+}-IMAC columns were conditioned using 30 µl of loading buffer consisting of 80% acetonitrile (ACN)/6% trifluoroacetic acid (TFA) and centrifugation at 600-1500 rpm for 2-

4 min. The samples were dissolved in 80% ACN/6% TFA, transferred to the spin tips in two-steps and centrifuged at 600-800 rpm for a total of 40 min. Then, the columns were sequentially washed with 30 µl of washing buffer 1 (50% ACN, 0.5% TFA containing 200 mm NaCl) and 30 µl of washing buffer 2 (50% ACN/0.1% TFA), being centrifuged at 800-1500 rpm for 1-3 min. The bound peptides were eluted into a new tube (already containing 35 µl of 10% formic acid) with 20 µl of 10% ammonia by centrifugation at 1000 rpm for 20 min. A final elution was performed with 2 µl of 80% ACN/2% formic acid for 10 min. The collected eluate was further acidified by the addition of 3 µl of 100% formic acid and stored at -20°C prior to nano-LC-MS/MS analysis.

6. Liquid Chromatography-tandem Mass Spectrometry (LC-MS/MS)

Peptide fractions were analyzed on an Orbitrap Velos (Thermo Fisher Scientific, Bremen) that was coupled to an Agilent 1200 HPLC system (Agilent Technologies, Waldbronn). Both the trap (20 mm×100 µm I.D.) and the analytical column (35 cm×50 µm I.D.) were packed in-house using Reprosil-pur C18 3 µm (Dr. Maisch). Peptides were trapped at 5 µL/min in 100% solvent A (0.1 M acetic acid (Merck)). Elution was achieved with the solvent B (0.1 M acetic acid in 80% acetonitrile) at 100 nL/min. The 180 minutes gradient used was as follows: 0-10 min, 100% solvent A; 10.1-115 min, 10–24% solvent B; 115-160 min, 24–50% solvent B; 160-163 min, 50-100% solvent B; 163-164 min, 100% solvent B; 164-165 min, 0-100% solvent A; 165-180 min, 100% solvent A. Nanospray was achieved using a coated fused silica emitter (New Objective, Cambridge, MA) (o.d. 360 µm; i.d. 20 µm, tip i.d. 10 µm) biased to 1.7 kV. The mass spectrometer was operated in the data dependent acquisition mode. A MS2 method was used with a FT survey scan from 350 to 1500 m/z (resolution 30,000; AGC target 5E5). The 10 most intense peaks were subjected to HCD fragmentation (resolution 7500; AGC target 3E4, NCE 45%, max. injection time 500 ms, dynamic exclusion 60 s). Predictive AGC was enabled.

7. Database search and validation

Raw data was converted to .mgf file type with Proteome Discoverer (version 1.3, Thermo). Mascot (version 2.3.02, Matrix science) was used to search the MS/MS data against the *S. elongatus* PCC 7942 Uniprot database (version 4-2010) including a list of common contaminants and concatenated with the reversed versions of all sequences (5,826 sequences). Trypsin was chosen

as cleavage specificity allowing two missed cleavages. Carbamidomethylation (C), TMT6plex (K) and TMT6plex (N-term) were set as fixed modifications. The variable modifications used were oxidation (M), phosphorylation (STY and H). The database searches were performed using a peptide tolerance of 50 ppm and a fragment mass tolerance of 0.05 Da (HCD). The 50-ppm mass window was chosen to allow random assignment of false positives that were later removed by filtering using the instruments actual mass accuracy (10 ppm). Quantification was performed using centroid peak intensity with the 'reporter ions quantifier' node. For all experiments only unique peptides were considered for protein quantification. To achieve a false discovery rate < 1% the following filters were applied: high confidence; peptide length 6–35; peptide score > 20; maximum search engine rank 1; and peptide mass deviation 10 ppm.

References

1. Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006) Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell* 127, 635–648
2. Moritz, A., Li, Y., Guo, A., Villen, J., Wang, Y., MacNeill, J., Kornhauser, J., Sprott, K., Zhou, J., Possemato, A., Ren, J. M., Hornbeck, P., Cantley, L. C., Gygi, S. P., Rush, J., and Comb, M. J. (2010) Akt-RSK-S6 Kinase Signaling Networks Activated by Oncogenic Receptor Tyrosine Kinases. *Sci. Signal.* 3, ra64–ra64
3. Matthews, H. R. (1995) Protein kinases and phosphatases that act on histidine, lysine, or arginine residues in eukaryotic proteins: a possible regulator of the mitogen-activated protein kinase cascade. *Pharmacol. Ther.* 67, 323–50
4. Garnak, M., and Reeves, H. C. (1979) Phosphorylation of Isocitrate dehydrogenase of Escherichia coli. *Science* 203, 1111–1112
5. Deutscher, J., and Saier, M. H. (2005) Ser/Thr/Tyr protein phosphorylation in bacteria - for long time neglected, now well established. *J. Mol. Microbiol. Biotechnol.* 9, 125–31
6. Eymann, C., Becher, D., Bernhardt, J., Gronau, K., Klutzny, A., and Hecker, M. (2007) Dynamics of protein phosphorylation on Ser/Thr/Tyr in Bacillus subtilis. *Proteomics* 7, 3509–3526
7. Macek, B., Gnad, F., Soufi, B., Kumar, C., Olsen, J. V, Mijakovic, I., and Mann, M. (2008) Phosphoproteome analysis of E. coli reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. *Mol. Cell. Proteomics* 7, 299–307
8. Schmidl, S. R., Gronau, K., Pietack, N., Hecker, M., Becher, D., and Stülke, J. (2010) The phosphoproteome of the minimal bacterium Mycoplasma

pneumoniae: analysis of the complete known Ser/Thr kinome suggests the existence of novel kinases. *Mol. Cell. Proteomics* 9, 1228–1242

9. Takai, N., Nakajima, M., Oyama, T., Kito, R., Sugita, C., Sugita, M., Kondo, T., and Iwasaki, H. (2006) A KaiC-associating SasA-RpaA two-component regulatory system as a major circadian timing mediator in cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* 103, 12109–14
10. Iwasaki, H., Williams, S. B., Kitayama, Y., Ishiura, M., Golden, S. S., and Kondo, T. (2000) A KaiC-Interacting Sensory Histidine Kinase, SasA, Necessary to Sustain Robust Circadian Oscillation in Cyanobacteria. *Cell* 101, 223–233
11. Markson, J. S., Piechura, J. R., Puszynska, A. M., and O'Shea, E. K. (2013) Circadian control of global gene expression by the cyanobacterial master regulator RpaA. *Cell* 155, 1396–408
12. Hanaoka, M., Takai, N., Hosokawa, N., Fujiwara, M., Akimoto, Y., Kobori, N., Iwasaki, H., Kondo, T., and Tanaka, K. (2012) RpaB, another response regulator operating circadian clock-dependent transcriptional regulation in Synechococcus elongatus PCC 7942. *J. Biol. Chem.* 287, 26321–7
13. Taniguchi, Y., Takai, N., Katayama, M., Kondo, T., and Oyama, T. (2010) Three major output pathways from the KaiABC-based oscillator cooperate to generate robust circadian kaiBC expression in cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* 107, 3263–8
14. Schmitz, O. (2000) CikA, a Bacteriophytochrome That Resets the Cyanobacterial Circadian Clock. *Science (80-.).* 289, 765–768
15. Guerreiro, A. C. L., Benevento, M., Lehmann, R., van Breukelen, B., Post, H., Giansanti, P., Maarten Altelaar, a F., Axmann, I. M., and Heck, A. J. R. (2014) Daily rhythms in the cyanobacterium synechococcus elongatus probed by high-resolution mass spectrometry-based proteomics reveals a small defined set of cyclic proteins. *Mol. Cell. Proteomics* 13, 2042–55
16. Forchhammer, K., and Tandeau de Marsac, N. (1994) The PII protein in the cyanobacterium Synechococcus sp. strain PCC 7942 is modified by serine phosphorylation and signals the cellular N-status. *J. Bacteriol.* 176, 84–91
17. Harrison, M. A., Tsinoremas, N. F., and Allen, J. F. (1991) Cyanobacterial thylakoid membrane proteins are reversibly phosphorylated under plastoquinone-reducing conditions in vitro. *FEBS Lett.* 282, 295–9
18. Piven, I., Ajlani, G., and Sokolenko, A. (2005) Phycobilisome linker proteins are phosphorylated in Synechocystis sp. PCC 6803. *J. Biol. Chem.* 280, 21667–72
19. Nishiwaki, T., Satomi, Y., Nakajima, M., Lee, C., Kiyohara, R., Kageyama, H., Kitayama, Y., Temamoto, M., Yamaguchi, A., Hijikata, A., Go, M., Iwasaki, H., Takao, T., and Kondo, T. (2004) Role of KaiC phosphorylation in the circadian clock system of Synechococcus elongatus PCC 7942. *Proc. Natl. Acad. Sci. U. S. A.* 101, 13927–13932
20. Yang, M., Qiao, Z., Zhang, W., Xiong, Q., Zhang, J., Li, T., Ge, F., and Zhao, J. (2013) Global phosphoproteomic analysis reveals diverse functions of serine/threonine/tyrosine phosphorylation in the model cyanobacterium Synechococcus sp. strain PCC 7002. *J. Proteome Res.* 12, 1909–23

21. Spät, P., Maček, B., and Forchhammer, K. (2015) Phosphoproteome of the cyanobacterium Synechocystis sp. PCC 6803 and its dynamics during nitrogen starvation. *Front. Microbiol.* 6, 248
22. Mikkat, S., Fulda, S., and Hagemann, M. (2014) A 2D gel electrophoresis-based snapshot of the phosphoproteome in the cyanobacterium Synechocystis sp. strain PCC 6803. *Microbiology* 160, 296–306
23. Zhou, H., Low, T. Y., Hennrich, M. L., van der Toorn, H., Schwend, T., Zou, H., Mohammed, S., and Heck, A. J. R. (2011) Enhancing the identification of phosphopeptides from putative basophilic kinase substrates using Ti (IV) based IMAC enrichment. *Mol. Cell. Proteomics* 10, M110.006452
24. Erickson, B. K., Jedrychowski, M. P., McAlister, G. C., Everley, R. A., Kunz, R., and Gygi, S. P. (2015) Evaluating multiplexed quantitative phosphopeptide analysis on a hybrid quadrupole mass filter/linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* 87, 1241–9
25. Vener, A. V., Ohad, I., and Andersson, B. (1998) Protein phosphorylation and redox sensing in chloroplast thylakoids. *Curr. Opin. Plant Biol.* 1, 217–223
26. Hansson, M., and Vener, A. V (2003) Identification of three previously unknown in vivo protein phosphorylation sites in thylakoid membranes of Arabidopsis thaliana. *Mol. Cell. Proteomics* 2, 550–559
27. Haldrup, A., Jensen, P. E., Lunde, C., and Scheller, H. V. (2001) Balance of power: A view of the mechanism of photosynthetic state transitions. *Trends Plant Sci.* 6, 301–305
28. Allen, J. F. (2003) Botany. State transitions--a question of balance. *Science* 299, 1530–1532
29. Murata, N. (1969) Control of excitation transfer in photosynthesis. I. Light-induced change of chlorophyll a fluorescence in Porphyridium cruentum. *Biochim. Biophys. Acta* 172, 242–251
30. Zhang, J., Sprung, R., Pei, J., Tan, X., Kim, S., Zhu, H., Liu, C.-F., Grishin, N. V, and Zhao, Y. (2009) Lysine acetylation is a highly abundant and evolutionarily conserved modification in Escherichia coli. *Mol. Cell. Proteomics* 8, 215–25
31. Berg, J. M., Tymoczko, J. L., and Stryer, L. (2002) *Biochemistry* (W. H. Freeman and Company, England). 5th Ed.
32. Albuquerque, C. P., Smolka, M. B., Payne, S. H., Bafna, V., Eng, J., and Zhou, H. (2008) A multidimensional chromatography technology for in-depth phosphoproteome analysis. *Mol. Cell. Proteomics* 7, 1389–1396
33. McAllister, T. E., Hollins, J. J., and Webb, M. E. (2013) Prospects for stable analogues of phosphohistidine. *Biochem. Soc. Trans.* 41, 1072–7
34. Rippka, R., Deruelles, J., Waterbury, J., Herdman, M., and Stanier, R. (1979) Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J. Gen. Microbiol.* 110, 1–61
35. Zhou, H., Ye, M., Dong, J., Corradini, E., Cristobal, A., Heck, A. J. R., Zou, H., and Mohammed, S. (2013) Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. *Nat. Protoc.* 8, 461–80

Chapter 5

Outlook

In a postgenomic era, proteomics now appears as an established and reliable field for the analysis of proteins in biological systems, constituting an important step to functionally understand and integrate biological processes. In this thesis I describe a series of global mass spectrometry (MS)-based quantitative proteomics studies, demonstrating some of its applications. These include, global inquiry of proteins or protein complexes simultaneously, as well as their post-translational modifications (PTMs), in this case in cyanobacterium. The continuity of quantitative MS-based techniques depends on its evolution and development. Thus, I will now discuss the present and future trends in this field, which I believe will continue to add biological information in many fields of study, including the biological processes involved in cyanobacteria.

In MS-based quantitative proteomics, there is not one ideal workflow, so in order to choose the right techniques for each step, one has to consider different aspects: the complexity and dynamic range of the organism; the type of experiment (is it a global and/or time-series) and adequate data analysis.

The complexity and dynamic range of the organism's proteome will influence its coverage. The steps preceding and including MS analysis will influence this aspect. Cyanobacteria, as prokaryotes, are among the organisms with the smallest genome; nevertheless pre-fractionation techniques are necessary to facilitate better proteome coverage. Nowadays, it is common practice to use two-dimensional liquid chromatography (LC) techniques for pre-fractionation. These can include different combinations, as mentioned in the first chapter of this thesis. In this work and many others, offline set-up of strong cation exchange (SCX) was chosen as a first dimension prior to reversed phase (RP)-LC. The use of ultra-high pressure (UHP) LC-based separation will additionally contribute to an increased sensitivity and accuracy, through the increase of separation power (1–3). A potentially interesting option would be the use of a chip-based LC separation set-up (4–6). This is an emerging technique which can be suitable for non-expert users because of its practicality. However, it still needs much improvement, in terms of cost-efficiency and MS-compatibility. The type of MS instrumentation should also be chosen accordingly. For a global proteome study of the sort presented here, a high-resolution and high accuracy instrument should be used, and high sensitivity and speed should be the goal. Instruments such as the Orbitrap and Q-TOF can nowadays characterize around 5,000-10,000 proteins, in human cells (7, 8). A direct extrapolation of these numbers for bacteria, would translate into 100% proteome coverage, if it wasn't for the

dynamic range issue. The recently introduced Orbitrap Fusion™ Tribrid™ (9), with quadrupole, linear ion trap and orbitrap analyzers, results in increased proteome identifications and coverage. Its increased acquisition rate enabled the identification of nearly all proteins in Yeast. This could be extrapolated to cyanobacteria and constitute an potential increase in proteome coverage, but in this particular case one has to consider also the highly abundant photosynthesis-associated proteins, which can still conceal the low abundant proteins (10, 11). Sample or cellular fractionation strategies, or even depletion approaches, might still be necessary to implement in this situation. Despite the instrumental advances, a data-dependent type of acquisition (DDA) will always be limited by the random selection of the most abundant peptide-ions for sequencing. An alternative is a data-independent acquisition (DIA) mode, which allows for the sequencing of theoretically all ions in a parallel fashion, potentially leading to full coverage (12, 13). The challenge here is the specificity and quantification accuracy, since it depends on the perfect alignment between the peptide and fragment elution peaks, which is still problematic for low abundant peptides in complex biological backgrounds.

To choose the right quantification technique, the goal has to be defined. Besides considering basic aspects like accuracy and selectivity, one should also think of the multiplexing aspect. For instance, in the case described in the second chapter of this thesis there was a time-series of 20 points. In this case tandem mass tags (TMT) were used, which allowed multiplexing of 6 samples. These isobaric tag methods, such as iTRAQ, have been usually and successfully applied to investigate proteome responses to different conditions simultaneously. At the time of our experiments we had to perform 4 experiments to cover the total of 20 time-points. Nowadays, the different label-tags can go up to 8-plex (iTRAQ) and 10-plex (TMT) (14), which reduces the manual labor. The downside of this technique is the inconvenience of the interference issue, responsible for a decrease in the quantification accuracy, due to the co-isolation of unwanted peptide ions for fragmentation and consequently quantification. Another issue to consider is the size of the label, which can influence a proper peptide ion fragmentation. An alternative to this could be the metabolic labeling, which ensures better coverage, but only for triplex experiments (15). On the other hand, cyanobacteria in particular present an additional difficulty, since several species obtain the N and C through fixing pathways. This means that one would have to control the gas and liquid fluxes in the cultures, becoming an expensive experiment. Luckily, some species do not perform N_2 fixing, making the approach more simple. For

the other species, obtaining a full coverage of the proteome dynamics might mean taking a label-free approach. This technique might be generally considered as less accurate, but has seen many developments lately and introduces many advantages for multiplexed and global proteomics experiments. Here, in theory, all that is confidently identified can be quantified, provided that you optimize the analysis conditions. Additionally, if we consider the use of DIA type of set-up mentioned above, a faster analysis time than DDA can be obtained, with a reproducibility and accuracy that has the potential to be similar to selected reaction monitoring (SRM) type of analysis (highly sensitive MS targeted approach discussed in Chapter 1) (16). An intensity based approach (centered on peptide ion intensities as measured by MS) can deliver better overall performance than spectral count (based on the number of peptide ions observed for each protein) (17, 18); provided that you have a high resolution and accuracy instrument, a robust LC set-up with stable elution and the right software, with peak alignment incorporated (to match the eluting peptides from different experiments). It's also important to find the right balance between accurate quantification, and better protein coverage. The first relies on MS (peptide ion) scan number and the second on MS/MS (fragment ion) scan number, and increasing one sacrifices the other. A so called accurate mass and retention time tag approach can be applied (19, 20), with a set of two experiments, one with focus on obtaining better identification and another on better quantification. Nowadays, there are already several commercially available, as well as free-licensed, software packages able to automatically deal with this type of data, namely MaxQuant (21), Progenesis and Mascot distiller.

After an enormous amount of data generated, an adequate data analysis software should be considered. The experimental design should also fit from the beginning to the right statistical analysis of significance, for instance the amount of replicates is crucial especially for label-free approaches. For circadian studies in particular, besides analyzing significant changes in abundance one has to analyze significant cyclic abundances, to identify true circadian oscillations. In transcriptomics this area is quite developed and is supported by numerous strategies. Common examples are the COSOPT (22) and Fisher's G-test (23) algorithms, which match the abundance profiles with variable cosine curves or search for periodicity, respectively. For protein analysis the available strategies are scarcer, depending on individual development and preferences. Only recently, Perseus software (from MaxQuant package), has integrated a circadian analyzer. This is based also

on cosine curve matching and includes FDR calculation, which is a good start and an incentive for circadian proteomics studies.

Another important aspect of data analysis is the proper extraction of biological significance. This is partly determined by the ease of data analysis, but also the ease of complementary data accessibility, with both fields in need of improvements. In particular for cyanobacteria, there is lack of information concerning protein PTM and functional characterization, and interaction. Many advances have already been made for analyzing the data generated in a fairly automated fashion, through software like Perseus and tools and databases such as Cytoscape (24), String (25), David (26) and Uniprot (http://www.uniprot.org). On the other hand, appropriate data dissemination is also important, since it will contribute for proper function assignment of uncharacterized proteins, for example. Certain initiatives have moved forward in order to storage data in different formats, such as PRIDE (27) and ProteomeXchange (http://www.proteomexchange.org). All of these databases at different ends of data-flow have made and are still making continuous efforts to improve data accessibility.

Ultimately, to properly understand cellular regulation as a whole, it being circadian responses or other processes, one has to combine different levels of information. The integration of data from the different “omics” fields is increasing and new fields such as Proteogenomics have been created to focus on the complementarity of information between those. Proteogenomics in particular has been used to improve the annotation of the genome, since protein evidence makes the ultimate proof of new splicing variants and new protein-coding genes (28). Now that we can cover a substantial part of the proteome, development is still needed to improve the bioinformatics tools, which will facilitate this systems-biology view. Initiatives, such as Circadiomics (29), which groups metabolomics and transcriptomics data for the human cell, could also be adopted for the cyanobacteria field.

In summary, the continuous improvements in mass spectrometry based-proteomics will keep adding to the cyanobacterial biological knowledge in the future. Considering this, it’s important to adopt a specific workflow to each type of study. For example, a label-free experiment is very useful for the comparison of many different conditions or in time-series experiments; a DIA type of set-up can be applied to a discovery study, but also to a validation study. Indeed, due to the many methods and workflows available, MS-based proteomics can be very versatile and be applied to different types of

experiments. A particular example, related to the work developed in this thesis, would be to use targeted type of experiments, such as SRM or DIA, to follow-up on observed cyclic protein abundances. In conjunction with immunoprecipitation strategies, it could be used to validate the newly presented protein-protein interactions, as well as link interesting cyclic proteins to their regulators. Moreover, it can be a useful addition to functional studies for the investigation of the uncharacterized proteins found in the global studies. It can also be applied in the investigation of the mechanisms behind post-transcriptional, translational and post-translational regulation, through the comparison of transcript abundance, newly synthetized and global protein abundances. MS could also be used to investigate PTM cross-talk and explore what other PTMs might be involved in circadian regulation, using for example different phosphorylation, acetylation, or methylation enrichment strategies. MS could contribute as well to the validation of such observations and likewise to results obtained by complementary technologies such as genomics or transcriptomics. The application of one or more MS improvements across the experimental workflow together with the established and new bioinformatic tools, which can extract and integrate several degrees of data, will significant aid cyanobacterial studies, being quantitative or qualitative studies.

References

1. Hyung, S.-W., Kim, M.-S., Mun, D.-G., Lee, H., and Lee, S.-W. (2011) The effect and potential of using a temperature controlled separation column with ultra-high pressure microcapillary liquid chromatography/tandem mass spectrometry on proteomic analysis. *Analyst* 136, 2100–5
2. Motoyama, A., Venable, J. D., Ruse, C. I., and Yates, J. R. (2006) Automated ultra-high-pressure multidimensional protein identification technology (UHP-MudPIT) for improved peptide identification of proteomic samples. *Anal. Chem.* 78, 5109–5118
3. Nagaraj, N., Kulak, N. A., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O., Vorm, O., and Mann, M. (2012) System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Mol. Cell. Proteomics* 11, M111.013722
4. Lee, J., Soper, S. A., and Murray, K. K. (2009) Microfluidic chips for mass spectrometry-based proteomics. *J. Mass Spectrom.* 44, 579–93
5. Mohammed, S., Kraiczek, K., Pinkse, M. W. H., Lemeer, S., Benschop, J. J., and Heck, A. J. R. (2008) Chip-Based Enrichment and NanoLC-MS/MS Analysis of Phosphopeptides from Whole Lysates. *J. Proteome Res.* 7, 1565–71

6. Vollmer, M., Hörth, P., Rozing, G., Couté, Y., Grimm, R., Hochstrasser, D., and Sanchez, J.-C. (2006) Multi-dimensional HPLC/MS of the nucleolar proteome using HPLC-chip/MS. *J. Sep. Sci.* 29, 499–509
7. Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Pääbo, S., and Mann, M. (2011) Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* 7, 548
8. Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A., Herzog, F., Rinner, O., Ellenberg, J., and Aebersold, R. (2011) The quantitative proteome of a human cell line. *Mol. Syst. Biol.* 7, 549
9. Senko, M. W., Remes, P. M., Canterbury, J. D., Mathur, R., Song, Q., Eliuk, S. M., Mullen, C., Earley, L., Hardman, M., Blethrow, J. D., Bui, H., Specht, A., Lange, O., Denisov, E., Makarov, A., Horning, S., and Zabrouskov, V. (2013) Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identi fi cation Rates.
10. Barrios-Llerena, M. E., Chong, P. K., Gan, C. S., Snijders, A. P. L., Reardon, K. F., and Wright, P. C. (2006) Shotgun proteomics of cyanobacteria--applications of experimental and data-mining techniques. *Brief. Funct. Genomic. Proteomic.* 5, 121–32
11. Gan, C. S., Reardon, K. F., and Wright, P. C. (2005) Comparison of protein and peptide prefractionation methods for the shotgun proteomic analysis of Synechocystis sp. PCC 6803. *Proteomics* 5, 2468–2478
12. Egertson, J. D., Kuehn, A., Merrihew, G. E., Bateman, N. W., MacLean, B. X., Ting, Y. S., Canterbury, J. D., Marsh, D. M., Kellmann, M., Zabrouskov, V., Wu, C. C., and MacCoss, M. J. (2013) Multiplexed MS/MS for improved data-independent acquisition. *Nat. Methods* 10, 744–6
13. Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012) Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* 11, O111.016717
14. Pichler, P., Ko, T., Holzmann, J., Mazanek, M., Taus, T., Ammerer, G., and Mechtler, K. (2010) Peptide Labeling with Isobaric Tags Yields Higher Identification Rates Using iTRAQ 4-Plex Compared to TMT 6-Plex and iTRAQ 8-Plex on LTQ Orbitrap. *Anal. Chem.* 82, 6549–6558
15. Ong, S.-E. (2002) Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteomics* 1, 376–386
16. Gallien, S., Duriez, E., and Domon, B. (2011) Selected reaction monitoring applied to proteomics. *J. Mass Spectrom.* 46, 298–312
17. Grossmann, J., Roschitzki, B., Panse, C., Fortes, C., Barkow-Oesterreicher, S., Rutishauser, D., and Schlapbach, R. (2010) Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods. *J. Proteomics* 73, 1740–1746
18. Choi, H., Glatter, T., Gstaiger, M., and Nesvizhskii, A. I. (2012) SAINT-MS1: protein-protein interaction scoring using label-free intensity data in affinity purification-mass spectrometry experiments. *J. Proteome Res.* 11, 2619–24

19. Monroe, M. E., Tolić, N., Jaitly, N., Shaw, J. L., Adkins, J. N., and Smith, R. D. (2007) VIPER: an advanced software package to support high-throughput LC-MS peptide identification. *Bioinformatics* 23, 2021–3
20. Fang, R., Elias, D. A., Monroe, M. E., Shen, Y., McIntosh, M., Wang, P., Goddard, C. D., Callister, S. J., Moore, R. J., Gorby, Y. A., Adkins, J. N., Fredrickson, J. K., Lipton, M. S., and Smith, R. D. (2006) Differential label-free quantitative proteomic analysis of Shewanella oneidensis cultured under aerobic and suboxic conditions by accurate mass and time tag approach. *Mol. Cell. Proteomics* 5, 714–25
21. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–72
22. Straume, M. (2004) DNA microarray time series analysis: automated statistical assessment of circadian rhythms in gene expression patterning. *Methods Enzymol.* 383, 149–66
23. Wichert, S., Fokianos, K., and Strimmer, K. (2004) Identifying periodically expressed transcripts in microarray time series data. *Bioinformatics* 20, 5–20
24. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–504
25. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., and von Mering, C. (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–52
26. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57
27. Vizcaino, J., Vizcaíno, J. A., Côté, R. G., Csordas, A., Dianes, J. a, Fabregat, A., Foster, J. M., Griss, J., Alpi, E., Birim, M., Contell, J., O'Kelly, G., Schoenegger, A., Ovelleiro, D., Pérez-Riverol, Y., Reisinger, F., Ríos, D., Wang, R., and Hermjakob, H. (2013) The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* 41, D1063–D1069
28. Brosch, M., Saunders, G. I., Frankish, A., Collins, M. O., Yu, L., Wright, J., Verstraten, R., Adams, D. J., Harrow, J., Choudhary, J. S., and Hubbard, T. (2011) Shotgun proteomics aids discovery of novel protein-coding genes, alternative splicing, and "resurrected" pseudogenes in the mouse genome. *Genome Res.* 21, 756–67
29. Patel, V. R., Eckel-Mahan, K., Sassone-Corsi, P., and Baldi, P. (2012) CircadiOmics: integrating circadian genomics, transcriptomics, proteomics and metabolomics. *Nat. Methods* 9, 772–3

Chapter 6

Summary
Samenvatting
Curriculum Vitae
List of publications
Acknowledgements

Summary

The major aim of the work presented in this thesis was to provide new insights for the molecular mechanisms of the *S. elongatus* circadian clock in particular, using quantitative proteomics and phosphoproteomics.

The **first** chapter is composed of two parts, which introduce the technical and biological aspects of this thesis. The first part addresses mass spectrometry and quantitative proteomics techniques, including the ones used in the course of this thesis. The second part describes what is known so far about cyanobacterial circadian rhythm, focusing on different targeted and global investigations.

The **second** chapter describes the large-scale analysis of *S. elongatus* proteome dynamics, which constitutes one of the most complete proteome analysis to date. The relative protein abundance of hundreds of proteins was monitored over 48 h under light and dark cycles (12:12 h), through the analysis of 20 samples. Multidimensional separation, TMT 6-plex isobaric labeling and high-resolution mass spectrometry approaches enabled this investigation. The analysis of the proteins showing a cyclic abundance profile revealed 77 proteins with 24-h period and an even smaller group with 12-h oscillations in abundance. This work also represents some of the clearest evidence for post-transcriptional regulation of many of the processes involved in the circadian rhythm and the clear need to analyze these processes at the protein level.

The **third** chapter describes the global analysis of macromolecular assemblies' dynamics. For this purpose, a label-free MS centered approach was conjugated with the pre-fractionation of protein assemblies by Size Exclusion Chromatography (SEC). To unveil *S. elongatus* adaptation between light and dark conditions, the samples were collected at the height of the day, and the depth of the night. Following several well-known complexes such as the RNA polymerase, the ribosome and complexes involved in photosynthesis, we observe for the first time that these complexes change not only in abundance, but also in constituency upon illumination variations. This work shows that the diurnal adaptation can occur at the level of higher organization of protein complexes, and partially explains why this 'light-dependent' bacterium shows less rhythmic oscillations at the protein level.

The **forth** chapter represents a first step towards the ultimate investigation of cyclic/circadian variations in phosphoproteins abundances. To study protein phosphorylation dynamics in light/dark (LD) conditions, *S. elongatus* cultures were grown in 24-h LD cycles, similarly to the experiments described in the second chapter. A mass spectrometry based approach was implemented, the 6-plex tandem mass tags (TMT) method was chosen for quantitation and Ti^{4+}-IMAC was chosen for phosphopeptide enrichment. This study showed that the chosen combination of quantitative proteomics and phosphopeptide enrichment was less suited to reveal the global phosphorylation dynamics of this organism. Still several known and novel phosphorylated proteins and sites were identified and their possible meaning is discussed.

The **fifth** chapter comprises an outlook addressing the applicability of quantitative mass spectrometry for the study of the cyanobacterial proteome. New trends and advances in mass spectrometry (MS) instrumentation, as well as in pre-MS methods and post-MS analysis are presented and discussed. Applications of these methods in the context of cyanobacterial studies are suggested, demonstrating that MS-based approaches can positively contribute for the advancement of biological knowledge in this species.

Samenvatting

Het werk in dit proefschrift beschrijft uitgebreide kwantitatieve proteomics en fosfoproteomics studies met als doel het verkrijgen van nieuwe inzichten in de moleculaire mechanismen achter het circadiaanse ritme van *S. elongatus*.

Het **eerste** hoofdstuk is samengesteld uit twee delen, waarin de technische en biologische aspecten van dit proefschrift worden geïntroduceerd. Het eerste deel beschrijft massa spectrometrie (MS) en kwantitatieve proteomics technieken in het algemeen, alsmede de specifieke technieken gebruikt in dit proefschrift. Het tweede deel verdiept zich in de bestaande kennis van het circadiaanse ritme van cyanobacterium met een focus op specifieke en globale studies die hiervoor zijn uitgevoerd.

In het **tweede** hoofdstuk wordt de analyse van de dynamica van het *S. elongatus* proteoom bescheven onder invloed van licht en donker. Deze studie vormt op dit moment de meest complete proteoom analyse van *S. elongatus*, waarbij de relatieve abundantie van honderden eiwitten is gevolgd over 48 uur onder licht en donker cycli (12:12 uur). Met een combinatie van multidimensionale scheiding, TMT 6-plex isobaric labeling en hoge resolutie MS, zijn 20 verschillende tijdspunten gemeten. Analyse van eiwitten met een cyclisch expressie profiel, leidde tot de vondst van 77 eiwitten met 24-uurs oscillaties en een kleinere groep met oscillaties van 12 uur. Verder laat deze studie een duidelijk post-transcriptionele regulatie zien van de verschillende ritmische processen en dus de noodzaak dit soort studies op het eiwit niveau uit te voeren.

Het **derde** hoofdstuk beschrijft de studie naar de dynamiek in eiwit complexen vorming onder licht en donker condities. Om te bestuderen hoe *S. elongatus* zich aan lichte en donkere condities aanpast werden er op het midden van de dag en midden in de nacht samples genomen. Deze samples zijn onder milde condities geprepareerd en vervolgens geanalyseerd met een combinatie van pre-fractionering van de eiwitcomplexen d.m.v. Size Exclusion Chromatography (SEC) en kwantitatieve massa spectrometrie. Van een aantal bekende complexen als RNA polymerase, het ribosoom en verschillende complexen met een functie in fotosynthese, zien we dat deze complexen niet alleen veranderen in expressie, maar ook in samenstelling. In dit werk hebben we laten zien dat de proteoom veranderingen van *S.*

elongatus gedurende de verschillende fasen van de dag vooral gereguleerd worden op het niveau van eiwit complexen en in mindere mate op eiwit expressie niveau.

Het **vierde** hoofdstuk gaat in op de regulatie van ritmische moleculaire processen door fosforylering van eiwitten. Hiervoor hebben we S. elongatus gegroeid in 24 cycli van dag en nacht condities. zoals beschreven in hoofdstuk twee. Vervolgens hebben we de samples met 6-plex tandem mass tags (TMT) gelabeld voor kwantificatie en de fosfopeptiden verrijkt met behulp van Ti^{4+}-IMAC. Deze studie toonde aan dat deze combinatie van kwantitatieve proteomics en de verrijking van fosfopeptiden minder geschikt was om de globale dynamica van fosforylering te bestuderen.
Desalniettemin hebben we in deze studie verschillende gefosforyleerde eiwitten kunnen identificeren, waarvan een aantal niet eerder zijn gerapporteerd. In dit hoofdstuk worden de mogelijke functies van deze fosforyleringen beschreven.

Het **vijfde** hoofdstuk geeft een vooruitzicht van de mogelijke toepassingen van kwantitatieve MS voor de studie van dynamische processen in het proteoom, met de nadruk op het cyanobacteria proteoom. Nieuwe methoden, voor zowel massa spectrometrie als voorbewerkingsmethoden en data analyse worden bediscussieerd in dit hoofdstuk.
Mogelijke toepassingen van deze behandelde technieken voor studies naar verschillende facetten van cyanobacteria worden gesuggereerd, waarbij duidelijk wordt dat massa spectrometrie nog veel kan bij dragen aan de biologische kennis over dit organisme.

Curriculum Vitae

I received a Bachelor of Science degree in Biochemistry from Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa in June 2008. Here, I received training in various scientific areas allied with Biochemistry. I continued my academic path obtaining a Master degree in Medical Biochemistry from Faculdade de Ciências, Universidade de Lisboa in November 2010. During the research that led to a Master thesis, I started to get familiar with Proteomics and gained knowledge in related equipment, such as Mass Spectrometry. While my interest in mass spectrometry-based proteomics increased, I was invited to integrate the Biomolecular Mass Spectrometry and Proteomics group, at the University of Utrecht, as a PhD candidate. During the work supervised by Prof. Dr. Albert Heck and Dr. Maarten Altelaar, I had the opportunity to expand my knowledge and experience in Proteomics research based on quantitative mass spectrometry techniques, from sample preparation to mass spectrometry analysis. My research topics focused on the application of quantitative mass spectrometry methodologies for the study of cyanobacterial light/dark dependent rhythms.

List of publications

Da Costa, G., **Guerreiro, A.**, Correia, C. F., Gomes, R. J., Freire, A., Monteiro, E., Barroso, E., Coelho, A. V, Outeiro, T. F., Freire, A. P., and Cordeiro, C. (2010) A non-invasive method based on saliva to characterize transthyretin in familial amyloidotic polyneuropathy patients using FT-ICR high-resolution MS. *Proteomics Clin. Appl.* 4, 674–678

Da Costa, G., Gomes, R. A., **Guerreiro, A.**, Mateus, É., Monteiro, E., Barroso, E., Coelho, A. V, Freire, A. P., and Cordeiro, C. (2011) Beyond Genetic Factors in Familial Amyloidotic Polyneuropathy: Protein Glycation and the Loss of Fibrinogen's Chaperone Activity. *PLoS One* 6, e24850

Guerreiro, A.*, da Costa, G.*, Gomes, R. A., Ribeiro-Silva, C., Gilberto, S., Mateus, E., Monteiro, E., Barroso, E., Coelho, A. V, Ponces Freire, A., and Cordeiro, C. (2012) α-Synuclein aggregation in the saliva of familial transthyretin amyloidosis: a potential biomarker. *Amyloid* 19, 74–80. *equal contribution

Guerreiro, A.C.L., Benevento, M., Lehmann, R., van Breukelen, B., Post, H., Giansanti, P., Maarten Altelaar, A.F., Axmann, I.M., and Heck, A.J.R. (2014) Daily rhythms in the cyanobacterium synechococcus elongatus probed by high-resolution mass spectrometry-based proteomics reveals a small defined set of cyclic proteins. *Mol. Cell. Proteomics* 13, 2042–55

Guerreiro, A.C.L., Penning, R., Raaijmakers, L., Axman, I.M., Heck, A.J.R. and Altelaar, A.F.M. Monitoring light/dark association dynamics of multiprotein complexes in cyanobacteria using size exclusion chromatography-based proteomics. Submitted to the *Journal of Proteomics*

Acknowledgements

My story in the Netherlands starts more than 4 years ago thanks to an invitation from Albert. I would like to thank you first, for giving me this amazing and unique opportunity, which was doing a PhD in your group. Since the beginning, you believed in me and trusted that I would be able to finish my work, even when I wasn't sure.

Maarten is certainly the second person to thank to. You guided me in every step of the way until the finish line. You were always able to maintain a relaxed environment when difficulties appeared and to promote my independence. I would like to thank you for your time and patience; for your teachings in the different forms of science, from experimental to the writing.

This thesis would also not exist without collaborations and for that I would like to thank Ilka, as well as Anne Rediger, Stefanie Hertel and Anika Wiegard for their essential contributions and for welcoming me in their lab for a short while.

During these years of PhD I was surrounded by wonderful people who helped me dealing with my work and also with the fact that I was in a foreign country. So, I would like to thank all the other previous and present lab members, whom without I wouldn't have survived or have such a wonderful time. Here are a few reasons why I would like thank you…

Harm, for setting up the terrain for my work, and together with Soenita, Mirjam and Arjan, for teaching me the first steps in the lab and making me feel that any of my wishes were possible. Marco H and Teck, for welcoming me in their office. Together with all my office mates, Basak, Sara C., Chiara, Nicolas, Violette, whit whom I shared great moments of stress and joy. Serena for introducing me for the first time to Dutch bureaucracy. Shabaz for sharing his immense knowledge in mass spectrometry. Corine and Martje for always being available for helping with paperwork and for making sure that there was always an International lunch, Labuitje and Sinterkerst dinner party! Geert, for keeping my PC in shape. Bas, Henk and Linsey for their very important bioinformatics and statistical advices. Jamila, for the translation help (for which I should also thank all the other Dutch people! ☺). Renske for your precious help in many things, it's not by chance you're a co-author. Marco B. for teaming up with me on OR3. Luc, for the interesting (sometimes philosophical ☺) discussions we had during our common project. The nice and friendly lab managers, Nadine and Gianluca, or should I say the fun and

party colleagues. Eleonora and Sara, for motivating me to go to the gym and welcoming me to the Expresso Club. A club for coffee making and socializing (many times in Italian ☺), with prestige members such as, Anna R., Angela, Andrea, Andrey, Simone, Domenico… My Strasbourg companions: the *gourmet* and brave guide, Lucrèce; the party guy, Fabio(lon); and Clement(ino), the bar *connoisseur* and chef. It was great fun, spending time with you there and in Utrecht in many dinners, barbeques… Matina, you appeared later, but your joy and spontaneity immediately and continuously, contaminated me and cheered me up. David, fueste mi alumno y ahora mi amigo. Lástima que las sesiones de baile hayan empezado tarde. Alba, my fellow Iberian, por sermos vecinas sólo nos podríamos haber dado bien! Compartí muchos buenos momentos contigo… Y espero seguir compartiendo.
Besides, but also including, all the above there were other people with whom I shared fun moments during the 'Friday beers' and table football sessions, or lunch and other fun events, which helped me relax: Piero and Christian, Erik and Pepijn, Letícia (team CPLP! ☺), Ayse, Joost, Dona, Mao, Suzy, Yang Yang and Liana (the mama's),Thin Thin, 'Fun' Fan, Anja, Philip and Michiel, Martin, Aneika, Wei, Anita, Séverine, Celine, Thierry, Esther, Kyle, Vojtech, Celia, Vincent, Guambo, Anna H., Richard… Although I didn't spend much time with some of you, I definitely have lots of nice and fun memories, which I thank you for.

O caminho até aos Países Baixos começa muito tempo antes e foi uma viagem alucinante que não dependeu só de mim. Agradeço em primeiro lugar aos meus pais, Clara e Joaquim, pelo seu apoio constante desde o início da minha carreira académica, que começou aos 6 anos! É realmente por uma série de eventos iniciados e perpetuados por vocês, que aqui cheguei. Agradeço também o facto de me terem apoiado na decisão mais recente que tomei e que me levou a esta aventura, mesmo sabendo que as saudades e as preocupações iam ser muitas. Agradeço à minha irmã Rita, que é única, no sentido estrito e lato. Por todos os momentos que passámos antes e depois da nossa separação, pela energia positiva que me transmitias e fazias emitir, enfim, pela tua irmandade incondicional. Apesar de esta aventura ter sido um dos melhores momentos da minha vida, foi difícil estar longe de vocês…
Agradeço ao Miguel, meu companheiro, meu amigo. Sei que não deve ter sido fácil apoiar a minha decisão de ir para longe de ti, bem como termos ficado separados por 1 ano; e sei isto porque também não foi fácil para mim.

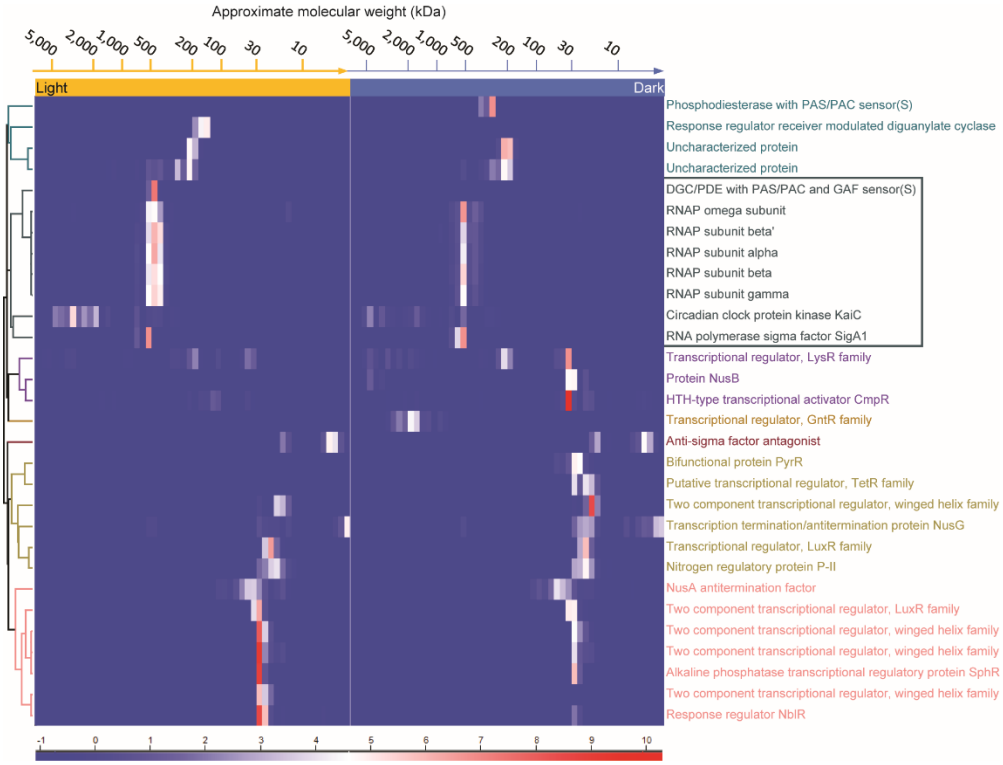
Agradeço o facto de teres estado sempre presente, quer longe, quer perto de mim nos últimos anos. O teu apoio e amor ajudaram a manter-me sã.
Agradeço também à minha tia e madrinha, Conceição, pelas dicas e ensinamentos, no início e final desta fase. Foram muito úteis e até possivelmente determinantes.

Dank je wel! Thank you! Vielen Dank! Grazie! ¡Gracias! Merci! Eskerrik asko! Σας ευχαριστούμε! Спасибо!Teşekkür ederiz! 谢谢大家！Terima kasih! ขอขอบคุณ! Děkuji vám! Obrigada!

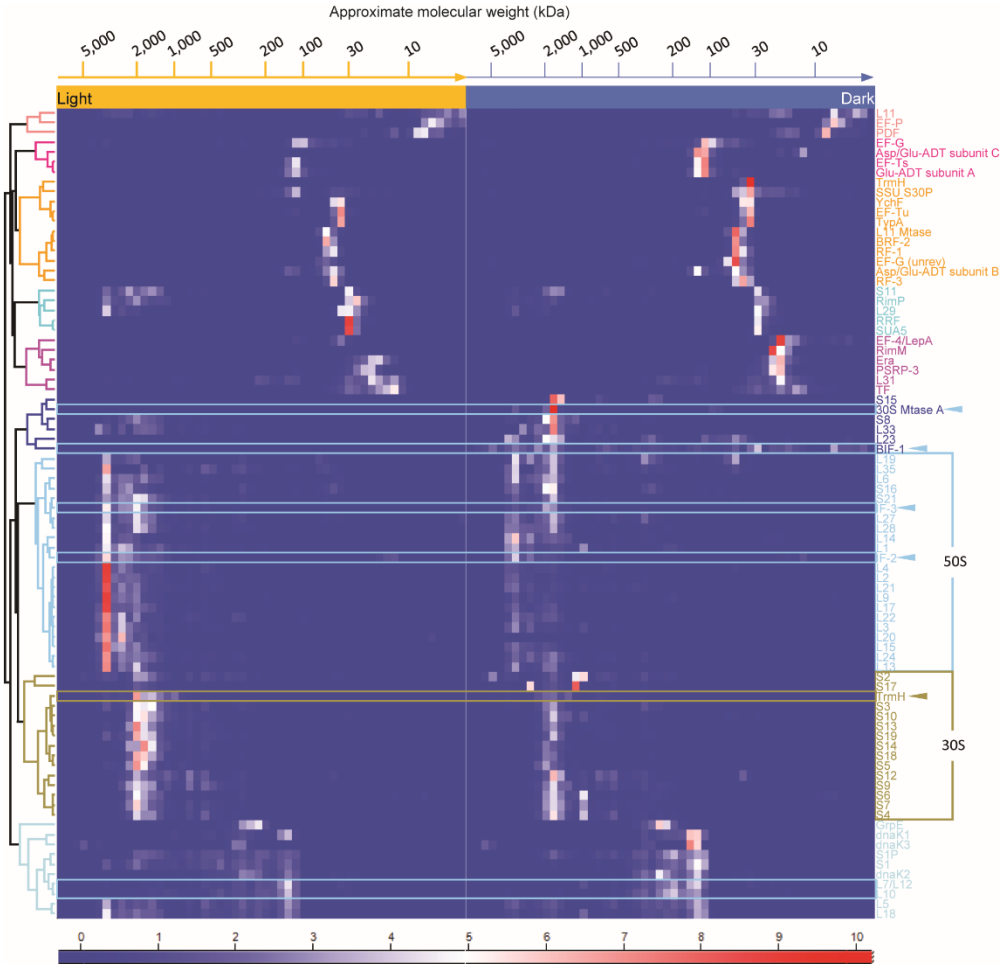
Chapter 7

Appendix

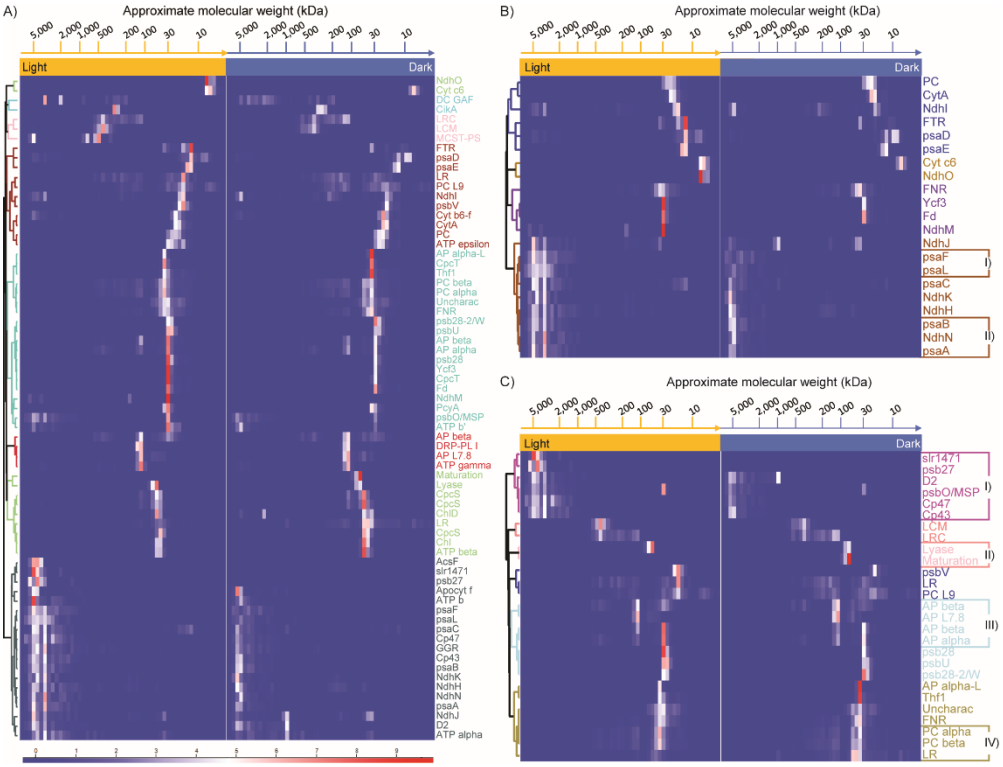
Supplementary figures to Chapter 3



Supplementary figure 1 - Hierarchical clustering of proteins involved in transcription. Heatmap depicting unsupervised hierarchical clustering taking the standardized (Z-score) LD profiles in parallel, using Perseus software, as described in the methods section. Red corresponds to higher protein abundance and blue to lower abundance. The RNA polymerase enriched cluster is highlighted with a light-blue box.



Supplementary figure 2 – Hierarchical clustering of ribosomal and translation related proteins. Heatmap depicting unsupervised hierarchical clustering taking the standardized (Z-score) LD profiles in parallel, using Perseus software, as described in the methods section. Red corresponds to higher protein abundance and blue to lower abundance. The 50S and 30S enriched clusters are highlighted with a light-blue or brown box, respectively. Interesting protein elution profiles are highlighted as well.



Supplementary figure 3 - Hierarchical clustering of proteins from the photosystem and phycobilisome complexes. Heatmaps depicting unsupervised hierarchical clustering taking the standardized (Z-score) LD profiles in parallel, using Perseus software, as described in the methods section. Red corresponds to higher protein abundance, blue to lower abundance and white to zero. A) All photosynthesis related proteins. B) Proteins from the photosystem I (PSI) and NAD(P)H-quinoneoxireductase complexes (NDH-1); highlighted SEC LD profiles: I) PSI interface and II) PSI core. C) Proteins from the photosystem II (PSII) and phycobilisome complexes; highlighted SEC LD profiles: I) PSII core proteins, II) Phycobilisome rod proteins (PC), III) Phycobilisome core proteins (AP) and IV) Chromophorylation related proteins.