Protein complexes in chlorophyll biosynthetic enzymes

Peterson Wulff, Ragna

2011

Protein complexes in chlorophyll biosynthetic enzymes

LUNDS
UNIVERSITET

Ragna Peterson Wulff

Opponent:

Assoc. Prof. Thomas Kieselbach, Umeå University

Committee:

Prof. Sara Linse, Lund University

Prof. Anette Henriksen, Novo Nordisk A/S, Denmark

Assoc. Prof. Andreas Blennow, University of Copenhagen

Dept. of Biochemistry and Structural Biology
Lund University
P.O. Box 124
S-221 00 Lund, Sweden
ISBN 978-91-7422-288-3

Printed in Sweden by Media-Tryck, Lund University
Lund 2011

Preface

List of Papers

I. **ATP-induced conformational dynamics in the AAA+ motor unit of magnesium chelatase**
Lundqvist J, Elmlund H, Peterson Wulff R, Berglund L, Elmlund D, Emanuelsson C, Hebert H, Willows R, Hansson M, Lindahl M and Al-Karadaghi S
Structure (18) 354-365 (2010)

II. **Magnesium chelatase subunit BchH interacts with the ATPase inactive side of the BchID AAA+ complex of *Rhodobacter capsulatus***
Peterson Wulff R, Lundqvist J, Axelsson E, Sirijovski N, Al-Karadaghi S, Emanuelsson C and Hansson M
Manuscript

III. **NADPH-dependent thioredoxin reductase and 2-Cys peroxiredoxins are needed for the protection of Mg-protoporphyrin monomethyl ester cyclase**
Stenbaek A, Hansson A, Peterson Wulff R, Hansson M, Dietz KJ and Jensen PE
FEBS Letters 582 (18) 2773-2778 (2008)

IV. **The activity of barley NADPH-dependent thioredoxin reductase C is independent of the oligomeric state of the protein: Tetrameric structure determined by cryo-electron microscopy**
Peterson Wulff R, Lundqvist J, Rutsdottir G, Hansson A, Stenbaek A, Elmlund D, Elmlund E, Jensen PE and Hansson M
Biochemistry (50) 3713-3723 (2011)

Additional Papers

Aqueous two-phase partitioning for proteomic monitoring of cell surface biomarkers in human peripheral blood mononuclear cells
Everberg H, Peterson R, Rak S, Tjerneld F, Emanuelsson C
Journal of Proteome Research 5 (5), 1168-1175 (2006)

Enhanced enzymatic conversion of softwood lignocellulose by poly(ethylene glycol) addition
Börjesson J, Peterson R and Tjerneld F
Enzyme and Microbial Technology (40), 754-762 (2007)

A new method for isolating physiologically active Mg-protoporphyrin monomethyl ester, the substrate of the cyclase enzyme of the chlorophyll biosynthetic pathway
Gough SP, Rzeznicka K, Peterson Wulff R, da Cruz Francisco J, Hansson A, Jensen PE and Hansson M
Plant Physiology and Biochemistry 45 (12), 932-936 (2007)

The role of subsite +2 of the Trichoderma reesei β-mannanase TrMan5A in hydrolysis and transglycosylation
Rosengren A, Hägglund P, Anderson L, Orozco PP, Peterson Wulff R, Nerinckx W and Stålbrand H
Biocatalysis and Biotransformation (Accepted manuscript)

Contributions to the Papers

I. RPW designed and performed the cross-linking experiments. PRW expressed and purified proteins. RPW analysed the data and took part in drafting of the manuscript.

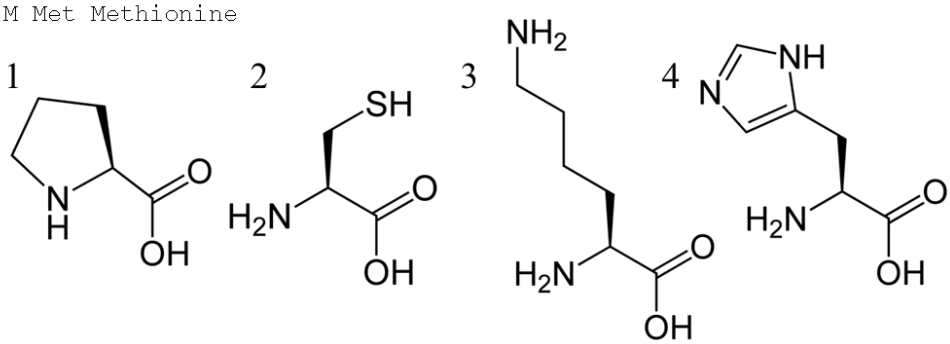
II. RPW designed the study and performed the cross-linking experiments. RPW expressed and purified proteins. RPW analysed the data and drafted the manuscript.

III. RPW took part in the experiments and drafting of the manuscript.

IV. RPW took part in designing the study and performed experiments. RPW analysed data and drafted the manuscript.

Amino acids

Non-polar	**Polar**	**Charged**
A Ala Alanine	G Gly Glycine	D Asp Aspartate (–)
V Val Valine	S Ser Serine	A Glu Glutamate (–)
L Leu Leucine	T Thr Threonine	K Lys Lysine(3) (+)
I Ile Isoleucine	C Cys Cysteine(2)	R Arg Arginine (+)
P Pro Proline(1)	Y Tyr Tyrosine	H His Histidine(4) (+)
F Phe Phenylalanine	N Asn Aspargine	
W Trp Tryptophan	Q Gln Glutamine	
M Met Methionine		



Abbreviations

ALA	5-aminolevulinic acid
ADP	Adenosine diphosphate
ATP	Adenosine-5'-triphosphate
Bchl	Bacteriochlorophyll
BchIDH	Mg-chelatase holoenzyme
BS^3	Bis(sulfosuccinimidyl) suberate
Chl	Chlorophyll
Cryo-EM	Cryo-electron microscopy
DTSSP	3,3´-dithiobis(sulfosuccinimidylpropionate)
ESI	Electrospray ionization
MALDI	Matrix-assisted laser desorption/ionization
MPE	Mg-chelatase monomethyl ester
MS	Mass spectrometry
m/z	Mass-to-charge
NADPH	Nicotinamide adenine dinucleotide phosphate
Prx	Peroxiredoxin
TOF	Time-of-flight
Trx	Thioredoxin
UV	Ultra violet

Short names for different proteins

BchI	Mg-chelatase subunit I of bacteria like *Rhodobacter capsulatus*
BchD	Mg-chelatase subunit D of bacteria like *Rhodobacter capsulatus*
BchH	Mg-chelatase subunit H of bacteria like *Rhodobacter capsulatus*
BchE	Cyclase in bacteria
BchM	Methyltransferase of bacteria like *Rhodobacter capsulatus*
ChlI	Mg-chelatase subunit I in plants and in cyanobacteria
ChlD	Mg-chelatase subunit D in plants and in cyanobacteria
ChlH	Mg-chelatase subunit H in plants and in cyanobacteria
Gun4	Genomes uncoupled 4
NTRC	NADPH-dependent thioredoxin reductase C

Contents

Summary

Proteins are found on the inside, in the membrane, on the surface and on the outside of cells. They form complicated structures and they interact with other molecules and proteins. Protein complexes and protein-protein interactions are challenging to investigate and in the beginning of protein research most studies were done with single proteins, often in water. Although, *in vivo* proteins rarely function alone.

To study protein complexes, two enzymes in the chlorophyll biosynthetic pathway were selected, Mg-chelatase in *Rhodobacter capsulatus* (bacteria) and the MPE cyclase complex in *Hordeum vulgare* (barley) and *Arabidopsis thaliana* (mouse-ear cress). Chlorophyll is a pigment formed through a complicated reaction path. Chlorophyll biosynthesis takes place in chlorophyll-producing organisms. The first committed step towards chlorophyll biosynthesis is performed by the enzyme complex Mg-chelatase. Mg-chelatase inserts a Mg^{2+} ion into the porphyrin substrate. The pathway is continued by a methyltransferase and thereafter the MPE cyclase complex which performs a complicated ring-closure in the porphyrin.

Mg-chelatase is composed of three proteins, BchI (40 kDa), BchD (60 kDa) and BchH (130 kDa). A cryo-electron microscopy model of the BchID complex (7.5 Å) revealed a two-tired hexameric ring structure with an arrangement of the subunits as a trimer of dimers. The transient full complex of Mg-chelatase, BchIDH, was chemically cross-linked and BchH was found to interact with the D-side of the BchID complex.

The MPE cyclase complex was more difficult to study and two of the three core components of the complex are still unknown. An interesting enzyme, NADPH-dependent thioredoxin reductase C (NTRC), was found to stimulate the MPE cyclase reaction together with a 2-Cys peroxiredoxin. NTRC was characterised further with regards to function and structure. The enzyme consists of a fusion between a NADPH-dependent thioredoxin reductase polypeptide and a thioredoxin polypeptide in the C-terminal. The three-dimensional structure of NTRC was determined with cryo-electron microscopy (10.0 Å) and revealed a tetramer.

1 Protein complexes

The background information of the topic "protein complexes" will start with the question of where proteins meet (section 1.1) and with the definition of what a protein complex is (section 1.2). A small introduction to some of the physical forces that hold molecules such as proteins together follow (section 1.3). The next question discusses why some proteins form complexes (section 1.4) and the last section deals with what methods to use to identify interacting proteins and protein complex (section 1.5).

1.1 Where do proteins meet?

Proteins are highly complicated compounds present in all living species. They consist of one or more polypeptides folded in a specific manner to facilitate a biological function. Throughout the years, most biochemical studies have focused on individual proteins. Nonetheless, proteins hardly ever function alone.

The abundance of proteins in an average cell can be estimated to millions, although no one knows the exact number. Proteins situated on the cell surface are less abundant than proteins on the inside of the cell (Lodish, Berk et al. 2000). Although less abundant, the cell surface proteins are often involved in regulation and activation of the cells. This is because many of these proteins function in the communication system of the cell, which logically is located in the cell membrane that is the border between the inside and the outside of the cell.

In medical research, the cell surface proteins are of interest in for example autoimmune diseases, immunotherapy and cancer therapy (Monney, Sabatos et al. 2002; Ho and Kim 2011; Zhang, Ma et al. 2011). In drug discovery, the interaction surface of protein complexes is of interest (Wells and McClendon 2007). A challenge in drug discovery is to find a small molecule with high affinity and selectivity towards a specific surface or active "pocket" on the target protein mimicking the site where protein complex formation occurs (Wells and McClendon 2007).

In plants there are different compartments such as the chloroplasts. The arena for soluble proteins to interact with other proteins can be either on the outside of the

cell, in the lumen, or in the stroma. Membrane proteins in plants are found in the thylakoid membrane or in the envelopes. In prokaryotes, organisms without cell nucleus, soluble proteins are found in the cytoplasm and membrane protein in the plasma membrane or in the cell wall.

1.2 What is a complex?

What is a complex? I will start to define another question: What is a protein-protein interaction? A general and simple answer would be molecular contact between proteins in a cell or in a living organism. Previously discussed in the literature is the distinction of two proteins sharing a "functional contact" but not a "physical contact". An example of proteins that are functionally interacting is the proteins in the ribosomes. The proteins share a functional contact but all proteins in a specific complex are not in physical contact (Mackay, Sunde et al. 2007; Chatr-Aryamontri, Ceol et al. 2008; De Las Rivas and Fontanillo 2010). Proteins can definitely interact even though the proteins are not physically connected.

Back to the first question: What is a protein complex? A protein complex and a protein-protein interaction should be specific, i.e. not just a passing protein inside or outside the cell that accidentally interacts with another protein in a non-specific way. Furthermore, a protein complex can exist in different timescales. Some proteins assembly in a stable complex, such as the subunits I and D of Mg-chelatase. Other protein complexes and protein-protein interactions are suggested to be more transient although highly specific, such as the interaction of the ID complex with BchH of Mg-chelatase (Paper I and II).

A protein can form a complex with its substrate and with cofactors (coenzyme or prosthetic group). And a protein without its cofactor is referred to as an apo-protein. The cofactor can either be an inorganic compound (metal ion) or an organic compound such as FAD, biotin or coenzyme A.

Yet another question: Can aggregated proteins be regarded as protein complexes? Generally, the definition of an aggregated protein is a denatured or misfolded protein. But there are cases when proteins "aggregate" into higher order of multi-oligomeric complex structures in a specific manner as a response to for example regulation or for some unknown reason. They can possibly be considered as protein complexes. Such an example is amyloids, the phenomenon of fibrillation, which have a specific structure but they are regarded as aggregated protein (Hellstrand, Boland et al. 2010).

Another type of interesting proteins are chaperones. A well-studied type of chaperones is heat-shock proteins (Hsp). They are often ring-structured homo-

oligomers and complex by the fact that they often are composed of a (double) ring formation of hexamers (e.g. heat-shock protein 21 and heat-shock protein 100) or heptamers (e.g. heat-shock protein 60) (Kirstein, Moliere et al. 2009; Mayer 2010; Lambert, Koeck et al. 2011). Chaperones are up-regulated in the cell as a response to stress, such as heat-stress or draught. Chaperones are not only keeping proteins in their native state and refolding denatured protein but can also serve as a platform for protein subunits to assembly on (Ellis 2006; Mayer 2010). Commonly, chaperones are believed not to be directly involved in the biological process of the protein.

1.3 What interactions hold a protein complex together?

What interactions (forces) hold a protein complex together? This is an important question to be asked. For a complex formation to even occur, it has to be energetically favourable. There are both intramolecular forces, acting within the protein, as well as intermolecular forces, between proteins in a complex. Some forces to be considered are hydrogen bonds, the hydrophobic effect, Coulombic interactions and van der Waals force.

A hydrogen bond is an interaction between a hydrogen atom and an electronegative atom (oxygen, nitrogen, chloride or fluoride). The hydrogen atom needs also to be bound to another electronegative atom. Hydrogen bonds help proteins form their well-known organised secondary structure, helices and sheets. Hydrogen bonds are also creating tertiary and quaternary structure in a protein and between subunits in a protein complex. Another force, the hydrophobic effect, is important for protein folding and structure. For soluble proteins and protein complexes, the hydrophobic amino acids (alanine, valine, leucine, isoleucine, methionine, phenylalanine and tryptophan) have their side-chains pointing towards the core of the structure, to avoid the hydrophilic water solution and thereby stabilising the structure (McNaught and Wilkinson 1997).

Another type of interaction is electrostatic interaction between electrically charged particles, called Coulomb´s law. The force (F) is explained by:

$$F = k_e \frac{q_1 q_2}{r^2}$$

where k_e is a proportionality constant, q_1 and q_2 are two charged points and r is the separation distance between q_1 and q_2.

In Coulombic interactions, the acidic (aspartate and glutamate) or basic (lysine, arginine and histidine) amino acid residues are important. The pH of the solution

will decide whether the residues will be charged or not and consequently they will interact according to Coulomb´s law. By adding salt to the solution, the high charge-charge interaction energy can be decreased (Baldwin 2007). In van der Waals force, the attractions between molecules or surfaces are the sum of all attractive or repulsive forces between them, excluding covalent bonds and electrostatic interactions. The force is considered weak, compared to for example covalent bonds and hydrogen bonds but they are still very important for protein-protein interactions (McNaught and Wilkinson 1997).

1.4 Why form a protein complex?

Why do proteins form complexes? One can think of several answers to that question. The first obvious answer would be that more than one protein or subunit of a protein is needed to perform an enzymatic step or reaction. Therefore, to achieve a functional protein, a complex formation or oligomerization of the protein is needed.

Another reason to form a complex can be to facilitate a reaction pathway and a way of channeling a substrate between enzymes in a pathway. This is especially important in cases when the substrate is toxic to the cell or organism, as in the case with porphyrins (Eckhardt, Grimm et al. 2004). The substrate channeling then protects the cell from the toxic substance.

Some (soluble) proteins have a hydrophobic surface and cannot be stable unless in its oligomeric state, e.g. a ring structure forming a channel, so the hydrophobic parts of the protein are turned inwards the ring structure. Other proteins need a partner protein with which to form complex to be stable, that is to stay in its native form, at least *in vitro* as in the example of BchD of Mg-chelatase (Jensen, Gibson et al. 1999; Axelsson, Lundqvist et al. 2006).

Complex formation can also be due to regulation. As a way to regulate a reaction step or a pathway, proteins form high-molecular well-organized complexes, and therefore have a biological function (Trost, Fermani et al. 2006). Oligomerization can also be induced in vitro due to the method used for purification as in the example of the S100B protein, which is purified with a His-tag and the imidazole molecule causes oligomerization with no biological function (Thulin, Kesvatera et al. 2011). Another benefit with forming a complex is achieving stability, for example with a metal ion to stabilize the structure of the complex. One example of such a structure is an iron-sulfur cluster (Meyer 2008). However, the reason why some proteins form high molecular mass complexes is still unknown.

1.5 How to discover a complex

Complexes can be discovered by a large variety of techniques and there are many methods available for studies on multi-protein complexes and protein-protein interactions, by molecular, biochemical or structural approaches. A selection of methods have been reviewed by Sali and colleagues (Sali, Glaeser et al. 2003). Some techniques are more suitable for high throughput analyses (for review see De Las Rivas and Fontanillo 2010) while others are more relevant when focusing on interactions in a specific protein complex (for review see Rappsilber 2011). A method that provides data based on physical contact between proteins is regarded as a "binary" method and data based on interactions, both direct and indirect, is regarded as a "co-complex" method (De Las Rivas and Fontanillo 2010).

An example of a popular large-scale analysis method giving "co-complex" data is tandem-affinity purification coupled to mass spectrometry (TAP-MS) where a known protein is tagged and expressed in suitable cells and can form complex under physiological conditions (Rigaut, Shevchenko et al. 1999; Berggard, Linse et al. 2007; Gavin, Maeda et al. 2011). Another "co-complex" method is co-immunoprecipitation based on antibody recognition (Co-IP) (Bonifacino, Dell'Angelica et al. 2001). In this method a known protein is targeted with an antibody and only tightly bound proteins are identified. Co-IP has become more of a standard when identifying new (unknown) protein complexes (Mackay, Sunde et al. 2007). Data collected with Co-IP can be found for different complexes in a database for protein-protein interactions called MINT – Molecular INTeraction database (Cesareni, Chatr-Aryamontri et al. 2008; Chatr-Aryamontri, Zanzoni et al. 2008; Ceol, Chatr-Aryamontri et al. 2010).

A "binary" method for screening protein-protein interactions is the yeast two-hybrid technique developed by Fields and Song (Fields and Song 1989). It is a powerful technique giving data based on molecular biology (Fields and Song 1989; Suter, Kittanakom et al. 2008). A drawback is the high percentage of "false positives" that are estimated to as much as 70% (Deane, Salwinski et al. 2002). A small-scale analysis method based on "binary" data is cross-linking combined with mass spectrometry (cross-linking/MS) which will be described in more detail in section 2.4.

Many protein complexes have already been identified, but there are many more for scientists around the world to discover.

2 Material and methods

A selection of methods used during the thesis work will be presented here. First presented is preparation of barley etioplasts (section 2.1). Methods on different ways to analyse the proteins and protein complexes are presented after that, starting with chromatography (section 2.2) where three techniques are focused on: size-exclusion chromatography, high performance liquid chromatography and reversed-phase liquid chromatography. The following sections deal with the basics in mass spectrometry (section 2.3), cross-linking/mass spectrometry (section 2.4), examples of chemical cross-linkers (section 2.5) and software to identify cross-links (section 2.6). Finally, there is a brief description on structure determination with cryo-electron microscopy (section 2.7).

2.1 Preparation of barley etioplasts

Chloroplasts that have not been exposed to light are referred to as etioplasts. Etioplasts are found in plants that have been grown in the dark and they are colored yellow rather than green. To prepare barley etioplasts, plants are grown in the dark for about seven days in room temperature. The top of the etiolated seedlings are collected and mixed with buffer in a blender. After centrifugation steps, a small pellet is obtained and solubilized in a suitable buffer. All preparation steps to obtain etioplasts need to be performed under dim green light not to form chlorophyll and chloroplasts (Hansson, Kannangara et al. 1999).

2.2 Chromatography

Chromatography was first invented by the Russian botanist Mikhail Tswett in the beginning of the 20th century during his studies on plant pigments, chlorophyll and carotenoids (Tswett 1906; Tswett 1906). Chromatography is a way of for example separating a mixture of proteins or purifying a sample. The sample is dissolved in a mobile phase and carried through a stationary phase. Chromatography can be divided into column chromatography or planer chromatography (e.g. thin-layer and paper). Column chromatography can be divided into different groups based on separation mechanism (ion-exchange and

size exclusion), mobile phase properties (gas or liquid), other special techniques (e.g. reversed-phase) or affinity.

In ion-exchange chromatography, the analytes are separated according to their charges and can be adjusted with the pH of the buffer. Size-exclusion chromatography separates the analytes according to the size of the analytes or rather the hydrodynamic diameter. Size-exclusion chromatography, or gel filtration, is a very gentle method which allows the protein to stay in its native state.

In gas chromatography the mobile phase is a gas and the stationary phase is packed in a column. In liquid chromatography the mobile phase is a liquid and the stationary phase is packed in a column in a similar way to in gas chromatography. One example of liquid chromatography is high performance liquid chromatography (HPLC). The sample in a liquid mobile phase is forced through a packed column under high pressure.

Reversed-phase liquid chromatography is a form of HPLC. The stationary phase used in reversed-phase liquid chromatography consists of non-polar beads. For peptides and proteins, C18 (n-octadecylsilyl) or C8 (n-octylsilyl) coated silica are commonly used. The mobile phase (e.g. water/acetonitrile) is more polar than the stationary phase. Reversed-phase liquid chromatography (often nano-liquid chromatography) is commonly used when a protein has been run on an SDS-PAGE, digested with protease, and the peptides need further separation before analysis with a mass spectrometer.

As the name implies, in affinity chromatography the analyte, often with a tag, reacts very specifically with a molecule attached to the column resin, e.g. an over-expressed protein can be tagged and further easily purified with an affinity column (Paper I-IV).

In this thesis I have mainly used two different chromatographic techniques; size-exclusion chromatography and HPLC (reversed-phase liquid chromatography). Size-exclusion chromatography (Paper IV) is used to characterize the protein in the study (NADPH-dependent thioredoxin reductase C – NTRC), to determine the native structure of the protein and as a purification step before cryo-electron microscopy. HPLC is used (Paper III) to detect chlorophyll intermediates accumulated upon feeding with ALA and reversed-phase nano-liquid chromatography (Paper I and Paper II) is used to fractionate peptides onto steel plate before analysis with mass spectrometry (MS) which will be described further in section 2.3.

2.3 Biological mass spectrometry

Mass spectrometry (MS) is used for determining the mass of a chemical compound, e.g. peptides, carbohydrates or intact proteins. The mass spectrometer determines the mass-to-charge (m/z) ratio of a charged analyte. The analyte needs therefore to be ionized. The principles of the mass spectrometer is the ionization of the sample, separation of ions by m/z ratio with an analyser, the detector, program conversion of m/z to a mass fingerprint in MS mode or to detection of fragment fingerprint in MS/MS mode (Figure 1). Thereafter the obtained data is interpreted with help from data analysis software.

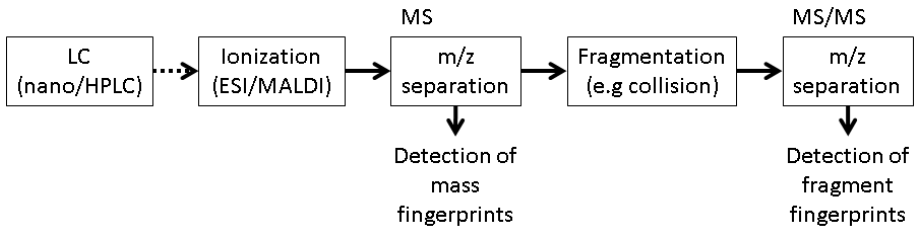


Figure 1. Schematic picture of the principles of a mass spectrometer. An LC, either nano-LC or HPLC, can be coupled to the mass spectrometer, directly (online) or indirectly (offline). The ionization source, ESI or MALDI, is used to ionize the sample before separation based on mass-to-charge ratio (m/z). Thereafter mass fingerprints can be detected. For a mass spectrometer in MS/MS mode further fragmentation is needed, with collision induced dissociation (CID) and fragment fingerprints can be detected.

A liquid chromatography system (LC) can be coupled to the mass spectrometer (Figure 1), either online (directly coupled) or offline (separately) (Paper I and II). MS/MS, or tandem MS, gives the possibility to obtain information about for example the amino acid sequence of a selected peptide. The first mass analyser selects a precursor ion mass which is further fragmented, for example with collision induced dissociation (CID) (Figure 1). The second mass analyser separates the fragments according to their m/z ratio and detects the fragment fingerprint. The peptide fragment fingerprints are analysed according to the special patterns for different fragment ions of which the most common are referred to as b or y ions (Roepstorff and Fohlman 1984).

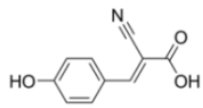
2.3.1 Important modules of the instrument

There are many different options and combinations for the two important modules of the mass spectrometer, the ion source and the mass analyser. The introduction of soft ionization techniques for mass spectrometry came in the late 1980ies. Electrospray ionization (ESI) was developed by John B. Fenn and colleagues (Fenn, Mann et al. 1989), soft laser desorption (SLD) by Koichi Tanaka and

colleagues (Tanaka, Waki et al. 1988) and matrix-assisted laser desorption/ ionization (MALDI) by Michael Karas, Franz Hillenkamp and colleagues (Karas, Bachmann et al. 1987). John B. Fenn and Koichi Tanaka were awarded a shared Nobel Prize in Chemistry in 2002 for their important discoveries on the soft ionization techniques ESI and SLD, while Karas and Hillenkamp were not. However, SLD is not used anymore while the MALDI technique is.

Ion source

The two dominating techniques used in chemical ionization is electrospray ionization (ESI) or matrix-assisted laser desorption/ ionization (MALDI). When using ESI, the analyte is sprayed into a fine aerosol. A typical solvent used is water mixed with a volatile organic compound (e.g. acetonitrile or methanol) and to decrease the size of the droplet, acetic acid (or another compound that increase the conductivity) is added.



α-cyano

In MALDI, the analyte is co-crystallized with a matrix, typically α-cyano-4-hydroxycinnamic acid (α-cyano), 2,5-dihydroxybenzoic acid (DHB) or sinapinic acid. For a peptide sample, α-cyano is often used as matrix (Paper I and II). In MALDI, the first step is desorption of the sample. Desorption is usually achieved by using a UV laser beam. Energy is absorbed by the matrix and protons are added or removed. The proton is supposed to be transferred to (or from) the sample molecule, e.g. the peptide, and a positively or negatively charged molecule is created and single charged ions are the most common ones to detect.

Mass analyser

Popular mass analysers to use in a mass spectrometer are (triple) quadrupole, orbitrap or time-of-flight (TOF) and will be described in brief. A quadrupole mass analyser is composed of four metal rods and each of the two opposing rods is connected electrically. A voltage is applied and only ions of a specific m/z ratio travelling between the rods will reach the detector. Triple quadrupoles have three sets of quadrupoles in a linear series where the first and the third are mass filters and the second one is used as a collision cell making it possible to analyse MS/MS fragmentation of the sample (Yost and Enke 1978; Morrison 1991).

In an orbitrap, ions in an electric field are trapped in an orbit around a spindle-shaped electrode. Ions with a specific m/z ratio will move in rings around the spindle and move back and forth. The ion oscillation will create an image current which the detector plates recognize and mass spectra are obtained. Orbitraps have a high mass accuracy of about 1-2 ppm (Hu, Noll et al. 2005; Scigelova and Makarov 2006).

In a TOF analyser, the ions are accelerated by an electric field in a vacuum tube and the velocity of the ions depends on the m/z ratio. Lighter ions travel faster

than heavy ions. The time it takes to travel a specific distance for a specific ion is measured by the detector and thereafter converted to the m/z ratio. A TOF-TOF instrument has two flight tubes connected in series with a collision chamber in between and MS/MS can be performed (Wollnik 1993).

2.3.2 MALDI-TOF(-TOF)

The combination of ionization and mass analyser used in this thesis (Paper I, II and IV) is the popular MALDI-TOF-TOF, or also called tandem time-of-flight. Commonly, a MALDI-TOF is used in MS mode to detect (peptide) mass fingerprints (Figure 1). This is a way to rapidly identify e.g. proteins in a solution or extracted from an SDS-PAGE, if the protein´s genome is submitted to the database. If the instrument is equipped with two flight tubes (TOF-TOF), it is possible to detect (peptide) fragment fingerprints in MS/MS mode (Figure 2). It will reveal information of the amino acid sequence of the detected peptide.

A MALDI-TOF can also be used in quantitative proteomics, revealing quantitative data on peptides (or proteins) in a sample and show differences between samples (Bantscheff, Schirle et al. 2007). Quantitative proteomics has become a popular technique in different disciplines of protein science reviewed elsewhere (Domon and Aebersold 2006; Cravatt, Simon et al. 2007; Wilm 2009). Depending on strategy, two or more samples can be compared and they are often labeled with different tags making it possible to distinguish the peptides in the mass spectrometer (Xie, Liu et al. 2011).

MALDI-TOF instruments are common in analytical laboratories to study a great number of organic compounds and in biochemistry to study peptides and proteins. The latest trend using the instrument of MALDI-TOF is in the field of microbiology. It is now possible to rapidly and accurately identify microorganisms such as fungi, bacteria or algae and it has more and more become a standard in clinical microbiology laboratories (Seng, Drancourt et al. 2009).

2.4 Chemical cross-linking/mass spectrometry

Cross-linking/MS is a method which can be used to gain structural information of peptides within or between proteins by linking them together covalently. The technique is based on “binary” data (see section 1.5). It can be used for example to improve a low resolution cryo-electron microscopy (cryo-EM) structure determination by providing it with high resolution information of peptides in close proximity (Bohn, Beck et al. 2010; Chen, Jawhari et al. 2010; Rappsilber 2011).

A typical sample is a recombinant protein in its native form. One protein, a dimer/oligomer of a protein or a protein complex is studied. A selected cross-linker is added to the protein solution and the reaction is immediate. Optionally an SDS-PAGE can be performed to visualize the cross-linked protein (complex) and thereafter excise the protein from the SDS-PAGE. The protein is subjected to proteolytic digestion, often with trypsin. The digested peptides are commonly separated with a reversed-phase column (C18) nano-LC system and thereafter analysed with mass spectrometry (MS).

One challenge with cross-linked peptides is to detect the modified peptides in the mass spectrometer. The cross-linked peptides are low abundant compared to unmodified peptides in the digested sample. By using isotopically-coded cross-linkers the mass spectral detection and the analysis of the results can be improved (Jin Lee 2008; Leitner, Walzthoeni et al. 2010; Singh, Panchaud et al. 2010). Other challenges with cross-linking/MS are the necessity for cross-linking reactive sites to be available and accessible in the protein (complex) and most important, the two reactive sites have to be close enough in the protein (complex) to link together.

Cross-links can be divided into three categories: type 0 (dead-ends), type 1 (intra-peptide link) and type 2 (inter-peptide links) (Figure 2). To obtain a dead-end cross-link (Figure 2, type 0), only one end of the cross-linker has found a reaction site. The other end of the cross-linker is hydrolysed. This occurs when two cross-linking reactive sites are not within cross-linking distance or if the reactive site is situated toward the surrounding solution. The detection of dead-end cross-links which also appear as type 1 or type 2 cross-links is common. The cross-linker (see section 2.5) can react in the presence of other primary amines like Tris. The cross-linking reaction is terminated with Tris and therefore some of the cross-linkers will not react with another lysine but with Tris instead.

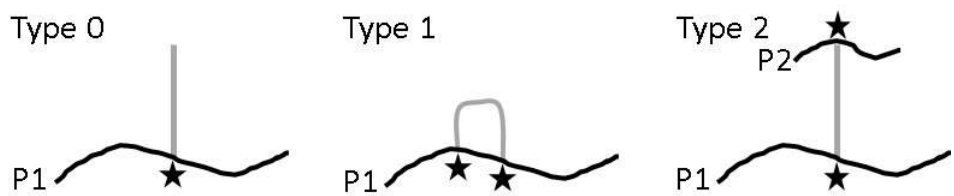


Figure 2. Categories of cross-links in peptide(s). Cross-linked peptides can be divided into three categories: type 0 (dead-end link), type 1 (intra-peptide link) and type 2 (inter-peptide link). One peptide (P1) or two peptides (P1 and P2) have reacted with the cross-linker (grey line). The star indicates where lysine residues are situated in the peptides.

An intra-peptide link, also called internal link (Figure 2, type 1), can be detected if two reactive sites are very close in the amino acid sequence, separated by only two or three residues. It becomes a cyclic link within one peptide with no proteolytic digestion site between the cross-linking reactive sites. The type 2 cross-links,

inter-peptide links (Figure 2, type 2), are obtained when the cross-linker has reacted with two separate peptides (P1 and P2) situated within cross-linking distance (King, Jones et al. 2008). The last category (type 2) can be further divided into intra-protein links and inter-protein links. Intra-protein links are cross-links found within a protein or a subunit and inter-protein links are between proteins or subunits.

The type of cross-links that gives the most important information on the arrangement or the structure of the protein (complex) is type 2 cross-links when the cross-linker has reacted with two different peptides, either within or between subunits.

2.5 Chemical cross-linkers

The chemical cross-linkers used in the experiments in this thesis work are 3,3´-dithiobis(sulfosuccinimidylpropionate) (DTSSP) and the isotope-coded bis(sulfo-succinimidyl) suberate (BS^3). DTSSP (Figure 3A) is a homobifunctional cross-linker containing N-hydroxysulfosuccinimide esters in both ends that react at pH 7-9 with primary amines in the side chain of lysine residues and at the N-terminal of the protein. The spacer arm of the cross-linker is about 12 Å. DTSSP contains a disulphide bridge which can be cleaved with reducing agent and therefore separates the cross-linked peptides. This can be useful as a control when identifying cross-linked peptides in the data analysis.

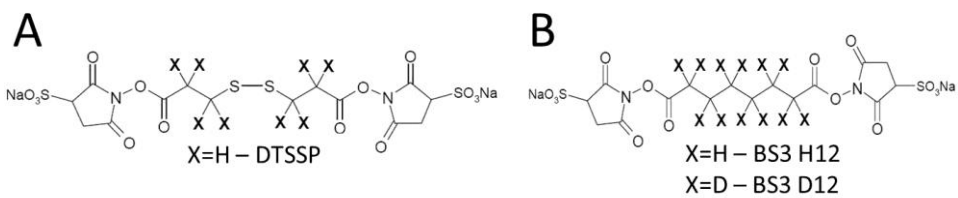


Figure 3. The homobifunctional cross-linkers DTSSP and BS^3. The cross-linkers have N-hydroxysulfosuccinimide esters in both ends that react with primary amines in the side chain of lysine residues and at the N-terminus of the protein. (A) DTSSP with a disulphide bridge which can be cleaved with reducing agent and (B) BS^3 H12/D12 is a non-cleavable analogue to DTSSP, without a disulphide bridge.

BS^3 is a non-cleavable analogue to DTSSP, lacking the DTSSP disulphide bridge, and therefore unaffected by reducing agents such as dithiothreitol (DTT) (Jin Lee 2008). BS^3 H12/D12 (Figure 3B) is isotope-coded with light (H12) and heavy (D12) forms of the reagent, which differ with 12 deuterium atoms (Creative Molecules Inc. 2008).

2.6 Software – tools to identify cross-links

There are many available programs to identify cross-linked peptides e.g. Pro-CrossLink (Gao, Xue et al. 2006), GPMAW (Peri, Steen et al. 2001), xQuest (Rinner, Seebacher et al. 2008). Several more are reviewed elsewhere (Mayne and Patterton 2011). No standard data analysis method has yet crystallized among the many available programs and different research groups use in-house software to facilitate the identification of cross-linked peptides (Chen, Jawhari et al. 2010). One software called FINDX (available at http://findxlinks.blogspot.com/) is designed to identify DTSSP or BS^3 (also for isotope-coded DTSSP and BS^3) cross-linked peptides derived from one or several proteins. The program is suitable for MALDI-TOF(-TOF) users and handles data from MS and/or MS/MS mode. A strong benefit with using MALDI-TOF(-TOF) and FINDX is the optional two-step method, where the first step is identifying possible cross-links in MS mode and thereafter running MS/MS mode and possibly verifying the first list of detected cross-links from MS mode. This is in contrast with ESI-MS users who have to run both MS and MS/MS simultaneously. Furthermore, the dominance of single charged peptides makes MALDI-MS workflow for cross-linked peptides easier to interpret than ESI-MS workflow where analyte molecules are found in several different charge states. Isotope-coded cross-linkers create a doublet pattern in the mass spectrum which simplifies the identification of cross-linked peptides. The doublet peaks are separated with 12 Da in the example of BS^3 H12/D12.

2.7 Cryo-electron microscopy

Cryo-EM is a method in structural biology using a microscope with a particle beam of electrons that illuminate a protein to produce an image. An advantage with cryo-EM compared to x-ray crystallography is that the protein can be studied in its native state rather than crystallized. However, the resolution of cryo-EM is less than that for x-ray crystallography. Therefore, large protein complexes are preferably studied rather than single proteins. The protein needs to be in a homogenous solution. Often the sample is purified with affinity chromatography and subsequently with size-exclusion chromatography to ensure that only one oligomeric form of the protein complex is analysed (Paper IV). Thereafter the sample is centrifuged to remove contaminating precipitates in the solution. The protein complex is quickly frozen to a low (cryogenic) temperature (< -150°C) which will create a snapshot of the native protein complex in the solution. Selected single particles obtained are analysed with a computerized processing technique used to analyse images (Paper I). The single-particle processing is a method to reduce the noise from the individual images. If an x-ray structure of the studied protein is available it can be fitted into the obtained image density. A resolution

down to 10 Å of a reconstructed protein complex is considered to be very good. The resolution in cryo-EM is determined by Fourier shell correlation (van Heel and Schatz 2005; Sousa and Grigorieff 2007). As a comparison, the resolution obtained with x-ray crystallography is typical in the range of 1.5-3 Å.

3 Chlorophyll biosynthesis

Oxygenic photosynthesis is performed in plants (e.g. *Hordeum vulgare* and *Arabidopsis thaliana*, see section 3.2.2 and 3.2.3), algae and some bacteria (e.g. cyanobacteria). The reaction converts carbon dioxide (CO_2) and water to (mainly) sugar ($C_6H_{12}O_6$) and oxygen (O_2):

$$6CO_2 + 6H_20 \xrightarrow{Light} C_6H_{12}O_6 + 6O_2$$

The primary product for photoautotrophs (photosynthetic organisms) is sugar which serves as an energy source for the organism. The waste product of the reaction is oxygen. Nearly all living species on Earth are dependent on the reaction either directly, like plants, or indirectly by the fact that the photoautotrophs keep the oxygen and (in a perfect world) the carbon dioxide levels balanced.

For the photosynthesis reaction to occur, energy is absorbed from (sun) light through a pigment called chlorophyll, present in plants, algae and cyanobacteria. Light energy is absorbed and transferred through an electron transport chain to a special reaction centre in the photosystems, where photosynthesis takes place.

Organisms without dependency on oxygen, such as heliobacteria, phototrophic green bacteria and phototrophic purple bacteria (e.g. *Rhodobacter capsulatus*, see section 3.2.1) can also perform photosynthesis. This is performed by an anaerobic reaction. Light energy is captured, ATP is synthesised and light energy is converted to chemical energy. In a similar way as in aerobic photosynthetic organisms, light energy is absorbed by a pigment, bacteriochlorophyll.

The chlorophyll pigment, as well as the equivalent bacteriochlorophyll in bacteria, belongs to a group of organic aromatic compounds called tetrapyrroles or porphyrins. A tetrapyrrole is composed of four pyrrole rings linked together (Figure 4). Porphyrins are highly toxic compounds if found free in the cell (Eckhardt, Grimm et al. 2004).

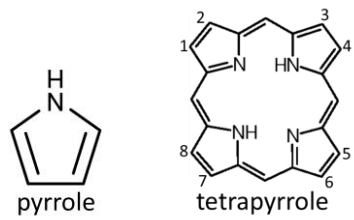


Figure 4. The structure of a pyrrole and a fully conjugated tetrapyrrole. The tetrapyrrole´s eight reactive sites, two on each pyrrole, are numbered.

The tetrapyrrole biosynthetic pathway (Figure 5) starts with the amino acid glutamate in plants and in bacteria. Glutamate is turned into glutamyl-tRNA by glutamyl-tRNA synthetase (1), followed by a reaction by glutamyl-tRNA reductase (2) to glutamate-1-semialdehyde (GSA), which is converted to 5-aminolevulinic acid (ALA) by the GSA aminotransferase (3). In non-plant eukaryotes, including mammals and yeast, and purple bacteria from the α subdivision (*Rhodospirillum*, *Rhizobium* and *Rhodobacter*), ALA is formed by ALA synthase through a condensation reaction between succinyl-CoA and glycin, (4). ALA is a common precursor to all known tetrapyrrole found in living organisms. The formation of ALA, through either of the two pathways, is the major rate-limiting step for both chlorophyll and heme biosynthesis.

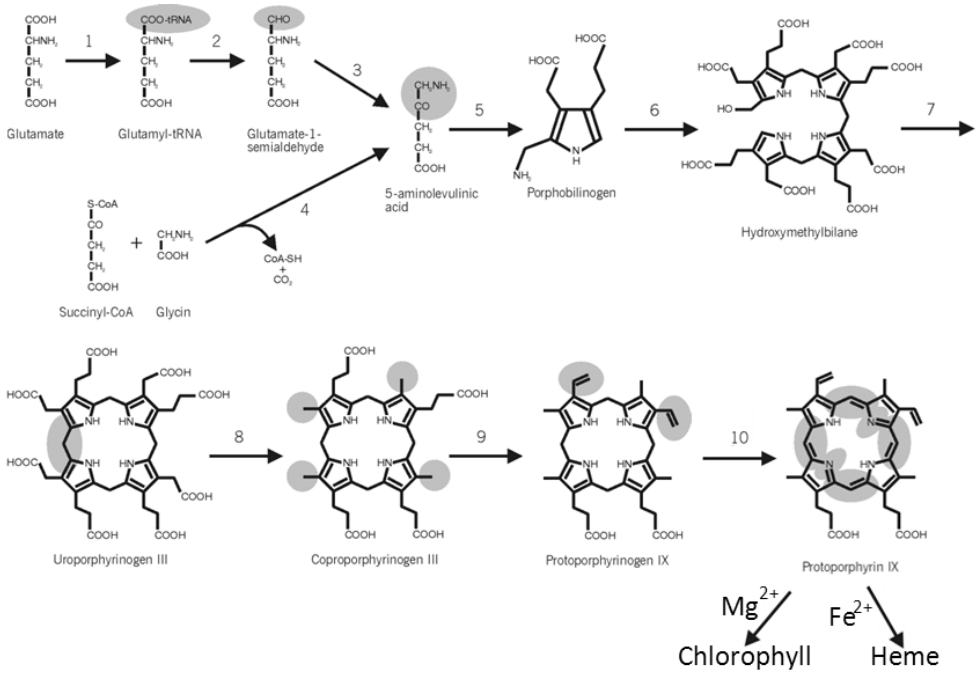


Figure 5. The common tetrapyrrole biosynthetic pathway to protoporphyrin IX. The numbers on the arrows refer to the enzyme responsible for the reaction step. (1) glutamyl-tRNA synthetase, (2) glutamyl-tRNA reductase, (3) GSA aminotransferase, (4) ALA synthase, (5) ALA-dehydratase, (6) PBG deaminase, (7) uroporphyrinogen III synthase, (8) uroporphyrinogen III carboxylase, (9) coproporphyrinogen oxidase and (10) protoporphyrinogen oxidase. After the formation of protoporphyrin IX the pathway is branched. If a Mg^{2+} is inserted into protoporphyrin IX the pathway continues towards chlorophyll and if Fe^{2+} is inserted heme b has been formed.

The first tetrapyrrole, uroporphyrinogen III, is formed through a ring closure by uroporphyrinogen III synthase (7). Eight ALA molecules are required in total. Four carboxyl groups on the porphyrin are turned into methyl groups by

uroporphyrinogen III carboxylase (8) and coproporphyrinogen III is formed. The additional two carboxyl groups are oxidised into vinyl groups by coproporphyrinogen oxidase (9) and form protoporphyrinogen IX. The last common enzyme for the tetrapyrrole biosynthetic pathway is protoporphyrinogen oxidase (10) which oxidises protoporphyrinogen IX into aromatic protoporphyrin IX, which is coloured. If Fe^{2+} is inserted by ferrochelatase into protoporphyrin IX, heme of b-type has been formed, which can be further converted to heme of o-, a-, c- or d-type. If a Mg^{2+} is inserted by Mg-chelatase, the pathway continues towards chlorophyll (Mochizuki, Tanaka et al. 2010).

The chlorophyll biosynthetic pathway (Figure 6) is formed through a complicated series of enzyme reactions. The first committed step towards chlorophyll biosynthesis is performed by Mg-chelatase (11).

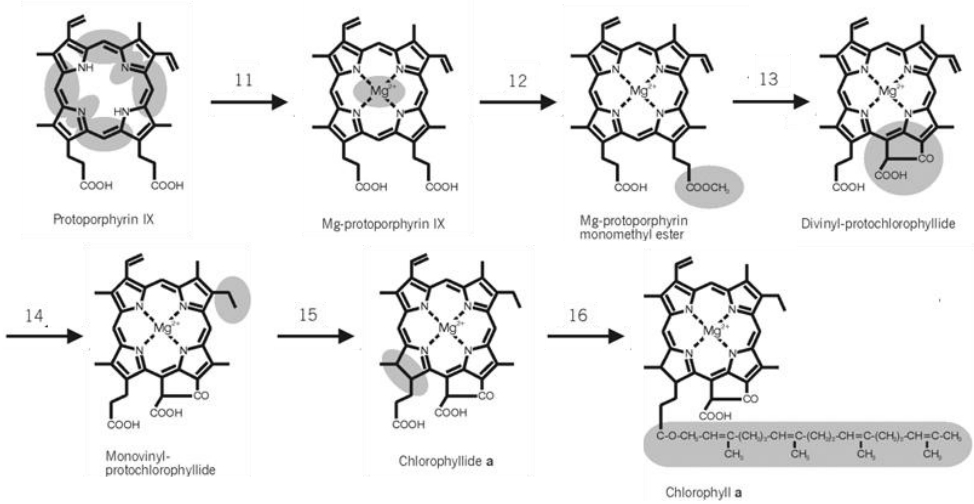


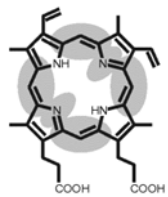
Figure 6. The chlorophyll biosynthetic pathway – from protoporphyrin IX to chlorophyll *a*. The numbers on the arrows refer to the enzyme responsible for the reaction step. (11) Mg-chelatase, (12) Mg-protoporphyrin IX methyltransferase, (13) Mg-protoporphyrin IX monomethylester cyclase, (14) divinyl reductase, (15) light-dependent (or light-independent) protochlorophyllide oxido-reductase and (16) chlorophyll synthase.

Mg-protoporphyrin IX is methylated on the carboxyl group by Mg-protoporphyrin IX methyltransferase (12) and the Mg-protoporphyrin IX monomethylester (MPE) cyclase (13) performs the formation of the fifth ring to divinyl-protochlorophyllide. A reduction is performed by divinyl reductase (DVR) (14). The light-dependent protochlorophyllide oxidoreductase (LPOR) (15) is present in chlorophyll producing organisms except for anoxygenic photosynthetic bacteria which have the light-independent protochlorophyllide oxidoreductase (DPOR) (15). The enzymes perform a reduction of a double bond in one of the pyrrole rings. The last enzyme, chlorophyll synthase (16), attaches a phytol chain and chlorophyllide *a* is converted to chlorophyll *a* (Tanaka and Tanaka 2007).

Protoporphyrin IX is an anabolic intermediate at the branch point between the chlorophyll and the heme biosynthetic pathways (Figure 5). The substrate is shared by the enzymes magnesium chelatase and ferrochelatase. As the names imply, Mg-chelatase and ferrochelatase insert a Mg^{2+} ion or an Fe^{2+} ion, respectively, into the porphyrin ring structure. With a magnesium ion coordinated in the porphyrin the pathway continues towards chlorophyll with several additional steps. Iron-protoporphyrin is also called heme b, i.e. ferrochelatase is the last enzyme of the heme pathway although heme b can be further converted to other types of heme and also to linear tetrapyrroles (phycobillins).

3.1 Substrates

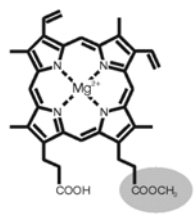
3.1.1 Protoporphyrin IX



Protoporphyrin IX

Protoporphyrin IX is commercially available but is rather difficult to dissolve in water and it easily forms aggregates (Willows and Hansson 2003). A solution of protoporphyrin IX is usually prepared in dimethyl sulfoxide (DMSO) and NH_3 is added in order to obtain an alkaline pH. Protoporphyrin IX has two vinyl groups at position 2 and 4. The vinyl groups are relatively reactive and make the porphyrin more unstable. Therefore, deuteroporphyrin IX is often used instead of protoporphyrin IX in biochemical assays with Mg-chelatase and ferrochelatase. Deuteroporphyrin IX has two hydrogen atoms instead of the vinyl groups, which makes deuteroporphyrin more stable and easier to handle and dissolve. Experiments with deuteroporphyrin IX has shown the same results as with protoporphyrin IX concluding that this non-physiological porphyrin is a good substrate of Mg-chelatase (Karger, Reid et al. 2001).

3.1.2 Mg-protoporphyrin IX monomethylester



Mg-protoporphyrin monomethyl ester

Mg-protoporphyrin IX monomethylester (MPE) is the substrate of the MPE cyclase complex. Physiologically active MPE substrate can be obtained from the *Rhodobacter capsulatus* mutant line DB575 (Gough, Rzeznicka et al. 2007). A quick method which gives large quantities of substrate is to freeze dry the cells before extraction. This is also a way to avoid contamination of carotenoids which has been a problem previously (Gough, Petersen et al. 2000; Gough, Rzeznicka et al. 2007). The isolation of MPE from *R.*

capsulatus DB575 has also the advantage that the physiological isomer of MPE is obtained. This isomer has the methylester linked to the carboxyl group at position 6 of the porphyrin skeleton (Figure 4). Chemical synthesis of MPE from Mg-protoporphyrin IX results in an isomeric mixture with 50% probability of having the methyl ester group bound to position 6 or 7 (Figure 4).

3.2 Selected model organisms

In order to study protein complexes both in the bacteriochlorophyll (Paper I and Paper II) and in the chlorophyll (Paper III and Paper IV) biosynthetic pathway, I have used the purple non-sulfur photosynthetic bacterium *Rhodobacter capsulatus* and the two plants *Hordeum vulgare* (barley) and *Arabidopsis thaliana* (mouse-ear cress).

3.2.1 *Rhodobacter capsulatus*

Rhodobacter capsulatus is an anoxygenic phototrophic bacterium which can perform bacteriochlorophyll-dependent photosynthesis (Nitschke and Rutherford 1991). It belongs to the group of purple non-sulfur bacteria which have been assigned a broad variety of biochemical, morphological and metabolic properties (Imhoff, Truper et al. 1984). Light energy is converted to chemical energy by the involvement of protein-pigment complexes between light harvesting complexes (LHC) and the pigments bacteriochlorophyll and carotenoids (McEwan 1994). The pigments give purple non-sulfur bacteria the characteristic colours ranging between purple, red and orange. All genes essential for photosynthesis of *R. capsulatus* are located in a 45 kb gene cluster (Marrs 1981). The cluster also contains the Mg-chelatase genes. The arrangement of the genes in one large gene cluster is very interesting from an evolutionary point of view but it has also been very helpful for scientists who got an excellent opportunity to connect genes with enzymatic functions in the bacteriochlorophyll as well as chlorophyll biosynthetic pathways.

3.2.2 *Hordeum vulgare*

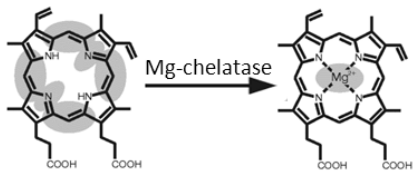
Hordeum vulgare (En. Barley. Sw. Korn) belongs to the *Poaceae* (grass) family and is classified as an angiosperm which is a plant that produces a fruit containing a seed. Barley is the fourth largest crop worldwide and is mostly used for animal feed and within the beer industry. It is also used as human additive preferentially in Asia, but has become an increasingly popular ingredient in the Western world in health food products. The genome of barley (5.3 Gbp) has so far not been

sequenced. Therefore, much of the genetic work has to rely on comparisons to related grass species like *Oryza sativa* (rice) with a genome of 0.39 Gbp and *Brachypodium distachyon* with a genome of 0.27 Gbp, which genomes rather recently have been fully sequenced (Goff, Ricke et al. 2002; The International Brachypodium Initiative 2010). The rice and the Brachypodium genomes, 0.39 and 0.27 Gbp respectively, have together a total genome size that is less than the size of one of the seven barley chromosomes.

3.2.3 *Arabidopsis thaliana*

Arabidopsis thaliana (En. Mouse-ear cress. Sw. Backtrav) is a small flowering (angiosperm) plant which belongs to the *Brassicaceae* family. Other well-known plants in the *Brassicaceae* family are e.g. *Brassica napus* (En. Rapeseed. Sw. Raps) and *Brassica oleracea*, including e.g. broccoli and cabbage (Sw. Kål). Arabidopsis has a short life cycle making it a very popular model organism. The small plant is inexpensive to propagate and it has a small genome of about 0.12 Gbp. The genome of Arabidopsis was the first plant genome to be sequenced (Dennis and Surridge 2000).

3.3 Mg-chelatase



Mg-chelatase is a multi-protein complex composed of three proteins, named BchI/ChlI (I subunit), BchD/ChlD (D subunit) and BchH/ChlH (H subunit), in the bacteriochlorophyll and the chlorophyll biosynthetic pathway, respectively. The enzyme is responsible for insertion of magnesium into protoporphyrin IX resulting in Mg-protoporphyrin IX. ATP and Mg^{2+} are needed for the reaction to take place (Beale 1999; Willows and Hansson 2003). The optimal ratio for the three components of the Mg-chelatase complex in an activity assay has been determined to 2:1:4 of I, D and H, which means that the I and H subunit need to be in access over the D subunit (Jensen, Gibson et al. 1998). The D subunit has been considered a hexameric platform for BchI to assemble on (Axelsson, Lundqvist et al. 2006).

3.3.1 The I subunit – BchI

BchI is the smallest of the three proteins of the Mg-chelatase multi-protein complex with a molecular mass of 40 kDa. BchI of *R. capsulatus* has been crystallized to a resolution of 2.1 Å (Fodje, Hansson et al. 2001). It has been shown that BchI can perform ATPase activity (Jensen, Gibson et al. 1999; Reid, Siebert et al. 2003; Reid and Hunter 2004). It has been proposed that BchI forms a

stable dimer (Willows, Gibson et al. 1996). The structural arrangement of BchI into hexameric complexes, or rather trimers of dimers (Paper I), reveals that it belongs to a class of proteins referred to as AAA+ proteins (Fodje, Hansson et al. 2001).

AAA+ proteins

AAA+ proteins stands for ATPases Associated with various cellular Activities ("Triple A" proteins). They are present in all types of organisms and are most commonly composed of a hexameric ring structure (Neuwald, Aravind et al. 1999; Vale 2000). Based on amino acid sequence homology of the important motifs and three-dimensional structure similarity, the AAA+ proteins are classified into different "clades". BchI belongs to the pre-sensor II insert clade also including the minichromosome maintenance family of helicases, the dynein/midacin family of ATP-dependent motors and the MoxR family of molecular chaperones (Iyer, Leipe et al. 2004; Erzberger and Berger 2006; Bae, Chen et al. 2009). A common feature for AAA+ proteins is the binding and hydrolysis of a nucleotide e.g. ATP that induce a conformational change in the protein (Al-Karadaghi, Franco et al. 2006; Paper I). The AAA module, conserved in the AAA family of proteins, contains a Walker A and Walker B motif. The Walker motifs are involved in binding and hydrolysis of ATP (Walker, Saraste et al. 1982). The ATP hydrolysis performed by an ATPase enzyme, such as BchI of Mg-chelatase, can simply be described as:

$$ATP + H_2O \rightarrow ADP + P_i$$

where P_i is a phosphate unit.

The target of a AAA+ protein is another protein and despite the similarities of AAA+ proteins, the biological functions differ (White and Lauring 2007; Sauer and Baker 2011). There are different ways for the AAA+ protein to recognize the substrate protein, either directly or via an adaptor protein (Dougan, Mogk et al. 2002; Mogk, Dougan et al. 2004).

3.3.2 The D subunit – BchD

BchD is the medium sized protein of the Mg-chelatase multi-protein complex with its 60 kDa. When *R. capsulatus* BchD is produced in *Escherichia coli* it forms inclusion bodies which have to be solubilised in urea. BchD is refolded by a quick dilution in a mixture containing BchI and ATP. BchD is a AAA-like protein with a high degree of sequence homology in its N-terminal part to the BchI protein (Paper I). The important Walker motifs in the ATPase active site is however incomplete resulting in an ATPase inactive D subunit. BchD can form a hexameric ring structure, even without ATP present and it has been proposed to act as a

hexameric platform used for the assembly if BchI into a two-tired hexameric ID complex (Axelsson, Lundqvist et al. 2006).

The C-terminus of BchD is homologues to an I domain of an integrin. This class of proteins binds to other proteins and thereby mediate cell-cell and cell-matrix interactions (Hynes 2002). The I domain mediates the binding through its metal ion adhesion site (MIDAS), which provides five of the six coordination sites for a Mg^{2+} ion or a Mn^{2+} ion (Hynes 1992). The sixth ligand is provided by the interacting protein. The MIDAS motif has been suggested to be involved in the coordination or binding of the metal ion (Springer 2006). Others have suggested that the MIDAS motif, with its bound Mg^{2+} ion, is involved in the binding of BchD to BchI and/or BchH (Fodje, Hansson et al. 2001).

Further, BchD contains an acidic proline-rich region connecting the AAA-like N-terminal end with the integrin-I-like domain in the C-terminus. There is no structural data of the acidic proline-rich region. The acidic proline-rich region has been proposed to be involved in protein-protein interactions (MacArthur and Thornton 1991). The acidic proline-rich region has also been suggested to be important for stabilising the quaternary structure of oligomeric protein complexes (Bergdoll, Remy et al. 1997).

3.3.3 The H subunit – BchH

The H subunit of Mg-chelatase is the largest of the three proteins in the multi-protein complex with a molecular mass of 130 kDa. When BchH is produced as a recombinant protein in *E. coli* it is coloured red due to the bound protoporphyrin IX which is the substrate of the Mg-chelatase (Gibson, Willows et al. 1995).

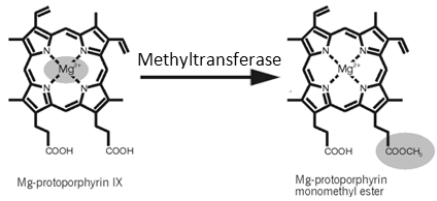
Protoporphyrin IX has been shown to be bound to BchH of Mg-chelatase. When *Rhodobacter sphaeroides* BchH is produced in *E. coli*, a red colour appears, which indicates that the subunit co-expresses with the substrate protoporphyrin IX (Gibson, Willows et al. 1995). Furthermore, experiments when BchH is pre-incubated with protoporphyrin IX have shown a shorter lag time and faster product formation (Jensen, Gibson et al. 1998).

3.3.4 Genomes uncoupled 4 – Gun4

The genome uncoupled 4 protein (Gun4) is a small protein, 22 kDa, that has been identified to bind porphyrins (Larkin, Alonso et al. 2003). The crystal structure has been determined to 1.5 Å (Davison, Schubert et al. 2005). Experiments with the Gun4 protein show a high affinity for protoporphyrin IX. Addition of Gun4 to the

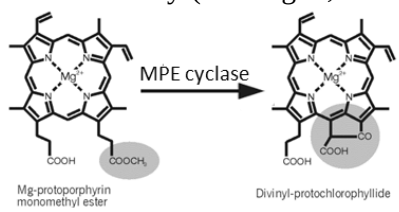
Mg-chelatase activity assay dramatically increased the reaction rate, i.e. the insertion of the magnesium ion into the porphyrin (Davison, Schubert et al. 2005). In plants and in cyanobacteria Gun4 is considered a carrier of protoporphyrin IX and Mg-protoporphyrin IX. Possibly this protects the cells from photooxidative damage caused by porphyrins. In anoxygenic photosynthetic bacteria, Gun4 does not exist, but BchJ has been suggested to play the role of a porphyrin carrier (Sawicki and Willows 2010).

3.4 Methyltransferase



Methyltransferase, also S-adenosyl-L-methionine (SAM) Mg-protoporphyrin IX methyltransferase, is a 25 kDa soluble protein associated to the membranes. The gene name in plants is *chlM* and in *R. capsulatus bchM*. It is found associated to the envelopes and to the thylakoids of *A. thaliana* (Block, Tewari et al. 2002). Methyltransferase moves a methyl group from SAM to the carboxyl group in Mg-protoporphyrin IX to form the next porphyrin referred to as Mg-protoporphyrin IX monomethylester (MPE) (see section 3.1.2), which is the substrate of the MPE cyclase system (Shepherd, McLean et al. 2005). Furthermore, the methyltransferase in *Nicotiana tabacum* (tobacco) was discovered to interact with Mg-chelatase (Alawady, Reski et al. 2005). This has previously been shown for BchM in *R. capsulatus* where BchH of Mg-chelatase was observed to stimulate the SAM Mg-protoporphyrin IX methyltransferase activity (Hinchigeri, Hundle et al. 1997).

3.5 Cyclase



The reaction step performed by the cyclase is one of the most complicated in the chlorophyll biosynthetic pathway. Starting with the substrate MPE, it requires a six-electron oxidation and the addition of one oxygen to perform the ring closure of the fifth ring in the tetrapyrrole before the product, protochlorophyllide, is obtained (Beale 1999; Bollivar 2006). At least three intermediates of the porphyrin have been proposed in the cyclase reaction. A hydroxyl group is created followed by a keto-group before the ring closure takes place (Beale 1999).

To further complicate the reaction step performed by the cyclase, two different enzymes, or enzyme systems, can complete the reaction. One system is represented by BchE, which is found in bacteriochlorophyll producing bacteria like *R. capsulatus*. So far BchE has only been connected to the cyclase reaction through its mutant phenotype, which accumulates MPE. The other cyclase system is found in chlorophyll-producing organisms and a few aerobic bacterio-

chlorophyll-producing bacteria (Beale 1999; Bollivar 2006). Most likely this system is a multi-protein complex composed of at least three core subunits that performs the enzymatic reaction.

3.5.1 Xantha-l and Viridis-k

Rzeznicka and colleagues (Rzeznicka, Walker et al. 2005) discovered that one of the subunits is a membrane bound protein encoded by the *Xantha-l* gene in barley (Xantha is Greek for yellow). The Xantha-l protein has a molecular mass of about 40 kDa. It has been determined that the membrane component encoded by Xantha-l contains a diiron component, which is not an iron-sulfur cluster (Berthold, Andersson et al. 2000; Rzeznicka, Walker et al. 2005). Yet, there is no expression system to obtain recombinant Xantha-l protein. A barley mutant, *Viridis-k*, was also assigned to be deficient in a membrane bound component of the MPE cyclase complex (Viridis is Latin for light green). However, after several attempts to identify *Viridis-k*, the gene remains unknown. For example, the barley *NTRC* gene was sequenced from the *viridis-k* deficient mutants, vir-k.23 and vir-k.170, but no mutation was found. In addition, at least one more subunit that is soluble is needed to achieve a functioning enzyme (Rzeznicka, Walker et al. 2005).

3.5.2 NADPH-dependent thioredoxin reductase C

NADPH-dependent thioredoxin reductase C (NTRC) is a newly discovered soluble protein of about 55 kDa (Serrato, Perez-Ruiz et al. 2002). In barley it is composed of a NADPH-dependent thioredoxin reductase polypeptide in the N-terminal fused with a thioredoxin part in the C-terminal. The two domains are separated by a linkage region, in barley of about 35 amino acids (Paper IV). In the sequence there are found FAD-binding motifs, NADPH-binding motifs and conserved cysteines in active site (Serrato, Perez-Ruiz et al. 2002).

The MPE cyclase reaction involves atmospheric oxygen and Xantha-l contains an iron non-sulfur cluster (Walker, Mansfield et al. 1989). Due to the fact that Xantha-l contains an iron, highly reactive H_2O_2 can be produced by the MPE cyclase reaction. Based on this, a H_2O_2 scavenger system is needed. It has been demonstrated that NTRC and 2-Cys peroxiredoxin can stimulate cyclase activity. The two enzymes, NTRC and 2-Cys peroxiredoxin, have been suggested to function as a scavenging system for reactive oxygen species in the chloroplasts, especially during darkness (Paper III).

Thioredoxins act on a number of proteins and pathways and therefore they have low specificity. In plants, thioredoxins have been reported to be involved in the Calvin cycle, ATP synthesis, fatty acid biosynthesis, starch metabolism and stress

coupled reactions (Schurmann and Buchanan 2008). Proteins in the tetrapyrrole biosynthetic pathway, such as Mg-chelatase, glutamate 1-semialdehyde aminotransferase and uroporphyrinogen decarboxylase, have also been identified as possible target proteins to thioredoxins (Balmer, Koller et al. 2003; Sturm, Jortzik et al. 2009) and NTRC has been reported to be involved in starch synthesis in chloroplasts and amyloplasts (Michalska, Zauber et al. 2009).

Rather recently, the I subunit of Mg-chelatase in *A. thaliana* was found to be a target protein for thioredoxins in chloroplasts. The ATPase activity of a recombinant I subunit was inactivated by oxidation but could be recovered through reduction by thioredoxin (Ikegami, Yoshimura et al. 2007). Therefore, the stimulatory effect of recombinant barley NTRC and 2-Cys peroxiredoxin was tested on Mg-chelatase activity. No stimulation was observed, neither for the *R. capsulatus* BchIDH system nor for the barley Mg-chelatase activity.

3.5.3 2-Cys Peroxiredoxin

Peroxiredoxins (Prx) belong to a family of antioxidant proteins. Peroxiredoxins are small, only 17-20 kDa. There are different classes of peroxiredoxins (1-Cys, 2-Cys, Prx II and Prx Q). In the 2-Cys peroxiredoxins, there are two conserved cysteine residues in each of the two subunits of the dimer (Dietz 2003). 2-Cys peroxiredoxin are found in chloroplasts (Baier and Dietz 1997) and also in the soluble fraction of etioplasts (Paper IV). 2-Cys peroxiredoxin has been discovered to protect plants against photo-oxidative damage by a Cys-thiol-based mechanism to reduce peroxides including H_2O_2 (Baier and Dietz 1999; Wood, Poole et al. 2003). Thioredoxins can reduce the oxidized peroxiredoxins. The unique combined protein complex of thioredoxin/NADPH-dependent thioredoxin reductase, NTRC, can reduce 2-Cys peroxiredoxin both *in vivo* and *in vitro* (Sueoka, Yamazaki et al. 2009).

4 Results and Discussion

This section will present and discuss the results obtained during my research on protein complexes in the chlorophyll biosynthetic pathway. First, the results on the Mg-chelatase complex will be described (section 4.1) and thereafter on the MPE cyclase complex (section 4.2).

4.1 The Mg-chelatase complex

The understanding of the Mg-chelatase multi-protein complex has grown over the years. The I subunit performs the ATP hydrolysis which fuels the insertion of the magnesium ion into the porphyrin. BchD has been reported to act as a platform for BchI to assemble. Probably because of the transient interaction between the ID complex and the H subunit, no one has been able to say where the actual interaction point is between the subunits.

The porphyrin substrate has affinity for the H subunit and the H subunit has been reported to interact with other enzymes involved in the chlorophyll biosynthetic pathway, such as Gun4 and in the bacteriochlorophyll pathway, such as BchJ. Gun4 (see section 3.3.4) has previously been characterised and has been discovered to have affinity for the porphyrin (Larkin, Alonso et al. 2003; Davison, Schubert et al. 2005). BchJ was previously proposed to be the divinyl reductase (Suzuki and Bauer 1995) but was recently reported to be involved in substrate binding in the Mg-chelatase reaction (Sawicki and Willows 2010). Mg-chelatase, and especially BchH, has also been reported to interact with the methyltransferase (Gorchein 1972; Hinchigeri, Hundle et al. 1997; Shepherd, McLean et al. 2005).

Results presented in this thesis contribute to the understanding of the Mg-chelatase complex. The first contribution regards the interaction sites between the I and D subunits and the cryo-EM structure model of the ID complex (section 4.1.1). Further the ID complex was cross-linked with BchH making it possible for the first time to detect protein-protein interaction sites for the full IDH complex. The result shows that BchH binds to BchD, which is the ATPase inactive side of the BchID AAA+ complex (section 4.1.2).

4.1.1 The BchID complex

Three different nucleotides were used to study cryo-EM structural conformations of the ID complex of Mg-chelatase. The first nucleotide, ATP, can be hydrolysed to ADP by BchI of the ID complex (Jensen, Gibson et al. 1999). The product of the reaction, ADP, was used as well as a non-hydrolysable ATP analogue called adenylyl imidodiphosphate (AMPPNP).

The structure of the *R. capsulatus* I subunit of Mg-chelatase has previously been determined with x-ray crystallography (Fodje, Hansson et al. 2001) but unfortunately not BchD. Based on a high degree of overall sequence similarity between BchI and the N-terminal AAA-like module of BchD the x-ray structure of BchI of *R. capsulatus* was suitable to use as a template for homology modelling of the AAA-like module of BchD.

The cryo-EM reconstruction of the ID_{ADP} complex was resolved to 7.5 Å (Paper I). Several domains were revealed and parts of the secondary structure could be distinguished in the density (Figure 7A). The reconstruction showed a ring structure composed of six I monomers on top of six D monomers. The three-fold symmetry demonstrates that the BchID complex is arranged as a trimer of dimers. Interestingly, within each D dimer pair, the monomers are rotated about 45° relative to each other.

The reconstruction of the ID complex, using the ATP analogue AMPPNP, was resolved to 14 Å. The ID_{AMPPNP} structure model showed a compressed complex (100 Å) in comparison to the ID_{ADP} complex (120 Å). The reconstruction of ID_{ATP} complex was resolved to 13 Å. This data set was heterogeneous due to the on-going hydrolysis of ATP which resulted in different structural conformations (Paper I).

To validate the low resolution cryo-EM model, cross-linking of the ID complex was performed (Paper I). The ID_{ADP} complex, with the best resolution from cryo-EM (7.5 Å), was used for the cross-linking experiments. The cross-linker DTSSP was used (see section 2.5). DTSSP has a length of 12 Å and is classified as a homobifunctional cross-linker with a reducible disulphide bridge. DTSSP reacts with primary amines, like lysine residues and the N-terminal of proteins. To verify the data, three different samples were compared, the subunits without cross-linker, the cross-linked ID complex sample and the sample where the cross-linked ID complex had been reduced with dithiothreitol (DTT). When DTT is added the cross-linked complex is cleaved within the linker and the proteins are observed as monomers on an SDS-PAGE. The cross-linked peptides were analysed with software called General Protein/Mass Analysis for Windows (GPMAW).

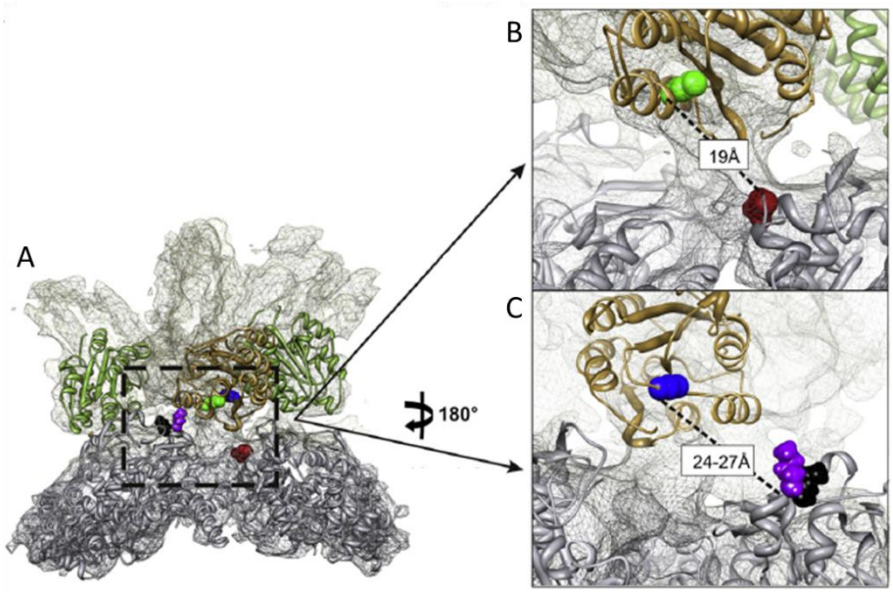


Figure 7. The location of the identified cross-linked lysine residues in BchI and BchD in the BchID cryo-EM model. (A) The BchI structure obtained by x-ray crystallography (Fodje, Hansson et al. 2001) has been aligned with the cryo-EM model of BchI (lower part) and the AAA-like domain of BchD (upper part). (B) Cross-link between the N-terminal of BchI and $K504_D$ and (C) cross-linking region between $K130_I$-$K373_D$ and $K133_I$-$K373_D$.

Cross-linking of the ID_{ADP} complex revealed an interaction site at the dimer-dimer interface of BchI (Figure 7B). The N-terminus of BchI and $K504_D$ was cross-linked (Table 1, No 1). Cross-links were also detected and mapped into the density between the C-terminal integrin I domain of BchD and BchI (Figure 7C), $K373_D$-$K130_I$ and $K373_D$-$K133_I$ (Table 1, No 2-4).

Table 1. DTSSP cross-linked peptides within the ID complex of Mg-chelatase identified with GPMAW. The cross-links marked with an asterisk (*) are also viewed in the cryo-EM model (Figure 7). The theoretical mass of the peptides (MH+) corresponds to a cross-linked peptide. Cross-link number five is from a region in BchD without any suitable template to use in the cryo-EM model and therefore impossible to map into the density.

No	MH+	Cross-linked peptides		Cross-linked lysines	
		BchI	BchD		
1*	3034	M_1TTVARLQPSASGAK$_{16}$	E_{496}LAGEQAT**K**VAR$_{507}$	N-term	K504
2*	1826	A_{127}IS**K**GEK$_{133}$	R_{372}**K**EMSDR$_{378}$	K130	K373
3*	2382	K_{131}GE**K**AFEPGLLAR$_{142}$	R_{372}**K**EMSDR$_{378}$	K133	K373
4*	2781	A_{127}IS**K**GE**K**AFEPGLLAR$_{142}$	R_{372}**K**EMSDR$_{378}$	K130	K373
5	1717	D_{313}HL**K**R$_{317}$	G_{324}**K**LEDDAK$_{331}$	K316	K325

The last cross-link detected (Table 1, No 5) is between $K316_I$-$K325_D$. The lysine in BchD is situated in the acidic proline-rich region connecting the AAA-like N-terminal domain with the integrin I domain of the C-terminus. Unfortunately, there is yet no high-resolution structural data available of this region and it is therefore impossible to map the $K316_I$-$K325_D$ link into the density. Nevertheless, it has previously been suggested that acidic proline-rich regions are involved in protein-protein interactions with I or H subunits (Walker and Willows 1997).

A trimeric structure of BchI suggests that one ID complex can hydrolyse a maximum of three ATP molecules at a time. Based on previously published data, 15 ATP molecules need to be hydrolysed per metal insertion (Reid and Hunter 2004; Sawicki and Willows 2008). Accordingly, each BchI homodimer will have to hydrolyse several ATP molecules to fuel the insertion of Mg^{2+} into the protoporphyrin IX.

4.1.2 The IDH complex

Previously a structural model of a stable ID_{ADP} double hexamer has been presented to a resolution of 7.5 Å based on cryo-EM data (Paper I). However, no activity is obtained with ADP and ATP hydrolysis is needed for the insertion of the Mg^{2+} ion into protoporphyrin IX (Reid and Hunter 2004; Sawicki and Willows 2008). It was also found that the cryo-EM reconstructions of the ID complex showed different conformations with different nucleotides, i.e. ADP, ATP or the ATP-analogue AMPPNP (Figure 8A).

To investigate the enzyme activity with AMPPNP, an activity assay was made (Paper II). AMPPNP was not able to perform the enzymatic reaction. Therefore, only ATP seems to be useful if the Mg-chelatase holoenzyme, i.e. Mg-chelatase with all three subunits (I, D and H), is the target for structural studies or studies by cross-linking.

Negative stain EM images of Mg-chelatase (Figure 8B and C) reveal formations of large complexes. In the sample where ATP was added to the mixture of BchIDH, two particle sizes were visible, the smaller one (Figure 8B, black boxes) is probably the BchID complex and the larger particles (Figure 8B, white boxes), BchIDH complexes. The larger particles of Mg-chelatase complexes were only obtained in the presence of ATP and not with the non-hydrolysable analogue AMPPNP (Figure 8C).

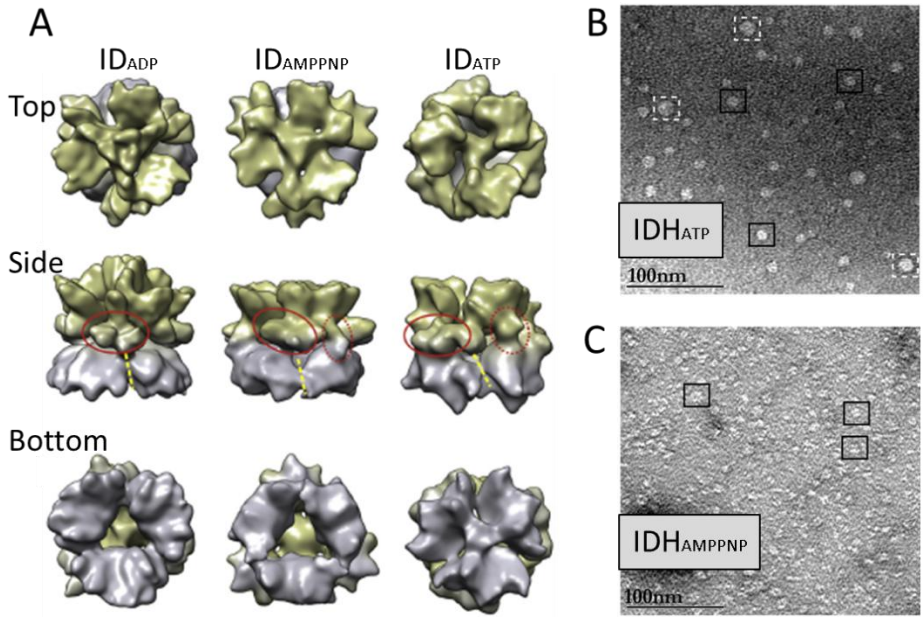


Figure 8. Single-particle reconstruction of the ID complex and negative stain EM images of the IDH complex. (A) Structural models of the ID complex with ADP, AMPPNP or ATP, based on cryo-EM. The top views show BchD and the bottom views show BchI. The side views of the models show the C-terminal integrin I like domain of BchD in different conformations (circles). Negative stain EM images of Mg-chelatase show different size of the complexes with (B) ATP or (C) AMPPNP present.

To trap the Mg-chelatase enzyme as a complete IDH complex, cross-linking was performed (Paper II). This was done using the isotope-labeled cross-linker BS^3 H12/D12 (see section 2.5). The cross-linker BS^3 differs from the previously used DTSSP by not having a cleavable disulphide bridge. This change was necessary because Mg-chelatase needs DTT to perform the chelation of the magnesium ion into the porphyrin. DTT is a reducing agent, which easily cleaves a disulphide bridge. The cross-linked BchIDH appeared in the well when an SDS-PAGE was run. Some of the proteins, I, D and H, were visible as monomers (Paper II).

Many of the detected cross-links appeared as dead-ends (see section 2.4 and Figure 2), which gave a sequence coverage of about 50% for all three subunits. The analysis showed that lysines in BchI were not detected as inter-protein links but mainly as intra-protein or dead-ends. What dominated the sample were inter-protein cross-links between BchD and BchH and intra-protein cross-links within BchD (Table 2). Fifteen cross-links were assigned as type 2 cross-links, i.e. either inter-protein links (between two subunits) or intra-protein links (within one subunit). Of the fifteen type 2 cross-links were six detected as between, or possibly within, the BchD subunits (Table 2, No 1-6). Two of the cross-links were immediately assigned as inter-protein cross-link, $K440_D$-$K440_D$ and $K504_D$-

$K504_D$ (Table 2, No 1-2). The two cross-links support our previously published EM-model of the BchID complex as a trimer of dimers, where two subunits form a dimer pair, facing each other (Paper I).

Table 2. The magnesium chelatase cross-linked peptides identified with MS. The theoretical mass of the cross-linked peptides (MH+) corresponds to the masses of two peptides from the magnesium chelatase subunits BchI, BchD or BchH and the mass of the H12/D12 isotope-labelled cross-linker BS^3. The peptides have been identified using the FINDX software based on MS and MS/MS data.

No	MH+	Cross-linked peptides		Cross-linked lysines		I/D/H	Validation
1	1804.0	S_{435}LTQTKR_{441}	S_{435}LTQTKR_{441}	K_{440}	K_{440}	D/D	a, b, c
2	2682.4	E_{502}L…KVAR_{513}	E_{502}L…KVAR_{513}	K_{504}	K_{504}	D/D	a, b, c
3	1302.8	K_{323}GK_{325}	S_{435}LTQTKR_{441}	K_{323}	K_{440}	D/D	a, b
4	1546.8	K_{323}GK_{325}	R_{371}RKEMSDR_{378}	K_{323}	K_{373}	D/D	a
5	2285.2	G_{324}KLEDDAK_{331}	E_{502}L…KVAR_{513}	K_{325}	K_{504}	D/D	a, b, c
6	2243.2	S_{435}LTQTKR_{441}	E_{496}L…KVAR_{507}	K_{440}	K_{504}	D/D	a, b, c
7	1428.8	G_{324}KLEDDAK_{331}	K_{140}LR_{142}	K_{325}	K_{140}	D/H	a, c
8	1158.6	K_{323}GK_{325}	S_{434}KVAER_{439}	K_{323}	K_{435}	D/H	a, b
9	1687.9	G_{324}KLEDDAK_{331}	T_{522}KWLK_{526}	K_{325}	K_{523}	D/H	a, b
10	1747.9	K_{373}EMSDR_{378}	R_{433}SKVAER_{439}	K_{373}	K_{435}	D/H	a, b
11	2032.1	R_{372}KEMSDR_{378}	R_{433}SKVAERK$_{440}$	K_{373}	K_{435}	D/H	a, b
12	2043.1	R_{371}RKEMSDR_{378}	K_{991}SFAYGR_{997}	K_{373}	K_{991}	D/H	a, b
13	2046.2	K_{323}GK_{325}	Q_{155}M…KMLK$_{167}$	K_{323}	K_{164}	D/H	a, b
14	2142.2	G_{324}KLEDDAK_{331}	V_{425}V…LRR_{433}	K_{325}	K_{430}	D/H	a, b, c
15	1740.0	M_{428}AKLRR_{433}	K_{991}SFAYGR_{997}	K_{430}	K_{991}	H/H	a

a. MS: Δ 12Da doublet peaks detected in FINDX.
b. MS: Δ 12Da doublet peaks in FINDX with peak intensities of 1:1 and mass error < 7 ppm.
c. MS/MS: Δ 12Da peaks in fragments by manual inspection.

The other BchD-BchD cross-links were investigated further (Paper II) and all of them (Table 2, No 3-6) were partially or fully mapped into the density of the homology-based cryo-EM model. Three of them (Table 2, No 3-5) were $K323/325_D$ situated in the acidic proline-rich region in the N-terminal and interacted with lysines in the integrin-I-like domain in the C-terminal of BchD. That particular region has previously been suggested for protein-protein interactions (Walker and Willows 1997).

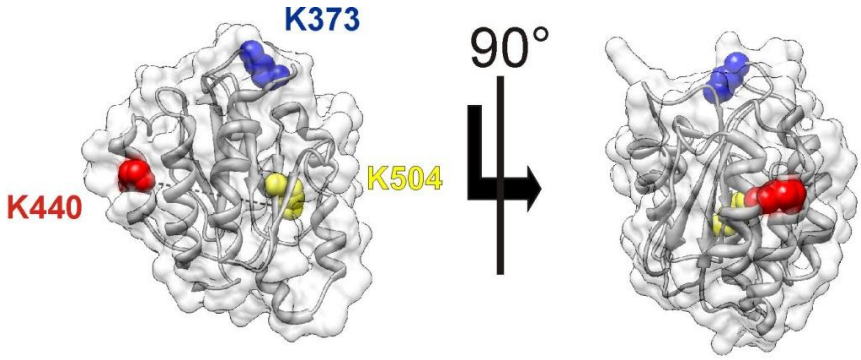


Figure 9. A homology-based model of the integrin-I-like domain of BchD. The cryo-EM model of BchD suggests the locations for $K440_D$, $K373_D$ and $K504_D$ to be on different sides of the domain.

The lysines that are situated in the integrin-I-like domain of BchD $K440_D$, $K373_D$ and $K504_D$ are suggested to be on different sides of the domain. The distance between them within one subunit of BchD is no more than 24 Å, which is within cross-linking distance (Figure 9). Despite that, the cross-link $K440_D$-$K504_D$ (Table 2, No 6) is considered to be an inter-protein cross-link. For an intra-protein cross-link to be true in this case, the cross-linker has to go through the structure to be able to cross-link with the other reactive site which is not likely to occur.

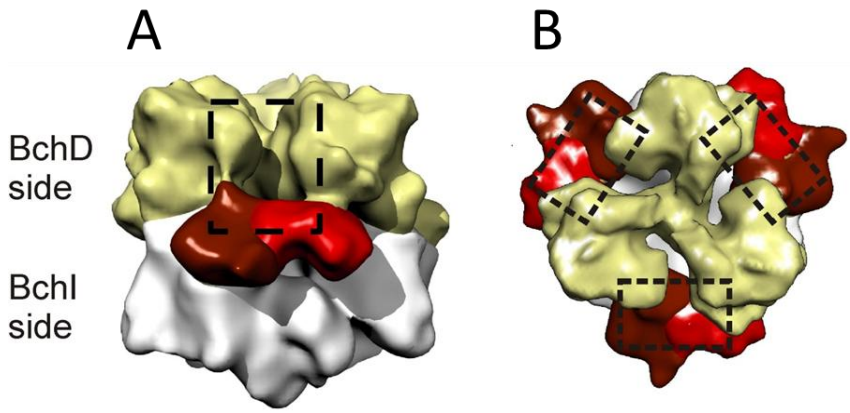


Figure 10. Interacting region of the BchID complex with BchH of Mg-chelatase. (A) The BchID complex is shown from the side and (B) from the top of the BchD hexamer. The dark/light red areas represent the C-terminal integrin-I-like domains of a BchD dimer pair. Cross-linked peptides were analysed with MALDI-MS/MS and detected with software FINDX.

Several cross-links were detected between BchD and BchH (Table 2, No 7-14). Both of the BchD lysines involved, $K323_D$ and $K373_D$, are situated in a grove between two BchD dimers in the suggested cryo-EM model of BchID (Figure 10, black box).

The IDH complex seems to be transient since purification of Mg-chelatase from plants and photosynthetic bacteria does not result in a pure holoenzyme. Rather the different subunits are obtained in different fractions (Kannangara, Vothknecht et al. 1997; Paper II). Until now, no one has been able to show how ID+H are interacting in the full complex of Mg-chelatase. Common for AAA+ proteins is that the target protein interacts with the ATPase active side, which in the case of BchIDH would be that BchH interacts with BchI, where ATPase hydrolysis takes place. Cross-linking has the benefit of catching a snapshot of what is going on in a complex between proteins, and for BchIDH, the H subunit is found to interact with the D subunits (Paper II). It is important to remember that absence of data in cross-linking is suggestive but, in a way, inconclusive. That is, only based in the fact that no data is found does not fully conclude that an interaction does not exist.

The three-fold symmetry of the BchID complex (Paper I) suggests that a 6:6:3 stoichiometry of BchI:BchD:BchH is probable. Although other AAA+ proteins, like unfoldases and proteases with a six-fold symmetry, only handle one substrate molecule at a time (Sauer and Baker 2011).

Furthermore, one cross-link was detected within BchH, $K430_H$-$K991_H$ (Table 2, No 15). $K430_H$ is situated in the N-terminal end of BchH and $K991_H$ in the C-terminal part. The N- and C-terminals have previously been shown to be situated in different domains of BchH and an EM-model was determined to 29 Å (Sirijovski, Lundqvist et al. 2008). The domains are referred to as the "thumb" domain (N-terminal) and the "finger" domain (C-terminal). The thumb domain represents amino acids 1-733 and the finger domain amino acids 734-1194. Without the porphyrin substrate present, the domains are in an open conformation. The thumb and the finger domains move together in a closed conformation when the substrate is present. The cross-link between the two domains supports the previous observation that the N- and C-terminal parts of BchH close its conformation when protoporphyrin IX binds. However, it can presently not be excluded that the observed $K430_H$-$K991_H$ cross-link is between two different H subunits.

4.2 The MPE cyclase complex

The aerobic MPE cyclase complex has puzzled scientists for a long time. No one knows how many proteins that are involved nor how the fifth ring of the porphyrin is completed. As presented in this thesis, two enzymes, NADPH–dependent thioredoxin reductase C (NTRC), together with the partner protein 2-Cys peroxiredoxin, were discovered to stimulate the MPE cyclase reaction. NTRC was characterized in regards to function (section 4.2.1) and structure (section 4.2.2)

followed by a summary on the knowledge about MPE cyclase components and their interacting partners (section 4.2.3).

4.2.1 NTRC – Function

To understand more about proteins involved in the chlorophyll biosynthetic pathway, bioinformatic screening was done using the *Arabidopsis thaliana* Co-Response Data base (Max-Planck-Institute of Molecular Plant Physiology, http://csbdb.mpimp-golm.mpg.de/csbdb/dbcor/ath.html). The database gives information about genes that co-express with a given gene, for example a gene of the chlorophyll biosynthetic pathway. One gene that was co-expressed with Arabidopsis *CHL27*, encoding the known MPE cyclase component, was the gene coding a newly discovered enzyme called NADPH–dependent thioredoxin reductase C (NTRC) (section 3.5.2).

To determine whether NTRC is connected to the chlorophyll biosynthetic pathway, *A. thaliana* seeds deficient in the *NTRC* gene were grown. The *ntrc* mutants showed a similar pale-green phenotype as a mutant deficient in CHL27 (Paper III). Next, the *A. thaliana ntrc* mutant plant was fed with ALA to analyse if the *NTRC* gene was directly affected in the chlorophyll biosynthetic pathway. ALA feeding of mutants is a common approach to reveal chlorophyll biosynthetic mutants (von Wettstein, Gough et al. 1995). The *ntrc* mutant accumulated Mg-protoporphyrin and MPE upon feeding with ALA (Paper III), which suggests that *NTRC* has a strong connection to chlorophyll biosynthesis.

A functioning MPE cyclase complex activity assay was developed using barley etioplasts. The barley *NTRC* gene was expressed in *E. coli* and the recombinant protein was purified (Paper III and Paper IV). The NTRC protein has a bright yellow colour due to the cofactor FAD (Paper IV). The barley NTRC polypeptide is 86% identical to that of rice. Due to the high degree of identity, the two proteins are expected to have similar properties (Paper IV).

2-Cys peroxiredoxin is an enzyme involved in protection against reactive oxygen species and is a H_2O_2 scavenger in chloroplasts. It is found in the soluble fraction of etioplasts together with NTRC (Paper IV). A function of NTRC is to regenerate the oxidized 2-Cys peroxiredoxin. A stimulation of MPE cyclase activity was only achieved when both NTRC and 2-Cys peroxiredoxin were present (Paper III and Paper IV). A suggestion is that NTRC and 2-Cys peroxiredoxin functions as a scavenging system for reactive oxygen species in the chloroplasts, especially during darkness.

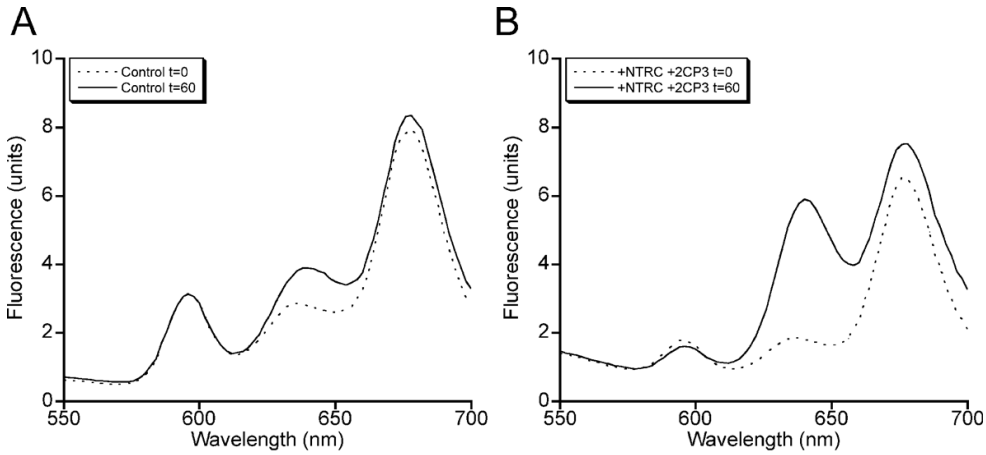


Figure 11. The ability of barley NTRC and 2-Cys peroxiredoxin to stimulate MPE cyclase activity. (A) Low cyclase activity could be detected if only a total plastid extract was used in an assay. (B) The activity could be stimulated 4-fold by addition of NTRC in combination with 2-Cys peroxiredoxin. The product of the cyclase reaction is protochlorophyllide, which has an emission maximum at 640 nm. The MPE substrate peaks at 595 nm and chlorophyllide at 670 nm. The dotted and solid lines show the reactions at 0 and 60 min, respectively.

When the substrate MPE is added to barley etioplasts, a low activity is visible, i.e. a small detectable conversion of MPE to protochlorophyllide (Figure 11A). The addition of NTRC in combination with 2-Cys peroxiredoxin stimulates the MPE cyclase activity approximately four times (Figure 11B). The product of the MPE cyclase complex reaction is protochlorophyllide (see section 3.5), which has an emission maximum at 640 nm. The MPE substrate of the cyclase (see section 3.1.2) peaks at 595 nm and chlorophyllide, which is the product of the protochlorophyllide oxido-reductase reaction, peaks at 670 nm.

Due to the fact that the deletion of *NTRC* in *A. thaliana* does not fully prevent chlorophyll biosynthesis it implies that *NTRC* is not essential for functioning biosynthesis of chlorophyll. The MPE cyclase complex is composed of at least two membrane subunits, encoded by *Xantha-l* (known gene product) and *Viridis-k* (unknown gene product). In addition, at least one soluble protein is required. The *NTRC* gene was sequenced in the two mutants, viridis-k.23 and viridis-k.170, deficient in a membrane component of the MPE cyclase complex. No mutations were found and the result clearly demonstrates that *NTRC* is not allelic to *Viridis-k*. The prepared etioplasts were divided into a supernatant and pellet fraction. The detergent-treated pellet fraction was representing the membrane fraction of the etioplasts. In the MPE cyclase activity assay NTRC cannot substitute the soluble fraction of etiolated plastids (TPS). Therefore, there has to be at least one more soluble protein needed for the MPE cyclase complex to function (Paper IV).

4.2.2 NTRC – Structure

The physical and structural properties of NTRC were further characterised. The recombinant barley NTRC has a theoretical mass of 55 kDa including a poly-His tag. On an SDS-PAGE, NTRC migrated as a 63 kDa protein (Figure 12A, lane N, arrow), similar to what has previously been shown with rice NTRC (66 kDa) (Perez-Ruiz, Gonzalez et al. 2009). Another band migrated at 130 kDa and was identified as a dimer of NTRC (Figure 12A, lane N, arrow).

When the protein was analysed with size-exclusion chromatography, it showed a mixture of two oligomeric states. Peaks of 200 kDa and 500 kDa were detected, where the smaller one was the most abundant (Figure 12B). To investigate if the fractions were stable in their oligomeric forms or if they were in equilibrium the fractions were isolated. After several days at 4°C, the oligomeric forms remained, which showed that they indeed are stable (Paper IV). Furthermore, addition of 250 µM NADPH or 5 mM DTT had no effect on the oligomeric states of barley NTRC contrasting previously shown results with rice NTRC (Perez-Ruiz, Gonzalez et al. 2009).

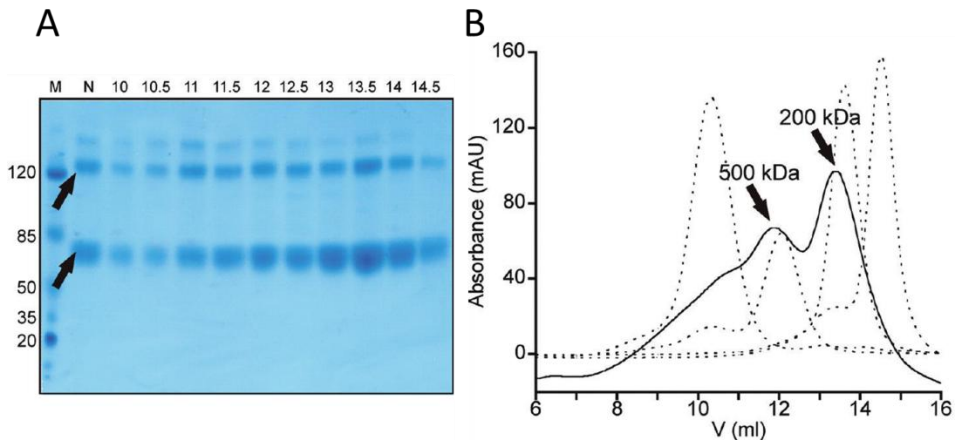


Figure 12. Size-exclusion chromatography of purified recombinant barley NTRC. (A) SDS-PAGE analysis of the barley NTRC fractions from panel B. NTRC appeared as monomers and dimers. Lane M is a molecular mass marker and N is purified NTRC. (B) Chromatogram of Superose 6 size-exclusion chromatography of NTRC (black line) and four size standard proteins: thyroglobulin (10.3 mL, 669 kDa), ferritin (12.0 mL, 440 kDa), aldolase (13.6 mL, 158 kDa), and conalbumin (14.5 mL, 75 kDa) (dotted lines). Two elution peaks are visible at 13.4 ml and 11.9 ml and correspond to masses of 200 kDa and 500 kDa, respectively.

To determine if the two oligomeric forms of barley NTRC were biologically functioning they were both tested in a MPE cyclase activity assay. Both the 200 kDa and the 500 kDa barley NTRC oligomers could stimulate the MPE cyclase activity to the same extent (Paper IV).

Moreover, the more abundant of the two oligomeric forms, the 200 kDa oligomer of NTRC, was selected for structural determination with electron microscopy. The purified recombinant barley NTRC was prepared for cryo-EM (Paper IV). The cryo-EM model obtained shows an NTRC tetramer arranged as two pairs of homodimers (Figure 13). The structure was resolved to 10.0 Å. It has four ear-shaped domains, coloured green and red. They are in an arrangement as a dimer of dimers. One monomer within a dimer has been coloured light green or light red, while the other monomer has a dark green or dark red colour (Figure 13, lower right).

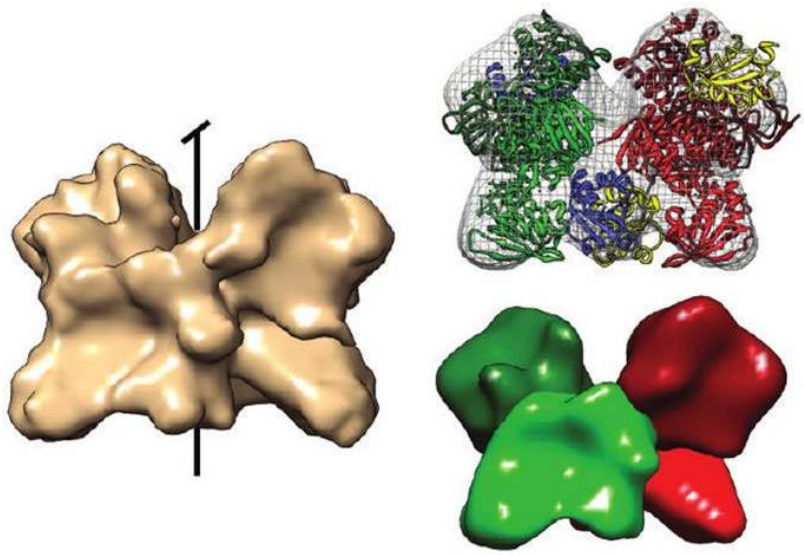


Figure 13. Three-dimensional reconstruction of the barley NTRC complex as determined by cryo-EM and single-particle reconstruction. The reconstitution, resolved to 10.0 Å, shows a tetrameric complex formed by two homodimers. The structure has four ear-shaped domains arranged as a dimer of dimers. The two dimers are coloured green and red. One monomer within a dimer has a light green or light red colour, while the other monomer has a dark green or dark red colour. In the top right corner of the model, the crystal structure of a dimeric complex of NTR, thioredoxin, and the NADP+ analogue AADP+ from *E. coli* are docked into the density of the cryo-EM reconstruction. In the ribbon representations, the two thioredoxins in each homodimer are coloured blue and yellow.

As a template, the crystal structure of a dimeric complex of NTR, thioredoxin, and the NADP+ analogue AADP+ from *E. coli* (Protein Data Bank entry 1f6m) was docked into the density of the cryo-EM reconstruction (Figure 13, upper right). Two thioredoxins in each homodimer, in the ribbon representations, are coloured blue and yellow.

4.2.3 MPE cyclase components

Many proteins seem to be involved in the complicated reaction step that forms the fifth ring of the porphyrin (see section 3.5). In barley, the *Xantha-l* gene was found to encode a membrane associated protein involved in the ring-closure, performed by the MPE cyclase complex (Rzeznicka, Walker et al. 2005). At the same time, a barley mutant deficient in the *Viridis-k* locus was assigned as a membrane associated component of the MPE cyclase complex. Very little is known about Viridis-k. The two membrane components, Xantha-l and Viridis-k, cannot perform the cyclation step alone, at least one soluble component is needed (Rzeznicka, Walker et al. 2005).

Several proteins have been suggested to be the Viridis-k or the soluble component. As already mentioned (section 4.2.1), NTRC and 2-Cys peroxiredoxin can stimulate the MPE cyclase reaction (Paper III and Paper IV). It would also be possible that MPE cyclase is connected to ferredoxin and ferredoxin reductase.

Another idea is that the Mg-chelatase is involved, perhaps through the H subunit, which has affinity for the porphyrin. Although, when the barley H subunit (Xantha-f) has been added to the MPE cyclase reaction, no stimulation was observed (data not shown). Gun4, that stimulates the Mg-chelatase reaction, has not yet been tested on the MPE cyclase reaction. Maybe Gun4 is the soluble component.

So many hours of hard research has focused on finding the soluble component of the MPE cyclase complex. Most work has been based on classical biochemical methods, such as precipitation, different chromatographic columns, SDS-PAGE, iso-electric focusing, two-dimensional SDS-PAGE, a lot of mass spectrometry and going through protein databases. Hsp70 emerged as a possible candidate but did not stimulate MPE cyclase activity. There is no other trustworthy candidate yet to be tested. One possible way to find the soluble component of the MPE cyclase complex can be to do co-immunoprecipitation (Co-IP) based on antibody recognition (see section 1.5). There are antibodies that recognize the known MPE cyclase component Xantha-l which can be used. Only tightly bound proteins in complex with Xantha-l would be identified. In this way, other interacting partners in the MPE cyclase complex might be identified.

5 Concluding remarks

There is so much for a Biochemist to discover here on Earth and so little time for each Scientist to do research.

The vital green pigment – chlorophyll. How is the pigment synthesized and by which proteins? Which cofactors are involved? Is one single protein holding the porphyrin throughout the pathway or is it passed around? Can the potentially toxic porphyrin be left free in the cell? Can the answer be that all enzymes in the chlorophyll biosynthetic pathway come together in one big super complex, interconnected to carefully handle the porphyrin?

If you have come this far, reading my Thesis, you know by now that there are so many more questions to be answered regarding the little, but utterly important, molecule of chlorophyll and the enzymes and multi-protein complexes involved in the chlorophyll biosynthetic pathway. In the world of Biochemistry there are many more complexes to discover and one of the challenges will be to identify transient complexes.

6 Conclusions of Papers

Paper I – Structure 2010

- Three different nucleotides were used to study conformations of the ID complex of Mg-chelatase: ATP (substrate that can be converted to ADP), ADP (product) and AMPPNP (a non-hydrolysable ATP analogue). The ID_{ADP} complex was also chemically cross-linked and analysed with mass spectrometry.
- A high degree of overall sequence similarity between the BchI and the AAA-like module of BchD suggested the x-ray structure of *R. capsulatus* BchI to be a suitable template for homology modelling of the AAA-like module of BchD.
- The reconstruction of ID_{ADP} (resolved to 7.5 Å) revealed several domains and some secondary structures could be distinguished in the density.
- The reconstruction showed a ring structure composed of six BchI monomers with a three-fold symmetry giving a trimer of dimer arrangement.
- Similar to the BchI subunit, six BchD monomers form a trimer of dimers. Unlike BchI, within each BchD dimer pair, the monomers are rotated about 45° relative to each other.
- Cross-linking of the ID_{ADP} complex revealed interaction sites between the integrin I domains of BchD and the dimer-dimer interface of BchI.
- The reconstruction of ID_{AMPPNP} was resolved to 14 Å. The ID_{AMPPNP} showed a compressed complex (100 Å) in comparison to ID_{ADP} (120 Å).
- The reconstruction of ID_{ATP} was resolved to 13 Å. Several conformations were found, which are likely to reflect structural changes due to ATP hydrolysis.
- The structural arrangement of trimers of dimers suggests that one ID complex can hydrolyse a maximum of three ATP molecules at a time. Based on previously published data (Reid and Hunter 2004; Sawicki and Willows 2008), 15 ATP molecules need to be hydrolysed per metal insertion. Each BchI homodimer will have to hydrolyse several ATP molecules to fuel the insertion of Mg^{2+} into the porphyrin.

Paper II – Manuscript 2011

- Negative stain particles of BchIDH reveal formation of complexes larger than the BchID complexes. The larger Mg-chelatase complexes were only obtained in the presence of ATP and not with the non-hydrolysable analogue AMPPNP.
- Six cross-links were detected between, or possibly within, D subunits.
- A homology-based model of the integrin-I-like domain of BchD suggests the location for $K440_D$, $K373_D$ and $K504_D$ to be on different sides of the domain. Observed links between any of these residues are likely to be between different BchD subunits. Otherwise, the suggested structural model of the integrin-I-like domain has to be revised.
- The inter-protein cross-links $K440_D$-$K440_D$ and $K504_D$-$K504_D$ support our previously published EM-model of the BchID complex as a trimer of dimers.
- The cross-link $K430_H$-$K991_H$ supports the previous observation that the N-terminal and C-terminal parts of BchH close its conformation when protoporphyrin IX binds. However, it has to be confirmed if the cross-link is an inter-subunit or an intra-subunit cross-link.
- Several cross-links were detected between BchD and BchH. The BchD lysines involved in cross-links with BchH are all situated in a grove between two BchD dimers in the suggested EM model of BchID.
- The three-fold symmetry of the BchID complex suggests that a 6:6:3 stoichiometry of BchI:BchD:BchH is possible.

Paper III – FEBS Letters 2008

- NADPH–dependent thioredoxin reductase C (NTRC) was found to co-regulate with chlorophyll enzymes
- The *ntrc* mutant gave a pale-green phenotype similar to a mutant deficient in the diiron component of the MPE cyclase complex.
- The *ntrc* mutant accumulated Mg-protoporphyrin and MPE upon feeding with ALA.
- Addition of NTRC in combination with 2-Cys peroxiredoxin stimulates the activity of the MPE cyclase complex.
- NTRC and 2-Cys peroxiredoxin are suggested to function as a scavenging system for reactive oxygen species in the chloroplasts.
- The deletion of the *NTRC* gene does not fully prevent chlorophyll biosynthesis which implies that *NTRC* is not essential for functioning biosynthesis of chlorophyll.

Paper IV – Biochemistry 2011

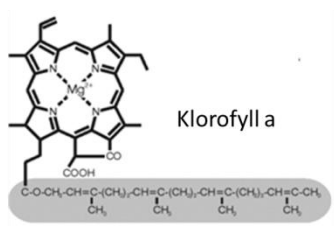
- The *NTRC* gene was found in barley and the mature polypeptide is 86% identical to the rice *NTRC* gene product. Due to the high degree of identity, the two proteins are expected to have similar properties.
- Recombinant production of barley NTRC gave bright yellow soluble protein. The colour was due to the cofactor FAD.
- The recombinant NTRC has a mass of 55 kDa (including His tag) and migrated as a 63 kDa protein on an SDS-PAGE, similar to previously shown with rice NTRC (66 kDa). Another band at 130 kDa, identified as NTRC, was suggested to be a dimer.
- Size-exclusion chromatography showed a mixture of two oligomeric states of NTRC. Peaks of 200 kDa and 500 kDa were detected, where the first one was most abundant.
- A stability test of isolated assembly forms of NTRC showed that the two oligomeric forms were stable.
- Addition of 250 μM NADPH or 5 mM DTT had no effect on the oligomeric states of barley NTRC unlike previously shown with rice NTRC.
- The 200 kDa oligomer of NTRC was selected for structure determination with cryo-EM and the model obtained shows a tetramer composed of two homodimers with 10.0 Å resolution.
- NTRC together with 2-Cys peroxiredoxin stimulate the cyclase reaction step in the chlorophyll biosynthetic pathway.
- The tetramer and the 500 kDa oligomer of NTRC show the same ability to stimulate the cyclase reaction.
- NTRC and 2-Cys peroxiredoxin cannot substitute for the soluble fraction of barley etioplasts, which suggests that there must be at least one more soluble component in the MPE cyclase complex.

Svensk sammanfattning

Titeln på avhandlingen är ”Proteinkomplex i klorofyllbiosyntetiska enzym”.

Proteiner består av långa kedjor av aminosyrer sammanlänkade med en så kallad peptidbindning. Proteinet är således en organisk molekyl och den har en hög molekylvikt som mäts i kilodalton (kDa). Proteiner kan agera som enzymer dvs de underlättar en biokemisk rektion. Ett protein kan vara funktionellt som en enhet och kallas då monomer. Ibland behövs flera proteiner, antingen av samma sort eller av olika proteiner, för att proteinet ska vara aktivt. De sitter då ihop i ett proteinkomplex. Två proteiner som sitter ihop kallas dimer, tre en trimer, fyra en tetramer, osv. Är de två proteinerna i dimeren av samma sort talar man om en homodimerer. Är de av olika sort så är det en heterodimer. Man kan också tala om oligomerer när det är två eller flera proteiner som sitter samman eller när man inte vet hur många som ingår.

De proteinkomplex som jag valt att studera ingår i klorofyllbiosyntesen. De är Mg-chelatase från en bakterie (*Rhodobacter capsulatus*), samt MPE cyclase-komplexet från korn (*Hordeum vulgare*) och från backtrav (*Arabidopsis thaliana*). Alla tre valda organismer är klassiska modellsystem inom biokemisk analys. En syntes betyder att ”sätta ihop” och i biokemisk mening är det en stegvis process där en molekyl bildas, som exempelvis klorofyll.



Klorofyll är ett pigment och består av en komplicerad organisk ringstruktur, en porfyrin, med en magnesiumjon i mitten och en sidokedja som kallas fytol. Klorofyll absorberar blått och rött ljus men inte grönt ljus och därav uppfattas växter som gröna. Detta är alltså den färg som reflekteras av klorofyllpigmentet. Klorofyll är involverat i den så viktiga fotosyntesen, som växter, men indirekt även övrigt liv på jorden, är beroende av. I fotosyntesen omvandlar växter vatten och koldioxid till kolhydrater och syre.

Mg-chelatase består av tre olika proteiner med olika storlek, BchI (40 kDa = litet), BchD (60 kDa = mellanstort) och BchH (130 kDa = stort). De tre proteinerna kan tillsammans genomföra det första enzymatiska steget i klorofyllsyntesen där startmaterialet heter protoporfyrin IX och produkten heter Mg-protoporfyrin IX.

Mg-chelatase sätter in en magnesiumjon i mitten av porfyrinen. I avhandlingen visas hur BchI och BchD sitter ihop i en dubbel ringformation med sex stycken enheter av BchI och BchD. För första gången visas också att BchH interagerar med BchD i ID-komplexet och inte med BchI.

MPE cyclase-komplexet består av minst tre stycken kärnkomponenter, varav endast en (Xantha-l) i nuläget är känd. Det enzymatiska steget som MPE cyclase-komplexet genomför är från Mg-protoporfyrin monometylester (MPE) till Mg-protoklorofyllide och är en ringslutning av den femte ringen i porfyrinen. Det är ett mycket komplicerat reaktionssteg och utöver de tre kärnkomponenterna behövs även andra proteiner. Två av dem presenteras i avhandlingen som NADPH-beroende tioredoxin reduktas C (NTRC) och 2-Cys peroxiredoxin. De två proteinerna kan tillsammans öka hastigheten på det enzymatiska steget som MPE cyclaset genomför. NTRC har också karakäriserats med avseende på sin tre-dimensionella struktur och då upptäcktes att NTRC bildar en oligomer bestående av fyra stycken proteiner som länkats samman.

Avhandlingen består av den sammanfattande delen som ”kappan” utgör (kapitel 1-6) samt fyra stycken artiklar bifogade därefter, som avhandlingen baseras på. Första artikeln (Paper I) är en strukturanalys av ID-komplexet i Mg-chelatase. Artikel nummer två (Paper II) är en studie av proteininteraktioner inom IDH-komplexet i Mg-chelatase. Nästa artikel (Paper III) handlar om NTRC och 2-Cys peroxiredoxin som visar sig vara viktiga för MPE cyclase-reaktionen. Sista artikeln (Paper IV) är en karaktärisering av NTRC med avseende på funktion och struktur.

Acknowledgements

References

Al-Karadaghi, S., R. Franco, et al. (2006). "Chelatases: distort to select?" Trends Biochem Sci **31**(3): 135-142.

Alawady, A., R. Reski, et al. (2005). "Cloning and expression of the tobacco CHLM sequence encoding Mg protoporphyrin IX methyltransferase and its interaction with Mg chelatase." Plant Mol Biol **57**(5): 679-691.

Axelsson, E., J. Lundqvist, et al. (2006). "Recessiveness and dominance in barley mutants deficient in Mg-chelatase subunit D, an AAA protein involved in chlorophyll biosynthesis." Plant Cell **18**(12): 3606-3616.

Bae, B., Y. H. Chen, et al. (2009). "Insights into the architecture of the replicative helicase from the structure of an archaeal MCM homolog." Structure **17**(2): 211-222.

Baier, M. and K. J. Dietz (1997). "The plant 2-Cys peroxiredoxin BAS1 is a nuclear-encoded chloroplast protein: its expressional regulation, phylogenetic origin, and implications for its specific physiological function in plants." Plant J **12**(1): 179-190.

Baier, M. and K. J. Dietz (1999). "Protective function of chloroplast 2-cysteine peroxiredoxin in photosynthesis. Evidence from transgenic Arabidopsis." Plant Physiol **119**(4): 1407-1414.

Baldwin, R. L. (2007). "Energetics of protein folding." J Mol Biol **371**(2): 283-301.

Balmer, Y., A. Koller, et al. (2003). "Proteomics gives insight into the regulatory function of chloroplast thioredoxins." Proc Natl Acad Sci U S A **100**(1): 370-375.

Bantscheff, M., M. Schirle, et al. (2007). "Quantitative mass spectrometry in proteomics: a critical review." Anal Bioanal Chem **389**(4): 1017-1031.

Beale, S. I. (1999). "Enzymes of chlorophyll biosynthesis." Photosynthesis Research **60**: 43-73.

Bergdoll, M., M. H. Remy, et al. (1997). "Proline-dependent oligomerization with arm exchange." Structure **5**(3): 391-401.

Berggard, T., S. Linse, et al. (2007). "Methods for the detection and analysis of protein-protein interactions." Proteomics **7**(16): 2833-2842.

Berthold, D. A., M. E. Andersson, et al. (2000). "New insight into the structure and function of the alternative oxidase." Biochim Biophys Acta **1460**(2-3): 241-254.

Block, M. A., A. K. Tewari, et al. (2002). "The plant S-adenosyl-L-methionine:Mg-protoporphyrin IX methyltransferase is located in both envelope and thylakoid chloroplast membranes." Eur J Biochem **269**(1): 240-248.

Bohn, S., F. Beck, et al. (2010). "Structure of the 26S proteasome from Schizosaccharomyces pombe at subnanometer resolution." Proc Natl Acad Sci U S A **107**(49): 20992-20997.

Bollivar, D. W. (2006). "Recent advances in chlorophyll biosynthesis." Photosynth Res **90**(2): 173-194.

Bonifacino, J. S., E. C. Dell'Angelica, et al. (2001). "Immunoprecipitation." Curr Protoc Mol Biol **Chapter 10**: Unit 10 16.

Ceol, A., A. Chatr-Aryamontri, et al. (2010). "MINT, the molecular interaction database: 2009 update." Nucleic Acids Res **38**(Database issue): D532-539.

Cesareni, G., A. Chatr-Aryamontri, et al. (2008). "Searching the MINT database for protein interaction information." Curr Protoc Bioinformatics **Chapter 8**: Unit 8 5.

Chatr-Aryamontri, A., A. Ceol, et al. (2008). "Protein interactions: integration leads to belief." Trends Biochem Sci **33**(6): 241-242; author reply 242-243.

Chatr-Aryamontri, A., A. Zanzoni, et al. (2008). "Searching the protein interaction space through the MINT database." Methods Mol Biol **484**: 305-317.

Chen, Z. A., A. Jawhari, et al. (2010). "Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry." EMBO J **29**(4): 717-726.

Cravatt, B. F., G. M. Simon, et al. (2007). "The biological impact of mass-spectrometry-based proteomics." Nature **450**(7172): 991-1000.

Davison, P. A., H. L. Schubert, et al. (2005). "Structural and biochemical characterization of Gun4 suggests a mechanism for its role in chlorophyll biosynthesis." Biochemistry **44**(21): 7603-7612.

De Las Rivas, J. and C. Fontanillo (2010). "Protein-protein interactions essentials: key concepts to building and analyzing interactome networks." PLoS Comput Biol **6**(6): e1000807.

Deane, C. M., L. Salwinski, et al. (2002). "Protein interactions: two methods for assessment of the reliability of high throughput observations." Mol Cell Proteomics **1**(5): 349-356.

Dennis, C. and C. Surridge (2000). "*Arabidopsis thaliana* genome. Introduction." Nature **408**(6814): 791.

Dietz, K. J. (2003). "Plant peroxiredoxins." Annu Rev Plant Biol **54**: 93-107.

Domon, B. and R. Aebersold (2006). "Mass spectrometry and protein analysis." Science **312**(5771): 212-217.

Dougan, D. A., A. Mogk, et al. (2002). "AAA+ proteins and substrate recognition, it all depends on their partner in crime." FEBS Lett **529**(1): 6-10.

Eckhardt, U., B. Grimm, et al. (2004). "Recent advances in chlorophyll biosynthesis and breakdown in higher plants." Plant Mol Biol **56**(1): 1-14.

Ellis, R. J. (2006). "Molecular chaperones: assisting assembly in addition to folding." Trends Biochem Sci **31**(7): 395-401.

Erzberger, J. P. and J. M. Berger (2006). "Evolutionary relationships and structural mechanisms of AAA+ proteins." Annu Rev Biophys Biomol Struct **35**: 93-114.

Fenn, J. B., M. Mann, et al. (1989). "Electrospray ionization for mass spectrometry of large biomolecules." Science **246**(4926): 64-71.

Fields, S. and O. Song (1989). "A novel genetic system to detect protein-protein interactions." Nature **340**(6230): 245-246.

Fodje, M. N., A. Hansson, et al. (2001). "Interplay between an AAA module and an integrin I domain may regulate the function of magnesium chelatase." J Mol Biol **311**(1): 111-122.

Gao, Q., S. Xue, et al. (2006). "Pro-CrossLink. Software tool for protein cross-linking and mass spectrometry." Anal Chem **78**(7): 2145-2149.

Gavin, A. C., K. Maeda, et al. (2011). "Recent advances in charting protein-protein interaction: mass spectrometry-based approaches." Curr Opin Biotechnol **22**(1): 42-49.

Gibson, L. C., R. D. Willows, et al. (1995). "Magnesium-protoporphyrin chelatase of Rhodobacter sphaeroides: reconstitution of activity by combining the products of the bchH, -I, and -D genes expressed in Escherichia coli." Proc Natl Acad Sci U S A **92**(6): 1941-1944.

Goff, S. A., D. Ricke, et al. (2002). "A draft sequence of the rice genome (Oryza sativa L. ssp. japonica)." Science **296**(5565): 92-100.

Gorchein, A. (1972). "Magnesium protoporphyrin chelatase activity in Rhodopseudomonas spheroides. Studies with whole cells." Biochem J **127**(1): 97-106.

Gough, S. P., B. O. Petersen, et al. (2000). "Anaerobic chlorophyll isocyclic ring formation in *Rhodobacter capsulatus* requires a cobalamin cofactor." Proc Natl Acad Sci U S A **97**(12): 6908-6913.

Gough, S. P., K. Rzeznicka, et al. (2007). "A new method for isolating physiologically active Mg-protoporphyrin monomethyl ester, the substrate of the cyclase enzyme of the chlorophyll biosynthetic pathway." Plant Physiol Biochem **45**(12): 932-936.

Hansson, A., C. G. Kannangara, et al. (1999). "Molecular basis for semidominance of missense mutations in the XANTHA-H (42-kDa) subunit of magnesium chelatase." Proc Natl Acad Sci U S A **96**(4): 1744-1749.

Hellstrand, E., B. Boland, et al. (2010). "Ab aggregation produces highly reproducible kinetic data and occurs by a two-phase process." ACS Chem Neurosci. **1**: 13-18.

Hinchigeri, S. B., B. Hundle, et al. (1997). "Demonstration that the BchH protein of *Rhodobacter capsulatus* activates S-adenosyl-L-methionine:magnesium protoporphyrin IX methyltransferase." FEBS Lett **407**(3): 337-342.

Ho, M. and H. Kim (2011). "Glypican-3: a new target for cancer immunotherapy." Eur J Cancer **47**(3): 333-338.

Hu, Q., R. J. Noll, et al. (2005). "The Orbitrap: a new mass spectrometer." J Mass Spectrom **40**(4): 430-443.

Hynes, R. O. (1992). "Integrins: versatility, modulation, and signaling in cell adhesion." Cell **69**(1): 11-25.

Hynes, R. O. (2002). "Integrins: bidirectional, allosteric signaling machines." Cell **110**(6): 673-687.

Ikegami, A., N. Yoshimura, et al. (2007). "The CHLI1 subunit of *Arabidopsis thaliana* magnesium chelatase is a target protein of the chloroplast thioredoxin." J Biol Chem **282**(27): 19282-19291.

Imhoff, J., H. Truper, et al. (1984). "Rearrangement of the Species and Genera of the Phototrophic ‘‘Purple Nonsulfur Bacteria”." International Journal of Systematic Bacteriology **34**(3): 340-343.

Iyer, L. M., D. D. Leipe, et al. (2004). "Evolutionary history and higher order classification of AAA+ ATPases." J Struct Biol **146**(1-2): 11-31.

Jensen, P. E., L. C. Gibson, et al. (1998). "Determinants of catalytic activity with the use of purified I, D and H subunits of the magnesium protoporphyrin IX chelatase from Synechocystis PCC6803." Biochem J **334 (Pt 2)**: 335-344.

Jensen, P. E., L. C. Gibson, et al. (1999). "ATPase activity associated with the magnesium-protoporphyrin IX chelatase enzyme of Synechocystis PCC6803: evidence for ATP hydrolysis during Mg2+ insertion, and the MgATP-dependent interaction of the ChlI and ChlD subunits." Biochem J **339 (Pt 1)**: 127-134.

Jin Lee, Y. (2008). "Mass spectrometric analysis of cross-linking sites for the structure of proteins and protein complexes." Mol Biosyst **4**(8): 816-823.

Kannagara CG, UC Vothknecht et al. (1997). "Magnesium chelatase association with ribosomes and mutant complementation studies identify barley subunits Xantha-G as a functional counter-part of Rhodobacter subunit BchD." Mol Gen Genet **254**: 85-92

Karas, M., U. Bachmann, et al. (1987). "Matrix-assisted ultraviolet laser desorption of non-volatile compounds." International Journal Of Mass Spectrometry and Ion Processes **78**: 53-68.

Karger, G. A., J. D. Reid, et al. (2001). "Characterization of the binding of deuteroporphyrin IX to the magnesium chelatase H subunit and spectroscopic properties of the complex." Biochemistry **40**(31): 9291-9299.

King, G. J., A. Jones, et al. (2008). "Identification of disulfide-containing chemical cross-links in proteins using MALDI-TOF/TOF-mass spectrometry." Anal Chem **80**(13): 5036-5043.

Kirstein, J., N. Moliere, et al. (2009). "Adapting the machine: adaptor proteins for Hsp100/Clp and AAA+ proteases." Nat Rev Microbiol **7**(8): 589-599.

Lambert, W., P. J. Koeck, et al. (2011). "Subunit arrangement in the dodecameric chloroplast small heat shock protein Hsp21." Protein Sci **20**(2): 291-301.

Larkin, R. M., J. M. Alonso, et al. (2003). "GUN4, a regulator of chlorophyll synthesis and intracellular signaling." Science **299**(5608): 902-906.

Leitner, A., T. Walzthoeni, et al. (2010). "Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics." Mol Cell Proteomics **9**(8): 1634-1649.

Lodish, H., A. Berk, et al. (2000). Molecular Cell Biology. New York, W. H. Freeman and Company.

MacArthur, M. W. and J. M. Thornton (1991). "Influence of proline residues on protein conformation." J Mol Biol **218**(2): 397-412.

Mackay, J. P., M. Sunde, et al. (2007). "Protein interactions: is seeing believing?" Trends Biochem Sci **32**(12): 530-531.

Marrs, B. (1981). "Mobilization of the genes for photosynthesis from Rhodopseudomonas capsulata by a promiscuous plasmid." J Bacteriol **146**(3): 1003-1012.

Mayer, M. P. (2010). "Gymnastics of molecular chaperones." Mol Cell **39**(3): 321-331.

Mayne, S. and H. Patterton (2011). "Bioinformatics tools for the structural elucidation of multi-subunit protein complexes by mass spectrometric analysis of protein-protein cross-links." Briefings in Bioinformatics.

McEwan, A. G. (1994). "Photosynthetic electron transport and anaerobic metabolism in purple non-sulfur phototrophic bacteria." Antonie Van Leeuwenhoek **66**(1-3): 151-164.

McNaught, A. and A. Wilkinson, Eds. (1997). IUPAC. Compendium of Chemical Terminology Oxford Blackwell Scientific Publications.

Meyer, J. (2008). "Iron-sulfur protein folds, iron-sulfur chemistry, and evolution." J Biol Inorg Chem **13**(2): 157-170.

Michalska, J., H. Zauber, et al. (2009). "NTRC links built-in thioredoxin to light and sucrose in regulating starch synthesis in chloroplasts and amyloplasts." Proc Natl Acad Sci U S A **106**(24): 9908-9913.

Mochizuki, N., R. Tanaka, et al. (2010). "The cell biology of tetrapyrroles: a life and death struggle." Trends Plant Sci **15**(9): 488-498.

Mogk, A., D. Dougan, et al. (2004). "Broad yet high substrate specificity: the challenge of AAA+ proteins." J Struct Biol **146**(1-2): 90-98.

Monney, L., C. A. Sabatos, et al. (2002). "Th1-specific cell surface protein Tim-3 regulates macrophage activation and severity of an autoimmune disease." Nature **415**(6871): 536-541.

Morrison, J. D. (1991). "Personal reminiscences of forty years of mass spectrometry in Australia." Organic Mass Spectrometry **26**: 183-194.

Neuwald, A. F., L. Aravind, et al. (1999). "AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes." Genome Res **9**(1): 27-43.

Nitschke, W. and A. W. Rutherford (1991). "Photosynthetic reaction centres: variations on a common structural theme?" Trends Biochem Sci **16**(7): 241-245.

Perez-Ruiz, J. M., M. Gonzalez, et al. (2009). "The quaternary structure of NADPH thioredoxin reductase C is redox-sensitive." Mol Plant **2**(3): 457-467.

Peri, S., H. Steen, et al. (2001). "GPMAW--a software tool for analyzing proteins and peptides." Trends Biochem Sci **26**(11): 687-689.

Rappsilber, J. (2011). "The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes." J Struct Biol **173**(3): 530-540.

Reid, J. D. and C. N. Hunter (2004). "Magnesium-dependent ATPase activity and cooperativity of magnesium chelatase from Synechocystis sp. PCC6803." J Biol Chem **279**(26): 26893-26899.

Reid, J. D., C. A. Siebert, et al. (2003). "The ATPase activity of the ChlI subunit of magnesium chelatase and formation of a heptameric AAA+ ring." Biochemistry **42**(22): 6912-6920.

Rigaut, G., A. Shevchenko, et al. (1999). "A generic protein purification method for protein complex characterization and proteome exploration." Nat Biotechnol **17**(10): 1030-1032.

Rinner, O., J. Seebacher, et al. (2008). "Identification of cross-linked peptides from large sequence databases." Nat Methods **5**(4): 315-318.

Roepstorff, P. and J. Fohlman (1984). "Proposal for a common nomenclature for sequence ions in mass spectra of peptides." Biomed Mass Spectrom **11**(11): 601.

Rzeznicka, K., C. J. Walker, et al. (2005). "*Xantha-l* encodes a membrane subunit of the aerobic Mg-protoporphyrin IX monomethyl ester cyclase involved in chlorophyll biosynthesis." Proc Natl Acad Sci U S A **102**(16): 5886-5891.

Sali, A., R. Glaeser, et al. (2003). "From words to literature in structural proteomics." Nature **422**(6928): 216-225.

Sauer, R. T. and T. A. Baker (2011). "AAA+ proteases: ATP-fueled machines of protein destruction." Annu Rev Biochem **80**: 587-612.

Sawicki, A. and R. D. Willows (2008). "Kinetic analyses of the magnesium chelatase provide insights into the mechanism, structure, and formation of the complex." J Biol Chem **283**(46): 31294-31302.

Sawicki, A. and R. D. Willows (2010). "BchJ and BchM interact in a 1 : 1 ratio with the magnesium chelatase BchH subunit of *Rhodobacter capsulatus*." FEBS J **277**(22): 4709-4721.

Schurmann, P. and B. B. Buchanan (2008). "The ferredoxin/thioredoxin system of oxygenic photosynthesis." Antioxid Redox Signal **10**(7): 1235-1274.

Scigelova, M. and A. Makarov (2006). "Orbitrap mass analyzer--overview and applications in proteomics." Proteomics **6 Suppl 2**: 16-21.

Seng, P., M. Drancourt, et al. (2009). "Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry." Clin Infect Dis **49**(4): 543-551.

Serrato, A. J., J. M. Perez-Ruiz, et al. (2002). "Cloning of thioredoxin h reductase and characterization of the thioredoxin reductase-thioredoxin h system from wheat." Biochem J **367**(Pt 2): 491-497.

Shepherd, M., S. McLean, et al. (2005). "Kinetic basis for linking the first two enzymes of chlorophyll biosynthesis." FEBS J **272**(17): 4532-4539.

Singh, P., A. Panchaud, et al. (2010). "Chemical cross-linking and mass spectrometry as a low-resolution protein structure determination technique." Anal Chem **82**(7): 2636-2642.

Sirijovski, N., J. Lundqvist, et al. (2008). "Substrate-binding model of the chlorophyll biosynthetic magnesium chelatase BchH subunit." J Biol Chem **283**(17): 11652-11660.

Sousa, D. and N. Grigorieff (2007). "Ab initio resolution measurement for single particle structures." J Struct Biol **157**(1): 201-210.

Springer, T. A. (2006). "Complement and the multifaceted functions of VWA and integrin I domains." Structure **14**(11): 1611-1616.

Sturm, N., E. Jortzik, et al. (2009). "Identification of proteins targeted by the thioredoxin superfamily in Plasmodium falciparum." PLoS Pathog **5**(4): e1000383.

Sueoka, K., T. Yamazaki, et al. (2009). "The NADPH thioredoxin reductase C functions as an electron donor to 2-Cys peroxiredoxin in a thermophilic cyanobacterium Thermosynechococcus elongatus BP-1." Biochem Biophys Res Commun **380**(3): 520-524.

Suter, B., S. Kittanakom, et al. (2008). "Two-hybrid technologies in proteomics research." Curr Opin Biotechnol **19**(4): 316-323.

Suzuki, J. Y. and C. E. Bauer (1995). "Altered monovinyl and divinyl protochlorophyllide pools in bchJ mutants of *Rhodobacter capsulatus*. Possible monovinyl substrate discrimination of light-independent protochlorophyllide reductase." J Biol Chem **270**(8): 3732-3740.

Tanaka, K., H. Waki, et al. (1988). "Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry." Rapid Commun Mass Spectrom **2**(8): 151-153.

Tanaka, R. and A. Tanaka (2007). "Tetrapyrrole biosynthesis in higher plants." Annu Rev Plant Biol **58**: 321-346.

The International Brachypodium Initiative (2010). "Genome sequencing and analysis of the model grass Brachypodium distachyon." Nature **463**(7282): 763-768.

Thulin, E., T. Kesvatera, et al. (2011). "Molecular determinants of S100B oligomer formation." PLoS One **6**(3): e14768.

Trost, P., S. Fermani, et al. (2006). "Thioredoxin-dependent regulation of photosynthetic glyceraldehyde-3-phosphate dehydrogenase: autonomous vs. CP12-dependent mechanisms." Photosynth Res **89**(2-3): 263-275.

Tswett, M. S. (1906). "Adsorptionanalyse und chromatographische Methode. Anwendung auf die Chemie des Chlorophylls. (Adsorption analysis and chromatographic method. Application to the chemistry of chlorophyll.)." Berichte der Deutschen botanischen Gesellschaft **24**: 384-393.

Tswett, M. S. (1906). "Physikalisch-Chemische Studien über das Chlorophyll. Die Adsorption. (Physical-chemical studies of chlorophyll. Adsorption.)." Berichte der Deutschen botanischen Gesellschaft, **24**: 316-326.

Vale, R. D. (2000). "AAA proteins. Lords of the ring." J Cell Biol **150**(1): F13-19.

Walker, C. J., K. E. Mansfield, et al. (1989). "Incorporation of atmospheric oxygen into the carbonyl functionality of the protochlorophyllide isocyclic ring." Biochem J **257**(2): 599-602.

Walker, C. J. and R. D. Willows (1997). "Mechanism and regulation of Mg-chelatase." Biochem J **327 (Pt 2)**: 321-333.

Walker, J. E., M. Saraste, et al. (1982). "Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold." EMBO J **1**(8): 945-951.

van Heel, M. and M. Schatz (2005). "Fourier shell correlation threshold criteria." J Struct Biol **151**(3): 250-262.

Wells, J. A. and C. L. McClendon (2007). "Reaching for high-hanging fruit in drug discovery at protein-protein interfaces." Nature **450**(7172): 1001-1009.

White, S. R. and B. Lauring (2007). "AAA+ ATPases: achieving diversity of function with conserved machinery." Traffic **8**(12): 1657-1667.

Willows, R. D., L. C. Gibson, et al. (1996). "Three separate proteins constitute the magnesium chelatase of Rhodobacter sphaeroides." Eur J Biochem **235**(1-2): 438-443.

Willows, R. D. and M. Hansson (2003). Mechanism, Structure, and Regulation of Magnesium Chelatase. The Porphyrin Handbook, Chlorophylls and Bilins: Biosynthesis, Synthesis, and Degradation, Elsevier Science (USA). **13**.

Wilm, M. (2009). "Quantitative proteomics in biological research." Proteomics **9**(20): 4590-4605.

Wollnik, H. (1993). "Time-of-flight mass analyzers." Mass Spectrometry Reviews **12**(2): 89-114.

von Wettstein, D., S. Gough, et al. (1995). "Chlorophyll Biosynthesis." Plant Cell **7**(7): 1039-1057.

Wood, Z. A., L. B. Poole, et al. (2003). "Peroxiredoxin evolution and the regulation of hydrogen peroxide signaling." Science **300**(5619): 650-653.

Xie, F., T. Liu, et al. (2011). "Liquid chromatography-mass spectrometry-based quantitative proteomics." J Biol Chem **286**(29): 25443-25449.

Yost, R. A. and C. G. Enke (1978). "Selected ion fragmentation with a tandem quadrupole mass spectrometer." Journal of American Chemical Society **100**: 2274-2275.

Zhang, Y., C. J. Ma, et al. (2011). "Tim-3 negatively regulates IL-12 expression by monocytes in HCV infection." PLoS One **6**(5): e19664.