Three dimensional structure of the light-harvesting chlorophyll *a/b* protein complex from plant chloroplasts

Dissertation
zur Erlangung des Doktorgrades der Naturwissenschaften

vorgelegt beim Fachbereich
Biochemie, Pharmazie und Lebensmittelchemie
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Matteo Lamborghini
aus Ferrara
Italien

Frankurt am Main, im Jahre 2002

Vom Fachbereich Biochemie, Pharmazie und Lebensmittelchemie der Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan:	Prof. Dr. Walter E. Müller
1. Gutachter:	Prof. Dr. Bernd Ludwig
2. Gutachter:	Prof. Dr. Werner Kühlbrandt

Table of contents

Abbreviations

ADA	N-(2-Acetamido)-2-iminodiacetic acid
HEPES	N-(2-hydroxyethyl)piperazine-N-(2-ethanesulfonicacid)
Tris	Tris-hydroxymethil)-aminomethane
Tricine	N-tris-(hydroxymethyl)-methylglycine
EDTA	Ethylenediaminetetraacetic Acid
PEG	Polyethylene glycol
DMSO	Dimethyl Sulfoxide
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
TX-100	Triton X-100
NG	n-Nonyl-β-D-glucoside
OG	n-Octyl-β-D-glucoside
MGDG	Monogalactosyl diacyl glycerol
SQDG	Sulphoquinovosyl diacyl glycerol
DGDG	Digalactosyl diacyl glycerol
PG	Phosphatidyl glycerol
PC	Phosphatidyl choline
TLC	Thin-layer chromatography
3D	Three-dimensional
2D	Two-dimensional
Chl	Chlorophyll
BChl	Bacteriochlorophyll
EM	Electron microscopy
bR	Bacteriorhodopsin

Abstract

The light-harvesting chlorophyll *a*/*b* protein complex (LHC-II) is the major collector of solar energy in all plants and it binds about half of the chlorophyll in green plants. LHC-II is a trimer in the photosynthetic membrane; each monomer consists of 232 amino acids, binds and orients a minimum of 12 chlorophyll molecules and three caroteinoids (two luteins and one neoxanthin) for light-harvesting and energy transfer.

Although, the structure of LHC-II has been determined at 3.4 Å resolution by electron microscopy of two-dimensional crystals (Kühlbrandt *et al.,* 1994), this is not sufficient to allow a complete understanding of the mechanism of energy transfer from LHC-II to the reaction centre, since the effective resolution in the *z* dimension is 4.9 Å. In fact, the chemical difference between Chl *a* and Chl *b*, which has a formyl group instead of the methyl group at the 7-position in the chlorin ring, is too small to be detected at this level of resolution. In addition, the orientation of the chlorophyll tetrapyrroles have not been determined unambiguously. This information is essential for a detailed understanding of the energy transfer within the complex and to the reaction centres of photosystem II and I (PSII and PSI).

X-ray crystallography of three dimensional (3D) crystals may yield a more complete structure at high resolution. 3D crystals have been grown from LHC-II isolated from pea leaves using a standard purification procedure (Burke *et al*., 1978). The thylakoid membranes are solubilised in Triton X-100 and further purified by sucrose gradient ultra centrifugation. The LHC-II fraction is salt precipitated and pellets resuspended at the chlorophyll a/b ratio 2.8 mg/ml in 0.9 % Nonyl-glucoside.

Crystals are currently obtained by vapour diffusion in hanging drops. These crystals are thin hexagonal plates, have a fairly large unit cell and diffract quite weakly. The high level of the background is due both to the detergent, necessary for protein solubilisation, and lipids, required for the trimer and crystals formation.

However, three data sets, each from one single crystal have been collected up to 3.2 Å resolution over a rotation range of 135°. The crystals were exposed to a very highly

collimated and brilliant beam (ID-14 EH1 at ESRF, Grenoble, France) and were kept under a stream of cold nitrogen to prevent radiation damage. Data were successfully integrated using the program XDS by Kabsch (1993). The crystals were found to belong to the space group $P6_322$ and have unit cell dimensions of a=128.45, b=128.45, c=135.32, $\alpha=\beta=90^{o}$, $\gamma=120$.

The solution of the phase problem was tackled by molecular replacement using, as a search model, the LHC-II structure solved by electron cryo-microscopy studies of two-dimensional crystals (Kühlbrandt *et al.* 1994).

Three different programs were tested: the most used AMoRe (Navaza *et al.*, 1994) and the brute force based program Brute (Fujinaga & Read 1987) did not found a correct solution. However, using a six-dimensional search method implemented in the program EPMR (Kissinger *et al.*, 1999) the final monomer orientation resulted to be correct with respect to the three-fold symmetry related monomers. Moreover, the resulting trimer was found to be the same as the model from the two-dimensional crystals analysis. Furthermore, crystals of LHC-II appear to be stacks of the two-dimensional crystals, therefore belonging to type I crystals, according the classification of Michel (Michel, 1983). The LHC-II crystals are stacked layers, which might not be precisely in register but translationally displaced with respect to one other, so that interference between layers may occasionally perturb the crystals symmetry.

The refinement procedure has been performed using the program package CNS 0.1 by Brünger *et al.* (1998). Data reduction resulted in 10727 independent X-ray reflections. The parameters to be refined are the three positional co-ordinates (x, y, z) and the temperature factor (B) for each of the 1952 atoms of the LHC-II monomer, making a total of almost 8000 parameters. Therefore, the ratio of observations to parameters is ~1.3, which is a poor over determination. At the current resolution, an R-factor lower than 30% cannot be expected, but on the other hand the R-free is still to high. Nevertheless, a preliminary fitting of new structural features was attempted. This showed some interesting new features of the LHC-II structure.

The α-helix–C is found to have an unwound region in the middle of the lipid bilayer. The biological reason for this structural feature is probably related to the arrangement and co-ordination of the chlorophyll molecules and a third carotenoid, which has been discovered, during this work, close to the α-helix C. This α-helix C also seems to be at least four amino acids longer with respect to the proposed fitting of the structure calculated by electron microscopy studies.

As a result of this thesis the phytyl chains have been assigned to eleven chlorophylls out of twelve. The chlorophyll chlorin rings are placed in two separate layers parallel to the membrane plane; one plane, build of seven chlorophylls, is closer to the stromal side, the other constituted of five chlorophylls is at the lumenal side. The phytyl chains are found to be in the space between these two layers, which is a highly hydrophobic region .

In addition, one carotenoid, most likely a neoxanthin molecule and not visible in the map derived from electron crystallography studies, has been fitted. This molecule appears to be close to the α-helix-C and almost perpendicular to the membrane plane. Moreover, it has been found to be in van der Waals distance to Chl *b*5 and Chl *b*6. This structural information is corroborated by recent single point mutation analysis of *in vitro* reconstituted LHC-II. These experiments performed by two different group showed that the binding sites of the Chl *b*5 and Chl *b*6 are linked to the neoxathin binding site (Rogl, 2000).

All these new features are of great interest. Indeed, the assignment of the phytyl chains will allow the calculation of the Q_x and Q_y dipole moments, and consequently the orientation of the chlorophylls with respect to the apoprotein, to each other and to the carotenoid molecules. These are essential information to perform further spectroscopic experiments and calculate a detailed LHC-II model in order to fully understand the energy transfer within the LHC-II complex and the path to the reaction centre of PSII and PSI.

Chapter 1

Introduction

1.1 Photosynthesis

Photosynthesis is the main source of energy for life in the biosphere. The conversion of light energy into chemical energy is catalysed by membrane-bound pigment protein complexes called photosynthetic reaction centres. Two different types of photosynthetic reaction centres are known: type I and type II. They are distinguished according to their terminal electron acceptors. If the terminal electron acceptor is an iron-sulphur cluster, the centre is classified as type I whereas, if it is a quinone, it is a type II reaction centre (Heathcote, 2001). Bacteria, carrying out anoxygenic photosynthesis, have only one reaction centre, while in the photosynthetic membrane of cyanobacteria, red, green algae and plants, both type I and type II reaction centres are present. They use the absorbed light energy, in form of photons, to translocate electrons along a chain of cofactors across the thylakoid membrane. This results in a transmembrane difference in the electrical potential and in proton (H^+) concentration, which drives ATP synthesis and reduction of $NADP^+$ to $NADPH$. In a consequent reaction, called the dark reaction, $NADPH$ and ATP are used to convert CO_2 to carbohydrates (Mitchell, 1976). In this process the fundamental role of gathering energy is played by the antenna complex which is a pigment binding protein.

1.2 Pigment molecules

The pigment molecules are responsible for the uptake of photons. The carotenoids are a group of polyunsaturated hydrocarbons having characteristic alternating single-double bonds (fig. 1.1). This conjugated π electron system is responsible for their absorption properties in the visible region of the solar spectrum. Among these are the xanthophylls lutein, neoxantin, violaxantin and zeaxantin. Chlorophylls are molecules formed by a cyclic tetrapyrrole, called chlorin, with a network of alternating double bonds and four central nitrogens, which coordinates a magnesium atom. A phytol chain is attached at the tetrapyrrole ring, which is responsible for the lipophilic nature of the chlorophyll. Higher plants contain only Chl *a* and Chl *b*. The difference between the two is a methyl group instead of a formyl group in position C-7. This chemical difference gives them a slightly different absorption spectrum. Indeed, Chl *a* shows a maximum absorption at 440 and 679 nm, while the Chl *b* has a maximum peak at 477 and 646 nm (Frank *et al*., 1999).

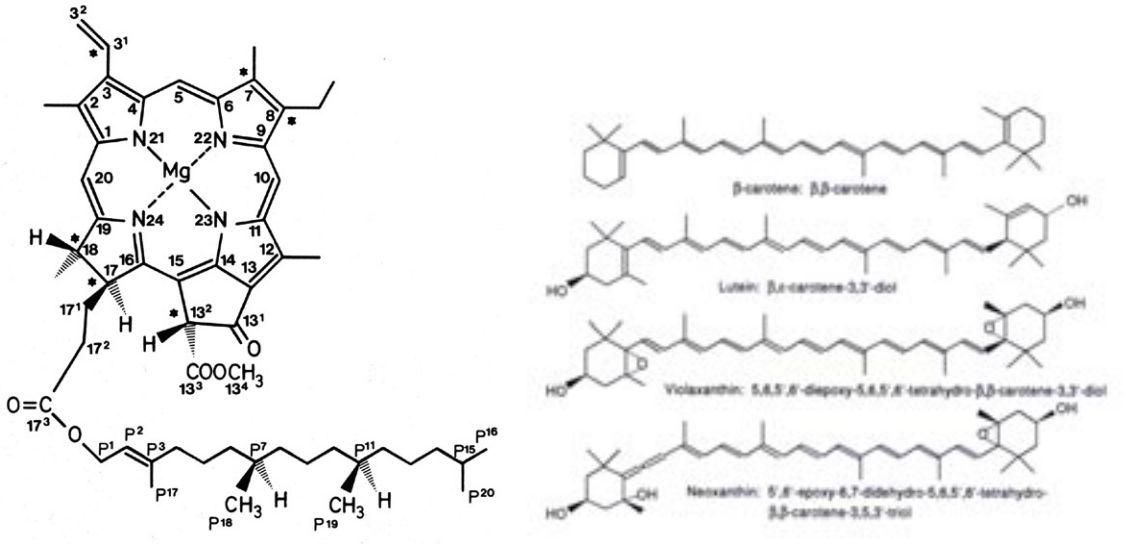


Fig. 1.1: Schematic drawing of chlorophyll a and xanthophylls molecules involved in energy absorption and transfer.

1.3 Mechanisms of oxygenic photosynthesis

The site of photosynthesis in eukaryotes (algae and higher plants) is the chloroplasts. Chloroplasts are organelles comprised of two envelope membranes, an aqueous matrix

called stroma and an internal membrane called thylakoid. All the light-harvesting and energy transducing functions are located in the thylakoid membrane, which encloses an aqueous compartment known as the thylakoid lumen. In higher plants the thylakoids are organised in stacked grana and non-stacked stroma membrane regions.

Photosynthesis occurs in two distinct phases: the light reaction, which uses light energy to generate NADPH and ATP, and the dark reaction, which is indeed a light-independent reaction, and uses NADPH and ATP to drive the synthesis of carbohydrates from CO_2 and H_20.

In plants and cyanobacteria the light reaction occurs in series following a so called "Z" scheme. In the thylakoid membrane the proteins, which are involved in the photosynthesis process are: the PSI and PSII and their relative minor antenna complexes, the major antenna chlorophyll a and b protein complex (LHC-II), the cytochrome b_6f protein complex and the ATP synthase (fig1.2).

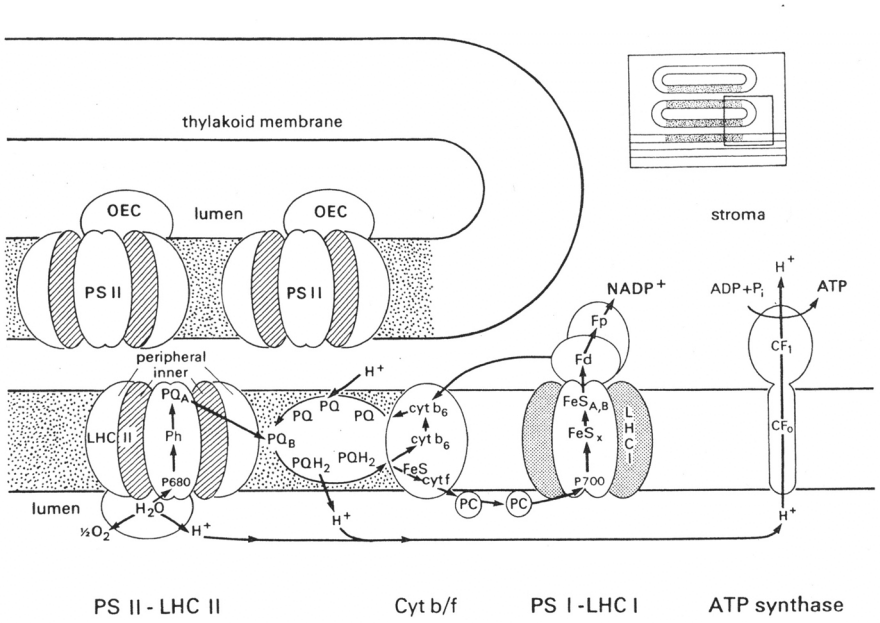


Fig. 1.2: Scematic representation of the thylakoids membrane organisation.

The conversion of light to chemical energy is associated with charge separation across the thylakoid membranes. The initial event is the absorption of a photon by the antenna system (LHC-II). The antenna chlorophyll molecules pass the energy of an absorbed photon by exciton transfer, from molecule to molecule until the photon reaches a

photosynthetic reaction centre. There the photon is trapped since reaction centre chlorophyll, even if are chemically identical to antenna chlorophyll have slightly lower excited state energies because of their different environment (van Grondelle, 1994). Once the photon reaches the PSII, there an electron is expelled from the primary electron donor, a chlorophyll a called $P680$. The primary donor is located near the lumenal side of the membrane, bound between the D1 and D2 subunits proteins. Once the $P680$ is oxidised to the $P680^{\bullet+}$ cationic radical, the electron moves to an electron acceptor. This is the primary acceptor, pheophytin (Ph). The electron then reduces a secondary and a tertiary acceptor. These are two molecules of plastoquinone: the Q_a and Q_b associated with D2 and D1, respectively. The plastoquinone Q_b can bind 2 electrons and 2 protons in his reduced state, called plastoquinol, which can freely diffuse within the membrane. Diffusing into the lipids bilayer the plastoquinol transfers electrons to the cytochrome $b_6 f$ complex. The $b_6 f$ transfers the electrons via the iron-sulphur binding protein (Rieske) to the copper protein called plastocyanin, thereby pumping 2 protons into the thylakoid lumen. At last, a plastocyanin passes two electrons to $P700$ or PSI reaction centre.

After primary charge separation initiated by excitation of the chlorohyll dimer $P700$, the electrons are moved along the electron transfer chain, which is formed by the Chl *a* (A_0) the phylloquinonen (A_1) and the three [4Fe4S] cluster referred to as F_x, F_a and F_b. Once the electron is at the stromal side, it is donated by F_b to the ferredoxin (flavodoxin) and finally transferred to the $NADP^+$ reductase. Completion of the reaction cycle is made by the re-reduction of the $P700^{\bullet+}$ by plastocyanin at the inner (lumenal) side of the membrane. The electron carried by the plastocyanin derives from the PSII reaction via a pool of plastoquinones and the cytochrome $b_6 f$ complex.

In PSII the $P680^{\bullet+}$, formed during charge separation, is a strong oxidant. This cationic radical $P680^{\bullet+}$ will regain the electrons from water through a complex containing a cluster of four Mn atoms. The Mn cluster undergoes several different steps of oxidation, in which it extract 4 electrons, 4 protons and 1 oxygen from 2 molecules of H_2O. Finally, the oxygen diffuses through the membrane out of the chloroplasts and is released into the atmosphere.

In order to better elucidate the electrons transfer chains across the thylakoids membrane it is relevant to know the exact position of the pigment molecules within the lipid bilayer, in a way that the electron path can be followed step by step. In the electron chain across the thylakoids membrane a central role is played by the reaction centre of the photosystem I and II and by the antenna complex. High resolution structural analysis provides the exact positions of pigment molecules; in this respect a great achievement has been reached since the structure of the PSI and PSII have been resolved at high resolution (Jordan *et al.*, 2001; Zouni *et al*., 2001).

1.4 Photosystem

Recently, high-resolution structures of photosystem I and photosystem II from cyanobacteria have been published, revealing new aspects in the molecular organisation of the pigment molecules coordinated by the apoprotein. The organisation of the chlorophyll molecules and of the others pigment molecules is of great intereset in order to understand the photons path from the antenna associated antenna system to the reaction centre.

Indeed, the work of Zouni *et al*. (2001) shows the structure of PSII isolated from the cyanobacterium *Synechococcus elongatus* at 3.8 Å resolution. The crystal contains 17 subunits including: the *psb*A and *psb*D, which are the reaction centre proteins D1 and D2 respectively; the inner antenna subunits, *psb*C and *psb*B, named CP43 and CP47 respectively, which bind chlorophyll *a* molecules and the α- and β-subunits of the cytochrome b-559, encoded by the *psb*E and *psb*F genes, respectively. The arrangement of the α-helices as well as the assignment of the subunits correspond to a model proposed from a projection map at 8 Å resolution (Rhee *et al*., 1998b) and confirmed by three dimensional analysis of two dimensional crystals at 8 Å (Rhee *et al*., 1998a). The PSII core complex is formed by the D1, D2, CP43, CP47 proteins and some smaller subunits encoded by *psb*H, *psb*I, *psb*J, *psb*K, *psb*L, *psb*M, *psb*N and *psb*X (for a review on the PSII binding proteins: Barber *et al.*, 1997). The PSII is a dimer in the crystals as

well as in the thylakoids membrane and the two monomers are arranged with a pseudo two fold rotation symmetry perpendicular to the membrane plane.

The crystal structure of PSI at 2.5 Å resolution (Jordan *et al*., 2001) provides a basis for understanding how the high efficiency of photosystem in light capturing and electron transfer is achieved (Fromme *et al*., 2001). PSI from *Synechococcus elongatus* consists of 12 individuals proteins (13 in plants) named according to their genes, *psa*A to *psa*X. PSI coordinates and binds 96 chlorophyll *a* molecules, 22 carotenoids, three [4Fe4S] clusters and two phylloquinones. At this high resolution structure three phospholipids and one galactolipid were revealed. In cyanobacteria, PSI is present as trimer and the internal three-fold rotation axis coincides with the crystallographic axis. One trimeric monomer is formed by a heterodimer of the *psa*A and *psa*B subunits. The complex is a trimer in the crystals as well as *in vivo* (fig. 1.3).

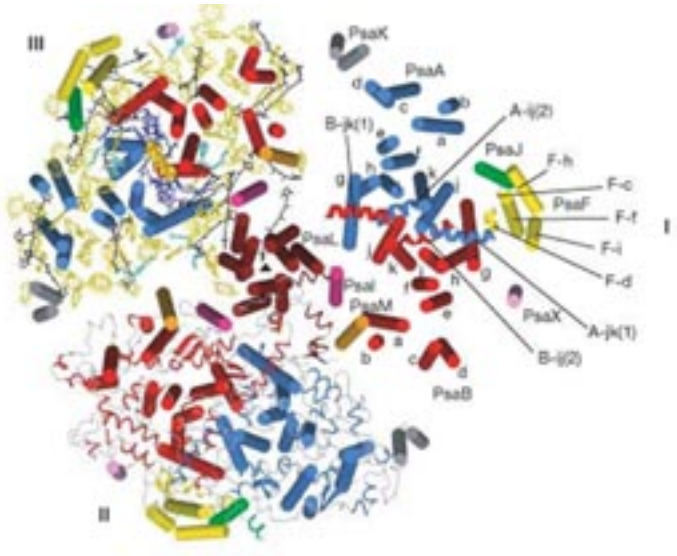


Fig. 1.3: Structural model of PSI viewed perpendicular to the membrane plane from the stroma side. The monomer are numbered I, II and III; arrangement of the transmembrane helices as cylinders. Chlorophylls in yellow, carotenoids in black and lipid in turquoise (Jordan *et al.*, 2001).

Looking at the high resolution structure it can be seen that the arrangement of the α-helices of D1 and D2 resembles that of sub-units L and M found in the reaction centre of purple bacteria (Deisenhofer *et al*., 1985) as well as the five carboxy-terminal helices of

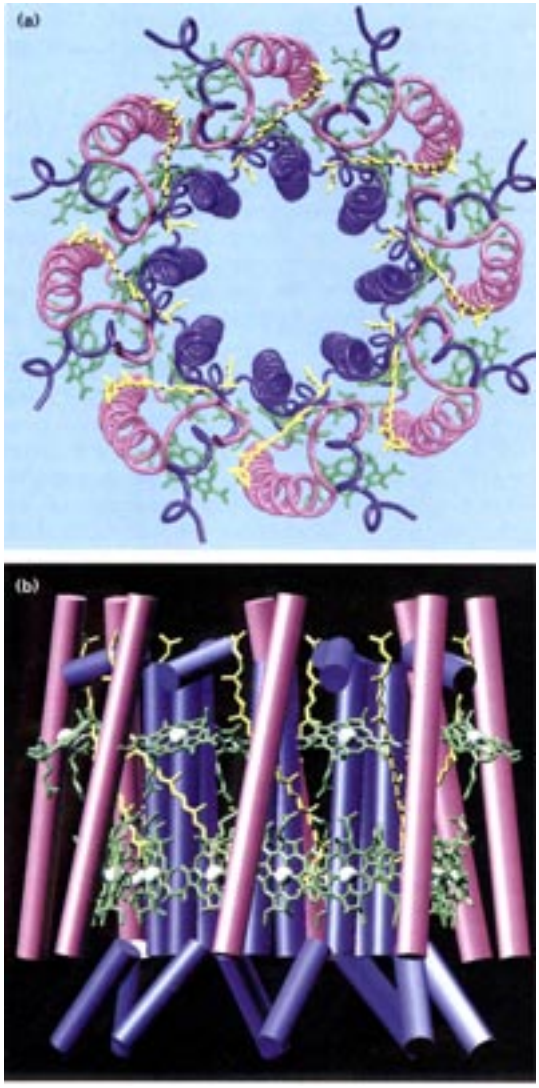
PsaA and PsaB in PSI. This supports the hypothesis of the evolution of the reaction centre from a common ancestor (Baymann *et al*., 2001).

1.5 Antenna systems

The main task of the antenna complexes is to absorb light energy and further funnel it to the photosynthetic reaction centre. Different organisms has been developed several forms of antenna systems during evolution. Some of the antenna system have been crystallised and the solution of the structure reveals many details about the function of these systems.

The light-harvesting in purple sulphur and non-sulphur bacteria occurs mainly via two different integral membrane antenna system: the B875 (LH1) and B800-850 (LH2), where the number relates to the maximum of light absorption (Codgell, 1985).
A low resolution projection map of two dimensional crystals of RC and LH1 isolated from *Rhodobacter (Rb.) sphaeroides* (Jungas *et al*., 1999) reveals a central position of the RC surrounded by a open circle of 12 $\alpha\beta$ trans-membrane subunits. This is comparable with a previous projection map of LH1 from *Rs. rubrum,* at 8.5 Å resolution (Karrasch *et al*., 1995), which presents a close circle of 16 subunits each composed of an $\alpha\beta$-heterodimer. The circle has an internal diameter dimension of 68 Å, which allows to accommodate the reaction centre inside the ring.
The high resolution X-ray structure of LH2 isolated from the purple non-sulphur photosynthetic bacteria *Rps. acidophila* has been solved at 2.5 Å resolution (McDermott *et al*., 1995). It consists of nine identical $\alpha\beta$-subunits arranged into a ring. Each subunit co-ordinates one bacteriochlorophyll (BChl) a800, two BChl a850 and one carotenoid, the rhodopin-glucoside. The chlorophylls are disposed in two distinct layers: one formed by B800 and the second by B850. The nine BChla 800 are located in the centre of the membrane, while the eighteen B850 are close to the external surface and are positioned perpendicular to the membrane plane.

A different antenna complex LH2 isolated from *Rs. molischianum* has been crystallised and the structure solved at 2.4 Å resolution (Koepke *et al*., 1996) fig. 1.4a, b.



The structural arrangement is similar to the LH2 of *Rps. acidophila*, however this complex shows a C8 symmetry ring, but similar to the previous structure the chlorophyll molecules are arranged into two rings. Sixteen B850 BChla almost perpendicular to the membrane plane and the eight B800 BChl almost parallel to the membrane plane. In this complex the main carotenoid is the lycopene, which extends in the lipids bilayer between the B800 and B850 rings.

Energy transfer occurs from the B800 ring to the BChla 850 ring where the excited state is delocalised over a large area. The energy is then available to be transferred to the LH1 complex B875 and finally to the reaction centre (Freer *et al*., 1996).

Fig. 1.4a, b: LH2 octameric complex a) view from above and b) side view. In green, bacteriochlorophyll; lycopene in yellow.

A different system has developed in cyanobacteria, blue-green algae, and the eukaryotic algae Rhodophyta (red algae) which contain phycobilisomes (PBS) as light-harvesting protein-pigment complexes. These supramolecular aggregates are attached on the outer surface of the thylakoids or photosynthetic membranes (fig. 1.5). Phycobilisomes absorb light in the range of 450 to 650 nm and are formed of two structural parts: the core and the rods. Both part resemble disc-like structures composed of phycobiliproteins, which themselves are heterodimers of A and B subunits. Both subunits bind covalently one or

more pigment molecules, the bilins, which are linear (open-chain) tetrapyrroles. Phycobiliprotein $\alpha\beta$-monomers are arranged around a 3-fold symmetry axis and form an ($\alpha\beta$)-trimer. Two trimers are associated face-to face forming an ($\alpha\beta$)6-hexamer, which is the functional unit in the native antenna. Organisms containing phycobiliproteins have the capability of altering the composition and the size of the phycobilisome antennas with respect to spectrum of incident sunlight. The crystallographic structure of phycobilisome antenna isolated from different cyanobacteria have been solved at atomic resolution (Brejc *et al*., 1995; Duerring *et al*., 1991).

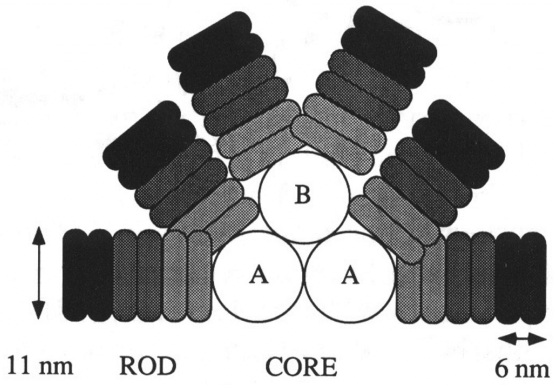


Fig. 1.5: Schematic representation of phycobilisome. Hemidiscoidal PBS with tricylindrical allophycocianin core. Six peripheral rods are normally bound to the core. In the rods, a hexameric phycocyanin complex is always located at the rod-core linkage position (Bryant, 1994).

Recently, it has been reported (Bibby *et al*., 2001; Boekema *et al*., 2001) that cyanobacteria that grow in iron-deficient environment express a gene which has a high amino acid sequence homology with the *psb*C (CP43) of photosystem II. This protein, called CP43', has an unknown function, however, forms a complex that consists of a ring of 18 CP43' molecules around a PSI trimer, increasing considerably the light-harvesting ability. Considering that the synthesis and assembly of phycobilisomes requires a great amount of iron, the expression of the CP43' in cyanobacteria is accompanied by an increased light-harvesting ability. It has been shown for the first time that a circular antenna organization around the reaction centre occurs in the cyanobacteria, which strikingly resembles the antenna system of anoxygenic purple photosynthetic bacteria.

The antenna complex of the green photosynthetic bacteria is located in unique extra-membrane cylindrical-shaped structures named chlorosomes (Staehelin *et al*., 1980) fig. 1.6. These are approximately 100 nm long and capture light very efficiently with some 200,000 bacteriochlorophyll (BChl) *c*, *d* or *e*-molecules per chlorosome. The energy transfer proceeds from the chlorosomes to the reaction centre via a so called base plate Fenna-Matthews-Olson BChlA (FMO) protein. This is a water soluble chlorophyll-containing protein, which was isolated from *Chlorobium limicola* and its three dimensional structure was solved by X-ray crystallography at 2.8 Å resolution (Fenna & Matthews, 1975). The FMO protein consist of three identical subunits each containing a core of seven bacteriochlorophyll molecules, which are irregularly arranged, however, enclosed within an envelope of protein. The bacteriochlorophyll porphirine rings lie in a plane that is almost parallel to the disk-shape formed by the trimer. These bacterial proteins function as an intermediate in the transfer of excitation energy from the antenna pigment to the reaction centre.

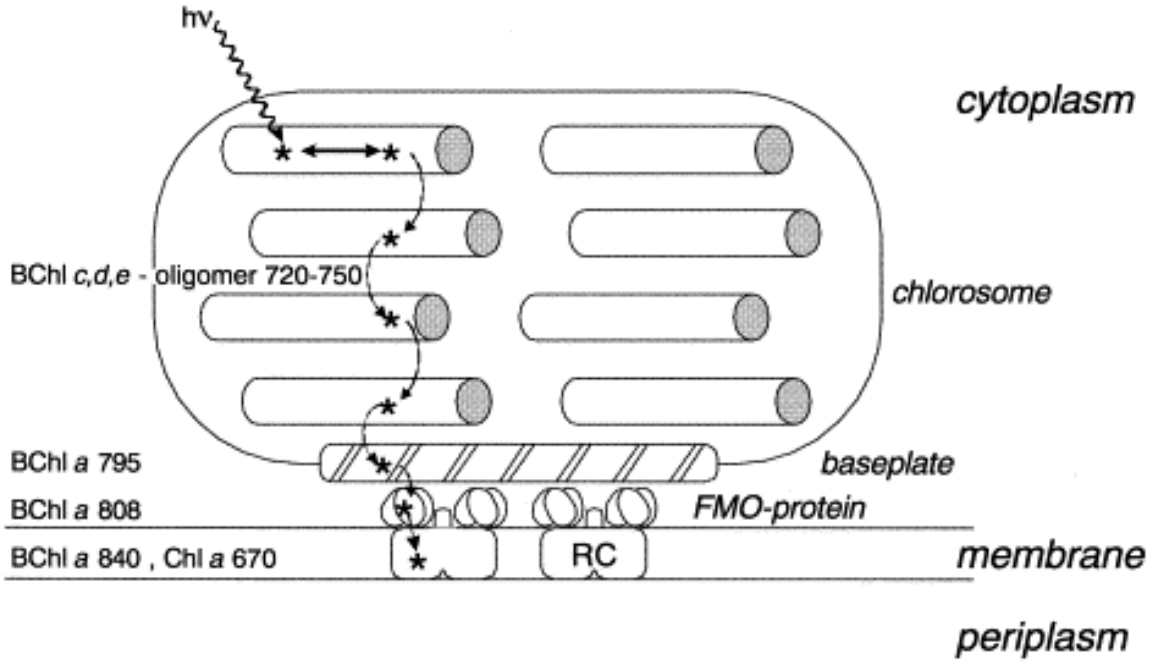


Fig. 1.6: The antenna system of green sulfur bacteria. The absorption maxima is reported for each pigment molecule (Hauska *et al*., 2001).

In addition to the membrane LH2, Dinoflagellates have developed a soluble antenna capable of light-harvesting in the blue–green range, between 470 to 550 nm. This is the

peridinin-chlorophyll-protein, which is a soluble boat-shaped antenna protein constituted of three domains (Hofmann *et al*., 1996). Each domain binds one chlorophyll a and four peridinin molecules.

In green plants and algae, the antenna function is accomplished by the light-harvesting chlorophyll *a/b* binding proteins (LHC). This is a group of intrinsic nuclear encoded thylakoid proteins. Among them the most abundant is the LHC-II or *Lhcb*1, which is responsible for the binding of almost 50 % of all chlorophyll on Earth. Before describing further structural details of the LHC-II, a complete, however short, description of all the chlorophyll *a/b* binding proteins (CAB) is reported here.

The light antenna complex of higher plants can be differentiated in peripheral and associated to the reaction centre. In plants the antenna apparatus of the reaction centre is formed by peripheral and closely associated antenna proteins. A list is given below and the nomenclature will be used according to Jansson (1994; 1999).
Some LHC proteins are part of PSI but most are part of PSII; the corresponding genes are named *Lhca* and *Lhcb*, respectively. They all have three hydrophobic regions spanning the thylakoid membrane, and the first and third helices are homologous to each other suggesting that they could have been derived from an internal gene duplication (Green & Kühlbrandt, 1995).
All the Chl *a/b* proteins (CAB) are structurally and evolutionary related showing a high level of homology. The conserved domains are related to the stress-inducible early light inducible proteins, ELIP, and the high-light inducible protein from cyanobacteria (Paulsen, 1995). All together the CAB proteins are ten different types, however, the strong similarity among them suggest a common ancestor. Indeed, it is supposed that the CAB protein, having three domain-spanning region, derived first from a duplication of a protein, which had two membrane spanning domains and a consequent deletion of the fourth domain. This hypothesis has become more convincing looking at the *psb*S protein (Funk *et al*., 1994). This is a four helix protein of which the function in the thylakoid membrane is not fully understood. It has been proposed that it is necessary for non-photochemical quenching even though it does not take part directly in light harvesting

and photosynthesis (Li *et al*., 2000). Moreover, this protein shows a high homology between helix one and three and conversely between helices two and four.

The LHCI is a group of protein associated with the PSI and their relative gene are named *Lhca*1-4. *Lhca*1 encodes a mature protein of 22 kDa, while *Lhca*4 encodes for 21 kDa protein (Jansson, 1994). They are associated together in the LHCI-730, which shows a maximum fluorescence emission at 730 nm, binds 14 chlorophyll molecules, 10 Chl *a* and 4 Chl *b* (Schmid *et al*., 1999) The *Lhca*2 correspond to a protein of 23 kDa and the *Lhca*3 gene to a 24 kDa polypeptide. These two latter proteins form the LHCI-680 complex, which exhibits maximum fluorescence at 680 nm (Jansson, 1994). At the moment no structural information is available about these two complexes, which serve as antenna for PSI and might additionally play an intermediate role in energy transfer from LHC-II to PSI.

*Lhcb*1, *Lhcb*2 and *Lhcb*3 are genes encoding the polypeptides of the major LHC-II, which is a trimer in the thylakoid membrane and in detergent solution. Due to the high numbers of *Lhcb* genes a clear stoichiometry and composition of the LHC-II in chloroplast membrane is not yet clear.

The LHC-II trimer also plays an important role in the energy distribution between PSII and PSI. It is thought that LHC-II, upon phosphorylation of a threonin residue close to the N-terminus, detaches itselves from PSII and attaches to PSI (Allen, 1992). Recently, other phosphorylation sites have been proposed (Dilly-Hartwig *et al*., 1998). In the thylakoid membrane the electron transport chain connect PSII and PSI, which need to have an equal rate of electron transport. Maximum efficiency is only achieved if PSII and PSI receive light energy at the same relative rate. Redox activation of LHC-II by a kinase protein balances the light-harvesting capacity of the two photosystems. If energy transfer in PSII happens at higher rate than in PSI, then plastoquinone is reduced: as a consequence the kinase is activated. Phosphorylated LHC-II transfers energy to PSI at damage of PSII, achieving an energy distribution balance. On the contrary whenever the platoquinone is oxidised the kinase protein is inactivated allowing dephosphorylation of

LHC-II. As such, the energy distribution between PSII and PSI is a self-regulated mechanisms.

The *Lhcb*4-6 gene products are usually referred to as CP29, CP26 and CP24, according to their apparent molecular weight (Bassi *et al*., 1990), which are probably monomeric forms and present in single copy per PSII unit (Jansson, 1994). Pigment stoichimetry reported for CP29 is of 6 Chl *a* and 2 Chl *b*, 1 lutein and 1 violaxantin molecules (Giuffra *et al.*, 1997); CP26 coordinates 6 Chl *a* and 3 Chl *b*, and 1 lutein and sub-stoichiometric amount of violaxantin (Bassi *et al.*, 1993). The smaller CP24 is found associated together with 5 Chl *a* and Chl *b*, 2 lutein and 1 violaxantin molecules (Sandona *et al.*, 1998). In all three proteins the neoxanthin is present in substoichiometric amount (less than 0.5).

1.6 Supramolecular organisation

A model for the photosynthetic apparatus of purple bacteria has been drawn from the high resolution structure of the bacterial reaction centre (Deisenhofer *et al*., 1985), the LH2 complex and a low resolution projection map of the LH1. In this presentation Papiz *et al*. (1996) proposed that, the RC is surrounded by the antenna complex LH1 or B875, which is in turn surrounded by several LH2 complexes forming densely packed and extended "lakes" of complexes, which are functionally connected, however these larger supra complexes patches have not been observed in detergent solubilised thylakoid membrane.

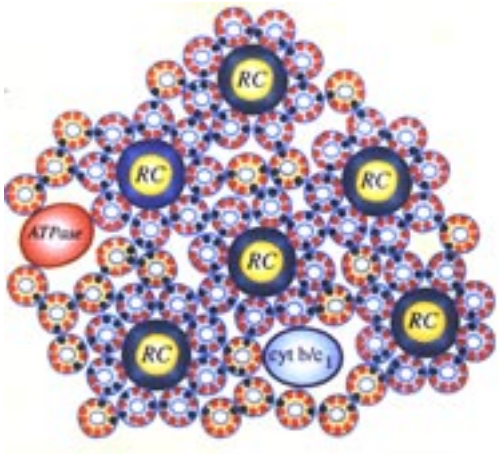


Fig. 1.7: A schematic representation of photosynthetic units embedded in patches of LH2 and other membrane proteins; RC, reaction centres (large yellow circle), LH1(blue rings) surrounded by LH2 (from Papiz *et al.*, 1996).

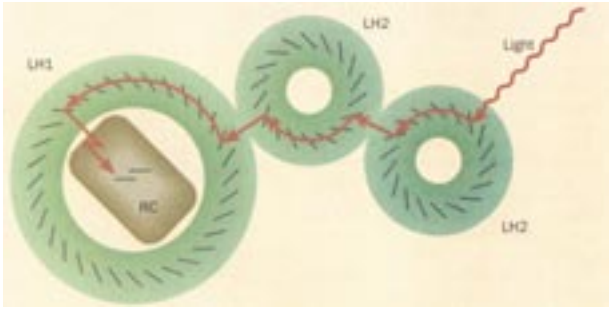


Fig. 1.8: LH2 absorbs light energy and passes to LH1; the final destination is the special pair of RC (from Kühlbrandt, 1995).

The solar energy is first collected by the LH2 complexes B800-850 and the transferred to the neighbour B875, which is strongly associated with the RC. It has been proposed that the densely packed B850 and B875 function as “storage rings” where delocalised energy is ready to be transferred from any point to the "special-pair" of the RC (Freer *et al*., 1996) (Kühlbrandt, 1995) (Fleming & van Grondelle, 1997).

The organisation of the plant photosynthetic supramolecular complex was reported at a resolution of 24 Å calculated by single particle analysis of images obtained by electron cryomicoscopy (Nield *et al*., 2000). This large multisubunit complex is a dimer and capable of light harvesting, as well as of water oxidation. Each monomer consists of the PSII core and the peripheral antenna proteins: a single monomer of the CP26 and CP 29 and the trimeric LHC-II (fig.1.1). The super complex isolated by mild detergent treatment of the thylakoid membrane retains the three major lumenal bound extrinsic protein of the OEC; the 33 kDa the 23 KDa and the 17 kDa proteins (Barber & Kühlbrandt, 1999).

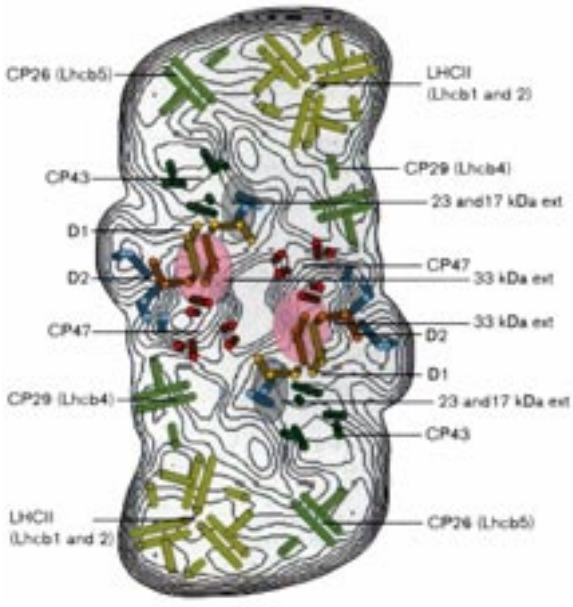


Fig. 1.9: Model the LHCII-PSII supercore complex organisation viewed from the lumenal side. Structure based on single particle analyses (from Barber & Kühlbrandt, 1999).

An independent study of (Boekema *et al*., 1999) showed a supramolecular organisation of the PSII complex, which is similar to the one in fig 1.9. However, their conclusion was that LHC-II can only be bound to the dimeric PSII core complex, if CP29 is bound. Furthermore they propose that CP29 is the first to bind at the PSII core complex, followed by CP26 and finally CP24, which is loosely bound to the super complex and not found in the sample of fig. 1.9. The technique used in these studies is of great interest in order to obtain information on the organisation of the peripheral antenna proteins in the PSII grana membranes. The overall organisation may elucidate the flow of the excitation energy form the peripheral antenna to the PSII reaction centre in the thylakoids membrane. Improved membrane isolation containing the mentioned PSII super-complex, the PSI and the cytochrome $b_6 f$ complex as well as the ATP synthase may yield a map of the *in vivo* situation.

1.7 LHC-II structure

The structure of LHC-II has been determined by electron cryomicroscopy of two dimensional crystals (Kühlbrandt *et al*., 1994). The resolution is 3.4 Å in plane, however only 4.9 Å in the direction perpendicular to the membrane layer or in other words along the *z* axis in the crystallographyc coordinate system.

The apoprotein is constituted of 232 amino acids and 36 % of these form the three major trans-membrane helices referred to as A, B and C fig. 1.10 (Kühlbrandt & Wang, 1991). The N-proximal helix is 51 Å long and is formed by 35 amino acids, which are assembled in 9.5 turns, and it will be referred to as B according to (Kühlbrandt & Wang, 1991). The second helix, named A, is 43 Å long and is formed of 30 residues, which span the lipid bilayer in 8 turns. Both helices A and B are tilted with 32º respect to the membrane normal, forming a coil-coiled region in the centre of the membrane. The maximal distance between the helices is 8.5 Å. A third helix C is 31 Å long, formed by 21 residues which make 5.5 turns. This helix is tilted 9º with respect to the membrane normal. A fourth amphipathic helix is present near the C-terminus domain and is named D. It is embedded in the lumenal side of the lipid membrane, formed by 8 residues,

which run in almost three turns parallel to the membrane plane. Remarkable is the two-fold internal symmetry between helices A and B with respect to the first 24 amino acids, eg the ion pairs Arg 70 and Arg 185, Glu 65 and Glu 180. These two ion pairs are thought to stabilise the transmembrane region of the proteins: the helices are held together by two strong ionic interactions as the positive Arg side chain extends from the centre of each helix near to the negative charged Glu side chain of the opposite helix.

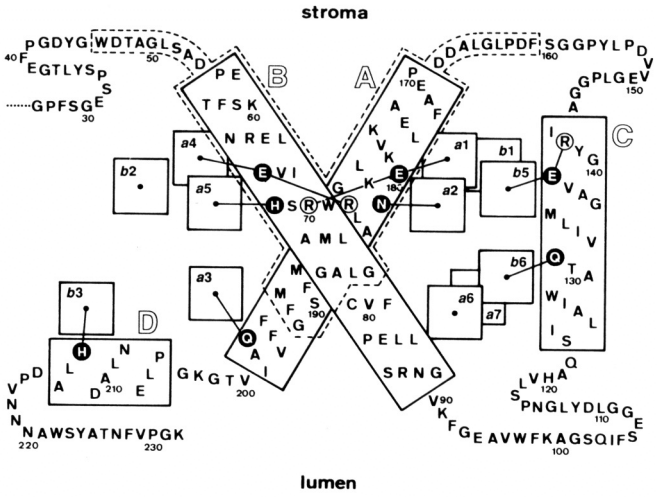


Fig. 1.10: Plan of the LHC-II polypeptide as in Kühlbandt *et al.* (1994).

The extramembranous domains constitute almost 65% of the whole apoprotein: the C- and N-termini and the loops connecting the four trans-membrane helices. The electron density relative to the region of the membrane surface was not as clear as the trans membrane region. However, in the EM map the loop between helices A and C is completely fitted into the electron density. On the other hand, the loop connecting helices B to C was tentatively fitted, especially between residues Ala 100 and Pro 116 (Kühlbrandt *et al*., 1994).

Moreover, the first 25 amino acids could not be resolved and the N-terminus is known to play a most important function for the complex trimerisation. Between residues position 17 to 22, the motif WYGPDR has been proposed to be the binding site of the lipid phosphatidyl glycerol (PG) (Hobe *et al.*, 1995), which is known to be essential for the trimerisation of LHC-II. This observation is corroborated by the recent work of Veverka

et al. (2000), which shows that a peptide containing this same motive is sufficient to bind the lipid molecule PG.

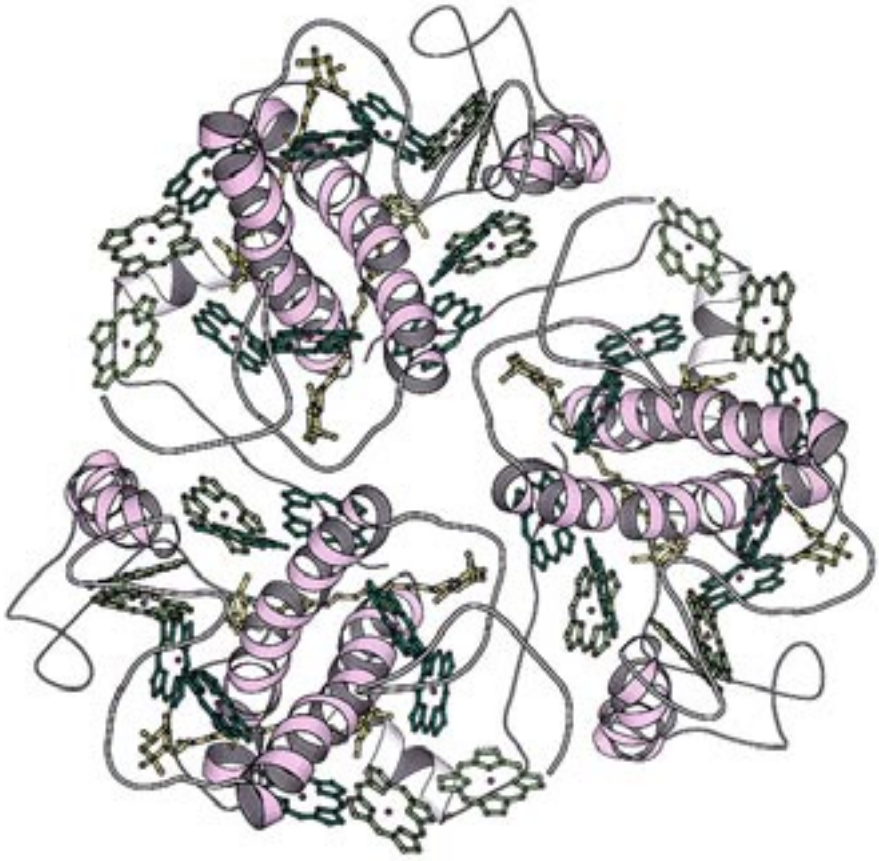


Fig. 1.11: Structure of the LHC-II trimer. Colour code: dark green, Chl *a*; light-green Chl *b*; luteins yellow. Viewed from the stroma side (Kühlbrandt, 1994)

In the EM map only twelve chlorophyll molecules and two xanthophyll molecules are found per each LHC-II monomer. On the other hand, chromatographic analysis of LHC-II usually detected between 12 to 15 chlorophyll molecules and three carotenoid molecules per monomer of LHC-II. Additionally, the stoichiometry of chl *a* and *b* is not unambiguously determined, it is usually reported to be between 1.0 and 1.3 depending on the purification conditions of the trimer. Usually 7 to 8 chlorophyll a and 5 to 7 chlorophyll *b* are found per LHC-II monomer (Dainese & Bassi, 1991; Kühlbrandt & Wang). The chlorophylls are not covalently bound to the complex, usually the fifth

ligand to the central Mg of the chlorin ring is provided by the side chain of Glu, as well as His or Gln residues. Nevertheless, in the present structure for some chlorophylls the ligands are not defined. One possibility is that these chlorophyll molecules are held in place by hydrogen-bonds or by carbonyl backbone atoms to neighbouring chlorophyll molecules (Kühlbrandt & Wang, 1991). The recent atomic resolution of PSI showed that the Mg ions of the chlorophylls are coordinated either directly by atoms of amino acid side chains (like the His, Gln, Asp, Glu or Tyr), by the peptide carbonyl oxygen or indirectly by water molecules; as well it was found out that a lipid molecule, PG, provides the fifth ligand of the Mg atoms of a Chl a molecule (Fromme *et al.*, 2001). In LHC-II the chlorophylls are organised into two layers parallel to the membrane plane, one is more proximal to the lumenal side, while the other to the stroma side; the chlorin rings are almost perpendicular to the membrane plane. The chlorophylls that are not evident in the structure could be situated in the polypeptide periphery and so removed during the LHC-II isolation procedure.

The electron density map reveals two 30 Å long region densities, which have been assigned to two lutein molecules, the most abundant carotenoid component in the LHC-II. These lie parallel to the central double helices and are positioned in the groove formed by the coil-coiled regions, which has a very hydrophobic character. The minimal distance between them is at the centre of the lipid bilayer and equal to 11 Å. The head groups are at almost equal distance to both the lumenal and stromal side of the membrane. The two luteins form a internal cross-brace in the centre of the complex, which provides a strong link between the peptide loops at both surfaces.

However, the identity of the chlorophyll molecules is not certain, since the diffracting resolution of the crystal is not high enough to identify the difference between Chl *a* and Chl *b*, moreover the carotenoids were assigned to lutein according to the biochemical characterisation of the pigments composition. Therefore, the chlorophylls were assigned according to the fact that the energy absorbed by a Chl *b* is transferred to Chl *a* within less that 1 ps where it stays for 1 to 3 ns, meaning that energy transfer from Chl *b* to Chl *a* is much more rapid than triple formation in the chlorophylls. Therefore triplet

quenching is required only for Chl *a*. Thus, the seven chlorophylls closest to the xanthophylls were designated to Chl *a*, while the others were assigned to Chl *b*.

Whenever an energy quanta is absorbed by a carotenoid or a chlorophyll molecule, they reach an excited state. Conversion of the energy will bring the excited molecule to its ground level energy state. This can be achieved by emission of energy in form of fluorescence or by heat dispersion, or the energy quanta can be transferred to the next neighbouring molecule. In antenna system the distance and the orientation between the pigment molecules is optimal to allow ultra fast energy transfer (Fleming & van Grondelle, 1997). Indeed each of the assigned Chl *b* is in van der Waals contact with each of the Chl *a*. The calculated energy transfer between them is of the magnitude order of the 200 fs (Kühlbrandt *et al*., 1994).

1.8 The crystallisation experiment

Many advances in our understand of the biological system at the molecular level have been made possible through knowledge of the detailed atomic structure. A fundamental requisite is the production of crystals of suitable size and order for structure determination. Even though the recent increased availability of powerful X-ray sources, fast data collection instruments and computational systems have allowed to solve many macromolecule structures, obtaining large, well-ordered single crystals, is still the rate-limiting step. Protein crystallisation is often a trial-and-error procedure where the task of the crystallographer is to bring about super-saturation a solution containing the macromolecule of interest, in such a way that well-ordered crystal will form.

In spite of the large number of crystals of soluble proteins that are currently available, membrane proteins crystals, grown from detergent solutions, are not routinely obtained. It is necessary to remove membrane protein from their native state by detergent solubilisation, however in this process the proteins will be exposed to solution condition

which have different properties from their native state. Moreover, the length of the detergent hydrocarbon tail influences the crystal lattice: in fact, a short tail, forming a small detergent micelle, favours the hydrophilic protein-protein interactions. On the other hand, these detergents tend to have a more aggressive behaviour affecting the stability of sensible proteins and increasing difficulties in finding successful crystallisation conditions (Garavito *et al.*, 1986; Kühlbrandt, 1988).

Regardless of the technique used the important requirement for successful crystallisation are purity, stability and homogeneity of the protein. Crystals form because of interaction between molecules. Proteins having a small hydrophilic part are suitable samples for two dimensional crystallisation. Indeed, the hydrophobic interactions among the protein in the plane of the membrane are favourable and the crystals array resemble the protein organisation in the lipid bilayer. Such an organisation has been studied in the purple membrane of the *Halobacterium salinarum*, which are formed of large patches of bacterio-rhodopsin (Henderson *et al*., 1990); a second example is LHC-II isolated from thylakoid membrane of plant chloroplasts .

The three dimensional structure of the protein is achieved by electron microscopy and image analysis. However, by using this technique is not straightforward to obtain atomic resolution information. Indeed, there is a systematic absence of information, the missing cone, due to the difficulty in tilting the grid speciment up to 90º in the electron-microscope. Consequently, the structural information tend to be anisotropic. Nevertheless, the power of electron microscopy relies on the relative small amount of pure sample required to obtain crystals. Moreover, it is an effective approach when large complexes should be studied. For instance, mild detergent solubilised thylakoid membranes yields a complex which contains the PSII reaction centre and several associated minor and major antenna proteins. This only approach provide a descriptive picture close to the real situation in the thylakoid membranes (Nield *et al.*, 2000).

There are two types of tree dimensional crystals, which are classified as type I and type II (Michel, 1983). The first type is a result of stacks of two dimensional arrays. In this case it seems difficult that this crystal would pack properly in a third dimension, since the forces which favour hydrophobic interaction among the neighbouring protein destabilises

the hydrophilic interaction between consecutive layers. Membrane proteins, which have large extra membrane parts are most likely of forming highly ordered three dimensional crystal of the type II. The hydrophobic part of membrane protein is bound by a detergent belt, which might cause a stereo impediment to the attachment of the next consecutive molecule. In order to favour the crystallisation of reluctant membrane protein it is possible to alter and modify the protein to be crystallised. For instance, by cleavage of extensive flexible loop, by single point mutation of amino acids which could favour some protein-protein contacts; by forming fusion of protein by co-crystallisation with inhibitors or antibodies. For instance, antibodies in form of the Fab or of the Fv fragment are attached to the hydrophilic part of the molecule in order to augment this portion, favouring the crystals contacts and hopefully the formation of a regular three dimensional array. Indeed antibodies may form a stable and rigid interaction with the antigen and act as a spacer for the detergent belt micelle around the hydrophobic protein portion (Ostermeier *et al*., 1995).

Recent example of this successful approach is the structure determination at 2.0 Å of the K^+-channel. In this structure, all protein contacts within the crystals are formed between neighbouring Fab fragments, in a manner that the K^+-channel and its detergent belt are not involved in crystals contacts. Another example of the efficacy of this methodology is the structure of the cytochrome *bc*1 complex from yeast (Hunte *et al.*, 2000). In these crystals the Fv fragment mediates the contact between two consecutive extra membrane portion of the protein complex.

A detailed structural analysis may give information about biological mechanisms like for example, proton translocation or electron transfer across lipid membrane. For instance several structures have been reported on bacterio-rhodopsin (bR), which is a proton pump light-driven integral membrane protein of a bacteria. This protein has been extensively studied both by electron microscopy (Henderson *et al.*, 1990; Grigorieff *et al.*, 1996) and more recently by X-ray studies (Luecke *et al.*, 1999). It has been possible to identify the amino acid involved in the proton transport across the membrane.

Last year has seen the determination of several membrane proteins structures solved at high resolution by X-ray of three dimensional crystals (Hunte *et al.*, 2000: Lancaster *et*

al., Zhou, Y. *et al.*, 2001). Indeed, even if the crystallisation procedure of membrane proteins has not become a routine process and each problem has to find its own solution, it seems that the methodology and the approach is clearer and hopefully will provide insights into complex biological processes.

1.9 The task of the project

Task of this work has been the production of LHC-II three dimensional crystals of size and order suitable for high resolution X-ray diffraction analysis. LHC-II is the main antenna protein complex of the thylakoid membrane, which is isolated from plant chloroplasts and almost the most abundant membrane protein on earth.

A structure of LHC-II was resolved at 3.4 Å resolution derived by electron cryo-microscopy studies of two dimensional crystals (Kühlbrandt *et al.*, 1994). The general organisation of the protein complex and the position of the pigment molecules was discovered, however, some questions are not yet resolved. Indeed, the current structure does not distinguish between all the chlorophyll molecules identities coordinated by the apoproteins. In fact, the difference between Chl *a* and Chl *b*, a methyl group instead of a formyl group at the 7-position of the chlorin ring, is too small to be resolved at this level of resolution. Moreover, the chlorophyll molecules are oriented in the lipid bilayer according to their phytyl chains, which strongly anchor the chlorin rings in the lipid bilayer. The phytyl chains are not assigned yet, but are essential in order to attribute dipole moments to the chlorophylls , and finally their orientation inside the complex. The number of chlorophyll molecules is not clearly determined: indeed it has been found to be between 12 and 15 (Butler & Kühlbrandt, 1988). This inconsistency is mainly due to different purification protocols, however, in the EM model 3.4 Å 12 chlorophyll have been attributed. Conversely, the Chl a/b ratio of the complex ranges between 1.2-1.4 depending on the isolation procedure and as well on the methodology utilised to determine the chlorophyll content. In addition, a recent work based on single point

mutation analysis has shown that a chlorophyll molecule identified as a Chl *b* could indeed be a Chl *a* (Rogl, 2000), as well other three chlorophyll molecules assigned to Chls a should be confirmed by high resolution structure analysis.

Moreover, biochemical analysis always indicated that LHC-II monomers contain three or four carotenoid molecules: two luteins, a nexoxanthin and other in sub-stoichiometric amounts. By EM studies two luteins molecules were identified in the groove of the coil-coiled formed by α-helices A and B. These two luteins play as quencher of harmful Chls triplet and are as well essential for the structural stability of the monomer. Apparently the neoxanthin does not have any structural function, as proved by refolding studies, indeed only two luteins are necessary to refold a structural intact monomers and hence trimers. However its position in the monomer respect to the neighbour chlorophyll should be indentified.

The LHC-II binds several lipids molecules: among them PG plays an essential role, being necessary to form stable and intact trimer and it is known to bind at the N-terminus. On other essential lipid is DGDG, which is bound at the periphery of the monomer and is necessary to form three dimensional crystals. High resolution data will eventually reveal the presence and ordered position of these lipids. Moreover in the EM model the first 25 amino acids at the N-terminus are not assigned, this portion is well known as well to be phosphorylated by an activated mechanism which acts as a switch and move the LHC-II complex closer to the PSI within the thylakoid membranes.

This work aims at discovering the detailed structure of the LHC-II isolated from pea chloroplasts following a standard procedure, which was already successfully utilised for the formation of two dimensional crystals. In order to understand the ultra fast energy transfer between the pigment molecules in LHC-II, the unambiguous assignment of Chl *a* or Chl *b*, their orientation in the lipid bilayer and the structural relation among them and between the two luteins and the third undiscovered carotenoids is required: with the hope to obtain such information, a 3D crystallisation trial was set up!

Chapter 2

Materials and Methods

2.1 Biochemical methods

Pea plants (*Pisum sativum*) were grown for 15 to 18 days under a constant light of 1.21×10^3 Lux, in a period of ten hours light and fourteen dark in a growth room, where temperature and humidity were set at 20°C and 45% respectively. Pea seeds were sowed on a substrate of vermiculite.

2.1.1 Isolation of thylakoids membranes

Thylakoid membranes were isolated as described in (Burke *et al*., 1978) and modified by (Kühlbrandt *et al*., 1983) from pea leaves.

Leaves were ground in a Warring blender in 33 mM sorbitol and 10 mM tricine at pH 7.8 (100 ml buffer for 35 gr of leaves). The leaves extract was filtered (8 layers of cheese cloth, 1 layer cotton, 2 layers cheese cloths) and then centrifuged (15 min at 6000 rpm, rotor Beckman, JA-14). The pellets, consisting of chloroplasts and cell debris, were re-suspended in 10 mM NaCl and 1 mM tricine (pH 7.8) and washed by centrifugation. Chloroplasts were then incubated at 4° C for 30 min in 10 mM NaCl, 1 mM tricine (pH 7.8) and 5 mM EDTA, to be disrupted by osmotic shock. The suspension was homogenised and washed in the same buffer. To de-stack the thylakoids, chloroplast membranes were suspended in 5 mM EDTA, 100 mM sorbitol (pH 7.8) and incubated at pH 6.0 for 15 min at room temperature. The membranes

were centrifuged (15 min at 6000 rpm, rotor Beckman JA-14) and re-suspended in 100 mM sorbitol.

2.1.2 Isolation of trimeric LHC-II

Thylakoid membranes in sorbitol solution were centrifuged (15 min at 6000 rpm, rotor Beckman JA-14) and the pellets re-suspended in water, homogenised and adjusted to a final chlorophyll concentration of 0.5 mg/ml. The detergent Triton X-100 (Calbiochem) at 0.5% w/v, was added and left for 30 min at 20 °C; then the protein solution was centrifuged for 20 min at 18000 rpm (rotor Beckman, JA-30.50 Ti). The solubilised proteins were loaded onto a linear sucrose density gradient (0.1 M-1.0 M sucrose) and subjected to ultra-centrifugation (18 hours, 38K rpm, Beckman, Ti-45). A red fluorescent and dark green band corresponding to the LHC-II was removed from the sucrose gradient and incubated with 300 mM KCl for 30 min at 4 °C. The suspension was centrifuged (15 min at 18000 rpm, Beckman, JA-30.50 Ti) and pellets washed twice in 100 mM KCl, and once with distilled water. The pellets were re-suspended in detergent solution at a chlorophyll concentration of 3.0 mg/ml, aliquoted, flash frozen in liquid nitrogen and stored at –80 °C. A list of the detergent used is given below (Table 2.1):

Detergents	CMC (mM)	Concentration (mM)
n-Hexyl-β-D-glucoside	250	270
n-Heptyl-β-D-glucoside	70	84
n-Octyl-β-D-glucoside	19	35
n-Nonyl-β-D-glucoside	6.5	30
n-Decyl-β-D-glucoside	2.2	6.6
n-Dodecyl-β-D-glucoside	0.19	2.5
n-Heptyl-β-D-thioglucoside	29	35
n-Nonyl-β-D-thioglucoside	2.9	6.5
n-Decyl-β-D-maltoside	1.8	5.6
n-Dodecyl-β-D-maltoside	0.17	1.2

This isolated trimeric LHC-II was both used directly to set crystallisation experiments or further purified by ion exchange chromatography.

2.1.3 Anion Exchange chromatography

After the sucrose gradient separation, trimeric LHC-II was precipitated in 300 mM KCl, washed twice in water and then suspended in 86 mM octyl-glucoside (OG) at a concentration of Chl [A+B]= 3.0 mg/ml. A total volume of 5 ml sample was loaded onto a Mono-Q ion exchange column HR16/10 (Pharmacia-Amersham) connected to an FPLC system (Pharmacia). Previously the column had been conditioned with buffer A. The sample was eluted, on a linear gradient, utilising buffer B up to a final concentration of 500 mM NaCl and a total volume of 200 ml. The elution speed was 0.5 ml/min and the experiment was performed at room temperature. The elution profile was detected at three different wavelengths: for the apo-protein λ=280 nm and λ=646.6 nm and λ=663.6 nm for Chl *a* and Chl *b*, respectively.

Buffer A
Tris-HCl 20 mM pH 7.5
NaCl 40 mM
Octyl-glucoside 26 mM

Buffer B
Tris-HCl 20 mM pH 7.5
NaCl 500 mM
Octyl-glucoside 26 mM

The eluted sample fractions were collected and concentrated to a final concentration of Chl [A+B]=3.0 mg/ml, using a 30 kDa MWCO filters (Centriprep and Centricon, Amicon).
In a parallel experiment the eluted fractions were diluted five fold and incubated with 300 mM KCl in a cold room for 30 min. The precipitated protein was washed twice by centrifugation with water and the pellets finally re-suspended either in 30 mM

nonyl-glucoside (NG) or 86 mM OG at the Chl [A+B]=3.0 mg/ml. Samples were flash frozen and stored at –80 ºC.

2.1.4 SDS-PAGE

Protein purity was checked by SDS-PAGE performed according to the system developed by Fling and Gregerson (1986). The acrylamide concentration used was 15% (v/v). The protein, diluted in the sample buffer, was incubated for 5 min at 95 ºC. The electrophoresis was performed at a constant voltage of 20 mA. The fixing and staining of protein bands was done according to the standard procedure of Sambrook *et al*. (1989) (all reagents were from ICN, Biomedicals).

2.2 Spectroscopic analysis

2.2.1 Pigment analysis

In order to assess the pigment concentration, the established method of Porra *et al.* (1989) was used. A sample of protein, in its trimeric form, was diluted 1 to 1000 in a solution of 80% aqueous acetone, which contains 2.5 mM sodium phosphate buffer at pH 7.8 to minimise conversion of chlorophyll to pheophytin. The sample solution was spun for 3 min at 14000 x *g*, in a tabletop centrifuge, to remove the apo-protein. The supernatant was transferred into a quartz cuvette and the OD was measured at wavelengths of 663.6 nm and 646.6 nm for Chl *a* and Chl *b* respectively.
The chlorophyll concentration is calculated according to the following equations:

$$\text{Chl } a \ [\mu g/ml] = 12.25A^{663.6} - 2.55A^{646.6}$$

$$\text{Chl } b \ [\mu g/ml] = 20.31A^{646.6} - 4.91A^{663.6}$$

The absorption was corrected for scattering by measuring the absorbance at 750 nm.

2.2.2 Absorption spectra

Room temperature absorption spectra of LHC-II were recorded using a Lambda Bio 40 spectrophotometer (Perkin Elmer). The protein sample was suspended in 10 mM Hepes pH 7.8 and placed in a 1 cm quartz cuvette. The spectral bandwidth was set to 2 nm, while the absorption was measured between 400 to 700 nm.

2.2.3 Fluorescence spectra

According to the function of LHC-II an energy transfer from Chl *b* to Chl *a* should occur. The extent of energy transfer is a measure of the intactness of the complex. It can be followed by exciting Chl *b* at 470 nm and then measuring the emission from Chl *a* and Chl *b*. The emission maximum of Chl *b* and Chl *a* is at 660 nm and at 680 nm, respectively. In a properly functional complex the Chl *b* transfers energy to a nearby Chl *a*. If the complex is degrading, whenever Chl *b* is excited, an increase in Chl *b* fluorescence is observed because less of the energy will be transferred to Chl *a*, due to the partial disruption of the structure. This is also detectable by a concomitant decay of Chl *a* fluorescence. Room temperature fluorescence emission spectra were recorded in a Hitachi F-4500 spectrofluorimeter. Chl *b* was excited at 469 nm. The excitation slit was set to 1 nm band width, while the emission slit was adjusted to 2.5 nm. The cuvette was 1 cm long in the direction of the excitation beam and 0.5 in the perpendicular emission direction.

2.3 Lipid isolation

2.3.1 Lipid purification by silicic acid column chromatography

The thylakoid membranes, at a concentration of Chl [A+B]=1.5 mg/ml, were suspended in 10 mM NaCl, 5 mM EDTA and 1 mM tricine pH 7.8. The total amount of lipids in solution was calculated considering that 3 mg of chlorophyll correspond approximately to 10 mg of acyl-lipids (Murphy & Woodrow, 1983). The total solvent volume (chloroform: methanol, 2:1 v/v) was added in proportion of approximately 20

ml solvent to 1 mg lipids. The membranes were homogenised with a glass homogeniser, filtered through a folded paper filter and stored in a volumetric cylinder to allow separation between solvent and aqueous phase. The lipid extract was transferred to a rotary evaporator and washed twice with chloroform.

Column preparation

The amount of silicic acid required was 75 mg per 1 mg lipids and was suspended in 10 volumes of distilled water. Then the suspension was adjusted to pH 8.0 with 1 M NaOH; the silicic acid was washed four times with 10 volumes of distilled water and allowed to settle for 30 min after each wash. Then it was dried in an oven at 110 ºC over night. After that the silicic acid was cooled to room temperature, suspended in five volumes of chloroform and packed in a column (Biorad) with an internal diameter of 2.5 cm, then washed again with 2 bed volumes of chloroform.

2.3.2 Lipid separation

The lipid extract in chloroform was applied onto the column. Initially pigments were eluted with 9-10 bed volumes of chloroform/acetone (4:1, v/v). The lipids were eluted according to the following procedure (Bratt & Åkerlund, 1993):
MGDG was eluted with 5 bed-volumes of chloroform/acetone (1:1, v/v), DGDG and SQDG eluted with 40 bed-volumes of acetone; finally PG and PC eluted with 10 volumes methanol. Eluted fractions were washed twice in acetone and dried in a rotary evaporator.
To purify individual lipids, a second separation step was required. A Biorad column (1 cm diameter, length 50 cm) was packed as described above. DGDG-SQDG sample was re-suspended in acetone and applied onto the column. To remove residual pigments, chloroform was used for initial washing, followed by 40 bed-volumes acetone. DGDG eluted within the first 20 bed-volumes of acetone solution.
Onto a third silicic acid column, PG and PC, suspended in acetone, were loaded. The elution was performed applying 10 volumes methanol. PG was eluted within the first 5 bed-volumes. Fractions were washed twice in acetone and dried in a rotary evaporator. Then they were re-suspended in chloroform at an approximate

concentration of 0.6 mg/ml, aliquotated in glass test-tubes, which were up with nitrogen gas and stored at -20 ºC. The sample purity was checked by two-dimensional thin layer chromatography (2D TLC).

2.3.3 Lipid determination by two-dimensional thin layer chromatography (2D TLC)

The lipid content of isolated LHC-II, and lipids purified from pea leaves were analysed by 2D TLC. On a pre-coated plate (Silica gel 60, MERC) either 2 μl of protein samples, at the Chl [A+B] concentration of 3 mg/ml, or purified lipids were applied.

The plate was positioned into a closed tank, pre-conditioned with the solvent in order to produce a homogeneous diffused vapour. In the first dimension, lipids separation occurred using a solution containing chloroform-methanol-water (65 : 25 : 5, v/ v/ v); in the second dimension the solvent contained chloroform-methanol-aqueous ammonia 25% (65 : 35 : 5, v/ v/ v). Lipids and detergent were detected using the non-specific and non-destructive reaction with iodine vapour.

To detect phospholipids 1.3% molybdenum oxide in 4.2 M sulphuric acid (Sigma spray reagent M3389) diluted in 4.2 M sulphuric acid (1:1, v/v) was uniformly sprayed over the plates. Glycolipids were detected with α-naphtol (Sigma), at the concentration of 5 mg/ml in methanol-water (1:1, v/v). The plate was sprayed until the surface was homogeneously wet. After air-drying, the plate was sprayed with 95% of sulphuric acid, and then heated on a hot plate until the spots became visible.

2.4 Crystallisation of LHC-II

2.4.1 Crystallisation by the vapour diffusion technique

Protein crystallisation relies on the principle that a protein solution in the thermodynamically unstable supersaturated-phase moves to equilibrium by transition to the thermodynamically more favourable crystalline phase. The protein solution is slowly brought to supersaturation by evaporating water until nucleation and crystal

growth can occur. All crystallisation trials were performed by the vapour diffusion method either in hanging or in sitting drops

2.4.2 Screening crystallisation conditions

Using a 24 multi-well tissue culture plate (Linbro, ICN Biomedical, Inc.) fig. 2.1, crystallisation experiments were performed, with a combinatorial screening strategy, in the following way: for each experiment a buffer, a salt and a precipitant were chosen from the ones listed below. All chemicals were prepared in stock solutions, filtered and stored at 20ºC. A buffer, for example 50 mM citric acid was chosen at pH 5.6 for the whole plate. To each A, B, C and D row a different salt concentration was applied (for example in row A 50 mM, in B 100 mM, in C 200 mM and in D 500 mM NaCl). Finally, columns were screened for a precipitant at different concentrations; PEG 400 was used for column 1 at 4%, in column 2 at 8%, etc. up to 30 % in the last column 6. The solutions were mixed in the reservoir well and water was added up to a total volume of 1 ml. Using this strategy buffers, salts and precipitants could be tested in several different combinations.

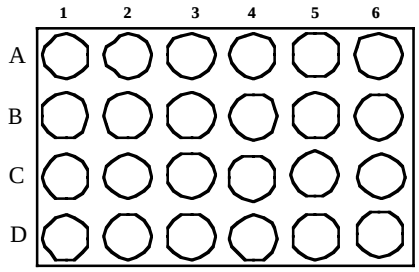


Fig. 2.1: schematic representation of a 24 well Linbro plate. To each well a letter (A, B, C, D) in the corresponding raw, and a number (1 to 6) each column were assigned.

Buffers:

Potassium hydrogen phosphate	pH 4.2-8.0
Sodium dihydrogen phosphate	pH 4.2-8.0
Citric acid	pH 4.8-5.6
Maleic acid	pH 5.8-6.2
ADA	pH 6.2-6.8
Hepes	pH 7.0-7.4
Tris-Buffer	pH 8.0-8.5

All buffers were screened in steps of 0.2 pH units. The salts: NaCl, KCl and $(NH_4)_2SO_4$ were used at the concentration between 50 to 500 mM. Precipitants, PEG-400, -1500, -2000 and -4000 were screened between 4% and 30%. (Buffers and salts were purchased from Sigma-Aldrich, while PEGs were from Merck).

2.4.3 Sample preparation

LHC-II purified either by linear sucrose gradient or by anion exchange chromatography was used for crystallisation experiments. Gradient-purified LHC-II was directly used in crystallisation trails, while the latter LHC-II was treated as described below.

Purified DGDG was aliquotated in a 1 ml Eppendorf tube and left overnight in a centrifuge evaporator to remove all chloroform. LHC-II purified by anion-exchange chromatography was added to the lipids at the ratio of 37 molecules DGDG per monomer, and incubated in a tabletop shaker at room temperature for 30 min. Samples were then ready for setting up crystallisation experiments.

2.4.4 Crystallisation trials

A small volume of LHC-II, about 2 to 4 μl, was mixed with the mother liquor at a ratio of 1:1 (v/v) and then transferred onto a siliconised cover slide (22 mm ∅, Hampton Research). The cover slides were positioned on the top of a well, which was already filled with the reservoir solution. To prevent evaporation of the mother liquor from the wells, a thin layer of Baysilone high viscosity silicon grease (Bayer) was spread over the rim of each well. All steps were carried out in an air conditioned room at a constant temperature of 20 °C. A stereo light microscope (Leica MZ 12) was used to examine the crystals and monitor their growth.

2.4.5 Optimised crystallisation conditions for LHC-II

Screening for optimal crystallisation condition yielded the following protocol: Crystallisation of trimeric LHC-II, suspended in 30 mM NG, was performed by vapour diffusion in hanging drops. Immediately before a crystallisation experiment, the sample was incubated for 15 min with 10 mM ADA buffer pH 6.4, and 10 mM NaCl, producing a partially precipitated solution. The precipitate was removed by centrifugation (2 min at 14000 x g) on a tabletop centrifuge. A 2 μl drop of the supernatant was applied on the cover slide and was equilibrated against a reservoir solution containing 10 mM ADA buffer pH 6.4, 100 mM NaCl and 9% PEG 4000.

2.4.6 Cryo-crystallography

Upon X-ray irradiation two different kinds of crystal degradation can occur: one is dose-dependent, which is a consequence of the direct X-ray-photon irradiation of the sample which leads to bond breakage in the molecules. The second is time-dependent and is associated with the presence of large amount of water in the crystals. Water radiolysis is a source of radicals (hydroxyl, hydro-peroxyl and oxygen) that give rise to radical chain reactions, leading to local damage or to the complete disintegration of the crystal.

The primary radiation damage caused by incident radiation cannot be avoided, however the diffusion process of the free radicals within the crystal can be slowed down. Experiments performed at cryogenic temperature considerably reduce secondary damage by limiting diffusion of free radicals within the crystal.

One most important aspect of cryo-crystallography is cooling the crystal without damaging it. Indeed, when a crystal of a biological macromolecule is cooled to cryogenic temperature, a 9% volume increase of the water occurs because of freezing. This results in a damage of the crystals due to mechanical stress. Cryo-protectants are therefore introduced to slow the crystallisation rate of the water, both within the crystal and within the surrounding solvent, reducing the possible disruption of the crystal by ice formation during the cooling procedure. A list of the test cryo-protectants and the used concentrations is given below (Table 2.2):

Compound	Concentration (%)
Glycerol	13-35 (v/v)
Ethylene Glycol	15-35 (v/v)
Polyethylene glycol (PEG) 400	25-35 (v/v)
Butane-2,3-diol	8 (v/v)
Glucose	25 (v/v)
2-Methyl-2,4-pentanediol (MPD)	20-30 (v/v)
Sucrose	25 (v/v)
Dimethyl Sulfoxide (DMSO)	25-35 (v/v)

The viscosity of the solution is high enough to induce a water vitrified-state at accessible cooling rates, so that the crystal is encased in a "rigid glass". Incompatibility of a crystal with a cryo-protectant solution can induce crystal cracking or its dissolution. In any case the osmotic shock experienced by a crystal, when exposed to a cryo-protectant buffer often increases mosaic spread.

2.4.7 LHC-II crystals harvesting and flash freezing

The cover slides, removed from the Linbro plates well, were placed on a glass support and positioned under a stereo light microscope. On the top of the drop containing the crystals, 10 to 20 μl of cryo-protectant solution, 30 % DMSO, was applied. The thin crystals plates of LHC-II were harvested from the drop using 10 μm thick nylon loops attached to a steel pin. This latter is moulded into a steel base, which is the cap of the cryo vial containers. The diameter of the loop was chosen larger than the crystal size, varying between 0.3 to 0.4 mm, so it was easier to scoop out the crystal from the drop. It was necessary to gently move the loop inside the drop upwards so that the crystal could be positioned on the meniscus of the drop. The crystal was then withdrawn moving the loop out of the drop and held in place by surface tension of the thin film of liquid. As general rule it is preferable to maintain the plane of the loop perpendicular to the surface of the drop to minimise the withdrawn liquid. The LHC-

II crystal's shape has a favourable surface-to-volume ratio (S/V) resulting in a fast cooling process.

The crystal caps are attached to a chrome plated steel wand with a magnet inside one end of the wand. A single crystal was harvested and immediately transferred in liquid nitrogen for flash freezing. The crystal cap vial was filled with liquid nitrogen and held with a pair of round-nose pliers. The loop was held with the magnetic rod and screwed into the vial. Care had to be used to maintain both loop and vial under the level of liquid nitrogen to avoid rapid increasing of the temperature and boiling inside the cap, which would damage the crystal. Finally the crystals were stored in a Dewar (Taylor-Warton) and were ready to be transferred to the synchrotron (all products were purchased from Hampton Research).

2.5 Data collection

The experiments were performed using a very highly collimated and brilliant beam source at the European Synchrotron Radiation Facilities, ESRF, beam line ID 14 (end-stations EH-1 and EH-3, Grenoble, France) and at DESY, beam line X-31 and BW7 A, (Hamburg, Germany).

2.5.1 Measurement of X-ray data at cryogenic temperature

The crystal caps were attached, almost vertical, to the magnetic end of the goniometer-head, using an arc part in such a way that the crystal was kept under a stream of cold nitrogen (Engel *et al.*, 1996). Once the crystal was placed in the right position the arc was removed allowing a complete rotation of 360^0 around the spindle axis. After beam alignment, the crystal was placed to the centre of the beam aperture using a CCD camera. Usually a first short, face on, exposure of the crystals was taken to evaluate its diffraction properties.

For the hexagonal space group, if the crystals is rotated around the a or the b axis, the required rotation range to collect a complete data set is 90º, but only 30º if the crystal is rotated around the c axis. The thin hexagonal plates were rotated along a or b in

either 0.5° or 1.0º steps. Each diffraction image was recorded by the rotation-oscillation method by (Arndt & Wonacott, 1977).

2.5.2 Preliminary analysis of the data

The first important step in each data collection strategy relies on the experience of the crystallographer. A very first impression of the quality of the diffracting crystal by looking at the image is required in order to adjust parameters such as:
i) the oscillation angle, that can be too small or too large; ii) the diffraction limits, to correctly position the detector with respect to the crystal; iii) the right exposure time, and iv) if experiments are set in cryogenic conditions, to discard crystals which show ice rings. For each crystal it is necessary to assess if it is worth collecting a data set and to adjust optimal data collection parameters.

2.6 Data processing

The Unix based program XDS was used for data processing, which was performed in three different steps (Kabsch, 1988a; Kabsch, 1988b; Kabsch, 1993):

1) Determination of crystal orientation, cell parameters and space group: indexing.
2) Integration of the diffraction spots from the images and simultaneous refinement of crystal, beam and detector parameters: integration.
3) Data reduction, that is placing all data on a common scale, merging multiple observations to give an unique data set.

2.6.1 Indexing

The program XDS assigns Miller indices *hkl* using images covering the first 5-degrees of crystal rotation. The crystal orientation, relative to the laboratory frame, and unit cell parameters are automatically determined from the list of observed spots. (Vectors are searched in reciprocal space; three non-coplanar seed vectors are selected with the

minimum angle between them. Several different clusters are calculated and a quality index is assigned for the best fit of the 14 different Bravais lattices).

2.6.2 Integration

Once the spot positions are correctly determined the intensity of the reflections are integrated. Integration of the intensity in a spot is the separation of the X-ray background from the reflection using a profile fitting procedure. By profile fitting a curve is fitted to each Bragg reflection and the integrated area under the curve is taken to be the spot intensity, moreover an error estimate of the intensity is assigned. The total intensity of a *hkl* reflection is:

$$I_{hkl} = \frac{\lambda^3}{\omega V^2} \times \left(\frac{e^2}{mc^2} \right) \times V_{cr} \times I_o \times L \times P \times A \times \left| F_{(hkl)} \right|^2$$

where:

I_{hkl} is the integrated intensity of the recorded reflections
λ is the wavelength of the incident beam
ω is the angular scanning velocity
e is the electron charge of mass m and c the constant light velocity
V is the volume of the unit cell
V_{cr} is the volume of the crystal
I_o is the intensity of the incident beam
L is the Lorentz factor depending on the data collection technique
P is the polarisation factor
A is the absorption factor
$F_{(hkl)}$ is the structure factor for each *hkl* reflection

2.6.3 Data reduction

2.6.3.1 Scaling

During data collection the working settings are kept constant as much as possible. However, some slight variation of the incident beam intensity (due to beam decay or re-injection at a synchrotron) and in the diffracting volume (when using the rotation method) occur, so that the images are generally recorded under slightly different conditions. Moreover, it is often the case that several data sets have to be merged. Relative scaling is done on the basis of the common reflections between data sets. Scaling is usually performed by applying a scale factor K and a temperature factor B to each image.

These parameters are refined by minimising the residual:

$$R = \sum_h \sum_i w_{hi} \left(\sqrt{I_{hi}} - \langle I_h \rangle / K_{hi} \right)^2$$

where:

I_{hi} is the ith measurement of reflection h

w_{hi} is the weight for that observation

$\langle I_h \rangle$ is the weight mean intensity for each reflection

$$K_{hi} = K_j \exp\left(-2B_j \sin^2 \theta_h / \lambda^2\right)$$

where:

K_j and B_j are the scale and temperature factors for image j on which I_{hi} was measured

θ_h is the Bragg angle for reflection h

λ is the radiation wavelength

2.6.3.2 Merging

Once all the observations are brought to a common scale, multiple observations are reduced to a weighted mean intensity and standard deviation, partially recorded reflections are summed together, outliers are rejected and the symmetry-equivalent observations are merged into a unique set. For each reflection *hkl*, an I and σ value

is reported. At this stage XDS scales the data sets in bins based on resolution and produces statistical measures allowing judgement of quality and completeness of the data.

2.7 Structure Solution

2.7.1 The phase problem

Once the production of highly pure protein is achieved and well diffracting crystals are obtained, the main task in X-ray crystallography is to solve the "Phase problem". This arises from the loss of phase information in recording the diffraction pattern.
The electron density in a crystal can be obtained by calculating the Fourier summation:

$$\rho(xyz) = \frac{1}{V}\sum_{hkl}|F(hkl)|\exp\left[-2\pi i(hx+ky+lz)+i\alpha(hkl)\right]$$

where
$\rho(xyz)$ is the electron density at every (x y z) position in the unit cell
$|F(hkl)|$ is the structure factor amplitudes of the reflections (hkl)
$i\alpha(hkl)$ is the phase angle
$F(hkl)$ is the Fourier transform of $\rho(xyz)$, and vice versa

The amplitudes value $|F(hkl)|$ can be derived because $I(hkl) = |F(hkl)|^2$, even though no information is available on the phase angles. Four different techniques have been developed in order to overcome the phase problem in X-ray crystallography.

1) Multiple Isomorphous Replacement (MIR) or Single Isomorphous Replacement (SIR), in which heavy atoms are attached to the protein molecules in the crystals. Their scattering properties are exploited to determine the initial phases.
2) Multiple Anomalous Dispersion (MAD) or Single Anomalous Dispersion (SAD), where anomalous scatterers are naturally present or inserted in the protein structure

itself. An example is the use of selenomethionine, which can be incorporated into the protein by molecular biology techniques. This is usually only possible for recombinant protein.
3) The molecular replacement method, which uses a known protein structure as a starting model for calculating a first set of phases.
4) Direct methods in which calculations are performed on a first set of random phases, which are then refined and updated repetitively. It is only useful for small molecules and high resolution data.

An exhaustive description of proposed solution to the phase problem can be found in several books and reviews (Drenth, 1994) (George H. & Lyle H., 1968) (McRee, 1993) (Giacovazzo *et al.* 1992). Here, only the molecular replacement methodology will be further discussed.

2.7.2 Molecular replacement

Crystallographic structure determination by molecular replacement (MR) uses a structural model that is fitted into the unit cell of the unknown structure providing initial approximate phases (Rossmann, 1972; Rossmann & Blow, 1962). A six parameters search, three rotation angles and three translations, describes the transformation of the set of co-ordinates of the known structure to the other. To avoid the long computing time of a systematic six-dimension search, the problem is usually divided into two stages. The first is a rotation search based on vectors close to the origin of the Patterson space, i.e. vectors between atoms within the protein molecule. The second stage looks for vectors between different molecules in the cell. If X is the set of vectors representing the atoms of the original molecule and X^I the transformed one, the transformation can be written:

$$X^I = [C]X + t$$

where C is the matrix that rotates the co-ordinates X into the new orientation and t is a translation vector. The three rotational and translation parameters that define the

orientation and position of the molecule in the asymmetric unit are calculated using the Patterson superposition method. The method searches the maximum correlation between observed Patterson vectors and those calculated for the search model.
As a search model the structure of LHC-II derived from electron cryomicroscopy of two-dimensional crystals was used (Kühlbrandt *et al*., 1994).

An initial attempt in undertaking a first set of phases was to utilise the program AMoRe (Automated Package for Molecular Replacement, Navaza *et al.*, 1994). A second trial to find a solution to the translation problem was to make use of the "brute force" based program BRUTE (Fujinaga & Read, 1987). A third effort to crack the phase problem was undertaken with the program EPMR (Kissinger *et al*., 1999), which by employing an evolutionary method, avoids the long calculation time of a simultaneous six-dimensional search. The experimental amplitudes are the input of the program EPMR, together with cell symmetry and the model co-ordinates. In analogy with evolutionary processes in biology this program uses a search algorithm through a population of trial solutions. A starting population of trial MR solutions is generated by assigning random values of the six parameters for the search model in the unit cell. Each member of the population is randomly compared to other individuals; those with the highest correlation coefficient are chosen and used in the next cycle. The process is repeated a fixed number of generations, in example 50, considering a selected maximum population size of 300 individuals.
The linear correlation coefficient between observed and calculated structure factors is defined:

$$C = \frac{\sum (|F_{obs}| - |\overline{F_{obs}}|)(|F_{calc}| - |\overline{F_{calc}}|)}{\left[\sum (|F_{obs}| - |\overline{F_{obs}}|)^2\right]^{1/2} \left[\sum (|F_{calc}| - |\overline{F_{calc}}|)^2\right]^{1/2}}$$

2.8 Refinement and model building

2.8.1 Refinement

Once an initial model is available, it has to be refined against the experimental data and the final structure has to be built. The aim of the refinement is to find a model in

which the position of the atoms gives calculated structure factors F_{calc} as closes as possible to the observed structure factors F_{obs}.

The refinement can be performed by minimising the crystallographic R-factor, which is defined as the difference between the observed and the calculated structure factor amplitude:

$$R = \frac{\sum_{hkl} \left\| F_{obs} \right| - k \left| F_{calc} \right\|}{\sum_{hkl} \left| F_{obs} \right|}$$

The refinement of the model is achieved by optimising a cross-validated maximum-likelihood function. This target function is based on the assumption that the quality of a model is judged by its consistency with the observation; meaning that to a certain given correct model the observations would have a high probability of having the same value. To assess if an atomic model is in good agreement with the observed diffraction data and not an artefact due to a mathematically under-determined system, a second R-factor is introduced. This is called the R-free, which is calculated over a fraction of data that have been excluded from the refinement process. This fraction usually corresponds to 10% of the whole data. The R-free gives an indication of the quality of refinement so that the most appropriated refinement strategy can be followed. A refinement is regarded as successful if the R-free decreases roughly in parallel with the working R-factor.

During refinement, flexible constraints may be imposed to some functions of the parameters in order to permit only realistic deviations of their values from given standard ones. These functions are used as supplementary observations, since the crystal may diffract poorly and the X-ray data are limited. The parameters are data on planarity of chemical groups, chirality, bond lengths, bond and torsion angles and contact distances. These restrains, concerning ideal values and their “rigidity”, may be imposed in order to maintain a realistic model during refinement.

All these sterochemical parameters, regarding the apo-proteins, were obtained by Engh and Huber (Engh & Huber, 1991). On the other hand, for chlorophylls and carotenoids, the parameters were calculated with the program XPLOR2D (Kleywegt & Jones, 1997) using the high-resolution structure of the single molecule found in the Hic-up database (www.xray.bmc.uu.se/hicup).

The Crystallography and NMR System program package, CNS 1.0 (Brünger *et al*., 1998) was used for the whole refinement procedure.
CNS incorporates the weighted observed structure factor amplitude and all stereo-chemical information into an artificial and empirical energy function. This target function is defined to be equivalent to the potential energy of the system:

$$E_{\text{total}} = w_A \sum E_{\text{crystallographic}} + w_B \sum E_{\text{bond distances}} + w_C \sum E_{\text{bond angles}} + w_D \sum E_{\text{torsion angles}} + w_E \sum E_{\text{non-bonded contacts}} + w_F \sum E_{\text{planar group}} + w_G \sum E_{\text{chiral volume}}$$

where w terms are used for the relative weighting of the different terms.
The energy for the crystallographic terms comes from the difference in F_{obs} and F_{calc}. The energy for stereo-chemical parameters is derived as a difference between the actual value and the ideal value of the atom position as calculated in the high-resolution structures. The total energy increases when the atoms deviate from their ideal position.
The refinement program operates to minimise the overall energy by calculating the co-ordinate shift which will give the lowest energy. The equation is re-evaluated and more shifts are applied, and if the energy decreases at each cycle the process will eventually converge.

2.8.2 Refinement protocol

The refinement protocol followed the molecular dynamics approach proposed by Brünger *et al*. (1997) (Brünger *et al.*, 1987). Two consecutive cycles of rigid body refinement followed by a simulating annealing were performed on the initial model of LHC-II. At last a minimisation step was calculated.

2.8.2.1 Rigid Body refinement

When the geometry of a group is accurately known the entire group can be treated as a rigid entity. In the case of the structure solution solved by molecular replacement, a

first refinement is performed considering the whole protein as a rigid body group. A second rigid body refinement is performed with rigid subunits of the proteins; like for example the trans-membrane region and the chlorophylls in LHC-II. The purpose is to reduce the numbers of parameters in the early stages of refinement.

2.8.2.2 Simulating annealing

The simulating annealing resembles the physical process for which a solid is heated up to a point where it is turned into a viscous liquid phase, thus all the particles are randomly oriented. Then a slow cooling procedure occurs in such a way that all particles rearrange themselves in the lowest energy state. The target function, defined above (§ 2.8.1), describes the total potential energy of the system to whom the simulating annealing procedure can be applied. The simulating annealing algorithm optimises a maximum-likelihood function according to a parameter defined as the "temperature". The term “temperature” does not have any physical meaning and yet simulation starts at ambient or at higher “temperature”. The system is heated up, starting at 300 K up to 4000-5000 K, allowing the atoms to move freely from their original positions. In a second step the system is cooled down to the initial temperature value. Even though this procedure does not have any biological equivalent, changing the temperature of the system brings the model out of local minima, allowing the structure to rearrange and eventually to find a global minimum. The molecular dynamics steps produces several different possible conformations of the model. Moreover, the constrains, imposed by the X-ray data, limits the conformations to those which really improve the R-factor.

2.8.3 Calculation of the electron density map and model building during refinement

The final experimental result of a crystallographic experiment is an electron density map, which is the calculation of the three-dimensional electron density function on a regular three-dimensional grid. The grid is usually set equal to one third of the resolution. The electron density map was calculated by the Fast Fourier Transform

(FFT) in the package program CNS 1.0 from the α_{calc} phases and observed structure factor amplitudes. The initial density map is used for the initial structure interpretation. On the other hand, this electron density map is usually biased towards the model, thus, to guide the model building and the structure refinement, difference Fourier maps and omit maps were calculated.

2.8.3.1 Difference Fourier maps

The "difference Fourier map" is a useful and convenient tool of refining a structural model. The difference $F_{obs} - F_{calc}$ map, where the F_{obs} are the observed structure factors and F_{calc} are those calculated from the model, shows positive peaks in the electron density where atoms should be added to the model and negative peaks indicated atoms that should be removed from the model.

A map with coefficients $2F_{obs} - F_{calc}$ is used to guide the construction of the model. A $2F_{obs} - F_{calc}$ map not only shows densities for the errors in the model, but also densities for the regions that are correct. This procedure is most helpful during interpretation of the map and consequent model building. This type of map is similar to the F_{obs} map, yet with less bias towards the model.

Once the model has been modified and errors are corrected or new features added, a new and hopefully better $2F_{obs} - F_{calc}$ map is calculated. This new map will be less noisy since the phases have hopefully improved.

2.8.3.2 Omit maps

The omit map is an electron density map with the observed structure factor amplitudes for the whole model but with phase angles calculated only from the part of the structure that is known to be correct. The omit map is based on the fact that the part of the model to be examined is removed from the phase calculation and the rest of the model is then used to phase the portion of the map of interest. In this way there will not be a bias from the part of model which was left out. Only a small portion of

the model is omitted each time, at most 10%, therefore it is necessary to repeat the process many times to examine the whole structure.

2.9 Phase improvement by density modification

The initial electron density map is usually quite noisy and this noise results in positive density features in the solvent region. Insufficient quality of the initial electron density map prevents a clear interpretation and an unambiguous tracing of the molecule, thus it is often necessary to improve the map phases. One of the most powerful techniques for phase improvement is to modify the electron density by solvent flattening (Wang, 1985).

2.9.1 Solvent flattening

Density modification by solvent flattening improves phases by modifying the density in real space by imposing constraints on the value of the electron density function. Indeed, crystals of macromolecules contain continuos solvent regions where the electron density is essentially constant and, from high resolution electron density maps of crystal structures, it is known that the density within the solvent region is rather low. This is because the molecules of the solvent are attached to the protein in form of a single or double layer and the free solvent molecules have a very low signal. Therefore, a set of crystallographic phases generating a flat map in the solvent regions has a higher probability to be correct. However, in order to perform solvent flattening a molecular envelope has to be defined, which is represented by a mask function. In the program CNS a mask defining the protein/solvent boundary is automatically calculated. A new set of phases on the modified model are obtained and these are used to replace the derived (MR) phases. At last, phases from density modification and observed amplitudes are used to calculate a new electron density map.

2.10 Graphical representation of protein structure

The visualisation of the electron density map of the protein model and its manual modification is possible using the graphic program O (Jones & Kjeldgaard, 1993). Making use of this program, on an SGI graphics workstation, it is possible both to fit a missing part of the model into the electron density or to change the orientation of side chain residues.

All pictures were generated using the MOLSCRIPT (Kraulis, 1991) and the most recent version BOBSCRIPT program package (Esnouf, 1997) (Esnouf, 1999). The three dimensional figures were created utilising the program Raster3D (Bacon & Anderson, 1988; Merritt & Bacon, 1997).

Chapter 3

Results

In this chapter, crystallisation experimental results are reported as well as the data analysis and reduction of three data sets collected at the synchrotron. Particular attention is given to some new molecular structure features, which were revealed by X-ray crystallography.

3.1 Isolation and biochemical characterisation of trimeric LHC-II

Once the LHC-II was purified either by linear sucrose gradient or by additional ion exchange chromatography, it was routinely necessary to verify if the protein sample was pure, active and functioning. The purity of the complex was checked by SDS-page and functionality by fluorescence emission spectra.

A Comassie blue brilliant stained gel is shown in fig. 3.1, where samples from two different preparations were loaded. The samples show a major band at ~ 25 kDa and a faint band at slightly lower molecular weight. The main band corresponds to the most abundant monomeric form of LHC-II. The LHC-II trimer was then used for factorial combinatorial screening. On the other

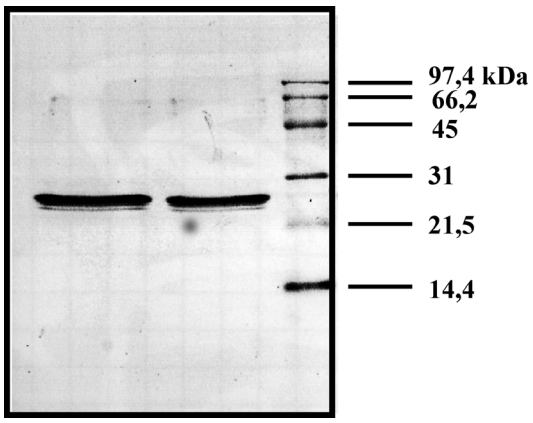


Fig. 3.1: LHC-II purified by sucrose gradient ultra centrifugation. Two different sample are shown.

hand in order to improve the protein homogeneity further purification was performed by column chromatography.

3.1.1 Anion exchange column chromatography

Anion exchange chromatography on a Mono-Q column was performed with linear salt gradient. The elution profile is shown in fig. 3.2; four different major peaks were resolved.

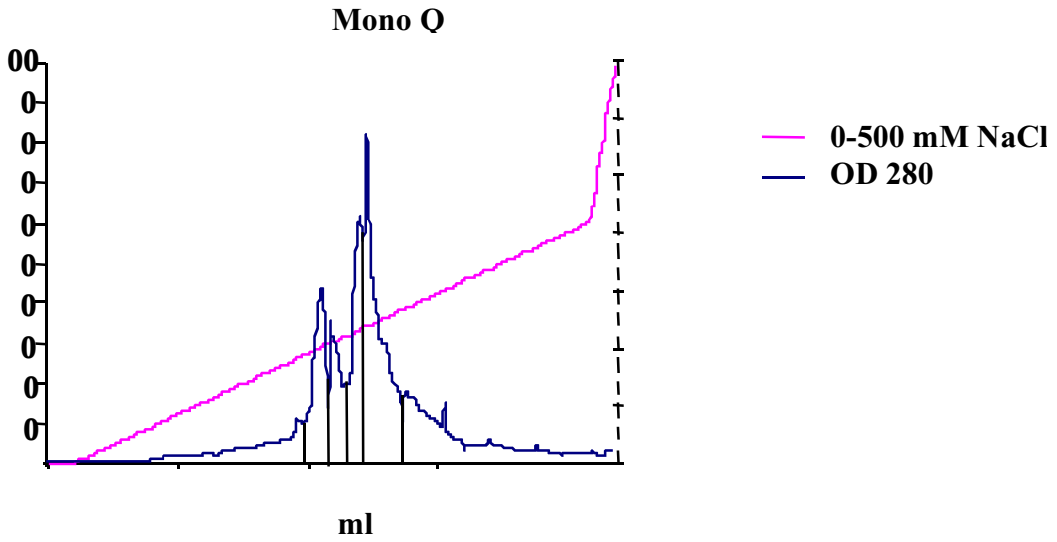


Fig. 3.2: Elution profile of LHC-II sample. Four peaks elute at salt concentration ranging from 125 to 200 mM NaCl.

Of these fractions a SDS-page and lipids analysis by 2D TLC was performed, and the Chl *a/b* ratio was measured. In fig 3.3 a schematic representation of the lipid analysis is shown. On the left panel the lipids of the trimer purified by linear sucrose gradient are revealed. In the same figure, panel on the right, the representation refers to the four eluted fractions. There were no differences between the samples, on the other hand these fractions revealed some difference when compared to the control. In fact, the extensive washing removed MGDG, SQDG and DGDG from the protein. Yet, PG was still present which is known to be essential to maintain the protein in its trimeric form. Another lipid is thought to bind to LHC-II; this is PC, however it was not detected during TLC experiments. One strong signal in form of a large spot was given by the detergent OG. In other words, there were no differences among the four

eluted fractions. All four fractions were lacking three of the lipids, if compared to the control sample.

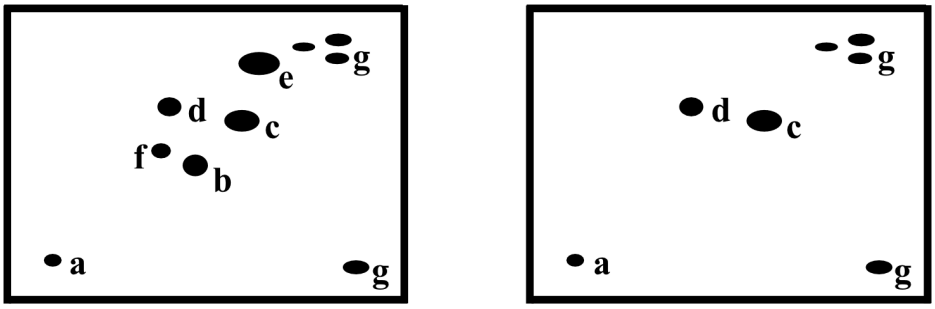


Fig. 3.3: 2D TLC: Left panel: LHC-II isolated from sucrose gradient. Right panel: LHC-II sample after anion exchange chromatography. Spots referred to as: a) Origin, b) DGDG, c) Octyl-glucoside, d) PG, e) MGDG, f) SQDG and g) free pigments.

Of these samples a gel stained with Comassie brilliant blue is shown in fig. 3.4. Eluted fractions are displayed and as a control LHC-II purified by linear sucrose gradient was used. The four samples resemble the control. Indeed, for each fraction two bands were revealed in the gel and differences are mainly due to the protein concentration of the loaded samples. One major band, corresponding to the 25 kDa monomer, and one band at lower molecular weight are evidently stained and are comparable with the sample of fig. 3.1. However, in sample 1 the lighter polypeptide is less abundant with respect to the heavier polypeptide. In samples 2 to 4 the two polypeptides forms are present in roughly equal amounts .

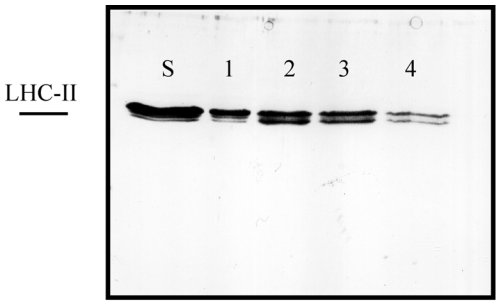


Fig. 3.4: LHC-II purified by only linear sucrose gradient is used as standart reference S. The elutions fraction are referred to as 1 to 4.

3.1.2 Spectroscopic analysis

Room temperature fluorescence emission spectra were measured to find out if the trimer in detergent solution was properly functioning. The proper energy transfer from Chl *b* to the Chl *a* in LHC-II, is shown in fig. 3.5, indicating that the complex was structurally intact.

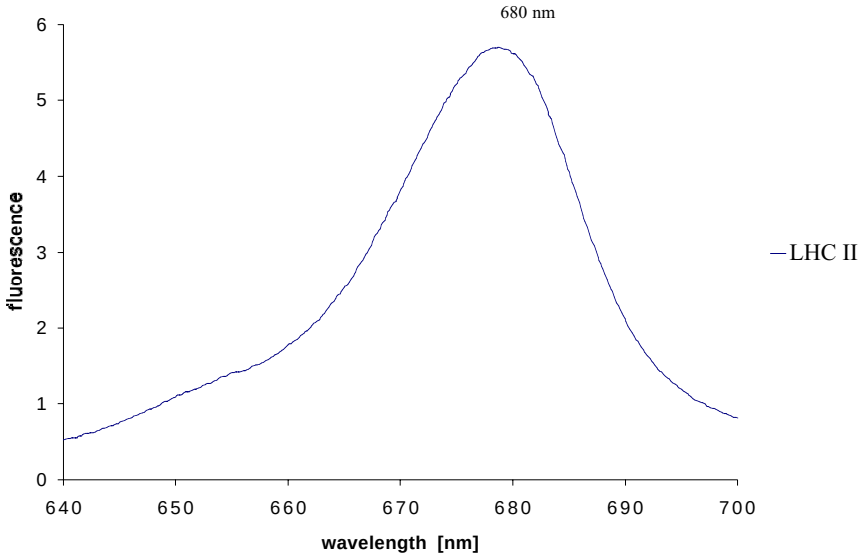


Fig. 3.5: Emission fluoresence spectra at room temperature of LHC-II. Excitation at 496 nm (Chl *b*).

When Chl *b* is excited at 460 nm, a peak occurs at 680 nm (fig. 3.5), which corresponds exactly to the maximum wavelength of Chl *a* emission, proving energy transfer within the LHC-II trimer. The room temperature absorption spectra was recorded to check the chlorophyll *a* and *b* ratio and sample composition (fig. 3.6).

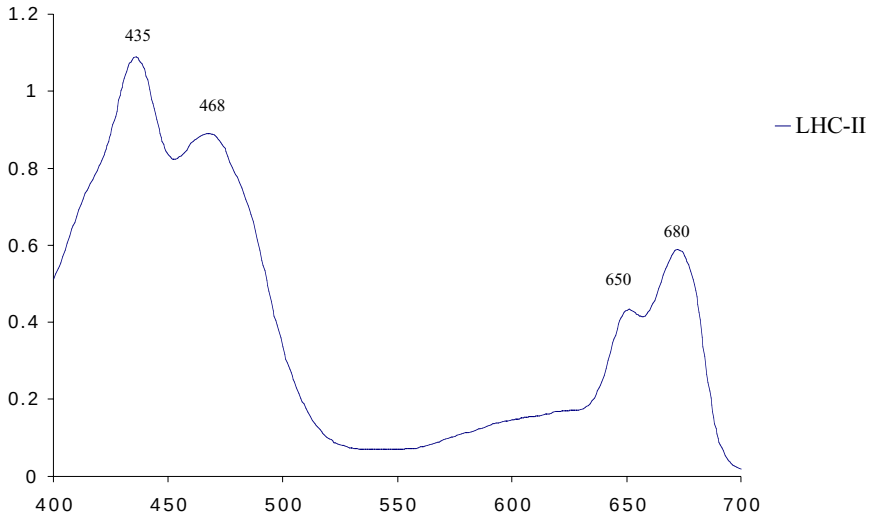


Fig. 3.6: Absorption spectra of LHC-II at room temperature.

Usually the sample was considered pure and suitable for crystallisation experiments at chlorophyll *a/b* ranging between 1.2 to 1.3.

3.2 Crystallisation

Protein fractions obtained by a linear salt-gradient in column chromatography were either concentrated utilising centri-prep or were salt precipitated and then re-suspended in detergent solution. Moreover, in order to promote crystal formation, as extensively studied by Nu§berger *et al*. (1993), the lipid DGDG is essential. Therefore LHC-II purified by ion exchange chromatography was incubated with DGDG prior to crystallisation.
The lipid purification by silicic acid column chromatography yielded to a pure sample of DGDG, which was checked by 2D TLC. On the other hand a small fraction of pigments, in form of free chlorophyll, was present in the solution. Even though other four lipids were purified, SQDG, MDGD, PC and PG, only DGDG was incorporated into the trimer. Yet, during the factorial screening, there were no conditions which could have suggested an improvement in the screening procedure. At the same time crystallisation experiments were performed using the LHC-II purified by solely sucrose gradient, which lead to more promising results. As a results of the crystallisation experiments only LHC-II purified by linear sucrose density gradient centrifugation was used in subsequent experiments.

In fact, LHC-II isolated by sucrose gradient, when forced out of solution, forms crystals within a week in several different forms. The combinatorial screening approach was used to test as many as possible crystallisation conditions obtaining information about the protein behaviour. Crystals were obtained by the vapour diffusion technique.
The peculiar feature of LHC-II of precipitating easily is of great use when several detergents would to be tested. Indeed, during the purification and after the sucrose gradient, the KCl precipitated trimeric protein can be re-suspended in any desired detergent. Of all the detergents tested (see⁄2.1.2) only NG yielded crystals. All the others resulted either in aggregation or in good indications that no further crystallisation experiments were possible. The NG concentration was 30 mM.

It was possible to observe crystal formation usually at physiological pH ranging from 6 to 7.5, in any of the ADA, HEPES, sodium or potassium phosphate buffers. On the

other hand, NG solubilised LHC-II does not form crystals both at lower or at high pH, so that Citric and Tris buffers induced mainly protein aggregation

Dark green, squared block-shaped crystals, grew within 2-5 days with phosphate buffer at pH between 6 and 7.5, NaCl and KCl at an ionic strength of 50 to 500 mM and at a protein concentration of Chl [A+B]=3.5 mg/ml (fig. 3.7a). However, from the previous work of Nu§berger (1994), it was already known that these octahedral crystals diffract at best up to 6-7 resolution when exposed to X-rays.

At protein concentration above 3.5 mg/ml, arrow-shaped crystals grew already within a couple of hours after setting up the crystallisation trials (fig. 3.7b). In these trials the buffers were used in a pH range between 6 and 7 and PEG as a precipitant at different concentrations. Indeed, these crystals are of the same type of the octahedral, but grow faster.

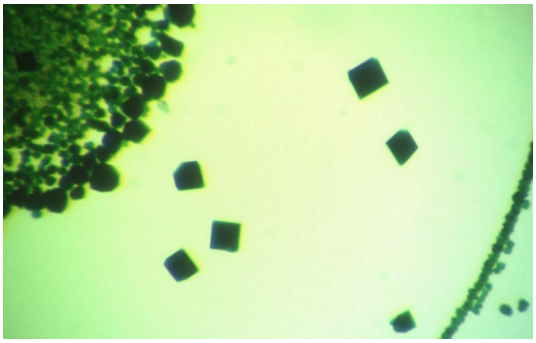


Fig. 3.7a

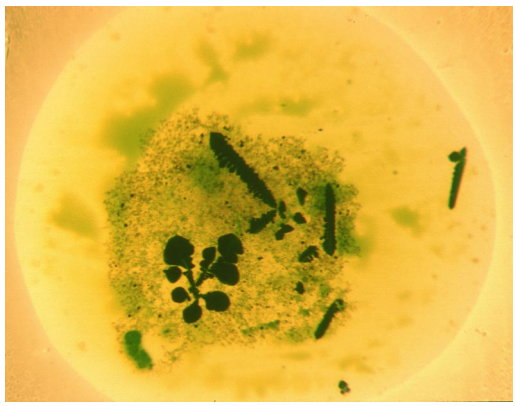


Fig. 3.7b

It was then concluded that the most suitable detergent to crystallise LHC-II was NG at the concentration of 30 mM and at the Chl [A+B] concentration of 3.0 mg/ml; that the buffer should be between pH 6 and 7 and PEG used as a precipitant. The last problem was then to slow down the process of controlled precipitation allowing the trimer to pack orderly in a three dimensional array. Thus the protein concentration was screened. The best results were obtained by not mixing the reservoir solution with the protein drops.

Very small thin plates crystals formed at first when using 20 mM HEPES, pH 7.4, NaCl 200 mM and 10% PEG 1500, fig. 3.7c.

Changing the buffer and pH to ADA pH 6.2 resulted in bigger hexagonal crystals, even if the protein still had the tendency to precipitate in presence of high ionic strength, 200 mM NaCl, as seen in fig. 3.7d.

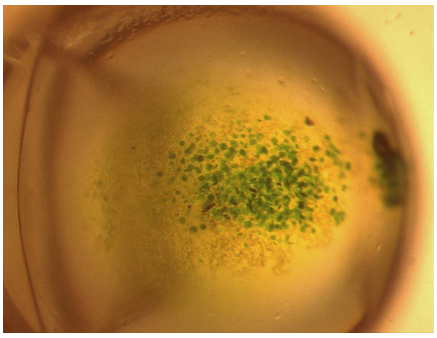


Fig. 3.7c

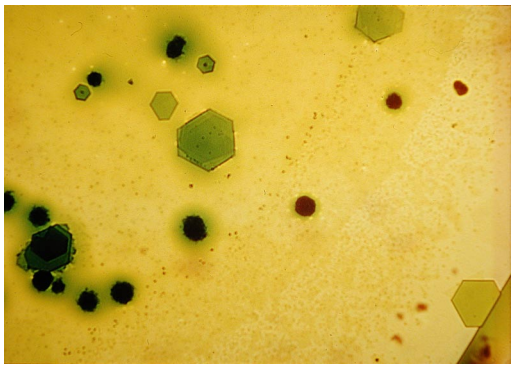


Fig. 3.7d

When using a protein concentration of Chl [A+B]=3.0 mg/ml and ADA buffer pH 6.4, 100 mM NaCl and 14% PEG 4000, hexagonal plates formed in one week, however still in presence of crystalline formation like stars, fig. 3.7e.

At the same protein concentration but at lower ionic strength, NaCl 50 mM, and precipitant, 9% PEG 4000, hexagonal crystals grew bigger and thicker compered to the previous. Nevertheless little octahedral crystals grew in the drop as well fig 3.7f and 3.7g.

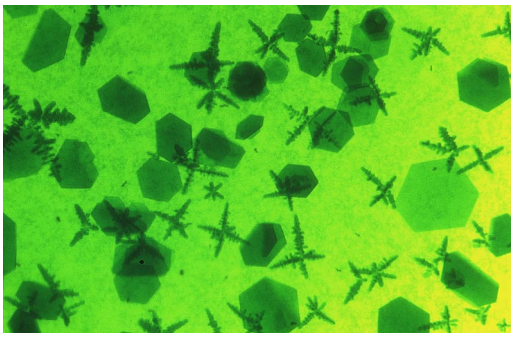


Fig. 3.7e

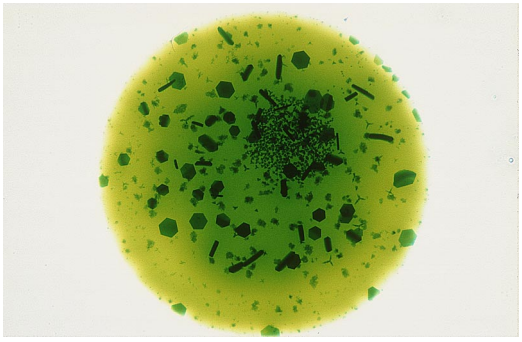


Fig. 3.7f

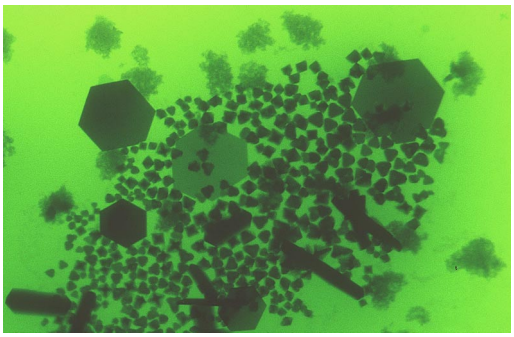


Fig. 3.7g

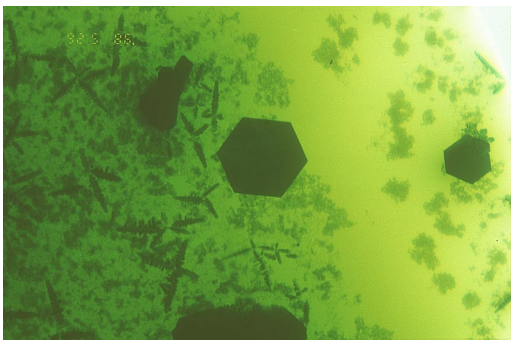


Fig. 3.7h

As a further improvement, before setting up the crystallisation trials, the protein sample was incubated with 10 mM NaCl and left at room temperature for 15 min and then spun for 5 min. Crystals appeared as thin hexagonal plates large between 0.3 to 0.4 mm across and 10 to 20 μm in thickness, both in sitting and in hanging drop.

Occasionally, some precipitation or poorly shaped crystals would still be present, fig. 3.7h.

Under optimised conditions, crystals were reproducible and only few of roughly the same size and shape formed in each drop within a week. To this end trimeric LHC-II at Chl [A+B]=2.8 mg/ml, solubilised in 30 mM NG, in the presence of 10 mM NaCl, 10 mM ADA buffer pH 6.4 was equilibrated against 9% PEG 4000 and NaCl 50 to 80 mM, fig 3.7i and l.

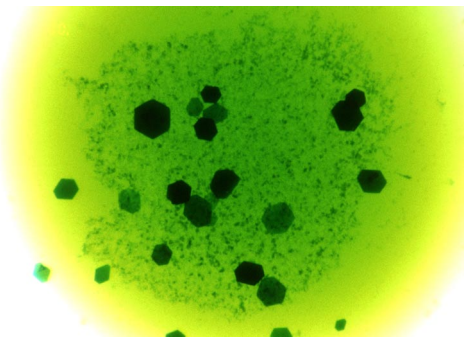


Fig. 3.7i

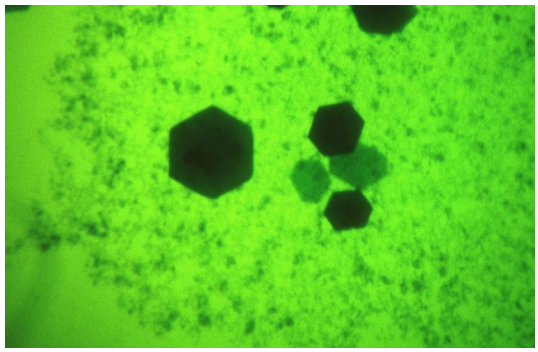


Fig. 3.7l

The pH of ADA buffer proved to be the most important parameter. Indeed, a shift of only 0.2 units, to higher or lower pH, already prevented regular crystal formation.

In sitting drop technique thin plates grew on the surface of the plastic well and were extremely difficult to be removed without damaging them. Therefore only the hanging drop technique was used to produce crystals for data collection.

3.2.1 Harvesting crystals and flash freezing

LHC-II crystals were extremely unstable and fragile, thus it was necessary to find conditions for freezing them without loss of order.

The eight most commonly used cryo-protectants were tested (⁄2.4.6). It has found that DMSO has the least harmful influence on the crystal. After applying 10 to 20 μl of 30% DMSO on the top of the drop, it was most important to harvest the crystals in a very short time. The time required between crystal picking and flash cooling is one of the most critical parameters in a cryo-crystallography experiment. The time should be less than one second and only a skillful hand can perform a good freezing. However, the high viscosity of the cryo-protectant prevented its fast mixing with the drop solution and the consequent dissolving of the crystals. Moreover, the crystals

were usually stacked on the cover slides and only with fast movements of the loop upwards it was possible to move a crystal in the drop so that it could be fished and frozen immediately. After 10 to 20 seconds DMSO diffuses into the drop; a typical observable behaviour was that the crystal plates opened along the edges as if they were pages of a book, then the thin layers dissolved. Therefore only one crystal could be harvested from each drop. Once the crystals were frozen and stored it was possible to judge the quality of the freezing procedure only when a diffraction pattern was recorded at the synchrotron. Usually the presence of ice indicates crystal damage, suggesting that the process of freezing had not been fast enough to produce vitreous amorphous ice, but instead a hexagonal or a cubic in crystals. The formation of these two forms of ice increases the solution volume, resulting in physical damage to the crystals, and which in turn leads to a wider rocking curve and a consequent loss in resolution. Data could not be recorded from such a crystal and it was necessary to mount a new one.

3.3 X-ray data collection

All together nine different experiments were performed at synchrotron radiation facilities.

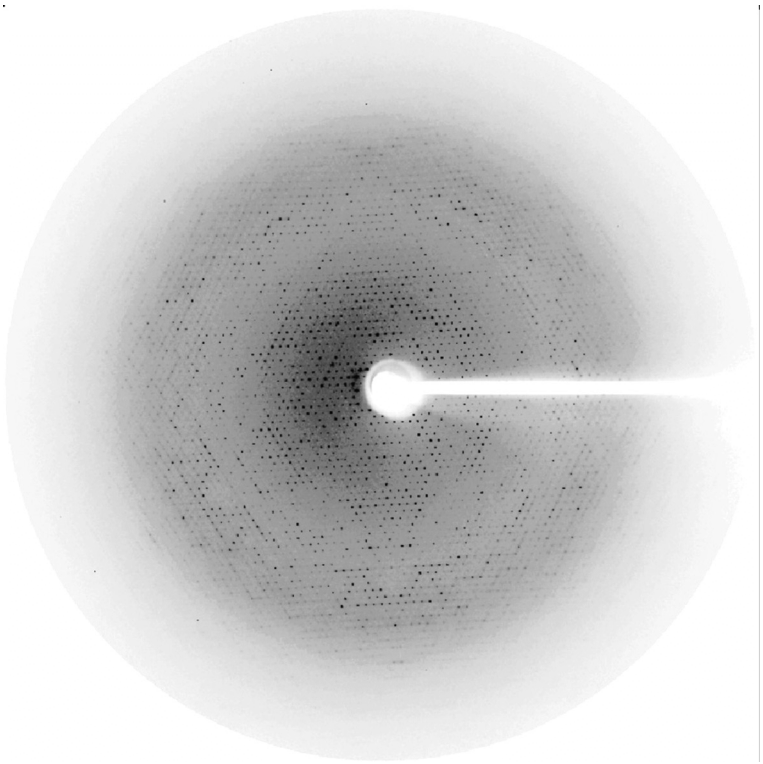


Fig. 3.8:Diffraction pattern of LHC-II crystals recorded at ID14 EH-1.

In total more than 400 crystals were tested at the ESRF, beam lines ID14 station EH-1 and EH-3. Few experiments were also performed at the EMBL outstation at DESY, beam lines BW7 and X-31. All experiments were performed under cryogenic conditions; crystals exhibited a variable diffraction behaviour. Most of them diffract up to 3.4 when oriented, in the experimental frame, in a way that the *c* axis is parallel to the incident X-ray beam and either the *a* or the *b* axis is along the spindle axis. When the crystal is rotated by 90° around the spindle axis, thus the incident beam is parallel with the *a-b* plane, the resolution reduces dramatically and only few crystals showed good diffraction in this orientation. At last, among all crystals tested only three were judged useful for data collection. All them were obtained from different protein preparations. Thus three data sets were collected and referred to as lhc-I, lhc-II and lhc-III. Data collection was performed using a rotation range of 0.5… for crystal lhc-I, while for crystal lhc-II and lhc-III 1… was used in oscillation diffracting technique. For crystal lhc-I a total number of 270 frames were recorded covering a rotation range of 135…, while for crystal lhc-II 124 frames were recorded and for crystal lhc-III only 110. In fig. 3.8 the diffraction image of the lhc-I crystal is shown: the resolution is 3.2 . The statistics of the measured data sets is summarised in table 3.1.

Data set	lhc-I	lhc-II	lhc-III
Wavelength ( )	0.934	0.934	0.9311
Unit cell ( )			
a	127.2	128.59	127.98
b	127.2	128.59	127.98
c	131.8	135.8	134.72
Space Group	P6322	P6322	P6322
Measured Reflections	124,437	72,569	89,444
Unique	10,907	6,399	8,469
Completeness %	92.3 (77.7)	83.8 (77.7)	91.2(91.5)
I/σ	13.02	10.12	10.5
R-factor	18.3	22.4	18.2
R-meas	18.9	21.9	19.1
Rmrgd-F	8.6	13.7	9.4
Resolution	3.2	3.2	3.4
Reflecting range	2.2	2.8	2.8
Beam line	ID 14 EH-1	ID 14 EH-1	ID 14 EH-3

Table 3.1: Data collection and data processing statistics obtained from three different crystals of LHC-II recorded at ESRF ID14 EH-1.

3.3.1 Crystal characterisation

The crystallographic characterisation of LHC-II crystals reveals unit cell constants of a=128.45, b=128.45, c=135.32, $\alpha=\beta=90\ldots,\gamma=120\ldots$. Considering a monomeric molecular weight of about 28 kDa the Matthews coefficient is 4.06 $\AA^3/Da$ and the solvent content 69.45%. A list of systematic absences is reported here below. Along the c^* axis, the (00l) reflections, for which l=2n+1, are weak or missing with the only violation of (0 0 29) at 4.6 resolution. This suggest with the space group $P6_322$.

H	K	L	RESOLUTION	INTENSITY	SIGMA	INTENSITY/SIGMA	#OBSERVED
0	0	8	16.854	0.4423E+05	0.6348E+04	6.97	1
0	0	9	14.981	0.6496E+02	0.1689E+03	0.38	1
0	0	10	13.483	0.7345E+03	0.3502E+03	2.10	1
0	0	11	12.257	0.1047E+03	0.1431E+03	0.73	2
0	0	14	9.631	0.3091E+05	0.2032E+04	15.21	2
0	0	15	8.989	-0.1245E+02	0.1135E+03	-0.11	2
0	0	17	7.931	0.6646E+03	0.2757E+03	2.41	1
0	0	18	7.491	0.6184E+04	0.3688E+03	16.77	2
0	0	19	7.096	0.1616E+02	0.8736E+02	0.18	2
0	0	20	6.742	0.4074E+04	0.2969E+03	13.72	2
0	0	21	6.421	0.1019E+03	0.9991E+02	1.02	2
0	0	22	6.129	0.3444E+04	0.2416E+03	14.25	2
0	0	23	5.862	0.1629E+03	0.1189E+03	1.37	2
0	0	24	5.618	0.3556E+04	0.2650E+03	13.42	2
0	0	25	5.393	0.4357E+03	0.1367E+03	3.19	2
0	0	26	5.186	0.9556E+04	0.6523E+03	14.65	1
0	0	29	4.649	0.1260E+04	0.2550E+03	4.94	1
0	0	30	4.494	0.4133E+03	0.1352E+03	3.06	2
0	0	31	4.349	0.9902E+02	0.1222E+03	0.81	2
0	0	32	4.213	0.2094E+03	0.1255E+03	1.67	2
0	0	33	4.086	0.1132E+03	0.1139E+03	0.99	2
0	0	34	3.966	0.2168E+03	0.1180E+03	1.84	2
0	0	35	3.852	0.2428E+03	0.1188E+03	2.04	2
0	0	36	3.745	0.9303E+02	0.1217E+03	0.76	2
0	0	38	3.548	0.1343E+03	0.1163E+03	1.16	2
0	0	40	3.371	0.7336E+02	0.1127E+03	0.65	2
0	0	41	3.289	0.2247E+03	0.1152E+03	1.95	2
0	0	42	3.210	0.2188E+03	0.1133E+03	1.93	2
0	0	43	3.136	0.2916E+03	0.1505E+03	1.94	1
0	0	44	3.064	0.1033E+03	0.1126E+03	0.92	2

3.3.2 Data quality

Judging the quality of a data set by looking only at the high-resolution limit is not always sufficient. Indeed, data should be complete at low resolution (to about 10),

because the low resolution data give information about the overall shape of the molecule and of the solvent, meanwhile high resolution data provide information about the detailed features of the protein.

At the end of data reduction the program XDS displays an output which describes the quality of the data as a function of resolution.
The most complete data set is lhc-I; at 3.2 resolution, the overall completeness is 92.3%, while at the highest resolution shell, between 3.3 and 3.2 , it is 77.7%. However, at low resolution, between 10 to 6 , the completeness is 99.9%.
Crystal lhc-II data are 83.8% overall complete. In the highest resolution shell (3.4-3.2) the completeness is 77.5% but only 83.9% at lower resolutions, between 10 and 6 .
The third data set, lhc-III, has an overall completeness of 91.2%, 91.5% in the high resolution shell, between 3.6 and 3.4 , and 88.8% for low resolution data (10-7).
For each data set, three different R-factors, as a function of resolution, are computed. Usually R_{sym} is used to indicate the quality of merging according to crystallographic symmetry, however in XDS a new kind of R-factor is introduced and referred to as R_{meas} (Diederichs & Karplus, 1997). The R_{meas} of the lhc-I and lhc-III data set are almost the same with 18.9 and 19.1%, while for lhc-II the value is slightly higher 21.9%. These values are comparable with the R-factor, which is a measure of the quality of the measured intensities. One more R-factor is provided to judge the data quality. It is defined as R_{mrgd-F} and is a measure of the quality of the reduced amplitudes. It is useful, since most crystallographic calculations, such as phasing and structure refinement are done using structure factors amplitude instead of using intensities. The values computed for lhc-I and lhc-III, 8.6 and 9.4% are low, yet for lhc-II, 13.7% was the highest value. However, these R-factors indicate that the data are usable for structure determination.
One other important parameter is the reflecting range, defined as the approximate rotation angle required for a strong spot, recorded perpendicular to the rotation axis, to pass completely through the Ewald sphere. It is also known as mosaicity. This was found to be 2.2… for crystals lhc-I, while a higher value, 2.8…, was computed for lhc-II and lhc-III.
Once the data were scaled to a common base, a single list of reflections is produced. Next, it is necessary to attribute a phase angle to each intensity.

3.4 Structure solution

3.4.1 Molecular replacement

An initial set of phases for the structure determination was obtained by molecular replacement using the LHC-II model at 3.4 x 4.9 resolution deduced from electron cryomicroscopy studies of two-dimensional crystals (K hlbrandt *et al*., 1994). The molecular replacement method is based on finding the maximum correlation between observed and calculated Patterson vectors from the search model. Three different programs were used in order to find a possible orientation of the LHC-II model in the unit cell, which had resulted from the data processing.

Using the program AMoRe (Navaza *et al.*, 1994) a solution was not found, in fact any of the molecular positions calculated during several cycles of superposition of the intermolecular vectors failed the search in the consequent translation step.

A second attempt was made with the program BRUTE (Fujinaga & Read, 1987). As well as for the former program no solution was found even if refinement of the rotation function was performed during a translational search. A last attempt to solve the phase problem was done with the program EPMR (Kissinger *et al*., 1999), which avoids separation of rotational and translational function searches by introducing a six-dimension search using an evolutionary optimisation algorithm.

By using a single monomer per asymmetric unit as a search model, the evolutionary program EPMR has found a most probable solution. The program found a population of the ten best solutions, which were defined by a correlation coefficient and an R-factor.

From these only one file was selected: the best.pdb . It consists of experimental amplitudes and calculated phases, has and a correlation coefficient of 36.1% and an R-factor of 54%. A second run of the EPMR was then performed using the trimer as a search model. The same solution was found.

3.4.2 Result of molecular replacement

When using a monomer as a search model the final orientation was given as a molecular position which resulted to be correct with respect to the three-fold

symmetry related monomers. Moreover the resulting trimer was found to be the same as the model from two-dimensional crystal analysis (K hlbrandt *et al*., 1994). Furthermore, crystals of LHC-II appear to be stacks of the two-dimensional crystals, therefore belonging to type I crystals, according the classification of Michel (Michel, 1983). The trimer has an internal three-fold symmetry and the trimers are symmetry related by a two-fold symmetry axis, which is parallel to the membrane plane. Thus in a single layer one trimer points up and the adjacent trimer points down. A six-fold screw axis is perpendicular to the crystal plane, so that one trimer is rotated 60… with respect to the corresponding trimer in the parallel crystal layer, and shifted half a unit cell along the *z* axis.

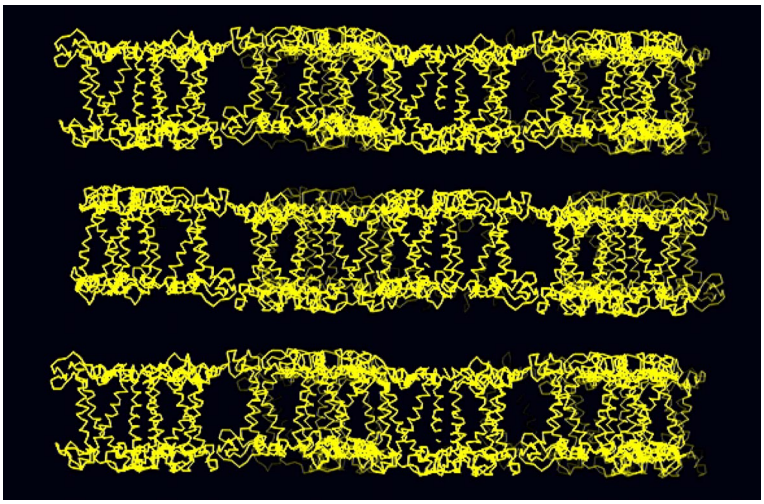


Fig. 3.9a: Side view of three consecutive layers of LHC-II trimer.

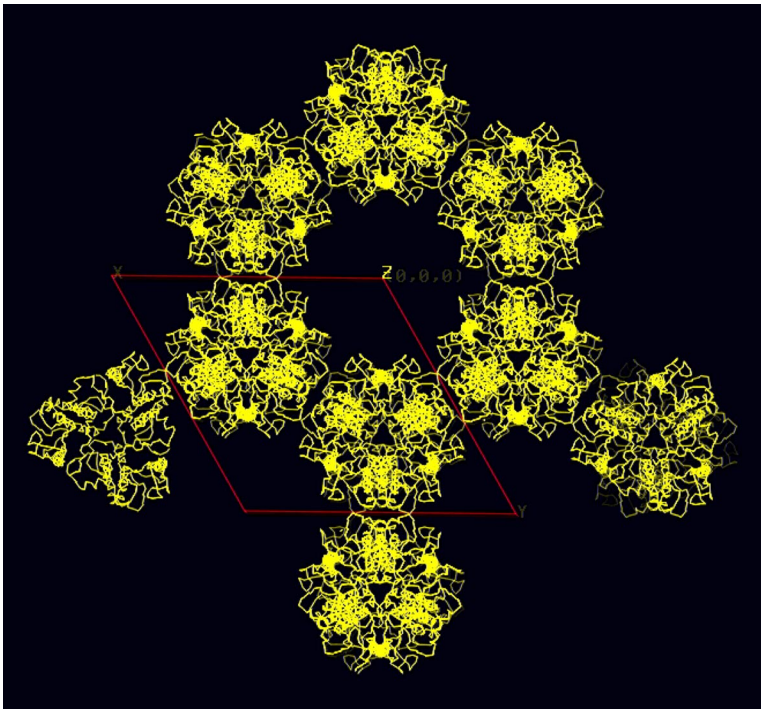


Fig. 3.9b: Top view of LHC-II trimer (in red the unit cell).

The crystal packing, in the space group $P6_322$, is shown in fig. 3.9a (side view) and fig. 3.9b (along *z* axis). Two consecutive protein layers are at the minimal distance of 14 . This initial model had to be improved by crystallographic refinement procedures.

3.5 Refinement

Refinement is often a critical task in the process of structure determination. The refinement was performed according to the refinement protocol of Br nger *et al.* (1997). The first step of the refinement was a cycle of rigid body refinement. The whole LHC-II trimer was considered to be almost in the right position and rigid body refinement was performed minimising a most likelihood function, between the R_{obs} and the R_{calc}. The starting R-factor value was 54%. After this first cycle the R-factor decreased to 50.25% while R-free was 53%.

As a second step, one more cycle of rigid body was performed. Yet, in this second cycle the single α-helices (A, B, C and D as in fig. 12a, b) were considered as single entities, so that only three rotational and three translation parameters were calculated for each one to a total of 24 independent parameters. In addition, the twelve chlorophylls were treated as single entities, thus another 72 parameters were introduced. Yet, after 200 steps of minimisation cycles the R-factor and R-free did not change in value.

A further step in the refinement protocol was the simulated annealing: in order to obtain a better fit, the electron densities of all atoms are subject to a movement in a way that allow atoms to move from their original position. At this stage the R-factor and R-free improved to 38.75 and 46.33% respectively. Eventually, one more cycle of minimisation, according to the maximum likelihood function, has been performed and the final R-factor was 35.7%. The R-free improved as well to 45.4%.

3.5.1 Omit map calculation

In order to verify if new features in the electron density map were not only biased towards the model omit maps were calculated. The co-ordinates relative to the Chl *a*6

were removed from the input pdb file and a cycle of refinement, following the protocol described in ⁄ 2.7.2, was performed. A new difference $2F_{obs} - F_{calc}$ map was calculated and positive peaks were found corresponding to the position of the omitted Chl *a*6, fig. 3.10 a and b.

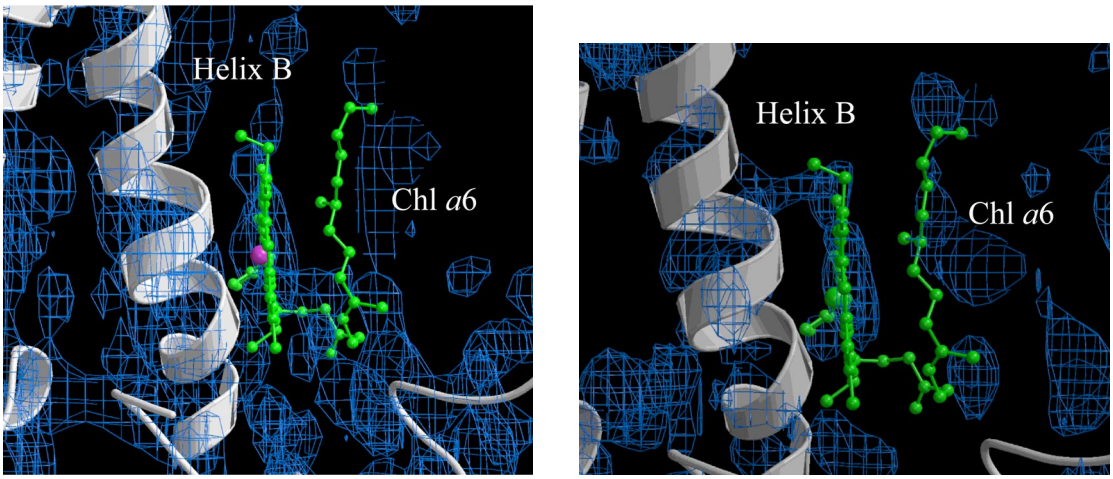


Fig. 3.10a: A $2F_o - F_c$ and a (figure on the right) $F_0 - F_c$ electron density of the omitted Chl *a*6.

The same experiment was performed removing ten amino acids of the central part of α-helix B (fig. 3.11a), from residue Glu 65 to Gly 75, from the initial model. A cycle of refinement was performed as for the previous example. In fig. 3.11b and 3.11c peaks of the $2F_{obs} - F_{calc}$ electron density map are displayed. It was concluded that all of the new features in the electron density map were not due to the initial model and therefore the MR solution was clearly correct. However, the electron density map was still not clear enough to draw and fit new structure features so that it was necessary to improve the phase angle.

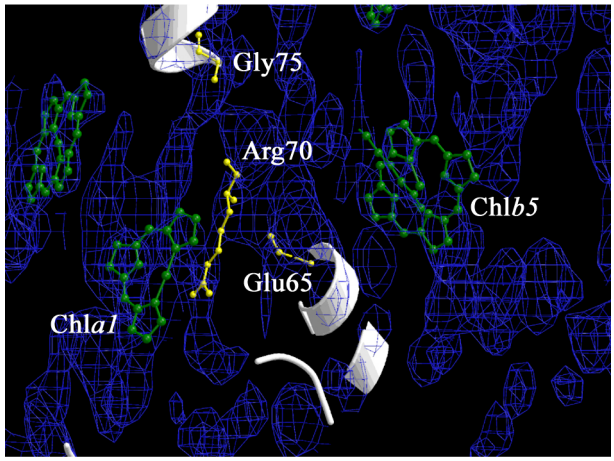


Fig. 3.11a: $2F_o - F_c$ electron density map at 1.8 σ level of the model. The amino acids between Gly75 and Glu65 are removed for clarity.

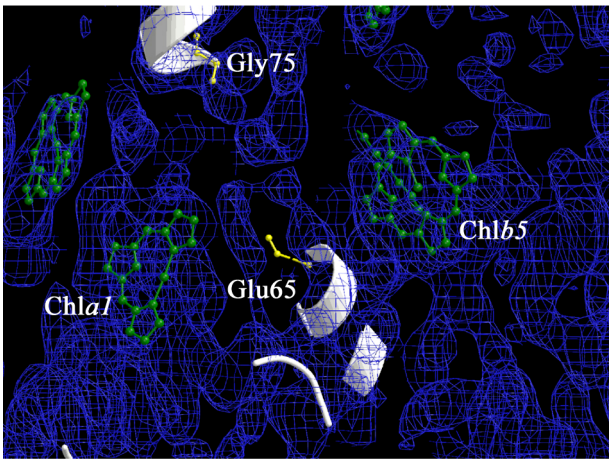


Fig. 3.11b: $2F_o - F_c$ electron density map at 1σ level of 10 amino acid omitted map.

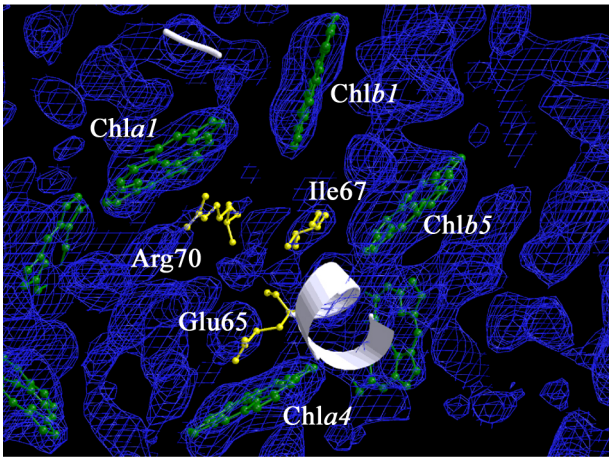


Fig. 3.11c: $2F_o - F_c$ electron density map at 1σ level of 10 amino acid omitted map, top view from the stroma side.

3.5.2 Solvent flattening

The electron density map is extremely sensitive to errors in the phases, which can strongly bias the structure solution towards the model. The most widely used method of phase improvement is solvent flattening. Only one cycle of solvent flattening was performed with the program CNS, and the resulting electron density map showed clearer features, revealing some new aspects which are described here below.

3.6 New features in LHC-II structure

Although, the structure model worked out by K hlbrandt *et al* . (1994) derived by electron cryomicroscopy of two-dimensional crystals is confirmed, some new structural features appear as a results of this work. In the following description the

published LHC-II structure will be referred simply to as "EM model". Consistently, with the EM model the α-helices will be referred to as A, B, C and D (fig. 12a, 12b) as well as the pigments: lutein 1 and 2 and the 12 chlorophylls named either Chl *a* or Chl *b*.

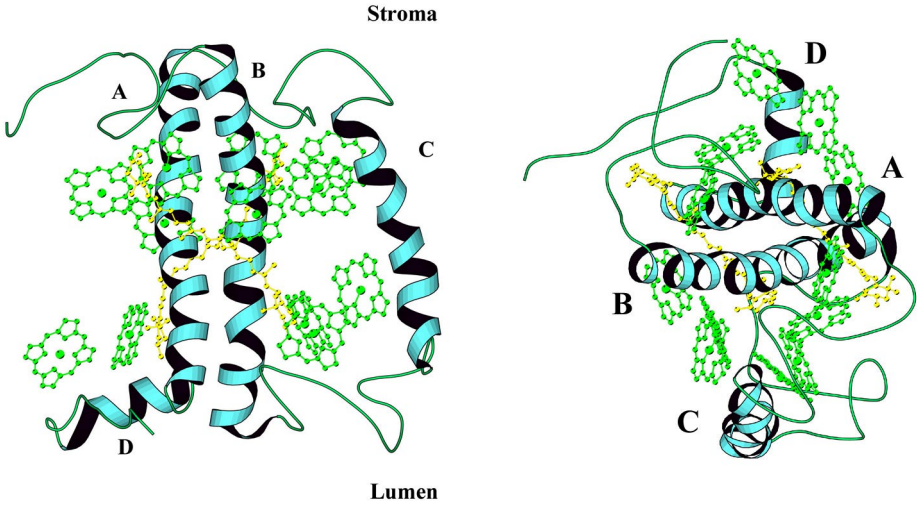


Fig. 3.12a-b: Overall view of LHC-II. Helices are labelled A to D. Chl tetrarrapyrroles in green and lutein molecules in yellow (K hlbrandt *et al.*, 1994).

At the membrane surface on the lumenals side the BC loop runs between residues Val 90 to Gln 122. However, in the EM model a clear fitting was made up to Ala 100 and from Pro 116. As well in the X-ray map it is evident that this loop should have a different orientation. Indeed, there is a clear density only up to Val 90 and again from Lys 118, however the current resolution is not enough to propose an unambiguous fitting of the residues in between.

The AC loop is between residues Ala 144 and Asp 169 and in the EM model it was entirely fitted in the electron density. On the contrary in the X-ray map around 5 amino acids (from the residue Gly 145 to Glu 150) there is not electron density and as well as in the case of BC loop a new fitting should be made.

In the EM model the first 25 amino acids of the N-terminus were not fitted in the density map. The solution of the X-ray data shows that after residue Gly 26, the electron density prolongs with respect to the density of the model (fig. 3.13a), therefore one more residues along the amino-acid chain could be fitted: according to the sequence of *Lhcb*1 in pea, this amino-acids is Lys 25 (fig. 3.13b).

The C terminus protrudes in the lumen, the X-ray electron density is well defined up to the Ala 214, corresponding to the last residues of the amphipathic α-helix D. However, the position of the residues between Try 222 and Asp 227 seem to be less clear, consequently the fitting of the last five amino acids from the Phe 228 to Lys 232 has to be confirmed at higher resolution.

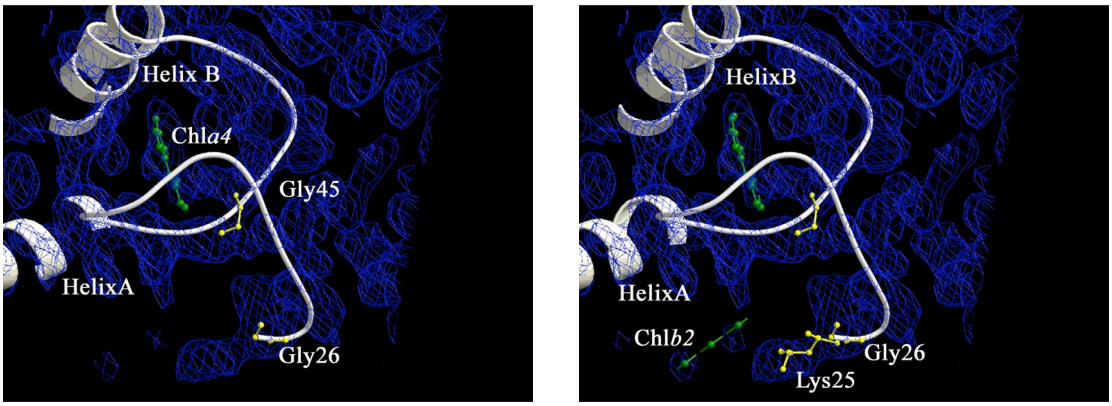


Fig. 3.13a-b: A $2F_o - F_c$ electron density map at the N-termini (stroma). After Gly 26 the electron density continues (a) and a new amino acid Lys25, coloured in yellow is fitted.

The α-helices A and B of the EM model correspond properly to the X-ray density map, however, with respect to the α-helix C some major differences are discovered .

3.6.1 Helix C

The electron density corresponding to α-helix C shows a discontinuity in the trans-membrane region. The fitting of the helix, as derived from the EM model does not match properly with the density calculated from the X-ray data. In fact, there is an evident and unexpected unwound region at level of Val 132, thus two short portions of α-helix span the lipid bilayer. This is shown in fig 3.14 where the α-helix C of the model is in white (a) and the new proposed modified α-helices are in red (b).
The upper α-helix runs from Leu 134 to Ala 144. While a second α-helix, closer to the lumenal side, runs from Gln 131 to Ala 120. In respect to the EM model four more amino acids residue have been fitted. Indeed in the EM model the α-helix C was fitted between Ser 123 to Ile 143. The X-ray density continues further at both side of the α-helix.

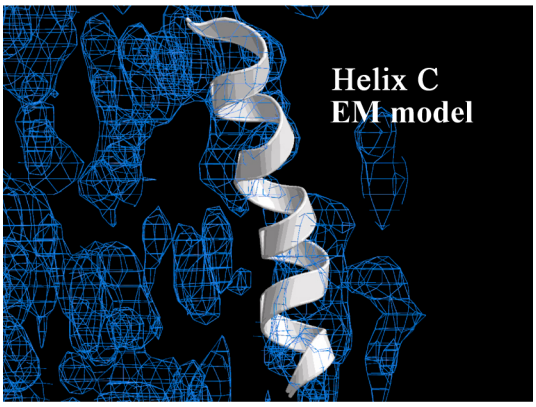


Fig. 3.14a: Helix C as in the EM model.

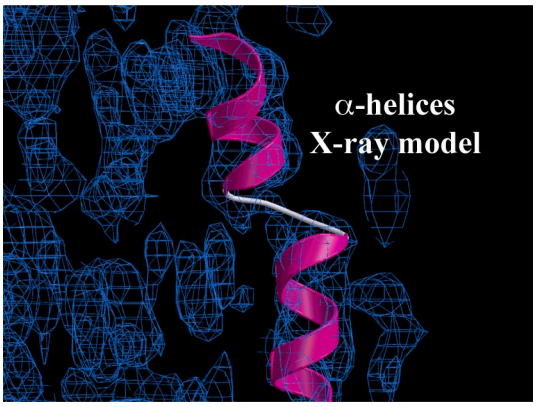


Fig. 3.14b: New fitted α-helices.

On the stroma side only one more amino acid could be fitted: the Ala 144. Instead at the lumenal side three more amino acids residues are fitted and these are the Gln 122, Ala 121 and His 120. It is, however not clear how the amino acid chain continues after Lys 118. However, the bulky side chains of the amino acids residues Glu 139 and Gln 131, which provide Chl b6 and Chl b5 ligands, respectively, have to be slightly moved to allow a this new orientation of the α-helix C. Indeed the bulky side chains of the residues were used in the EM to drive a first the fitting of the peptide chain (K hlbrandt *et al.*, 1994).

3.6.2 Carotenoids

As a results of X-ray data analysis an electron density was discovered and attributed to a third carotenoid molecule, which appears to be parallel to the α-helix C (fig. 3.15a). This density looks to be straight, linear and almost perpendicular to the membrane plane; it extends between the Tyr 141 and Ser 123, and it is 28 long. Moreover, it is on the opposite side of the α-helix C with respect to the α-helices A and B domain and hence exposed to the external part of the LHC-II monomer. According to biochemical analysis this carotenoid should be a neoxanthin, which is indeed always found in LHC-monomer in stoichiometric amount. The carotenoid is at 7.5 away from the α-C of Leu 134 and bents at both the extremities closer to Tyr 141 and to Ser 123 (fig. 3.15b).

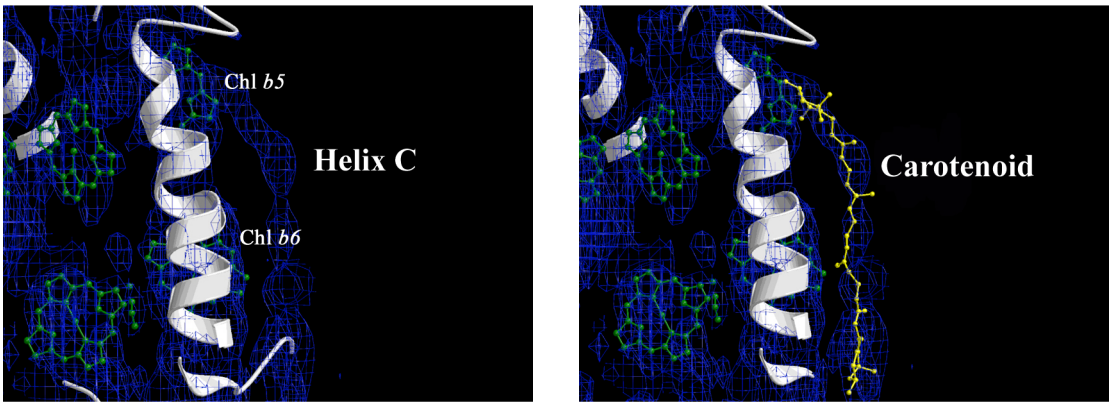


Fig. 3.15a-b: The electron density on the side of helix C (picture on the left); a carotenoid is fitted most probably a neoxanthin (picture on the right)**.**

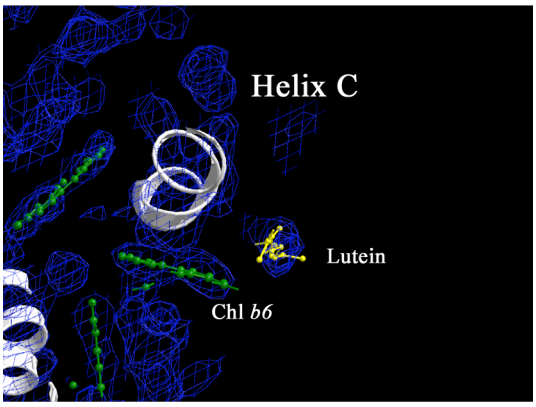


Fig. 3.15c: Top view of the helix C and lutein.

At the stroma side, the head group of the carotenoid is at the same height of Chl b5, which is a 3.4 away. On the lumenal side the carotenoid is 4.5 far from Chl b6. In addition, the closest chlorophyll a is Chl a6: the chlorin ring is at 14.5 away while the phytytl chain is only 10 away from the carotenoid. The position of the two others carotenoid, assigned to lutein molecules in the EM model, is confirmed as well in the X-ray electron density map.

3.6.5 Chlorophylls

The phytyl chains, which gives to the chlorophyll molecules their lipid character, are fitted and built into the electron density map for all the chlorophylls except for one. Moreover, beside some positional shifts with respect to their position, twelve chlorophylls are confirmed and no other density was discoverd which could be attributed to additional chlorophyll molecules. They are placed in two parallel layers.

One of them is close to the lumen while the other one is close to the stroma. Even though the resolution improved to 3.2 , yet it does not allow to distinguish between chlorophyll *a* and *b*. The different features which were found in the map, calculated after solvent flattening, are listed for each chlorophyll here below.

Chlorophyll *a*1

The Chl *a*1 is embedded in a region between Pro 156 and Phe 173 and no further density is revealed around the porphyrin ring, which could indicated the position of the phytyl chain. Although the Chl *a*1 position was confirmed the chlorin ring is

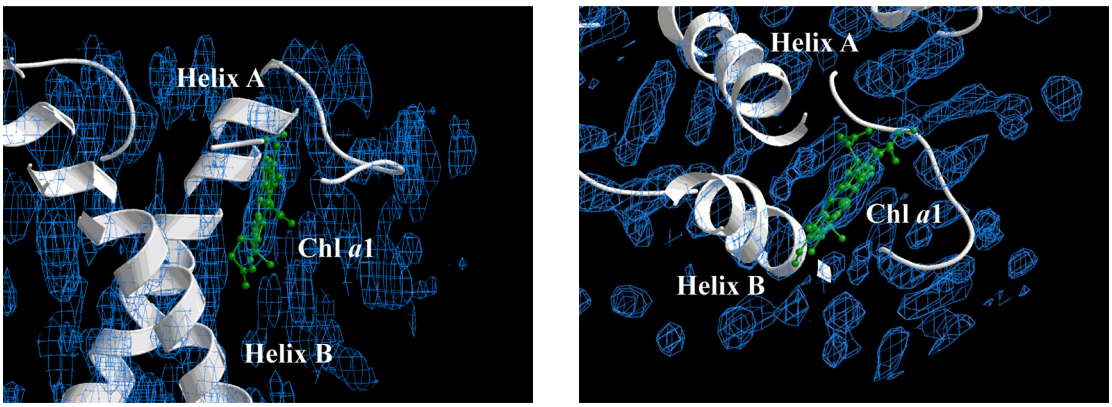


Fig 3.15: Chl a1, side view and top from the lument side.

shifted of 2.5 . The distance to Glu 180 and to Arg 70 is about 4.5 . The head group of the lutein is 4 away.

Chlorophyll *a*2

Close to Chl *a*2 chlorin ring an electron density was discovered and attributed to a phytyl chain. The Chl *a*2 phytyl chain is perpendicular to the membrane plane and bent over the porphyrin ring pointing towards the inner part of the lipid bilayer. Chl *a*2 is co-ordinated by Asn 183 (fig. 3.16a, b) and is located between two consecutive trimers in the crystal, the porphyrin ring is perpendicular to the plane of Chl *b*2, which is 9.9 away. From the position that has been assigned in the EM model, Chl *a*2 is shifted 1.5 away from Chl *b*2. As a consequence the distance to the co-ordinating Asn 183 is 3.2 compared to the 3.66 described previously.

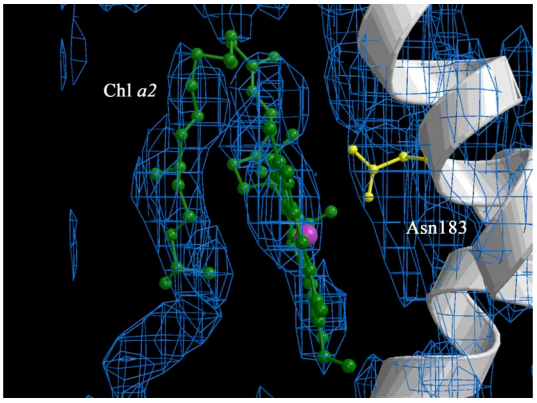


Fig. 3.16a: Chl a2 side view.

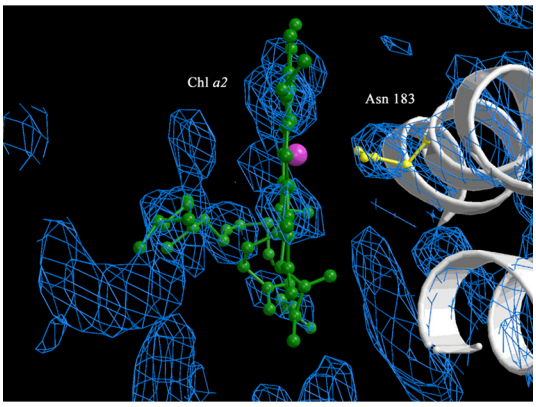


Fig. 3.16b: Chl a2 top view.

Chlorophyll *a*3

The Chl *a*3 phytyl chain has been completely assigned (fig. 3.17a, b and c). The electron density is straight up the chlorin ring and bents at position C8. It is extended up to Arg 185 with a total length of 22 . The phytyl chain is 5.6 away from the chlorin ring of Chl *b*2 and on the other side of Chl *a*4. Chl *a*3 is co-ordinated by Gln 197. The porphyrin ring is moved by 1 towards the helix D, while keeping the same position in plane if compared to the position assigned in the EM model.

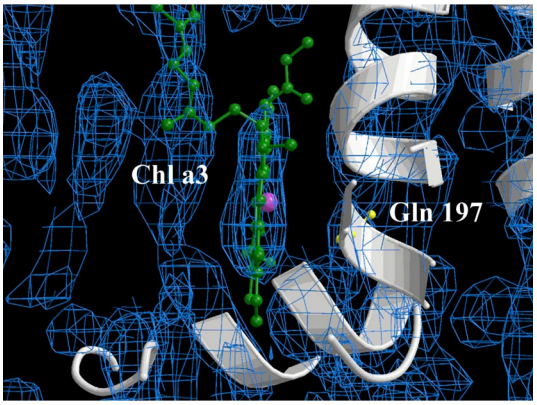


Fig. 3.17a: Chl a3 side view.

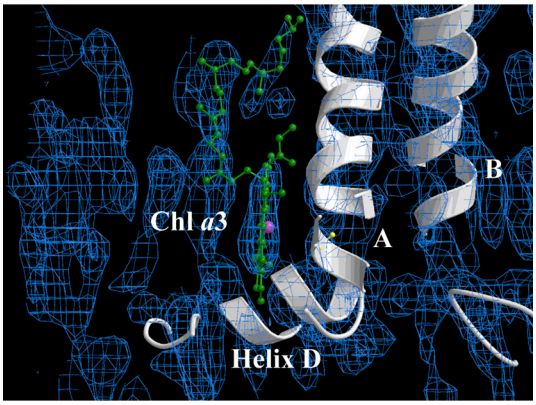


Fig. 3.17b: Chl a3 side view.

The distance between the pair Chl *a*3 and Chl *b*3 is 8.8 .

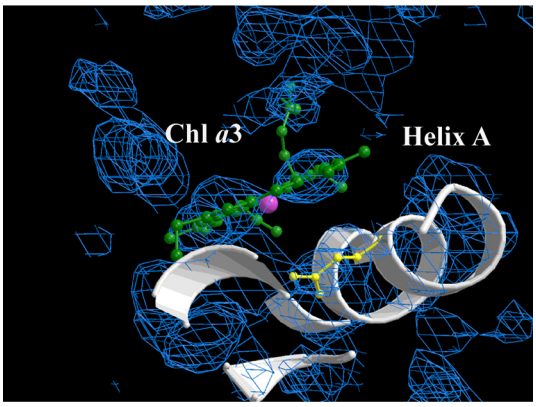


Fig. 3.17c: Chl a3 top view.

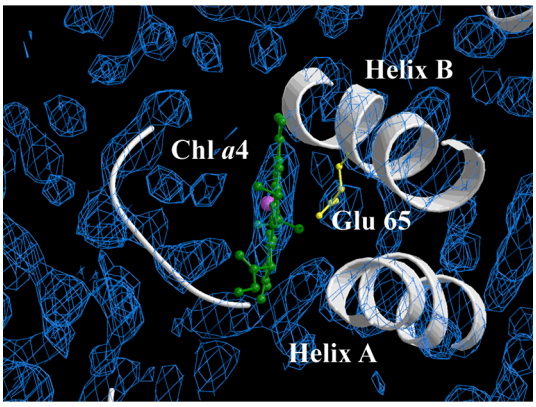


Fig. 3.18: Chl a4 top view.

Chlorophyll *a*4

Chl *a*4 sits in the upper part of the monomer and is co-ordinated by Glu 65 (fig. 3.18). Unfortunately for this chlorophyll the fitting of the phytyl chain is not obvious. However, it is likely that the hydrophobic carbon chain extends downwards pointing to the nearby symmetry related monomer. Moreover, the porphyrin ring is found in the same plane but shifted by 1 towards Chl *a*5, if compared with the EM model. The distance between Chl *a*4 and Chl *a*5 is 10.3 . The lutein 2 is 3.7 away from the central Mg.

Chlorophyll *a*5

The fitting of the whole phytyl chain was unambiguous since a continue density is prolonging from the chlorin ring of Chl *a*5, which is co-ordinated by His 68 (fig. 3.19a, b) and sits close to the centre of the trimer facing the internal three-fold symmetry axis. The phytyl chain has 'S' shape and extends on the opposite side respect to His 68 and to Chl *b*5. It points towards the centre of symmetry in two turns while keeping a distance to the ring of 5.6 . Moreover, Chl *a*5 is moved aside by about 1.5 in direction opposite to Chl *b*5 when compared to its coordinates in the EM model. The porphyrin ring of Chl *a*5 lies in a plane which forms an angle of 45… with respect to the plane of Chl *b*5, in contrast with the model where the angle between the two planes is 30…. The Mg-to-Mg distance is 9.6 and the lutein 2 is 6.3 away from Chl *a*5.

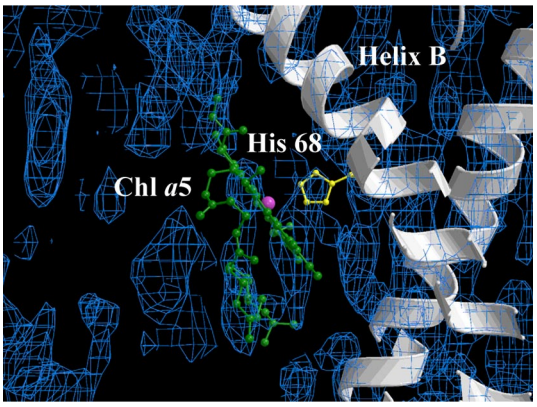


Fig. 3.19a: Chl a5 side view.

Chlorophyll *a*6

The phytyl chain was unambiguously traced up to C13 since the electron density appears to be continuous protruding from the chlorin ring. The Chl *a*6 has one of the best defined electron density. The phytyl chain bends at position C3, pointing straight upwards to the centre of the lipid bilayer. The carbon chain is perpendicular to the membrane plane and parallel to the plane of Chl *a*6; moreover it stays at 3.8 from the porphyrin ring of Chl *a*6 and 4.7 from Chl *b*6. Chl *a*6 is (fig. 3.20a, b) coordinated by Gly 78 and the head group of lutein 2 is 2.3 away. The planes of Chl *a*6 and Chl *b*6 chlorin rings form an angle of 80… between them. Thehlorophylls are situated between the two α-helices, C and B. Chl *a*6 is shifted 2.5 from the position to which it was previously assigned, towards Chl *b*6. The chlorophylls Mg-to-Mg distance is 8.8 , whereas the edges-to-edge distance is 2.8 .

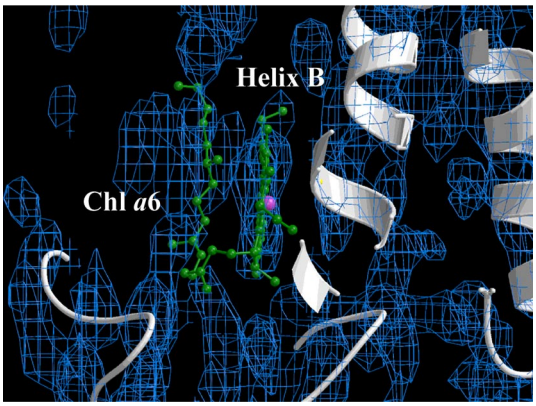


Fig. 3.20a: Chl a6 side view.

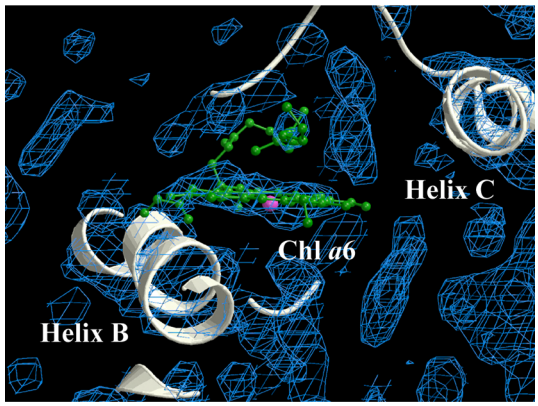


Fig. 3.20b: Chl a6 top view.

Chlorophyll *a*7

The phytyl chain of Chl *a*7 is fitted in a density for all the length of the twenty carbons. This chlorophyll is between the α-helix C and the α-helices A and B domain and at the interface between two consecutive monomer in the same trimer. In that regions a strong density is present, which could be easily attributed to lipids. Indeed, the phytyl chain tracing is not unique. The carbon tails is straight pointing upwards far from a hydrophilic area, however one other solution could have been chosen and an improved fitting could assign a different path.

Chl *a*7 is in the lower part of the monomer, next to the lumen (fig. 3.21a, b) and faces the central part of the trimer. It is 8.9 apart from Chl *a*6.

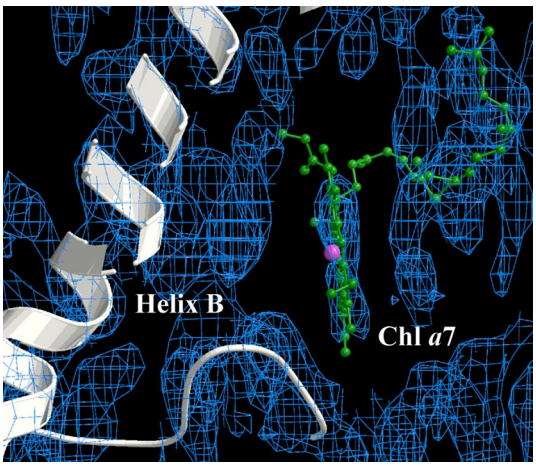


Fig. 3.21a: Chl a7 side view.

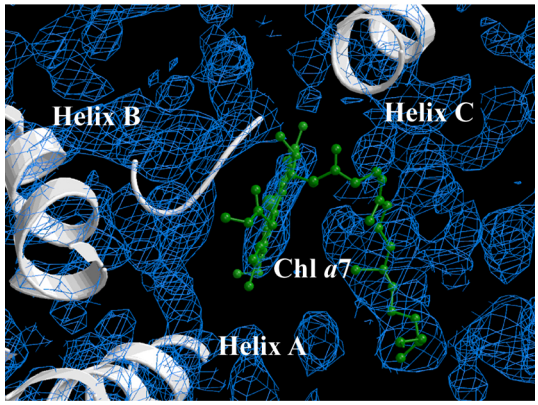


Fig. 3.21b: Chl a7 top view.

The two chlorophylls are almost perpendicular to each other, indeed the planes of the chlorin rings form an angle of 81…. Between Chl *a*6 and *a*7 there is the head group of lutein 2, which is 5.1 and 3.3 apart from the chlorophylls, respectively. Chl *a*7 is shifted 1.2 upwards with respect to the EM model and the distance to Chl *b*5 is 12.0 . The angle between the plane of these two chlorophylls is 30….

Chlorophyll *b*1

Chl *b*1 reveals a strong density relative to the chlorin ring and the phytyl chain, which was fitted in all its length (fig. 3.22a, b). Chl *b*1 is located in the upper part of the monomer, towards the stroma, in an area between helix C and A and exposed to the external part of the crystallographic trimer. The electron density attributed to the phytyl chain was continuous from the chlorin ring and was relative easy to follow the

fitting of the carbons atoms, which point downwards to the centre of the lipid bilayer. The phytyl chain has a high hydrophobic character and it is unlucky that it points towards the external aqueous region. Moreover, it extends opposite to helix C. It bends at position C6 pointing upwards to the trimeric centre of symmetry, close to residue Ser 160; it is perpendicular to the membrane plane and parallel to the chlorophyll ring, to which it has a distance 8.4 . Chl *b*1 is at 12.4 from Chl *a*1 and the planes of these porphyrin rings form an angle of 135…Chl *b*1 is moved 1.7 up towards the loop and residue Gly 158 compared to the two-dimensional crystals model. In the EM model the Mg-to-Mg distance is 10.5 , while in the X-ray structure is 12.4 however the reciprocal orientation is maintained. The central Mg of Chl *b*1 is at 10.2 to lutein 1.

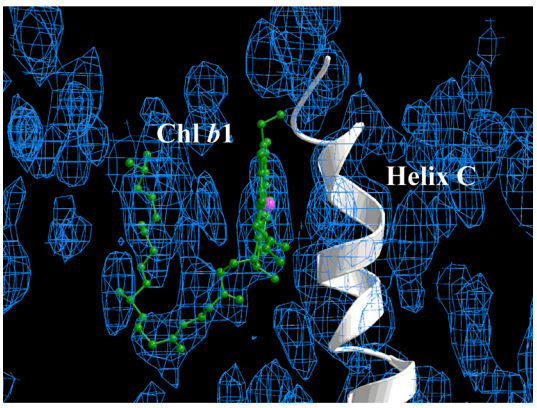


Fig. 3.22a: Chl b1 side view.

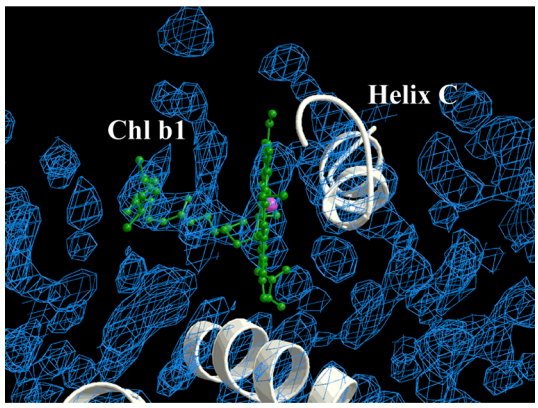


Fig. 3.22b: Chl b1 top view.

Chlorophyll *b*2

The electron density of the Chl *b*2 is well defined and the phytyl chain was fitted in a S shaped density up to C16 and is opposite to helix A (fig. 3.23a ,b). This chlorophyll is in the same position with respect to the assigned one in the EM model. Chl *b*2 is positioned on the side of the monomer, between the N-terminus and helix A. Chl *b2* is 9.9 away from Chl *a*2, whereby the plane of its chlorin ring is parallel to helix A and perpendicular to the ring of Chl *a*2. In this model Chl *b*2 lies in a plane which is rotated along the axis perpendicular to the membrane plane and passing through the central Mg atom, of 77… with respect to the plane assigned in the EM model (fig. 3.23c, d). Thus the Chl *a*2 and Chl *b*2 form an angle of 117….

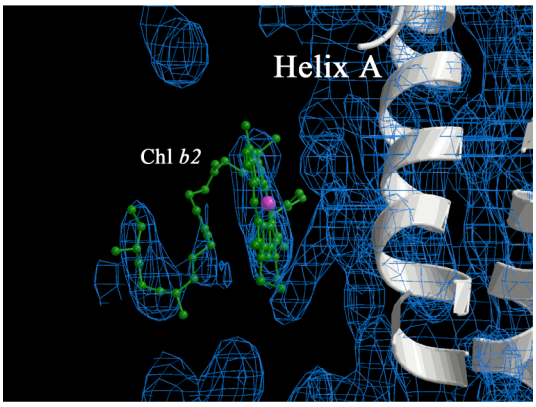


Fig. 3.23a: Chl b2 side view.

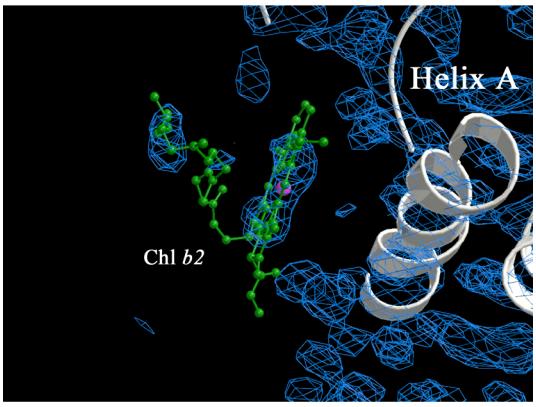
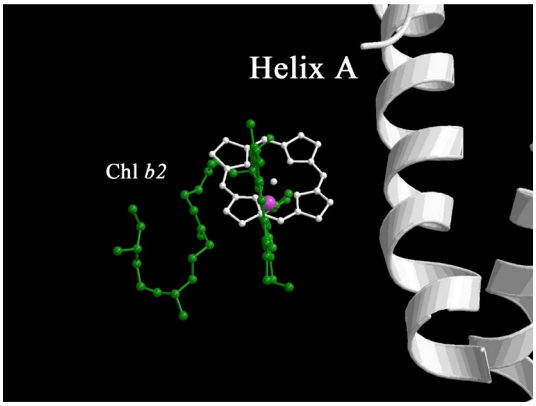


Fig. 3.23b: Chl b2 top view.



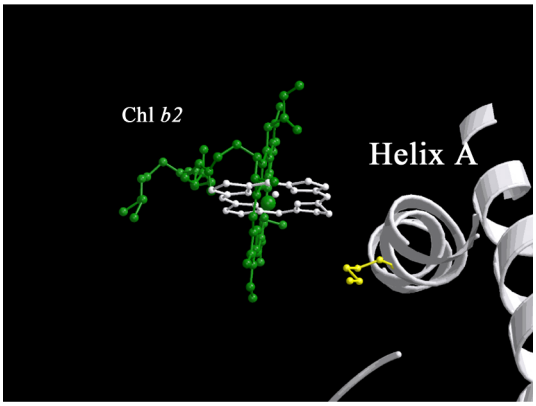


Fig. 3.23c, d: Chl b2, in white Chl b2 from EM model, in green the fitting of the X-ray data side and top view, respectively.

Chlorophyll *b*3

In Chl b3 the density attributed to the phytyl chain allows to fit carbons only up to C7, which extend straight over the chlorophyll head pointing up close to residue Phe 28. Chl *b*3 is co-ordinated by residue His 212 that is 6.6 away (fig. 3.24 a, b). Compared to the EM model the central Mg is moved by 4.3 in plane and the plane of the chlorin ring is parallel to the previous one and yet shifted by 1.8 . The previous distance to His 212 was 3.7 . The angle formed by the planes of Chl *a*3 and *b*3 is 90… while the previous was 72….

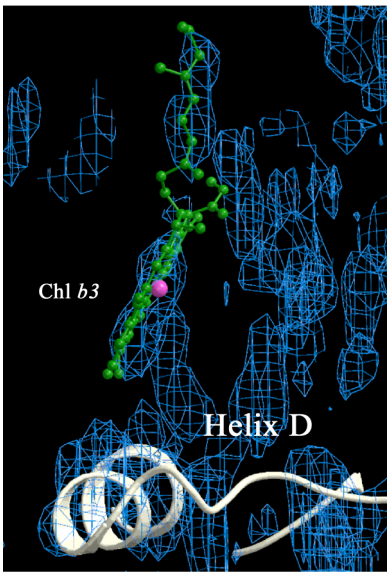


Fig. 3.24a: Chl b3 side view.

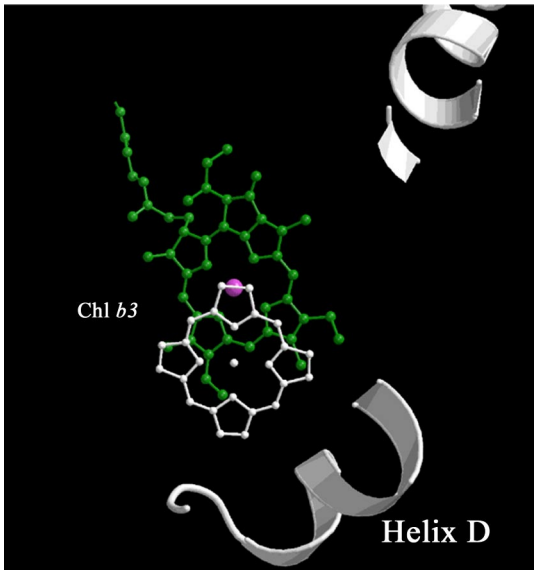


Fig. 3.24b: Chl b3: white EM, green the X-ray model.

Chlorophyll *b*5

The Chl *b*5 is defined as well by a clear density, however, the phytyl chain was fitted only up to C7, while further density is not resolved in this map. The phytyl chain seems to point down in a region where no other density is resolved; it bends over the chlorophyll head pointing to the centre of the lipid bilayer (fig. 3.25a, b). Compared to the EM model this chlorophyll maintains the same position both relative to Chl *a*5 and to the apoprotein. Chl *b*5 is located on the top of the monomer between helices C and B, and it is co-ordinated by the Glu 139, which is 2.9 away.

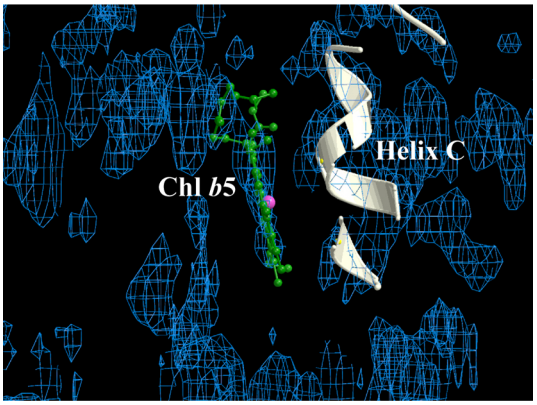


Fig. 3.25a: Chl b5 side view.

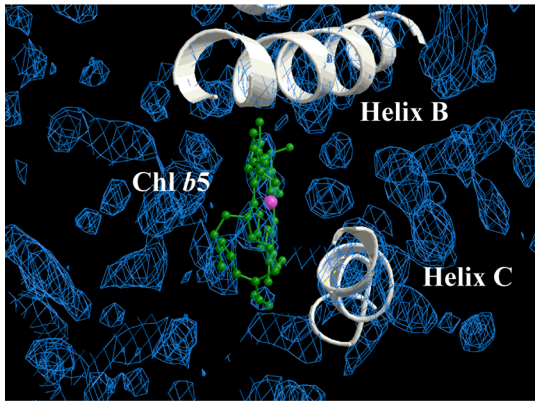


Fig. 3.25b: Chl b5 top view.

Chlorophyll *b*6

The Chl *b*6 is defined by a strong density, on the other side the phytyl chain is tentatively fitted up to C9. In fact, Chl *b*6 is surrounded by density which are not occupied by any other part of the proteins and hence a close electron density area was attributed to the side chain. The phytyl chain points toward the loop between helices B and C (fig. 3.26), close to residue Gly 114 and Asn 115, which are on the lumenal side. This chlorophyll is shifted of 3.1 in opposite direction of Chl a6, compered if the EM model. However these two chlorophylls keep and maintain their relatively distance and orientation both between them and in the position in the monomer. The new position of Chl *b*6 is close to the electron density found and attributed to a third carotenoid, a neoxanthin. The distance between the phytol chain to the Mg is 3.8 while to the Chl b6 is 5,5 . The Chl *b*6 is co-ordinated by Gln 131, it is close to Chl *a*6 and to Chl *a*7 as already discussed above.

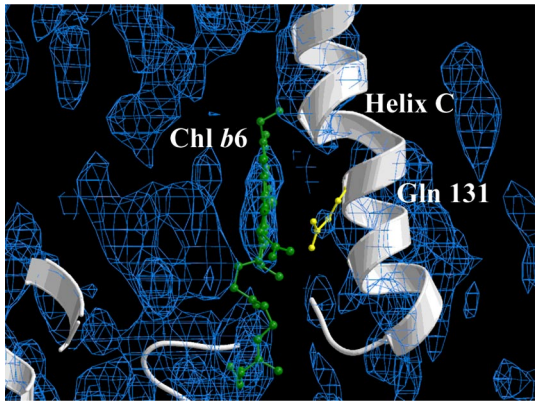


Fig. 3.26a: Chl b6 side view.

Chapter 4

Discussion

4.1 Biochemistry

LHC-II isolated from thylakoid membranes contains three slightly different monomers, which are the translational products of the genes *Lhcb*1-3 (Jansson, 1994; 1999). The most abundant of them has the highest molecular weight of 25 kDa and is the product of the *Lhcb*1 gene. In principle, the LHC-II trimer may therefore be a permutation of all three monomer types. Recently, it has been published that seven different trimeric LHC-II sub-complexes were isolated by non-denaturating isoelectric focusing (ndIEF) from spinach BBY membrane preparations (Jackowski & Pielucha, 2001). The stochiometry of the *Lhcb*1, *Lhcb*2 and *Lhcb*3 gene products was different in each fraction. In carnation and in *Arabidopsis* the *in vivo* existence of the $Lhcb(1)_3$ homotrimer as well as $Lhcb(1)_2$/*Lhcb*2, $Lhcb(1)_2$/*Lhcb*3 and *Lhcb*1/*Lhcb*2/*Lhcb*3 heterotrimers is already known (Jackowski *et al*., 2001). Since the expression of the *Lhcb*1-3 genes is highly regulated by the light intensity, the multiplicity of LHC-II trimers provides the PSII antenna apparatus with the capability of fine tuning in response to variations of the incident light intensity (Walters & Horton, 1999).

Ideally, protein crystallography and particularly membrane protein crystallography requires highly homogenous protein preparations to obtain well-ordered crystals and high resolution data. The purification procedure described in this work yielded LHC-II trimers with reproducible protein concentration and Chl a/b ratio. However, the SDS Comassie-

blue stained gel (fig. 3.1) reveals two bands for each sample. This indicated a heterogeneity of the crystallised protein, which could not be removed by purification. The most abundant band (25 kDa) is the product of the *Lhcb*1 gene. The fainter band has been attributed to a product of the *Lhcb*2-3 gene. Several attempts have been made during this thesis to improve the purification protocol and to obtain a more homogeneous preparation; for example, anion exchange chromatography was performed to separate the various trimer populations. The Mono-Q elution profile (fig. 3.2) shows four different fractions of LHC-II released by the linear salt gradient. There are slight differences among the four fractions. They all show a comparable amount of the main *Lhcb*1 gene product (fig. 3.4) and a second band at slightly lower molecular weight, which is present in different amounts. However, the overall composition of the samples is comparable to the composition of those purified only by sucrose gradient ultracentrifugation.

After anion exchange chromatography, in order to further characterise the fractions, a lipid analysis by 2D TLC was performed for each the sample (fig. 3.3). All the samples showed the same lipid composition, nonetheless three of the major thylakoid membrane lipids were missing from all the fractions.

Thylakoid membranes have specific and characteristic lipid composition: about 50 % is MGDG, 30 % is DGDG and about 10% PG. Each of these lipids plays a specific but different role in the membrane. DGDG binds at the periphery of the trimers and is involved in the contact between adjacent trimers (Nußberger *et al.*, 1993). Moreover, DGDG promotes and allows the trimer to pack into two-dimensional arrays and is essential for the formation of the 3D crystals (Nußberger, 1994). PG is thought to bind at the protein-protein interface in the trimer; it is tightly associated with the complex and is essential for trimer formation (Nußberger, 1994). It is thought that, the head group exposed to the aqueous solution interacts with the polypeptide while the fatty acid tails stay within the hydrophobic part of the protein embedded into the lipid bilayer. Single point mutation analysis confirmed that the sequence between Trp 16 and Arg 21, which contains the PG binding site, is essential to form the trimer (Hobe *et al.*, 1995). In addition, the work of Veverka *et al.* (2000) shows that Gly 18, Asp 20 and Arg 21 are directly involved in the binding of the PG. MGDG has been recently reported to drive

the stacking of the bilayer to form the grana in reconstituted LHC-II lipid membranes (Lee, 2000). In fact, once de-lipidated LHC-II is mixed with MGDG stacked lamellar structures are observed (Simidjiev *et al.*, 2000); on the contrary, once de-lipidated LHC-II is incorporated either with DGDG or PG, the grana-like stackes of membrane are not formed. In addition, MGDG does not promote crystallisation in either 2 or 3 dimensions (Nußberger, 1994).

During anion exchange chromatography almost 90 % of the thylakoid lipids were lost (Nußberger *et al.*, 1993). The extensive washing previous and during column elution removed DGDG, SQDG and MGDG in all four fractions (fig. 3.3). An analogous experiment by Nußberger (1994) resolved five different fractions. The first eluted fraction contained all the thylakoid lipids, the second fractions lost SQDG and PC, while the last three fractions bound only PG. In the experiment performed by Nußberger *et al.* (1993) the first LHC-II fraction was eluted at the void volume, while in this thesis the column has been extensively washed (with almost 100 ml detergent-containing buffer) before initiating the salt gradient and the consequent release of the protein. This means that during the step of column washing, the buffer has removed many of the lipids bound to the LHC-II trimer. Indeed, the two experiments have different results since the elution volume utilised in this work is five times larger than the one used by Nußberger (1994); this is the explanation why the four fractions are completely de-lipidated. However, in both experiments the trimers still bind PG.

Finally, it was found that the successful crystallisation experiments required LHC-II purified only by sucrose gradient ultracentrifugation, which have the lipids associated with the complex and therefore no further lipids were added before the crystallisation experiments.

4.2 Crystallisation

LHC-II is an integral membrane protein which can be easily precipitated in presence of low concentrations of KCl (200 mM) or $MgCl_2$ (10 mM) (Burke *et al.*, 1978; Kühlbrandt *et al.*, 1983). Removal of detergent from a soluble sample of the protein induces LHC-II

to form crystalline sheets with hexagonal symmetry as was shown by freeze-fracture analysis (Kühlbrandt *et al.*, 1983).

LHC-II forms octahedral 3D crystals under several different conditions as has been shown, Kühlbrandt (1987) and § 3.2 of this thesis. These dark green octahedral crystals diffract at most up to 6-7 Å and are not suitable for high resolution structural analysis (Nußberger, 1994). The main task was to obtain 3D crystals of type I (Michel, 1983) which can be thought of as stacks of 2D protein crystals. The LHC-II stack crystals were shown to be in register even if displacement of one layer with respect to one other was occasionally observed (Kühlbrandt *et al.*, 1983). Moreover, these crystals were proved to diffract electrons to a resolution higher than 4 Å (Kühlbrandt, 1987). To achieve more suitable type I crystals a first approach made use of the sparse matrix screening of Zeelen (Bergfors, 1999) and the one purchased at Hampton. Neither gave clear results. However, once small hexagonal crystals were obtained, size improvement took long time since type I crystals are held together by a delicate balance both between hydrophobic as well as hydrophilic interactions. In fact, in the layer accommodating the protein, lipids and detergent molecules participate at hydrophobic and hydrophilic contacts, while in the third dimension only hydrophilic interaction occurs (Michel, 1983). In case of LHC-II a balance between polar and apolar interactions was achieved by adding a small amount of NaCl (10 mM) and equilibrating the drop against 9% PEG 4000. However, the crystals showed a high sensibility to pH.

Furthermore, the improvement of the crystallisation conditions resulted in thin plates, which were shown to be fragile and sensitive to any changes of their environment. As a results, it was not possible to find a harvesting buffer and the only way to harvest a crystals from the drop was to use a loop. As a consequence data collection had to be performed in cryogenic conditions. This was also necessary to reduce radiation damage, which would have been severe at room temperature due to the small size of the crystals, moreover the crystals were too small for data collection in-house with a lab X-ray source.

Among several cryo-protectants tested (table 2.2) only DMSO at the concentration of 30 % (v/v) was useful for data collection, indeed it showed the least harmful effect on the crystals. Moreover, the harvesting procedure had to be as fast as possible and only a thin layer of cryo-protectant, covering the crystal surface was sufficient to prevent ice

formation. Indeed, cryo-protectants are supposed to diffuse inside the crystals and to mix with the solvent solution (Rodgers, 1994; Garman & Schneider, 1997) . On the other hand a long incubation time dissolves the LHC-II crystals and a short freezing time was essential not to disrupt the crystal and to prevent damage. Once the DMSO diffuses in LHC-II crystals the osmotic shock experienced by the crystals resulted in cracking and dissolution. In fact, the crystals diffracted at best at 3.2 Å resolution only during the latest two experiments at the synchrotron, implying that over time better experience was gained and the freezing procedure was optimised.

Nonetheless, it is not a requirement to freeze the crystals for data collection at synchrotron radiation. Indeed, two structures have been recently determined at high resolution using protein crystals mounted in a capillary as for the fumarate reductase (Lancaster *et al.*, 1999), or for the cytochrome *bc*1 complex (Hunte *et al*., 2000). On the other hand, these crystals were larger than LHC-II plates and relatively stable in the harvesting buffer. At room temperature the crystals usually suffer more of radiation damage, but for the fumarate reductase a complete data set was collected from a single crystal, whereas several crystals had to be used for cytochrome *bc*1 complex. In case of LHC-II, each data set was collected from one single crystal under a stream of cold nitrogen.

The low resolution of the diffraction with the X-ray beam parallel to the LHC-II plate suggested poor crystal contacts along the *z* dimension of the unit cell. In fact, when exposed face-on to X-ray radiation, the crystals usually diffracted almost to 4 Å, which was always assumed as a good indication of proper packing at least in the ab plane. The diffraction pattern allows a quick assessment of the quality of the crystal. Indeed, LHC-II crystals showed quite a reproducible diffraction pattern, however so far not at resolution beyond 3 Å and always anisotropic. The data sets were processed utilising the program package XDS (Kabsch, 1993). In table 3.1 the data quality is reported with respect to the R_{mrgd-F}, which indicates the quality of the amplitude of measured intensity. The R_{mrgd-F} values worked out by Diederichs & Karplus (1997), of three different data set of a crystal of the enzyme urease diffracting at 2 Å, are ranging between 9.5 and 10.3%, which correspond to a R_{sym} between 5.5 and 6.0. Usually the R_{mrgd-F} is higher than the R_{sym}.

Concerning the LHC-II crystals the R_{mrgd-F} value is of 9.3 %, which is a good indication of the successful data reduction process.

The two-dimensional crystals of LHC-II belong to layer group p321 and the unit cell is a=b=129.5 Å while the thickness is about 48 Å (Kühlbrandt *et al*., 1994). The three dimensional crystals belong to the hexagonal crystal system and the space group is $P6_322$. In table 3.1 the unit cell dimensions are reported as well. The 2D and 3D have similar ab dimensions of the unit cell. The c dimension is larger than the double of a lipid bilayer, which is as well the thickness of the unit cell of 2D crystals. Therefore, two trimers are accommodate per unit cell in the 2D crystals, while six trimers are present per unit cell in the 3D crystals. Moreover, the lattice parameters of the three 3D crystals are just slightly different. Yet, the c dimension of lhc-I crystal is shorter by almost 3 Å compered to crystal lhc-III, and 4 Å compared to crystal lhc-II. In addition, lhc-I crystal has been measured at the last experiment and showed the least anisotropicity. It is likely that a better and faster freezing caused less disruption of the crystal packing.

One major aspect revealed by data processing is the crystal mosaicity, which is relatively high: 2.2º, for data set lhc-I, while 2.8º both for lhc-II and lhc-III. The large mosaicity denotes that a reflection is spread across several crystallographic frames, consequently the precise position of each spot is not easily determined. However, the data could be processed. Conversely, low mosaicity is an indication of good crystal packing, since it measures the angle necessary for a spot to pass through the Ewald sphere. In reciprocal space a spot represents a set of parallel layers of the crystals lattice. In other words the shape of the spot reflects the order of the actual crystal lattice. In addition, the LHC-II crystals are stacked layers, which might not be precisely in register but translationally displaced with respect to one other, so that interference between layers perturbs the crystals symmetry. The more 2D layers forms the 3D crystals the more packing distortion can be introduced. This is corroborated by the fact that, crystals of LHC-II, isolated from spinach leaves, grow bigger in size and were thicker than the LHC-II crystals from pea, therefore the mosaicity degree was even larger than the one reported in table 3.1 (Zeelen, personal communication).

Mosaicity provides indication of improper crystal packing and this applies both to soluble and membrane proteins. However, membrane proteins are more susceptible to this effect.

The reason may be related to the following: the membrane proteins can be handled only after detergent solubilisation, which interferes with the crystals growth. In addition, membrane proteins are often subject to conformational changes, like membrane transporters, or can have small hydrophilic portions, which do not easily form molecular contacts and hence a strong crystal packing.

The LHC-II crystals have been frozen and this procedure always produces a mechanical stress inducing sometime strong distortion to the packing and increases the mosaicity (Garman & Schneider, 1997). At the synchrotron, a quick and easy experiment was performed in order to improve both mosaicity and resolution. It has been reported that thawing and rapid freezing cycles of the crystals under the stream of liquid nitrogen could decrease the degree of mosaicity as well as increase the diffraction limit (Yeh & Hol, 1998). This process is called flash annealing. However, when it was tried on LHC-II crystals a loss of diffraction power was observed and no improvement of the mosaicity degree was noticed. The successful flash annealing has so far been reported only for soluble proteins. Once mounted in a loop LHC-II crystals are held in place by surface tension of the solution, which is mainly 30 % DMSO. Once the stream of cold nitrogen is stopped the crystals are thawed together with the cryoprotectant. The DMSO diffuses into the crystals, resulting in an osmotic shock and a consequent disruption of the crystal contacts. In other words, during flash annealing the LHC-II crystals would be exposed to similar conditions as when they are left for a long time in the cryo-protectant solution, before flash freezing.

An integral membrane protein, which has a similar crystallographic characteristic behaviour to LHC-II is bacteriorhodopsin (bR). This is a light-driven proton pump in Halobacteria and forms crystalline patches in the cell membrane (Henderson *et al.*, 1990). To my knowledge only the bR crystals, diffracting up 2.9 Å resolution, have mosaicity values comparable to LHC-II. In fact, the bR crystals, for which structure resolution is carried out, are reported to have mosaicity ranging between 1.3º to 2.0º (Essen *et al.*, 1998). Indeed, looking at the data quality of other protein structure determination of crystals of both soluble and membrane proteins, under cryogenic conditions, the mosaicity is usually not reported meaning that it is a negligible value and usually lower

than 1 degree. For example, the 30S ribosomal subunit, crystals are reported to have mosaicity lower than 0.64 (Clemons *et al*., 1999). The bR protein forms well ordered two dimensional crystals and the structure was solved by electron crystallography studies (Henderson *et al*., 1990; Kimura *et al*., 1997). Three dimensional crystals were reported to diffract to 3.6 Å resolution, however, due to a poor packing along the third dimension the resolution of diffracted X-ray beam perpendicular to the c axis was 6 Å (Schertler *et al*., 1993). However, these crystals were not stacks of two dimensional protein layers. Recently, three dimensional crystals were obtained and the structure was solved at 2.5 Å (Pebay-Peyroula *et al*., 1997) inducing protein crystallisation in a newly designed bi-continuos lipidic cubic phase (Landau & Rosenbusch, 1996). Pebay-Peyroula *et al.* (1997) reported that 9 crystals out of 10 exhibited a large mosaicity and in addition they observed that thin crystals showed the least degree of mosaicity. More recently an improvement of the crystal resolution to 1.9 Å was achieved (Belrhali *et al*., 1999) and to 1.55 Å (Luecke *et al*., 1999); this latter work revealed a mosaicity of only 0.61º. The stacking along the c axis is due to polar interactions between the cytosolic and extra-cellular loops of adjacent layers. The new approach of the crystallisation established by Landau *et al*. (1996) seems to be most promising for LHC-II since the extra membrane part is formed only by relative short loops as is in the case of bR. Further crystallisation experiments by lipid cubic phase should be performed with the hope to improve the crystals packing and finally obtaining data at resolution higher than 3.2 Å.

4.3 Molecular replacement and crystals packing

The structure of LHC-II has been resolved at nearly atomic resolution by electron cryo-microscopy of two dimensional crystals which diffracted to 3.4 Å in the ab plane accommodating the trimer (Kühlbrandt *et al.*, 1994). However, the resolution along the *z* axis, perpendicular to the membrane plane, is only 4.9 Å.

The 3D crystals instead diffracted up to the resolution of 3.2 Å and the data completeness is 92.3 %. This should be sufficient to improve the current electron microscopy model and to discover some aspects that were not revealed previously.

Two major points were in favour of solving the "phase-problem" by molecular replacement (MR). First because a structure model of the LHC-II calculated by electron cryo-microscopy was available and structural information from an EM model were already used to solve an X-ray structure. For example, the X-ray structures of bR at 2.9 Å (Essen *et al.*, 1998) and at 2.5 Å (Pebay-Peyroula *et al.*, 1997) as well as the atomic structures worked out by Luecke *et al.* (1999) and by Belrhali *et al*. (1999) were solved using the phases from the electron microscopy model of Grigorieff *et al*. (1996). One other possibility of undertaking this problem would have been to attribute precise phases by multiple isomorphous replacement (MIR). However, preliminary experiments to add heavy atoms both to the crystallisation solution and directly to the drop containing grown crystals already LHC-II crystals failed. In the first case no crystals formation occurred. In the second case, as expected the thin LHC-II already formed were disrupted.

Much time and effort have been spent during this thesis on finding an appropriate solution by MR. In fact, the MR procedure computes an initial search for the correct rotation orientation of the model and then a search for the correct translation in the experimental unit cell. The procedure is usually divided into two different steps to avoid enormous computational cost. Furthermore, during the rotational step only a subset of all Patterson vectors is superimposed; thus the high correlation found in the rotational search might not correspond to the correct solution, and then the translation search is highly sensitive to any orientation error that occurred in the rotation search.

A first attempt was made with the program AMoRe (Navaza *et al.*, 1994). The program AMoRe is one of the most frequently used and common programs for solving the phase problem by molecular replacement (Stock *et al*., 1999; Zhou *et al*., 2001). It relies on a large number of solutions for each rotational search (in this experiment up to 100) and the best of those in terms of quality are chosen as input of the subsequent translational search. In all the trials made, the proposed solutions of the rotational search were not providing a strong peak or indication of an evident solution. Additionally, the best solution provided by AMoRe were unrealistic due to overlapping symmetry mates.

The second program used was the most recent version of Brute (Fujinaga & Read, 1987). This program, based on a brute force approach, did not provide any reasonable solution either, all of which showed overlapping of the symmetry-related monomers.
Finally, only by utilising a six-dimensional search a solution to the molecular replacement was obtained. The results were chosen according to the highest correlation coefficient, as defined in § 2.6.2, and the lowest R-factor. The solution of the molecular replacement by EPMR showed a correlation coefficient (CC) value of 36.1 % and a R-factor of 54 %. Both the correlation coefficient and the R-factor were comparable with respect to the EPMR solution of a cytochrome c' where structural data, ranging between 15 to 4 Å, were used to solve a structure with CC of 51 % and a R-factor of 59.8 % (Kissinger *et al*., 1999). Recently the structure of a twinned crystal of a 288 amino acid soluble protein, the methylene-tetrahydromethanopterin-dehydrogenase (MtdA) isolated from *Methylobacterium extorquens*, has been solved. The crystals diffracting at the resolution of 2.5 Å contained three trimers in the asymmetric unit. The phase-problem was solved by molecular replacement using the program EPMR. In this case, the successful solution, for the first trimer, was given at the CC of 20 % and at the initial R-factor of 84 %. However, including the second and third trimer, CC increased to 52 and 67% and the R-factor decreased to 69 and 72 %, respectively. The structure was finally refined using the CNS program package (Brünger *et al*., 1998) to a final value of R-factor of 23.7% and R-free of 27.5% (Warkentin, personal communication).

In the case of LHC-II, the best solution found was correct with respect to the fact that each monomer was related to the other two monomers and no structural overlapping occurred. A second and independent phase search was performed using the whole trimer as a search model. As in the case of the monomer, the symmetry-related trimers were found to be in correct position with respect to each other and, most convincing to coincide with their symmetry related mates around the three fold axis. However, it might be that the solution found with MR is not the best estimation of the true phases consequently including a bias towards the known EM structure. In case a reasonable model is available, MR is usually thought as the fastest solution to the “phase problem”,

however multiple-isomorphous replacement or anomalous dispersion provide unbiased phases.

Figure 3.8 represents the α-carbon chains of LHC-II, where the typical organisation of type I crystals is depicted. There is no evident contact between two consecutive layers, which have a minimum distance of 14 Å. It is unlikely that only the solvent fill the space between two consecutive layers. However, it can be that two dimensional arrays form 3D stacks by polar interaction mediated by low ionic strength. In the light microscope, once crystallisation conditions were optimised, the crystals looked to have always the same shape and dimension. However, more than 400 crystals were tested and only three showed a reasonable crystal packing in the third dimension. Supposedly the protein layers stack by contacts in the third dimension and only occasionally the layers are completely in register and properly packed. This may explain why, the harvesting procedure was crucial to maintain this delicate equilibrium, and was only successful three times. The gentle harvesting using the loop probably did not impair the crystals array. Subsequently the flash freezing would force the layers in a fixed position.

4.4 Refinement

Refinement was a difficult and time-consuming task in the process of structure determination, since LHC-II crystals have a fairly large unit cell and a comparatively large number of reflections with a low signal-to-noise ratio. Moreover, crystals diffract only up to 3.2 Å resolution. Data reduction resulted in 10727 independent X-ray reflections. The parameters to be refined are the three positional co-ordinates (x, y, z) and the temperature factor (B) for each of the 1952 atoms of the LHC-II monomer, making a total of almost 8000 parameters. Therefore, the ratio of observations to parameters is ~1.3, which is a poor over determination. This is the reason why as many as possible additional observations are necessary. The stereo-chemical data on the amino acids are stored in the CNS program package library, however for the chlorophyll and the carotenoid molecules the constrains relative to bond lengths and angles had to be derived from a high resolution structure. A further additional parameter, which can be considered

an "observation" is the fact that the bulk solvent area is highly disordered and it should appear as a flat region in the electron density map. As a consequence a solvent flattening procedure was applied to the map.

The refinement of LHC-II at 3.4 Å calculated by electron microscopy studies resulted in a R-factor of 33.0% and R-free of 37.9 % (Kühlbrandt *et al*., 1994). For the 3D crystals the refinement protocol from Brünger *et al.*, (1987) was followed and an initial rigid body refinement was performed. First, the whole molecule was considered as an entire entity and only six parameters (three translational and three rotational) had to be refined. During this procedure the R-factor improved from the initial value of 54 % to 50.25%. A second step of rigid body was performed with the difference to the previous, that the four α-helices A, B, C and D as well as the two luteins and the twelve chlorophyll molecules were assigned to single entities, in other words allowed to move independently with respect to the other. Unfortunately, even if during the second cycle of refinement a larger number of parameters had to be refined, the R-factor did not improve as much as it did during the first refinement, where only few parameters were considered. At last the simulating annealing improved the R-factor from 50 % to 35 % and the corresponding R-free to 45%. At this point the phase improvement by solvent flattening was performed and a new density map calculated. This new map became clearer with respect to the previous and evidently new features were seen. However, when MR is used, it is not an easy task to overcome the phase bias and to obtain an R-factor of a value low enough to have a reliable map. Indeed, if any differences are observed from the starting model it would be difficult to decide if these were real. Conversely, if no difference were observed there will remain the doubt of a strong model bias. Hence, in this case one useful method of lowering the phase bias is the omit map calculation.

According to the Fourier-transform property every point in real space is influenced by every point in reciprocal space and vice versa. In case that the rest of the model is correct, the phases calculated for the omitted portion will be close to true phase. Problems arise when the model is incorrect, since the phase bias are spread throughout the whole model. To overcome this problem it is better to remove the portion before refining the model. Using a first orientation of LHC-II monomer in the unit cell the coordinates relative to the Chl *a*6 were removed and after a cycle of refinement the

density map was once again calculated. This map results in a density peak in correspondence to Chl *a*6. The same procedure was performed for the ten amino acids of the central part of α-helix B. The interpretation was that any of the new features was not caused by a bias of the model nor by noise. At the current resolution, an R-factor lower than 30% cannot be expected, but on the other hand the R-free is still to high. Nevertheless, a preliminary fitting of new structural features was attempted. This showed some interesting new features of the LHC-II structure.

4.5 The apo-protein

In LHC-II, three trans-membrane α-helices A, B, C and the amphipatic α-helix D count for only 36 % of the whole protein: the loop area is therefore a major part of the structure. This large part is not unambiguously determined in the EM model (Kühlbrandt *et al.*, 1994). The loop between the B and C α-helices has 32 amino acid residues. The loop between α-helices C and A consists of 26 residues; 54 amino acids form the N-terminus and 17 amino acids the C-terminus. The α-helices A and B fit very well to the X-ray density map calculated after solvent flattening, but new interesting aspects are revealed near the middle of the α-helix C.

4.5.1 Solvent-exposed parts of the protein

The N-terminus is at the stroma side and is directly involved in the regulatory function of LHC-II. The N-terminus is phosphorylated by a kinase during the regulation of the light energy transfer between the PSII and PSI (Allen, 1992). Whenever an excess of solar energy is delivered to PSII an acidification of the lumenal side occurs, which leads to an activation of the kinase. The mechanism proposed involves a lateral movement of LHC-II from the stacked grana to the lamella region, where the PSI is mainly present. It is clear that an elucidation of the exact position of the N-terminus is of great interest. The N-terminus is poorly defined in the LHC-II model calculated by EM, so that the first 25

amino acids could not be accommodated. In the X-ray map calculated at 3.2 Å, a prolongation of the electron density allowed a fit of one more amino acid (Lys 25). It has been already described that the N-terminus is essential for the trimerisation of the complex (Hobe *et al*., 1995) and it is relevant to assign further amino acid positions. This could also bring elucidation on the position of the phospholipid PG, which is known to be essential to keep the trimer intact (Nußberger, 1994). In addition, it would be interesting to see if this lipid is exactly positioned parallel to the three fold symmetry axis of the trimer and to visualise the position of the N-terminus with respect to the neighbouring monomers.

On the stromal side, the AC loops have been assigned by EM studies between residues Ala 144 to Asp 169. However, the assignment of residues between Ala 144 to Leu 148 does not correspond to any electron density of the X-ray map. Moreover from Gly 149 to Leu 155 some amino acids do not seem to be in the best fitting position either. This maybe due to an average of different diffraction patterns resulting from different population of monomers within the crystals. In fact, between Ala 144 and Leu 155 the sequences of the *Lhcb*1-3 genes has some of the lowest homology, whilst between and Leu 155 to Asp 169 it is well conserved at least for the pea genes.

On the lumen side the BC loop have been assigned in the EM model between Val 90 to Gln 123, however between the residues Ala 100 to Pro 116 an unambiguous fitting was not proposed. The X-ray model at 3.2 Å does not reveal any strong density in this region and so loop cannot be assigned yet. Possibly, this relative long loop is too flexible to obtain a clear structure information even if there is an high homology between the *Lhcb*1-3 gene.

On the C-terminus, the superposition of the EM model and the X-ray map revealed that the loop between Val 200 to Pro 205 is properly fitted, however, after Asp 215 the apoprotein probably follows a different path. It will be necessary to obtain better data at higher resolution to finally identify the extra membrane portion of the LHC-II complex.

4.5.2 Helix C

A new aspect of the structure of LHC-II was discovered by the X-ray structure. This new features concerns the membrane-spanning α-helix C.

The EM model structure analysis proposes a fitting of the α-helix between Ser 123 and Ile 143 spanning the membrane layer in 5.5 turns. Indeed, the electron density calculated at 3.2 Å resolution by X-ray crystallography does not have a continuos cylindrical shape, instead there are two short cylinders along the same density. At position of Ile 133 (referred to as assigned in the EM model) the α-helix is unwound. The new fitting proposes to split the α-helix C into two short α-helices. The three amino acids involved in the unwinding region are Val 132, Ile 133 and Leu 134. These three residues are strongly conserved among the *Lhcb*1-3 genes (see appendix A). The α-helix C is almost perpendicular to the membrane plane (Kühlbrandt *et al.*, 1994). In the electron microscope experiment the maximum tilting angle was 60º, resulting in a loss of information referred to as the missing cone and affecting the resolution, which is 4.9 Å in the *z* direction, perpendicular to the plane of the 2D crystals. Due to this fact all the electron densities appear to be elongated in the *z* direction. This means that if the two α-helices are elongated along *z* their own electron density will merge in the centre of the lipid bilayer and only a long density will be seen: consequently the unwound part was obscured. The less anisotropic three dimensional crystals are not affected by any elongation along the *z* direction so that this new features was discovered.

The biological reason for this structural feature is probably related to the arrangement and co-ordination of the chlorophyll molecules and a third carotenoid, which has been discovered, during this work, close to the α-helix C. In LHC-II , along the α-helix C are two residues, the Gln 131 and Glu 139, which directly co-ordinate Chl *b*6 and Chl *b*5, respectively (Kühlbrandt *et al.*, 1994) and recently confirmed by single point mutation analysis (Rogl & Kühlbrandt, 1999). In addition, a carotenoid is in van der Waals contact with the hydrophobic residue Lue 134. In the LHC-II structure, the distortion of α-helix C could be explained by the vicinity of the chlorophyll and carotenoid molecules in a way that these pigments impose a restriction to the helix position and configuration. Indeed, the Gln 131 and Glu 139 do not directly participate in the unwound region, however they

could be forced in the place by the chlorophyll molecules and consequently stretch the α-helix. Anyway, the unwound region has to be confirmed at better refined value.
It has been reported that α-helices are deformable and not to be considered only rigid and linear rods (Schulz & Schirmer, 1979). Moreover, Barlow and Thornton (1988) showed that 85% of 48 α-helices, of known protein structures, refined between 1 and 2 Å resolution, are in some way distorted. It has been reported a kink in the region of an isoleucine residue in the α-helix of the erythrocruorin enzyme (Steigemann & Weber). This kink is considered to be necessary since the residue Ile makes contact with the haem group, which imposes its own restriction on the helix position. An additional example that was reported is about two kinks found in two helices of the thermolysin enzyme. The kinks are in the region of a valine and of a glutamic acid, both residues are involved in the co-ordination of the Zn active site (Barlow and Thornton, 1988).
In the EM model, the α-helix C starts at Ser 123 on the lumenal side. However, in the X-ray map the electron density continues further allowing one more turn to be fitted on the lumenal side up to His 120. As a consequence the BC loop would begin at Val 119, while in the EM model it starts at Gln 122 and in addition this loop would have a completely different orientation assignment. No new fitting of this regions has been carried out. In the EM derived structure, at the stromal side the helix C terminates at Ile 143. In contrast, in the X-ray map a prolongation of the density is revealed, and one more amino acid, Ala 144, could be accommodated. The AC loop would then start at the position of Gly 145, which may give a more flexible character to the loop.

4.5.3 The third carotenoid

Next to the α-helix C a long, well-defined electron density was revealed. This density is continuous, straight and 28 Å long. This is similar to the 30 Å long region assigned to the lutein molecules in the EM model (Kühlbrandt *et al*., 1994).
The carotenoid molecules in light-harvesting complexes fulfil three major function (Frank *et al.*, 1999).

As antenna pigments collect light energy mostly in the blue-green spectral region and transfer energy to nearby chlorophylls. Light energy is further funneled to the photosynthetic reaction centres.

As structural components they stabilise light-harvesting complexes. In fact the EM model of LHC-II revealed that the two lutein molecules span the lipid membrane and connect the cluster of chlorophylls, which are at both side of the thylakoid membrane. As well it was shown that the two α-helices A and B are stabilised by the two carotenoids molecules bound into the grooves of the superhelix. Moreover, their important stabilising function is suggested from the reconstitution experiments in vitro. These experiments proved a restricted preference of carotenoid to assist the assembly process and the correct folding of the apoproteins (Rogl, 2000; Yang *et al.*, 1999).

The carotenoids serve as well to prevent the formation of the triplet state: the head groups of the lutein molecules, a cyclohexane ring, are in van der Waals distance to Chl a molecules and they provide photoprotection by quenching excited triplet states in chlorophyll molecules and singlet states in oxygen (Kühlbrandt *et al.*, 1994).

It has been reported that three xanthophyll molecules, two luteins and one neoxanthin, are bound to LHC-II in high affinity sites while a forth site, with lower affinity can be occupied by a violaxanthin depending on the genotype and the physiological state of the plant (Ruban *et al.*, 1999). The location of the xanthophyll sites differs: two are internal, which binds luteins, and one is peripheral, which binds neoxanthin or violaxanthin. The binding of neoxanthin is a peculiar characteristic of LHC-II, which binds almost 90% of thylakoid neoxanthin while *Lhcb*4 and *Lhcb*5 bind sub-stoichiometric amounts of this pigment; *Lhca*1-4 and *Lhcb*6 protein do not bind it at all (Croce *et al.*, 1999). Rogl *et al.* (1999) demonstrated that LHC-II apoprotein can be refolded *in vitro* only in presence of two lutein molecules: the occupancy of the neoxanthin site is not essential for protein refolding. In the same work, single points mutation analysis suggested that the binding site of Chl *b*5 (Glu139), *b*6 (Gln 131) and *a*6 (the couple Leu77-Gly78) are close to the neoxanthin binding site. The yield of neoxanthin in the Chl b mutants was about half of the wild type while the mutation Chl a6 had lost all the neoxanthin. A parallel experiment of Croce *et al.* (1999), using Chl b5 and b6 mutants, is in excellent agreement with this work. It has to be taken into consideration that, the double mutant, Leu 77-Gly

78, promotes destabilisation of the trimers, i. e. that a mutation on α-helix B induces a distortion of the overall structure effecting other binding sites as well. Therefore, it cannot be concluded that these two residues have direct effect on the complete loss of neoxanthin (Rogl & Kühlbrandt, 1999).

The proposed neoxanthin position (Croce *et al.*, 1999) is next to α-helix C, and is 14.5 Å away from Chl a6, with Chl b6 in between. According to these authors, based on linear dichroism analysis, the carotenoid polyene chain forms an angle of 57 ± 1.5^{o} with respect to the normal of the membrane plane and the pigment is localised between the α-helix C and the domain of α-helices A and B. However, the X-ray structure solution shows this carotenoid almost perpendicular to the membrane plane. Other carotenoids span the lipid bilayer being less tilted with respect to the normal of the membrane plane: e. g. the rhodopin-glucoside in the LH2 isolated from *Rps. acidophila* (McDermott *et al*., 1995) or the lycopene in the LH2 isolated from *Rs. molischianum* (Koepke *et al.*, 1996). Moreover, the lycopene molecule, in the bacterial complex, is held in place by a cluster of aromatic group like Trp and Phe, which are in van der Waals contacts forming an anchor site for the chromophore. In LHC-II, the side chains of the Thr 130 and Leu 134 as well as Val 138 are exposed directly to the neoxanthin binding pocket and are in van der Waals contact to the carotenoids, which is surrounded by lipids on the opposite side. As already reported in § 3.1 many different lipids are bound to the LHC-II trimer. Nonetheless the long electron density is very similar to the density already assigned to the luteins in the centre of the monomer by EM studies. The carotenoid molecule have a strong delocalisation of the π electrons along the C chain, which will contribute to a strong X-ray diffracting signal. Such effect cannot be expect from the lipid molecule. Indeed, it is reported that the lipid DGDG is essential for the crystal formation and it is expected to be found in the electron density map. However, this elongated density which has been here discovered is next to the α-helix C which is not in close contact to the next trimer, at least in this crystal packing. Therefore it is unlikely that this electron density could be assigned to a lipid molecule.

Peterman *et al*. (1997) show that no direct energy transfer from carotenoid to Chl b happens, while it occurs between carotenoid and Chl a on ~220 fs time scale (Peterman *et al*., 1997) and in agreement with Kühlbrandt *et al*. (1994). By contrast Connelly *et al*. (1997) suggested energy transfer from the carotenoids to Chl b and then from Chl b to Chl a, proposing to change the assignment of carotenoids according to Kühlbrandt *et al*. (1994). Presently, there is no evident need of changing the current assignment of the pigment molecules, as the correctness of several chlorophyll assignment have been confirmed by single point mutation analysis (Rogl, 2000) but still at 3.2 Å resolution, it is not possible to distinguish between Chl *a* and *b*. As a result of the X-ray analysis the position of the neoxanthin molecule is in van der Waals contacts together with Chl *b*6 and Chl *b*5. In summary, a less tightly bound carotenoid in the vicinity of the Chl b explains the results of Connelly *et al*. (1997) whilst the assignment of the Chl a close to the central molecules of luteins is in agreement with the results of Peterman *et al.* (1997). Therefore, it is concluded that this density represents the third carotenoid in LHC-II.

4.5.4 The Chlorophylls

The LHC-II structure calculated at 3.4 Å by EM studies allowed for the first time the assignment of twelve chlorophylls. Solving the X-ray structure at 3.2 Å made it possible to finally assign phytol chains to most of the chlorophylls bound to LHC-II. The phytyl chains are essential structural information. Due to their non-polar composition they give the chlorophylls their strongly hydrophobic character. In LHC-II, the Chls are placed in two separate layers parallel to the membrane plane; one plane, build of seven chlorophylls, is near to the stromal side, the other constituted of five chlorophylls, is close to the lumen. The space between these two planes is highly hydrophobic and contains the phytyl chains. This is an important improvement of the earlier EM structure. In fact, the phytyl chains attributed to all the chlorophylls except for Chl *a*1, will allow the calculation of the orientation of the Q_x an Q_y transition dipoles within each of the chlorin rings and will greatly contribute to the elucidation of energy transfer mechanism within monomer and the LHC-II trimer.

The phytyl chains strongly anchor the chlorophylls to the membrane. They have different shapes and most of them are bent over the plane of the chlorophyll rings, which are close to the aqueous solution, however all the phytyl chains point toward the centre of the lipids bilayers. The phytyl chains were fitted entirely for Chl *a*2, *a*3, *a*4, *a*5, *a*7 and Chl *b*1. Only 14 carbons atoms were assigned to Chl *a*6 and *b*2, seven to Chl *b*3 and *b*5, and nine carbons to the phytol chain of Chl *b*6.

The phytyl chains of three chlorophylls are straight and stretched into the lipid bilayer, and these belong to Chl *a*4, *a*3 and *b*3. However, the Chl *b*3 phytyl chain is visibile only up to C7, the Chl *a*4 has less clear assignment and has to be confirmed by a better refinement value. In fact, Chl *a*4 is close to the centre of symmetry of the trimer and it was not possible to distinguish a continuos density aside the density of the chlorin ring. In addition, this area shows strong densities. They could belong to a lipid or the part of the N-terminus still missing at the lumenal side, however at the present stage it is not possible to assigne further structural features with confidence.

Apart from the ones described above, all phytyl chains are bent or folded over the chlorophyll porphirin rings. This conformational arrangement recalls the one of the BChlB in LH2 (McDermott *et al.*, 1995). In this protein, the chlorophyll phytyl chains are wrapped around the carotenoid and have mutual interaction between them, which strongly stabilises the complex (Freer *et al*., 1996). In LHC-II, the phytyl chains have a different organisation. Indeed, they are oriented towards the external part of the monomer, being opposite to the co-ordinating residues and in general in a direction opposite to the apo-protein skeleton. The two luteins are deeply inserted in the lipid bilayer and sit in the groove of the coil-coiled region of helices A and B, in a way that only the porphirin rings of the chlorophylls are facing the carotenoid. In this arrangement the phytyl chains are oriented towards the adjacent monomer of the same trimer: the carotenoids organisation is depicted in fig 4.1a and b.

The monomer has an internal pseudo two fold symmetry, perpendicular to the membrane plane, which relates the first 20 amino acids of the helices A and B and as well the Chl *a*3 and *a*6 (Green & Kühlbrandt, 1995). However, the phytyl chains seem to break this symmetry and to have different shape and orientation. The phytyl chain of Chl *a*3 is

straight up the chlorin ring while it is bent at C1 for Chl *a*6, giving the two chlorophylls different dipole moments which influence the spectroscopic properties of the chlorophyll. The densities for Chl *b*5 and *b*6 were found to be in the same plane as in the map derived by electron crystallography. After the refinement procedure the distance Glu 139-Mg Chl *b*5 is 2.9 Å, i. e. comparable to the distance calculated in the EM model (2.2 Å). On the other hand, the molecular distance between Gln 131 and Mg Chl b6 is 5.7 Å, which is evidently too long for a direct co-ordination and in disagreement with the corresponding EM model value of 2.9 Å. Nevertheless, the bulky side chain of the Gln 131 appears as a strong density in the X-ray density map. The recent structure of PSI at 2.5 Å resolution revealed a cluster of chlorophylls in which the fifth ligand of the central Mg is attributed to a water molecules (Jordan *et al*., 2001). At the current resolution of 3.2 Å, it is not possible to see water molecules in the density map, however the distance between the residual side chain and the central Chl *b*6 Mg is ~6 Å, which is long enough to accommodate a water molecule allowing two hydrogen bonds at each sides.

For all the chlorophylls the Mg-Mg distance in the EM model are almost the same that have been found in the X-ray map. However a major difference exists in the orientation of Chl *b*2.

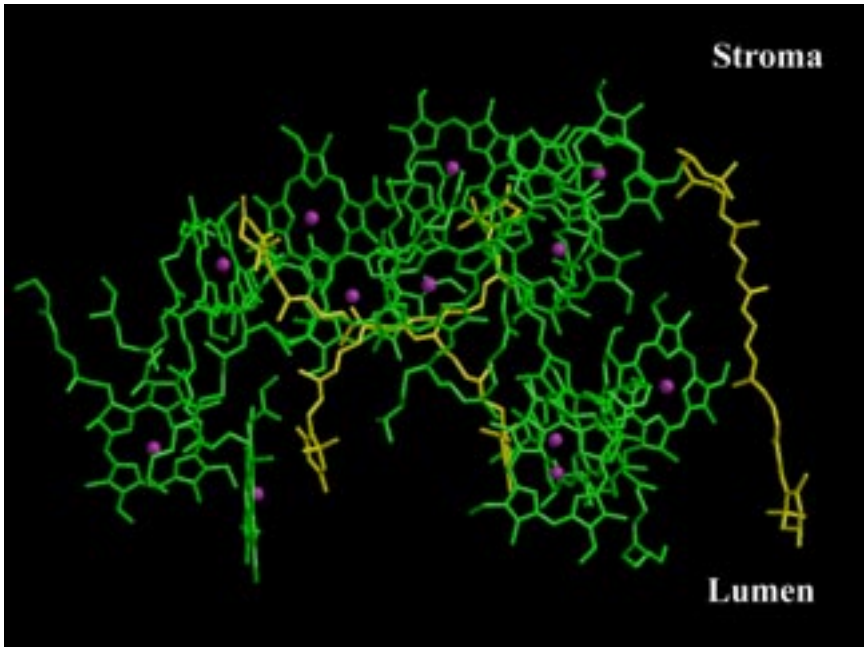


Fig.4.1a: Side view of the pigment molecules in LHC-II monomer.

In fact, this chlorophyll is in the same position as in the EM model, however rotated of about the c axis of almost 90º. In this new position the Mg-Mg distance of Chl *a*2 to Chl *b*2 is 9.3 Å, while in the EM model it was 8.3 Å. Due to the new position of Chl *b*2 the chlorophylls are almost perpendicular to each other like other pairs of chlorophyll, for example Chl *a*1-*b*1, Chl *a*3-*b*3 as well as Chl *a*5-*b*5 and Chl *a*6-*b*6.

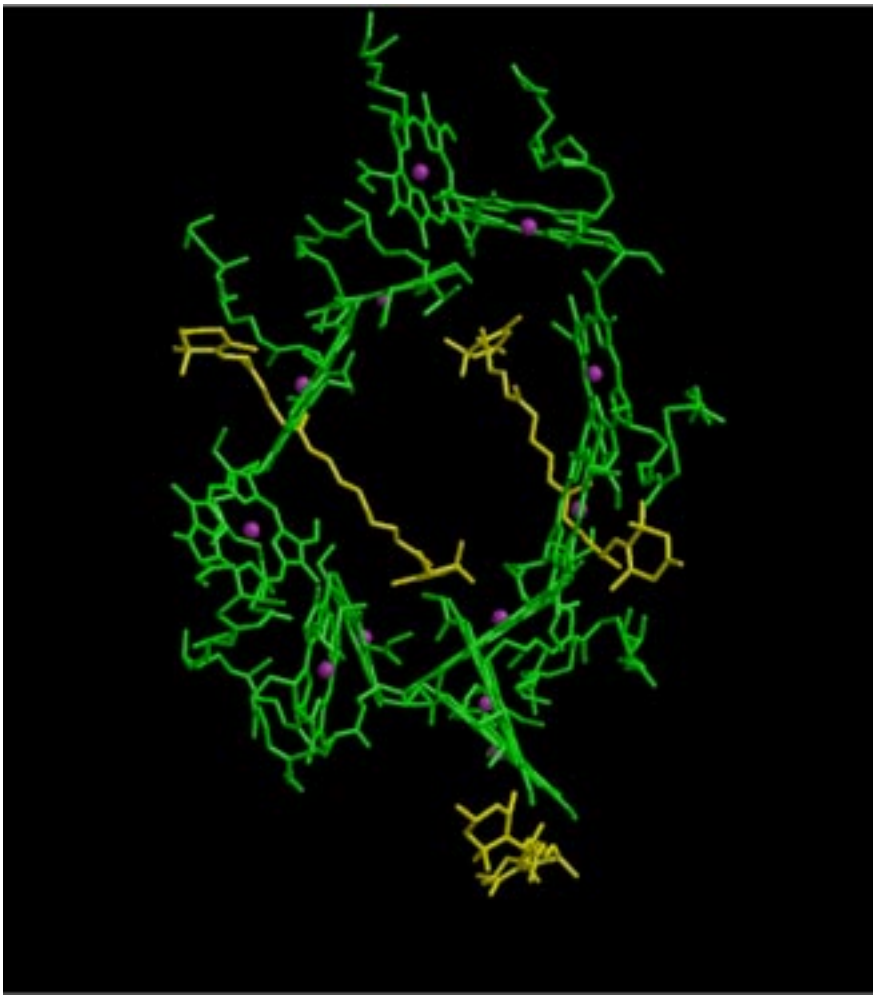


Fig. 4.1.b: Top view from the stroma side of the pigment molecules (colour code: green, chlorophyll; yellow, luteins; magenta, Mg atoms).

The assignment of seven Chls *a* and five Chls *b* was made considering the requirement photo-protection of the LHC-II (Kühlbrandt *et al.*, 1994). At current resolution of 3.2 Å, it is not possible to distinguish between Chl *a* and Chl *b*. However, the identity of most chlorophyll molecule assigned in the EM model were confirmed by single point mutations analysis. Rogl (2000) proved Chl *a*1, *a*2 and *a*3 to be Chls *a*, whilst Chl *b*5

and *b*6 are Chls *b*. On the contrary Chl *b*3 seems most likely to be a Chl *a*. The identity of Chl *a*4, *a*5 and *a*6 is not unambiguous determinated. In fact, preliminary data confirms the identity of *a*5 as Chl *a* but *a*4 could be a Chl *b*. The remaining Chls *a*7, *b*1 and *b*2 were not studied by this technique since the binding side chains are not identify yet. For this reasons the identity of Chls a and Chls b is not changed from the EM model.

4.6 Outlook

In order to fully understand the energy transfer within LHC-II, it is required to know the exact identity and position of the chlorophyll and of carotenoid molecules in the monomer. This will be achieved only when crystals diffracting to resolution higher than 2.0 Å will be obtained.

As has been extensively discussed, LHC-II crystals are affected by mosaicity, due to imperfect packing in the c direction. Obviously, another crystal form, with different molecular contacts, would be likely to lead to an improvement. Manipulation of already grown crystals may improve the diffraction quality, although the experience so far has not been encouraging. Recently, a method has been reported, which induces shrinkage of crystals both of soluble and membrane proteins, and causes transformation to higher order (Kiefersauer, *et al.*, 2000). It would be worth to try this technique with LHC-II crystals.

As more other membrane protein crystals, the LHC-II crystals have a high level of solvent content: 69.45 % calculated according to Matthews (1968). Kiefersauer and co-workers (2000) have demonstrated that crystal transformation in terms of unit cell can improve the diffraction resolution. For instance, crystals of CO dehydrogenase from an aerobic bacteria, *Oligotropha carboxidovorans*, showed a diffraction limit at 3 Å and after changing the humidity inside the capillary, where the crystal was sitting, the resolution achieved the limit of 1.8 Å. As already discussed the LHC-II crystals are sensitive to any environmental changes and the experienced gained till now shows a dissolving of the crystals when exposed to a new solution. However, harvesting the crystals with a loop is a gentle approach which might not affect the diffraction properties.

The loop would then be transferred into a capillary and connected to the equipment described by Kiefersauer *et al.* (2000). The dehydration occurring by evaporation of the solvent would probably induce the protein layers to stay closer and thereby favour the interactions between the hydrophilic part of the loops. Studies have shown that the best diffracting LHC-II crystal is the one with shortest c dimension of the unit cell (Table 3.1). If a control over the crystals transformation could be achieved it would surely bring an improvement in the diffraction quality.

However, it is most likely that improvements in the immediate future will come from X-ray analysis of three dimensional crystal grown of recombinant *Lhcb* product. Indeed, Rogl *et al.* (1999) optimised a system for the *in vitro* expression and refolding of *Lhcb* genes. Utilising this protocol, enough amount of LHC-II biologically active is prepared for crystallisation trials. Moreover, the successful expression and refolding of the protein encoded by *Lhcb*2 gene had been achieved and, using the crystallisation conditions developed during this work, small hexagonal plates are formed (Standfuß, personal communication). At present these crystals diffract X-ray radiation to the limit of 3.5 Å resolution, but they are still highly anisotropic.

Molecular biology technique can also aid in resolving the phase angles for the X-ray reflections. A different approach from MR to solve the phase problem is by multiple anomalous dispersion (MAD). Utilising a particular strain of *E. coli* which was made able to grow on medium with selenomethionin instead of methionin, it is possible to express a protein which systematically incorporate the selenomethionin instead of the methionin. In LHC-II monomer there are five methionin residues, which will be substituted by selenomethionin providing a sufficient to perform a successful MAD experiment. The selenium resembles the sulphur and the resulting selenomethionin proteins appear to be essential isostructural with natural methionin. It is likely that in the next future LHC-II crystals containing selenio-methionine will soon be available and will provide unbiased model phase, leading to a better refined model.

4.7 Conclusions

Due to the work outlined here, LHC-II X-ray model at 3.2 Å resolution is available. While no substantial differences with respect to the EM model calculated at 3.4 Å have been found a number of important new features have been discovered fig. 4.2a and b:

1) all the phytol chains have been attributed to eleven chlorophyll molecules out of twelve

2) a third carotenoid has been assigned

3) it has been discovered that the helix C has a short unwound region near its centre.

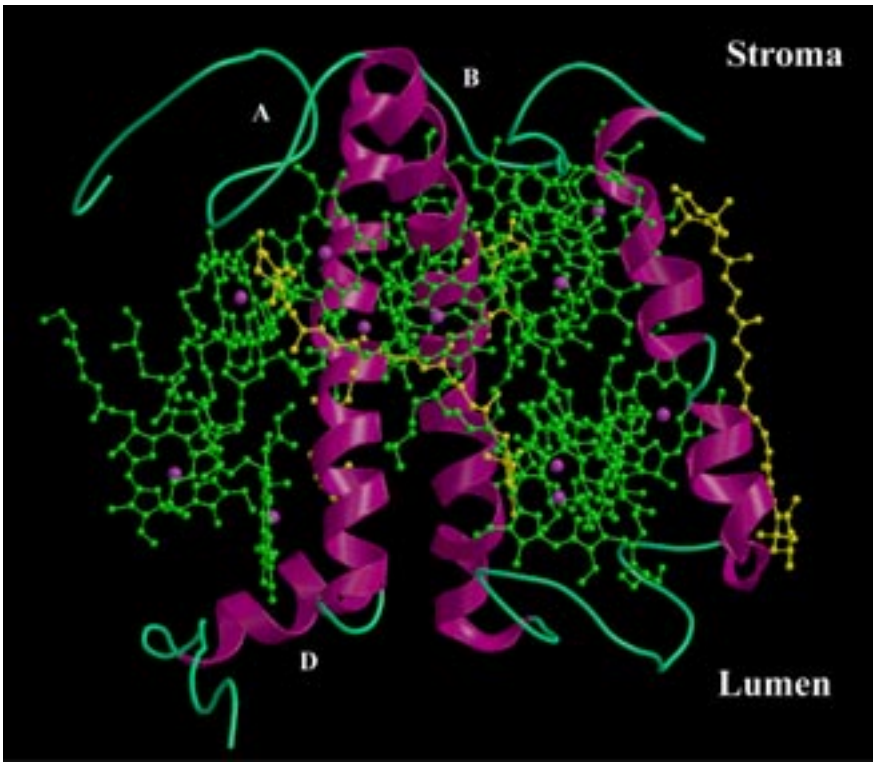


Fig 4.2a: Side view of the LHC-II monomer.

The X-ray structure is based on data collected to a slightly higher resolution and yet the overall structure of the LHC-II apo-protein has the same organisation. It has to be underlined that all these new structural features have to be confirmed, since both the R-factor and the R-free are at the moment too high to assume that this model is absolutely correct.

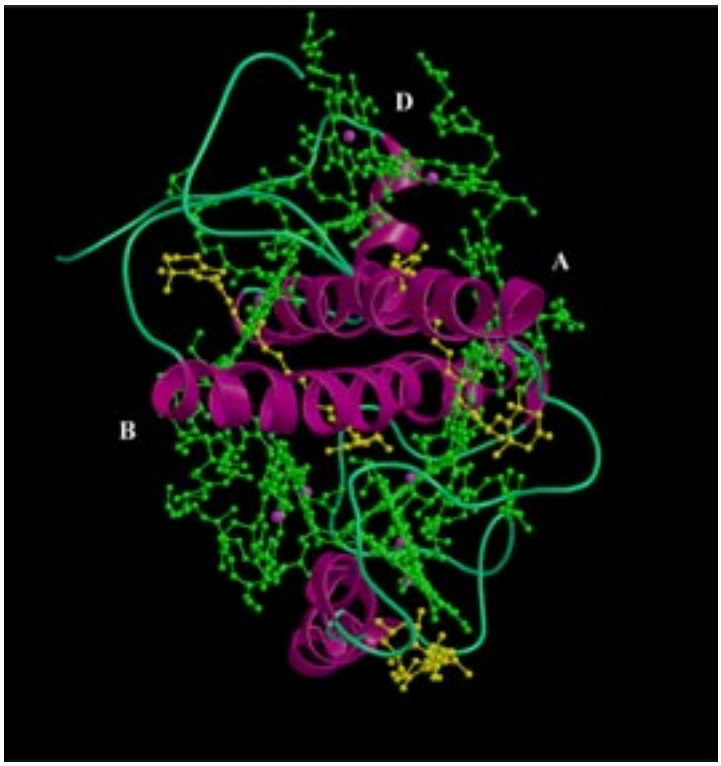


Fig 4.2b: Top view from the storma side of the LHC-II monomer (colour code: helices in mangenta pigments as in fig.4.1 a and b).

One interesting structural aspect is the unwound region of α-helix C, which also seems to be at least four amino acids longer in the X-ray model with respect to the EM model. The α-helix C seems to be substituted by two shorter α-helices. If these two new α-helices will be confirmed at improved refinement level, it will be necessary to attribute new names to the α-helices.

Improvement in resolution led to the final identification of the phytol chains for all the chlorophyll molecules except for Chl *a*1. Indeed, the phytol chains were assigned in full length for six chlorophylls, for the other chlorophylls the chains were only partially assigned. Nevertheless, it is not yet possible to establish the identity of the chlorophyll molecules.

Moreover, an additional carotenoid molecule, neoxanthin, was assigned. This molecule was identified as an integral component of the LHC-II complex by biochemical analysis and here it has finally been located near the α-helix C and is therefore found to be in van der Waals contacts with Chl *b*5 and Chl *b*6. In the structure, the position of the neoxanthin together with the assignment of the chlorophyll phytol chains and their

relative dipole moments will open the possibility of performing spectroscopic experiments elucidating the path of the energy transfer within the LHC-II monomer and neighbouring trimeric complex.

Appendix

Sequence alignment of the *Lhcb*1-3 gene from *Pisum sativum*

```
lhcb1   RKSATTKKVASSGSPWYGPDRVKYLGPFSGESPSYLTGEFPGDYGWDTAGLSADPETFSK 60
lhcb2   RRTVKS----APESIWYGPDRPKYLGPFSEQIPSYLTGEFPGDYGWDTAGLSADPETFAR 56
lhcb3   GTSKFT----MGNDLWYGPDRVKYLGPFSAQTPSYLTGEFPGDYGWDTAGLSADPEAFAK 56
          :  :       . ****** ******* : ************************:*::

lhcb1   NRELEVIHSRWAMLGALGCVFPELLSRNG-VKFGEAVWFKAGSQIFSEGGLDYLGNPSLV 119
lhcb2   NRELEVIHSRWAMLGALGCTFPELLEKNG-VKFGEAVWFKAGSQIFAEGGLDYLGNPNLI 115
lhcb3   NRALEVIHGRWAMLGALGCITPEVLQKWVRVDFKEPVWFKAGSQIFSEGGLDYLGNPNLV 116
        ** *****.**********  **:*.:   *.* *.**********:**********.*:

lhcb1   HAQSILAIWATQVILMGAVEGYRIAG-GPLGE----VVDP----LYPGG-SFDPLGLAD- 168
lhcb2   HAQSILAIWATQVVLMGFVEGYRVGG-GPLGEG----LDP----LYPGG-AFDPLGLAD- 164
lhcb3   HAQSILAVLGFQIVLMGLVEGFRINGLPDVGE----GND-----LYPGGQYFDPLGLAD- 166
        *******: . *::*** ***:*: *   :**      *     *****  ********

lhcb1   DPEAFAELKVKELKNGRLAMFSMFGFFVQAIVTGKGPLENLADHLSDPVNNNAWSYATNF 228
lhcb2   DPDSFAELKVKELKNGRLAMFSMFGFFVQAIVTGKGPIQNLYDHVADPVANNAWAFATNF 224
lhcb3   DPVTFAELKVKEIKNGRLAMFSMFGFFVQAIVTGKGPLENLLDHLDNPVANNAWVYATKF 226
        ** :********:************************::** **: :** **** :**:*

lhcb1   VPGK 232
lhcb2   VPGQ 228
lhcb3   VPGA 230
        ***
```

Bibliography

Allen, J. F. (1992). How does protein phosphorylation regulate photosynthesis? *Trends Biochem. Sci.* **17**, 12-17.

Arndt, U. W. & Wonacott, A. J. (1977). *The rotation method in crystallography*, Nort-Holland, Amsterdam.

Bacon, D. J. & Anderson, W. F. (1988). *Molec. Graphics* **6**, 219-220.

Barber, J. & Kühlbrandt, W. (1999). Photosystem II. *Current Opinion in Structural Biology* **9**, 469-475.

Barber, J., Nield, J., Morris, E. P., Zheleva, D. & Hankamer, B. (1997). The structure, function and dynamic of photosytem two. *Physiol. Plant* **100**, 817-827.

Bassi, R., Rigoni, F. & Giacometti, G. M. (1990). Chlorophyll binding proteins with antenna function in higher plants and green algae. *Photochem. Photobiol.* **52**, 1187-1206.

Bassi, R., Pineau, B., Dainese, P. & Marquardt, J. (1993). Carotenoid-binding protein of photosystem II. *Eur. J. Biochem.* **212**, 297-303.

Baymann, F., Brugna, M., Mühlenhoff, U. & Nitschke, W. (2001). Daddy, where did (PS)I come from? *Biochim. Biophys. Acta* **1507**, 291-310.

Belrhali, H., Nollert, P., Royant, A., Menzel, C., Rosenbusch, J. P., Landau, E. M. & Pebay-Peyroula, E. (1999). Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular view of the purple membrane at 1.9 Å resolution. *Structure* **7**, 909-917.

Bergfors, T. M. (1999). Protein Crystallization: Techniques, Strategies and Tips. IUL Biotechnology Series, Internationl Univeristy Line, La Jolla.

Bibby, T. S., Nield, J. & Barber, J. (2001). Iron deficiency induces the formation of an antenna ring around trimeric photosystem I in cyanobacteria. *Nature* **412**, 743-745.

Boekema, E. J., Hifney, A., Yakushevska, A. E., Piotrowski, M., Keegstra, W., Berry, S., Michel, K. P., Pistorius, E. K. & Kruip, J. (2001). A giant chlorophyll-protein complex induced by iron deficiency in cyanobacteria. *Nature* **412**, 745-748.

Boekema, E. J., van Roon, H., van Breemen, J. F. L. & Dekker, J. P. (1999). Supramolecular organization of photosytem II and its light-harvesting antenna in partially solubilized photosystem II membranes. *Eur. J. Biochem.* **266**, 444-452.

Bratt, C. E. & Åkerlund, H.-E. (1993). Isolation of pigment-free bulk lipids from thylakoids. *Biochim. Biophys. Acta* **1165**, 288-290

Brejc, K., Ficner, R., Huber, R. & Steinbacher, S. (1995). Isolation, crystallisation, crystal structure analysis and refinement of allophycocyanin from cyanobacterium *Spirulina platensis* at 2.3 Å resolution. *J. Mol. Biol.* **249**, 424-440.

Brünger, A. T., Adams, P. D. & Rice, L. (1997). New applications of simulated annealing in X-ray crystallography and solution NMR. *Structure* **5**, 325-336

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Cryst. D* **54**, 905-921.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R-factor refinement by molecular dynamics. *Science* **235**, 458-460.

Burke, J. J., Ditto, C. L. & Arntzen, C. J. (1978). Involvement of the light-harvesting complex in cation regulation of excitation energy distribution in chloroplasts. *Arch. Biochem. Biophys.* **187**, 252-263.

Butler, P. J. G., and Kühlbrandt, W. (1988). Determination of the aggregate size in detergent solution of the ligth-harvesting chlorophyll a/b-protein complex from chloroplast membranes. *Proc. Natl. Acad. Sci.* U.S.A. **85**, 3797-3801.

Clemons, W. M., May, J. L. C., Wimberly, B. T., McCutcheon, J. P., Capel, M. S. & Ramakrishnan, V. Structure of a bacterial 30S ribosomal subunit at 5.5 Å resolution. 1999 *Nature* **400**, 833-840.

Connelly, J. P., Müller, M. G., Bassi, R., Croce, R. & Holzwarth, A. R. (1997). Femtosecond Transient Absorption Study of Carotenoid to Chlorophyll Energy Transfer in the Light-Harvesting Complex II of Photosystem II. *Biochemistry* **36**(2), 281-287.

Croce, R., Remelli, R., Varotto, C., Breton, J. & Bassi, R. (1999). The neoxanthin binding site of the major light harvesting complex (LHCII) from higher plants. *FEBS Letters* **456**, 1-6.

Dainese, P. & Bassi, R. (1991). Subunit stoichiometry of the chloroplast photosystem II antenna system and aggregation state of the component chlorophyll a/b binding proteins. *J. Biol. Chem.* **266**, 8136-8142

Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. (1985). Structure of the protein subunits in the photosynthetic reaction centre of *Rhodopseudomonas virids* at 3 Å resolution. *Nature* **318**, 618-624.

Diederichs, K. & Karplus, P. A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nature Struct. Biol.* **4**, 269-275.

Dilly-Hartwig, H., Allen, J. F., Paulsen, H. & Race, H. L. (1998). Truncated recombinant light-harvesting complex II proteins are substrates for a protein kinase associated with photosytem II core complexes. *FEBS Lett.* **435**, 101-104.

Drenth, J., (1994). Principles of Protein X-ray Crystallography. Springer-Verlag, New York.

Duerring, M., Schmidt, G. B. & Huber, R. (1991). Isolation, crystallisation, crystal structure analysis and refinement of constitutive C-phycocyanin from the chromatically adapting cyanobacterium *Fremyella diplosiphon* at 1.66 Å resolution. *J. Mol. Biol.* **217**, 577-592.

Engel, C., Rik, W., Tucker, P. A. (1996). A removable arc for mounting and recovering flash-cooled crystals. *J. Appl. Cryst*. **29**, 208-210.

Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst. A* **47**, 392-400

Esnouf, R. M. (1997). *J. Mol. Graphics* **15**, 132-134.

Esnouf, R. M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D* **55**, 938-940.

Essen, L.-O., Siegert, R., Lehmann, W. D. & Oesterhelt, D. (1998). Lipid patches in membrane protein oligomers: Crystal structure of the bacteriorhodopsin-lipid complex. *PNAS* **95**, 11673-11678.

Fenna, R. E. & Matthews, B. W. (1975). Chlorophyll arrangement in a bacteriochlorophyll protein from *Chlorobium limicola*. *Nature* **258**, 573-577.

Fleming, G. R. & van Grondelle, R. (1997). Femtosecond spectroscopy of photosynthetic light-harvesting systems. *Curr. Opin. Struct. Biol.* **7**, 738-748.

Fling, S. P. & Gregerson, D. S. (1986). Peptide and protein molecular weight determination by electrophoresis using a high -molarity Tris-buffer system without urea. *Anal. Biochem.* **155**, 83-88

Frank, H. A., Young, A. J., Britton, G. & Richard J, C. (1999). *Photochemistry of Carotenoids*, Kluwer Academic Publisher.

Freer, A., Prince, S., Sauer, K., Miroslav, P., Hawthornthwaite-Lawless, A., McDermott, G., Cogdell, R. & Isaacs, N. W. (1996). Pigment-pigment interactions and energy transfer in the antenna complex of the photosynthetic bacterium *Rhodopseudomonas acidophila*. *Structure* **4**(449-462).

Fromme, P., Jordan, P. & Krauß, N. (2001). Structure of photosystem I. *Biophis. Biochem. Acta* **1507**, 5-31.

Funk, C., Schröder, W. P., Grenn, B. R., Renger, G. & Andersson, B. (1994). The intrinsic 22 kDa protein is a chlorophyll-binding subunit of photosystem II. *FEBS* **324**, 261-266.

Fujinaga, M. & Read, R. J. (1987). Experiences with a new translation-function program. *J. Appl. Cryst.* **20**, 517-521.

Garavito, R. M., Markovic-Housley, Z. & Jenkins, J. A. (1986). The growth and characterization of membrane protein crystals. *J. Cryst. Growth*, **76**, 701-709.

Garman, E. F. & Schneider, T. (1997). Macromolecular Cryocrystallography. *J. Appl. Cryst.* **30**, 211-237.

George H., S. & Lyle H., J. (1968). *X-ray Structure Determination*, The Macmillan Company, London.

Giuffra, E., Zucchelli, G., Sandonà, D., Croce, R., Cugini, D., Garlaschi, F. M., Bassi, R. & Jennings, R. C. (1997). Analysis of Some Optical Properties of a Native and Reconstituted Photosystem II Antenna Complex, CP29: Pigment Binding Sites Can Be Occupied by Chlorophyll a or Chlorophyll b and Determine Spectral Forms. *Biochemistry* **36**, 12984-12993.

Green, B. R. & Kühlbrandt, W. (1995). Sequence conservation of light-harvesting and stress-response proteins in relation to the three-dimensional molecular structure of LHCII. *Photosynth Res* **44**, 139-18.

Grigorieff, N., Ceska, T. A., Downing, K. H., Baldwin, J. M. & Henderson, R. (1996). Electron-crystallographic refinement of the structure of bacteriorhodopsin. *J. Mol. Biol.* **259**(3), 393-421.

Hauska, G., Schoedl, T., Remigy, H. & Tsiotis, G. (2001). The reaction center of green sulfur bacteria. *Biochim. Biophys. Acta* **1507**, 260-277.

Heathcote, P. (2001). Type I photosynthetic reaction centres. *Biochem. Biophys. Acta* **1507**, 1-2

Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E. & Downing, K. H. (1990). A model for the structure of bacteriorhodopsin based on high resolution electron cryomicroscopy. *J. Mol. Biol.* **213**, 899-929.

Hendrickson, W. A., Horton, J. R. & LeMaster, D. M (1990). Selenomethionyl proteins produced for analysis by multiwavelenght anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO Journal **9**, 1665-1672.

Hobe, S., Foerster, R., Klingler, J. & Paulsen, H. (1995). N-Proximal sequence motif in light-harvesting chlorophyll a/b binding protein is essential for the trimerization of light-harvesting chlorophyll a/b complex. *Biochemistry* **34**, 10224-10228.

Hofmann, E., Wrench, P. M., Sharples, F. P., Hiller, R. G., Welte, W. & Diederichs, K. (1996). Structural Basis of Light Harvesting by Carotenoids: Peridinin-Chlorophyll-Protein from *Amphidinium carterae*. *Science* **272**, 1788-1791.

Hunte, C., Koepke, J., Lange, C., Roßmanith, T. & Michel, H. (2000). Structure at 2.3 Å resolution of the cytochrome bc1 complex from the yeast Saccharomyces cerevisiae co-crystallized with an antibody Fv fragment. *Structure* **8**, 669-684.

Jackowski, G., Kacprzak, K. & Jansson, S. (2001). Identification of Lhcb1/Lhcb2/Lhcb3 heterotrimers of the main light-harvesting chlorophyll a/b protein complex of Photosystem II (LHC II). *Biochim. Biophys. Acta* **1504**(2-3), 340-345.

Jackowski, G. & Pielucha, K. (2001). Heterogeneity of the main light-harvesting chlorophyll a/b-protein complex of photosystem II (LHCII) at the level of trimeric subunits. *Photochem. Photobiol.* **64**, 45-54.

Jansson, S. (1994). The light-harvesting chlorophyll a/b-binding proteins. *Biochim. Biophys. Acta* **1184**, 1-19.

Jansson, S. (1999). A guide to the Lhc genes and their relatives in Arabidopsis. *Trends Plant Sci.* **4**(6), 236-240.

Jordan, P., Fromme, P., Witt, H.-T., Klukas, O., Saenger, W. & Krauß, N. (2001). Three-dimensional structure of cyanobacterial photosytem I at 2.5 Å resolution. *Nature* **411**, 909-917.

Jones, T. A. & Kjeldgaard, M. (1993). *O-The Manual*, Uppsala University, Uppsala, Sweden.

Jungas, C., Ranck, J.-L., Rigaud, J.-L., Joliot, P. & Vermeglio, A. (1999). Supramoleculra organization of the photosynthetic apparatus of *Rhodobacter sphaeroides*. *EMBO* **18**, 534-542.

Kabsch, W. (1988a). Automatic indexing of rotation diffraction patterns. *J. Appl. Cryst.* **21**, 67-71.

Kabsch, W. (1988b). Evaluation of single crystal X-ray diffraction data from a position sensitive detector. *J. Appl. Cryst.* **21**, 916-924.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795-800.

Karrasch, S., Bullough, P. A. & Ghosh, R. (1995). The 8.5 Å projection map of the light-harvesting complex I from *Rhodospirillum rubrum* reveals a ring composed of 16 subunits. *The EMBO Journal* **14**(4), 631-638.

Ke, B. (2001). Photosynthesis. Kluwer Academic Publishers, Dordrecht, The Netherlands

Kiefersauer, R., Than, M. E., Dobbek, H., Gremer, L., Melero, M., Strobl, S., Dias, J. M., Soulimane, T. & Huber, R. (2000). A novel free-mounting system for protein

crystals: transformation and improvement of diffraction power by accurately controlled humidity changes. *J. Appl. Cryst.* **33**, 1223-1230.

Kimura, Y., Vassylyev, D. G., Miyazawa, A., Kidera, A., Matsushima, M., Mitsuoka, K., Murata, K., Hirai, T. & Fujiyoshi, Y. (1997). Surface of bacteriorhodopsin revealed by high-resolution electron crystallography. *Nature* **389**, 206-211.

Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). Rapid automated molecular replacement by elvolutionary search. *Acta Cryst. D* **55**, 484-491.

Kleywegt, G. T. & Jones, T. A. (1997). Model-building and refinement practice. *Meth. Enzymol.* **227**, 208-230.

Koepke, J., Hu, X., Muenke, C., Schulten, K. & Michel, H. (1996). The crystal structure of the light-harvesting complex II (B800-850) from *Rhodospirillum molischianum*. *Structure* **4**(5), 581-597.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950

Kühlbrandt, W. (1987). Three-dimensional crystals of the light-harvesting chlorophyll a/b protein complex from pea chloroplasts. *J. Mol. Biol.* **194**, 757-762.

Kühlbrandt, W. (1988). Three-dimensional crystallization of membrane protein. *Q. Rev. Biophys.* **21**, 429-477.

Kühlbrandt, W. (1995). Many wheels make light work. *Nature* **374**, 497-498.

Kühlbrandt, W. (1995). Structure and function of bacterial light-harvesting complexes. *Structure* **3**, 521-525

Kühlbrandt, W., Thaler, T. & Wehrli, E. (1983). The structure of membrane crystals of the light-harvesting chlorophyll a/b protein complex. *J. Cell Biol.* **96**, 1414-1424.

Kühlbrandt, W. & Wang, D. N. (1991). Three dimensional structure of plant light-harvesting complex determined by electron crystallography. *Nature* **350**, 130-134.

Kühlbrandt, W., Wang, D. N. & Fujiyoshi, Y. (1994). Atomic model of plant light-harvesting complex by electron crystallography. *Nature* **367**, 614-621.

Kühlbrandt, W. (1994) Structure and function of the plant light-harvesting complex, LHC-II. *Curr. Op. Struct. Biol.* **4**, 519-528

Lancaster, C. R. D., Kröger, A., Auer, M. & Michel, H. (1999). Structure of fumarate reductase from *Wolinella succinogenes* at 2.2 Å resolution. *Nature* **402**, 377 - 385.

Landau, E. M. & Rosenbusch, J. P. (1996). Lipidic cubic phases: A novel concept for the crystallization of membrane proteins. *Proc. Natl. Acad. Sci U.S.A.* **93**, 14532.

Lee, A. G. (2000). Membrane lipids: it's only a phase. *Current Biology* **10**, 377-380.

Li, X.-P., Björkman, O., Shih, C., Grossman, A. R., Rosenquist, M., Jansson, S. & Niyogi, K. K. (2000). A pigment-binding protein essential for regulation of photosynthetic light harvesting. *Nature* **403**, 391 - 395.

Luecke, H., Schobert, B., Richter, H. T., Cartailler, J. P. & Lanyi, J. K. (1999). Structure of bacteriorhodopsin at 1.55 angstrom resolution. *J. Mol. Biol.* **291**(4), 899-911.

McDermott, G., Prince, S. M., Freer, A. A., Hawthornthwaite-Lawless, A. M., Papiz, M. Z., Cogdell, R. J. & Isaacs, N. W. (1995). Crystal structure of an integral membrane light-harvesting complex from photosynthetic bacteria. *Nature* **374**, 517-521.

Merritt, E. A. & Bacon, D. J. (1997). Raster3D: Photorealistic molecular graphics. *Methods Enzymol.* **277**, 505-524.

Michel, H. (1983). Crystallization of membrane proteins. *TIBS* **8**, 56-59.

Mitchell, P. (1976). Possible molecular mechanisms of the protonmotive function of cytochrome systems. *J. Theor. Bol.* **62**, 327-367.

Murphy, D. J. & Woodrow, I. E. (1983). *Biochim. Biophys. Acta* **725**, 104-112.

Navaza, J. (1994). AMoRe : An automated package for molecular replacement. *Acta Cryst. A* **50**, 157-163.

Nield, J., Orlova, E. V., Morris, E. P., Gowen, B., van Heel, M. & Barber, J. (2000). 3D map of the plant photosystem II supercomplex obtained by cryoelectron microscopy and single particle analysis. *Nature struct. biol.* **7**(1), 44-47.

Nußberger, S. (1994). Untersuchungen zur oligomeren Struktur des Lichtsammlerkomplexes der Pflanze, dissertation am Lehrstuhl für Bioiphysik der TU München.

Nußberger, S., Doerr, K., Wang, D. & Kühlbrandt, W. (1993). Lipid-protein interactions in crystals of plant light-harvesting complex. *J. Mol. Biol.* **234**, 347-356.

Ostermeier, C., Iwata, S., Ludwig, B. & Michel, H. (1995). Fv fragment-mediated crystallization of the membrane protein bacterial cytochrome c oxidase. *Nat. Struct. Biol.* **2**, 842-846.

Papiz, M. Z., Prince, S. M., Hawthornthwaite-Lawless, A. M., McDermott, G., Freer, A. A., Isaacs, N. W. & Cogdell, R. J. (1996). A model for the photosynthetic apparatus of purple bacteria. *Trends Plant Sci.* **1**, 198-206.

Paulsen, H. (1995). Chlorophyll a/b-binding proteins. *Photochem. Photobiol.* **62**, 367-382.

Pebay-Peyroula, E., Rummel, G., Rosenbusch, J. P. & Landau, E. M. (1997). X-ray structure of bacteriorhodopsin at 2.5 Angstroms from microcrystals grown in lipidic cubic phases. *Science* **277**, 1676-1681.

Peter, G. F. & Thornber, J. P. (1991). Biochemical composition and organization of higher plant photosystem II light-harvesting pigment-proteins. *J. Biol. Chem.* **266**, 16745-16754.

Peterman, E. J. G., Monshouwer, R., van Stokkum, I. H. M., van Grondelle, R. & van Amerongen, H. (1997). Ultrafast singlet excitation transfer from carotenoids to chlorophylls via different pathways in light-harvesting complex II of higher plants. *Chem. Phys. Lett.* **264**, 279-284.

Porra, R. J., Thompson, W. A. & Kriedmann, P. E. (1989). Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b with four different solvent: verifications of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochim. Biophys. Acta* **975**, 384-394.

Rhee, K.-H., Morris, E. P., Barber, J. & Kühlbrandt, W. (1998a). Three-dimensional structure of the plant photosystem II reaction centre at 8 Å resolution. *Nature* **396**, 283-286.

Rhee, K.-H., Morris, E. P., Zheleva, D., Hankamer, B., Kühlbrandt, W. & Barber, J. (1998b). Two-dimensional structure of plant photosystem II at 8 Å resolution. *Nature* **389**, 522-526.

Rodgers, D. W. (1994). Cryocrystallography. *Structure* **2**, 1135-1140.

Rogl, H. (2000). Struktur und Funktion des Lichtsammelkomplexes LHC-II der Höheren Planzen, Johann Wolfang Goethe-Universität.

Rogl, H. & Kühlbrandt, W. (1999). Mutant trimers of light-harvesting complex II exhibit altered pigment content and spectroscopic features. *Biochemistry* **38**, 16241-1622.

Rossmann, M. G. (1972). *The Molecular Replacement Method*, Gordon & Breach Science, New York, London, Paris.

Rossmann, M. G. & Blow, D. M. (1962). The detection of subunits within the crystallographic asymmetric unit. *Acta Cryst. A* **15**, 24-31.

Ruban, A. V., Lee, P. J., Mark Wentworth, M., Young, A. J. & Horton, P. (1999). Determination of the Stoichiometry and Strength of Binding of Xanthophylls to the Photosystem II Light Harvesting Complexes. *J. Biol. Chem.* **274**, 10458-10465.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular cloning: A Laboratory Manual* (ed, n., Ed.), Cold Spring Harbor, NY

Sandonà, D., Croce, R., Pagano, A., Crimi, M. & Bassi, R. (1998). Higher plants light harvesting proteins. Structure and function as revealed by mutation analysis of either protein or chromophore moieties. *Biochem. Biophy. Acta,* **1365**, 207-214.

Schertler, G. F. X., Bartunik, H., Michel, H. & Oesterhelt, D. (1993). Orthorhombic Crystal Form of Bacteriorhodpsion Nucleated on Benzamidine Diffracting to 3.6 Å Resolution. *J. Mol. Biol.* **234**, 156-164.

Schmid, V. H. R., Cammarata, K. V., Bruns, B. U. & Schmidt, G. W. (1999). In vitro reconstitution of the photosystem I light-harvesting complex LHCI-730: Heterodimerization is required for antenna pigment organization. *PNAS* **94**, 7667-7672.

Schulz, G. E. & Schirmer, R. H. (1979). Principles of Protein Stucture, Springer-Verlag, New-York.

Simidjiev, I., Stoylova, S., Amenitsch, H., Jávorfi, T., Mustárdy, L., Laggner, P. & Holzenburg, A. (2000). Self-assembly of large, ordered lamellae from non-bilayer lipids and integral membrane proteins *in vitro*. *PNAS* **97**, 1473-1476.

Staehelin, L. A., Golecki, J. R. & Drews, G. (1980). Supramolecular organization of chlorosomes (chlorobium vesicles) and their membrane attachment sites in *Chlorobium limicola*. *Biophys. Biochem. Acta* **589**, 30-45.

Steigemann, W. & Weber E. (1979). Strucutre of erythrocruorin in different ligand states refined at 1.4 Å resolution. J. Mol. Biol. **127**, 309-38

Stock, D., Leslie, A. G. W. & Walker, J. E. (1999). Molecular architecture of the rotary motor in ATP synthase. *Science* **286**, 1770-1705.

Veverka, V., Hrabal, R., Durchan, M. & Stys, D. Studies of phospholipid binding to N-terminal domain of membrane protein light-harvesting complex II. **523**, 281-287

Walters, R. G. & Horton, P. (1999). Structural and functional heterogeneity in the major light-harvesting complexes of higher plants. *Photosynth. Res.* **61**, 77-90.

Wang, B.-C. (1985). Resolution of Phase Ambiguity in Macromolecular Crystallography. *Methods Enzymol* **115**, 90-112

Yang, C., Kosemund, K., Cornet, C. & Paulsen, H. (1999). Exchange of Pigment-Binding Amino Acids in Light-Harvesting Chlorophyll a/b Protein. Biochemistry 38, 16205-16213

Yeh, J. I. & Hol, W. G. J. (1998). A flash-annealing technique to improve limits and lower mosaicity in crystals of glycerol kinase. *Acta Cryst.* **D54**, 479-480.

Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. (2001). Chemistry of ion coordination and hydration revealed by a K+ channel-Fab complex at 2.0 Å resolution. *Nature* **414**, 43 - 48.

Zouni, A., Witt, H.-T., Kern, J., Fromme, P., Krauß, N., Saenger, W. & Orth, P. (2001). Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 Å resolution. *Nature* **409**, 739-743.

Zusammenfassung in deutscher Sprache

Der Chlorophyll *a* und *b* bindende Lichtsammelkomplex II (LHC-II) ist der Hauptantennenkomplex höherer Pflanzen. Er dient der Lichtabsorption und Energieweiterleitung und bindet fast die Hälfte der Chlorophylle in grünen Pflanzen. In der Thylakoidmembran liegt LHC-II als Trimer vor. Jedes Monomer besteht aus 232 Aminosäuren und bindet spezifisch mindestens 12 Chlorophyllmoleküle (Chl) und drei Carotinoide (zwei Luteine und ein Neoxanthin).

Die Struktur von LHC-II wurde mittels Elektronenmikroskopie anhand von zweidimensionalen Kristallen mit einer Auflösung von 3.4 Å bestimmt. Diese Auflösung reicht jedoch nicht aus, um den Energieübertragungsmechanismus von LHC-II zu Photosystem II (PSII) zu verstehen, besonders da die tatsächliche Auflösung in der z-Dimension nur 4.9 Å beträgt. Der chemische Unterschied zwischen Chl *a* und Chl *b* besteht aus einer Formylgruppe anstelle einer Methylgruppe an der siebten Stelle des Chlorin-Rings und ist zu gering, um bei dieser Auflösung sichtbar zu werden. Zudem ist die Orientierung der Phorphyrinringe der Chlorophylle nicht eindeutig zu bestimmen. Diese Kenntnis ist jedoch wichtig für ein detailliertes Verständnis der Energieübertragung innerhalb des Komplexes und auf die Reaktionszentren von PSI und PSII.

Um die Auflösung der Struktur zu verbessern und die EM Daten zu vervollständigen wurde Röntgenkristallographie an dreidimensionalen Kristallen (3D) durchgeführt. LHC-II wurde nach einem Standardprotokoll (Burge et al. 1978) aufgereinigt. Die in Triton X-100 solubilisierten Thylakoidmembranen wurden mittels eines linearen Saccharosegradienten aufgetrennt. Die fluoreszierende LHC-II Bande wurde mit KCl präzipitiert, gewaschen und in Nonyl-glucosid aufgenommen. Die Kristalle wurden mittels Dampfdiffusion in hängenden Tropfen erzeugt. Die Kristalle waren dünne hexagonale Platten, welche eine verhältnismäßig große Einheitszelle aufwiesen und schwach beugten. Die Kristalle waren zwischen 0.3 bis 0.4 mm groß und 10 bis 20 μm dick. Der hohe Hintergrund beruhte sowohl auf der Verwendung von Detergenz zur Solubilisierung des Proteins, als auch auf in der Präparation vorhandenen Lipiden. LHC-

II bindet verschiedene Lipide: SQDG, MGDG, DGDG, PG und PC. Wie frühere Arbeiten gezeigt haben sind DGDG und PG für die 3D Kristallisation (Nußberger, 1994) notwendig, weshalb auf eine Entfernung der Lipide vor der Kristallisation verzichtet wurde.

Die Datenaufnahme erfolgte an gefrorenen Kristallen um Strahlungsschäden zu verringern. Dies ermöglichte einen kompletten Datensatz über einen Rotationsbereich von 135º von einem einzelnen Kristall aufzunehmen. Insgesamt wurden 3 Datensätze mittels eines hoch kollimierten und brillianten Strahls (ID-14 EH1 am ESRF, Grenoble, Frankreich) mit einer Auflösung bis 3.2 Å aufgenommen.

Die Datenanalyse erfolgte mittels des Programms XDS von Kabsch (1993). Dieses Programm verwendet einen R-faktor (R_{mrgd-F}) wie er in Diederichs & Karplus (1997) definiert wurde. Dieser ist ein Qualitätsidikator für die gemessenen Amplituden. Die R-Faktoren und damit die Qualität der drei gemessenen Datensätze sind vergleichbar mit den R-Faktoren die von Diederich und Karplus beschrieben wurden und die Daten konnten daher für die Strukturaufklärung verwendet werden. Die Kristalle gehören der Raumgruppe $P6_322$ an und die Einheitszelle hat die Dimensionen a=128.45, b=128.45, c=135.32, $\alpha=\beta$=90º, γ=120º.
Ein Aspekt für die Datenanalyse ist die Mosaizität der Kristalle. Diese ist ein Indikator für eine gute Kristallpackung. In LHC-II Kristallen ist die Mosaizität relativ hoch. Dies kann auf die Kristallpackung von Typ I (Michel 1983) zurückgeführt werden. Solche Kristalle sind aus Lagen von 2D Kristallen aufgebaut, welche häufig leicht gegeneinander verschoben sind und damit die Kristallsymmetrie stören. Auch das Einfrieren von Kristallen führt im allgemeinen zu einer erhöhten Mosaizität.

Die Lösung des Phasenproblems erfolgte durch molekulare Ersetzung mit Hilfe der elektronenmikroskopischen Struktur. Hierbei wurde ein LHC-II Monomer als Suchmodell verwendet. Drei verschiedene Programme wurden getestet wobei jedoch die für molekulare Ersetzung am häufigsten verwendeten Programme AMORE (Navaza *et al.*, 1994) und Brute (Fujinaga & Read, 1987) keine sinnvolle Lösung ergaben. Erst mit

dem Programm EPMR (Kissinger *et al*., 1999) konnte das Phasenproblem gelöst werden. Dieses verwendet einen sechsdimensionalen Suchalgorithmus um mögliche Lösungen zu erhalten. Jeder Einzelkandidat wird mit einem Korrelationskoeffizienten (CC) und R-Faktor evaluiert. Die besten Kandidaten werden dann wiederum für einen neuen Durchlauf verwendet. Das Programm führt 50 aufeinanderfolgende Stufen mit jeweils 300 Einzelkandidaten aus. Die Endlösung hatte einen CC von 36.1 % und einen R-Faktor von 54% und war damit ein guter Ausgangspunkt für die weitere Verfeinerung der Struktur.

Für die Verfeinerung wurde das Programm CNS (Brünger *et al*. (1997) verwendet. Datenreduktion führte zu 10727 unabhängigen Reflektionen. Die zu verfeinernden Parameter sind die drei räumlichen Koordinaten und die B-Faktoren der 1952 Atome des LHC-II Monomers. Die Relation zwischen Reflektionen und Parametern beträgt damit nur 1.3 und ist daher sehr gering. Trotzdem wurde die Verfeinerung der Daten fortgeführt um ein vorläufiges Modell zu erhalten.

Der erste Schritt der Verfeinerung war eine Anpassung des Modells des Gesamtkomplexes an die Elektronendichten als eine Einheit (rigid body refinement) bei dem sechs Parameter (drei Translationen und drei Rotationen) verfeinert wurden. Dies verbesserte den R-Faktor auf 50.52 %.
Der zweite Schritt war eine Verfeinerung bei welchem die Helices, die Chlorophylle und die Carotinoide im Unterschied zum ersten Schritt jeweils als einzelne Einheiten angepaßt wurden. Auch wenn bei diesem Schritt mehr Parameter als beim ersten Schritt verwendet wurden führte es nur zu einer geringfügigen Verbesserung des R-Faktors.
Als letzten Schritt der Verfeinerung wurde ein „simulated annealing“ ausgeführt. Diese Methode fitted das Modell durch Simulation der Bewegung der einzelnen Atome wie bei Erhitzung und Abkülung, um lokale Energieminima zu vermeiden. Dies verbesserte den R-Faktor auf 35% und den R-free auf 45%. Diese Werte sind zu hoch für eine unzweideutige Auslegung der Elektronendichtekarte, bei einer Auflösung von 3.2 Å kann jedoch kein besserer R-Faktor als 30% erwartet werden und eine vorläufige Anpassung

des Modells wurde daher vorgenommen. Dieses Modell zeigte einige interessante neue Aspekte des LHC-II.

In der elektronenkristallographischen Struktur konnten die ersten 25 Aminosäuren des N-Terminus des Apoproteins nicht aufgelöst werden. Dieser kann von einer spezifischen Kinase phosphoryliert werden und spielt daher eine wichtige Rolle bei der Regulation der Energiebalance zwischen PSII und PSI (Allen, 1992). In der hier vorgelegten Arbeit konnte das vorhandene Strukturmodell um eine weitere Animosäure (Lys 25) erweitert werden.

In der EM Struktur erscheint Helix C als durchgehende Helix die die Membran senkrecht durchmißt. Aufgrund der höheren vertikalen Auflösung zeigt die Röntgenstruktur innerhalb der Helix C jedoch eine nicht-helicale Region in der Mitte der Lipid-Doppelschicht, durch die die ursprüngliche Helix C unterbrochen ist. Auch zeigt das neue Modell eine Verlängerung der Helix um 1 Aminosäure auf der stromalen Seite und um 3 Aminosäuren auf der luminalen Seite.

Als weiteres Ergebnis dieser Doktorarbeit konnten die Phytylseitenketten von elf Chlorophyllen lokalisiert werden. Die Porphyrinringe der Chlorophylle sind in zwei getrennten Lagen parallel zur Membranebene angeordnet. Eine Lage, bestehend aus sieben Chlorophyllen, ist nahe der Stromaseite lokalisiert. Die andere, geformt von fünf Chlorophyllen, befindet sich an der luminalen Seite. Die Phytylseitenketten sind in dem hydrophoben Bereich zwischen diesen zwei Ebenen angeordnet. Zudem konnte ein Carotinoid, höchstwahrscheinlich ein Neoxanthin-Molekül, welches mittels Elektronenkristallographie nicht aufgelöst wurde, eingepaßt werden. Dieses Molekül erscheint in der Nähe von Helix C fast senkrecht zur Membranebene. Des weiteren hat es sich erwiesen, daß es sich in van der Waals Kontakt mit Chl *b*5 und Chl *b*6 befindet.
Die genaue Lokalisation der Carotinoids präzisiert die Erkenntnis die mit Hilfe von Untersuchungen an Punktmutanten gewonnen wurde. Diese Experimente wurden von

zwei verschiedenen Gruppen ausgeführt und beweisen, daß die Bindungsstellen von Chl *b*5 und Chl *b*6 mit der Neoxanthinbindungsstelle direkt in Verbindung stehen (Rogl, 2000; Croce *et al.*, 1999)

Die neuen strukturellen Details sind besonders für zukünftige Untersuchungen sehr interessant. Die Anordnung der Phytylseitenketten wird die Berechnung der Q_x und Q_y Dipolmomente der Chlorophylle ermöglichen. Diese Informationen zusammen mit der Position und Orientierung des Neoxanthins werden neue spektroskopische Untersuchungen und ein detailliertes neues LHC-II Modell ermöglichen, was wiederum zu einem tieferen Verständnis der Energieübertragung innerhalb des LHC-II Komplexes und zu den beiden Reaktionszentren von PSII und PSI beitragen wird.

Acknowledgements

I wish to express my gratitude to Prof. Dr. Werner Kühlbrandt for giving me the opportunity of working in his group, for sustaining this project with ideas and discussion. Above all for his enthusiastic optimism throughout these years and for allowing my scientific education;

I also would like to thank:

Prof. Dr. Bernd Ludwig for accepting me as a PhD. student at the Johann Wolfgang Goethe-Universität;

Johan P. Zeelen for aiding at the early stage of the crystallisation experiments and for all the days and nights spent at the synchrotron;

Dr. Eberhard Warkentin for helping me to tackle the "phase-problem" and the refinement process, as well as for discussing results and problems; and Dr. Ulrich Ermler for proposing useful suggestions;

Dr. Wolfgang Kabsch for his clear introduction to XDS at the beginning of data analysis and Dr. Dean Madden for counseling during the latest time of the refinement process;

Dr. Hans Rogl and Jörg Standfuss for sharing with me their interest and curiosity on the LHC-II and for discussing our results; Dr. Ian Collinson for commenting upon any scientific problems and for bringing a friendly atmosphere in the lab;

Dr. Claudia Büchel for kindly reading and commenting upon the manuscript;

All the people of the Max-Planck-Institute for creating a pleasant environment and especially: Karen, Manfred, Jens, Mihna, Michael, David, Ulrike, Luise, Janet, Deryck, Tassilo and Lena; and Franziska for telling me always her point of view about many of the non-scientific problems.

Teresa for her friendly mood and continuous advice and support. My friend Hildur and the “Icelandic” attitude towards life; my generous friend Paolo; Gianni for revealing an easy-going approach to life; and particularly Luana for bearing and sharing my bad and good times and cheering up with her incredibly vitality and spirit;

Especially, Cécile for her continuous encouraging support, for discussing about the problems encountered during different phases of the thesis, for manifesting a frank and sincere friendship and yet being always at my side.

Lebenslauf

Name	Matteo Lamborghini
Geburtsdatum	13.3.1971
Geburtsort	Ferrara, Italien

Schulbildung

1985-1990	Liceo Scientifico "A. Roiti", Ferrara, Italien

Studium

1990-1997	Laurea in Scienze Agrarie an der "Universita degli Studi di Bologna", Bologna, Italien
1994-1995	Diploma in Microbiology an der "Kent University at Canterbury", Canterbury, England

Dissertation

Seit 1997	Doktorarbeit zum Thema "Three dimensional structure of the light-harvesting chlorophyll *a/b* protein complex from plant chloroplasts" in der Abteilung von Prof. Dr. Werner Kühlbrandt am Max-Planck-Institut für Biophsik in Frankfurt am Main.