THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE NATURAL SCIENCES

Structural Insights at Sub-Ångstrom, Medium and Low Resolution:

Crystallization of Trypsin, Bacterioferritin, Photosynthetic Reaction Center, and Photosynthetic Core Complex

WEIXIAO YUAN WAHLGREN

Department of Chemistry and Molecular Biology – Biochemistry
Göteborg, Sweden
2012

Thesis for the Degree of Doctor of Philosophy in the Natural Science

Structural Insights at Sub-Ångstrom, Medium and Low Resolution:

Crystallization of Trypsin, Bacterioferritin, Photosynthetic Reaction Center, and Photosynthetic Core Complex

Weixiao Yuan Wahlgren

Cover: Overview of 4-fold pore of bacterioferritin from *Blastochloris viridis*

ISBN 978-91-628-8539-7
Available online at http://hdl.hanle.net/2077/30181

Department of Chemistry and Molecular Biology
Biochemistry and Biophysics
University of Gothenburg
SE-413 90 Göteborg, Sweden

Printed by Kompendiet
Göteborg, Sweden 2012

Till Erik och Johan

Abstract

The catalytic action of serine proteases depends on the interplay of a nucleophile, a general base and a general acid. The catalytic triad is composed of serine, histidine and aspartate residues. The serine acts as a nucleophile while the histidine plays a dual role as the general base or acid at different steps of the reaction. However, the role of aspartate is unclear. I recovered an ultrahigh resolution (0.93 Å) X-ray structure of a complex formed between trypsin and a canonical inhibitor. At sub-ångstrom resolution, hydrogen atoms could be visualized, giving a clue to the protonation state of the catalytic residues. By comparing this with the theoretical electron density calculated by density theory functional, the protonation states of the catalytic histidine and aspartate are discussed. Hence, a refined mechanism for serine protease action is proposed in this thesis.

Photosystem harvests energy from sunlight with near 100% quantum yield. To study light-induced structural changes of the photosynthetic reaction center from purple non-sulfur bacterium *Blastochloris viridis* using X-ray crystallography, robust protein crystals with tight crystal packing are prerequisite. In this thesis, lipidic-sponge phase crystallization method was used and yielded well diffracting crystals for structure determination. Crystals showed a type I packing and a 1.86 Å resolution structure was determined with four lipid molecules captured in the structure. Moreover, I demonstrated that an occupied Q_B binding site can be obtained by co-crystallizing with UQ_2 using the sponge phase crystallization method. However, attempting to crystallize the reaction center-light harvesting 1 core complex, a 440 kDa membrane protein complex of total 54 putative subunits, it required different crystallization methods. Here, the resolution has been optimized to beyond 8 Å by using the lipidic bicelle crystallization method.

Conflict between the free but potential toxic Fe(II) and the insolubility of Fe(III) led to the evolution of bacterioferritin in bacteria, which functions as an iron storage and detoxification protein. Bacterioferritin from *Blastochloris viridis* (*Bv* Bfr) was crystallized and the structure was solved to 1.58 Å resolution. With the combination of X-ray structure, redundancy PCR and tandem mass spectrometry, the previously unknown amino acid sequence of *Bv* Bfr was determined. Conformational states of the ferroxidase center which undergoes reorganization upon different soakings were trapped. One water-like small ligand coordinated to the Fe_1 binding site was captured in the Fe(II)-soaked structure. By density functional theory calculations the character of this small legend was rationalized. In addition, the structure and mechanism of iron import of the protein was studied and discussed. Finally, the redox-state of the heme in the crystals with and without Fe(II)-soaking treatment was studied by single crystal UV-VIS microspectrophotometry, before and after the X-ray exposure.

Contribution report

Paper I. I was involved in data analysis, preparing figures and writing of the manuscript.

Paper II. I took a major part in the protein production, purification, crystallization and X-ray data collection. I took part in the crystallographic analysis and preparation of figures and manuscript.

Paper III. I took a major part in the work involving reaction center from *Blastochloris viridis* and the corresponding lipidic-sponge phase conditions.

Paper IV. I was involved in the entire project. I was responsible in sequencing, production, purification and crystallization of the protein. I planned soaking experiment, performed X-ray data collection, single crystal UV-VIS microspectroscopy, structural analysis and density functional theory calculation. I took a major part in interpretation of the results, preparing figures and writing of the manuscript.

List of original publications

This thesis is based upon the following papers:

I**. Wahlgren WY**, Pal G, Kardos J, Porrogi P, Szenthe B, Patthy A, Graf L, Katona G. (2011) The Catalytic Aspartate Is Protonated in the Michaelis Complex Formed between Trypsin and an in Vitro Evolved Substrate-like Inhibitor - A REFINED MECHANISM OF SERINE PROTEASE ACTION. *J. Biol. Chem*. **286** (5), 3587-3596.

II. Wöhri AB, **Wahlgren WY**, Malmerberg E, Johansson LC, Neutze R, Katona G. (2009) Lipidic Sponge Phase Crystal Structure of a Photosynthetic Reaction Center Reveals Lipids on the Protein Surface. *Biochemistry* **48** (41), 9831-9838.

III. Wöhri AB, Johansson LC, Wadsten-Hindrichsen P, **Wahlgren WY**, Fischer G, Horsefield R, Katona G, Nyblom M, Oberg F, Young G, Cogdell RJ, Fraser NJ, Engström S, Neutze R. (2008) A lipidic-sponge phase screen for membrane protein crystallization. *Structure* **16** (7), 1003-1009.

IV. **Wahlgren WY**, Omran H, von Stetten D, Royand A, van der Post S, Katona G. Structural characterization of bacterioferritin from *Blastochloris viridis*. Accepted for publication at *PLoS ONE.*

Related publications:

Johansson LC, Arnlund D, White TA, Katona G, et al. (2012) Lipidic phase membrane protein serial femtosecond crystallography. *Nature Methods* **9**, 263–265

Lundholm I, **Wahlgren WY,** Piccirilli F, Lupi S, Perucchi A, Katona G. Terahertz Absorption Change in a Photosynthetic Reaction Center upon Photoactivation*. Submitted to Biophysics Journal*

Other publication:

Rapali P, Radnai L, Suveges D, Harmat V, Tolgyesi F, **Wahlgren WY**, Katona G, Nyitray L, Pal G. (2011) Directed Evolution Reveals the Binding Motif Preference of the LC8/DYNLL Hub Protein and Predicts Large Numbers of Novel Binders in the Human Proteome. *PLoS ONE*. **6** (4)

Abbreviations

ADP	adenosine-5'-diphosphate
AFM	atomic force microscopy
ATP	adenosine-5'-triphosphate
BChl	bacteriochlorophyll a
BPhe	bacteriopheophytin a
Bv Bfr	bacterioferritin from *Blastochloris viridis*
Bfr	bacterioferritin
CHAPS	3-([3-cholamidopropyl]-dimethylammonio)-1-propanesulfonate
CHAPSO	3-([3-cholamidopropyl]-dimethylammonio)-2-hydroxy-1-propanesulfonate
DDM	n-dodecyl-b-D-maltopyranoside
DFT	density functional theory
DMPC	1,2-dimyristoyl-sn-glycero-3-phosphocholine
DMPG	1,2-dimyristoyl-sn-glycero-3-phospho-rac-glycerol
DMSO	dimethyl sulfoxide
FI	fluid isotropic
kDa	kilo dalton
LCP, Q	lipidic cubic phase
LDAO	lauryldimethylamine-N-oxide
LH1	light harvesting 1 complex
LH2	light harvesting 2 complex
LSP, L_3	lipidic-sponge phase
L_α	lamellar phase
MAG	monoacylglycerol
MO	monoolein
MPD	2-methyl-2,4-pentanediol
MQ_A, Q_A	menaquinone
MW	molecular weight
NMR	nuclear magnetic resonance
P	special pair
PDB	protein data bank
PEG	polyethylene glycol
Pi	inorganic phosphor
PPO	pentaerythritol propoxylate
P_{960}	special pair *Blastochloris viridis*
Q_B, UQ_B	ubiquinone
UQ_BH_2	ubiquinol
UQ_2	ubiquinone-2
UQ_9	ubiquinone-9
RC-$LH1_{vir}$	photosynthetic reaction center and light harvesting 1 core complex from *Blastochloris viridis*
RC	reaction center
RC_{vir}	reaction center from *Blastochloris viridis*
TM	transmembrane
Å	angstrom (10^{-10} m)

Table of Contents

1 Introduction

Have you ever wondered about the origin of life on Earth? Although we are far from certain about how life arose, generally it is accepted that the development of life started with a chemical evolution for nearly four billion years ago [1]. While it was a reducing atmosphere on Earth at that time, simple geologically occurring molecules reacted with each other and formed organic polymer complexes [2, 3]. It was followed by a stage in which these polymer-collections self-organized and formed replicating units. At some point in this process, the transition from a lifeless collection to a living system occurred. Eventually, through biological evolution the complex network of modern life was formed.

In a reducing atmosphere of the prebiotic Earth, reactions among the simple molecules are thought to have formed the organic precursors from which biological molecules such as polypeptides and polynucleotides developed [2, 3]. Minerals like clays may have played an important role as catalysts and catalyzed these chemical reactions.

1.1 The nature of enzymatic catalysis

A catalyst works by providing an alternative reaction pathway between substrate and product. Thus, the rate of the reaction is increased since this alternative route has a lower activation-energy, compared to the reaction path without the catalyst. The enormous variety of biochemical reactions that comprise life are nearly all mediated by a series of biological catalysts named enzymes. These biological catalysts are one of the remarkable outcomes of biological evolution and they differ from ordinary chemical catalysts in several important aspects. Compared with most efficient analogous chemical catalysises, enzymes can mediate reactions under milder reaction conditions, but with at least several orders of magnitude greater reaction rates. Furthermore, they have greater reaction specificity for both their substrates and products, compared to chemical catalysts and rarely have side products. Finally, they have a high capacity for regulation in response to external signals other than substrates and products [4]. Considering all these remarkable catalytic properties which enzymes have, one of the central questions of biochemistry rises: how do enzymes work?

Compared to an un-catalyzed reaction, a catalyst stabilizes the transition state of the reaction, thus lowering the height of the kinetic barrier and increasing the reaction rate. Enzymes are unique in that they are able to combine their specificity of substrate binding and their optimal arrangement of catalytic groups. To date, six types of catalytic mechanisms that enzymes employ have been classified: acid-base catalysis, covalent catalysis, metal ion catalysis, electrostatic catalysis, proximity and orientation effects and finally preferential binding of the transition state complex [4]. Serine Proteases are one of the best characterized group of enzymes with detailed examination of their catalytic mechanisms.

1.2 The origin of photosynthesis

The process of photosynthesis is possibly the most important chemical reaction on Earth which has led to the development of advanced life forms. The most accepted hypothesis for the origin of photosynthesis on Earth is that photosynthesis developed in Bacteria after divergence from Archaea and Eukarya [5]. Photosynthesis derives from five photosynthetic bacterial lineages. Among the four anoxygenic photosynthetic bacterial lineages, purple bacteria (nonoxygen-evolving type II photosystem), green sulphur bacteria (type I photosystem), green non-sulphur bacteria (nonoxygen-evolving type II photosystem) and heliobacteria (similar homodimeric type I reaction center), the purple bacteria were most likely the earliest emerging photosynthetic lineage [5, 6]. The fifth and only oxygenic photosynthetic bacterial lineage is cyanobacteria (both type I photosystem and oxygen-evolving type II photosystem) and which were proved to be late-evolving [6]. It has been widely accepted that the photosynthetic properties of eukaryotes in form of chloroplasts developed from cyanobacteria through endosymbiosis [7]. However, coinciding with Earth being dominated by cyanobacteria, oxygen level on Earth started to rise.

1.3 Iron storage and detoxification

Iron is one of the most abundant metals on Earth. It has two stable oxidation states, Fe (II) and Fe (III). Depending on the environment, the two states can be readily switchable which makes iron an extremely useful redox-mediator in biology [8]. Irons are involved in a variety of critical processes, such as respiration, photosynthesis, nitrogen fixation and DNA synthesis. Iron can be incorporated in protein molecules in the form of a heme, bound to sulfur in various types of iron-sulfur clusters, as mononuclear iron centers or as di-iron center [9].

In an oxidizing atmosphere, oxidation of the free ferrous ion to the solid ferric form initiates two problems. Fe (II) becomes rapidly oxidized by oxygen to Fe (III) and produces reactive oxygen species through the Fenton reaction, subsequently triggers the oxidative damage processes. On the other hand, Fe (III) has a low solubility of around 10^{-18} M in concentration under physiological conditions, thereby preventing it as an available iron source. Therefore, an efficient iron storage and release mechanism is required in a living organism, and ferritins fulfill this role.

$$Fe^{2+} + O_2 \rightarrow Fe^{3+} + O_2^{\cdot -}$$

$$2O_2^{\cdot -} + 2H^+ \rightarrow H_2O_2 + O_2 \qquad \text{(Fenton reaction)}$$

$$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + OH^- + \cdot OH$$

Ferritins constitute a broad superfamily of iron-storage proteins, widespread in all kingdoms of life. By controlling a reversible transition between the free ferrous ion in solution and the mineral ferric core inside its cavity, ferritins supply living cells with an effective iron concentration in the range of 10^{-3}- 10^{-5} M [10]. Additionally, by isolating the excess irons inside the cavity, away from oxidizing molecules, ferritins minimize the production of oxidative harmful species in the cell. Therefore, ferritins are presumed to have an iron detoxification function and they can be involved in cell redox-stress resistance [9]. Ferritins isolated from bacteria may also contain a heme, and are then called bacterioferritins (Bfr).

1.4 Scope of the Thesis

The aim of this thesis has been to combine different biochemical methods together with protein X-ray crystallography in order to obtain detailed structural and biochemical information of proteins. Furthermore, lipidic crystallization methods have been developed and evaluated in an attempt to obtain diffraction-quality crystal of membrane protein complexes.

Although serine proteases are one of the best characterized enzymes with respect to detailed description of the catalytic mechanisms, questions like protonation states of catalytic residues at different steps of the reaction still remained unsolved. An ultrahigh resolution (0.93 Å) X-ray structure of a co-crystallized complex formed between the serine protease, trypsin and one canonical inhibitor is discussed in this thesis. Electron density distribution of covalent bonds in the key catalytic residues were analyzed and compared with theoretic electron density produced from density theory functional calculations. Combined with partially visible densities corresponding to putative hydrogen atoms, the protonation states and function of these catalytic residues were discussed and a refined mechanism for serine protease action was proposed.

Photosynthesis, which converts sunlight into chemical energy, is one of the most important biological processes in nature. To study light-induced structural changes of the photosynthetic reaction center from purple non-sulfur bacterium *Blastochloris viridis* using X-ray crystallography, robust protein crystals with tight crystal packing is a prerequisite. Protein was purified from its native source with a modified purification protocol and lipidic-sponge phase crystallization method was applied. The reaction center was used as a model protein to develop a lipidic-sponge phase screen and optimize a modified sponge phase form.

It is still a challenge to study large membrane protein complexes using X-ray crystallography. Often, the bottleneck is to produce well-diffracting crystals. Methods including traditional vapour diffusion, lipidic-cubic phase, lipidic-sponge phase and bicelle crystallization have all been used attempting to crystallize reaction center - light harvesting 1 core complex (RC-$LH1_{vir}$). By using bicelle crystallization method the resolution of this 440 kDa membrane protein complex crystal structure has been improved.

Bacterioferritins provide an accessible storage of iron in a mineralized ferric form. Simultaneously, they reduce the concentration of toxic free Fe(II). Bacterioferritin from *Blastochloris viridis* (*Bv* BFR) was accidently crystallized while we worked on RC-$LH1_{vir}$ complex. With the combination of X-ray structure, redundancy PCR and mass spectrometry the previously unknown amino acid sequence of *Bv* Bfr was determined. Multiple soaking experiments were performed aiming to study the mechanism of iron transport of the protein. Conformational changes around the ferroxidase center upon different soaking treatments were captured by X-ray crystallographic structures. Methods like density functional theory calculations and single crystal UV-VIS microspectrophotometry were applied assisting the interpretation of the structural details.

With a wide range of methodologies being used, structural and biochemical details of three different classes of proteins have been investigated. Additional, different lipidic crystallization methods have been discussed while focusing on the crystallization of a large membrane protein complex.

2 Methodologies

2.1 Degenerate Polymerase chain reaction

Polymerase chain reaction (PCR) is a biochemical technology that amplifies NA to obtain numerous copies of a particular DNA sequence. The method consists of cycles of repeated DNA melting, primer annealing and enzymatic replication of the required DNA sequence. Degenerate PCR is fundamentally identical to ordinary PCR, except for one major difference: instead of using specific PCR primers with a given sequence, degenerate primers are used whereby one or more of its positions have several alternative nucleotides. This method has proven to be a tremendously powerful tool to identify new members of gene families [11].

Most genes within a family encode proteins sharing structural similarities. Bacterioferritin from *Blastochloris viridis* was accidently crystallized while attempting to crystallize reaction center-light harvesting 1 core complex of this organism (paper IV). In the absence of the genomic information for *Blastochloris viridis,* the initial guess of the amino acid sequence was derived directly from the electron density map. In addition, by aligning the amino acid sequences of proteins from related organisms, conserved and variable regions can be determined respectively. Once the sequences of the conserved regions are identified, the degeneracy of these amino acids can be determined. This can be done by taking the product of the degeneracy of each amino acid in the sequence. For example, valine has four codons (GTT GTC GTA GTG) and thus has a degeneracy of 4 while tryptophan has only one codon (TGG) and a degeneracy of 1. Methionine and tryptophan are the only amino acids that are coded by one unique codon, while serine, arginine and leucine are coded by six codons and therefore should be avoided if possible [12-14]. For PCR, two conserved regions for locating the forward and reverse primers are needed. Once the conserved amino acids regions with low degeneracy are chosen, they can be back-translated to the corresponding nucleotide sequences which are used as starting points for designing degenerate PCR primers. The so called "wobbles" are inserted in the PCR primers where there is more than one possibility of nucleotides and one pair of degenerate PCR primers is constructed. Degenerate PCR primers are normally 5 to 7 amino acids long and fairly close together (200 bp – 600 bp). PCR efficiency will drop if these regions are too far apart or the primers are too long. “Tails” can be added to the degenerate primers on the 5' ends and help to increase the PCR efficiency of these primers by increasing primer length and hence annealing temperature. Tails including restriction sites can be used for directional cloning. Alternatively, those ended with terminal G’s will encourage Taq polymerase to add overhanging A's for use in TA subcloning.

2.2 Protein crystallization

2.2.1 Protein crystal growth

The goal of crystallization is to produce well-ordered crystals which diffract when hitting with an X-ray beam. It is often the rate-limiting step in protein crystallography, particularly for membrane proteins. The growth conditions for a protein crystal are unpredictable and a strategy of systematic approach is essential.

The basic principle of protein crystallization is to alter the composition of the protein solution in a controlled way so that protein precipitates and forms diffracting-quality crystals. In order to get crystals the protein solution has to reach a supersaturated state, as described in the phase diagram (Figure 1). The supersaturated region can be divided into three zones, metastable, nucleation and precipitation zone. If the precipitation process occurs too fast, the protein solution jumps to the precipitation zone and amorphous precipitate which does not contain any internal order within the protein molecules is usually observed. To obtain crystals, nuclei have to be formed in the nucleation zone where small nuclei are formed containing a small amount of well-ordered protein molecules. Afterwards crystal growth spreads outwards from the nucleating site and protein solution stays in the metastable zone. The successful production of diffraction quality crystals depends on all parameters influencing crystallization, such as type and concentration of precipitants, protein concentration, pH, buffer, temperature, detergent, protein purity and more.

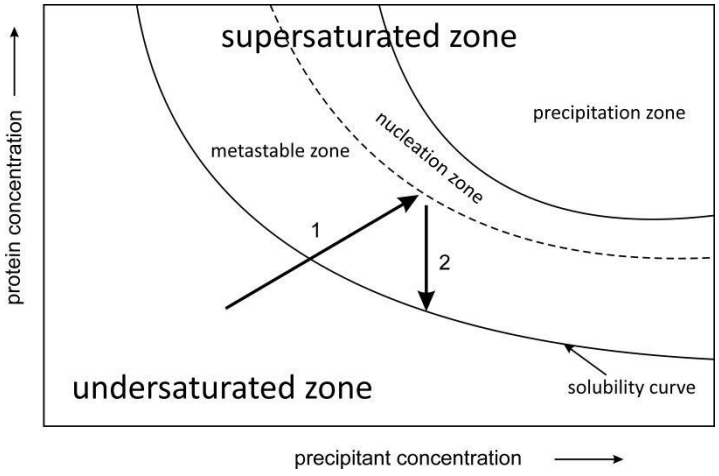

Figure 1. Phase diagram [15]. The concentrations of purified protein and precipitant solution in the droplet increase and reach the nucleation zone (1). Nucleation occurs and crystal starts to grow. The concentration of the protein in the solution decreases (2).

The most commonly used method for protein crystallization is vapour diffusion. There are three different experimental arrangements known as hanging drop, sitting drop and sandwich drop methods (Figure 2). The reservoir solution in hanging drop contains precipitant solution. A droplet containing purified protein and precipitant solution from the reservoir is placed onto a glass cover slide which is then sealed so that the drop is hanging above the reservoir solution. Initially, the droplet of protein solution contains lower concentration of precipitant than in the reservoir. Since the system strives towards equilibrium, water diffuses from the drop to the reservoir and both the protein and precipitant concentrations gradually increase to an optimal level for crystallization in the supersaturated state, where nucleation starts (Figure 1). The same principle is applied for sitting and sandwich drop.

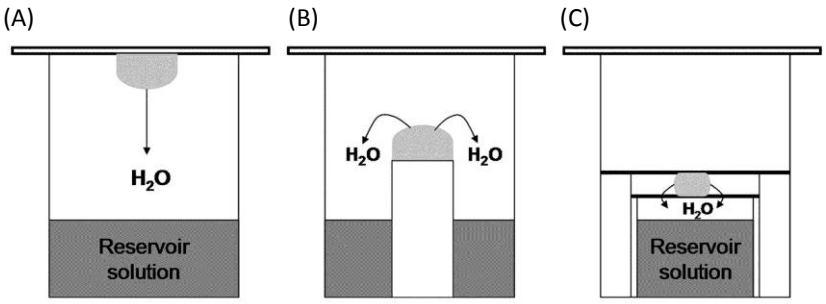

Figure 2. The vapour diffusion experiment. In a hanging-drop setup (A), the protein/ precipitant drop is turned upside-down. In a sitting-drop setup (B), a pedestal is used to separate the protein/ precipitant drop from the reservoir solution. In a sandwich-drop setup (C), the protein/ precipitant drop is located between two slices with spacer [16].

The key of forming a crystal is the crystal contacts where the protein molecules come in close contact and adhere to each other at specific points. This is particularly difficult for membrane protein.

2.2.2 Membrane protein crystallization

There are several challenges when working with membrane proteins, such as overexpression in high yields, purification with intact function and crystallization. The difficulties originate from the intrinsic nature of the membrane proteins, consisting of a hydrophilic part facing the aqua environment and a hydrophobic surface embedded in a lipid bilayer (Figure 3).

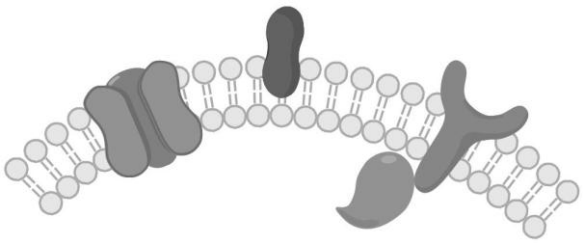

Figure 3. The biological membrane. The lipid bilayer is shown in light grey, with polar head groups as circles that are connected by fatty acid tails. Integral or peripheral membrane proteins are shown in dark grey [15].

With help of detergents, membrane proteins can be extracted from the native lipid bilayer or refolded in an aqueous solution and the same architecture arrangement of the hydrophilic/hydrophobic parts can be kept. In the aqueous solution, detergent molecules form a micelle around the hydrophobic surface of the protein, hence help stabilizing the membrane protein in aqueous solution (Figure 4). In nature, membrane proteins have evolved to be sufficiently stable in a membrane full of lipids. However, they have not evolved to be stable in an environment of detergents. Thus, membrane proteins often have poor stability in detergent solutions, especially in the detergents with short aliphatic chains and small or charged head groups, which are the best suited detergents for growing diffraction-quality membrane protein crystals. Consequently, detergents exclusively used for membrane protein purification often have longer aliphatic chains and big head groups, such as DDM, which are mild and more efficient at stabilizing the membrane proteins and keeping their intact function, but prevent the necessary crystal contacts.

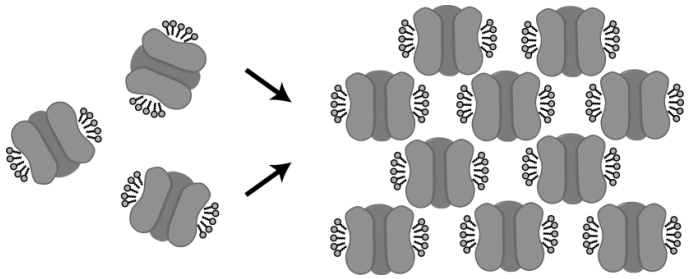

Figure 4. Solubilized membrane protein in solution and crystal form. Hydrophilic parts of the membrane proteins are covered by detergent molecules in solution (left). In crystal line, crystal contacts are mainly formed by hydrophilic parts of the membrane proteins (right) [15].

While purified detergent-solubilized membrane proteins are still required as a starting point, approaches, including lipidic-cubic phase, lipidic-sponge phase, and bicelle crystallization methods, have been developed to immerse purified membrane protein within a lipid-rich matrix before crystallization. This environment is hypothesized to contribute to the protein's long-term structural stability and thereby favor crystallization [17].

2.2.3 Lipidic-cubic phase crystallization (LCP)

During the last decade, several alternative methods of membrane protein crystallization have been established to solve the problems of using detergents. In 1996 the concept of the lipidic-cubic phase crystallization technique was introduced. The idea is to bring the membrane proteins back to the lipidic environment after they have been purified with help of detergents, thus facilitating the crystallization [18]. Monoacylglycerols (MAGs) form the most widely used and best studied bi-continuous lipidic-cubic phases and of these, monoolein (MO) is the most frequently used for membrane protein crystallization (Figure 5 A). When MO is mixed with water, depending on temperature and water/lipid ratio a range of phases can be formed.

Bi-continuous cubic phases (Q) are spontaneously formed when the protein solution is mixed with MO. The detergent molecules are presumably integrated into the lipids and the membrane protein molecules are incorporated into the highly curved cubic phase. By adding the precipitant solution or salt, water molecules are osmotically withdrawn from the interior part of the cubic phase. Gradually, the bilayer curvature increases and clusters together and locally flattens regions of the cubic phase into planar lamellar (L_α) stacks into which the proteins move [19]. When the protein molecules pack together and associate, they can eventually nucleate and grow into crystals (Figure 7 A).

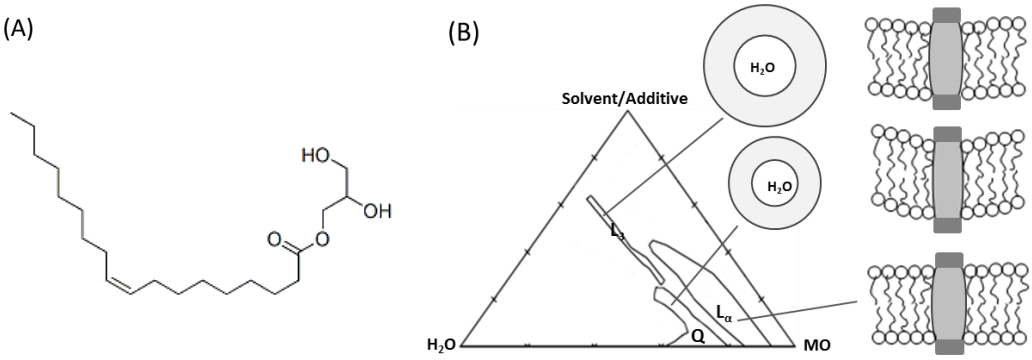

Figure 5. Phase map. MO (A) and schematic phase map for the solvent/additive-MO-H_2O system and schematic representation of the water channels with surrounding MO bilayers (B). The three phase regions are indicated: lamellar phase (L_α), cubic phase (Q) and sponge phase (L_3). To the right is shown a highly curved bilayer of the cubic phase, a reduced curvature of the bilayer of the sponge phase and a flat lamellar phase. The figure is adapted from [16].

2.2.4 Lipidic-sponge phase (LSP)

Almost ten years later, a liquid analogue of the LCP, the lipidic-sponge phase was developed by Caffrey et al. [20]. Usually, the protein diffusion rate in the lipid bilayer is low. In order to make it easier for the proteins to diffuse in the lipid bilayer, the highly curved cubic phase can be flattened out by several methods. One option is to add solvents or other additives, such as jeffamine, polyethylene glycol (PEG) or 2-methyl-2,4-pentanediol (MPD), to the MO/water system in order to create a sponge phase, L_3 phase [21]. In a MO-solvent-water system, it starts from the cubic phase. When solvent is added, the area exposed to the aqueous domains increases and the bilayer interaction decreases. The cubic phase swell until a liquid phase lacking long-range order is formed. The sponge phase can be regarded as a diluted or melted cubic phase with two-three times larger aqueous pores than in the cubic phase (Figure 5 B) [22]. Since it contains large aqueous pores, membrane proteins with larger hydrophilic domains may be incorporated easier in this phase than in the cubic phase.

2.2.5 Lipidic bicelle crystallization

Bicelles are a mixture of aliphatic long chain lipids, between 12 and 18 carbons, and short chain lipids or detergent, 6 to 8 carbons [23]. The morphology of bicelle is fairly adaptable depending on composition, temperature and hydration. The most recognized phase behavior of a bicelle mixture is a nano-disc with the long chain lipids present in majority in the disc plane and the short chain lipids or detergents mainly distributed in the torus of the disc (Figure 6 A). Within the last 20-30 years bicelles have been a membrane model system developed in the NMR field, where bicelles with neutral or charged lipids self-align parallel or perpendicular in a magnetic field depending on the charge of the lipid [24]. Bicelle sample can be rapidly spun at variable angles in a magnetic field [25]. Hence, orientation, structure and dynamic of membrane proteins inserted into these bicelles can be studied by both solid-state NMR and multidimensional liquid-state NMR spectroscopy [26, 27].

In 2001, bicelle crystallization was presented as another lipidic crystallization technique by Salem Faham and James U. Bowie [28]. Well-diffracting crystals of bacteriorhodopsin formed from bicelle, a method which is flexible and simple to use. Bicelle tends to form small bilayer disks at low temperature, and appears to form a perforated lamellar phase at higher temperature [29]. Detergent-purified membrane proteins can be readily reconstituted into the bilayered discoidal lipid-detergent assemblies, where they are maintained in a native-like bilayer environment (Figure 6 B). Due to the liquid-like nature of protein-bicelle mixture at low temperature like 4 °C, it is easy to measure the functionality of a membrane protein in bicelles before the crystallization setup to optimize the bicelle composition [30]. Moreover, protein–bicelle mixtures can be manipulated in the same easy way as membrane proteins-detergent solution, making bicelles compatible with all commercially available screens and standard equipment including high-throughput crystallization robots. A number of membrane proteins have now been successfully crystallized using the bicelle method, including bacteriorhodopsin, β2 adrenergic receptor, voltage-dependent anion channel, xanthorhodopsin and rhomboid protease [31-35].

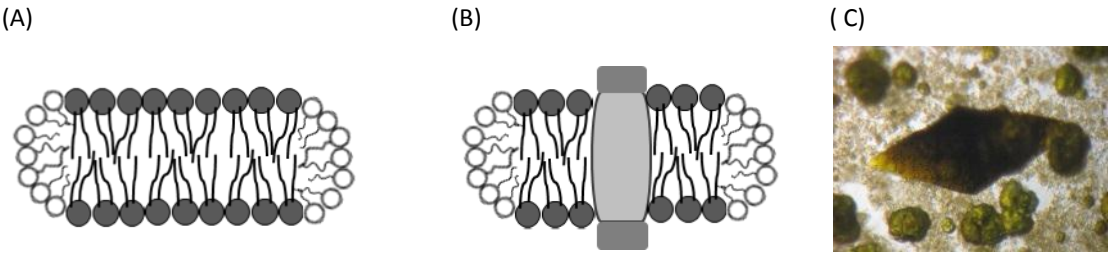

Figure 6. Bicelles. Bicelle in form of small bilayer disk (A) is formed by lipid with head group (dark grey) and double tails, and detergent molecule with head group (white) and single tails. Membrane protein can be reconstituted into the bicelle bilayered disk (B). Crystals of reaction center-light harvest 1 complex formed with bicelle crystallization method (C).

2.2.6 Crystal packing

Three-dimensional crystals of membrane proteins can pack in two different types: type I crystals and type II crystals (Figure 7). Protein molecules in type I crystals are packed in lipid monolayers and stabilized via hydrophobic interactions. The monolayers are then on top of each other and stabilized via both hydrophobic interactions and hydrophilic interactions between the polar groups of the proteins. They tend to diffract well and are more stable and robust. This is the type of crystals that usually forms when crystallizing with the lipidic phase methods. Type II crystals are often grown from the conventional detergent vapour diffusion technique. The hydrophilic regions of the proteins interact with each other and leave large solvent spaces around the hydrophobic parts.

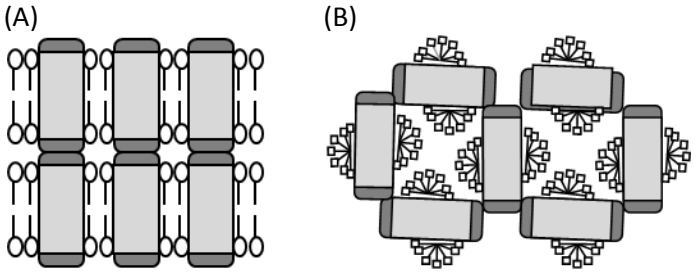

Figure 7. Crystal packing. Type I crystal packing (A): membrane proteins (grey) packed in lipid (white head group) monolayers, and type II crystal packing (B): membrane proteins interact with each other by their hydrophilic regions [16].

2.3 Protein X-ray crystallography

The most commonly used method for determining protein three-dimensional structure is X-ray crystallography. A crystal is a solid in which the constituent atoms, molecules or ions are packed in a regularly ordered, repeating pattern extending in three dimensions. When the crystal is exposed to X-rays, the photons interact with the core electrons and are scattered in all directions. The scattered beams are either reinforced or extinguished by constructive or destructive interference, producing a diffraction pattern. Bragg's law gives the condition for constructive interference:

$$n\lambda = 2d \sin\theta \qquad \text{(Equation 1)}$$

- n is an integer
- λ is the wavelength of x-ray beam used
- d is the distance between the planes in the atomic lattice
- θ is the angle between the incident x-ray beam and the scattering planes

The diffraction of X-rays through the closely spaced lattice of atoms in a crystal is recorded and then analyzed to reveal the nature of that lattice, which leads to an understanding of the molecular structure. To be able to solve the structure, a complete dataset needs to be collected by rotating the crystal in order to cover most parts of the reciprocal space. The data collected from a diffraction experiment is a reciprocal space representation of the crystal lattice.

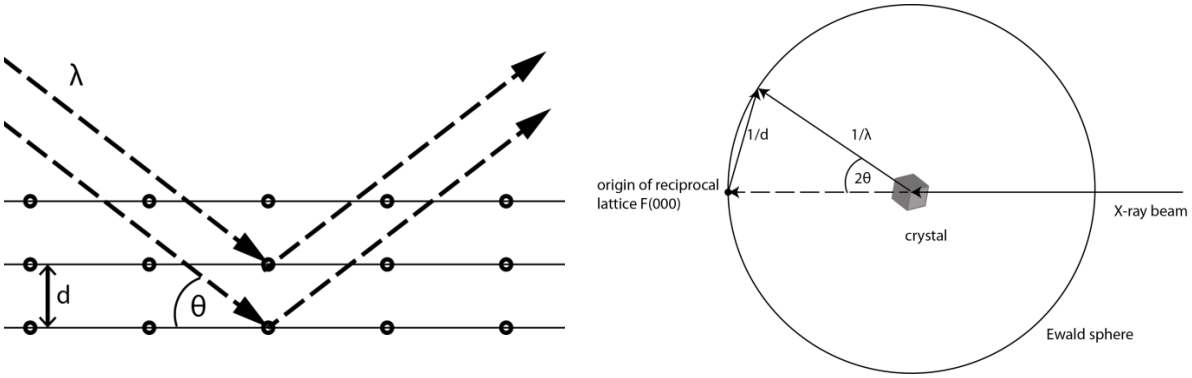

Figure 8. Illustration of Bragg's law [15].

2.3.1 Structure determination

In crystallography, the structure factor F_{hkl} (static structure factor) describes how a crystal scatters incident radiation resulting in diffraction spots. It is a mathematical function defining the amplitude and phase of a wave diffracted from crystal lattice planes characterized by Miller indices *h,k,l*. The electron density ρ(*x,y,z*) at a position *x,y,z* in the unit cell can be fully described by the inverse Fourier transformation of the structure factor:

$$\rho(x,y,z) = \frac{1}{V}\sum_{hkl}|F_{hkl}|e^{-2\pi i(hx+ky+lz)+i\alpha(hkl)} \qquad \text{(Equation 2)}$$

Here *V* is the volume of the unit cell, *h*, *k*, *l* are the Miller indices, $|F_{hkl}|$ is the amplitude of structure factor for a respective reflection *hkl*, α is phase angle and *x, y, z* define a certain position in the unit cell. However, only the intensity of each diffraction spot can be recorded experimentally and it is proportional to the square of the structure factor amplitude $|F_{hkl}|^2$.

2.3.2 The phase problem

The data obtained by X-ray diffraction lacks phase information, thus calculating the electron density directly from it is not possible. There are several ways developed to obtain initial phases. Experimental phasing methods like MAD (multiple anomalous diffraction) and SAD (single anomalous diffraction) can be used, together with more traditional methods such as MIR (multiple isomorphous replacement) and SIR (single isomorphous replacement) [36].

However, most protein structures are determined by using phase from an existing structure of a homologous protein or fragment. Molecular replacement is still the most frequently used technique to overcome the phase problem due to the remarkable growth of available protein structures in the Protein Data Bank (PDB). The Patterson function

$$P_{uvw} = \frac{1}{V^2}\sum_{hkl} F_{hkl}^2 \cos[2\pi(hu+kv+lw)] \qquad \text{(Equation 3)}$$

is an alternative expression of Equation 2. It represents an interatomic vector map containing one peak for each atom related to every other atom and is created by squaring the structure factor amplitudes and setting all phases to zero. Thus Patterson maps for the data derived from unknown structure and from model, the structure of a previously solved homologue, are both generated. Subsequently, rotation function is performed in Patterson space while translation search is applied in real-space to move the model through the unit cell. In the correct orientation and position within the unit cell, the two Patterson maps should be closely correlated and the initial phase can be obtained. The quality of the solutions can then be evaluated by a standard linear correlation coefficient or an R_{factor}.

2.3.3 Structure refinement and validation

Once phases have been obtained, the first electron density map of the unknown structure can be calculated and a preliminary model can be built into the density map. The model is then refined by iterating between automated refinement in reciprocal space and manual real space refinement. The agreement of the model with the data is evaluated with the help of the crystallographic R_{factor}:

$$R = \frac{\sum_{hkl}||F_{obs}(hkl)|-|F_{cal}(hkl)||}{\sum_{hkl}|F_{obs}(hkl)|} \qquad \text{(Equation 4)}$$

Here the experimental structure factor F_{obs} is from the observed data and the model structure factor F_{cal} is the structure factor calculated from the model. As refinement progresses, the value of the *R-factor* should decrease since ideally F_{obs} and F_{cal} should be identical. Although in practice, the value of 'the highest resolution divided by 10' is generally considered as a reasonable R_{factor}. Usually, an additional factor R_{free}, calculated as R_{factor} but the data used for calculation is a small fraction of reflections (usually 5%) that is never included in the refinement process, is commonly used as cross-validation. Thus the R_{free} becomes a less biased indicator for the quality of the model. Large discrepancies between these two values can be interpreted as implying that the refined structural model might be over-fitted to the data during the refinement.

2.3.4 Data-quality validation and high-resolution cut-off

Crystallographic data-quality is evaluated by an analogous indicator R_{merge} (or R_{sym})which measures the spread of *n* independent measurements of the intensity of an unique reflection, $I_i(hkl)$, around the average $\bar{I}(hkl)$ [37]:

$$R_{merge} = \frac{\sum_{hkl}\sum_{i=1}^{n}|I_i(hkl)-\bar{I}(hkl)|}{\sum_{hkl}\sum_{i=1}^{n}I_i(hkl)} \qquad \text{(Equation 5)}$$

Since $I_i(hkl)$ values influence $\bar{I}(hkl)$, R_{meas}, equals to R_{merge} times a factor of $\sqrt{n}/(n-1)$, was used as a replacement of R_{merge} to give an evaluation which is not dependent of the multiplicity [38]. Because the intensity of diffraction decreases with resolution, a high-resolution cut-off is applied to discard data which go beyond the cut-off resolution since they are often considered as noise and their inclusion may degrade the quality and add bias to the model. As a rule of thumb, typically data are truncated at a resolution before the R_{merge} (or R_{meas}) value exceeds 0.6 – 0.8, and before the signal-to-noise ratio ($\langle I/\sigma(I)\rangle$) drops below 2, although many uncertainties are associated with these criteria [39, 40].

At high-resolution data, background noise dominates the numerator part of the equation while the denominator approaches zero due to the average net intensity. Thus, the commonly applied

criterion of high-resolution cut-off is not as reliable as it appears. Data-quality *R*-values are not comparable to *R*-values from model refinement. In May 2012, Karplus and Diederichs proposed using Pearson correlation coefficient (CC), a measure of linear association between two variables, as a parameter that could potentially evaluate both data accuracy and the agreement between model and data on a common scale [41]. Pearson's CC has been efficiently used in crystallography in anomalous scattering data process [42].

Unmerged data is divided into two parts, each containing a random half of the measurements of each unique reflection, I_1 and I_2. The CC, denoted $CC_{1/2}$, is calculated between the average intensities of each subset, $\langle I_1 \rangle$ and $\langle I_2 \rangle$:

$$CC_{1/2} = Corr(I_1, I_2) = Corr[(I_1 - \langle I_1 \rangle), (I_2 - \langle I_2 \rangle)] \qquad \text{(Equation 6)}$$

Since $CC_{1/2}$ measures the correlation of one noisy data subset with another noisy subset, thus $CC_{1/2}$ should be expected to be underestimating the real information content of the data. The true signal would be measured by so called CC_{true}, the correlation between the averaged data set of these two subsets, thus less noisy due to the extra averaging, and the noise-free true signal. Although the true signal would normally not be known, the relation between $CC_{1/2}$ and CC_{true} can be derived with the assumption that errors in the two half-data sets are random and on average of similar size:

$$CC^* = \sqrt{\frac{2CC_{1/2}}{1+CC_{1/2}}} \qquad \text{(Equation 7)}$$

Here CC^* estimates the value of CC_{true} based on a finite-size sample. CC^* approaches 1.0 at low resolution and drops to near 0.1 at high resolution cut-off. Moreover, CC_{work} and CC_{free}, the standard and cross-validated correlations of the experimental intensities with the intensities calculated from the refined molecular model, can be calculated on the same scale and directly compared with CC^*. A CC_{work} larger than CC^* indicates over-fitting while a CC_{free} smaller than CC^* suggests that the model does not account for the entire signal in the data. A CC_{free} closely matching CC^* implies that data quality is limiting model improvement, such as at high resolution. CC^* can be used not only to evaluate data quality but also to link crystallographic model quality with data quality [41]. In this thesis, CC^* was used as criterion for high-resolution cut-off when low resolution X-ray crystallographic data of reaction center-light harvesting 1 complex was analyzed.

3 Trypsin: an example of serine protease catalysis (paper I)

3.1 Serine protease catalysis

Proteases are a group of proteolytic enzymes found in all organisms. They are functioning in digestion, posttranslational processing of secreted proteins, neurotransmitters and hormones, blood coagulation and complement fixation [43]. Many of these enzymes are of medical importance, and hence are potential drug targets. Basically, there are four distinct proteases based on the catalytic type namely: serine, cysteine, aspartic and metallopeptidases.

By delocalization of the nitrogen lone pair into the carbonyl group, the peptide bond is a strong linkage with a high degree of double-bond character. During hydrolysis of a peptide bond, all three heavy atoms of the peptide are directly involved in the catalytic reaction by interacting with appropriate catalytic groups of the enzyme (Figure 9) [44]. Thus, the collective action of catalytic groups is essential. In the first step, the strength of the peptide bond weakens by a nucleophile, such as a serine OH group or a water molecule, which attacks the carbonyl carbon atom of the peptide. With the help of a general base, which accepts the proton from the nucleophilic OH group, a tetrahedral intermediate forms. Subsequently, the intermediate state is stabilized by an electrophilic catalysis step. It is followed by the leaving of the amine group from the tetrahedral intermediate. Since the amine is a poor leaving group, its protonation by a general acid is essential. So the whole process is facilitated by a general base-general acid catalytic mechanism.

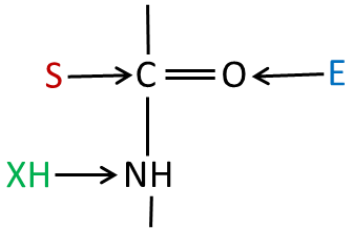

Figure 9. Common features of peptide bond cleavage by proteases. S represents the nucleophile and E the electrophile. XH acts as a general-acid which donates its proton to the leaving amine group NH.

Through kinetic, chemical, physical and genetic analyses, serine proteases are among the most studied protein families. In addition, the discovery of many three-dimensional structures has led to a better understanding of enzyme catalysis [45, 46]. As shown in Figure 10, all serine protease families are presumed to share a common catalytic mechanism, with a catalytic triad composed of Asp-His-Ser [45, 47]. Serine proteases cleave peptide bonds in two major steps: acylation and deacylation. When a suitable substrate binds to the enzyme, it induces a conformational change in the enzyme so a transient β-sheet forms. Consequently, the active site is positioned next to the scissile peptide bond. The catalytic serine attacks the scissile peptide bond and a short-lived tetrahedral intermediate forms. A proton of the serine is then transferred to the catalytic histidine. The tetrahedral intermediate is rapidly deacylated by the reverse of the acylation steps followed by the release of the resulting carboxylate product, thereby regenerating the active enzyme (Figure 10).

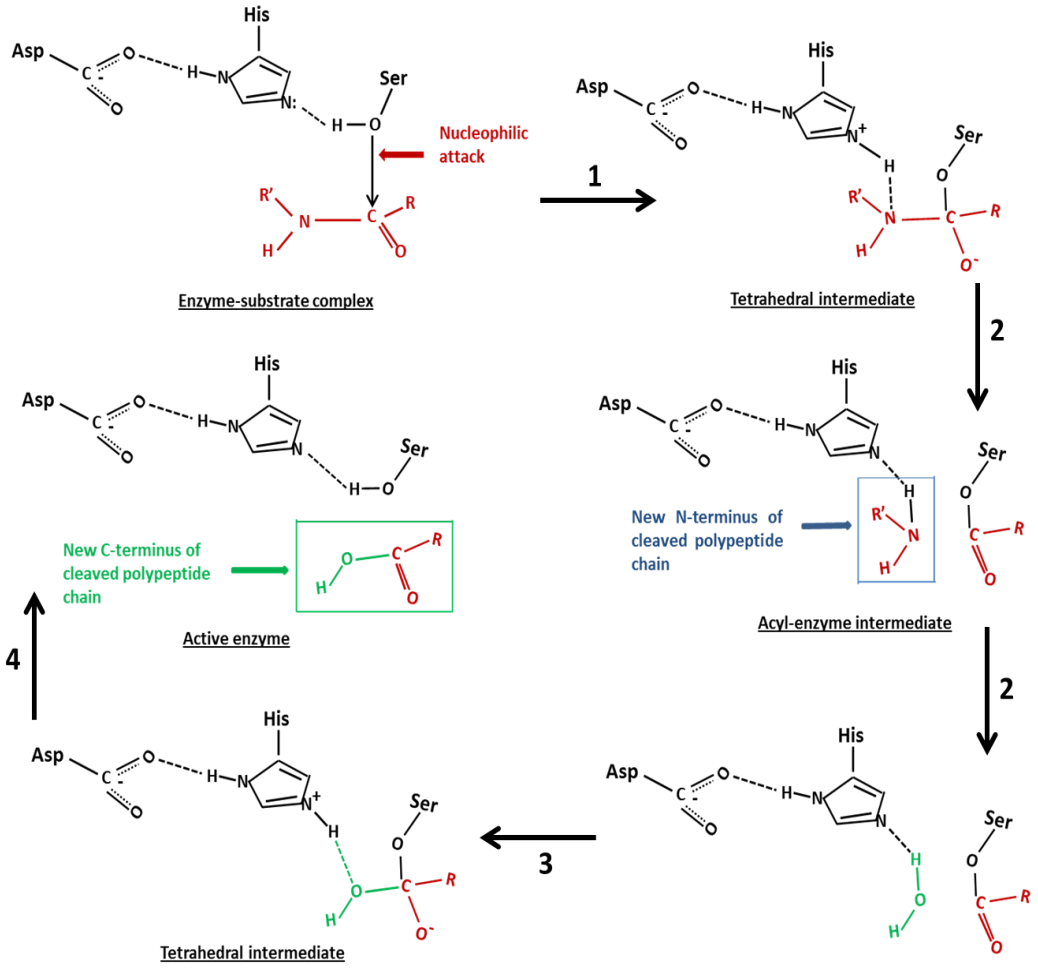

Figure 10. The catalytic mechanism of serine proteases (adapted from [4]) . The reaction involves **(step 1)** the nucleophilic attack of the active site Ser on the carbonyl carbon atom of the scissile peptide bond to form the tetrahedral intermediate and His accepts a proton from Ser; **(step 2)** the decomposition of the tetrahedral intermediate to the acyl-enzyme intermediate through general acid catalysis by the active site Asp-polarized His, followed by loss of the amine product and its replacement by a water molecule; **(step 3)** the reversal of step 2 to form a second tetrahedral intermediate; **(step 4)** the reversal of step 1 to yield the reaction's carboxyl product and the active enzyme.

Serine functions as the primary nucleophile. At different steps of the reaction process histidine functions as proton acceptor and donor. However, the role of aspartic acid is unclear. One theory is the charge relay mechanism suggested by Blow et al. who propose that the histidine becomes protonated by the OH group of serine, subsequently transferring one of its protons to the catalytic aspartate which is negatively charged in the free enzyme, thus neutralizing the active site [45]. However, the most critical point of this theory is based on the correct pK_α value and protonation state of the histidine. In the free, solvent-accessible form the pK_a of the histidine is around 7.0 - 7.5 while aspartate residue has an assumed pK_a around 3, which does not motivate the transfer of a proton from histidine to aspartate since histidine is a better proton keeper than aspartate [48, 49]. Site-directed mutagenesis of the aspartate to neutral asparagine was examined and catalytic rate was decreased by a factor of 10^4 which supported the role of a negative charged residue near the histidine [50]. Hence, the protonation state and the task of the buried aspartate residue at the catalytic triad are still under debate.

3.2 Co-crystallization

Recent technical advances in crystallographic analysis, particularly improvements in X-ray flux at third generation synchrotrons with high brilliance and highly focused synchrotron beam lines, have significantly improved the resolutions to better than 1 Å which now becomes achievable for many macromolecular crystal structures. With sub-ångstrom resolution, the level of visualized detail of the structure is increased, including the location of hydrogen atoms [51, 52]. Hydrogen atoms play a significant role in the catalytic mechanism of enzymes. By monitoring the hydrogen occupancy in structures, captured at a series of pH values, the pK_α can be directly determined [53]. Nevertheless, to obtain such ultrahigh resolution diffraction, the key step is to grow crystals of exceptional quality.

Protein complexes formed by components with weak binding strength usually show different potential binding modes, thus their crystals often suffer from disorder and do not diffract well [54]. Optimization of contact interfaces between a protein and its binding partner can lead to increased binding. In this scenario complexes formed are less dynamic which may improve crystal quality[55]. Wild-type SGTI (also named SGPI-1) is a weak inhibitor of bovine trypsin (BT) with an inhibition constant in micromolar range [56, 57]. Using a phage display assay, a set of SGTI mutants were selected that showed improved binding to bovine trypsin [56, 57]. The unique combination of mutations improved the inhibition constant of the mutant inhibitor SGPI-1-PO-2 from micromolar to picomolar range and made it possible to crystallize the complex [56]. From a crystallization buffer at pH 4.6, crystals of the complex were produced and diffracted to a ultrahigh 0.93 Å resolution.

3.3 Density functional theory

In physics and chemistry Density functional theory (DFT) is a quantum mechanical modeling method used to investigate electronic structure, principally ground state, of many-body systems in atoms, molecules and condensed phases. Thus, the properties of a many-electron system can be determined by using functionals of the spatially dependent electron density. DFT has been increasingly used in applications related to biological systems, complementing experimental investigations. Hence, it allows a close connection between theory and experiment, serving either to validate the conclusions that have been reached from the analysis of the experiments, or to distinguish between those possibilities that were left open. It often leads to critical clues about the geometric, electronic, spectroscopic properties of the systems being studied [58]. Therefore, the experimental coordinates of the simplified active site in the structure were optimized using DFT with the aim of supporting interpretation of the protonation state of catalytic residues observed from the experimental electron density (Figure 11).

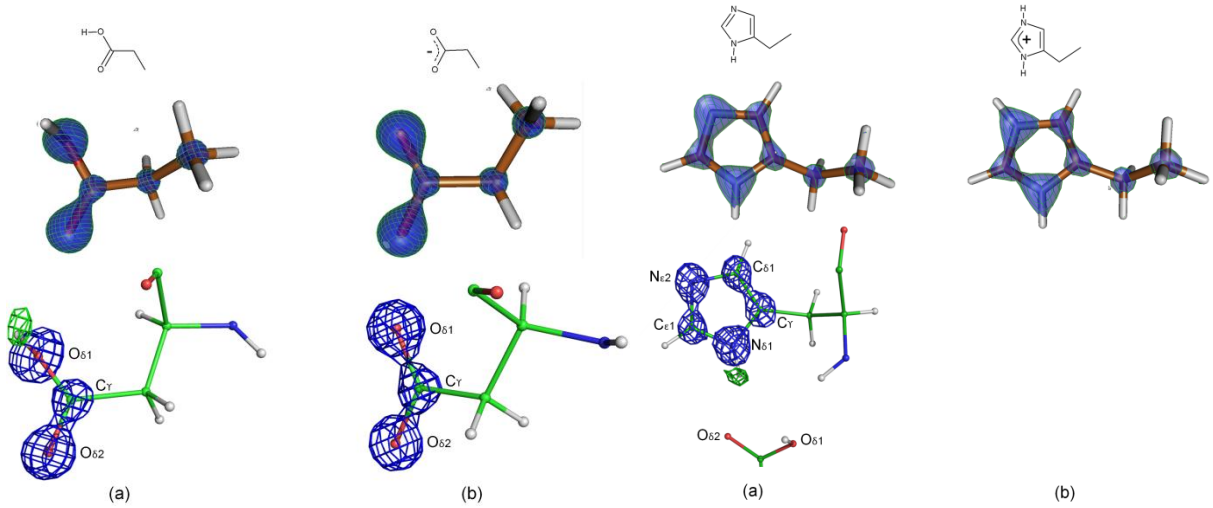

Figure 11. Comparison of the theoretical protonated and deprotonated carboxyl groups to experimental electron density in the crystal structure at 0.93 Å resolution (left). (a) electron density of Asp-102 compared with theoretical protonated propionic acid. (b) electron density of Asp-189 compared with theoretical deprotonated propionic acid. $2mF_{obs}$ - DF_{calc} electron density map (blue) is contoured at 4.5 σ, and mF_{obs} – DF_{calc} density (green) is contoured at 2.5 σ. Alternative protonation states of the imidazole ring as calculated by density theory and experimentally observed at residue His-57 in the electron density maps at 0.93 Å resolution (right). $2mF_{obs}$ - DF_{calc} electron density map (blue) is contoured at 3.5 σ, and mF_{obs} – DF_{calc} density (green) is contoured at 3.0 σ.

4 Photosynthetic reaction center and reaction center - light harvesting 1 core complex from *Blastochloris viridis* (paper II, III)

A huge amount of solar energy reaches the surface of the earth every day. In order to harvest and concentrate the solar energy, photosynthetic organisms have evolved light-harvesting systems. Non-sulfur purple photosynthetic bacteria are anaerobic phototrophic bacteria which capture energy from light by anoxygenic photosynthesis. They use carotenoids and bacteriochlorophylls to absorb blue-green light and far-red light, which have wavelengths beyond 750 nm. These light-absorbing pigments are non-covalently attached to specific integral membrane proteins named light-harvesting complexes (LH). There are two types of light harvesting complexes: the peripheral LH2 and the core LH1 which surrounds reaction center (RC) and together is referred as reaction center and light harvesting 1 core complex (RC-LH1). The light-harvesting antennas absorb light and funnel it into the reaction center where a chain of electron transfer reactions is initiated. The transfer of electrons is coupled by the release of protons from one side of the membrane to the other side generating an electrochemical proton gradient across the membrane. This proton motive force drives the membrane protein ATP synthase to produce Adenosine-5'-triphosphate (ATP) from adenosine diphosphate (ADP) and inorganic phosphate (Pi). In purple photosynthetic bacteria, nearly 100 % quantum yielded energy is harvested from solar lights and converted into ATP, the biochemical energy currency. Importantly, the system is environment-adaptive and maintains efficiency under changing conditions.

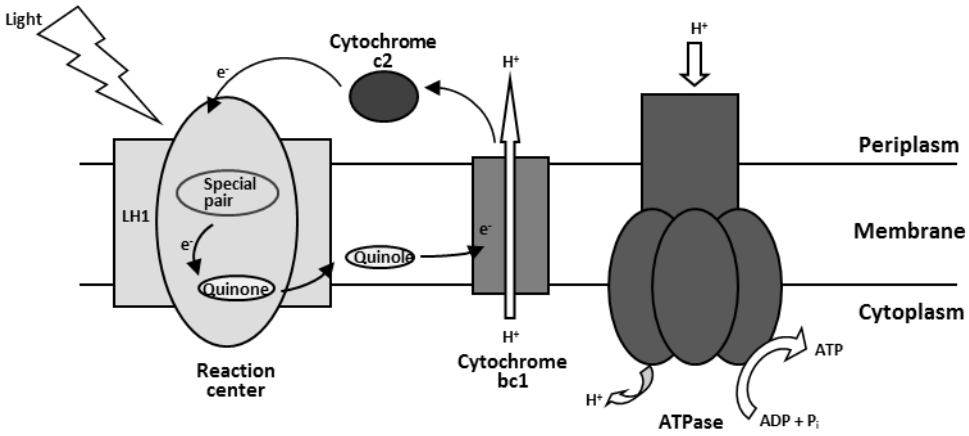

Figure 12. Schematic representation of light-induced electron transport and ATP synthesis in the membrane of *Blastochloris viridis*. The reaction center, light-harvesting 1, cytochrome c2, cytochrome bc1, ATPase are all integral inner membrane proteins [16].

4.1 Photosynthetic reaction center from *Blastochloris viridis* (paper II and III)

The first structure of a purple bacterial reaction center was the reaction center from *Blastochloris viridis* determined in 1985 representing the first membrane protein structure solved in atomic resolution by X-ray crystallography (Figure 13). The work performed by Deisenhofer, Huber and Michel was awarded the Nobel Prize for Chemistry in the year 1988 [59, 60]. *Blastochloris viridis* (*B.*

viridis) belongs to non-sulfur purple photosynthetic bacteria. The reaction center of *B. viridis* (RC_{vir}) consists of four protein subunits: Light (L), Medium (M), Heavy (H) and cytochrome C (C) subunit, a cytochrome with four covalently bound c-type hemes. L and M subunits possess five transmembrane helices each and form the transmembrane (TM) domain, while the H subunit is loosely anchored to the membrane by one α-helix [61-63]. The additional C subunit is located at the periplasmic side of the membrane. The M subunit is prolonged by 18 extra residues at its C-terminus and interacts with the C subunit. In RC_{vir} scaffold, there are 10 non-covalently bound co-factors, four bacteriochlorophyll-b (BChl), two bacteriopheophytin-b (BPhe), a menaquinone (MQ_A) as the primary electron acceptor, a mobile ubiquinone (UQ_B), a non-heme iron, a carotenoid, and four covalently bound hemes. Two of the BChls, named special pair (P_{960}), strongly interact with each other and have an absorbance maximum at 960 nm. The L and M subunits form a scaffold for the co-factors and position them in such a way as to achieve optimal efficiency in function. With a sequence identity of 33 %, structural similarity of L and M subunits give rise to a high degree of two-fold symmetry in the TM domain [61, 62, 64]. A two-fold rotational pseudo-symmetry axis running through P_{960} and the non-heme iron divides the co-factors of RC_{vir} into A and B branches (except the carotenoid molecule which locates at B branch only).

Upon illumination of reaction center, either by direct activation by light or by energy transferred from the surrounding light-harvesting antenna complex, one photon is absorbed by the P_{960}. It reaches its photo-activated state P_{960}^* and one electron from P_{960}^* is transferred through the bacteriopheophytin to the tightly bound menaquinone MQ_A in A branch side. This leaves the special pair with a positive charge and the MQ_A with a negative charge, the so called charge separated state $P^+Q_A^-$ [64]. P_{960}^+ is then reduced by the C subunit. Subsequently, the electron is transferred from MQ_A^- to the mobile ubiquinone UQ_B and the reaction cycle is ready for another excitation event. A second photon absorption yields a second electron transfer. Consequently, negatively charged UQ_B^{2-} is neutralized by the pick-up of two protons from the cytoplasm and UQ_BH_2 is released into the membrane.

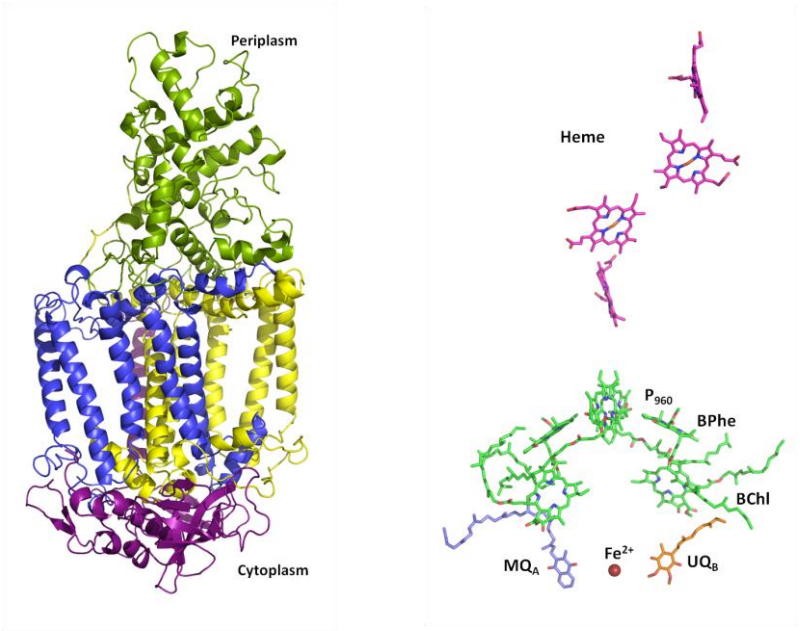

Figure 13. Ribbon representation of the structure of RC_{vir} (left): L in blue, M in yellow, H in purple and C in green. Cofactors (right).

The role of the additional C subunit is to re-reduce the photo-oxidized primary electron donor P_{960} with its covalently bound hemes. A soluble protein named cytochrome c_2 then cycles back electrons from the cytochrome bc_1 protein to the C subunit of RC_{vir}. The detailed mechanism about how the C subunit, with its four covalently bound hemes, re-reduces the special pair, makes the RC_{vir} a more complicated system compared with some reaction centers from other purple bacteria and has been subject of interest [65].

With time-resolved Laue diffraction experiments, structural changes occurring in the reaction center upon photo-activation can be studied in detail. To be able to tolerate both the laser light flash and the polychromatic X-ray beam during the time-resolved Laue experiments, tremendously robust and stable protein crystals are required. Crystals of membrane proteins obtained by detergent-based crystallization method are frequently fragile and poorly diffracting. With the aim of hunting high quality crystals suitable for such complicated experiment, the lipidic phase technique developed for crystallization of the reaction center from *Rhodobacter sphaeroides*, another purple bacterium, has been applied to the crystallization of RC_{vir} [66, 67].

4.1.1 The purification and lipidic-sponge phase crystallization of RC_{vir}

Following the published purification protocol for RC_{vir} did not yield adequate amounts of protein [68]. In the original protocol, up to 1% LDAO was added to the membrane solution and stirred for 10 minutes at room temperature in the dark. This was followed by an ultracentrifugation step, any remaining membrane pellet was again resuspended. The LDAO-solubilization / ultracentrifugation steps were repeated 3-4 times. The supernatant, except from the first extraction, contained mixtures of RC_{vir} and RC-$LH1_{vir}$ and could be loaded onto an anion-exchange column DE52. In practice, after the first ultracentrifugation it was difficult to resuspend the remaining membrane pellet to get a homogenous membrane solution and the whole process was very time-consuming. So instead of having 3-4 solubilization cycles, the protocol was optimized so that LDAO was increased to 4 % and the incubation time to 3 - 4 hours. It was followed by one ultracentrifugation step and the remaining membrane pellet was discarded. The solubilized membrane protein solution could be monitored by UV-VIS spectroscopy. Compared with UV-VIS measurement of the solubilized membrane sample from the original protocol, it was apparent that more RC_{vir} (A_{830}) was released from the $LH1_{vir}$ (A_{1012}) during the one-step solubilization process. Disrupting the whole RC-$LH1_{vir}$ core complex and releasing the RC_{vir} is the intention of using LDAO as detergent since it is a relatively harsh detergent compared with CHAPS and DDM. The purification protocol was optimized from the original one with some other modifications [68]. After the solubilization step, another anion-exchange chromatography column POROS HQ was chosen due to its high binding capacity which is not affected by the high detergent content in the solubilized membrane solution loaded on the column. It was then followed by a HiPrep 16/60 Sephacryl S-300 size-exclusion chromatography step. The final protein solution was concentrated to 20-25 mg/mL and the optical purity ratio (A_{280}/A_{830}) of the sample was approximately 2.1 before crystallization setup.

First, lipidic-cubic phase crystallization method was tried on RC_{vir} without any success. This was probably due to the high curvature, and 2-3 times smaller water channel of the cubic phase, compared to lipidic-sponge phase [22]. Owing to the fact that RC_{vir} has an additional C subunit which forms a long hydrophilic domain on the periplasmic side, larger hydrophilic pores between the lipid-bilayers are needed to be able to fit the C subunit in the monoolein/water system. Thus a swelled lipidic-cubic phase, lipidic-sponge phase, seemed to be a more suitable alternative.

The protein crystals of RC_{vir} were grown using the lipidic-sponge phase crystallization technique, with monoolein as the crystallization medium and jeffamine M-600 as the sponge-phase-inducing solvent. Melted monoolein was thoroughly mixed in a ratio of 3:2 (v/v) with 0.1M HEPES (pH 7.5) until a viscous, transparent LCP was obtained. The formed phase was then transferred into a glass vial and sponge-phase-inducing solution (1:4 ratio) was added containing jeffamine M-600 (16% jeffamine M-600, 1M HEPES pH 7.9, 0.7M Ammonium sulphate, 2.5% 1,2,3-Heptanetriol) which swelled the cubic phase to sponge phase. After phase separation overnight, the upper phase (sponge phase) was harvested. Hanging drop experiments with lipidic sponge phase were performed. RC_{vir}, sponge phase and additive (1 M tri-sodium citrate) were pipetted (1:1:1 ratio) onto a cover slide, which was then placed on top of sealed reservoir filled with 400µL solution (0.1 g/l Na/KPO_4, 0.55 M Na-acetate, 0.75 M HEPES pH 6.3) . The plates were stored in the dark at 20 °C. The obtained crystals diffracted to 1.86 Å with space group $P2_12_12$ and have type I crystal packing. Low- and high-dose X-ray data sets were collected and the radiation damage during data collection was investigated.

4.1.2 Lipids bound to the protein surface

Four lipid molecules were captured on the RC_{vir} surface: a covalently bound diacylglycerol lipid and three monoolein molecules, one of which occupied the Q_B ubiquinone binding pocket. The diacylglycerol is located at the N-terminal cysteine of the C subunit. Previous mass spectrometry studies identified a diacylglycerol molecule covalently attached to the cysteine side chain of the C subunit through a thioether bond [69]. The same covalent posttranslational modification of diacylglycerol was also modeled in two recently published RC_{vir} structures (PDB id 3T6D, 3T6E) obtained from detergent-based crystallization method [70]. Furthermore, a high-dose data set was collected to analyze the effect of radiation damage in the structure. Electron density recovered from this high-dose data set shows reduced occupancy of the sulfur atom and the lipid moiety of the diacylglycerol. Bonds involving heavier atoms seem to be more affected by radiation damage which can explain the radiation sensitivity of the thioether bond containing sulfur atom.

A very long (~36 Å) continuous electron density feature, approximately perpendicular to the membrane plane, was observed in both low- and high-dose data sets. It lies within a groove of the membrane region bordered by the transmembrane helix of the H subunit, the L subunit, and co-factors of the A-branch. LDAO molecule was modeled into the same binding site in earlier crystal structures (PDB entry 2PRC) [71]. Thus, in our RC_{vir} structures, the putative lipid exceeds the length of membrane-specific fatty acids in their fully extended conformation. With the aim of identifying this putative lipid, lipid extractions using a methanol-chloroform mixture with different ratios was performed to obtain lipids from both the RC_{vir} crystals and the purified RC_{vir} protein solution for mass spectrometry and thin-layer chromatography analysis. RC_{vir} crystals were fished from their crystallization drops avoiding as much mother liquid/lipid from the drop as possible and washed three times before being dissolved. Notably, it was difficult to dissolve the RC_{vir} crystal which was most likely due to the well-packed and robust character of crystals from LSP, and also to the fact that it was challenging to wash away lipid-like materials that surround the crystals in the drop. LDAO was identified in the protein solution, whereas in the crystal-dissolved sample MO molecules were predominant. Despite our efforts, no lipid found from the samples matches the length of the putative lipid found in the density map.

4.1.3 Q_B binding site (unpublished result)

The Q_B site is a difficult region to model due to the relatively weak binding of the ubiquinone in its active site. Native ubiquinone-9 (UQ_9) gets lost during the normal purification process and the occupancy on Q_B site is approximately 30% in some cases [71]. The low occupancy on the Q_B binding site can be increased by reconstitution with synthetic ubiquinone-2 (UQ_2) which has a significantly shorter isoprenoid tail, thus easier to fit into the Q_B binding pocket. The ubiquinone binding site of RC_{vir} was occupied with a ubiquinone-1 molecule in the original structure obtained from the native source using detergent-based crystallization methods [72]. Presently, there are several published structures of ubiquinone co-crystallized RC_{vir} with detergent-based crystallization method [71, 73, 74].

Attempting to model a ubiquinone molecule into the binding pocket of both high dose and low dose RC_{vir} structures failed. However monoolein, the lipid used as crystallization medium, is a monoacylglycerol with a small head group and it was modeled in the Q_B binding site without residual negative $F_o - F_c$ density (Figure 14 C). The combination of a much higher amount of LDAO and longer incubation time applied at the solubilization step may cause the total loss of ubiquinone at its binding pocket in our structures. In the RC_{sph} purification process, much lower concentration of LDAO (0.8 %) was used to solubilize the membrane protein for a short period of 10 minutes, which may be the reason why Q_B binding pocket was partially occupied by ubiquinone in the LSP structure of RC_{sph} [67]. Detergent solubilization, the extraction of membrane proteins from their native membrane, can cause structural lipids and weakly bound units to be lost.

Ultimately, the Q_B binding problem was solved by co-crystallization of synthetic ubiquinone-2 with RC_{vir} in lipidic-sponge phase (Figure 14). However, using the same batch of purified RC_{vir} protein and sponge phase, an attempt to co-crystallize with an inhibitor, herbicide terbutryn, did not produce crystals with the inhibitor in the Q_B binding site. A monoolein molecule was found bound in the Q_B pocket (Figure 14 C). For RC_{sph} it was suggested that the first two and part of the third isoprene units of the native ubiquinone-10 co-factor interact with the protein interior at both Q_A and Q_{-B} binding site [75]. Furthermore, one- and two-isoprene-substituted quinones bind more tightly than analogs substituted with saturated-alkyl tail substituents. Hence, it is proposed that the binding sites exhibit binding specificity for the native isoprene tails structure, and the tail-protein interaction is designed to prevent competitive interference of quinone function by amphiphilic species present in the native membrane [75]. By comparing the concentration of ubiquinone-2 (1.25 mM) and terbutryn (80 µM) used in the co-crystallization set up, a certain concentration of the co-crystallization molecule is presumably needed in order to compete against monoolein. Thus, this result exhibits that lipidic-sponge phase crystallization method is compatible with co-crystallization approaches (Table 1).

Table 1. Crystallographic data and refinement statistics.

	RC_{vir} – UQ_2	**RC_{vir} - MO**
Space group	$P2_12_12$	$P2_12_12$
Cell constants		
a, b, c (Å)	84.6, 139.7, 179.2	84.9, 139.5, 178.5
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	2.55	2.3
Total reflections	525334 (76270)	709163 (103030)
Completeness (%)[a]	99.3 (99.0)	100 (100)
R_{sym} (%)[a,b]	15.1 (57.2)	17.9 (65.6)
$\langle I/\sigma \rangle$[a]	9.3 (3.6)	8.9 (2.8)
R_{work} (%)[c]	16.96	15.9
R_{free} (%)[c]	21.2	19.5
R.m.s deviations from		
Bond lengths (Å)	0.0182	0.0232
Bond angles (°)	3.0	3.1
Ramachandran plot (% by PROCHECK)		
Most favored	95.07	95.51
Generously allowed	4.15	3.71
Disallowed	0.78	0.78

[a] Values in parentheses indicate statistics for the highest resolution shell.
[b] $R_{sym} = \sum |Io - \langle I \rangle| / \sum Io \times 100\%$, where Io is the observed intensity of a reflection and $\langle I \rangle$ is the average intensity obtained from multiple observations of symmetry related reflections.
[c] $R_{factor} = \sum ||F_{obs}|-|F_{calc}|| / \sum |F_{obs}| \times 100\ \%$.

The use of a 'relative quinone quality' factor (RQQ) has been suggested by C. R. D. Lancaster as a method of estimating the quality of Q_B-site models [76]. Co-factor quality factor (CQ) can be obtained by the equation $CQ = 100*Q*\exp(-B/4*d_{min}^2)$, where Q is the atomic occupancy, B the atomic temperature factor and d_{min} is the maximum resolution of the data set [77]. The molecular CQ value is then determined from the average of respective atomic CQ factors. The RQQ factor is achieved by dividing the molecular quinone CQ values by the respective special-pair CQ factor. For comparison with the literature, the modeled ubiquinone at Q_B binding pocket was reduced to quinone-1 model and the atoms of the phytyl tails of the special-pair were omitted. From our structure the calculated RQQ value of ubiquinone at Q_B binding site was estimated to be 38 %. Compared with RQQ values from two other RC_{vir} structures (PDB id: 1PRC and 2PRC) which are 15 % respective 57 %, the quality of our modeled ubiquinone at Q_B site appears to be acceptable [76].

As shown in Figure 14 B, the head group of UQ_2 is well defined in the electron density. Possible hydrogen bonds between side chain of Ser L223 - UQ_2, His L190 - UQ_2 and backbone of Ile L224 – UQ_2 are indicated by dashed black lines, which have a distance of 2.68 Å and 2.71 Å respectively. Compared to previously published structures with UQ_2 in the binding pocket the distances are very similar (2.80 Å and 2.61 Å respectively) [71]. The UQ_2 is found to bind in a proximal position which may correspond to the position in which a neutral ubiquinone accepts an electron from Q_A^-. Subsequently a tighter binding of the semiquinone is induced.

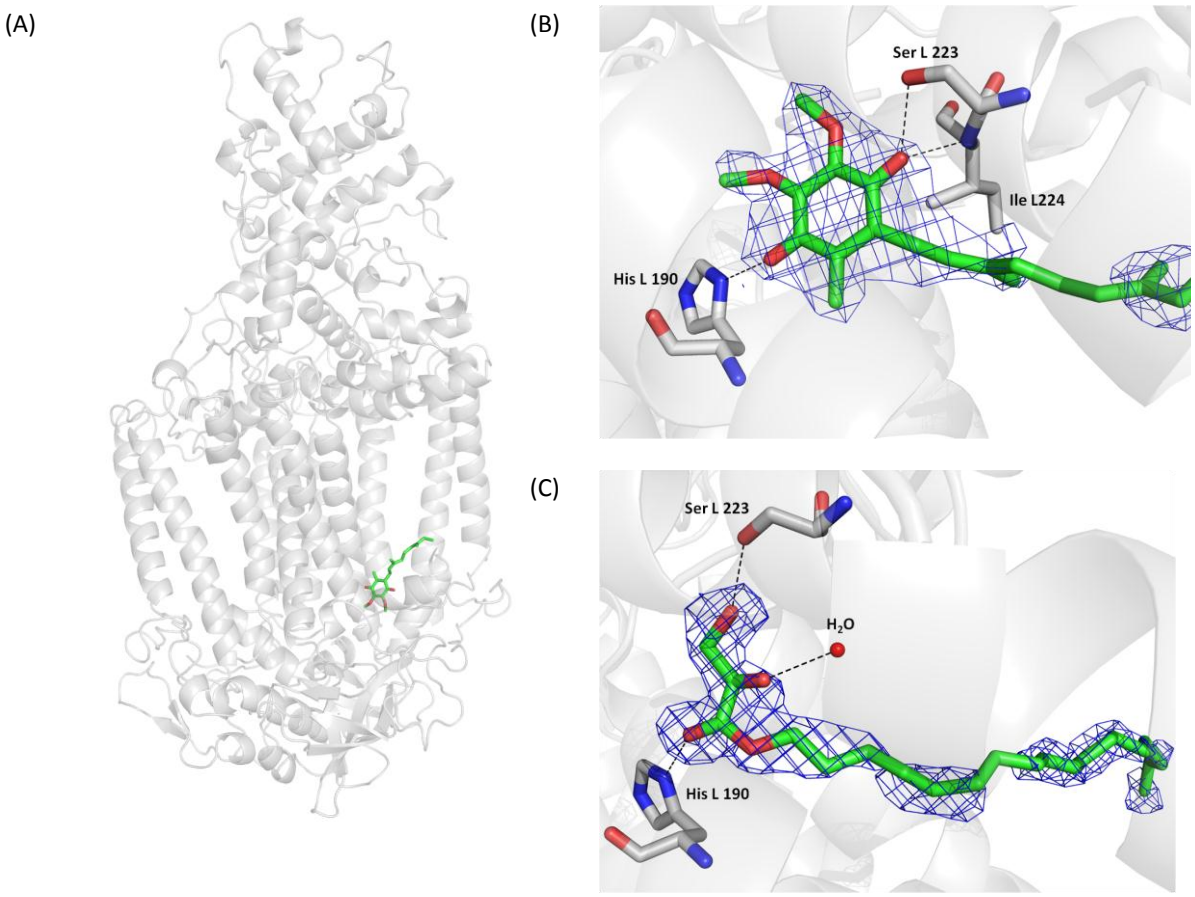

Figure 14. The Q_B binding site in RC_{vir}. An overview of RC_{vir} (gray) with a UQ_2 molecule (green) at the Q_B site (A) and zoom in view of the UQ_2 molecule located at its proximal position at the Q_B bind pocket (B). A monoolein molecule was found bound in the Q_B binding pocket upon co-crystallization attempt with low concentration of terbutryn (C). The $2F_o$-F_c map is contoured at 1 σ.

A putative second ubiquinone binding site (Figure 15) has been observed in the UQ_2 co-crystallized structure of RC_{vir}. This site is situated between the M and L subunits close to the membrane and soluble periplasmic domain. Published co-crystallized structures (PDB entries 3G7F, 2I5N, 1VRN, 5PRC) of RC_{vir} have electron density corresponding to the ligand at this position. Some of them have modeled a ubiquinone molecule at this position [78]. UQ_2 produced by the addition of bromine or mercury atoms across unsaturated double bonds in the fatty acid tails can be used to evaluate the identification of the ubiquinone molecule at this putative second binding site [79, 80].

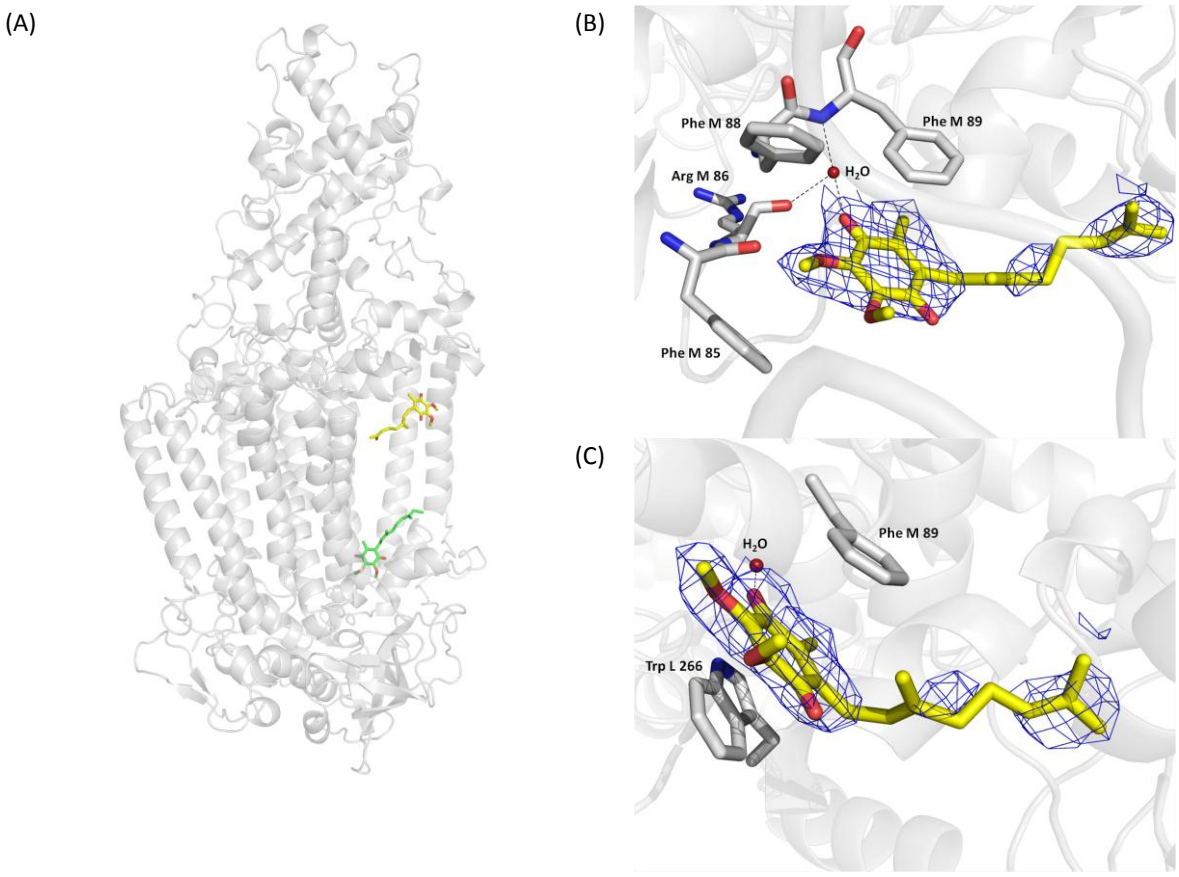

Figure 15. The putative second Q_B binding site in RC_{vir}. An overview (A) of RC_{vir} with UQ_2 molecule at the Q_B site (green) respective the putative binding site (yellow). Zoom in view of the UQ_2 molecule at the putative site (B and C). The 2Fo-Fc map is contoured at 1 σ.

4.2 Crystallization of reaction center - light harvesting 1 complex from *Blastochloris viridis*

Purple photosynthetic bacteria have proven to be excellent model systems to study the light-induced reactions of photosynthesis. The reactions take place in the photosynthetic membrane where the reaction center is surrounded by two types of antenna complexes, named LH1 and LH2 [81]. Both light harvesting complexes show a similar cyclic structure composed of multiple copies of hydrophobic apoproteins α and β, each of which has a single transmembrane (TH) helix. These α and β transmembrane helices are packed side-by-side forming the inner and outer rings respectively [82]. Two types of light-capture-co-factors, bacteriochlorophyll and carotenoid molecules, are arranged between these two rings. The LH1 antenna forms a stoichiometric complex with the RC, the RC-LH1 core complex. The photons from the solar energy trapped by the LH1 complex are further transferred to the special pair of the RC and the electron transfer process begins.

The nonameric structure of the LH2 complex from the purple bacterium *Rhodopseudomonas acidophila* was determined first time in 1995 and now has the highest resolution at 2 Å [83, 84]. A 4.8 Å RC-LH1 crystal structure from *Rhodopseudomonas palustris* (*R. palustris*), the only X-ray structure of RC-LH1 from purple bacteria, indicated that RC is enclosed by an oval LH1 ring consisting of 15 pairs of α and β apoproteins. An additional single transmembrane helix, 70-amino acid PufX, breaks the LH1 ring and is located next to the binding site of the mobile electron

acceptor ubiquinone in the reaction center. It was proposed that PufX may induce a specific orientation of the RC inside the LH1 and the break of the LH1 ring may provide a portal through which UQ_B can transfer electrons to cytochrome bc_1 complex [85]. Negative-stained native membranes of *Rhodobacter sphaeroides* (*R. sphaeroides)* were examined by electron-microscopy, and it is generally accepted that the RC-LH1 complex from *R. sphaeroides* is made up of a dimeric RC-LH1-PufX complex, an elongated s-shaped structure composed of two c-shaped LH1 structures facing each other with the open side and partially encircle one reaction center each [86]. Presumably, the open assembly allows the free movement of quinone /quinol between RC and cytochrome bc_1 complex (Figure 12). Nevertheless, electron-microscopy and atomic force microscopy (AFM) studies of complexes from some other photosynthetic bacteria have shown that LH1 complex forms a complete ring around the RC [87-92]. RC-LH1 from *Rhodospirillum rubrum* (*R. rubrum*) was detergent-solubilized, purified and crystallized in 2D crystals and projection structure was calculated with electron cryo-microscopy. No evidence for a 'quinone' channel or fixed opening in the LH1 ring, and no equivalent of PufX or any other protein component in the antenna have been found. However, a debate emerged whether RC can adopt four possible orientations within the LH1 ring in *R. rubrum* [88, 90].

In *B. viridis*, photosynthetic membranes are organized in a lamellar arrangement and are connected to the cell membrane by relatively small tunnel-like structures which can be characterized by a similar two-dimensional hexagonal lattice organization [93]. The photosynthetic apparatus of *B. viridis* has only one type of antenna complex and LH2 is lacking, which makes this photosynthetic apparatus particularly simple. The LH1 complex of *B. viridis* ($LH1_{vir}$) consists of oligomers of three types of polypeptides, α, β and γ, each of which has a single TM helix. So each subunit of $LH1_{vir}$ constitutes three TM helices. High resolution AFM of the native *B. viridis* membrane shows a closed ellipsoid of 16 $LH1_{vir}$ subunits which surround a single RC_{vir} which exhibits an asymmetric topography and intimately associates to the LH1 at the short ellipsis axis. Same as in *R. palustris*, the $LH1_{vir}$ subunits can rearrange into a circle after removal of the RC_{vir} from the core complex, providing evidence that ellipticity of the LH1 is RC-induced [91, 94]. It was proposed that the LH1 ellipticity may represent a strategy which combines broadening light absorption and more efficient energy transfer by bringing some of the bacteriochlorophylls at the antenna ellipsis closer to the special pair of RC. Meanwhile, compared to a circular LH1 complex the non-symmetric RC/LH1 interaction in an elliptical ring may favor quinone/quinol passage with a 'closed' LH1 complex [92].

Crystals of RC-$LH1_{vir}$ were produced by the detergent-based crystallization method and showed limited diffraction to 8 Å resolution [95]. The self-rotation function was calculated, indicating that the $LH1_{vir}$ is distributed with quasi-16-fold rotational symmetry around the RC_{vir} [95]. So far, the three-dimensional structure at atomic resolution of this complex is still absent. Whether the LH1 ring is complete in the presence of RC and how many orientations RC does pose in the center of LH1 are just two of many questions. Detailed structural information about how quinone/quinol molecules mobilize between the Q_B binding site of RC and cytochrome bc_1 complex or quinone/quinol pool outside the LH1 remains a mystery. An efficient reaction process requires high structural organization of the co-factors, residues from surrounding LH1 and RC and solvent / water molecules. The light-harvesting 1 complex surrounding the reaction center in *B. viridis* works as a light-energy collecting device; hence a three dimension structure of the RC-$LH1_{vir}$ core complex at higher resolution is desirable.

4.2.1 Purification of RC-LH1$_{vir}$ core complex

Cells of *B. viridis* strain ATCC 19567 were grown semi-anaerobically. The purification protocol of RC-LH1$_{vir}$ was adapted from S. Saijo et al., with certain modifications [95]. Membranes were obtained by sonication. After removing unbroken cells by centrifugation, membranes were pelleted and resuspended in 20mM Tris-HCl pH 8.5. Membranes were solubilized by addition of 5 % (w/v) CHAPS and the supernatant was loaded on a DEAE ion-exchange column pre-equilibrated with buffer A of 10mM Tris-HCl pH 8, 5% glycerol, 0.5% CHAPS. RC-LH1$_{vir}$ was eluted using a linear NaCl gradient from 0 to 0.5 M in buffer A. A final step of size-exclusion chromatography (Sephacryl S-400) was carried out in buffer A and RC-LH1$_{vir}$ fractions with ratio $A_{280}/A_{1012} < 0.76$ were pooled and concentrated by Vivaspin centrifugal concentrator with a 100-kDa molecular mass cut-off.

4.2.2 Lipidic-sponge phase screen

To simplify the application of lipidic-sponge phase as part of the standard search-for-diffracting crystal process applied to every membrane protein, a lipidic-sponge phase screen consisting of 48 different conditions has been developed and validated in our lab. The main concept of the screen is based on four sponge phase causing solvents, PEG 400, PEG 1500, PEG 4000 and jeffamine M600. The first three have proven to be successful reagents for membrane protein crystallization using detergent-based crystallization method and the fourth solvent, jeffamine M600, has been successfully used for crystallization of both RC$_{sph}$ and RC$_{vir}$ from LSP [96, 97]. A bacterial light harvesting complex II using LCPs swollen by the addition of 20% pentaerythritol propoxylate (PPO) to form a LSP has also been reported [20]. Similarly, a LCP crystal-form of the bacterial cobalamin transporter BtuB was improved by swelling the LCP to LSP by the addition of 10–12% 2-methyl-2,4-pentanediol (MPD) during crystallization [98, 99]. However, it appeared that sponge phases produced by PPO and MPD were not easy to reproduce, especially with MPD, which shows a very narrow L_3 phase area in the MO-water-MPD phase diagram. Thus, they were not included in this developed lipidic-sponge phase screen. Moreover, three different buffers MES, HEPES and Tris, and four different salts, NaCl, $MgCl_2$, Li_2SO_4 and $(NH_4)_2SO_4$, have been chosen for the screen. The three buffers span the pH range between 5.5 and 9.0 which can match requirements of different membrane proteins. And the four salts have proven to help liquefy the L_3 phase, thus making sponge phase easier to handle.

To prepare the PEG-containing sponge phases, MO powder was first heated to 50 °C and becomes liquid-like. With a pipette, appropriate amounts of MO can be measured and transferred into small glass vials along with PEG. The glass vials were sealed and incubated for 24 hours at 37 °C for homogenization. For the jeffamine-containing sponge phase, it was prepared as described in 4.2.1.

While purified detergent-solubilized membrane proteins are still required as a starting point for lipidic-phase crystallization, lipids which are important for stability and activity of membrane proteins may be removed from the protein surface during purification process. However, with both detergent-based and lipidic-cubic phase crystallization method it has been reported that membrane proteins were crystallized by adding specific lipids which are essential for their stability and function in the crystallization step [100-103]. Cholesterol is a lipid additive which can be easily incorporated into lipidic-sponge phase up to 5 % (w/w). Crystallization experiments of RC-LH1$_{vir}$ using a cholesterol-doped variant of the lipidic-sponge phase screen have also been performed.

To validate the screen, eleven different membrane proteins were tested. For eight out of the eleven membrane proteins initial crystal leads were observed, including RC-LH1$_{vir}$ which was obtained from the cholesterol-doped variant of the lipidic-sponge phase.

4.2.3 Further optimization of lipidic-sponge phase crystallization method (unpublished results)

The condition of the initial crystal lead obtained from cholesterol-doped lipidic-sponge phase was further optimized. Crystals with suitable size were screened for diffraction. Hundreds of crystals were screened and datasets from 10 crystals, obtained from various crystallization drops and from different purification batches, were collected and analyzed. Solutions from molecule replacement identified that they were crystals of RC$_{vir}$, thus with altered space group and cell dimensions compared with previous obtained results (Table 2). It appeared that only RC$_{vir}$ was crystallized from the original RC-LH1$_{vir}$ crystallization drops, although the purity of RC-LH1$_{vir}$ just before the crystallization setup was confirmed [95].

Table 2. Different crystal structure statistic table.

	RC 1	RC 2
Space group	C2	P2
Cell dimensions		
a, b, c (Å)	120.0, 85.0, 180.4	57.2, 82.0, 173.0
α, β, γ (°)	90.0, 96.4, 90.0	90.0, 93.5, 90.0
Resolution (Å)[a]	46.7 - 2.80 (2.87 – 2.80)	49.0 - 4.20 (4.31 – 4.20)
Total reflections[a]	83888 (6192)	67423 (5145)
Unique reflections[a]	37074 (2813)	11855 (890)
Completeness (%)[a]	82.9 (85.7)	99.6 (100.0)
R_{sym} (%)[a,b]	12.4 (89.2)	41.3 (72.4)
<I/σ>[a]	8.34 (1.18)	5.42 (2.70)

[a] Values in parentheses indicate statistics for the highest resolution shell.
[b] $R_{sym} = \sum |Io - \langle I \rangle| / \sum Io \times 100\%$, where Io is the observed intensity of a reflection and <I> is the average intensity obtained from multiple observations of symmetry related reflections.

One hypothesis was that a small portion of RC-LH1$_{vir}$ protein molecules lost the LH1 during the purification process and those LH1-less RC$_{vir}$ could not be completely separated by size-exclusion chromatography from the intact RC-LH1$_{vir}$. Therefore crystallization drops might contain a mixture of RC$_{vir}$ and RC-LH1$_{vir}$ proteins. Since it is easier to crystallize RC$_{vir}$, the majority of the crystals formed in the drops were in fact RC$_{vir}$. Additional purification steps have been attempted, including hydrophobic interaction chromatography, sucrose density centrifugation, attempting to dispose of the LH1-less RC$_{vir}$, thus improving the purity of RC-LH1$_{vir}$. These efforts gave neither any higher purity sample, nor any RC-LH1$_{vir}$ crystals.

To find a new RC-LH1$_{vir}$ crystal hit, a modified form of sponge phase was developed and validated by using RC$_{vir}$ as a test membrane protein. Instead of using the routine sponge-phase-inducing solution 16 % jeffamine M-600, 1M HEPES pH 7.9, 0.7M ammonium sulphate, 2.5 % 1,2,3-Heptanetriol which has been successfully applied to both RC$_{sph}$ and RC$_{vir}$, solution containing 16 % jeffamine M-600, 1M HEPES pH 7.9 was used to prepare the modified sponge phase while both ammonium sulphate and 1,2,3-Heptanetriol were excluded.

LCP containing RC_{vir} was then transferred into a glass vial and solution containing 16 % jeffamine M-600, 1M HEPES pH 7.9 was added (1:4 ratio). HEPES must necessarily be at a concentration of at least 0.8 M in order to create the sponge phase. After phase separation overnight, the upper brown-colored sponge phase containing RC_{vir} was harvested. Sitting drop experiments with 288 different reagent solutions from commercially available protein crystallization screens were performed using a Cartesian honeybee robot (Genomic Solutions Ltd). 500 nl of sponge phase (with RC_{vir} incorporated) were added on top of 500 nl of reagent solution in 96-well plates. The robot plates were stored at 20 °C in the dark. After two weeks, RC_{vir} crystals were obtained from a completely new condition 0.1 M Na-acetate pH 4.5, 0.1 M NaCl, 0.1 M MgCl, 30% PEG 400. The initial RC_{vir} crystal from robot plate diffracted to 5 Å and had space group P1 with type I crystal packing (Table 3).

Table 3. Cell parameters for the two types of space groups of RC_{vir}.

	RC_{vir} from sponge phase	RC_{vir} from modified sponge phase
Space group	$P2_12_12$	P1
Cell dimensions		
a, b, c (Å)	84.6, 139.7, 179.2	112.6, 98.7, 239.8
α, β, γ (°)	90, 90, 90	83.8, 78.5, 100.3

After the successful experiment with RC_{vir}, RC-$LH1_{vir}$ was incorporated into the modified sponge phase as described above. After overnight incubation, a beautiful transparent, green-colored sponge phase formed. A UV-VIS absorption spectrum was taken on the RC-$LH1_{vir}$-sponge phase (Figure 16). Unfortunately, the characteristic absorption peak at A_{1012} disappeared in the spectrum. Instead, a peak at 830 nm representing isolated RC_{vir} appeared. In the lipidic-sponge phase crystallization of light harvesting complex 2 (LH2) article, absorption spectrum was taken on LH2 incorporated in cubic phase [20]. It showed a similar spectrum with minor differences compared with LH2 in detergent solution. It was proposed that due to the large size of LH2 complex, reconstitution of the protein in the MO bilayers introduces strain into the bilayers and locally distorts the cubic phase. Meanwhile, it likely also strains the LH2 complex, as the minor differences showed in the UV-VIS absorption spectrum. The conclusion drawn from our UV-VIS spectra suggests that since RC-$LH1_{vir}$ is an even larger complex, the strains on the protein complex may become too strong so that the surrounding $LH1_{vir}$ complex ring is disrupted, leaving the subunits of RC_{vir} still intact and incorporated in the bilayers. This could also explain the previous RC_{vir} crystals produced from the RC-$LH1_{vir}$ sponge phase crystallization setup.

Moreover, PPO- and MPD-induced sponge phases which are not included in the developed lipidic-sponge phase screen have been applied on RC-$LH1_{vir}$ in a similar approach as the jeffamine-induced sponge phase. UV-VIS spectra showed similar results. The traditional detergent-based crystallization method was also applied on RC-$LH1_{vir}$ complex without any success. Obviously, a different approach was needed.

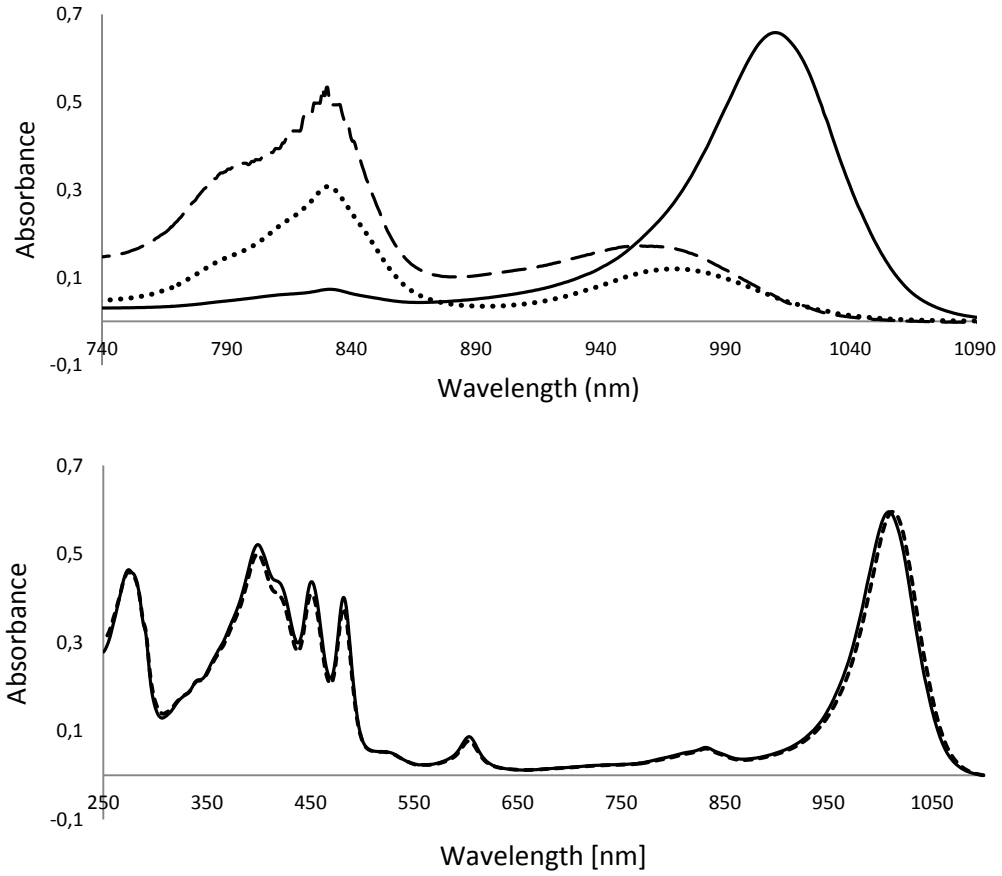

Figure 16. UV-VIS with RC-LH1$_{vir}$ incorporated in sponge phase (up): solid line shows the RC-LH1$_{vir}$ in micelle solution, dashed line shows RC-LH1 incorporated in sponge phase and dotted line shows RC in micelle solution. Purified RC-LH1$_{vir}$ incorporated in bicelle (down): solid line shows the RC-LH1$_{vir}$ in micelle solution and dotted line shows RC-LH1$_{vir}$ incorporated in bicelle.

4.2.4 Development of bicelle crystallization method (unpublished results)

Bicelles with lipids DMPC, DMPG and detergent CHAPS, CHAPSO, in all combinations, and with different lipid/detergent ratios were produced according to the published protocol [28]. The first step of producing bicelles is mixing the detergent with filtered MQ-water. Subsequently the detergent solution is mixed with the lipid and the bicelle mixture is subjected to numbers of cycles of vortexing, cooling, vortexing, heating and centrifugation. The resulting bicelle solution was added to the purified RC-LH1$_{vir}$ (10 – 15 mg/ml) in a series of different bicelle/protein ratio and mixed through pipetting. The bicelle-protein mixture was incubated on ice for 30 minutes prior ultracentrifugation to discard excess of lipid and protein precipitation. The whole procedure was performed on ice. The supernatant was subsequently used for UV-VIS absorption measurement and eventually for crystallization robot screening with commercially available precipitant solutions. RC-LH1$_{vir}$ has been shown to be fully functional in bicelles according to the UV-VIS absorption spectrum where the characteristic absorption peak at A_{1012} is still there, though red-shifted from the reconstituted bicelle sample (Figure 16). Such red-shift in UV-VIS absorption spectra could be seen from the isolated *B. viridis* membrane, compared with detergent solubilized RC-LH1$_{vir}$. The crystallization plates were incubated in darkness at 20 °C.

Crystal leads were discovered from bicelles (9 % w/v lipid) of DMPG/CHAPS with a protein/bicelle ratio of 3:1, and DMPG/CHAPSO with a protein/bicelle ratio of 2:1. After additive screens and further optimization, dark green colored crystals with a size of 0.5 mm x 0.3 mm appeared within 2 weeks. Crystals were fished with and without cryo-protectants before flash-freezing in liquid nitrogen for X-ray diffraction experiments. However, most of the crystals diffracted to 20 Å and the best diffraction was beyond 14 Å resolution. The crystal which diffracted to 14 Å was washed four times in crystallization precipitant solution and dissolved in MQ-water. UV-VIS spectrum was taken and SDS-PAGE was performed with the dissolved sample, and both confirmed that it contained RC-LH1$_{vir}$ (Figure 17).

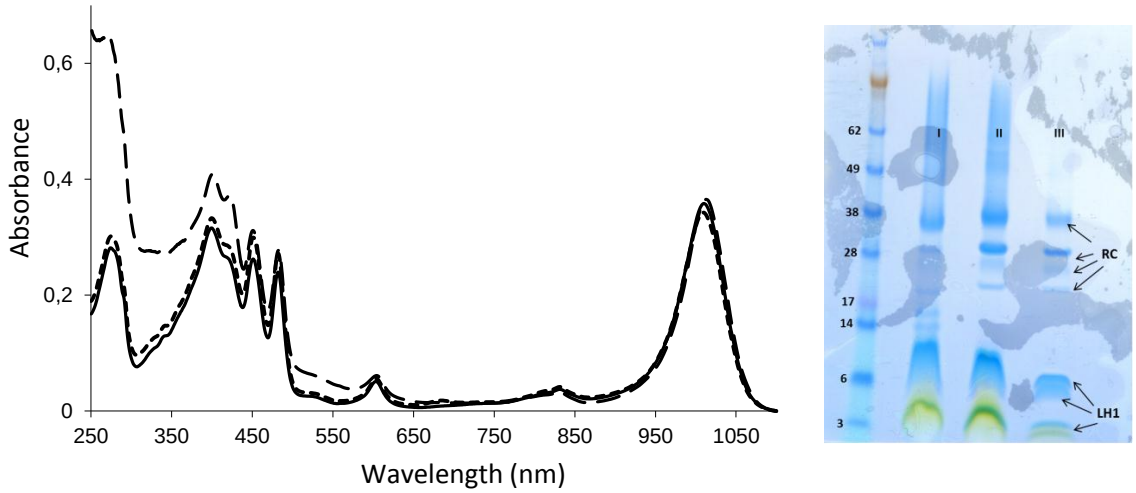

Figure 17. Solid line shows RC-LH1 in micelle solution. dotted line shows dissolved crystal from bicelle set up which diffracted to 17 Å, dashed line shows isolated membrane of *B. viridis* (left) and SDS-PAGE of dissolved crystals (right), line III was the sample of dissolved crystal which diffracted to 17 Å.

To improve diffraction quality, post-crystallization treatments have proven to be important for many crystal structure determinations of very large macromolecules [104-106]. Crystal dehydration with higher PEG concentrations, flash annealing and annealing on the loop treatments have been applied on RC-LH1$_{vir}$ crystals without any improvement on their diffraction qualities [107]. Thus, dehydration of RC-LH1$_{vir}$ crystals need to be optimized further in a controlled way such as by using a humidity control device [108].

One of the additives which improved the crystal diffraction quality was DDM. It was suggested that RC-LH1 complex from *R. rubrum* S1 was only stable in DDM [109]. However, the stability of RC-LH1$_{vir}$ in 19 different detergents was examined with time-dependence of UV-VIS absorption spectrum and it was suggested that RC-LH1$_{vir}$ was most stable in CHAPS [110]. Even though among the 19 detergents DDM, the most used detergent in membrane protein field, was not included in the study. Bicelles of DMPG/DDM and DMPC/DDM with different lipid/detergent ratio were developed with the aim of searching better diffracting RC-LH1$_{vir}$ crystals. Bicelle (16.7 % w/v lipid) of DMPG/DDM with a ratio of 16.7:20 has been proven to produce initial diffracting crystals of RC-LH1$_{vir}$ from robot plates to 7.6 Å resolution (Figure 18). Thus, optimization around the promising crystallization conditions, additive screens and post-crystallization methods are essential in order to improve the resolution further.

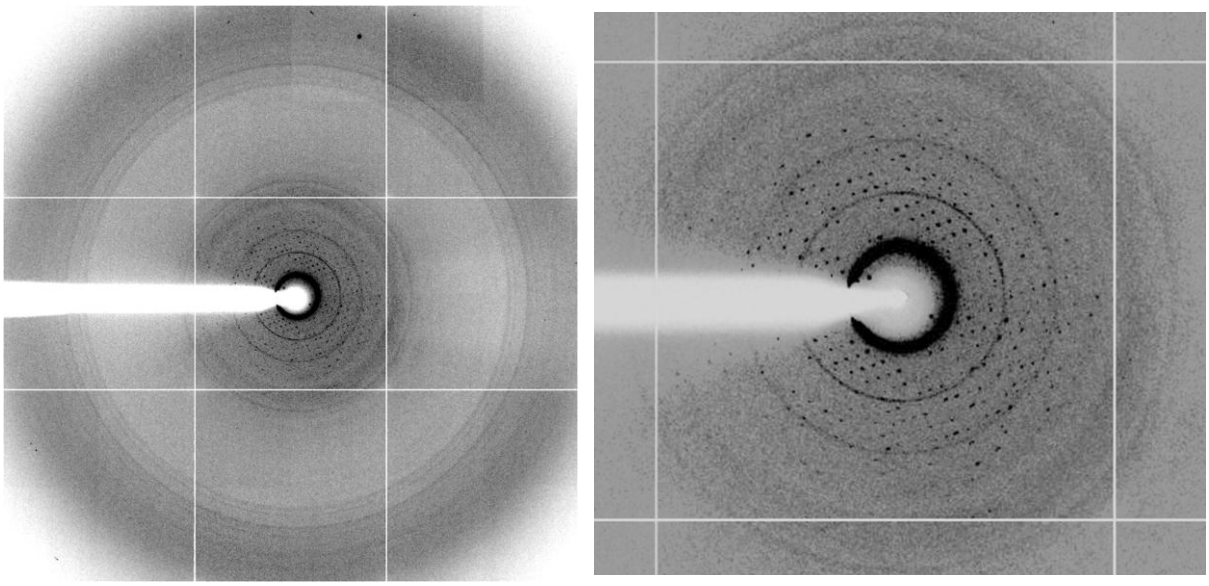

Figure 18. Diffraction pattern of RC-LH1$_{vir}$ crystal from DDM-DMPG bicelle (0.1M NaCl, 0.1M MgCl, 0.1M NaHEPES, pH 7.5, 12% PEG4000) and zoom in picture of the diffraction pattern (right).

4.2.5 Serial nano-crystallography with X-ray free-electron laser sources (unpublished results)

The integrated intensity of an X-ray diffraction peak from a crystal is proportional to the ratio of its diffracting volume to its unit cell volume. Therefore, the size of the crystals used for X-ray crystallography is of particular importance for large membrane protein complexes which often have a large unit cell volume. However, the larger and more complicated the protein structure becomes, the more challenging it is to isolate the protein complex intact and to grow large, well-diffracting crystals. Furthermore, X-ray damage to macromolecular crystals has become one of the major limitations for X-ray structure determination, especially since the advent of third generation synchrotron undulator sources [111]. Finding appropriate cryo-protectant conditions to minimize the radiation damage during data collection can be a challenge especially as to membrane protein crystals. However, even at cryogenic temperature (~100 K), radiation damage may still limit the final resolution of the structure [112].

In 2000, R. Neutze et al. predicted that structural information can be recovered from the scattering of intense femtosecond X-ray pulses by single protein molecule and small assemblies [113]. Ultra-short, high-intensity coherent X-ray pulses from free-electron lasers have an intensity more than a billion times of conventional synchrotron sources [114]. Experiments using such high X-ray dose rates and ultra-short exposures may capture useful structural information before radiation damage destroys the sample. Thus, it provides the possibility to achieve high-resolution structural information with X-rays, without the need for macroscopic crystals [113]. In the process of crystallization, many macromolecules, especially membrane proteins, do not form large crystals. However, they often form sub-micron crystals which are usually too small to generate useful diffraction data at a conventional synchrotron source. With this 'probe before destroy' method, structures could be determined from sub-micron crystals and nano-crystalline materials.

By combining serial nano-crystallography with X-ray free-electron laser sources, sub-micro and nano-crystals which contain a few hundred or thousand protein molecules of large membrane protein complexes, such as photosystem I and II, have been shown to produce useful molecular resolution electron density maps [115-118]. Likewise, interpretable diffraction data from micrometer-sized lipidic-sponge phase crystals of RC_{vir} have been successfully collected with an X-FEL beam using a sponge phase micro-jet [119]. Moreover, it not only solves the problem caused by radiation damage but also opens a window to apply time-resolved experiments on enzyme systems at room temperature. Using a variable delay trigger, such as photo-activation, it becomes possible to film dynamic structural changes [118].

To facilitate the delivery of a fully hydrated stream of nano-crystals into an X-ray X-FEL beam, the promising conditions of RC-$LH1_{vir}$ from DMPG/DDM bicelle have been experimented on using a large volume of batch crystallization setup. Bicelle composition of DMPG/DDM (8 % w/v lipid) with a ratio of 16.7:40 and a protein/bicelle ratio of 3:1 was used. Two volumes of protein-bicelle solutions were mixed with 1 volume of precipitant solution (0.1 M NaCl, 0.1 M $MgCl_2$, 0.1 M Na-HEPES pH 7.5, 12 % - 20 % PEG4000) and 0.5 volume of ethylene glycol. The batch setups were incubated in darkness at 20 °C for 5 weeks. X-ray diffraction pattern of RC-$LH1_{vir}$ was collected at the Coherent X-ray Imaging (CXI) instrument at the LCLS [120]. RC-$LH1_{vir}$ nano-crystals were delivered to the injector nozzle via a sample loop and injected into the LCLS X-ray beam at a flow rate of 15 µl min^{-1}. The liquid capillary of the nozzle had an inner diameter of 100 µm and the liquid was focused by coaxially flowing helium gas [121]. A diffraction spot up to 25 Å resolution was found from RC-$LH1_{vir}$ nano-crystal samples prepared from bicelle crystallization method (Figure 19).

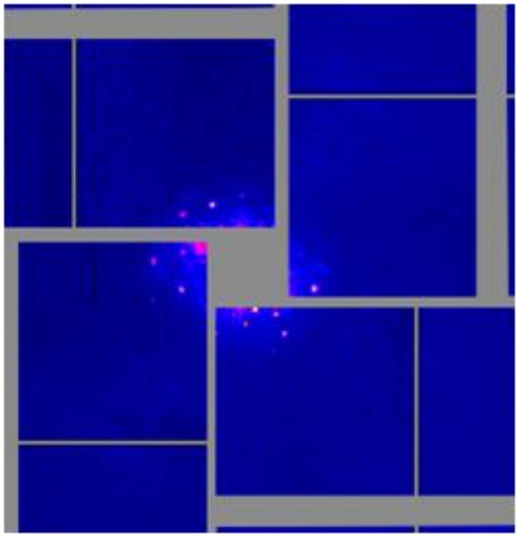

Figure 19. Diffraction pattern from the sample with presumed nano/micro-crystals of RC-$LH1_{vir}$ at the LCLS. Batch crystallization set up of RC-$LH1_{vir}$ in DMPG-DDM bicelle (0.1M NaCl, 0.1M MgCl, 0.1M NaHEPES, pH 7.5, 15% PEG4000).

4.2.6 Future perspective

Complexes formed by more than one component or subunit often display several binding alternatives with enhanced structural dynamic, which increases disorder within the unit cell of the crystals. For multi-subunit membrane proteins, native lipids have been repeatedly captured in crystal structures demonstrating their importance in protein conformational stability [122-125]. These native lipids survive processes like detergent membrane solubilization and protein purification steps, partially due to their strong interaction with the membrane proteins and possibly also owing to a 'mild' membrane protein preparation protocol used. An altered RC-LH1$_{vir}$ purification protocol can be tested using DDM in the solubilization and purification processes with shorter incubation time during the solubilization step.

Another way to minimize the structural dynamics of a membrane protein complex might be exogenously adding lipids back to the purified proteins. Cardiolipin has been suggested to be important for the thermal stability of reaction center from *R. spheroides*, but it has never been found in the membrane of *B. viridis* [122]. In all published crystal structure obtained from bicelle-like crystallization method, lipids with a neutral head group like DMPC and DOPC were used [28, 32-35, 126]. Two lipid components of bicelles, DMPC, DMPG and a mixture of both, have been experimented with in crystallization of RC-LH1$_{vir}$. Surprisingly, crystal leads were detected only from bicelles solely containing DMPG which has a negatively charged head group. Lipid components of membranes from *B. viridis* can be analyzed using, for instance, mass spectrometry. Subsequently, the 'key lipids' can be screened by a RC-LH1$_{vir}$ thermal stability test. Once the putative 'key lipids' are identified, they can be added to the protein environment throughout the purification and crystallization trials.

5 Structure of bacterioferritin from *Blastochloris viridis* (paper IV)

Bacterioferritin (Bfr) belongs to the iron storage ferritin superfamily. It differs from other ferritins by its ability to bind heme cofactors. Bfrs not only provide an accessible storage of iron in living cells, but also reduce the concentration of toxic free Fe(II) to mineralized Fe(III). Thus, it has an 'anti-oxidating' function when ferrous concentration becomes too high in a bacterial cell. The structures of Bfrs are highly conserved and are generally comprised of 24 identical subunits with two-, three- and four-fold symmetry forming a hollow, roughly spherical protein shell and the ferric irons are stored within the cavity (Figure 20 A). It is capable of accommodating up to 4500 iron atoms as an inorganic complex core. The inner core of Bfr is linked to the protein exterior by eight 3-fold pores and six 4-fold pores (Figure 21). Each subunit (~20 kDa) folds as a four α-helix bundle, with a heme located between two subunits whereby a methionine from each monomer provides an axial ligand to the heme iron (Figure 20 B). In each subunit there is an evolutionary well conserved ferroxidase site (Figure 20 C). It is capable of binding two irons and is therefore also referred to as the di-iron center. The ferroxidase active site consists of six highly conserved amino acids coordinating possibly two irons: His54 and Glu18 are terminal ligands to iron 1 (Fe_1), His130 and Glu94 are terminal ligands to iron 2 (Fe_2), and Glu127 and Glu51 are bridging ligands.

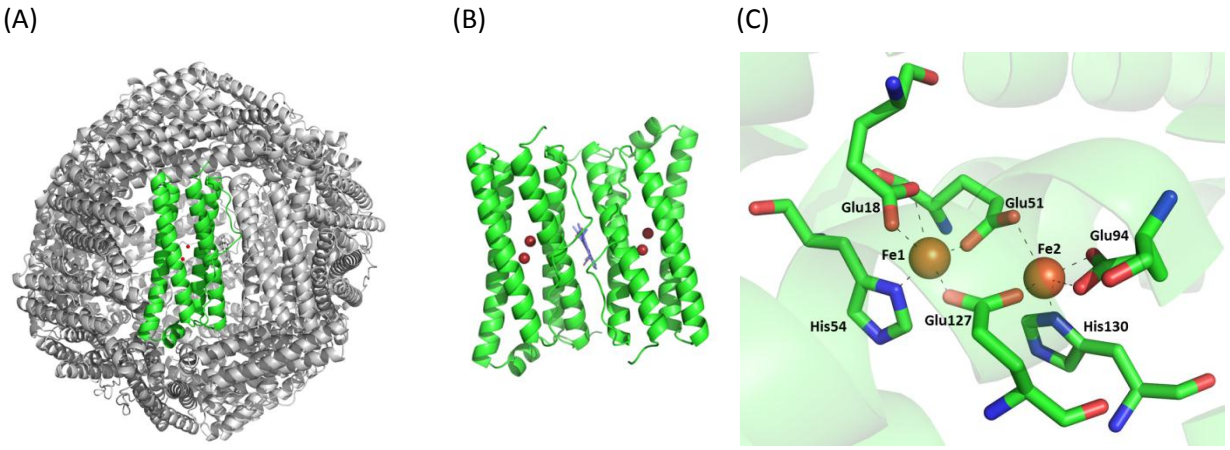

Figure 20. Bacterioferritin from *Blastochloris viridis*. Overview of the *Bv* Bfr with 12 dimers (A). Zoom in view of two subunits forming one homodimer (B). The location of the heme-b on the 2-fold axis is shown. The red spheres represent Fe ions located in the ferroxidase site of each monomer. Ferroxidase center with highly conserved residues binding two iron ions (C).

In Bfrs, the exact role of the ferroxidase site is unclear. In some Bfrs, the ferroxidase site has been suggested to function as a gated site for the transfer of iron ions into the cavity after the oxidation of ferrous to ferric iron [127, 128]. Nevertheless, for Bfr from *Escherichia coli* (*Ec* Bfr) a third non-heme iron binding site on the inner surface of the protein shell of each subunit, lying ~ 10 Å below the ferroxidase site, was discovered. It was proposed that iron ions enter the cavity through pores other than the ferroxidase pore and get oxidized at the inner shell of the central cavity. Electrons originating from Fe(II) oxidation are then transferred to the ferroxidase center, via this novel iron binding site. Hence, the ferroxidase site of *Ec* Bfr functions as a true catalytic co-factor [129].

To fulfill the cellular requirements for iron, Bfrs can quickly deposit excess irons within the central cavity and also release them back to cellular environment when it is needed. The inner cavity of a

Bfr is linked to the protein exterior by its 3- and 4-fold pores (Figure 21). In eukaryotic ferritins which lack heme molecules, it was suggested that irons enter and exit the protein via the 3-fold pores [130]. In some crystal structures, localized unfolding of the 3-fold pores was observed when highly conserved residues around the pores that influence pore gating were mutated [131-134]. The same kind of localized unfolding of 3-fold pores was proposed to occur in the presence of urea at physiological concentrations (1-10 mM). Thus, the iron mineral stored in the central core becomes accessible to cellular reductant [132, 135, 136]. Despite this, the exact role of 3- and 4-fold pores in Bfrs remains unclear.

(A) (B)

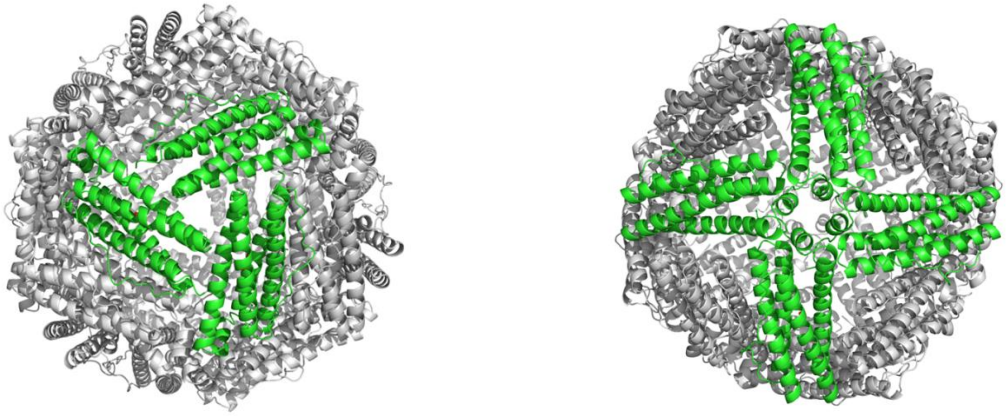

Figure 21. Overview of 3-fold pore (A) and 4-fold pore (B).

At present, eight unique bacterioferritin structures have been reported [128, 137-142]. Among all the Bfr structures, bacterioferritins from purple bacterial are especially well represented. Moreover, three of the eight structures were obtained unintentionally. Bfr appears easily to co-purify with other proteins [137, 142, 143]. In this work, crystallization of bacterioferritin from *B. viridis* (*Bv* Bfr) occurred accidentally while attempting to crystallize the RC-LH1$_{vir}$ from the same organism.

5.1 Discovery and structure determination

The purple bacterium *B. viridis* is an anaerobic photosynthetic bacterium with microaerophilic growth capacity [144]. The strain ATCC 19567 grows in liquid culture at 30 °C anaerobically in light with a generation time of 13 hours and under microaerophilic growth conditions in the dark with a generation time of 24 hours [144]. It was reported that illumination in the presence of oxygen inactivates photosynthetic pigments by photo-oxidation and inhibits growth [144]. In our lab, 10-liters cultivation bottles are regularly used for expression of photosynthetic apparatus from *B. viridis*. The bottle is filled to the top with the medium to get rid of any oxygen. Firstly, the culture is incubated in the dark for 36 hours allowing *B. viridis* to grow under microaerophilic condition and consume the limited oxygen dissolved in the liquid medium. The bottle is then moved to light and its metabolism switches to phototrophic growth and thylakoid membranes form. Mistakingly, one 12-liter bottle was used and the beforehand prepared medium could not fill the bottle leaving a big amount of oxygen in the cultivation environment. As usual, the culture was left in the dark for 36 hours before it was put under the light. After 48 hours illumination, the liquid culture showed a slightly altered green color and had lower cell density compared to cells grown under standard conditions. After cell harvesting, the cell pellet was portioned and stored at -20°C until required.

The first portion of the cell pellet was used initially to purify RC-LH1$_{vir}$. After membrane isolation, *Bv* Bfr and RC-LH1$_{vir}$ were solubilized with CHAPS, although bacterioferritin is a soluble protein. Both proteins were able to bind to DEAE, an ion-exchange column used to purify RC-LH1$_{vir}$, and eluted with a linear NaCl gradient. A final step of size-exclusion chromatography (Sephacryl S-400) failed to separate these two proteins due to their similarity in molecular size: *Bv* Bfr has a molecular weight of 480 kDa, while RC-LH1$_{vir}$ has a weight of 430 kDa plus the detergent molecules around. Purified and concentrated *Bv* Bfr is expected to be bright red colored, but owing to the dark green color of RC-LH1$_{vir}$, it was not possible to observe the contamination of *Bv* Bfr by eye. However, some kind of contamination was evident in the UV-VIS spectra because of overlapping absorption spectra from both RC-LH1$_{vir}$ and *Bv* Bfr (Figure 22). Ratio of A_{280}/A_{1012}, the purity indicator of RC-LH1$_{vir}$, was 0.83 instead of usual value of 0.76, although the ratio between absorption peaks corresponding to RC$_{vir}$ and LH1$_{vir}$ was as usual. Therefore the contaminant could not be RC$_{vir}$. In crystallization experiments, contaminants may not always influence the crystallization process, therefore the sample was tentatively used to set up robot plates with detergent-based crystallization method. In a number of crystallization drops red, cubic-shaped tiny crystals had grown after two weeks (Figure 22). Crystals were flash frozen in liquid nitrogen for diffraction without any cryo-protectant solution. The crystals diffracted to 2.8 Å and the first diffraction data set was collected at 100 K at beam line I02 at Diamond Light Source. Owing to the unique cubic space group of almost all the ferritin-family proteins, the identity of this unknown structure could be discovered easily, by searching at the Protein Data Bank. The structure was solved by molecular replacement using the Bfr model (PDB id: 3GVY) from *Rhodobacter sphaeroides* (*Rs* Bfr) which was also accidentally crystallized while attempting to crystallize cytochrome cbb$_3$ cytochrome oxidase [145].

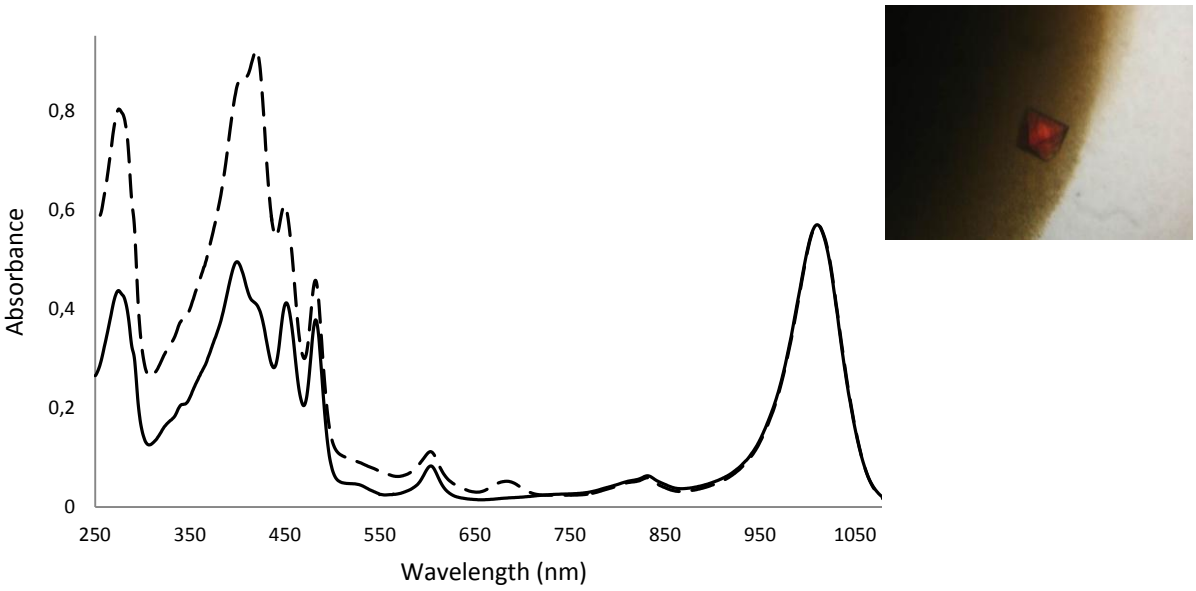

Figure 22. UV-VIS spectrum of purified RC-LH1 complex in detergent solution contaminated with *Bv* Bfr (left). Solid line shows purified RC-LH1$_{vir}$ complex in detergent solution without contamination ($A_{280/1012}$ = 0.76) and dashed line shows RC-LH1 complex contaminated with *Bv* Bfr ($A_{280/1012}$ = 1.43). Red-colored cubic *Bv* Bfr crystal with green-brown colored RC-LH1$_{vir}$ precipitant in the background (right).

After the identification of *Bv* Bfr, another portion of cell pellet was used to try to isolate 'purified' *Bv* Bfr with a modified RC-LH1$_{vir}$ purification protocol. Much less detergent was used with short

incubation times in the solubilization step to minimize the extraction of RC-LH1$_{vir}$ from the membrane. Fractions containing minimal amounts of RC-LH1$_{vir}$ were chosen from the final size-exclusion chromatography step. Initial robot crystallization hits were systematically screened using a crystallization parameters based on a pH vs precipitant concentration grid. Crystallization drops were setup manually with the intention of obtaining bigger and better diffracting *Bv* Bfr crystals. The second *Bv* Bfr data set was collected at beam line ID29 at ESRF, with a resolution of 1.8 Å. The structure recovered from this data set was termed the "as-isolated" *Bv* Bfr structure.

5.2 DNA and amino acid sequencing

Since the whole genome of *B. viridis* bacterium is not sequenced, the translated protein sequence of *Bv* Bfr was absent. With a near atomic-resolution of 1.8 Å, the identification of the amino acid sequence was derived directly from the electron density. Combining this preliminary estimated amino acid sequence and published codon usage lists of L, H, M and C subunits of RC$_{vir}$, a score system was created based on the degeneracy of each amino acid [146-148]. For instance, methionine and tryptophan have only one possible codon. Therefore they have the highest score, while leucine has six possible codons and four of them dominated usage in RC$_{vir}$ according to the codon usage lists, thus it has the lowest score. PCR efficiency varies depending on numerous of factors, such as the degeneracy and length of the primers, the distance between the forward-reverse primers, and the concentration of primers used in a PCR reaction. Thus, several pairs of primers were designed. Temperature gradient was applied to different pairs of primers to find their optimum annealing temperature. After successful amplification with degenerate PCR, the translated sequence of the PCR product covered 68.5 % of the protein sequence.

To acquire a pure *Bv* Bfr protein sample was difficult since our purification process could not completely separate *Bv* Bfr from RC-LH1$_{vir}$. However, to attempt mass spectrometry *Bv* Bfr crystals were fished from crystallization mother liquid and moved to fresh drops of water and washed 3 times before they were dissolved in water. This *Bv* Bfr protein sample was applied to tandem mass spectrometry to resolve sequences in the remaining regions of the N- and C-terminus. By the combination of electron density analysis, nucleotide sequencing and tandem mass spectrometry the complete amino acid sequence of *Bv* Bfr was recovered.

5.3 Soaking experiments

After the sequence of *Bv* Bfr had been completely recovered, we studied what conformational changes occur in *Bv* Bfr in different environmental conditions.

To investigate the existence of an additional non-heme iron binding site and the occurrence of a localized unfolding at 3-fold pores in this newly discovered *Bv* Bfr, some crystal soaking experiments were designed. More *Bv* Bfr crystals were produced using the purification protocol and crystallization condition described previously. Since *Bv* Bfr was crystallized in the presence of the RC-LH1$_{vir}$ protein complex which precipitated during the crystallization incubation, the precipitates were often attached to the *Bv* Bfr crystals. It was important to remove the RC-LH1$_{vir}$ aggregates from the crystals before they were moved to soaking drops.

Crystal Fe(II)-soaking experiments were performed aerobically by transferring native crystals of *Bv* Bfr into a freshly-prepared Fe(II)-soaking solution containing the reagent solution supplemented with 50 mM $FeSO_4$ and incubated for 15 minutes or one hour before flash-cooling in liquid

nitrogen. During incubation brown-colored precipitate formed in the drops indicating the oxidation of Fe(II) to insoluble Fe(III). Double-soaked crystals were prepared by moving Fe(II)-soaked crystals (one hour) to a freshly-prepared double-soaking solution of the same reagent solution containing 10 mM urea and incubated aerobically for 15 minutes before flash-cooling in liquid nitrogen. The brown insoluble Fe(III) precipitates were avoided when moving the Fe(II)-soaked crystals to the drops of double-soaking solution. Several data set from crystals of each soaking treatment were collected at beam line ID14-4 at ESRF and structures recovered from them are referred to as "Fe(II)-soaked" and "double-soaked" *Bv* Bfr structures with the highest resolution of 1.58 Å and 1.68 Å respectively.

5.4 Ferroxidase site

All three structures have a fully occupied Fe_1 site, thus a varied occupancy presented at the Fe_2 site. At Fe_2 site, the occupancy was 40 % in the "as-isolated" structures and 60 ± 10 % in the "double-soaked" structures while a fully occupied ferroxidase center was present in the "Fe(II)-soaked" structures. Furthermore, in the structure from which the crystals were incubated in the Fe(II)-soaking solution for only 15 minutes, a fully occupied ferroxidase center was observed. This indicates that even some fraction of Fe^{2+} ion in the freshly-prepared $FeSO_4$ solution was oxidized during the one hour incubation and might not be available to the protein. The occupancy of the ferroxidase center was not affected by this point.

In addition, different soaking conditions affected the conformation of coordinating residues at the di-iron center, such as Glu94, Glu127 and Glu51 (Figure 24). Upon double-soaking treatment, His130 rotated away from the Fe_2 site and opens access to the inner core of Bfr through the ferroxidase pore as illustrated on Figure 23 using the Hole2 representation [149].

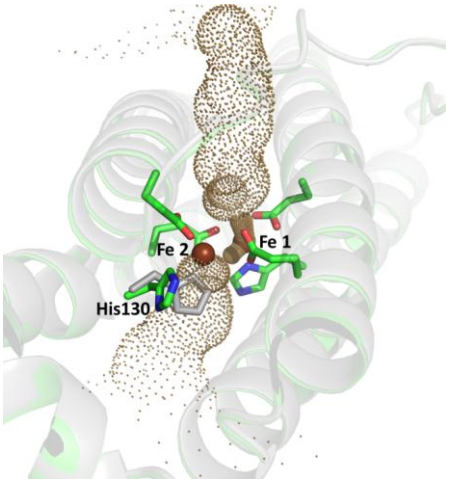

Figure 23. Ferroxidase site in the open and closed state. His130 from 'as-isolated' structure (gray) is rotated toward the ferroxidase center to facilitate binding of Fe_2, while in the double-soaked structure (green) the His130 is in its non-coordinative conformation. Hole2 channel through the ferroxidase site was generated using the double-soaked structure with both Fe_1 and Fe_2 removed.

Additionally, our soaking experiments trapped two different states of the ferroxidase active site. Density representing a small ligand is observed 2.4 Å away from Fe_1 in "Fe(II)-soaked" structure. Together with the coordinating side chains Fe_1 becomes octahedrally coordinated (Figure 24), while in "double-soaked" structure a water molecule was modeled at the density positioned 3.3 Å away from Fe_1. Density functional theory calculations were used to rationalize the character of the closely bound ligand in the Fe(II)-soaked structure.

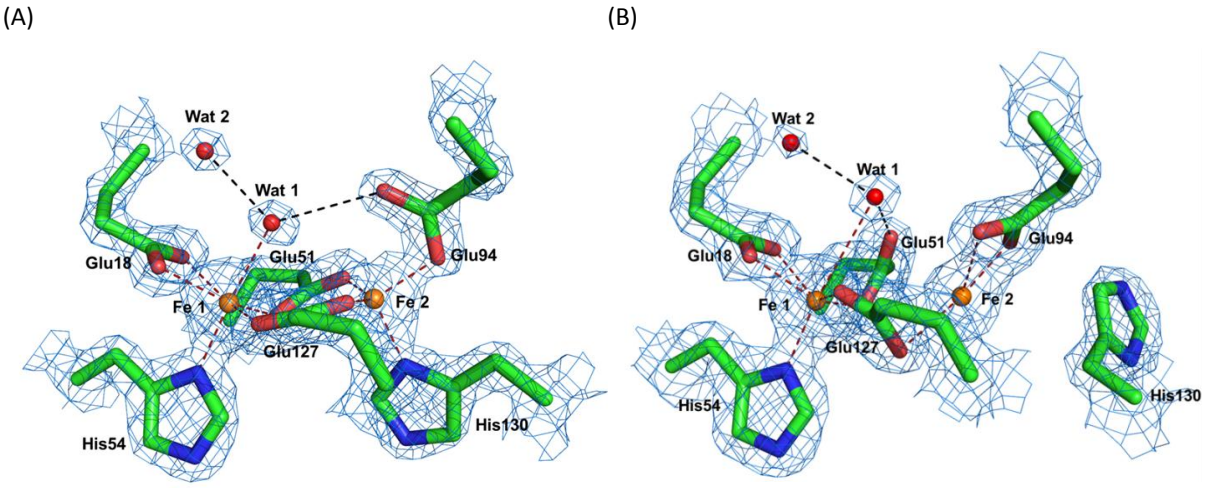

Figure 24. The ferroxidase center of electron density in 2mFo-DFc maps. The structure of Fe-soaked (A) and double-soaked (B) *Bv* Bfr active site. The 2mFo-DFc map (blue) is contoured at 2σ and 1.5σ, respectively. The dashed red lines show the coordinating bonds to the iron atoms and the dashed black lines indicate hydrogen bonds.

5.5 Three- and four-fold pore

In all three structures of *Bv* Bfr, the 3- and 4-fold pores are empty. Both pores have an hourglass shape from the outside of the protein shell down to the central cavity (Figure 25). From up to down, the narrowest part of the 3-fold pore has a diameter less than 1.2 Å and is capped by three Lys117 residues. Then the internal pore becomes wider and is lined by three Glu121 and Asn118 residues. For the 4-fold pore, the up entrance of the narrow part has a diameter less than 1.2 Å and is capped by hydrophobic Leu148 residues. It then widens to approximately 1.8 Å in diameter capped by the side chain of Gln151. Immediately after, the pore becomes broad to a diameter larger than 2.5 Å, enough for a water molecule to occupy. An iron (II) ion in an octahedral-coordinated state has a diameter of approximately 1.24 Å at low spin state (S = 1/2) and 1.56 Å at high spin state (S = 5/2) [150]. A pore with a diameter of 1.8 Å is wide enough for an iron ion to pass. So it might be possible that one iron ion coming from the central cavity gets coordinated by the oxygen atoms of four Gln151 residues and by an oxygen atom from a water molecule from beneath, which triggers the conformational change involving Leu148 to open up the pore and thereby releasing the iron.

The urea contained in double-soaking solution did not affect the static structural conformation of the 3-fold pore regions in *Bv* Bfr crystals as it did in eukaryotic ferritins (Figure 26). However, crystallographic B-factors of residues forming 4-fold pore are locally higher upon both Fe(II)- and double-soaked treatments and the changes are statistically significant. This might indicate that the structural dynamics of the 4-fold pore may assist the iron transport.

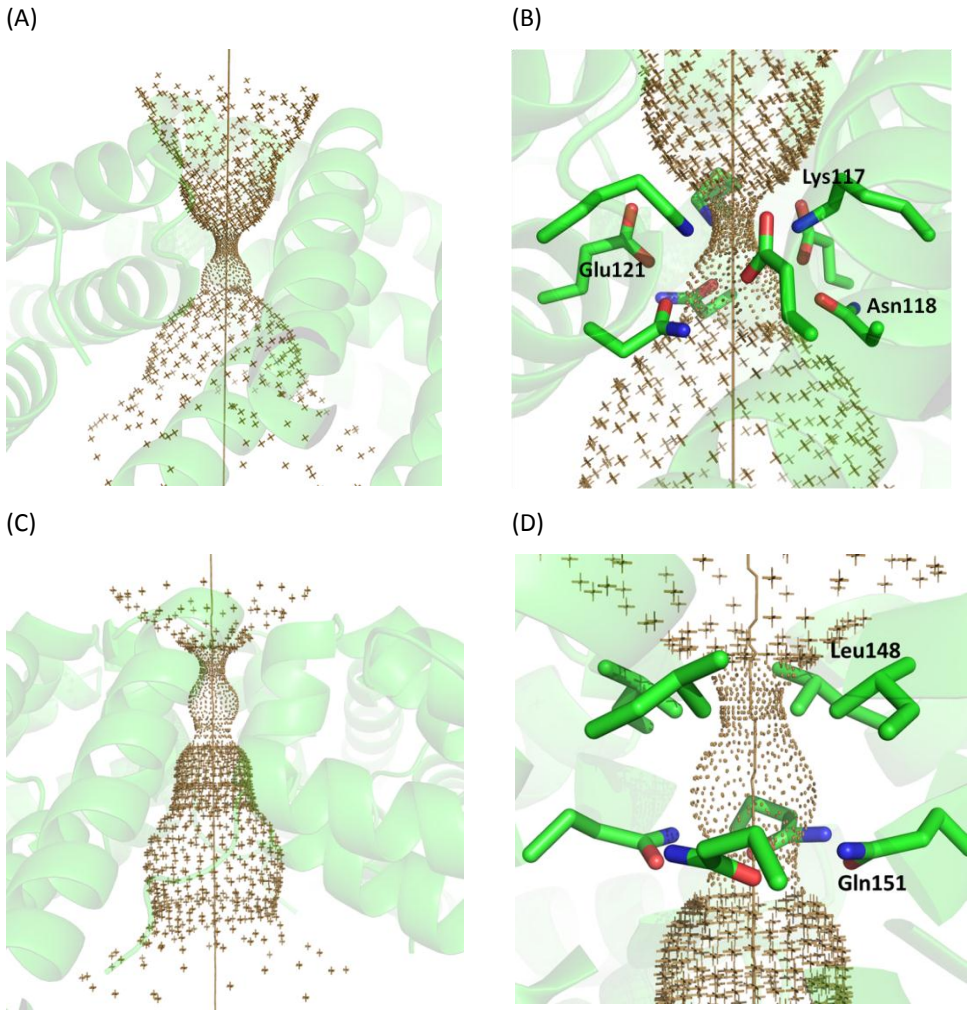

Figure 25. Hole2 representations of the 3-fold (A and B) and 4-fold pore (C and D). Overview of the *Bv* Bfr 3-fold (A) and 4-fold pore (C) and the respective constraining residues (B and D).

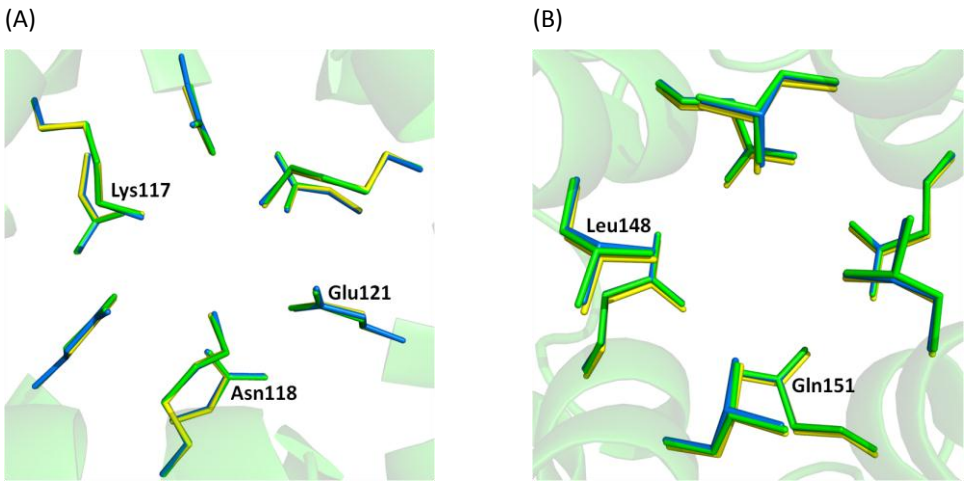

Figure 26. Superposition of the pore forming residues in the 3-fold and 4-fold pore. 3-fold pore (A) and 4-fold pore (B). The native 'as isolated' structure, the Fe(II) soaked and double soaked structures are marked with green, blue and yellow respectively.

5.6 Single crystal UV-VIS microspectrophotometry

Bacterioferritins are distinct from other ferritins by their ability to bind heme, hence the function of this co-factor is not clear. It has been suggested to play a role in facilitating release of iron in *Ec* Bfr by catalyzing electron transfer from outside of the protein to the internal cavity [151]. On the other hand, it was speculated that heme bound to *Ec* Bfr participates in some electron transfer process as part of ferroxidase activity, thus accelerating iron core formation [152]. In the three *Bv* Bfr structures, the heme-b molecules do not change conformation or position with response to the different treatments and their crystallographic B-factor is constant across the reported models. Since *Bv* Bfr has been co-purified with RC-LH1$_{vir}$, the electronic absorption spectrum on protein samples showed overlapping attributed by the presence of both proteins. In this case, UV-VIS microspectrometric on single crystals seems to be a more useful approach. The oxidation state of the heme-b in *Bv* Bfr was examined with a UV-VIS microspectrophotometer. Crystal UV-VIS absorption spectra was recorded before and after X-ray irradiation on "as-isolated", "Fe(II)-soaked" and "double-soaked" *Bv* Bfr crystals. The spectra from the "double-soaked" crystal had a high background, and is therefore not presented here [153].

There is always a conflict in choosing the optimum crystal size for X-ray data collection and microspectrometric measurements [154]. Since the focus of this experiment was on the microspectrometric studies, small *Bv* Bfr crystals were preferred. In addition to the size, the geometry of the crystal can further complicate measurements by having different refractive indexes for different wavelengths, so called chromatic aberration. Several different shapes of *Bv* Bfr crystals have been found during the crystallization trials, thus only the small cubic shaped crystals were chosen for this experiment. Due to the cubic shape of the crystals and the nature cubic symmetry of the protein, problems caused by orientation of the crystal among different measurements was reduced considerably. Two different fishing loops made of nylon or mylar were used to compare the spectral artifacts from the loops and loops of mylar gave less background. Mother liquid or soaking liquid was avoided when fishing the crystals to minimize the spectra background of RC-LH1$_{vir}$ from mother liquid, or $FeSO_4$ from soaking solution. To prevent ice forming around the crystals hence reducing the spectra noise further, crystals were dipped for 2-3 s in cryoprotected mother liquor or soaking solution containing 20 % glycerol before flash cooling. The crystal orientation was kept approximately the same for measurements performed before and after X-ray exposure to the crystal.

UV-VIS absorption spectra were recorded on "as-isolated" *Bv* Bfr crystals before and after 100 seconds X-ray exposure which was enough for one data set collection (Figure 27 A). For "Fe(II)-soaked" *Bv* Bfr crystal, spectra before and after one second X-ray exposure were measured (Figure 27 B). Heme shows a Soret band at around 418 nm, and α and β bands around 564 nm and 527 nm, respectively. Bis-methionine binding of heme is evidenced by a weak peak at around 740 nm. Here the measurements were focused on peaks at 564 nm which indicates an oxidized heme. As judged from a sharp peak at 558 nm and a broad peak at 564 nm in Figure 27 A, the hemes in "as-isolated" *Bv* Bfr crystals are in a mixture of redox states which was stable for 100 seconds in the X-ray beam. Interestingly, the "Fe(II)-soaked" crystals are completely oxidized before X-ray exposure but become fully reduced quite quickly, with as short as one second in the heavily attenuated X-ray beam (Figure 27B).

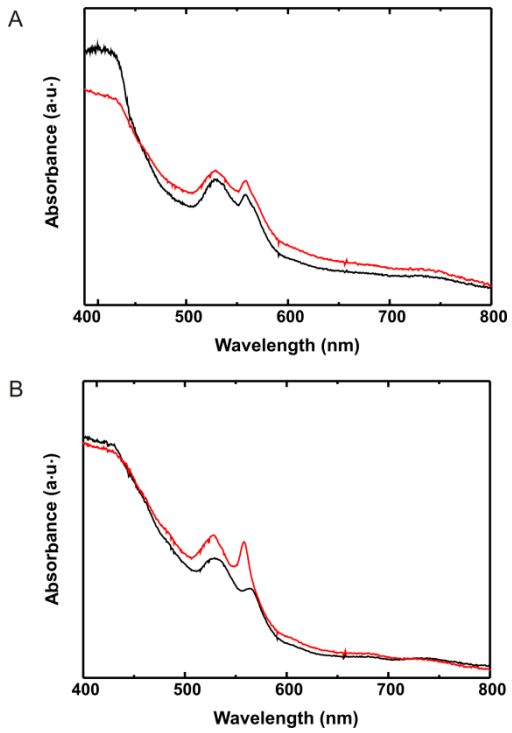

Figure 27. UV-VIS spectrum of *Bv* Bfr crystals. (A) UV/visible spectrum of "as-isolated" *Bv* Bfr crystals before (black) and after (red) 100 s X-ray exposure at beamline ID29, ESRF (5% transmission, X-ray flux: 9.6×1010 ph/s). (B) Crystal spectrum of "Fe(II)-soaked" *Bv* Bfr before (black) and after (red) 1 s X-ray exposure at beamline ID29, ESRF (5% transmission, X-ray flux: 1.2×1011 ph/s).

In an aerobic environment, why are the hemes of the "as-isolated" crystals in a mixture of redox states while in the "Fe(II)-soaked" crystals they are fully oxidized before the X-ray exposure? The "as-isolated" crystals were formed in crystallization drops dominated by RC-LH1 complex proteins which contain plenty of heme molecules and irons. They might compete with the oxygen dissolved in the drop liquids which led to a partially oxidized heme state. When they were moved to a drop of freshly prepared Fe(II)-soaking solution, the hemes were then easily oxidized. Furthermore, the increased radiation sensitivity in the "Fe(II)-soaked" crystals may be the result of a higher iron concentration in the surrounding buffer and the ferric mineral inside the central cavity which lead to the release of more photoelectrons upon X-ray exposure.

The role of heme in Bfrs is still debatable, thus its association with the electron transfer process seems to be a common speculation [151, 152]. Remarkably, an engineered *Ec* Bfr, constructed as a model for a component of photosystem II, has provided evidence that electron transfer between the zinc-chlorin e_6 bound to the heme binding site and the ferroxidase center reconstituted with two equivalents of manganese is possible via potentially redox-active Tyr25 in this artificial 'reaction center' [155].

Nature is an outstanding 'designer' of nearly everything, including macromolecules like proteins. With high-resolution X-ray crystallographic structures, conformational detail of proteins of interest can be visualized and studied at an atomic level. With the combination of other biochemical and biophysical methods, protein function and mechanism of action can be investigated in detail, hence helping design 'devices', like artificial solar cells, in the future.

6 Acknowledgements

If I have to use one word to summarize the time of my five years PhD study, it will be “fun”. I have to admit that 98 % of all the working days, it was fun to come to the lab in the morning, leave my lunch box in the refrigerator at the lunch corner and start to work. I have been very lucky in meeting many wonderful people.

First of all, I would like to thank my supervisor **Gergely Katona** for giving me the opportunity to work in his group. You made me a crystallographer. Your broad knowledge from cloning to computer simulation has kept surprising me. Thank you for patiently explaining everything to me. Also thanks **Richard Neutze**, my co-supervisor for inspiring me with your endless ideas. Many thanks to my former examiner **Jan Rydström** and examiner **Martin Billeter** for helping me with all my ‘problems’.

I start with our ‘Gergely-group’: **Annette**, **Ida**. It is ‘mysigt’ to have our weekly meetings. And the discussions after the meetings are as spirited. I had luck to have you two as group-mates and it was really fun to work together with you. Special thanks to Annette for sharing the same office with me. I believe we had a lot of fun behind the closing door. **Hadil** and **Elin**, thank you for doing your master thesis with me. Your hard works are two important parts of this thesis.

The ‘core-Neutze-group’: **Kristina**, **Annemarie**, **Urszula**, **Gerhard**, **Erik**, **Linda**, **David**, **Jennie**, **Cecilia** and **Peter**. Kristina, I really enjoyed being part of the protein engineering course. Thank you for giving me the opportunity. Annemarie, everything started by working with you with my master thesis. You introduced me to this group and taught me a lot of lab skills during my 6 months diploma work. Urszula, I hope you are a ‘Latte mamma’ because soon we will start to ‘fika i stan’. Gerhard, I remember those ESRF trips I had when I just moved to Gergely’s group. You are always so enthusiastic and willing to share your knowledge and experiences. Your ‘finishing-your-PhD’ document helped a lot. Linda, our SP discussions over lab bench always inspired me. David, thank you for taking care of the pod coffee machine. Linda, David, Erik, Jennie and Cecilia, I enjoyed your company in LCLS.

The ‘Susanna-group’: **Susanna**, **Anna**, **Mikael**, **Mike** and **Etienne**. Susanna, I am very grateful for your understanding and comfort. Anna, we started at Chalmers at the same time 10 years ago. It was great to have you as a classmate, whom I could borrow lecture notes from. Having you around the lab was even better since you are always so responsible. Mikael, thanks for a wonderful ‘protein engineering’ party. I really enjoyed it. Mike, Jennie, Annette, Ida, Rhawnie and Elin, I certainly had fun with you during the ECM.

The ‘Rosie-group’: **Rosie**, **Rhawnie** and **Elin**. Rosie, you are always positive, helpful and easy to talk to. Rhawnie, it is great to have you in the corridor. You are always so kind and positive. Elin, I feel we were a great team during your diploma work and I hope you had fun during those 10 months.

The ‘Sebastian-group’: **Sebastian, Petra, Alex, Sebastian, Stephan, Kiran, Yogi**. A group which keeps growing. Welcome all the new comers.

Karin, **Madde, Piero, Chaille.** Karin, thank you for all the ESRF trips. Madde, good luck with your thesis and house-renovation. Piero, your tiramisu was really delicious. And Chaille, thank you so much for proof-reading my thesis.

The 'NMR-center': **Göran**, **Vladislav**, **Anders**, **Cecilia**, **Helén**, **Johan**, **Maria**, **Linnéa**. I missed the Friday's sushi and fika at NMR-center. Johan, thank you for taking care of me during my first year of PhD at NMR center. I learned a lot from you. Anders, you are always so understanding and helpful. Helén, ska luncha med dig snart. Maria, **Diana**, **Jonas K**, **Jonas A**, **Daniel**, not bad we went through half of the 'Spin Dynamics' together.

Till min familj, Mikael, Erik och Johan. Jag är oerhört tacksam för allt stöd från dig, Mikael. Du är en fantastisk make och pappa. Jag är lycklig som får dela livet med dig. Erik och Johan, tack för att ni har gett mitt liv en ny dimension och jag älskar er!

7 References

1. Oparin, A.I. and S. Morgulis, *The origin of life*. 1938, New York: The Macmillan Company. viii p., 2 l., 270 p.
2. Miller, S.L. and H.C. Urey, *Organic Compound Synthesis on the Primitive Earth.* Science, 1959. **130**(3370): p. 245-251.
3. Miller, S.L., *A Production of Amino Acids under Possible Primitive Earth Conditions.* Science, 1953. **117**(3046): p. 528-529.
4. Voet, D. and J.G. Voet, *Biochemistry*. 3rd ed. 2004, New York: J. Wiley & Sons. 2 v. (xv, 1560 p.).
5. Xiong, J. and C.E. Bauer, *A cytochrome b origin of photosynthetic reaction centers: an evolutionary link between respiration and photosynthesis.* Journal of Molecular Biology, 2002. **322**(5): p. 1025-1037.
6. Xiong, J., et al., *Molecular evidence for the early evolution of photosynthesis.* Science, 2000. **289**(5485): p. 1724-1730.
7. Xiong, J. and C.E. Bauer, *Complex evolution of photosynthesis.* Annual Review of Plant Biology, 2002. **53**: p. 503-521.
8. Andrews, S.C., *Iron storage in bacteria.* Advances in Microbial Physiology, Vol 40, 1998. **40**: p. 281-351.
9. Crichton, R.R. and J.R. Boelaert, *Inorganic biochemistry of iron metabolism : from molecular mechanisms to clinical consequences*. 2nd ed. 2001, Chichester ; New York: Wiley. xxii, 326 p., 16 p. of plates.
10. Bertini, I., A. Sigel and H. Sigel, *Handbook on metalloproteins*. 2001, New York: Marcel Dekker. xxvii, 1182 p.
11. Telenius, H., et al., *Degenerate Oligonucleotide-Primed Pcr - General Amplification of Target DNA by a Single Degenerate Primer.* Genomics, 1992. **13**(3): p. 718-725.
12. Weyer, K.A., et al., *Amino acid sequence of the cytochrome subunit of the photosynthetic reaction centre from the purple bacterium Rhodopseudomonas viridis.* EMBO J, 1987. **6**(8): p. 2197-202.
13. Michel, H., et al., *The 'heavy' subunit of the photosynthetic reaction centre from Rhodopseudomonas viridis: isolation of the gene, nucleotide and amino acid sequence.* EMBO J, 1985. **4**(7): p. 1667-72.
14. Michel, H., et al., *The 'light' and 'medium' subunits of the photosynthetic reaction centre from Rhodopseudomonas viridis: isolation of the genes, nucleotide and amino acid sequence.* EMBO J, 1986. **5**(6): p. 1149-58.
15. Fischer, G., *Structure and Function of Aqua(glycero)porins* in *THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE NATURE SCIENCE,* 2011, University of Gothenburg: Online: http://hdl.handle.net/2077/24754.
16. Wadsten-Hindrichsen, P., *On the Crystallization of Membrane Proteins in Lipidic Sponge and Cubic Phases*, in *THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY,* 2007, Chalmers University of Technology.
17. Landau, E.M. and J.P. Rosenbusch, *Lipidic cubic phases: A novel concept for the crystallization of membrane proteins.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(25): p. 14532-14535.
18. Landau, E.M. and J.P. Rosenbusch, *Lipidic cubic phases: a novel concept for the crystallization of membrane proteins.* Proc Natl Acad Sci U S A, 1996. **93**(25): p. 14532-5.
19. Caffrey, M., *A lipid's eye view of membrane protein crystallization in mesophases.* Current Opinion in Structural Biology, 2000. **10**(4): p. 486-497.
20. Cherezov, V., et al., *Room to move: Crystallizing membrane proteins in swollen lipidic mesophases.* Journal of Molecular Biology, 2006. **357**(5): p. 1605-1618.
21. Engstrom, S., et al., *Solvent-induced sponge (L-3) phases in the solvent-monoolein-water system.* Colloid Science of Lipids, 1998. **108**: p. 93-98.

22. Ridell, A., et al., *On the water content of the solvent/monoolein/water sponge (L-3)phase.* Colloids and Surfaces a-Physicochemical and Engineering Aspects, 2003. **228**(1-3): p. 17-24.
23. Gabriel, N.E. and M.F. Roberts, *Spontaneous Formation of Stable Unilamellar Vesicles.* Biochemistry, 1984. **23**(18): p. 4011-4015.
24. Sanders, C.R. and R.S. Prosser, *Bicelles: a model membrane system for all seasons?* Structure with Folding & Design, 1998. **6**(10): p. 1227-1234.
25. Lancelot, N., K. Elbayed, and M. Piotto, *Applications of variable-angle sample spinning experiments to the measurement of scaled residual dipolar couplings and (15)N CSA in soluble proteins.* Journal of Biomolecular Nmr, 2005. **33**(3): p. 153-161.
26. Park, S.H., et al., *High-resolution NMR spectroscopy of a GPCR in aligned bicelles.* Journal of the American Chemical Society, 2006. **128**(23): p. 7402-7403.
27. Carlotti, C., F. Aussenac, and E.J. Dufourc, *Towards high-resolution H-1-NMR in biological membranes: magic angle spinning of bicelles.* Biochimica Et Biophysica Acta-Biomembranes, 2002. **1564**(1): p. 156-164.
28. Faham, S. and J.U. Bowie, *Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure.* Journal of Molecular Biology, 2002. **316**(1): p. 1-6.
29. Nieh, M.P., et al., *SANS study of the structural phases of magnetically alignable lanthanide-doped phospholipid mixtures.* Langmuir, 2001. **17**(9): p. 2629-2638.
30. Czerski, L. and C.R. Sanders, *Functionality of a membrane protein in bicelles.* Analytical Biochemistry, 2000. **284**(2): p. 327-333.
31. Faham, S. and J.U. Bowie, *Bicelle crystallization: A new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure.* Journal of Molecular Biology, 2002. **316**(1): p. 1-6.
32. Rasmussen, S.G.F., et al., *Crystal structure of the human beta(2) adrenergic G-protein-coupled receptor.* Nature, 2007. **450**(7168): p. 383-U4.
33. Ujwal, R., et al., *The crystal structure of mouse VDAC1 at 2.3 angstrom resolution reveals mechanistic insights into metabolite gating.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(46): p. 17742-17747.
34. Luecke, H., et al., *Crystallographic structure of xanthorhodopsin, the light-driven proton pump with a dual chromophore.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(43): p. 16561-16565.
35. Vinothkumar, K.R., *Structure of Rhomboid Protease in a Lipid Environment.* Journal of Molecular Biology, 2011. **407**(2): p. 232-247.
36. Dauter, Z., *New approaches to high-throughput phasing.* Current Opinion in Structural Biology, 2002. **12**(5): p. 674-678.
37. Arndt, U.W., R.A. Crowther, and J.F. Mallett, *A computer-linked cathode-ray tube microdensitometer for x-ray crystallography.* J Sci Instrum, 1968. **1**(5): p. 510-6.
38. Diederichs, K. and P.A. Karplus, *Improved R-factors for diffraction data analysis in macromolecular crystallography.* Nat Struct Biol, 1997. **4**(4): p. 269-75.
39. Evans, P.R., *An introduction to data reduction: space-group determination, scaling and intensity statistics.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 282-92.
40. Evans, P., *Scaling and assessment of data quality.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 1): p. 72-82.
41. Karplus, P.A. and K. Diederichs, *Linking crystallographic model and data quality.* Science, 2012. **336**(6084): p. 1030-3.
42. Schneider, T.R. and G.M. Sheldrick, *Substructure solution with SHELXD.* Acta Crystallogr D Biol Crystallogr, 2002. **58**(Pt 10 Pt 2): p. 1772-9.
43. Neurath, H., *The Versatility of Proteolytic-Enzymes.* Journal of Cellular Biochemistry, 1986. **32**(1): p. 35-49.

44. Polgar, L., *The catalytic triad of serine peptidases.* Cell Mol Life Sci, 2005. **62**(19-20): p. 2161-72.
45. Blow, D.M., J.J. Birktoft, and B.S. Hartley, *Role of a Buried Acid Group in Mechanism of Action of Chymotrypsin.* Nature, 1969. **221**(5178): p. 337-40.
46. Matthews, B.W., et al., *3-Dimensional Structure of Tosyl-Alpha-Chymotrypsin.* Nature, 1967. **214**(5089): p. 652-&.
47. Robertus, J.D., et al., *Subtilisin - Stereochemical Mechanism Involving Transition-State Stabilization.* Biochemistry, 1972. **11**(23): p. 4293-&.
48. Markley, J.L. and M.A. Porubcan, *The charge-relay system of serine proteinases: proton magnetic resonance titration studies of the four histidines of porcine trypsin.* J Mol Biol, 1976. **102**(3): p. 487-509.
49. Ash, E.L., et al., *A low-barrier hydrogen bond in the catalytic triad of serine proteases? Theory versus experiment.* Science, 1997. **278**(5340): p. 1128-1132.
50. Sprang, S., et al., *The three-dimensional structure of Asn102 mutant trypsin:role of Asp102 in serine protease catalysis.* Science, 1987. **237**: p. 905-909.
51. Ahmed, H.U., et al., *The determination of protonation states in proteins.* Acta Crystallogr D Biol Crystallogr, 2007. **63**(Pt 8): p. 906-22.
52. Schmidt, A., et al., *Trypsin revisited: crystallography AT (SUB) atomic resolution and quantum chemistry revealing details of catalysis.* J Biol Chem, 2003. **278**(44): p. 43357-62.
53. Vrielink, A. and N. Sampson, *Sub-Angstrom resolution enzyme X-ray structures: is seeing believing?* Curr Opin Struct Biol, 2003. **13**(6): p. 709-15.
54. Burgess, S.G., et al., *Probing the dynamic interface between trimethylamine dehydrogenase (TMADH) and electron transferring flavoprotein (ETF) in the TMADH-2ETF complex: role of the Arg-alpha237 (ETF) and Tyr-442 (TMADH) residue pair.* Biochemistry, 2008. **47**(18): p. 5168-81.
55. Williams, D.H., et al., *Understanding noncovalent interactions: ligand binding energy and catalytic efficiency from ligand-induced reductions in motion within receptors and enzymes.* Angew Chem Int Ed Engl, 2004. **43**(48): p. 6596-616.
56. Szenthe, B., et al., *When the surface tells what lies beneath: combinatorial phage-display mutagenesis reveals complex networks of surface-core interactions in the pacifastin protease inhibitor family.* J Mol Biol, 2007. **370**(1): p. 63-79.
57. Malik, Z., et al., *Proteinase inhibitors from desert locust, Schistocerca gregaria: engineering of both P(1) and P(1)' residues converts a potent chymotrypsin inhibitor to a potent trypsin inhibitor.* Biochim Biophys Acta, 1999. **1434**(1): p. 143-50.
58. Orio, M., D.A. Pantazis, and F. Neese, *Density functional theory.* Photosynthesis Research, 2009. **102**(2-3): p. 443-53.
59. Deisenhofer, J., et al., *X-Ray Structure-Analysis of a Membrane-Protein Complex - Electron-Density Map at 3a Resolution and a Model of the Chromophores of the Photosynthetic Reaction Center from Rhodopseudomonas-Viridis.* Journal of Molecular Biology, 1984. **180**(2): p. 385-398.
60. Deisenhofer, J., et al., *Structure of the Protein Subunits in the Photosynthetic Reaction Center of Rhodopseudomonas-Viridis at 3a Resolution.* Nature, 1985. **318**(6047): p. 618-624.
61. Williams, J.C., et al., *Primary Structure of the L-Subunit of the Reaction Center from Rhodopseudomonas-Sphaeroides.* Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, 1984. **81**(23): p. 7303-7307.
62. Williams, J.C., et al., *Primary Structure of the M-Subunit of the Reaction Center from Rhodopseudomonas-Sphaeroides.* Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, 1983. **80**(21): p. 6505-6509.
63. Donohue, T.J., J.H. Hoger, and S. Kaplan, *Cloning and Expression of the Rhodobacter-Sphaeroides Reaction Center-H Gene.* Journal of Bacteriology, 1986. **168**(2): p. 953-961.

64. Allen, J.P., et al., *Structure of the Reaction Center from Rhodobacter-Sphaeroides R-26 - the Cofactors.* Proceedings of the National Academy of Sciences of the United States of America, 1987. **84**(16): p. 5730-5734.
65. Wohri, A.B., et al., *Light-Induced Structural Changes in a Photosynthetic Reaction Center Caught by Laue Diffraction.* Science, 2010. **328**(5978): p. 630-633.
66. Katona, G., et al., *Lipidic cubic phase crystal structure of the photosynthetic reaction centre from Rhodobacter sphaeroides at 2.35 angstrom resolution.* Journal of Molecular Biology, 2003. **331**(3): p. 681-692.
67. Wadsten, P., et al., *Lipidic sponge phase crystallization of membrane proteins.* Journal of Molecular Biology, 2006. **364**(1): p. 44-53.
68. Gardiner, A.T., et al., *Electron paramagnetic resonance studies of zinc-substituted reaction centers from Rhodopseudomonas viridis.* Biochemistry, 1999. **38**(36): p. 11773-11787.
69. Deisenhofer, J. and H. Michel, *The Photosynthetic Reaction Center from the Purple Bacterium Rhodopseudomonas-Viridis - Nobel Lecture, December 8, 1988.* Chemica Scripta, 1989. **29**(3): p. 205-&.
70. Roszak, A.W., et al., *New insights into the structure of the reaction centre from Blastochloris viridis: evolution in the laboratory.* Biochemical Journal, 2012. **442**: p. 27-37.
71. Lancaster, C.R.D. and H. Michel, *The coupling of light-induced electron transfer and proton uptake as derived from crystal structures of reaction centres from Rhodopseudomonas viridis modified at the binding site of the secondary quinone, Q(B).* Structure, 1997. **5**(10): p. 1339-1359.
72. Deisenhofer, J., et al., *Crystallographic Refinement at 2.3-Angstrom Resolution and Refined Model of the Photosynthetic Reaction-Center from Rhodopseudomonas-Viridis.* Journal of Molecular Biology, 1995. **246**(3): p. 429-457.
73. Baxter, R.H.G., et al., *Time-resolved crystallographic studies of light-induced structural changes in the photosynthetic reaction center.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(16): p. 5982-5987.
74. Baxter, R.H.G., et al., *Cryogenic structure of the photosynthetic reaction center of Blastochloris viridis in the light and dark.* Acta Crystallographica Section D-Biological Crystallography, 2005. **61**: p. 605-612.
75. Warncke, K., et al., *Influence of hydrocarbon tail structure on quinone binding and electron-transfer performance at the QA and QB sites of the photosynthetic reaction center protein.* Biochemistry, 1994. **33**(25): p. 7830-41.
76. Lancaster, C.R., *Quinone-binding sites in membrane proteins: what can we learn from the Rhodopseudomonas viridis reaction centre?* Biochem Soc Trans, 1999. **27**(4): p. 591-6.
77. Arnold, E. and M.G. Rossmann, *Analysis of the structure of a common cold virus, human rhinovirus 14, refined at a resolution of 3.0 A.* Journal of Molecular Biology, 1990. **211**(4): p. 763-801.
78. Li, L., et al., *Nanoliter microfluidic hybrid method for simultaneous screening and optimization validated with crystallization of membrane proteins.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(51): p. 19243-19248.
79. East, J.M. and A.G. Lee, *Lipid Selectivity of the Calcium and Magnesium-Ion Dependent Adenosine-Triphosphatase, Studied with Fluorescence Quenching by a Brominated Phospholipid.* Biochemistry, 1982. **21**(17): p. 4144-4151.
80. Roszak, A.W., et al., *Brominated lipids identify lipid binding sites on the surface of the reaction center from Rhodobacter sphaeroides.* Biochemistry, 2007. **46**(11): p. 2909-2916.
81. Cogdell, R.J., et al., *How photosynthetic bacteria harvest solar energy.* Journal of Bacteriology, 1999. **181**(13): p. 3869-3879.

82. Walz, T., et al., *Projection structures of three photosynthetic complexes from Rhodobacter sphaeroides: LH2 at 6 angstrom LH1 and RC-LH1 at 25 angstrom.* Journal of Molecular Biology, 1998. **282**(4): p. 833-845.
83. Mcdermott, G., et al., *Crystal-Structure of an Integral Membrane Light-Harvesting Complex from Photosynthetic Bacteria.* Nature, 1995. **374**(6522): p. 517-521.
84. Papiz, M.Z., et al., *The structure and thermal motion of the B800-850 LH2 complex from Rps. acidophila at 2.0 (A)over-circle resolution and 100 K: New structural features and functionally relevant motions.* Journal of Molecular Biology, 2003. **326**(5): p. 1523-1538.
85. Roszak, A.W., et al., *Crystal structure of the RC-LH1 core complex from Rhodopseudomonas palustris.* Science, 2003. **302**(5652): p. 1969-1972.
86. Jungas, C., et al., *Supramolecular organization of the photosynthetic apparatus of Rhodobacter sphaeroides.* Embo Journal, 1999. **18**(3): p. 534-542.
87. Miller, K.R., *3-Dimensional Structure of a Photosynthetic Membrane.* Nature, 1982. **300**(5887): p. 53-55.
88. Stahlberg, H., et al., *Are the light-harvesting I complexes from Rhodospirillum rubrum arranged around the reaction centre in a square geometry?* Journal of Molecular Biology, 1998. **282**(4): p. 819-831.
89. Ikeda-Yamasaki, I., et al., *Projection map of the reaction center-light harvesting 1 complex from Rhodopseudomonas viridis at 10 angstrom resolution.* Febs Letters, 1998. **425**(3): p. 505-508.
90. Jamieson, S.J., et al., *Projection structure of the photosynthetic reaction centre-antenna complex of Rhodospirillum rubrum at 8.5 angstrom resolution.* Embo Journal, 2002. **21**(15): p. 3927-3935.
91. Scheuring, S., et al., *Nanodissection and high-resolution imaging of the Rhodopseudomonas viridis photosynthetic core complex in native membranes by AFM.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(4): p. 1690-1693.
92. Scheuring, S., *AFM studies of the supramolecular assembly of bacterial photosynthetic core-complexes.* Current Opinion in Chemical Biology, 2006. **10**(5): p. 387-393.
93. Konorty, M., et al., *Structural analysis of photosynthetic membranes by cryo-electron tomography of intact Rhodopseudomonas viridis cells.* Journal of Structural Biology, 2008. **161**(3): p. 393-400.
94. Scheuring, S., et al., *The photosynthetic apparatus of Rhodopseudomonas palustris: Structures and organization.* Journal of Molecular Biology, 2006. **358**(1): p. 83-96.
95. Saijo, S., et al., *Crystallization and preliminary X-ray studies on the reaction center-light-harvesting 1 core complex from Rhodopseudomonas viridis.* Acta Crystallographica Section F-Structural Biology and Crystallization Communications, 2005. **61**: p. 83-86.
96. Iwata, S., *Methods and results in crystallization of membrane proteins*. IUL biotechnology series. 2003, La Jolla, Calif.: International Unversity Line. xviii, 355 p., 10 p. of plates.
97. Newstead, S., S. Ferrandon, and S. Iwata, *Rationalizing alpha-helical membrane protein crystallization.* Protein Science, 2008. **17**(3): p. 466-472.
98. Misquitta, L.V., et al., *Membrane protein crystallization in lipidic mesophases with tailored bilayers.* Structure, 2004. **12**(12): p. 2113-2124.
99. Cherezov, V., et al., *In meso structure of the cobalamin transporter, BtuB, at 1.95 angstrom resolution.* Journal of Molecular Biology, 2006. **364**(4): p. 716-734.
100. Kurisu, G., et al., *Structure of the cytochrome b6f complex of oxygenic photosynthesis: tuning the cavity.* Science, 2003. **302**(5647): p. 1009-14.
101. Guan, L., et al., *Structural determination of wild-type lactose permease.* Proc Natl Acad Sci U S A, 2007. **104**(39): p. 15294-8.
102. Cherezov, V., et al., *High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor.* Science, 2007. **318**(5854): p. 1258-65.

103. Hanson, M.A., et al., *A specific cholesterol binding site is established by the 2.8 angstrom structure of the human beta(2)-adrenergic receptor.* Structure, 2008. **16**(6): p. 897-905.
104. Jenni, S., et al., *Structure of fungal fatty acid synthase and implications for iterative substrate shuttling.* Science, 2007. **316**(5822): p. 254-261.
105. Cramer, P., D.A. Bushnell, and R.D. Kornberg, *Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution.* Science, 2001. **292**(5523): p. 1863-1876.
106. Jordan, P., et al., *Three-dimensional structure of cyanobacterial photosystem I at 2.5 angstrom resolution.* Nature, 2001. **411**(6840): p. 909-917.
107. Heras, B. and J.L. Martin, *Post-crystallization treatments for improving diffraction quality of protein crystals.* Acta Crystallographica Section D-Biological Crystallography, 2005. **61**: p. 1173-1180.
108. Sanchez-Weatherby, J., et al., *Improving diffraction by humidity control: a novel device compatible with X-ray beamlines.* Acta Crystallographica Section D-Biological Crystallography, 2009. **65**: p. 1237-1246.
109. Brotosudarmo, T.H.P., et al., *The light intensity under which cells are grown controls the type of peripheral light-harvesting complexes that are assembled in a purple photosynthetic bacterium.* Biochemical Journal, 2011. **440**: p. 51-61.
110. Odahara, T., *Stability and solubility of integral membrane proteins from photosynthetic bacteria solubilized in different detergents.* Biochimica Et Biophysica Acta-Biomembranes, 2004. **1660**(1-2): p. 80-92.
111. Ravelli, R.B.G. and E.F. Garman, *Radiation damage in macromolecular cryocrystallography.* Current Opinion in Structural Biology, 2006. **16**(5): p. 624-629.
112. Mueller, M., S. Jenni, and N. Ban, *Strategies for crystallization and structure determination of very large macromolecular assemblies.* Current Opinion in Structural Biology, 2007. **17**(5): p. 572-579.
113. Neutze, R., et al., *Potential for biomolecular imaging with femtosecond X-ray pulses.* Nature, 2000. **406**(6797): p. 752-757.
114. Winick, H., *The linac coherent light source (LCLS): A fourth-generation light source using the SLAG linac.* Journal of Electron Spectroscopy and Related Phenomena, 1995. **75**: p. 1-8.
115. Neutze, R., et al., *Potential impact of an X-ray free electron laser on structural biology.* Radiation Physics and Chemistry, 2004. **71**(3-4): p. 905-916.
116. Boutet, S. and I.K. Robinson, *Coherent X-ray diffractive imaging of protein crystals.* Journal of Synchrotron Radiation, 2008. **15**: p. 576-583.
117. Chapman, H.N., *X-ray imaging beyond the limits.* Nature Materials, 2009. **8**(4): p. 299-301.
118. Kern, J., et al., *Room temperature femtosecond X-ray diffraction of photosystem II microcrystals.* Proc Natl Acad Sci U S A, 2012. **109**(25): p. 9721-6.
119. Johansson, L.C., et al., *Lipidic phase membrane protein serial femtosecond crystallography.* Nat Methods, 2012. **9**(3): p. 263-5.
120. Boutet, S. and G.J. Williams, *The Coherent X-ray Imaging (CXI) instrument at the Linac Coherent Light Source (LCLS).* New Journal of Physics, 2010. **12**.
121. DePonte, D.P., et al., *Gas dynamic virtual nozzle for generation of microscopic droplet streams.* Journal of Physics D-Applied Physics, 2008. **41**(19).
122. Fyfe, P.K., et al., *Disruption of a specific molecular interaction with a bound lipid affects the thermal stability of the purple bacterial reaction centre.* Biochimica Et Biophysica Acta, 2004. **1608**(1): p. 11-22.
123. Jormakka, M., et al., *Molecular basis of proton motive force generation: structure of formate dehydrogenase-N.* Science, 2002. **295**(5561): p. 1863-8.
124. Robinson, N.C., *Functional binding of cardiolipin to cytochrome c oxidase.* J Bioenerg Biomembr, 1993. **25**(2): p. 153-63.
125. Paradies, G., et al., *Reactive oxygen species affect mitochondrial electron transport complex I activity through oxidative cardiolipin damage.* Gene, 2002. **286**(1): p. 135-41.

126. Gourdon, P., et al., *HiLiDe-Systematic Approach to Membrane Protein Crystallization in Lipid and Detergent.* Crystal Growth & Design, 2011. **11**(6): p. 2098-2106.
127. Levi, S., et al., *Mechanism of ferritin iron uptake: activity of the H-chain and deletion mapping of the ferro-oxidase site. A study of iron uptake and ferro-oxidase activity of human liver, recombinant H-chain ferritins, and of two H-chain deletion mutants.* J Biol Chem, 1988. **263**(34): p. 18086-92.
128. Weeratunga, S.K., et al., *Structural studies of bacterioferritin B from Pseudomonas aeruginosa suggest a gating mechanism for iron uptake via the ferroxidase center.* Biochemistry, 2010. **49**(6): p. 1160-75.
129. Crow, A., et al., *Structural basis for iron mineralization by bacterioferritin.* J Am Chem Soc, 2009. **131**(19): p. 6808-13.
130. Hempstead, P.D., et al., *Comparison of the three-dimensional structures of recombinant human H and horse L ferritins at high resolution.* J Mol Biol, 1997. **268**(2): p. 424-48.
131. Treffry, A., et al., *Defining the roles of the threefold channels in iron uptake, iron oxidation and iron-core formation in ferritin: a study aided by site-directed mutagenesis.* Biochem J, 1993. **296 (Pt 3)**: p. 721-8.
132. Theil, E.C., X.S. Liu, and T. Tosha, *Gated Pores in the Ferritin Protein Nanocage.* Inorganica Chim Acta, 2008. **361**(4): p. 868-874.
133. Hasan, M.R., T. Tosha, and E.C. Theil, *Ferritin contains less iron (59Fe) in cells when the protein pores are unfolded by mutation.* Journal of Biological Chemistry, 2008. **283**(46): p. 31394-400.
134. Takagi, H., et al., *Localized unfolding at the junction of three ferritin subunits. A mechanism for iron release?* Journal of Biological Chemistry, 1998. **273**(30): p. 18685-8.
135. Liu, X.S., et al., *Peptides selected for the protein nanocage pores change the rate of iron recovery from the ferritin mineral.* Journal of Biological Chemistry, 2007. **282**(44): p. 31821-5.
136. Liu, X., W. Jin, and E.C. Theil, *Opening protein pores with chaotropes enhances Fe reduction and chelation of Fe from the ferritin biomineral.* Proc Natl Acad Sci U S A, 2003. **100**(7): p. 3653-8.
137. Cobessi, D., et al., *The 2.6 A resolution structure of Rhodobacter capsulatus bacterioferritin with metal-free dinuclear site and heme iron in a crystallographic 'special position'.* Acta Crystallogr D Biol Crystallogr, 2002. **58**(Pt 1): p. 29-38.
138. Macedo, S., et al., *The nature of the di-iron site in the bacterioferritin from Desulfovibrio desulfuricans.* Nature Structural Biology, 2003. **10**(4): p. 285-290.
139. Swartz, L., et al., *Redox-dependent structural changes in the Azotobacter vinelandii bacterioferritin: new insights into the ferroxidase and iron transport mechanism.* Biochemistry, 2006. **45**(14): p. 4421-8.
140. Janowski, R., T. Auerbach-Nevo, and M.S. Weiss, *Bacterioferritin from Mycobacterium smegmatis contains zinc in its di-nuclear site.* Protein Sci, 2008. **17**(7): p. 1138-50.
141. Gupta, V., et al., *Crystal structure of Bfr A from Mycobacterium tuberculosis: incorporation of selenomethionine results in cleavage and demetallation of haem.* PLoS One, 2009. **4**(11): p. e8028.
142. Nam, K.H., et al., *Crystal structure of bacterioferritin from Rhodobacter sphaeroides.* Biochem Biophys Res Commun, 2010. **391**(1): p. 990-4.
143. Liu, H.L., et al., *2.6 angstrom resolution crystal structure of the bacterioferritin from Azotobacter vinelandii.* Febs Letters, 2004. **573**(1-3): p. 93-98.
144. Lang, F.S. and D. Oesterhelt, *Microaerophilic Growth and Induction of the Photosynthetic Reaction Center in Rhodopseudomonas-Viridis.* Journal of Bacteriology, 1989. **171**(5): p. 2827-2834.
145. Nam, K.H., et al., *Crystal structure of bacterioferritin from Rhodobacter sphaeroides.* Biochemical and Biophysical Research Communications, 2010. **391**(1): p. 990-994.

146. Weyer, K.A., et al., *Amino-Acid-Sequence of the Cytochrome Subunit of the Photosynthetic Reaction Center from the Purple Bacterium Rhodopseudomonas-Viridis.* Embo Journal, 1987. **6**(8): p. 2197-2202.
147. Michel, H., et al., *The Heavy Subunit of the Photosynthetic Reaction Center from Rhodopseudomonas-Viridis - Isolation of the Gene, Nucleotide and Amino-Acid Sequence.* Embo Journal, 1985. **4**(7): p. 1667-1672.
148. Michel, H., et al., *The Light and Medium Subunits of the Photosynthetic Reaction Center from Rhodopseudomonas-Viridis - Isolation of the Genes, Nucleotide and Amino-Acid-Sequence.* Embo Journal, 1986. **5**(6): p. 1149-1158.
149. Smart, O.S., et al., *A novel method for structure-based prediction of ion channel conductance properties.* Biophys J, 1997. **72**(3): p. 1109-26.
150. Shannon, R.D., *Revised Effective Ionic-Radii and Systematic Studies of Interatomic Distances in Halides and Chalcogenides.* Acta Crystallographica Section A, 1976. **32**(Sep1): p. 751-767.
151. Yasmin, S., et al., *A new role for heme, facilitating release of iron from the bacterioferritin iron biomineral.* J Biol Chem, 2011. **286**(5): p. 3473-83.
152. Wong, S.G., et al., *Fe-heme bound to Escherichia coli bacterioferritin accelerates iron core formation by an electron transfer mechanism.* Biochem J, 2012.
153. Royant, A., et al., *Advances in spectroscopic methods for biological crystals. 1. Fluorescence lifetime measurements.* Journal of Applied Crystallography, 2007. **40**: p. 1105-1112.
154. Baxter, R.H.G., E. Krausz, and J.R. Norris, *Photoactivation of the photosynthetic reaction center of Blastochloris viridis in the crystalline state.* Journal of Physical Chemistry B, 2006. **110**(2): p. 1026-1032.
155. Conlan, B., et al., *Photo-catalytic oxidation of a di-nuclear manganese centre in an engineered bacterioferritin 'reaction centre'.* Biochimica Et Biophysica Acta-Bioenergetics, 2009. **1787**(9): p. 1112-1121.