Study of Protein Production, Folding, Crystallization and Structure:

Survival of Motor Neuron Protein and Fenna-Matthews-Olson Protein

by

Chadwick R Larson

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved November 2010 by the
Graduate Supervisory Committee:

James Allen, Chair
Wilson Francisco
Julian Chen

ARIZONA STATE UNIVERSITY

December 2010

ABSTRACT

Protein crystallization has become an extremely important tool in biochemistry since the first structure of the protein Myoglobin was solved in 1958. Survival of motor neuron protein has proved to be an elusive target in regards to producing crystals of sufficient quality for X-ray diffraction. One form of Survival of motor neuron protein has been found to be a cause of the disease Spinal Muscular Atrophy that currently affects 1 in 6000 live births. The production, purification and crystallization of Survival of motor neuron protein are detailed.

The Fenna-Matthews-Olson (FMO) protein from *Pelodictyon phaeum* is responsible for the transfer of energy from the chlorosome complex to the reaction center of the bacteria. The three-dimensional structure of the protein has been solved to a resolution of 2.0Å with the R_{work} and R_{free} values being 16.6% and 19.9% respectively. This new structure is compared to the FMO protein structures of *Prosthecocholoris aestuarii* 2K and *Chlorobium tepidum*. The early structures of FMO contained seven bacteriochlorophyll-*a* (BChl) molecules but the recent discovery that there is an eighth BChl molecule in *Ptc. aestuarii* 2K and *Cbl. tepidum* and now in *Pld. phaeum* requires that the energy transfer mechanism be reexamined. Simulated spectra are fitted to the experimental optical spectra to determine how the BChl molecules transfer energy through the protein. The inclusion of the eighth BChl molecule

within these simulations may have an impact on how energy transfer through FMO can be described.

In conclusion, a reliable method of purifying and crystallizing the SMNWT protein is detailed, the placement of the 8th BChl-*a* within the electron density and the implications on energy transfer within the FMO protein when the 8th BChl-*a* is included from the green sulfur bacteria *Pld. phaeum* is discussed.

ACKNOWLEDGMENTS

The author would like to express gratitude, respect and appreciation to Dr. Jim Allen. Dr. Allen allowed me to work freely and develop as a graduate student in the only way that I believe could have a worked for me. I would also like to thank Dr. Scott Lefler who graciously allowed me to teach for him over the last five years. I also owe a big thank you to Chenda Seng and Lisa Lauman for all the support, data analysis and discussions in the lab over my time as a graduate student. The entire Allen lab group has been amazing and I would like to thank all of them also.

I also need to thank Dr. Joe Stickler and Dr. Hilde van Gijssel of Valley City State University in North Dakota. Without them pushing me to become a better student at the undergraduate level and then a giving me a ‘gentle’ nudge out the door into graduate school, I would never have considered earning a Ph.D.

Lastly, my family and friends have been an inspiration to me. My parents deserve a huge amount of the credit for any accomplishments that I have achieved and hope to achieve in the future. My friends have been a rock for me to lean on during the tough times and I can’t express how much they have meant to me. I would never be where I am today without the love, support and unending patience that everyone in my life has shown me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

Chapter 1

Introduction

Spinal Muscular Atrophy

Spinal Muscular Atrophy (SMA) is an autosomal recessive disorder that is characterized by the degeneration of the motor neurons in the spinal cord. SMA affects one in 6000 live births and is the leading genetic cause of infant mortality according to FSMA.org. *(1)*. The Families of SMA website states that SMA is classified as a relatively common rare disorder *(1)*. The disease has four classifications based on the severity of the disease and the time of onset of symptoms. Type I, also called Werdnig-Hoffmann disease, is the most severe condition with the onset of symptoms coming within six months of birth and death occurring within two years of age. Type 2 is the intermediate form with the onset of symptoms occurring within 18 months of birth. Type 3 is the least severe and is sometimes called Kugelberg-Welander disease. The symptoms can appear anywhere from 1 year to 18 years of age. There are variable degrees of weakness in the proximal limbs of people suffering from this type *(1).* Type 4 or adult onset SMA is characterized by the disease symptoms appearing during adulthood up to around the age of 65. Table 1.1 summarizes the four types of SMA and some of the symptoms associated with each type.

SMA causes muscles to progressively become weaker as the disease progresses. The muscles that are affected include the arms, legs,

and torso muscles. When torso muscles are affected, difficulty in breathing and swallowing are a common and dangerous part of SMA. While the proximal muscles are severely affected by the disease, brain function is normal. Many of the individuals affected by SMA have above average intelligence and very good social skills *(1)*

SMA is a disease that is unique to humans because humans are the only species to date that has been identified to carry two copies of a gene name *SMN* (*2*). The Survival of Motor Neuron (SMN) protein has been found in a number of different organisms including mice and fish but it has been found that these organisms contain only one copy of the *SMN* gene. In a study (*3*) that inactivated the *SMN* gene in mice, none of the embryos survived until birth. The study also found that mice have only one copy of *SMN* and it was proposed that the SMA disease can only express itself in humans because of the second copy that is found on chromosome 5. Mouse models have been developed that mimic the symptoms of SMA and can then be used for studying the disease (*4*).

Source of Spinal Muscular Atrophy

In 1990, two groups found that the cause of SMA could be mapped to an unstable portion of chromosome 5. (*5-7*) The gene discovered at that location was termed Survival Motor Neuron (*SMN*) and it was found that it encoded a protein that is 294 amino acids in length (*8*). The gene was later found to occur in duplicate on chromosome 5 with the two copies being named *SMN1* and *SMN2*. *SMN1* encodes the full length Survival of

Motor Neuron (SMN) protein that is fully functional. This full length protein (SMNWT) has a theoretical weight of 31.8kDa (*9*). *SMN2* encodes a truncated version of the SMN protein termed SMNΔ7. This truncated protein is a cause of SMA and the severity of symptoms is closely linked to the copy number of the *SMN2* that is on the chromosome (*10*). The gene product from *SMN2* has been termed SMNΔ7 because exon 7 is not spliced into the protein. There is however a low percentage (10-15%) of the protein produced from *SMN2* that is the full length SMNWT (*11*). This truncated version has 282 amino acids and a theoretical weight of 30.4kDa *(8)*. The SMNΔ7 protein is alternatively spliced because of a single base pair mutation that occurs in the DNA sequence of *SMN1* and *SMN2* (*11*). There are a total of five nucleotide changes between the two genes but a Cytosine to Thymine conversion at codon 280 has been implicated as the cause for the skipping of exon 7 and ultimately the cause of SMA symptoms. Interestingly, this conversion is a silent mutation when it comes to the amino acid sequence coding of the protein but the mutation still causes exon 7 to be left out when the exons are spliced together *(12)*. The SMNΔ7 form was found to be much less stable than the SMNWT form (*13*) which could contribute to the disease pathology also. More than 95% of all patients with SMA have either a deletion of *SMN1* or a conversion of *SMN1* to *SMN2* (*8*). The other 5% of patients that suffer from SMA have a number of small mutations that mainly occur in exons 3 and 6 (*14*). There are also number of other mutations that

occur in the other exons but none to date have been found in exon 5 (*15*). Table 1.2 shows all of the known mutations that have been found to cause SMA. There is an inverse relationship of the severity of SMA symptoms, and the amount of SMNWT present, which is based on whether the *SMN1* is deleted or if it has been converted to *SMN2.* If the copy number of *SMN2* is high in an individual, then the symptoms of SMA will be less severe. A schematic of the differences between the splicing events for the two genes is shown in figure 1.1. A comparison of the amino acid sequences for SMNWT and SMNΔ7 is shown in figure 1.2. In a recent paper, a new isoform of SMN has been found. This form exists in axons exclusively and consists of only exons 1-3 and a portion of intron 3 (*16*).

SMN complex and function

SMNWT is a ubiquitously expressed protein that is found throughout the body in both the nucleus and the cytoplasm. The protein has been localized *in vivo* to structures found in the cytoplasm termed gems (*17).* The best described complex that SMN is a part of consists of 9 components: SMN, Gemin2-Gemin8 and a protein called UNR-interacting protein (UNRIP) (*18-23*). Although the components are known, the exact stoichiometry of the complex's components is still unknown. SMN interacts directly with Gemin2, Gemin3, Gemin5, and Gemin7 (*22*). The protein Gemin2 has been shown to interact strongly with SMN and is also an essential component of the SMN complex (*24*). Gemin2 interacts with the N-terminal of the SMN protein and is also involved in the

oligomerization of the SMN complex (*24*). Both SMN and Gemin2 have been found to contain self-association domains. Gemin4 doesn't bind directly to SMN but is a part of the SMN complex through an interaction with Gemin3 *(25)*. Gemin3 has been classified as a DEAD box RNA helicase (*26*). Gemin5 has recently been implicated as the part of the SMN complex that interacts directly with the small nuclear RNA (snRNA) sequences that are involved in small nuclear ribonucleoprotein (snRNP) biogenesis (*27*). The importance of snRNA and snRNPs to the pathology of SMA will be discussed below. Gemin6 has been found to only interact directly with Gemin7. Gemin6 and Gemin7 have had their structures solved and it was found that they are similar structurally to Sm proteins. This may play a part in the recruitment of Sm proteins to the SMN complex during snRNP biogenesis (*28*). Gemin8 binds to SMN directly and also with the Gemin6/Gemin7 dimer (*23, 29*).

The SMN complex is involved in the assembly of snRNPs within the cell. Each snRNP consists of one snRNA sequence, seven Sm proteins (B, D1, D2, D3, E, F, and G) and another set of proteins that are specific to that particular snRNP (*18*). The Sm proteins form a heptameric ring from three substructures that consist of Sm proteins binding to each other, specifically B-D3, D1-D2, and E-F-G (*30*).The snRNPs have been implicated in recognition of splice sites and the removal of introns from pre-mRNA sequences *(31*). The specific snRNAs identified that interact

with the Sm protein ring and the SMN complex are U1, U2, U4, U5, U11 or U12 (*30*).

While snRNP biogenesis can occur in either the nucleus or the cytoplasm of a cell, the snRNPs can only carry out the splicing of pre-mRNA within the nucleus (*18*). The snRNPs that are generated in the cytoplasm need to be imported into the nucleus, which is one of the crucial elements for generating the necessary pre-mRNA machinery within the nucleus (*32, 33*). The biogenesis requires that the heptameric ring of Sm proteins form around a conserved sequence of the snRNAs called the Sm site (*18*). Specifically the Sm site is a short single stranded sequence of the snRNA that is Uridine rich (*20*). It is interesting to note that the formation of the Sm core *in vitro* occurs spontaneously but the *in vivo* assembly reaction is ATP dependent (*20, 34, 35*). This difference between the ATP dependency of the *in vitro* and *in vivo* formation of the Sm core may be due to the fact that when the core is formed *in vitro* there is no competing RNA sequences. The *in vivo* formation reaction of the Sm core likely needs the SMN complex to correctly target and bind the snRNA sequences needed (*20, 36-38*). There is also evidence that the mutations that cause SMA are also a source of reduced binding ability of the snRNP biogenesis proteins *(39)*. Figure 1.3 shows a schematic of the sequence of events required for snRNP biogenesis. A representation of the SMN complex, Sm proteins and snRNA is shown in figure 1.4. To date, there is

no structure of how these three components fit together during the biogenesis of snRNPs.

There are a number of regions on SMNWT protein that have had their functionality characterized over the years. Exon2 can be divided into two separate regions, 2a and 2b. Exon 2a was found to interact with RNA (*40, 41*) and the exon2b region has been implicated as one of the self association domains of the protein along with another self association domain located in exon 6 (*42, 43*). Many of the mutations that cause SMA are in close proximity to exon 6, which would suggest that the dimerization or oligomerization the protein is important for its function *in vivo*. SMNWT also associates strongly with Gemin2 (formerly SIP1). It was found that Gemin2 interacted specifically with the N-terminus of SMNWT (*44*). The only structural part of SMN that has been successfully solved is a small section of the SMN protein termed the Tudor domain (*45, 46*). It was first done with Nuclear Magnetic Resonance Imaging and was later completed using X-ray crystallography. The Tudor domain consists of amino acids 92-144, with the majority of residues falling into Exon 3 of the protein. The tudor domain was found to interact with Sm proteins (*47*), which are found in the snRNP's when bound to SMN (*47*). The point mutation within the tudor domain of E134K showed that the binding of Sm proteins was drastically affected (*45*). This point mutation is one of the Type I causing mutations in the SMN protein. The C-terminus of SMNWT has also been linked to nuclear targeting of the protein in a mouse model for SMA (*48*). A

construct that lacked a portion of the C-terminus was found at a reduced level within the motor neurons of the mice models and this truncation also affected the formation of gems (*48*).

SMN's role in Spinal Muscular Atrophy

How the SMN protein causes SMA is still unclear but there are two things known about SMN's role in the disease: snRNP biogenesis has decreased activity without SMNWT and motor neurons are especially sensitive to the loss of the native function that the SMN complex performs (*49*). There are two different views on how these mutations and/or deletions can either affect motor neurons or the snRNP biogenesis. One view is that when snRNP formation is upset at a crucial developmental period, the genes that are responsible for motor neuron growth are also disrupted (*39, 50, 51,*). The other view currently being investigated is that SMN has a specific function in axons and this function is altered in SMA patients (*39, 50, 51, 52, 53, 54. 55*). These two views are similar and could be linked but there is not a consensus as to which is correct or if both are valid.

Photosynthesis

Plants and a number of bacteria species can harvest light energy from the sun and turn this energy into a usable form by a process called photosynthesis. The light energy that is used in this process must first be captured by the organism and then converted to a usable form. Chlorophyll-based photosynthetic organisms use structures called

antennas to aid the capture of light. An antenna does the same basic job as a satellite dish that provides TV programming to the average home. The satellite dish must harvest the scrambled TV signal and funnel the signal to the satellite box that is attached to a TV. The satellite box that is connected to the TV then converts the scrambled signal to a form that the TV can then project. An antenna in a plant or green sulfur bacteria acts in much the same way. The antennas harvests light energy and funnels this energy to the reaction center of the organism. This funneling of energy and subsequent transfer also quenches or deactivates the excited energy of the absorbed light so that it does not damage the reaction center (*56*).

There are many different kinds of antennas that have been discovered and described (*56*). Many photosynthetic organisms are classified and grouped by the kind of antennas that are present in those organisms. The antenna complex that will be focused on in this part of the chapter is complex called the Fenna Matthews Olson (FMO) protein (*57, 58*). FMO is a protein-pigment complex that is found in many species of green sulfur bacteria (*59*). This protein lies between two other structures (Figure 1.5), the chlorosome and the reaction center, and will be discussed briefly.

The cholorosome is one of the largest known complexes in nature (*60*). It is one of the key components that allow green sulfur bacteria to carry out photosynthesis in environments that have very little light. Cholorsomes are found in anoxygenic phototrophs and have been found

as deep as 100 meters below the surface of the black sea (*61*). The extremely large size of the complex, 100 – 200nm in length and up to 60nm in diameter, allows it to be seen by electon microscopy (*60*). An interesting aspect of the chlorosome is that the pigments are largely assembled without the aid of a protein scaffold. Most green plant light-harvesting complexes need a protein scaffold to ensure that the pigments are arranged correctly and can harvest light efficiently (*60*). The cholorsome transfers the light energy that it has absorbed to a complex called the Fenna Matthews Olson (FMO) protein.

FMO is a trimeric protein that contains eight bacteriachlorophyll-*a* within the folded protein (*62*). It has been extensively studied because of it water solubility. In a recent paper by Wen et. al. in 2009, it was demonstrated how the FMO protein was orientated compared to the chlorosome and the reaction center in green sulfur bacteria. The orientation of FMO has been a topic of debate for some time. The implication of the FMO orientation has an effect on how energy transfer simulations are treated. The simulations attempt to fit simulated spectra to the various experimental optical spectra that have been measured for the FMO protein(*64-66*). The fitting of spectra is a method for elucidating how the energy is transferred from one pigment to another within an antenna complex (*64*).

Once energy has been transferred from the chlorosome to FMO, the energy subsequently moves through the FMO complex and enters the

reaction center. The function of the reaction center is to turn light energy into chemical energy that the organism can use for survival. This function sustains nearly all life on earth at this point in history. By converting sunlight into a usable form, the reaction center and other photosynthetic complexes are responsible for growing trees for paper in lab books, the cotton that the US dollar is printed on and the energy that human beings use when reading and processing thoughts. The reaction center complexes of both plants and bacteria have been and will continue to be studied intensely for the foreseeable future.

Energy transfer concepts

How does this energy that is needed for creating an excited molecule make it to the reaction center? The answer lies in energy transfer. The higher energy wavelengths are absorbed the antenna complexes that are most distant from the reaction center. As the energy is passed from pigment to pigment, a small amount of the energy is lost as heat with each transfer (*56*). This downward trend of energy intensity is also helpful with forcing the energy to a focal point. This ‘funneling’ of energy also reduces how much energy escapes the system as wasted energy. There are two basic concepts call Föster energy transfer and exciton coupling. Both of these concepts will be discussed briefly.

The Föster theory of energy transfer deals mainly with weakly coupled molecules and in our case the pigments within the antenna complex. The theory was developed by Thomas Förster and has been

extensively studied for over 50 years (*67*). The Förster energy transfer theory is a nonradiative energy transfer that does not involve electrons (*68*). Energy is transferred from a donor molecule to an accepter molecule as shown in Figure 1.6. This theory has been the dominate description of energy transfer between molecules for distances of 20Å to 100Å (*69*). According to the equation:

$$k_e = k_f (\frac{R_o}{R})^6 \qquad (1.1)$$

the energy transfer is heavily dependent on the distance term R^6. As the distance between the two molecules increases, the transfer of energy becomes less efficient. R_o is called the critical distance and is the distance where energy transfer is at 50% efficiency (*56*). The distances between pigments in antenna complexes is normally at distances that discourage the transfer of electrons but close enough to allow efficient transfer of energy (*56*). Föster energy transfer is thought to dominate the larger distances but if the pigments are brought closer together, the theory by itself beaks down and other factors must be considered.

The second way that energy transfer can be viewed is by a process called exciton coupling. This is largely a qualitative process as it is presently understood (*70*). Exciton coupling occurs when two molecules, antenna pigments in our case, are within about 10Å of each other.

I would like to thank Chenda Seng for her valuable contributions in chapters 3 and 4. I would also like to thank Lisa Lauman and Chad Simmons for their contributing work in chapter 4.

	Onset	Symptoms
Type 1 (Werdnig-Hoffman)	Within 6 months of birth	• Death usually occurs within 2 years • Feeding and breathing problems
Type 2	18 months of earlier	• Problems breathing and moving
Type 3 (Kugelberg-Welander)	1 year to 18 years of age	• Variable degrees of weakness in the proximal limbs
Type 4	Adulthood to around 65 years of age	• Weakness of proximal limbs

Table 1.1. The four different types of Spinal Muscular Atrophy, the time of onset and what some of the common symptoms are displayed here. The distinction between the types is not always straightforward with the diagnosis being somewhat subjective.

Mutation	Exon involved	SMA type
A2G	1	III
D30N	2a	II
D44V	2a	III
W92S	3	I
V94G	3	II
G95R	3	III
A111G	3	I
I116F	3	I
Y130C	3	III
E134K	3	I
Q136E	3	I
A188S	4	I
P245L	6	III
L260S	6	II
S262G	6	III
S262I	6	III
M263R	6	I
M263R	6	II
S266P	6	II
Y272C	6	II
H273R	6	II
T274I	6	III
G275S	6	III
G279C	7	I
G279V	7	I

Table 1.2. The mutations that cause SMA are listed here. Most of the missense mutations fall in exons 3 and 6. This table is adapted from *Burghes et al 2009*.

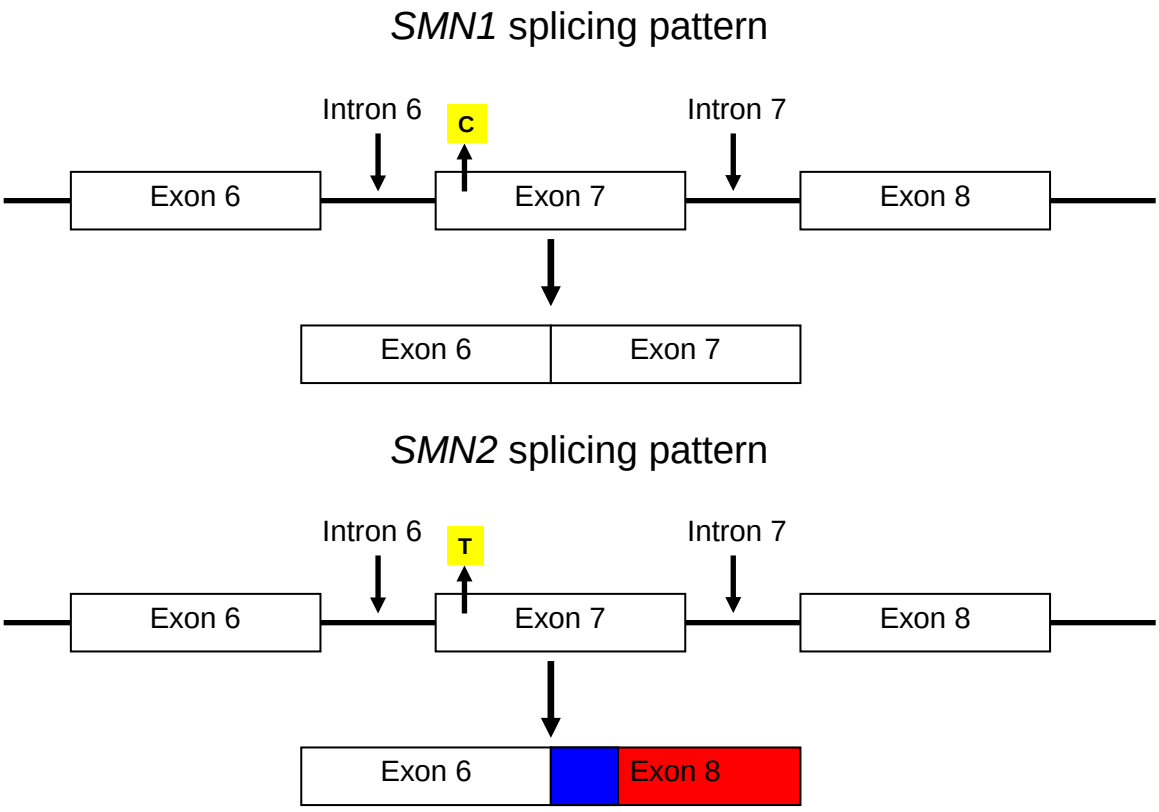

Figure. 1.1: This figure shows the gene splicing pattern for the genes *SMN1* and *SMN2.* The abberent splicing in *SMN2* is caused by the cytosine to thyamine change highlighted in yellow. *SMN2* also transcribes a part of exon 8 (blue) while *SMN1* does not and is the reason for SMNWT being 12 amino acids longer than SMNΔ7.

SMN WT

```
MAMSSGGSGG GVPEQEDSVL FRRGTGQSDD SDIWDDTALI KAYDKAVASF
KHALKNGDIC ETSGKPKTTP KRKPAKKNKS QKKNTAASLQ QWKVGDKCSA
IWSEDGCIYP ATIASIDFKR ETCVVVYTGY GNREEQNLSD LLSPICEVAN
NIEQNAQENE NESQVSTDES ENSRSPGNKS DNIKPKSAPW NSFLPPPPPM
PGPRLGPGKP GLKFNGPPPP PPPPPPHLLS CWLPPFPSGP PIIPPPPPIC
PDSLDDADAL GSMLISWYMS GYHTGYYMGF RQNQKEGRCS HSLN
```

SMN ND7

```
MAMSSGGSGG GVPEQEDSVL FRRGTGQSDD SDIWDDTALI KAYDKAVASF
KHALKNGDIC ETSGKPKTTP KRKPAKKNKS QKKNTAASLQ QWKVGDKCSA
IWSEDGCIYP ATIASIDFKR ETCVVVYTGY GNREEQNLSD LLSPICEVAN
NIEQNAQENE NESQVSTDES ENSRSPGNKS DNIKPKSAPW NSFLPPPPPM
PGPRLGPGKP GLKFNGPPPP PPPPPPHLLS CWLPPFPSGP PIIPPPPPIC
PDSLDDADAL GSMLISWYMS GYHTGYYMEM LA
```

Fig. 1.2. This figure shows an amino acid sequence comparison of the SMNWT and SMNΔ7 proteins. The mutation that affects the proteins activity occurs in the seventh and eighth exons. The difference between the two sequences is highlighted.

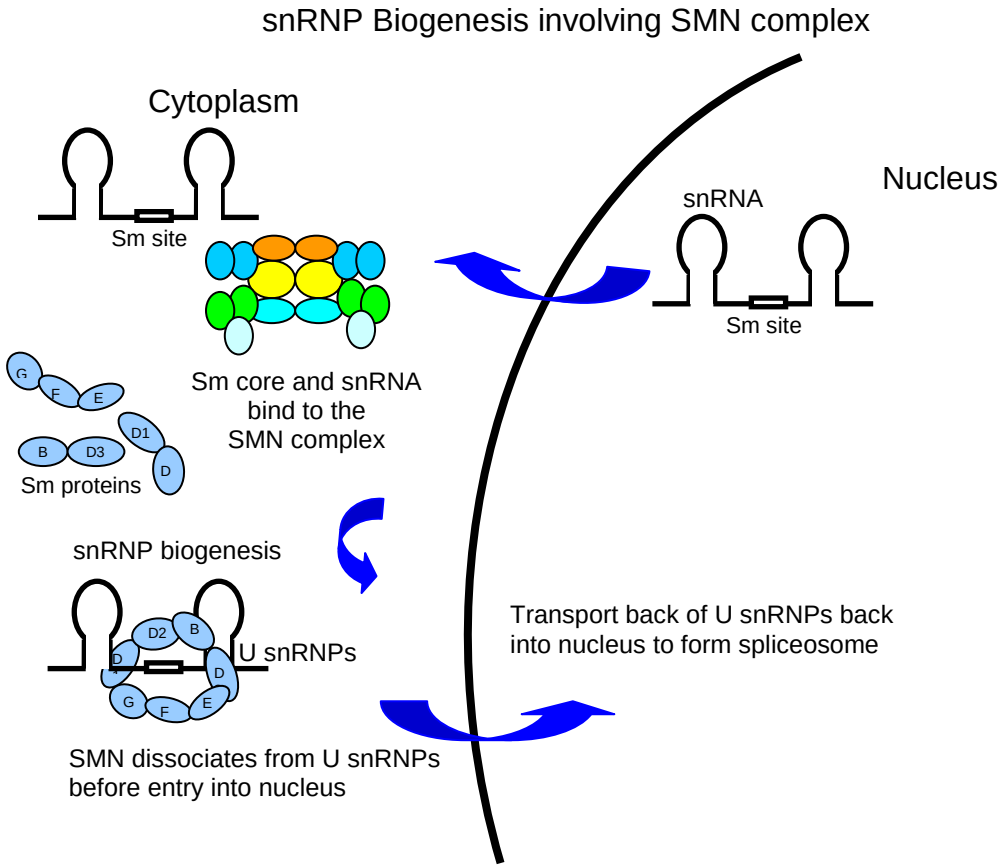

Fig. 1.3 This schematic shows how SMNWT interacts in the process of snRNP biogenesis which has been shown to be important for mRNA production (ref 12,14)

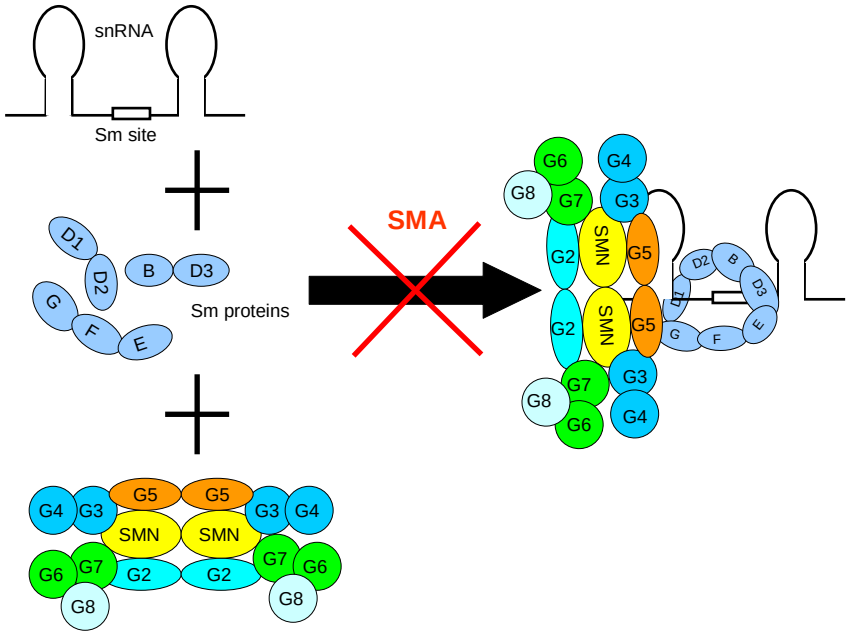

Fig. 1.4. A schematic of how the SMN complex guides and regulates the biogenesis of snRNPs. The Sm proteins won't form the heptameric ring *in vivo* without the help of the SMN complex. Gemin5 has been implicated as the protein of the complex that contacts the specific snRNA molecule (Battle et al 2006). The reaction is reduced in SMA patients

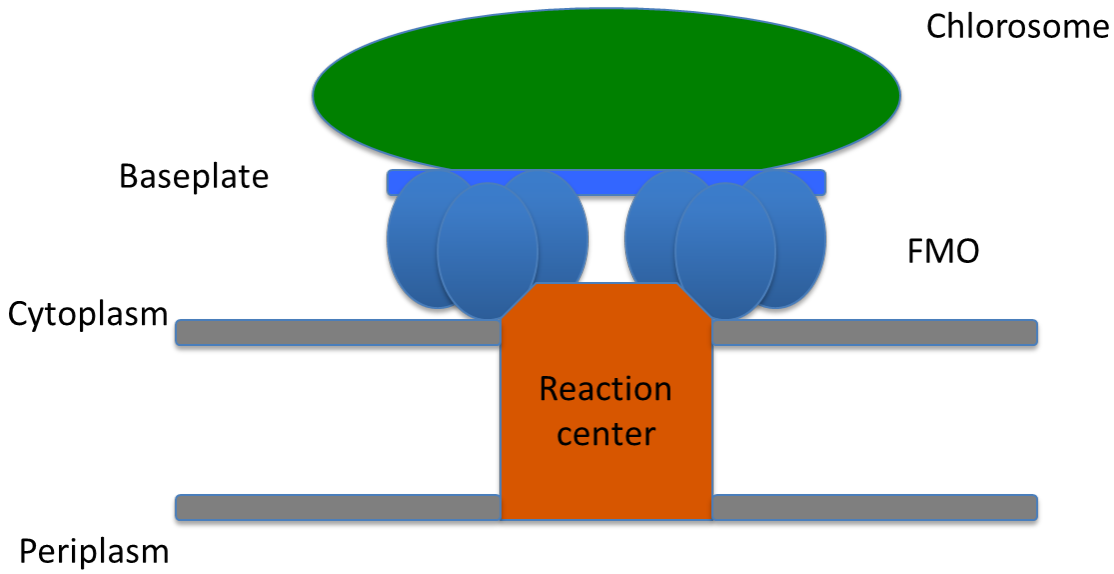

Figure 1.5. The orientation of the chlorosomes, Fenna Matthews Olson protein and the reaction center within green sulfur bacteria is shown schematically. The orientation of the FMO protein to the chlorosomes has been elucidated. Chlorosomes harvest light and subsequently funnel through FMO and then into the reaction center (*Wen et al 2008).* This figure was adapted from *Wen et al 2008.*

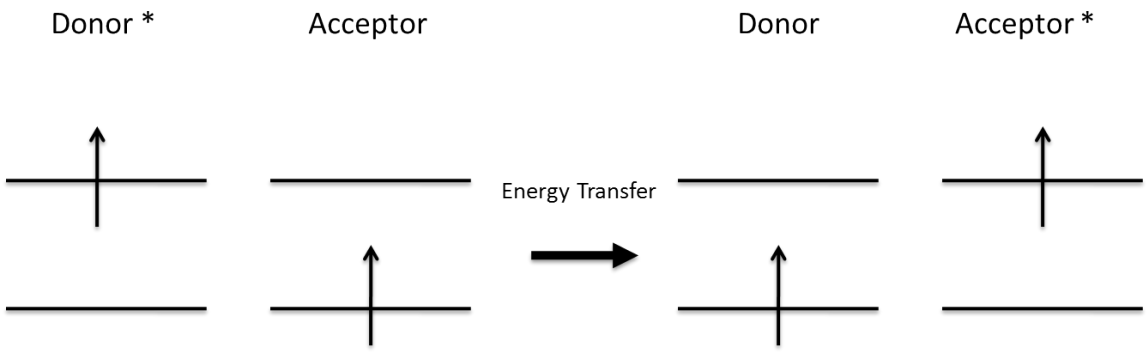

Figure 1.6. In this basic schematic of how Förster energy transfer occurs, a donor molecule is in the excited state (shown as a *). The acceptor molecule is at some relatively close distance away. When the energy transfer occurs, there is no transfer of electrons, only a transfer of energy which causes the acceptor molecule to move into an excited state. The energy transfer is dependent on the distance (R^6) between molecules and is shown in equation 1.1.

References:

1. www.fsma.org accessed 2010

2. Rochette, C. F., Gilbert, N., and Simard, L. R. (2001) *SMN* gene duplication and the emergenece of the *SMN2* gene occureed in distinct hominids: *SMN2* is unique to *Homo sapiens. Hum. Genet. 108*, 255-266.

3. Schrank, B., Gotz, R., Gunnersen, J. M., Ure, J. M., Toyka, K. V., Smith, A. G., and Sendtner, M. (1997) Inactivation of the survival motor neuron gene, a candidate gene for human spinal muscular atrophy, leads to massive cell death in early mouse embryos. *Proc. Natl. Acad. Sci. U. S. A. 94***,** 9920-9925.

4. Hsieh-Li, H. M., Chang, J. G., Jong, Y. J., Wu, N. M., Tsai, C. H., and Li, H. (2000) A mouse model for spinal muscular atrophy. *Nat. Genet. 24,* 66-70.

5. Brzustowicz, L. M., Lehner, T., Castilla, L. H., Penchaszadeh, G. K., Wilhelmsen, K. C., Daniels, R., Davies, K. E., Lepper, M., Ziter, F., Wood, D. , Dubowitz, V., Zerres, K., Hausmanowa-Petrusewicz, I., Ott, J., Munsat, T. L., & Gilliam, T. C. (1990) Genetic mapping of chronic childhood-onset spinal muscular atropy to chromosome 5q11.2-13.3. *Nature 344,* 540-541.

6. Melki, J., Abdelhak, S., Sheth, P., Bachelot, M. F., Burlet, P., Marcadet, A., Aicardi, J., Barois, A., Carrier, J. P., Fardeau, D., Ponsot, G., Billette, T., Angelini,C., Barbosa, C., Ferriere, G., Lanzil, G., Ottolini, A., Babron, M. C., Cohen, D., Hanauer, A., Clerget-Darpoux, F., Lathrop, M., Munnich, A., and Frezal, J. (1990) Gene for chronic proximal spinal muscular atrophies maps to chromosome 5q. *Nature 344,* 767-768.

7. Gilliam, T. C., Brzustowicz, L. M., Castilla, L. H., Lehner, T., Penchaszadeh, G. K., Daniels, R. J., Byth, B. C., Knowles, J., Hislop, J. E., Shapira, Y., Dubowitz, V., Munsat, T.L., Ott, J., and Davies, K. E. (1990) Genetic homogeneity between acute and chromic forms of spinal muscular atrophy. *Nature 345***,** 823-825.

8. Lefebvre, S., Buglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., Le Pasilier, D., Frezai, J., Cohen, D., Weissenbach, J., Munnich, A., and Melki, J. (1995) Identification and Characterization of a Spinal Muscular Atrophy-Determining Gene. *Cell 80,*155-165.

9. www.Uniprot.org accessed 2010

10. Lefebrvre, S., Burlet, P., Liu, Q., Bertrandy, S., Clermont, O., Munnich, A., Dreyfuss, G., and Melki, J. (1997) Correlation between severity and SMN protein level in spinal muscular atrophy. *Nat. Genet. 16,* 265-269.

11. Lorson, C. L., Hahnen. E., Androphy, E. J., and Wirth, B. (1999) A single nucleotide in the *SMN* gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl. Acad. Sci. U. S. A. 96,* 6307-6311.

12. Kashima, T., and Manley, J. L. (2003) A negative element in *SMN2* exon 7 inhibits splicing in spinal muscular atrophy. *Nature genet. 34*, 460-463.

13. Lorson, C. L., and Androphy, E. J. (2000) An exonic enhancer is required for inclusion of an essential exon in the SMN-determining gene *SMN.Hum. Mol. Genet. 9,* 259-265.

14. Alias, L., Bernal, S., Fuentes-Prior, P., Barceló, M. J., Also, E., Martínez-Hernández, R., Rodríguez-Alvarez, F. J. , Martín, Y., Aller, E., Grau, E., Peciña, A., Antiñolo, G., Galán, E., Roas, A. L., Fernández-Burriel, M., Borrego, S., Millán, J. M., Hernández-Chico, C., Baiget, M., and Tizzano, E. F. (2009) Mutation update of spinal muscular atrophy in Spain: molecular characterization of 745 unrelated patients and identification of four novel mutation in the *SMN1* gene. *Hum. Genet. 125*, 29-39.

15. Burghes, A. H. M., and Beattie, C. E. (2009) Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick? *Nat. Rev. Neurosci. 10*, 597-609.

16. Setola, V., Terao, M., Locatelli, D., Bassanini, S., Garattini, E., and Battaglia, G. (2007) Axonal-SMN (a-SMN), a protein isoform of the survival motor neuron gene, is specifically involved in axonogenesis. *Proc. Natl. Acad. Sci. U. S. A*. *140,* 1959-1964.

17. Liu, Q., and Dreyfuss, G. (1996) A Novel nuclear structure containing the survival of motor neurons protein. *EMBO J. 15,* 3555-3565.

18. Pellizzoni, L. (2007) Chaperoning ribonucleoprotein biogenesis in health and disease. *EMBO Rep.8*, 340-345.

19. Gubitz, A.K., Feng, W., and Dreyfuss, G. (2004) The SMN complex. *Exp. Cell Res. 296,* 51-56.

20. Pellizzoni, L., Yong, J., and Dreyfuss, G. (2002) Essential Role for the SMN Complex in the Specificity of snRNP Assembly. *Science 298*, 1775-1779.

21. Meister, G., Buhler, D., Laggerbauer, B., Zobawa, M., Lottspeich, F., and Fischer, U. (2000) Characterization of a nuclear 20S complex containing the survival of motor neurons (SMN) protein and a specific subset of spliceosomal Sm proteins. *Hum. Mol. Genet. 9,* 1977-1986.

22. Otter, S., Grimmler, M., Neuenkirchen, N., Chari, A., Sickmann, A., and Fischer, U. (2007) A Comprehensive Interaction Map of the Human Survival of Motor Neuron (SMN) Complex. *J. Biol. Chem. 282,* 5825-5833.

23. Carissimi, C., Saieva, L., Baccon, J., Chiarella, P., Maiolica, A., Sawyer, A., Rappsilber, J., and Pellizzoni, L. (2006) Gemin8 is a Novel Component of the Survival Motor Neuron complex and Functions in Small Nuclear Ribonucleoprotein Assembly. *J. Biol. Chem. 281,* 8126-8134.

24. Ogawa, C., Usui, K., Aoki, M., Ito, R., Itoh, M., Kai, C., Kanamori-Katayama, M., Hayashizaki, Y., and Suzuki, H. (2007) Gemin2 Plays an Important role in Stabilizing the Survival of Motor Neuron Complex. *J. Biol. Chem. 282,* 11122-11134.

25. Charroux, B., Pellizzoni, L., Perkinson, R. A., Yong, J., Shevchenko, A., Mann, M., and Dreyfuss, G. (2000) A Novel Component of the Smn Complex That Is Found in Both Gems and Nucleoli. *J. Cell Biol. 148*. 1177-1186.

26. Charroux, B., Pellizzoni, L., Perkinson, R. A., Shevchenko, A., Mann, M., and Dreyfuss, G. (1999) Gemin3: A Novel DEAD Box Protein that Interacts with SMN, the Spinal Muscular Atrophy Gene Product, and Is a Component of Gems. *J. Cell Biol. 147,* 1181-1193.

27. Battle, D. J., Lau, C-K., Wan, L., Deng, H., Lotti, F., and Dreyfuss, G. (2006) The Gemin5 Protein of the SMN Complex Identifies snRNAs. *Mol. Cell 23,* 273-279.

28. Ma, Y., Dostie, J., Dreyfuss, G., and Van Duyne, G. D. (2005) The Gemin6-Gemin7 Heterodimer from the Survival of Motor Neurons Complex Has an Sm Protein-like Structure. *Structure 13*, 883-892.

29. Carissimi, C., Saieva, L., Gabanella, F., and Pellizzoni, L. (2006) Gemin8 is required for the Architecture and Function of the Survival Motor Neuron Complex. *J. Biol. Chem. 281,* 37009-37016.

30. Raker, V. A., Plessel, G., and Lührmann, R. (1996) The snRNP core assembly pathway: identification of stable core protein heteromeric coplexes and an snRNP subcore particle *in vitro*. *EMBO J. 15*, 2256-2269.

31. Raker, V. A., Hartmuth, K., Kastner, B., and Lührmann, R. (1999) Spliceosomal U snRNP core Assembly: Sm Proteins Assemble onto an Sm Site RNA Nonanucleotide in a Specific and Thermodynamically Stable Manner. *Mol. Cell Biol.19*, 6554-6565.

32. Paushkin, S., Gubitz, A.K., Massenet, S., and Dreyfuss, G. (2002) The SMN complex, an assemblysome of ribonucleoproteins. *Curr. Opin. Cell Biol. 14,* 305-312.

33. Pellizzoni, L., Kataoka, N., Charroux, B., and Dreyfuss, G. (1998) A Novel Runction for SMN, the spinal Muscular Atophy Disease Gene Product, in Pre-mRNA splicing. *Cell 95,* 615-624.

34. Meister, G., Buhler, D., Pillai, R., Lottspeich, F., and Fischer, U. (2001) A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. *Nat. Cell Biol. 3,* 945-949.

35. Meister, G., and Fischer, U. (2002) Assisted RNP assembly: SMN and PRMT5 complexes cooperate in the formation of spliceosomal UsnRNPs. *EMBO J. 21,* 5853-5863.

36. Yong, J., Pellizzoni, L., and Dreyfuss, G. (2002) Sequence-specific interaction of U1 snRNA with the SMN complex. *EMBO J. 21*, 1188-1196.

37. Golembe, T. J., Yong, J., Battle, D. J., Feng, W., Wan, L., and Dreyfuss, G. (2005) Lymphotropic *Herpesvirus simiri* Uses the SMN Complex To Assemble Sm Cores on Its Small RNAs. *Mol. Cel. Biol. 25*, 602-611.

38. Golembe, T. J., Yong, J., and Dreyfuss, G. (2005) Specific Sequence Features, Recognized by the SMN Complex Identify

snRNAs and Determine Their Fate as snRNPs. *Mol. Cel. Biol. 25,* 10989-11004.

39. Pellizzoni, L., Charroux, B., and Dreyfuss, G. (1999) SMN mutants of spinal muscular atrophy patients are defective in binding to snRNP proteins. *Proc. Natl. Acad. Sci. U. S. A. 96,* 11167-11172.

40. Bertrandy, S., Burlet, P., Clermont, O., Huber, C., Fondrat, C., Thierry-Mieg, D., Munnich, A., and Lefebvre, S. (1999) The RNA-binding properties of SMN: deletion analysis of the zebrafish orthologue defines domains conserved in evolution. *Hum. Mol. Genet. 8,* 775-782.

41. Lorson, C.L., and Androphy, E.J. (1998) The domain encoded by exon 2 of the survival motor neuron protein mediates nucleic acid binding. *Hum. Mol. Genet. 7,* 1269-1275.

42. Young, P. J., Man, N. T., Lorson, C.L., Le, T. T., Androphy, E. J., Burghes, A. H. M., and Morris, G. E. (2000) The exon 2b region of the spinal muscular atrophy protein, SMN, is involved in self-association and SIP1 binding. *Hum. Mol. Genet. 9,* 2869-2877.

43. Lorson, C. L., Strasswimmer, J., Yao, J-M., Baleja, J.D., Hahnen, E., Wirth, B., Le, T., Burghes, A.H.M., and Androphy, E.J. (1998) SMN Oligomerization defect correlates with spinal muscular atrophy severity. *Nat. Genet. 19,* 63-66.

44. Liu, Q., Fischer, U., Wang, F., and Dreyfuss, G. (1997) The Spinal Muscular Atrophy Disease Gene Product, SMN, and Its Associated Protein SIP1 Are in a Complex with Spliceosomal snRNP Proteins. *Cell 90,* 1013-1021.

45. Selenko, P., Sprangers, R., Stier, G., Buhler, D., Fischer, U., and Sattler, M. (2001) SMN Tudor domain structure and its interaction with the Sm proteins. *Nat. Struct. Biol. 8,* 27-31.

46. Sprangers, R., Groves, M. R., Sinning, I., and Sattler, M. (2003) High-resolution X-ray and NMR Structures of the SMN Tudor Domain: Conformational Variation in the Binding Site for Symmetrically Dimethylated Arginine Residues. *J. Mol. Biol. 327,* 507-520.

47. Bühler, D., Raker, V., Luhrmann, R., and Fischer, U. (1999) Essential role for the tudor domain of SMN in spliceosomal U snRNP assembly: implications for spinal muscular atrophy. *Hum. Mol. Genet. 8,* 2351-2357.

48. Frugier, T., Tiziano, F. D., Cifuentes-Diaz, C., Miniou, P., Roblot, N., Dierich, A., Le Meur, M., and Melki J. (2000) Nuclear targeting defect of SMN lacking the C-terminus in a mouse model of spinal muscular atrophy. *Hum. Mol. Genet. 9,* 849-858.

49. Lorson, C. L., Rindt, H., and Shababi, M. (2010) Spinal muscular atrophy : mechanisms and therapeutic strategies. *Hum. Mol. Genet. 19*, R111-R118.

50. Eggert, C., Chari, A., Laggerbauer,B., and Fischer, U. (2006) Spinal muscular atrophy: the RNP connection. *TRENDS Mol. Med. 12,* 113-121.

51. Gabanell, F., Butchbach, M. E. R., Saieva, L., Carissimi, C., Burghes, A. H. M., and Pellizzoni, L. (2007) Ribonucleoprotein Assembly Defects Correlate with Spinal Muscular Atrophy Severity and Preferentially Affect a Subset of Spliceosomal snRNPs. *PLoS One 2,* e921.

52. Rossoll, W., Jablonka, S., Andreassi, C., Kroning A-K., Karle, K., Monani, U. R., and Sendtner, M. (2003) Smn, the spinal muscular atrophy-determining gene product, modulates axon growth and localization of β-actin mRNA in growth cones of motoneurons. *J. Cell Bio. 163*, 801-812.

53. McWhorter, M. L., Monani, U. R., Burghes, A. H. M., and Beattie, C. E. (2003) Knockdown of the survival motor neuron (Smn) protein in zebrafish causes defects in motor axon outgrowth and pathfinding. *J Cell Biol 162,* 919-931.

54. Carrel, T. L., McWhorter, M. L., Workman, E., Zhang, H., Wolstencroft, E. C., Lorson, C., Bassell, G. J., Burghes, A. H. M., and Beattie, C. E. (2006) Survival Motor Neuron Function in Motor Axons is Independent of Functions Required for Small Nuclear Ribonucleoprotin Biogenesis. *J. Neurosci. 26,* 11014-11022.

55. Fan, L., and Simard, L. R. (2002) Survival motor neuron (SMN) protein: role in neurite outgrowth and neuromuscular maturation during neuronal differentiation and development. *Hum. Mol. Genet. 11,* 1605-1614.

56. Blankenship, R. B. (2002) Molecular Mechanisms of Photosynthesis *London: Blackwell Science*

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

58. Matthews, B. W., Fenna, R. E., Bolognesi, M. C., and Schmid, M. F. (1979) Structure of a bacteriochlorophll *a*-protein from the green photosynthetic bacterium *Prosthecochloris aestuarii. J. Mol. Biol. 131*, 259-285.

59. Tsukatani, Y., Wen, J., Blankenship, R. E., and Bryant, D. A. (2010) Characterization of the FMO protein from the aerobic chlorophotroph, *Candidatus* Chloracidobacterium thermophilum. *Photosynth Res. 104,* 201-209.

60. Ganapathy, S., Oostergetel, G. T., Wawrzyniak, P. K., Reus, M., Chew, A. G. M., Buda, F., Boekema, D. A., Bryant, D. A., Holzwarth, A. R., and de Groot, H. J. M. (2009) Alternating *syn-anti* bacteriochlorophylls form concentric helical nanotubes in chlorosomes. *Proc. Natl. Acad. Sci. U. S. A. 106*, 8525-8530.

61. Frigaard, N-U., Chew, A. G. M., Julia, H. L., Maresca, J. A., and Bryant, D. A. (2003) *Chlorobium tepidum*: insights into the structure, physiology, and metabolism of a green sulfur bacterium derived from the complete genome sequence. *Photosyn. Res. 78*, 93-117.

62. Li, Y. F., Zhou, W., Blankenship, R. E., and Allen, J. P. (1997) Crystal structure of the bacteriochlorophyll *a* protein from *Chlorobium tepidum*. *J. Mol. Biol. 271*, 456-471.

63. Wen, J., Zhang, H., Gross, M. L., and Blankenship, R. E. (2009) Membrane orientation of the FMO antenna protein from *Chlorobaculum tepidum* as determined by mass spectroscopy based footprinting. *Proc. Natl. Acad. Sci. U. S. A. 106,* 6134—6139.

64. Brixner, T., Stenger, J., Vaswani, H. M., Cho, M., Blankenship, R. E., and Fleming, G. R. (2005) Two-dimensional spectroscopy of electronic couplings in photosynthesis. *Nature 434*, 625-628.

65. Johnson, S. and Small, G. (1991) Excited-state structure and energy-transfer dynamics of the bacteriochlorophyll a antenna complex from *Prosthecochloris aestuarii*. *J. Phys. Chem. 95*, 471—479.

66. Louwe, R. J. W., Vrieze, J., Aartsma, T. J., and Hoff, A. J. (1997) Toward an integral interpretation of the optical steady-state spectra

of the FMO-complex of *Prosthecochloris aestuarii.* 1. An investigation with linear-dichroic absorbance-detected magnetic resonance. *J. Phys. Chem. 101,* 11273—11279.

67. Förster, T. (1960) Transfer Mechanisms of Electronic Excitation Energy. *Rad. Res.Supplement. 2*, 326-339.

68. Selvin, P. R. (1995) Fluorescence Resonance Energy Transfer. *Methods Enzymol. 246.* 300-344.

69. Jong, S., Newton, M. D., and Silbey, R. J. (2004) Multichromophoric Förster Resonance Energy Transfer. *Phys. Rev. Lett. 92,* 218301.

70. van Amerongen, H. , van Grondelle, R. , and Valkunas, L. (2000) Photosynthetic Excitons. *World Scientific, London.*

Chapter 2

Purification, Folding and Crystallization of Survival of Motor Neuron protein

Recombinant Protein

Most protein that is used in research needs to be produced in sufficient quantities. Obtaining large quantities from the natural source that produces the protein presents some challenges and makes it difficult for doing structural studies. Another challenge that can be just as difficult is trying to separate the target protein form the other cellular proteins that are produced. One method that has become extremely popular is producing the target protein as what is called a recombinant protein. Recombinant protein is produced in mass quantities by an easily produced organism such as bacteria or yeast. This method has become common place and is one of biggest factors to the explosion of structural studies that have been done (*1*).

Whenever a recombinant protein is produce for a study, it must be purified from the other proteins in the host organism. There have been a number of different methods developed that allow efficient isolation of the target protein and here I will briefly talk about one method called affinity tags. Affinity tags are the method that has been chosen to isolate a number of different proteins including our SMN constructs.

There are basically two different classes of affinity tags that are in widespread use today. The first type of tag uses a peptide or a protein that

is fused to the target protein and recognizes a small molecule. The peptide or fused protein has a high affinity for some kind of a molecule that can be linked to a solid support. One very common example of this kind of tag is the poly-histidine sequence. The poly-histidine tag binds very efficiently to a variety of different transition metals (*2*). The second type of tag is a peptide or protein that recognizes a different protein or peptide that is immobilized on a solid support. One of the most commonly used proteins is called glutathione *S*-transferase (GST). The GST protein is commonly used as a fusion protein because it binds readily to the columns that have glutathione bound to the resin (*3*). There are other subcategories to this second type of tag. There are many different sizes of proteins that can be used and the target proteins characteristics can be useful in determining what size to use (*4*). Another subcategory to the second method is similar but it uses an antibody that is bound to the resin instead of a peptide/protein. The antibody is immobilized onto a resin and the antibody binds a specific protein that is again fused to the target protein (*5*). All of these techniques are well documented using the *Escherichia coli* bacterial system for protein expression. The SMN constructs that are produced all have a poly-histidine tag placed on the amino terminal of the protein. The isolation of the SMN constructs using this histidine tag will be discussed more below.

Protein from Inclusion bodies

When a eukaryotic protein is over-expressed in a bacterial system, which is usually the bacteria *Escherichia coli* (*E. coli*), the protein is often placed into inclusion bodies (*6-10*). Inclusion bodies were first described in *E. coli* in 1975 (*11*) but are still an active area of research today because of the widespread use and need of recombinant protein. Inclusion bodies are often used as a relatively pure source of a target recombinant protein. The general belief is that some of the protein that is placed into inclusion bodies may be correctly folded but a portion is also misfolded or unfolded (*12*). There are increasing accounts that state the protein found inside the inclusion bodies is just as active as the soluble fraction after refolding (*13*). There are a number of advantages that occur when an over-expressed protein is placed into inclusion bodies: 1) The level of protein production is extremely high, at times the production can raise above 30% of the entire cellular protein 2) Isolation of the inclusion bodies from the cell is relatively easy and straightforward 3) There is a lower level of degradation of the target protein because the inclusion bodies protect it from cellular processes 4) Inclusion bodies also shield the protein from the *E. coli* bacteria's proteases 5) The relative purity of the protein is a great advantage (*14*).

The general methodology for extracting the target protein involves isolating the inclusion bodies from the *E. coli* culture, solubilizing the inclusion bodies, which usually includes complete denaturation of all the protein, removal of the impurities and finally renaturation of the protein. As

in any system, there are also some key disadvantages that must be kept in mind. When the protein is completely denatured, it has a higher tendency to aggregate with itself then when it is in a partially folded state (*14*). The purity of the protein from inclusion bodies makes this problem more prevalent because proteins tend to aggregate with themselves and not other proteins (*15, 16*). Protein aggregation in general is a bad thing for proteins. There are some very debilitating diseases such as Alzheimer's disease and cystic fibrosis that have their root in either protein aggregation or in protein misfolding (*17*). When a protein is folded either *in vivo* or *in vitro*, the aggregates that are created are similar. This gives rise to the thought that the intermediate to blame for the aggregation is dependent on the amino acid sequence and not on the conditions of the folding environment (*18-21*)

During the isolation of the inclusion bodies, the *E. coli* cells must be lysed at some point to release the inclusion bodies and hence the protein of interest. The lysing of the cells can be achieved by a number of different methods. Three of the most popular methods are mechanical homogenation, french press and sonication. At this point there are also a number of hurdles that must be addressed so that the protein can retain its activity and native structure. After the cell is lysed, a number of cellular components have the potential of being adsorbed onto the surface of the inclusion bodies (*22*). This can be detrimental to the refolding process because other proteins contaminate the target protein and decrease the

refolding yield (*8, 23*). Another negative effect of the isolation process is that proteases can be copurified with the inclusion bodies during centrifugation (*24, 25*). Protein that has become aggregated is susceptible to proteolytic degradation both *in vivo* (*26*) and *in vitro* (*27, 28*). Many of these proteases can stay active *in vitro* even in the presence of the high concentrations of denaturants that are generally used to completely denature the protein before it can be refolded (*22*). The two most common denaturants used to denature proteins are urea and guanidium hydrochloride. At the high concentrations typically used to denature proteins, 8 molar and 6 molar respectively, little if any secondary structure is present (*29, 30*). Exactly how resistant a given recombinant protein is to proteolysis must be discovered empirically (*31*). There are some strategies that can be employed to aid in the purification such as adding lysozyme to aid in breaking the cells, reductant to stop nonnative sulfide bonds (*8* ,*32*) and ethylenediaminetetraacetic acid (EDTA) can be added to scavenge metal ions to stop unwanted oxidation of the protein (*8, 33*). There have been reports of up to 95% purity from *E. coli* inclusion bodies (*34*).

Protein Folding

The correct tertiary structure of a protein is of paramount importance when it comes to its function but how does a protein correctly fold *in vivo* and how does a denatured protein refold *in vitro*? The correctly folded structure of protein and how the folding occurs *in* vivo and *in vitro* has

been and will continue to be a leading area of research. Back in 1960's, Levinthal first introduced the famous Levinthal's paradox (*35*). This paradox says that a protein can't possibly sample at random, every conformation possible. A twist on a classical example demonstrates this. A protein of 101 amino acids has 3^{100} or 5×10^{47} different possible confirmations. If the protein is allowed to sample 50 billion confirmations every second (1 confirmation for every dollar of Bill Gates' fortune) the protein could sample 1.6×10^{18} confirmations per year. Even at this pace, to sample all possible conformations, it would still take 3×10^{29} years to find the correct conformation. This is obviously is not correct when some proteins can fold into their native fold on a roughly subsecond timescale if they are under 100 amino acids (*36*). This leads to another famous protein folding theory called Afinsen's dogma that was introduced over 40 years ago (*37*). This hypothesis states that the native configuration has the lowest Gibbs free energy value and that this energy value is determined entirely by the proteins amino acid sequence. These theories have been expanded upon and revised extensively over the last 40 years. In reality, the folding landscape of a protein is probably not one exact path that the protein must follow from the denatured state to the native fold but a series of different pathways. This view has been compared to a funnel (*38*) where there can be many different starting points and multiple pathways to the single native conformation (Figure 2.1). The folding landscape pathway theory can be divided into two views depending on the size of the

protein. The first view says that smaller proteins, less than 100 amino acids, generally proceed to their native state in a two-step process that involves one intermediate (*39*). The second view is for the proteins that have over 100 amino acids. Generally it is believed that these proteins find their native conformation through more than one intermediate (*39*). The correct tertiary structure of a protein can contribute to its function *in vivo* and also its activity, for the case of an enzyme, *in vitro*. It is known that the protein fold is tied to many of its activities including the transport of other molecules, regulation and how a cell may differentiate during its development (*40*). The native fold of a protein is usually has the highest stability in terms of thermodynamics (*41*). The question of how a protein locates this stable fold from countless other conformations is not straightforward. Depending on whether the protein is inside a cell or in a test tube can complicate how any protein achieves its native fold. The folding of a protein has to compete against other processes such as misfolding and aggregation that are considered unfavorable when studying protein characteristics (*30*). *In vivo* a protein is produced when the mRNA is transcribed by the ribosome into the polypeptide chain. The *in vivo* process of folding the protein may start immediately while the ribosome is doing the transcription. The *in vivo* process of protein folding also has the advantage of native conditions and chaperone proteins that aid in folding and also hinder misfolding and aggregation (*43*). In recent years, new techniques that utilize fluorescence have been developed that

allow researchers to follow protein within the cell (*44*). Nature has evolved a number of methods to overcome some disadvantages that occur when a protein is produced and folded *in* vivo. Typically the protein concentration within a cell is on the order of 300-400 g/L (*43*). This very large concentration of protein means that folding protein must be kept away from other proteins to minimize aggregation. Groups of proteins called molecular chaperones have been discovered that aid a cell correctly folding the protein that it produces (*45*). Other *in vivo* folding aids include interacting proteins that help to stabilize the proteins and a biological pH that can aid in interactions between proteins involved in a complex. This has been demonstrated in the SMN protein complex. Gemin2 aids the SMN complex by increasing the stability of the whole complex (*46*).

When a protein is folded *in vitro*, other strategies must be employed to aid the refolding. Some target proteins can be produced in a soluble folded form. Some proteins have the disadvantage of having to be completely unfolded and subsequently refolded from the denaturing conditions. A number of methods have been developed to aid in the refolding of recombinant proteins that have been obtained from inclusion bodies. None of these methods have proved to be a universal method for all recombinant proteins and the refolding conditions for any given protein cannot be predicted from the amino acid sequence alone and must be optimized individually (*47*). Here I will briefly review some of the popular

methods that are used frequently in protein labs and then discuss the current method used for the refolding of the SMN constructs.

The simplest method used to refold recombinant proteins is dialysis or dilution (*29*). This method can be done using a few different approaches. The first approach for dialysis uses a large amount of buffer that is free of any denaturant. When a protein in denaturing conditions is diluted to a point where the denaturant drops to a low concentration, it effectively leaves the protein in a state that allows it to refold in this single step. The second approach is to use step-wise buffer exchange that slowly steps the denaturant concentration down to a point where the protein can fold into its native tertiary structure. The second approach may allow the protein to sample more of its folding landscape. By slowly stepping down the denaturant concentration, the protein is allowed to fold into local energy minima that may be incorrect. The low barrier of the energy minima, caused by the presence of the denaturant that favors unfolding, allows the protein to escape the local energy minima and sample other folds that are closer to the native global minima for the protein (*37, 49*). This native global minima usually represents the structure that has the highest thermodynamic stability *in vivo*. A third approach is similar to the step-wise dialysis method but it instead is a continuous exchange of the buffers. This method also slowly lowers the denaturant concentration which also allows more of the folding landscape to be sampled. The dilution dialysis method has a number of disadvantages.

The best refolding yields require dilute concentrations of less than 10 mg/mL. This low concentration of protein will lower the amount of aggregation but the low concentration makes using this method difficult for structural work (*29*). The buffer that is used to refold the protein by this method must also be carefully chosen. The correct refolding buffer can aid in further minimizing aggregation (*50*). Buffers are essential for maintaining a constant pH in almost all biological activities *in vivo*. Buffers also serve the same purpose *in vitro,* although the pH may need to be held at a value that is outside the normal value seen inside a cell. There are many different buffers but some of the most common used in biochemistry include 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid(HEPES), TRIS and phosphate. A buffer solution may also have a number of 'additives' also added to it. These additives can help to stabilize the protein during folding and after folding (*48).*

On column refolding of recombinant protein from *E. coli* has become a very popular method. It is fast and easily adaptable for many different proteins. There are three basic methods for on column refolding: 1) the denatured protein is immobilized on the column: 2) size exclusion chromatography is used to simultaneously remove the denatured and separate out contaminates from the solution: 3) folding catalyst/chaperones can be immobilized to the column to promote the refolding (*29*). All three methods have their own advantages and disadvantages. An example of this advantage vs. disadvantage is in the

size exclusion chromatography method. This method separates out contaminates according to size but does not ensure that the protein molecules are kept the appropriate distance from each other to allow correct refolding.

Protein Crystallization

Over 25 years ago, recombinant protein technology was in its beginning stages (*50*). This technique has been improved greatly over time and is the source of almost all protein used in structural biology studies today. The advancement of recombinant protein has made structural determination much more feasible. The first protein was that had its structure elucidated by X-ray crystallography was Myoglobin in 1958 (*51*). When the protein data bank (pdb) was first put together in 1971 it contained only seven structures (*52*) and on January 1, 2010 there were over 64,000 protein structures deposited. Although this number is slightly elevated due to some proteins being solved more than once due to crystals that diffract to better a resolution and the same protein being solved but from different organisms. At the current time, X-ray crystallography is the only technique that can show the structure of a protein down to atomic detail.

As stated earlier, crystallization attempts usually begin after the protein is in an essentially homogenous form. There are a variety of methods that have been developed for making protein crystals but the most popular and successful method is called vapor diffusion. There are

two dominant forms of vapor diffusion: hanging drop and sitting drop as shown in Figure 2.2. Vapor diffusion works by placing the protein of interest in a solution that has a lower concentration of additives than does a reservoir solution that is in close vicinity to the protein (*53*). Vapor diffusion works because the water in the protein solution that has lower concentrations of additives will diffuse out of that solution and into the reservoir solution that has higher concentrations (*53*). Additives are important because they help to both stabilize the folded protein and force the protein to form the crystal.

A good quality crystal is difficult to obtain and a process called screening and optimization is usually employed. Screening involves testing a wide variety of conditions and different additives on the protein to see if any of them produce a crystal or other phase that could indicate there is a possibility of crystal formation. The systematic search consists of changing these conditions in a way as to rule out different variables until a single optimal condition is found. A perfect protein crystal is rarely found but if the right condition is found, a crystal of suitable size, order and quality can be produced. There are many excellent reviews excellent reviews on the defects that occur within protein crystals (*54, 55*) that lead to either disordered crystals that don't diffract X-rays well or other defects that make crystals unusable.

The formation and growth of protein crystals takes place in three stages: nucleation, growth and termination. Before nucleation can take

place, the protein concentration in the solution must increase to a point of supersaturation. This increase in protein concentration is what the vapor diffusion method accomplishes. Once the protein is at a supersaturated concentration (Figure 2.3), nucleation may occur. Nucleation is just one of the major hurdles in protein crystallography. There are two types of nucleation that may occur; homogeneous and heterogeneous. Homogeneous nucleation occurs when a random event of protein molecules becoming clustered at the same location to form what is called the critical nucleus (*53, 57*). Once a critical nucleus is formed, the growth stage of the crystal formation can proceed. The second way to form the critical nucleus is through heterogeneous nucleation. Heterogeneous nucleation occurs when a number of the protein molecules form around a foreign particle that will not make up the bulk of the protein crystal. The foreign particle can be many different things such as hair (both human and horse hair have been tried) and sephadex beads to name a few (*57*). There have been many attempts but no successes in discovering a particle that could trigger heterogeneous nucleation for all proteins, as this would take out one of the major hurdles in protein crystallography. The nucleation stage is critical for obtaining suitable crystals. If nucleation occurs too quickly and at too many sites, there may only be showers of needle crystals (Figure 2.4) that are unusable for X-ray diffraction. The upside to needle crystals is that they have the possiblility for being used in a technique called seeding. Protein crystallography doesn't normally rely

on this technique to start the nucleation phase but in recent years there have been many attempts to develop successful seeding techniques to help to the nucleation process (*56*). If the nucleation is too slow, the critical nucleus may not be reached and no crystal will be produced. Nucleation occurs at a stage of supersaturation above where the growth phase of the protein crystal takes place. The growth phase mainly occurs in a zone called the metastable zone (*50*). The growth phase of the crystal is where the protein molecules pack onto the crystal surface and the size and dimensions of the crystal increase.

Methods

Isolation of SMN from Inclusion bodies

The first step in our method of obtaining the SMN protein constructs is the growth of the *E. coli* cells that contain the desired protein vector. The *SMN* sequence is in a pRSET vector and the *E. coli* are of the BL21 production line of cells. A starter culture is first grown in super optimal broth (SOB) media at 37° C with ampicillin for 10 hours. This culture is then used to inoculate six liters of SOB media and shaken at 250 rpm for 10 -12 hours or until absorbance at 600 nm equals 0.8. When the appropriate cell density has been achieved, Isopropyl β-D-1-thiogalactopyranoside (IPTG) is then added to induce the over-expression of the SMN construct (either SMNWT or SMNΔ7). Harvesting of the cells is completed in a Sorval RC6 Plus. The cells are spun at 8000 rpm for ten

minutes to sediment the cells out of the media. The cell pellet is then frozen at -20° C until it can be used at a later time.

The frozen cell pellet is then solubilized in a PBS buffer overnight by gently stirring at 4° C. When the cells have become completely solubilized, they are then lysed to extract the inclusion bodies that contain desired protein. Lysis occurs by the addition of lysozyme, DNase (used to ensure no nucleic acids stick to the protein) and $MgCl_2$ to activate the DNase. This solution is allowed to stir at room temperature for approximately 20 minutes or until the solutions viscosity has been reduced significantly. Sonication of the cells on ice is done to ensure complete breakage of cells. When sonication is complete a detergent, Triton X100, is added and the solution is allowed to stir at room temperature for five minutes. A centrifugation step is then employed to extract the protein in the insoluble fraction of the solution. The centrifugation step is completed on the Sorval RC6Plus at 13,000 rpm for 30 minutes. The resulting supernatant is discharged and the cell pellet is saved for further use.

The insoluble fraction is then solublized overnight in Buffer A by gently stirring at 4° C. The Buffer A solution has a high concentration of Urea and will completely denature all the proteins in the cell pellet. Protein found in inclusion bodies usually in somewhat pure in content but all of the protein is not correctly folded. To ensure that none of the SMN is misfolded, all of the protein must first be unfolded and later refolded.

This solublized solution of the cell pellet is then again centrifuged at 13,000 rpm for 30 minutes to deposit out the insoluble fraction that does not contain our desired protein. The supernatant now contains a soluble fraction of our desired protein along with all the other soluble proteins. The SMN construct must be now be purified from this soluble fraction without the other protein from the cell.

In the method employed for extracting and folding of SMNWT and SMNΔ7 from the supernatant, the soluble fraction is first bound to Sp Sepherose beads for one to two hours. This method is known as Immobilized Metal Affinity Chromatography (IMAC) and has been in use since 1975 for the purification of proteins (*58*). Different transition metals can be used in IMAC but in our case, Nickel is the metal of choice. IMAC utilizes the fact that the histidine functional group has a very high affinity for the Nickel. This aids in purifying the desired protein from the rest of the soluble protein in the supernatant that does not have a histidine tag. The amino acid histidine has an imidazole ring for its side chain. The electron donor groups that are located on the imidazole ring, form coordination bonds with transition metals. This affinity for the metal allows the protein to bind to the matrix and thus purifying it from any proteins that do not bind to the matrix.

After binding, the next step is folding the protein while it is bound to the IMAC column. A wash step of Buffer A with a small concentration of imidazole is allowed to run on the column to attempt to wash away non-

specifically bound proteins that may have a low affinity for either the Nickel or the SMN construct. After the wash buffer has ran through the column, a gradual gradient exchange of 100% Buffer A to 100% Buffer B takes place to slowly remove the denaturant while the protein is bound to the column. This gradual removal of the denaturant allows the protein to sample different conformations in its folding landscape. Our Buffer B solution has the folding additive sucrose added to it. The sucrose helps to stabilize the folded protein while in solution (*12*). Gradually removing the denaturant has the advantage of giving the protein a chance to fold into an incorrect local energy minima conformation but also a chance to escape that energy minima on its way to a global energy minima (*48*). Obtaining protein with the correct fold is extremely important for crystallization attempts.

After the denaturant has been completely removed, the protein needs to be eluted off the column. This is done by using a high concentration of imidazole which will compete with the protein for binding to the Nickel on the beads. Figure 2.5. shows an sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) gel of the protein after the IMAC column.

The imidazole must be removed from the eluted protein sample and this is done by gently stirring in dialysis overnight at 4° C overnight. After dialysis has been completed, the protein must be concentrated and then further purified on a Size Exclusion Chromotography (SEC) column.

Purification of Protein

The SEC column will allow the separation of proteins that have different molar masses. The heaviest proteins run through the column much faster than the smaller proteins. A typical SEC elution graph for SMN is shown in Figure 2.6. Typically two peaks are shown in the elution curve off the column. The 1st peak could be a number of things including a dimer of the SMN construct, a higher order oligimer or an aggregation of protein molecules. The second peak is the SMN construct that is separated from other contaminates presumably.

Crystallization was then attempted and found to be extremely prep dependent. Different conditions were attempted and found that the best conditions were 100 mM Tris buffer at a pH of 8.8, polyethylene glycol 4000K at 40% w/v and the reducing agent *tris*(2-carboxyethyl)phosphine (TCEP) at a concentration of 6mM. The drop size was 2 µL and contained 1µL of reservoir and 1 µL of the concentrated protein solution. Other similar conditions often produced needle crystals which is indicative of nucleation occurring to quickly and in too many locations to produce a diffraction quality crystal. X-ray diffraction has been attempted but no usable data was obtained with the present crystals.

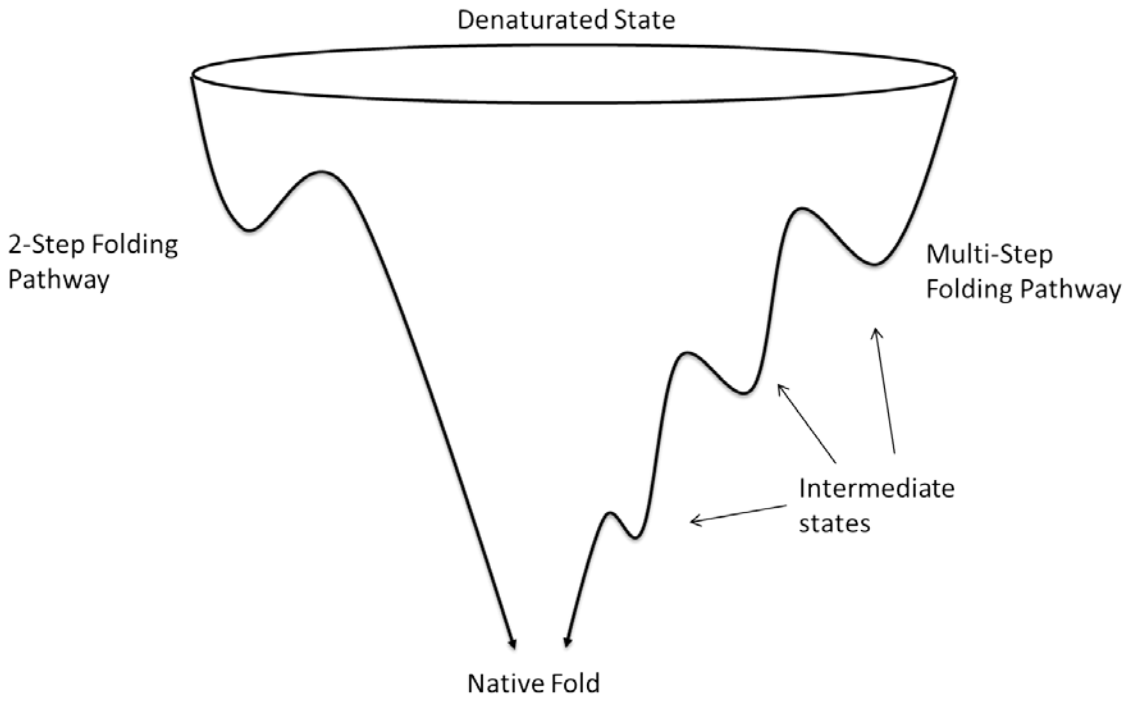

Figure 2.1. This figure shows the new views on how proteins fold. Smaller proteins of fewer than 100 amino acids will usually fold by the 2-step folding pathway with one intermediate step. Larger proteins of over 100 amino acids are thought to fold by moving though at least two intermediate states.

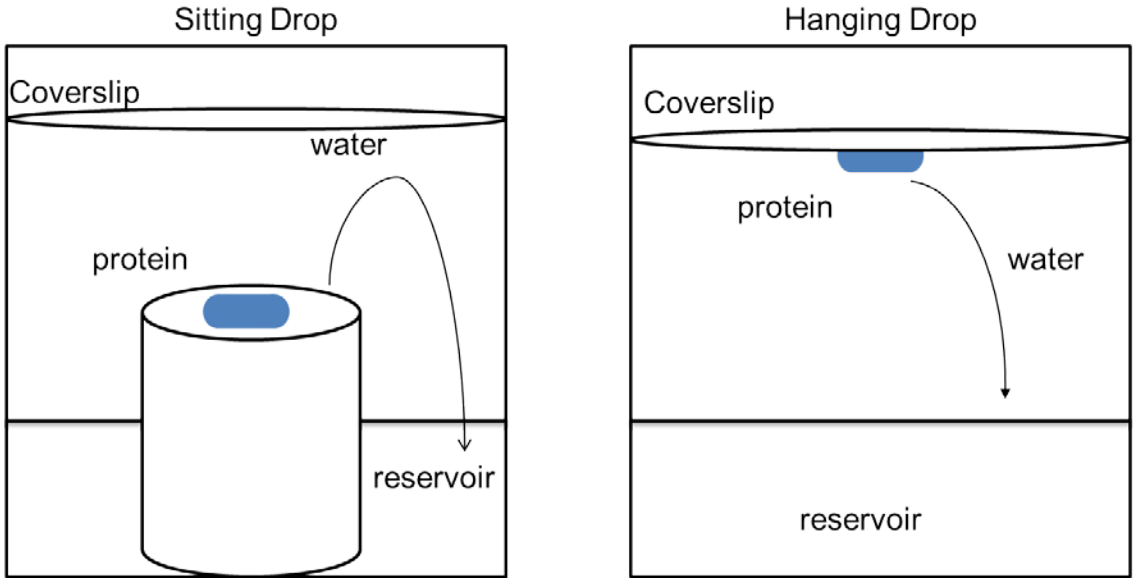

Figure 2.2. This figure shows the two common methods used for vapor diffusion. Water moves from the protein drop into the reservoir. This cause the concentration of both the protein and the additives in the drop to increase and eventually moving the protein into a point supersaturation.

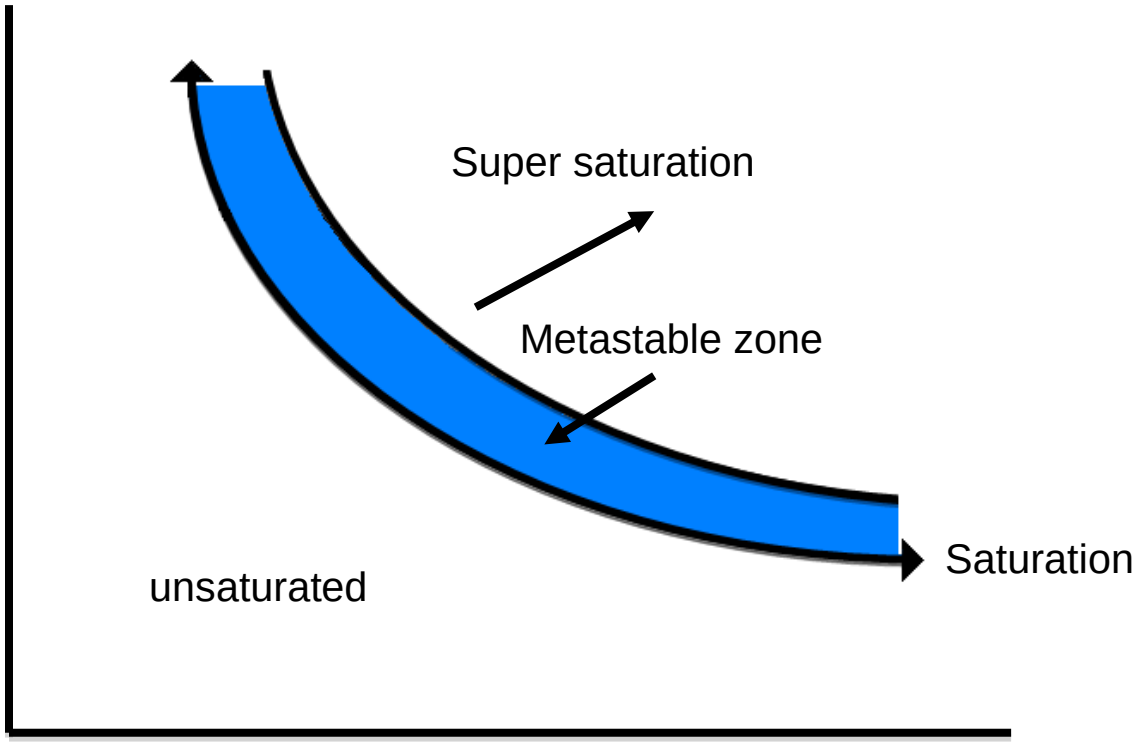

Figure 2.3. This figure shows a basic solubility curve. The supersaturation zone is where nucleation and aggregation occur. The metastable zone is where the growth of the crystal takes place. In the unsaturated zone neither nucleation or growth will occur. The crystal growth will terminate when the concentration of the protein drops below the saturation line.

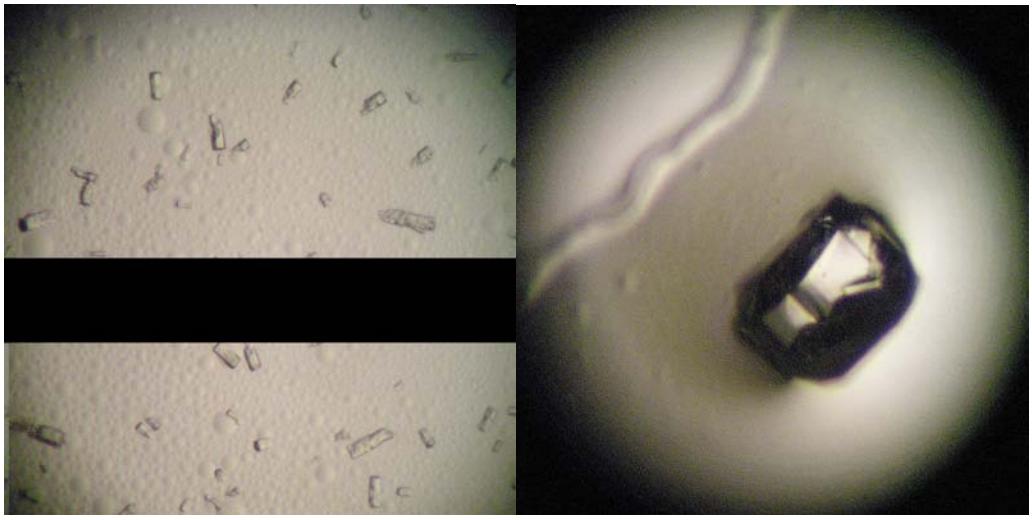

Figure 2.4. SMNWT has been crystallized successfully but no diffraction data has been collected to date. The picture on the right what occurs when nucleation happens very rapidly. The figure on the left is a large crystal that was formed when nucleation was occurring more slowly.

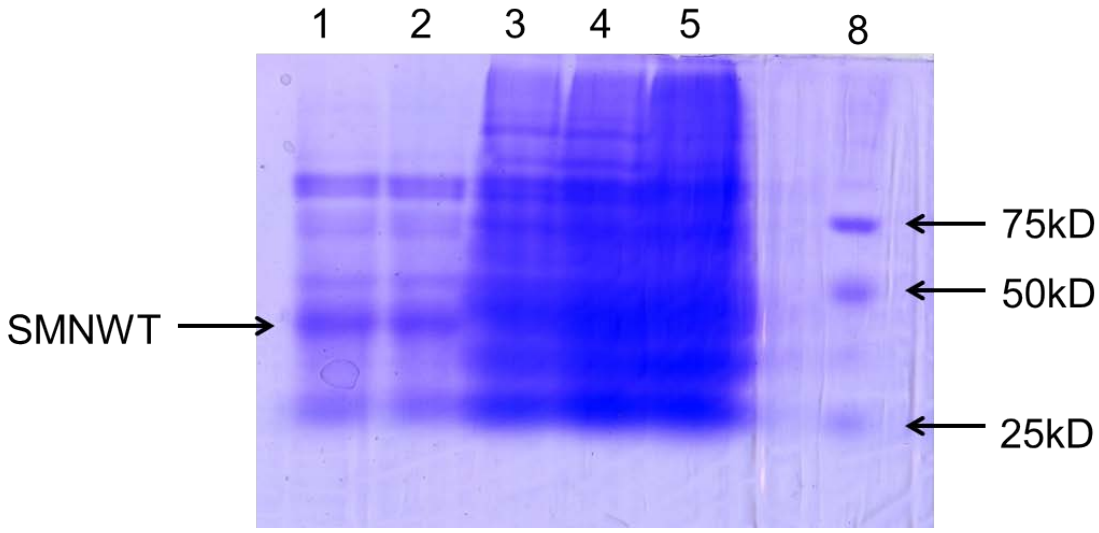

Figure 2.5. A typical SDS page gel of SMN WT is shown after purification with IMAC chromatography. Lanes 1 and 2 have the elution fractions while lanes 3, 4, and 5 have the supernatant from the sonication and clarification spin steps. Size exclusion must be ran to further purify the SMNWT protein for crystallization attempts.

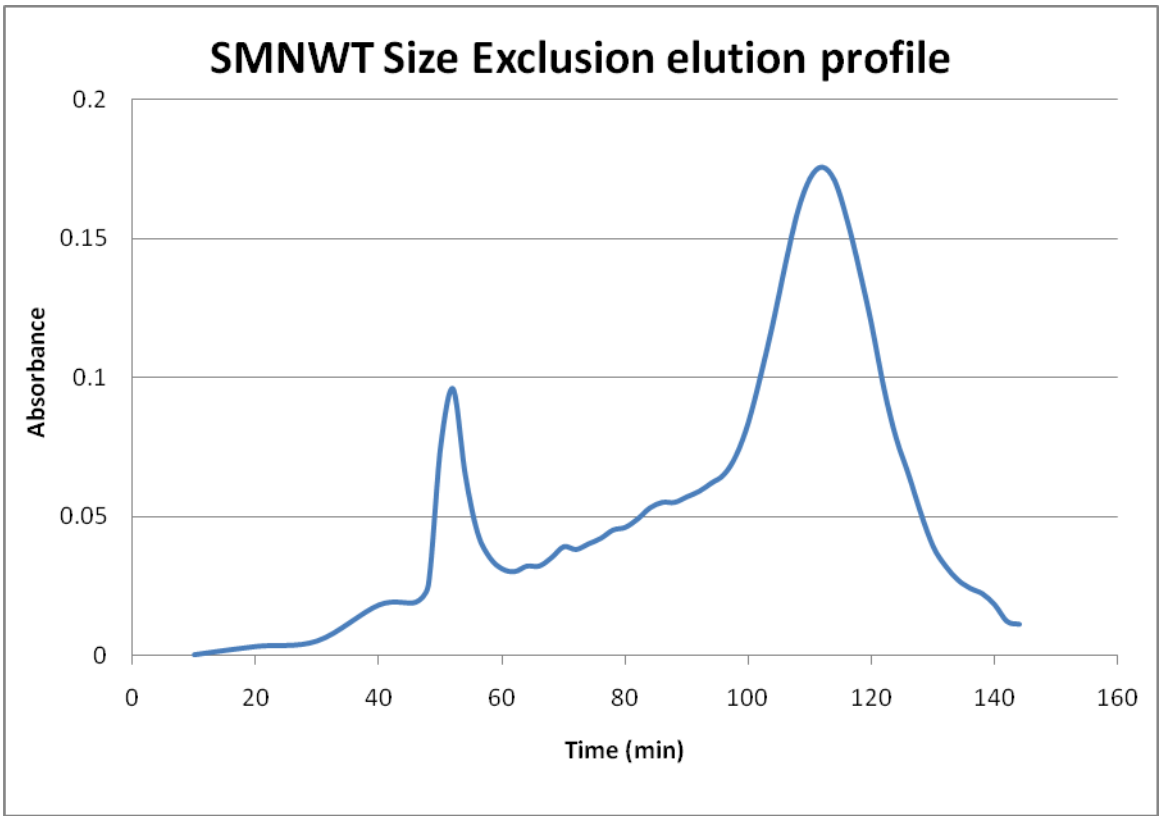

Figure 2.6. The typical size exclusion elution shows that there are two peaks. The second peak was found to be the pure form of the SMNWT monomer that was used in crystallization trials.

References

1. Joachimiak, A. (2009) High-throughput crystallography for structural genomics. *Curr. Opin. Struct. Biol. 19*. 573-584.

2. Hochuli, E., Döbeli, H., and Schacher, A. (1987) New Metal Chelate Adsorbent Selective for Proteins and Peptides Containing Neighbouring Histidine Residues. *J. Chromotography 411*, 177-184.

3. Smith, D. B., and Johnson, K. S. (1988) Single-step purification of polypeptides expressed in *Escherichia* coli as fusions with glutathione S-transferase. *Gene 67*, 31-40.

4. Stofko-Hahn, R. E., Carr, D. W., and Scott, J. D. (1992) A single step purification for recombinant proteins Characterization of a microtubule associcated protein (MAP 2) fragment which associates with the type II cAMP-dependent protein kinase. *FEBS Lett. 302*, 274-278.

5. Brizzard, B. L., Chubet, R. G., and Vizard, D. L. (1994) Immunoaffinity purification of FLAG epitope-tagged bacterial alkaline phosphatase using a novel monoclonal antibody and peptide elution. *Biotechniques 16*, 730-735.

6. Carrio, M. M., and Villaverde, A. (2001) Protein aggregation as bacterial inclusion bodies is reversible. *FEBS Lett. 489*, 29-33.

7. Carrio, M. M., and Villaverde, A. (2002) Construction and deconstruction of bacterial inclusion bodies. *J. Biotechnol. 96*, 3-12.

8. Clark, E. D. B., (1998) Refolding of recombinant proteins *Curr. Opin. Biotechnol. 9*, 157-163.

9. Fahnert, B., Lile, H., and Neubauer, P. (2004) Inclusion Bodies: Formation and Utilisation. *Adv. Biochem. Eng. Biotechnol. 89*, 93-142.

10. Kane, J. F., and Hartley, D. L. (1988) Formation of recombinant protein inclusion bodies in *Escherichia coli. Trends Biotechnol*. *6*, 95-101.

11. Prouty, W. F., Karnovsky, M. J., and Goldberg, A. L. (1975) Degradation of abnormal proteins in *Escherichia coli.* Formation of protein inclusions in cells exposed to amino acid analogs. *J. Biol. Chem. 250*, 1112-1122.

12. Baneyx, F., and Mujacic, M. (2004) Recombinant protein folding and misfolding in *Escherichia coli. Nature Biotechnol. 22*,1399-1408.

13. Martínez-Alonso, M., González-Montalbán, N., García-Fruitós, E., and Villaverde, A. (2009) Learning about protein solubility from bacterial inclusion bodies. *Microbial Cell Factories 8*, 4.

14. Singh, S. M., and Panda, A. K. (2005) Solubilization and refolding of bacterial inclusion body proteins. *J. Biosci. Bioeng*. *99*, 303-310.

15. Speed, M. A., Wang, D. I. C., and King, J. (1996) Specific aggregation of partially folded polypeptide chains: The molecular basis of inclusion body composition. *Nature. Biotechnol. 14*, 1283-1287.

16. Fink, A. L. (1998) Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Folding & Design 3*, R9-R23.

17. Luheshi, L. M., Crowther, D. C., and Dobson, C. M. (2008) Proteins misfolding and disease: from the test tube to the organism. *Curr. Opin. Chem. Biol. 12*, 25-31.

18. King, J., Haase-Pettingell, C., Robinson, A. S., Speed, M., and Mitraki, A. (1996) Thermolabile folding intermediates: inclusion body precursors and chaperonin substrates. *FASEB J.10*, 57-66.

19. Oberg, K., Chrunyk, B. A., Wetzel, R., and Fink, A. L. (1994) Native-like Secondary Structure in Inerleukin-1.beta. Inclusion Bodies Attenuated Total Reflectance FTIR. *Biochemistry 33*, 2628-2634.

20. Teschke, C. M., and King, J. (1993) Folding of the phage P22 coat protein in vitro. *J. Biol. Chem. 32*, 10839-10847.

21. Georgiou, G., Valax, P., Ostermeier, M., and Horowitz, P. M. (1994) Folding and aggregation of TEM β-lactamase: Analogies with the formation of inclusion bodies in *Escherichia coli*. *Protein Sci. 3*, 1953-1960.

22. Georgiou, G., and Valax, P. (1999) Isolating inclusion bodies from bacteria. *Methods Enzymol. 309*, 48-58.

23. Maachupalli-Reddy, J., Kelley, B. D., and Clark, E.D.B. (1997) Effect of Inclusion Body Contaminants on the Oxidation

Renaturation of Egg White Lysozyme. *Biotechnol. Prog. 13*, 144-150.

24. Babbitt, P. C., West, B. L., Buechter, D. D., Kuntz, I. D. and Kenyon, G. L. (1990) Removal of a Proteolytic Activity Associated with Aggregates Formed from Expression of Creatine Kinase in *Escherichia coli* Leads to Improved Recovery of Active Enzyme. *Nat. Biotechnol.* 8, 945-945.

25. Betton, J.-M., Sasson, N., Hofnung, M., and Laurent, M. (1998) Degradatikon *versus* Aggregation of Misfolded Maltose-binding Protein in the Periplasm of *Escherichia coli*. *J. Biol. Chem. 273*, 8897-8902.

26. Cubarsí, R., Carrió, M. M., and Villaverde, A. (2001) *In Situ* Proteolytic Digestion of Inclusion Body Polypeptides Occurs as a Cascade Process. *Biochem. Biophys. Res. Commun. 282*, 436-441.

27. Corchero, J. L., Viaplana, E., Benito, A., and Villaverde, A. (1996) The position of the heterologous domain can influence the solubility and proteolysis of β-galactosidase fusion proteins in *E. coli*. *J. Biotechnol.* **48**, 191-200

28. Carbonell, X., and Villaverde, A. (2002) Protein aggregated into bacterial inclusion bodies does not result in protection from proteolytic digestion. *Biotechnol. Lett. 24*, 1939-1944.

29. Middelberg, A.P.J. (2002) Preparative protein refolding. *Trends Biotechnol. 20*. 437-443.

30. Tsumoto, K., Ejima, D., Kumagai, I., and Arakawa, T. (2003) Practical considerations in refolding proteins from inclusion bodies. *Protein Expression Purif. 28*. 1-8.

31. Villaverde, A., and Carrió, M. M. (2003) Protein aggregation in recombinant bacteria: biological role of inclusion bodies. *Biotechnol. Lett. 25*. 1385-1395.

32. Fischer, B., Sumner, I., and Goodenough, P. (1993) Isolation, renaturation, and formation of disulfide bonds of eukryotic proteins expressed in *Escherichia coli* as inclusion bodies. *41.* 3-13.

33. Valax, P., and Georgiou, G. (1993) Molecular Characterization of β-Lactamase Inclusion Bodies Produced in *Escherichia coli*. 1. Composition. *Biotechnol. Prog*. *9*, 539-547.

34. Khan, R. H., Rao, K. B. C., Eshwari, A. N. S., Totey, S. M., and Panda, A. K. (1998) Solubilization of Recombinant Ovine Growth Hormone with Retention of Native-like Secondary Structure and Its Refolding from the the Inclusion Bodies of *Escherichia coli. Biotechnol. Prog. 14*, 722-728.

35. Levinthal, C. (1968) Are the pathways to protein folding. *J. Chim. Phys. 65.* 44.

36. Daggett, V., and Fersht, A. R. (2003) Is there a unifying mechanism for protein folding? *Trends Biochem. Sci. 28.* 18-25.

37. Sela, M., White, Jr. F. H., and Anfinsen, C. B. (1957) Reductive Cleaveage of Disulfide Bridges in Ribonuclease. *Science 125.* 691-92.

38. Dill, K. A., and Chan, H. S. (1997) From Levinthal to pathways to funnels. *Nat. Struct. Biol. 4.* 10-19.

39. Brockwell, D. J., and Radford, S. E. (2007) Intermediates: ubiquitous species on folding energy landscapes? *Curr. Opin. Struct. Biol. 17*. 30-37.

40. Radford, S. E., and Dobson, C. M. (1999) From computer simulations to human disease: emerging themes in protein folding. *Cell 97.* 291-298.

41. Dobson, C. M., Šali, A., and Karplus, M. (1998) Protein Folding: A Perspecitive from Theory and Experiment. *Angewandte Chemie 37.* 868-893.

42. Kramer, G., Ramachandiran, V., and Hardesty, B. (2001) Cotranslational folding – omnia mea mecum porto? *Int. J. Biochem. Cell Biol. 33.* 541-553.

43. Hartl, F. U., and Hayer-Hartl, M. (2009) Converging concepts of protein folding *in vitro* and *in vivo. Nat. Struct. Mol. Biol. 16.* 574-581.

44. Ignatova, Z., and Gierasch, L. M. (2004) Monitoring protein stability and aggregation *in vivo* by real-time fluorescent labeling. *Proc. Natl. Acad. Sci. U. S. A. 101.* 523-528.

45. Ellis, J. (1987) Proteins as molecular chaperones. *Nature 328.* 378-379.

46. Ogawa, C., Usui, K., Aoki, M., Ito, R., Itoh, M., Kai, C., Kanamori-Katayama, M., Hayashizaki, Y., and Suzuki, H. (2007) Gemin2 Plays an Important role in Stabilizing the Survival of Motor Neuron Complex. *J. Biol. Chem. 282,* 11122-11134.

47. Valejo, L. F., and Rinas, U. (2004) Optimized procedure for renaturation of recombinant human bone morphogenetic protein-2 at high protein concentration. *Biotechnol. Bioeng*. *85*, 601-609.

48. Dobson, C. M. (2003) Protein folding and misfolding. *Nature 426*, 884-890.

49. Ugwu, S. O., and Apte, S. P. (2004) The effect of Buffers on Protein Conformational Stability. *Pharm. Technol. March*, 86-113.

50. McPherson, A. (2004) Introduction to protein crystallization. *Methods 34.* 254-265.

51. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature 181*, 662-666.

52. Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, J. D., Westbrook, J. D., and Zardecki, C. (2002) The Protein Data Bank. *Acta. Cryst. D. 58*. 899-907.

53. Bolanos-Garcia, V. M., and Chayen, N. E. (2009) New directions in conventional methods of protein crystallization. *Prog. Biophy. Mol. Biol. 101.* 3-12.

54. Caylor, C. L., Dobrianov, I., Lemay, S. G., Kimmer, C., Kriminski, S., Finkelstein, K. D., Zipfel, W., Webb, W. W., Thomas, B. R., Chernov, A. A. and Thorne, R. E. (1999) Macromolecular impurities and disorder in protein crystals. *Proteins: Struct., Funct., Bioinf. 36*. 270-281.

55. Chernov, A. A. (1999) Estimates of internal stress and related mosaicity in solution grown crystals: proteins. *J Cryst. Growth 196*. 524-534.

56. Bergfors, T. (2003) Seeds to crystals. *J. Struct. Biol.142.* 66-76.

57. Saridakis, E., and Chayen, N. E. (2009) Towards a ‘universal’ nucleant for protein crystallization. *Trends Biotechnol. 27.* 99-106.

58. Porath, J., Carlsson, J., Olsson, I., and Belfrage, G. (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature 258*, 598-599.

Chapter 3

Structures of proteins and cofactors: X-ray crystallography

Introduction

Photosynthesis is the conversion of light energy into chemical energy in photosynthetic organisms. While each organism has a different specific pathway for the conversion process, the basis for all of these pathways is the capture of light, the transfer of energy among (bacterio)chlorophyll-containing light-harvesting complexes, and the eventual creation of a charge-separate state by the transfer of electrons from (bacterio)chlorophylls to electron acceptors in protein complexes. The ability of these molecules to perform energy and electron transfer is highly dependent on the precise spatial arrangement of the bacteriochlorophylls, the surrounding protein environment and the relative positions of the other cofactors. Therefore, an understanding of photosynthesis at a molecular level requires knowledge of the three-dimensional arrangement of the bacteriochlorophylls and chlorophylls in the proteins. In this article, the use of protein crystallography to determine these spatial arrangements is described (the crystallization process is reviewed in an accompanying article). The basic concepts of X-ray diffraction are briefly outlined, including the technical aspects of data collection and analysis, followed by examples of several structures of proteins from photosynthetic organisms that have been determined using crystallography.

Although tens of thousands of protein structures have been solved, the use of crystallography is still limited for integral membrane proteins, which hare proteins embedded in the cell membrane rather than free in solution. Since many of the critical proteins involved in photosynthesis are membrane proteins, the emphasis of this review will be on the elucidation of the structures of membrane proteins, with the structures of the bacterial reactioncenter, photosystem I, photosystem II, and the light-harvesting complexes I and II used as examples. The structures of bacterial reaction centers were the first membrane protein structures that were determined *(1,2)*, and many structures of the reaction center from *Rhodobacter sphaeroides* with mutations of other modifications have since been determined. While the structure of photosystem I is now well defined, the structural determination of photosystem II still remains an active area of research. Reaction centers, photosystem I, and photosystem II perform the primary energy conversion reactions in anoxygenic and oxygenic photosynthetic organisms, and these structures illustrate the beauty of the overall arrangement of these complexes as well as the difficulties in accurately determining the structures of large protein complexes.

In addition, this chapter briefly reviews the structures of the complexes that harvest the light and transfer the energy to the complexes discussed above. Purple bacteria predominately have two types of antenna, both of which are found in the cell membrane and surround the reaction center. Structures of both the light-harvesting complexes I and II

have been determined revealing a symmetric ring configuration that promotes energy transfer among the bacteriochlorophylls. Some green bacteria contain large assemblies known as chlorosomes that are attached to the cell membrane. Chlorosomes are coupled to the reaction center through the FMO protein, which is a water-soluble protein containing a number of chlorophylls. Although the FMO protein was the first bacteriochlorophyll-binding protein whose structures was solved *(3)*, the ability of the seven bacteriochlorophylls, which are not symmetrically arranged, to transfer energy among the various bacteriochlorophyll molecules is actively investigated as discussed below.

Basic concepts of diffraction

In this section, I will briefly summarize the basic concepts of diffraction. X-ray diffraction has been used for 100 years to solved the structures of molecules, with the first structures of proteins, namely, myoglobin and hemoglobin, being solved almost 50 years ago *(4,5)*. Diffraction is the technique of choice for the determination of the three-dimensional structures of proteins. Optical microscopy cannot be used for objects significantly smaller than the wavelength of light, 500 nm for green light, due to the intrinsic resolution of microscopes being coupled to the wavelength of the light. Use of X-rays with a wavelength of 0.1 nm is not feasible, as the materials used in conventional microscopes do not function at such short wavelengths because of the large energy

associated with this wavelength region. Therefore, structures are determined by using X-rays in a diffraction experiment.

The basic concept of diffraction can be understood by considering a simple array of molecules that are uniformly spaced in a line (Figure 3.1). When a wave impinges on the array, each molecule is assumed to simply scatter the wave, changing only the direction but not the energy of the wave. After scattering to a specific point at a large distance away from the array, the wave at that point will be the sum of all the scattered waves. In some cases, the waves will superimpose favorably while in other cases the waves will cancel. Consider the scattering from the neighboring molecules in which two scattered waves are identical except that one wave travels a longer distance Δ given by

$$\Delta = a \sin \theta, \tag{3.1}$$

where a is the spacing between the molecules, and θ is the angle between the incident wave and the scattered wave. When this path difference is equal to the wavelength, then the waves add constructively. In general, a diffraction peak will be observed when the path difference is a multiple of the wavelength, or

$$n\lambda = a \sin \theta, \tag{3.2}$$

where n is an interger.

For diffraction from a physical crystal, the scattering in three dimensions must be considered, although the analysis is essentially the same (Figure 3.2). The incident wave can be considered to strike a plane

of atoms in the crystal. Diffraction is observed when the difference in the path length between the scattered waves from two adjacent planes, which is determined by the distance between the two planes, *a*, and the incident angle, θ, is equal to a multiple of the wavelength according to Bragg's law:

$$n\lambda = 2a \sin \theta \quad (3.3)$$

All crystals, including crystals made from proteins, are composed of molecules that are precisely arranged in three-dimensional arrays. The smallest building block of the crystal is known as the unit cell, and the crystal can be considered to be composed of unit cells stacked next to each other. As originally done in the 1800's by Bravais, the arrangement of unit cells can be classified according to the three lengths and angles associated with each unit cell and their symmetry relationship. The possible arrangements range from triclinic cells that have no symmetry to the highly symmetrical cubic cells (Figure 3.3). In some cases, the symmetry is more involved, for example, unit cells can have twofold screw axis in which each molecule is related to each other by the combination of a twofold rotation and a translation along the *c* axis. Protein crystals commonly have screw axes because the crystal packing is imported compared to crystals with simple rotation axes. For example, the bacterial reaction center crystallizes in the orthorhombic space group $P2_12_12_1$, which has three perpendicular screw axes, and photosystem I crystallizes in $P6_3$, in which the proteins are related by a 60° rotation followed by a translation along the *c* axis.

Data Collection and analysis

Once crystals have been obtained, X-ray diffraction measurements are performed using either in-house systems or at a synchrotron. Laboratory instruments are useful for establishing the space group and measuring the initial data sets. However, for many proteins, the diffraction is weak requiring that the measurements use the much more intense X-ray beams available at synchrotrons. In either case, the crystals are aligned and exposed to the X-ray beam with the diffraction measurement in real time by an electronic detector. The space group can be immediately identified based on the arrangement of the diffraction peaks in the resulting image. The quality of the data is usually described by a termed called the resolution limit. The resolution limit represents the largest angle from the beam that can be reliably measured. In order to solve any given structure, data to a resolution limit beyond 3 Å is required with smaller numerical values for the resolution limit representing data that is collected to larger angles, and hence is more complete. Crystals yielding data with smaller resolution limits represent more measured data and hence better data, with data beyond a resolution limit of 1 Å being rare for protein crystals. For the bacterial reaction center, the initial structures were at resolution limits of ~3 Å, while the later structures were solved at higher resolution limits as discussed below. Several structures of photosynthetic complexes have been reported to much poorer resolution limits, notably photosystem II structures were initially determined to

resolution limits of 3.8–3.7 Å *(6, 7)*, resulting in limitations in the placement of the protein backbone and the cofactors, with the most recent structure being at a 3.5–3.0 Å resolution *(8, 9)*. In addition to the resolution limit, there are other measures of the quality, with the order being critical for many integral membrane proteins. Due to the presence of detergent molecules surrounding a significant portion of the protein, the interactions between the neighboring proteins are often limited resulting in relatively poorly ordered crystals. In terms of the diffraction, poorly ordered crystals are apparent by the measurement of diffraction spots that are not sharp but diffuse, with the resolution being sometimes restricted by the overlap of spots.

In order to determine the structure, it is necessary to measure the intensity of every diffraction spot. The specific space group determines the location of each diffraction peak, while the intensity is related to the composition of the unit cell. Since the data are recorded electronically, these measurements can be performed quickly. The difficulty in the interpretation lies in the nature of the diffraction as arising from complex rather than real terms. The diffraction can be considered to arise from the summation of vectors, called structure factors, which have both an amplitude and phase, but measurement of the intensity provides only the amplitude. Formally, in order to determine the structure, both are required in a Fourier series involving the summation of the structure factors, F_{hkl}, for all measured reflections, which are identified by the indices *hkl*:

$$\rho(x,y,z) = \sum_{hkl} F_{hkl}\, e^{-2\pi i(hx+ky+lz)} \qquad (3.4)$$

Notice that the summation yields not the position of the nuclei but rather ρ(*x,y,z)*, which is the density of electrons at a given point in space, since it is the electrons that scatter the X-rays. For this reason, the outcome of the data analysis is an electron density map into which a structural model, consisting of amino acid residues, must fit.

Three experimental approaches can be used to provide the missing information concerning the phases. The first, called MAD, or multiple anomalous dispersion, involves measurements at several wavelengths around the transition energy for a metal bound to the protein. Since the transition energy for each metal is very specific, the differences in the measured diffraction arise only due a specific metal resulting in contribution that arises only from the electrons in that metal. Once this metal is located, then an initial set of phases can be estimated and a model can be built. For a second approach termed MIR, or multiple isomorphous replacement, the protein is modified such that a metal is incorporated and the diffraction is compared to the protein without the metal, with the measurements all being at a single wavelength. In this case, measurements of the diffraction from crystals with different bound metals are needed to give accurate estimations of the phases. Finally, MR, or molecular replacement, can be utilized to solve the structures. In this case, there must be an existing structure that is highly homologous to the unknown structure. The homologous structure is essentially artificially

rotated and translated in the unit cell until the orientation and position match those of the protein in the crystal.

These approaches provide the means for generating the phases for the diffraction data and allowing the electron density to be calculated using Eq. 3.4. The outcome is a map of the electron density that is usually contoured at a level significantly above background. Fitting of the maps involves identifying the atoms that give rise to each region of the density. While this is largely manually done on a graphics terminal by a crystallographer, increasingly, the analysis of electron density maps is being performed directly by sophisticated programs. The polypeptide chains are evident as tubes of electron density with secondary structural elements, such as alpha helices, readily identifiable. Once the native structure has been determined, modifications by mutagenesis or biochemical techniques can be rapidly identified provided the structural differences are small. For example, a very large number of mutants of the reaction center from *R. sphaeroides* have been solved and the mobility of the secondary quinone was identified. Also, light-induced changes have been found involving the secondary quinone in the bacterial reaction center *(10, 11)*.

Structures of pigement-protein complexes from photosynthetic organisms

Bacterial reaction centers

The bacterial reaction centers from *Blastochloris viridis* and *Rhodobacter sphaeroides* were the first integral membrane proteins to

have their three-dimensional structures determined *(12)*. Bacterial reaction centers contain at least two protein subunits, which are termed the L and M subunits, and surround the cofactors *(13)*. Both these subunits consist of five transmembrane helices, with the helices from each subunit being related to each other by an approximate twofold symmetry *(1, 2, 14-18)* (Figure 3.4). Each transmembrane helix contains 24–31 amino acid residues that are largely hydrophobic. The reaction centers from purple bacteria have an additional subunit, which is termed the H subunit. Unlike the L and M subunits, the H subunit has only one transmembrane helix with a large extramembranous domain on the cytoplasmic side of the reaction center. Although the H subunit does not directly interact with any of the cofactors, the H subunit is required to stabilize the complex *(19)* and for assembly *(20-22)*.

The cofactors of the reaction center all lie within the L and M subunits, and are organized in two branches that are related to each other by the same twofold symmetry axis as for the protein subunits. On the periplasmic side of the reaction center are two bacteriochlorophyll molecules that overlap at the ring A position and serve as the primary electron donor. Each of the two branches also has a bacteriochlorophyll monomer, a bacteriopheophytin, and a quinone, with a single non-heme iron atom lying on the symmetry axis between the primary and secondary quinone. In wild type, another cofactor, a carotenoid, is asymmetrically positioned near one of the bacteriochlorophyll monomers. Light absorption

results in excitation of the bacteriochlorophyll dimer followed by electron transfer along the A-branch of cofactors to the primary quinone, and then to the secondary quinone. After the bacteriochlorophyll dimer is reduced by a cytochrome, a second light absorption and electron transfer results in a second reduction of the secondary quinone that is coupled to the transfer of two protons. The quinone then carries the electrons and protons to the membrane and is replaced with another quinone. Cyclic electron transfer is achieved through electron- and proton-transfer processes involving the cytochrome bc_1 complex. In some organisms, including *B. viridis*, the bacteriochlorophyll dimer is reduced by a large bound tetraheme cytochrome subunit that contains four hemes (Figure 3.4). After light excitation, the bacteriochlorophyll dimer is rapidly reduced by the closest heme, allowing the complex to absorb light again. For reaction centers that do not possess a tetraheme subunit, a cytochrome c_2 binds transiently to the reaction center to transfer the electron in about 1 μs *(23)*. The structure of the cytochrome c_2 bound to the reaction center of *R. sphaeroides* has been determined by using protein crystallography (Figure 3.4), and shows that the heme is located directly over the bacteriochlorophyll dimer in a position similar to that of the closest heme of the tetraheme cytochrome *(24)*. While electrostatic interactions are critical factors that determine the binding, the final arrangement of the cytochrome bound to the reaction center is determined by other interactions such as hydrophobic interactions *(25)*.

Photosystem II

Photosystem II is a unique protein, as it is the site for the oxidation of water to molecular oxygen in oxygenic photosynthesis *(26)*. This protein is a large multi-subunit complex with many cofactors that are all embedded in the thylakoid membranes of cyanobacteria and plants. While there are some differences between the complexes found in these different organisms, the core of photosystem II is largely well conserved. In cyanobacteria, photosystem II has 17 integral membrane protein subunits that span the membrane and three extrinsic membrane protein subunits. There have been several reports of the three-dimensional structure of photosystem II at different resolution limits and quality of models. The first reports were at resolution limits of 3.8 and 3.7 Å *(6, 27, 28)* and revealed the organization of the transmembrane helices and the approximate positions of most of the chlorophyll cofactors. Subsequently improved models were determined from data at resolution limits of 3.5–3.0 Å *(8, 9)* and revealed to a much more complete degree the organization of the protein subunits and positions of the amino acid residues and cofactors. While efforts are underway to improve the quality of the diffraction data, all these diffraction data have been measured from the same crystal form, $P2_12_12_1$. Central to photosystem II are two protein subunits, D1 and D2, which encase the cofactors that perform the normal light-induced electron-transfer reactions. As is found for the bacterial reaction center, these subunits and cofactors are symmetrically related

around an approximate twofold symmetry axis (Figure 3.5). Each of the D1 and D2 subunits contains five transmembrane helices, whose relative positions are homologous to those of the L and M subunits of the bacterial reaction center. Only one branch of cofactors, located on the D1 side of the complex, predominantly participates in the electron transfer. The two quinones serve the same roles as found in the bacterial reaction center, with the primary quinone initially receiving the electron and the secondary quinone serving as the two electron–proton gate.

While many of the functional features are very similar on the acceptor side for photosystem II and the bacterial reaction center, the donor side functions are much different. The two chlorophylls lying in the center are not as closely spaced together as found in the reaction center and the excitation is thought to be delocalized along all four chlorophylls found in the central core *(29)*. After electron transfer, the oxidized chlorophylls, P680, are reduced not by a cytochrome but rather by an intrinsic amino acid, tyrosine 161 of D1, which is termed Y_Z. The resulting tyrosyl radical is not stable and is rapidly reduced by the manganese cluster. After four electron equivalents are collected on the manganese cluster, oxidation of water occurs.

While the general scheme of water oxidation is understood, both the detailed mechanism of water oxidation and the precise structure of the manganese cluster remain in question. Part of the difficulty is that the current models are based on diffraction data with limited resolutions. The

other difficulty is that the diffraction data have been measured using intense X-rays that are available from synchrotrons and the X-ray damage causes reduction of the manganese cluster to a state not normally observed *(30)*. While efforts are underway to improve the quality of the diffraction data, transient optical and X-ray spectroscopy, coupled with biochemical and mutagenesis studies, has provided insight into the three-dimensional arrangement of the manganese cluster.

Photosystem I

In oxygenic photosynthesis, the light-induced biochemical pathways involving photosystem II are coupled to a second chlorophyll-containing complex, photosystem I *(31)*. Photosystem I is a very large complex, comprised multiple subunits and cofactors. In cyanobacteria, the complex has 96 chlorophylls and 22 carotenoids that serve both light-harvesting and electron-transfer functions. The structure has been determined initially at a resolution limit of 6.0 Å and subsequently to 2.5 Å *(32, 33)*. The initial electron density revealed the presence of a trimer, which has an approximate diameter of 220 Å, formed by transmembrane helices with some of the cofactors, in particular the iron–sulfur clusters. Once the higher resolution data was obtained, the amino acid residues were modeled, the cofactors were accurately positioned, and the remaining cofactors were identified. In addition to improvements in the crystal quality through seeding and modification of the crystallization conditions, the packing of the complexes is very open with a large solvent content and

significant improvement in the diffraction data was achieved after conditions for freezing the crystals were developed *(38)*. The protein crystallizes in the space group $P6_3$ with each subunit of the trimer related by a crystallographic threefold symmetry axis. The structure of the complex from plants has also been determined at an initial resolution limit of 4.4 Å, followed by a 3.4 Å structure, and shows many of the same features, with the additional features of interactions with the light-harvesting I complex *(27, 35)*.

Each monomer of photosystem I from the cyanobacterium *Thermosynechococcus elongatus* has 32 transmembrane helices from nine different subunits. Central to photosystem I are two subunits, PsaA and PsaB, which each have a molecular mass of 80 kDa and 11 transmembrane helices. As found for the bacterial reaction center and photosystem II, photosystem I has a core that has an approximate twofold symmetry axis relating both the protein subunits and cofactors. From PsaA and PsaB are five transmembrane helices that are closely interacting with the cofactors and surround the cofactors that perform the primary photochemistry (Figure 3.6). The electron-transfer cofactors are composed of chlorophyll and phylloquinone cofactors that are divided into an A and B branch. Two closely associated chlorophylls are found forming the primary electron donor although the two molecules are not identical, as the A-branch side is a chlorophyll *a'* while the B-branch side is a chlorophyll *a*. Another notable difference compared to the bacterial

reaction center and photosystem I is the presence of three redox active Fe_4S_4 clusters coordinated by cysteines. An iron–sulfur cluster, termed F_X, is found at the homologous location of the non-heme iron of the bacterial reaction centers and photosystem II. In addition, there are two other iron–sulfur clusters, F_A and F_B, that serve as electron acceptors from F_X and are located on the stromal subunit PsaC, with the electrons being transferred out of the complex by a ferredoxin.

Light-harvesting complexes I and II

Two types of light-harvesting complexes are found in purple bacteria, both of which are integral membrane proteins: a core light-harvesting I complex and a peripheral light-harvesting complex II, which is found in many but not all bacteria *(13)*. For both complexes, there are two polypeptide chains, alpha and beta, with each set of peptides being associated with two and three bacteriochlorophylls, for the light-harvesting complexes I and II, respectively. Light is captured initially by the light-harvesting complexes II, and the energy is transferred to the light-harvesting complex I in 2–4 ps, followed by the energy being transferred to the reaction center within 40 ps. The means by which the protein environment tunes the properties of the bacteriochlorophylls has long been a topic of study *(36)*. For example in *R. sphaeroides*, the light-harvesting complex I has absorbance bands at 875 nm, while the bands are at 800 and 850 nm for the light-harvesting complex II, despite both complexes containing the same cofactor, bacteriochlorophyll *a*.

The three-dimensional structure of one of these complexes was first solved at 2.5 Å for the light-harvesting complex II from *Rhodopseudomonas acidophila (37)*, with the diffraction data subsequently improving to 2.0 Å *(38)*. The crystals belong to the trigonal space group R32 with the asymmetric unit containing three sets of the alpha and beta peptides. By the crystallographic symmetry, each part of the asymmetric unit maps by a crystallographic threefold axis into a large ring structure composed of nine sets of alpha and beta peptides (Figure 3.7). The bacteriochlorophylls are arranged into two rings. One is near the periplasmic side of the protein and is composed of 18 closely interacting cofactors that are assigned with the 850 nm absorption band. A second ring of nine widely separated molecules is found near the cytoplasmic side and is assigned to the 800 nm band. The light-harvesting complex II from *Rhodospirillum molischianum (39)* has also been determined to a resolution limit of 2.5 Å. In that case, the space group is $P42_12$, which has no threefold symmetry axes, and the crystallographic symmetry results in a ring of protein and cofactors as found for *R. acidophila*, but the complex is an octomer rather than a nonamer. The organization of the cofactors is similar, although there are small differences in the various angles and distances.

In addition to the crystallographic studies of the light-harvesting complex II, structural studies have been performed on the light-harvesting complex I. Projection maps from electron microscopy studies of two-dimensional

crystals of the light-harvesting complex I from *Rhodospirillum rubrum,* show a ring arrangement of 16 sets of polypeptide pairs, but the resolution was too limited to provide a detailed picture *(40)*. A model of the reaction center light-harvesting complex I from *Rhodopseudomonas palustris* has been determined at a resolution limit of 4.8 Å *(41)*. That resulting structure consists of 15 pairs of alpha and beta subunits, each with two bacteriochlorophylls, with the 16th position of the ring occupied by a different set of peptides, which are proposed to be the PufX proteins. The presence of the additional peptides breaks the ring structure and is proposed to provide a portal for quinone exchange from the reaction center to the cytochrome bc_1 complex.

FMO protein

In photosynthetic systems, a number of different complexes harvest the light and transfer the energy to the complex, where the photochemistry occurs. Some light-harvesting complexes, such as the ones described above, are found in the cell membranes. Others are found associated with the membrane but with no protein subunits that span the membrane. One example, of a membrane-associated antenna is the FMO protein. The FMO protein participates in the light-harvesting process in green bacteria, as it facilitates energy transfer to the reaction center from large membrane-associated components called chlorosomes.

The FMO complex was originally solved from *Prosthecochloris aestuarii* at a resolution limit of 2.8 Å *(3, 42)*. The space group, $P6_3$,

produced closely associated trimers of the FMO that are thought to be the biological unit. This structure was the first of a bacteriochlorophyll-containing complex until the bacterial reaction center structures were determined (see above). Due to the lack of a protein sequence at that time (1975), the model was not completed with inclusion of the amino acid side chains until several years later at which time an improved 1.9 Å electron density map was available *(43)*. Subsequently, the FMO protein was solved from *Chlorobium tepidum* at a resolution limit of 2.2 Å *(44, 45)*. In that case, the space group was $P4_332$ but the crystallographic symmetry again produced the same trimeric arrangement of the protein subunits.

Overall, the FMO complex consists of a trimer of three identical subunits (Figure 3.8). Each subunit is folded into two beta sheets that have a “taco shell” arrangement with seven bacteriochlorophylls inside. At the open end of the taco shell are small alpha helices. The seven bacteriochlorophylls are arranged asymmetrically within the protein and have different types of coordination. The involvement of each bacteriochlorophyll molecule in the transfer of energy and contribution to the optical spectrum has been discussed by many research groups. Recent transient spectroscopic measurements showed that while each molecule can be excited the coupling among the molecules results in the energy being transferred along specific pathways and collected on a single bacteriochlorophyll, which presumably serves as the bridge to the reaction center *(46, 47)*. More recently, a structural model of FMO has

been proposed that contains an additional cofactor, an eighth bacteriochlorophyll, although the occupancy is low compared to the full occupancy of the other bacteriochlorophylls *(28)*.

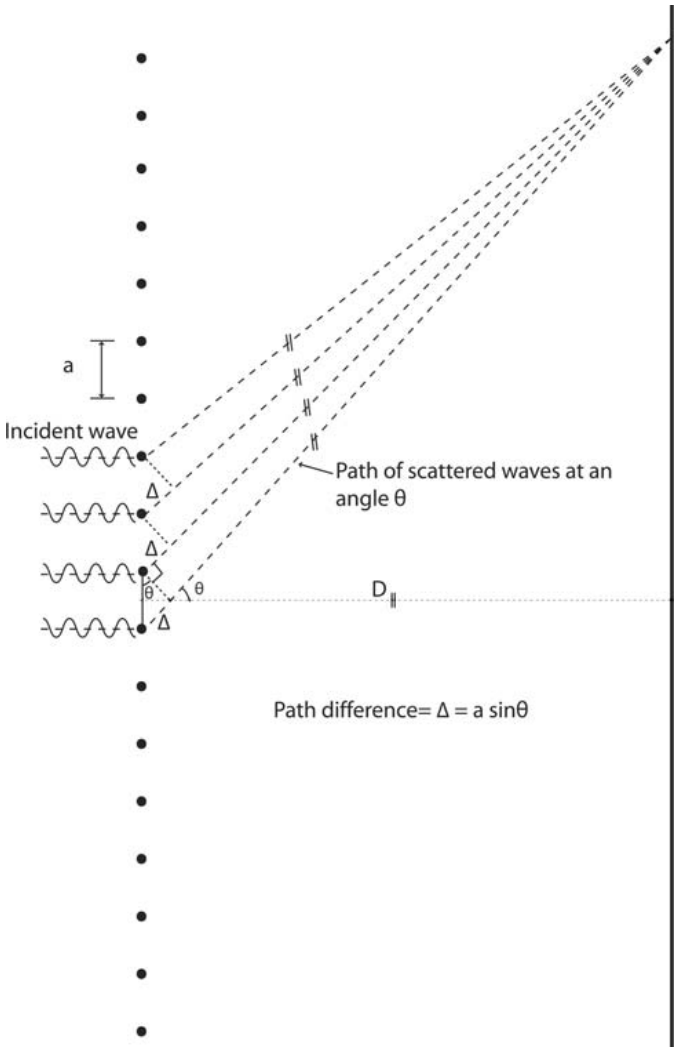

Figure 3.1: Diffraction from a one-dimensional array of molecules. Each molecule is spaced a distance a apart from the neighboring molecule. The incident wave is considered to scatter from each molecule to a point at a large distance D, with the scattered waves having different path lengths. At an angle h, the path length difference between neighboring molecules is given by a sinh. When the path length difference is equal to a multiple of the wavelength, then the waves add constructively and a diffraction peak is observed. For simplicity, only four of the scattered waves are shown in the figure.

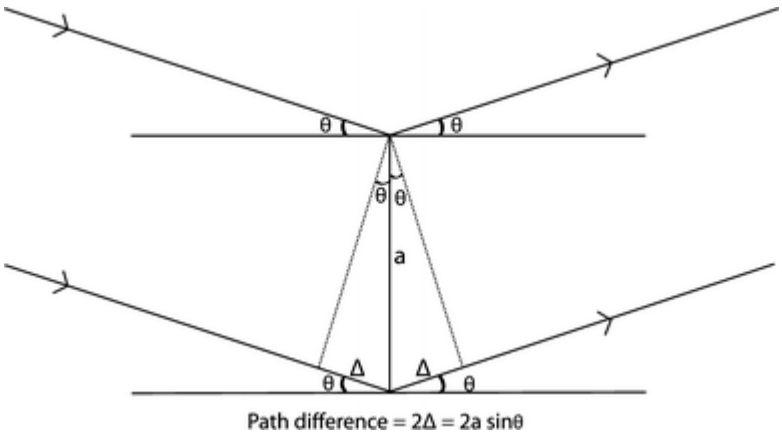

Figure 3.2: Diffraction from crystallographic planes. The crystal is considered to consist of a series of planes that can each scatter the incident wave. As was true for the one-dimensional array, the difference between the scattered waves is the difference in the pathlength between the neighboring planes, which equals 2*a* sin*è*. When this path difference equals a multiple of the wavelength, the waves add constructively and a diffraction peak is observed.

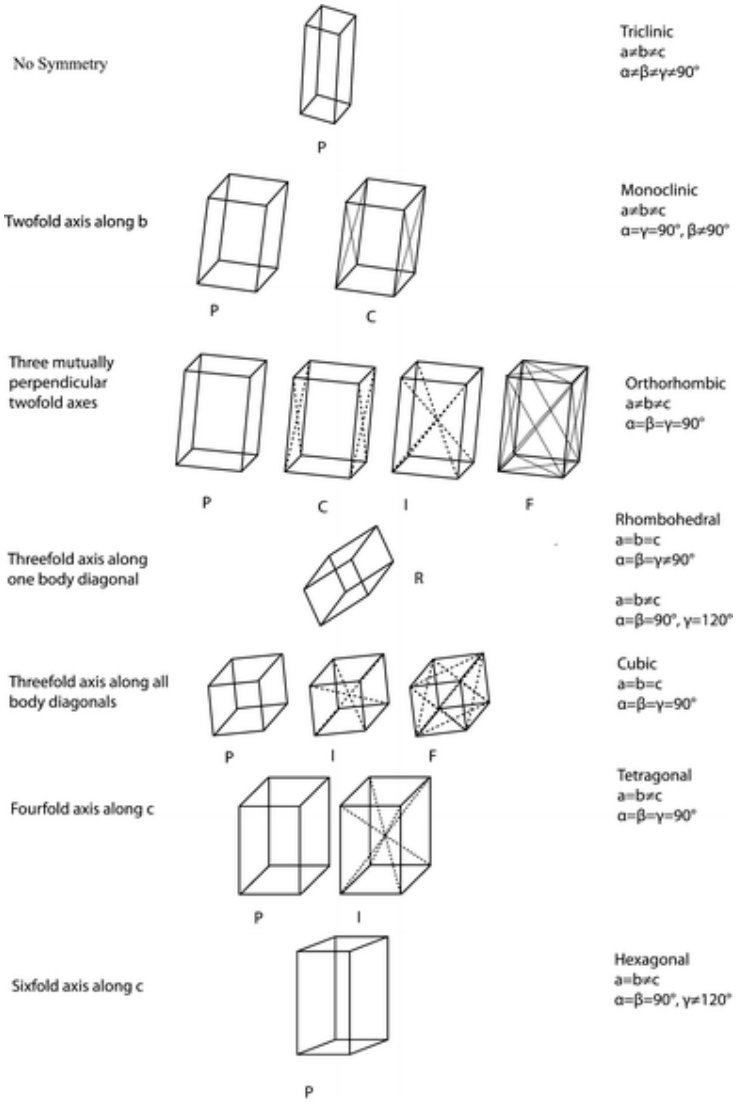

Figure 3.3: The different types of Bravais lattices. Shown are the seven types of unit cells and the allowed unique lattices. The relationships among the three characteristic cell lengths, *a*, *b*, and *c* are shown as well as for the three angles, *á*, *â*, and *ã*. For example, in triclinic cells, all the cell lengths differ while they are all equal in cubic cells. Also indicated are the symmetries for each cell type. Each cell type can have primitive (P), face-centered (C), body-centered (I), or all facecentered (F) cells. For example, all triclinic cells are primitive while cubic cells can be primitive, body centered, or face centered. The dashed lines show that for these different cell types, there can be additional lattice points. For example, the body-centered cells have an additional lattice point in the center.

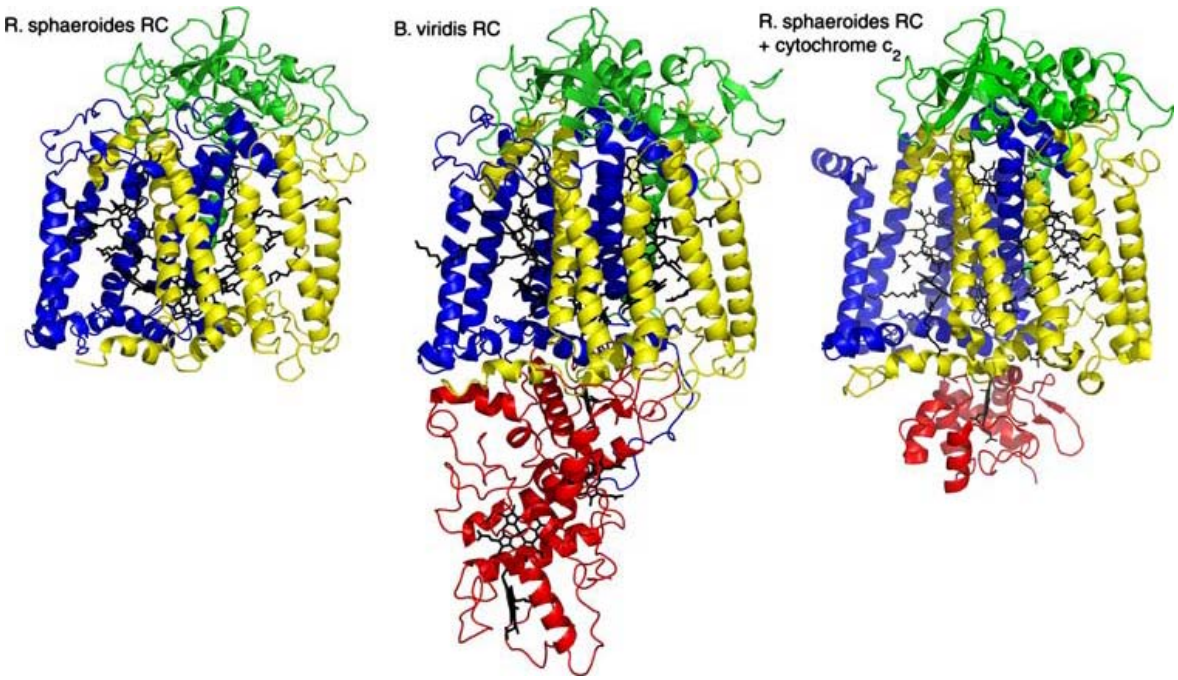

Figure 3.4: The three-dimensional structures of the reaction center from *Rhodobacter sphaeroides*, the reaction center from *Blastochlorsis viridis*, and the reaction center from *Rhodobacter sphaeroides* with a bound cytochrome c2. Shown are the backbones of the protein subunits, L (yellow), M (blue), H (green), cytochrome (red), and the cofactors (black). The views are approximately with the twofold symmetry axis of the protein in the plane of the paper. The coordinates are 4RCR (Allen et al. 1987), 4PRC (Deisenhofer et al. 1995), and 1L9B (Axelrod et al. 2002).

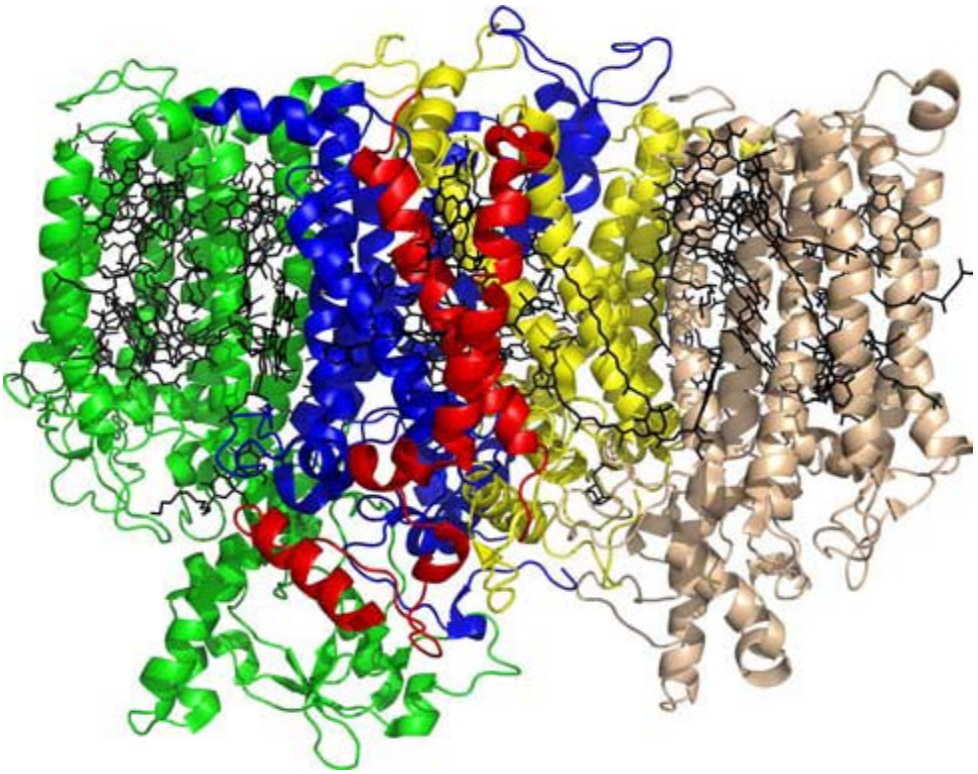

Figure 3.5: The three-dimensional structure of photosystem II from *Thermosynechococcus elongatus*. Shown are the backbones of several of the protein subunits, D1 (yellow), D2 (blue), cytochrome b559 alpha (red), CP43 (wheat), the core light-harvesting subunit (green), and cofactors (black). For clarity, some subunits and cofactors are not shown. Only one monomer of the biologically active dimer is shown. The view is approximately with the twofold symmetry axis of the protein in the plane of the paper. The coordinates are 1S5L (Ferreira et al. 2004).

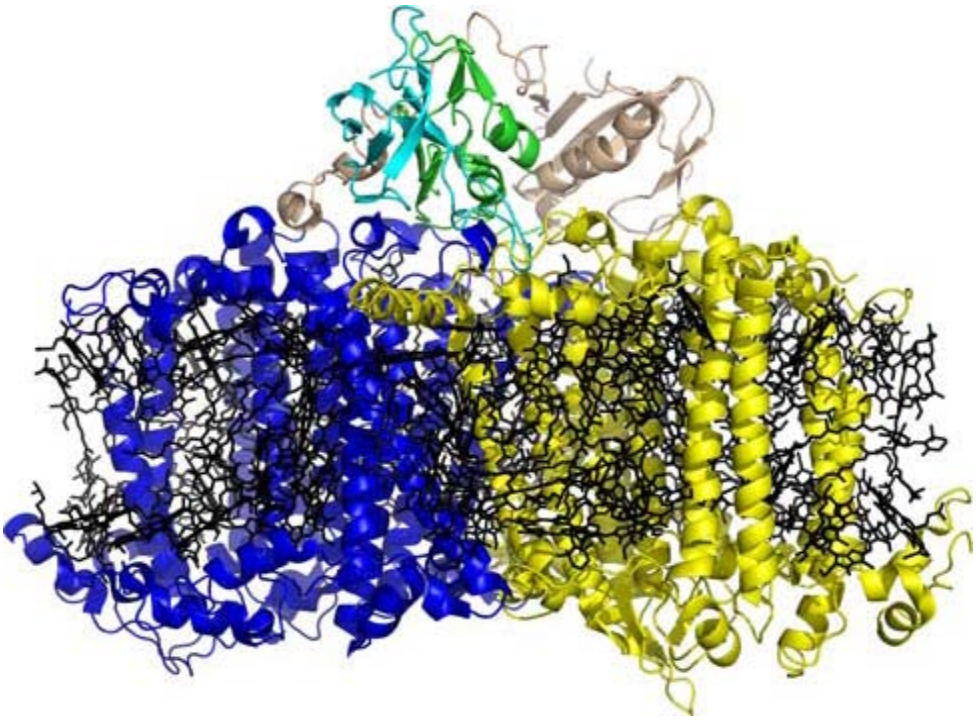

Figure 3.6: The three-dimensional structure of photosystem I from *Synchococcus elongatus*. Shown are the protein subunits: PsaA (yellow), PsaB (blue), PsaC (green), PsaD (cyan), PsaE (wheat), and the cofactors (black). For clarity, some subunits and cofactors are not shown. Only one monomer of the biologically active trimer is shown. The view is approximately with the threefold symmetry axis of the protein in the plane of the paper. The coordinates are 1JB0 (Jordan et al. 2001).

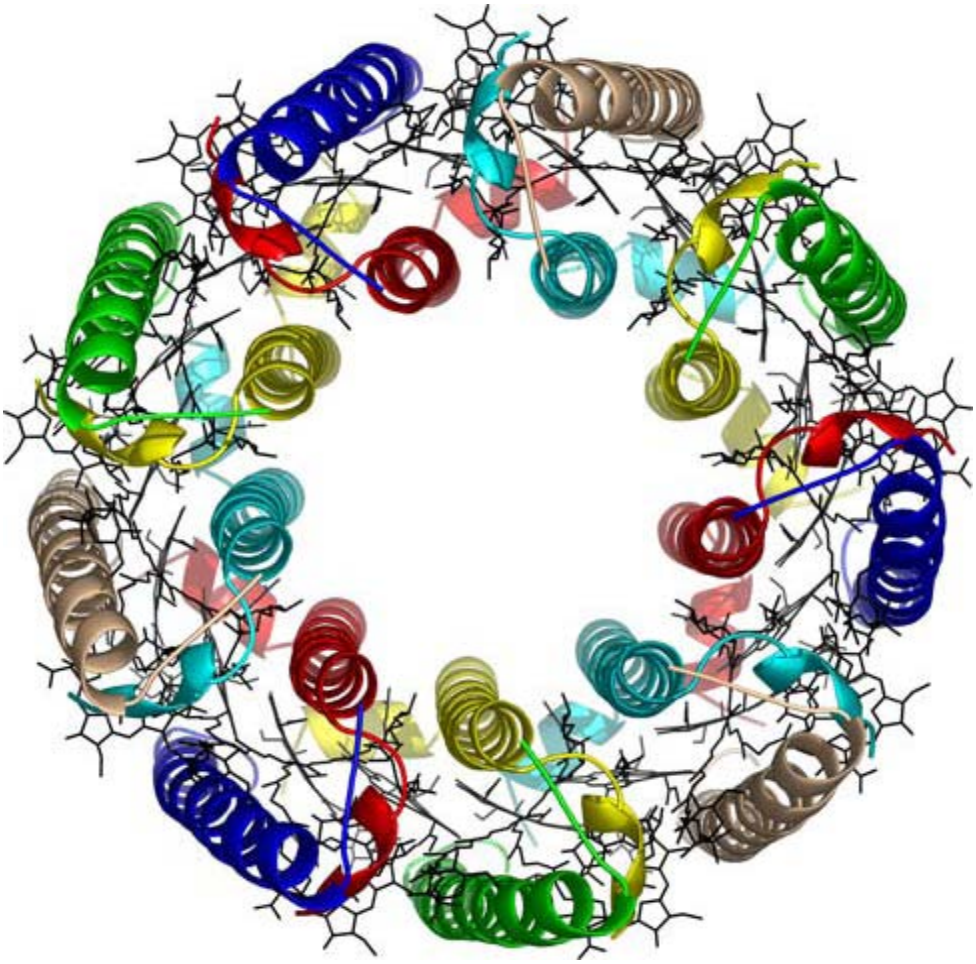

Figure 3.7: The three-dimensional structure of the light-harvesting complex II from *Rhodopseudomonas acidophila*. Shown is the biologically active complex with 18 protein subunits (multiple colors) arranged as two concentric rings with three bacteriochlorophylls (black) between each pair of subunits. The view is approximately down the ninefold symmetry axis of the biologically active nonamer. The coordinates are 1KZU (McDermott et al. 1995).

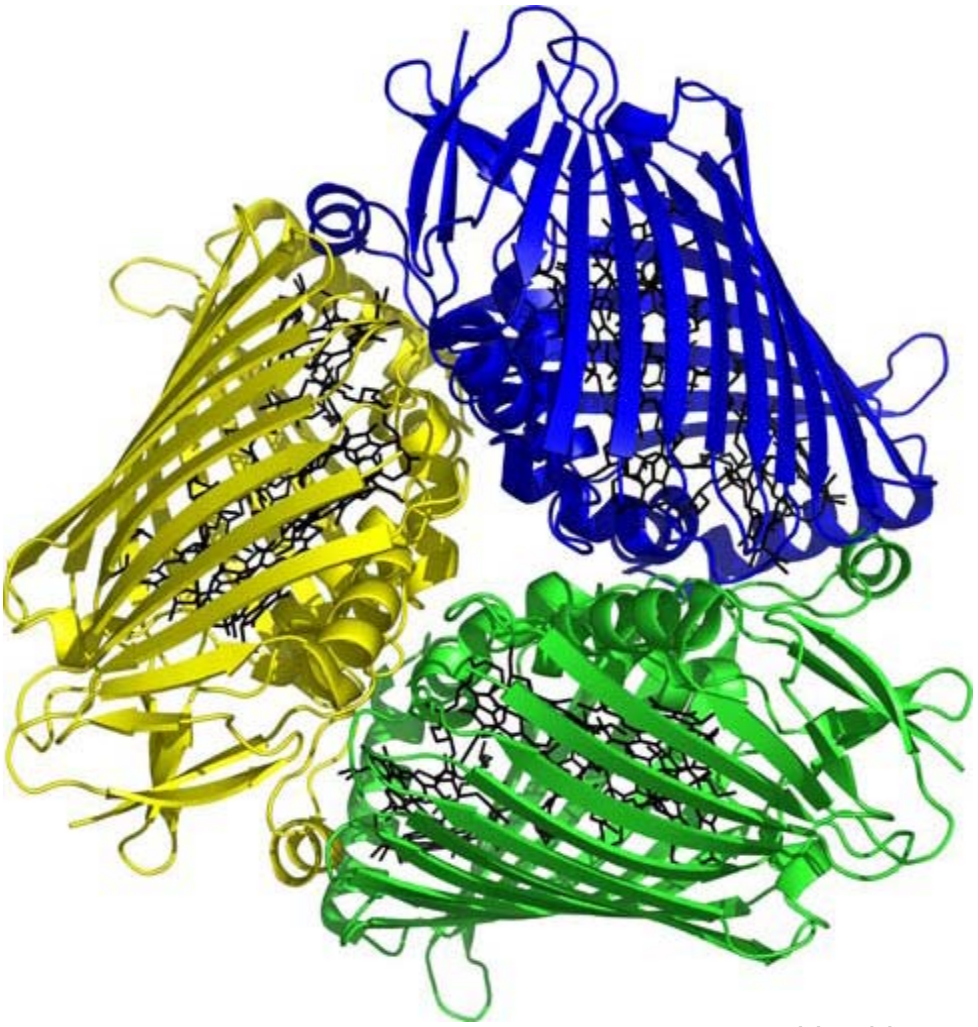

Figure 3.8: The three-dimensional structure of FMO from *Chlorobium tepidum*. Shown are the backbones of the three protein subunits (yellow, green, and blue) and the cofactors (black) of the trimer. The view is approximately down the threefold symmetry axis of the biologically active trimer. The coordinates are 1M50 (Camara-Artigas et al. 2003).

References:

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

2. Allen, J. P., Feher, G., Yeates, T. O., Komiya, H., and Rees, D. C. (1987) Structure of the reaction center from *Rhodobacter* sphaeroides R-26: the protein subunits. Proc*. Natl. Acad. Sci. USA 84*, 5730-5734.

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

4. Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C., and Shore, V. C. (1960) Structure of Myoglobin: A Three-Dimensional Fourier synthesis at 2Å. Resolution. *Nature 185*, 422-427.

5. Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G., and North, A. C. T. (1960) Structure of Haemoglobin: A Three-Dimensional Fourier Synthesis at 5.5Å. Resolution, Obtained by X-Ray Analysis. *Nature 185*, 416-422.

6. Zouni, A., Witt, H. T., Kern, J., Fromme, P., Krauss, N., Saenger, W., and Orth, P. (2001) Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8Å resolution. *Nature 49*, 739-743.

7. Kamiya, N., and Shen, J. R. (2003) Crystal structure of oxygen-evolving photosystem II from *Thermosynechococcus vulcanus* at 3.7Å resolution. *Proc. Natl. Acad. Sci. USA 100*, 98-103.

8. Ferreira, K. N., Iverson, T. M., Maghlaoui, K., Barber, J., and Iwata, S. (2004) Architecture of the photosynthetic oxygen-evolving center. *Science 303*, 1831-1838.

9. Loll, B., Kern, J., Saenger, W., Zouni, A., and Biesiadka, J. (2005) Towards complete cofactor arrangement in the 3.0Å resolution structure of photosystem II. *Nature 438*, 1040-1044.

10. Stowell, M. H. B., McPhillips, T. M., Rees, D. C., Soltis, S. M., Abresch, E., and Feher, G. (1997) Light-induced structural changes in photosynthetic reaction center: implications for mechanism of electron-proton transfer. *Science 276*, 812-816.

11. Baxter, R. H., Seagle, B. L., Ponomarenko, N., and Norris, J. R. (2005) Cryogenic structure of the photosynthetic reaction center of *Blastochloris* viridis in the light and dark. Acta *Crstallogr D 61*, 605-612.

12. Feher, G., Allen, J. P., Okamura, M. Y., and Rees, D. C. (1989) Structure and function of bacterial photosynthetic reaction centres. *Nature 339*, 111-116.

13. Hunter, N., Daldal, F., Thurnauer, M., and Beatty, J. T. (eds) (2008) *Springer-Verlag Publishers* Dordrecht, the Netherlands

14. Deisenhofer, J., Epp, O., Sinning, I., and Michel, H. (1995) Crystallographic refinement at 2.3Å Resolution and Refined Model of the Photosynthetic Reaction Centre from *Rhodopseudomonas viridis. J. Mol. Biol. 246*, 429-457.

15. Chang, C. H., El-Kabbani, O., Tiede, D., Norris, J., and Schiffer, M. (1991) Structure of the membrane-bound protein photosynthetic reaction center from *Rhodobacter sphaeroides. Biochemistry 30*, 5352-5360.

16. Ermler, U., Fritzsch, G., Buchanan, S. K., and Michel, H. (1994) Structure of the photosynthetic reaction centre from *Rhodobacter sphaeroides* at 2.65Å resolution: cofactors and protein-cofactor interactions. *Structure 2*, 925-936.

17. McAuley, K. E., Fyfe, P. K., Ridge, J. P., Isaacs, N. W., Cogdell, R. J., and Jones, M. R. (1999) Structural details of an interation between cardiolipin and an integral membrane protein. *Proc. Natl. Acad. Sci. USA* **96**, 14706-14711.

18. Camara-Artigas, A., Brune, D., and Allen, J. P. (2002) Interactions between lipids and bacterial reaction centers determined by protein crystallography. *Proc. Natl. Acad. Sci. USA. 99*, 11055-11060.

19. Debus, R. J., Feher, G., and Okamura, M. Y. (1985) LM complex of reaction centers from *Rhodopseudomonas sphaeroides* R-26: characterization and reconstitution with the H subunit. *Biochemistry 24*, 2488-2500.

20. Cheng, Y. S., Brantner, C. A., Tsapin, A., and Collins, M. L. P. (2000) Role of the H protein in assembly of the photochemical reaction center and intracytoplasmic membrane in *Rhodospirillum rubrum*. *J. Bacteriol. 182*, 1200-1207.

21. Tehrani, A., Prince, R. C., and Beatty, J. T. (2003) Effects of Photosynthetic Reaction Center H Protein Domain Mutations on Photosythetic Properties and Reaction Center Assembly in *Rhodobacter sphaeroides. Biochemistry 42*, 8919-8928.

22. Lupo, D., and Ghosh, R. (2004) The reaction center H subunit is not required for high levels of light-harvesting complex 1 in *Rhodospirillum rubrum* mutants. *J. Bacteriol 186*, 5585-5595.

23. Axelrod, H. L., and Okamura, M. Y. (2005) The structure and function of the cytochrome c 2: reaction center electron transfer complex from *Rhodobacter sphaeroides. Photosynth. Res 85*, 101-114.

24. Axelrod, H. L., Abresch, E. C., Okamura, M. Y., Yeh, A. P., Rees, D. C., and Feher, G. (2002) X-ray structure determination of the cytochrome c2: reaction electron transfer complex from *Rhodobacter sphaeroides*. *J. Mol. Biol. 319,* 501-515.

25. Miyashita, O., Okamura, M. Y., and Onuchic, J. M. (2005) Interprotein electron transfer from cytochrome c2 to photosynthetic reaction center: Tunneling across an aqueous interface. *Proc. Natl. Acad. Sci. USA 102*, 3558-3563.

26. Wydrzynski, T. J., and Satoh, K. (eds) (2005) Photosystem II: The Light-Driven Water: Plastoquinone Oxidoreductase. *Springer-Verlag Publishers*, Dordrecht, the Netherlands

27. Ben-Shem, A., Frolow, F., and Nelson, N. (2003) Crystal structure of plant photosystem I. *Nature 426*, 630-635.

28. Ben-Shem, A. , Frolow, F., and Nelson, N. (2004) Evolution of photosystem I-from symmetry through pseudosymmetry to asymmetry. *FEBS Lett 564*, 274-280.

29. Raszewski, G., Diner, B. A , Schlodder, E., and Renger, T. (2008) Spectroscopic properties of reaction center pigments in photosystem II core complexes: revision of the multimer model. *Biophys. J. 95,* 105-119.

30. Yano, J., Kern, J., Sauer, K., Latimer, M. J., Pushkar, Y., Biesiadka, J., Loll, B., Saenger, W., Messinger, J., Zouni, A., and Yachandra, V. K. (2006) Where water is oxidized to dioxygen: structure of the photosynthetic Mn4Ca cluster *Science 314*, 821-825.

31. Goldbeck, J. (ed) (2006) *Springer-Verlag Publishers* Dordrecht, the Netherlands

32. Krauss, N., Hinrics, W., Witt, I., Fromme, P., Pritzkow, W., Dauter, Z., Betzel, C., Wilson, K. S., Witt, H. T., and Saenger, W. (1993) Three-dimensional structure of system I of photosynthesis at 6 Å resolution. *Nature 361*, 326-331.

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

34. Krauss, N. (2008) In: Fromme, P. (ed) *Wiley-Blackwell*, 23-64

35. Amunts, A., Drory, O., and Nelson, N. (2007) The structure of a plant photosystem I supercomplex at 3.4Å resolution. *Nature 447*, 58-63.

36. Cogdell, R. J., Gall, A., and Köhler, J. (2006) The architecture and function of the light-harvesting apparatus of purple bacteria: from single molecules to *in vivo* membranes. *Q. Rev. Biophys. 39*, 227-324.

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

38. Papiz, M. Z., Prince, S. M., Howard, T., Cogdell, R. J., and Isaacs, N. W. (2003) The Structure and thermal Motion of the B800-850 LH2 Complex from *Rps. Aacidophila* at 2.0Å Resolution and 100K: New Structural Features and Functionally Relevant Motions. *J. Mol. Biol. 326*, 1523-1538.

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

40. Karrasch, S., Bullog, P. A., and Ghosh, R. (1995) The 8.5Å projection map of the light-harvesting compolex I from *Rhodospirillum rubrum* reveals a ring composed of 16 subunits. *EMBO J.14,* 631-638.

41. Roszak, A. W., Howard, T. D., Southall, J., Gardiner, A. T., Law, C. J., Isaacs, N. W., and Cogdell, R. J. (2003) Crystal structure of the

RC-LH1 core complex from *Rhodopseudomonas palustris Science 302*, 1969-1972.

42. Matthews, B. W., Fenna, R. E., Bolognesi, M. C., Schmid, M. F., and Olson, J. M. (1979) Structure of a bacteriochlorophyll *a*-protein from the green photosynthetic bacterium *Prosthecochloris aestuarii. J. Mol. Biol. 131*, 259-285.

43. Tronrud, D. E., and Matthews, B. W. (1993) In: Norris, J. , Deisenhofer, J. (eds) *Academic Press,* New York 13-21

44. Li, Y-F., Zhou, W., Blankenship, R. E., and Allen, J. P. (1997) Crystal structure of the bacteriochlorophyll *a* protein from *Chlorobium tepidum. J. Mol. Biol. 271*, 456-471.

45. Camara-Artigas, A., Blankenship, R. E., and Allen, J. P. (2003) The structure of the FMO protein from *Chlorobium* tepidum at 2.2Å resolution. *Photosynth. Res. 75*, 49-55.

46. Bixner, T., Stenger, J., Vaswani, H. M., Cho, M., Blankenship, R. E. , and Fleming, G. R. (2005) Two-dimensional spectroscopy of electronic couplings in photosynthesis. *Nature 434*, 625-628.

47. Read, E. L., Schlau-Cohen, G. S., Engel, G. S., Wen, J., Blankenship, R. E., and Fleming, G. R. (2008) Visualization of excitonic structure in the Fenna-Matthews-Olson photosynthetic complex by polarization-dependent two-dimensional electronic spectroscopy. *Biophys. J. 95*, 847-856.

Chapter 4

The three dimensional structure of the FMO protein from *Pelodictyon phaeum* and the implications for energy transfer

Abstract

The Fenna-Matthews-Olson (FMO) antenna protein from the green bacterium *Pelodictyon phaeum* mediates the transfer of energy from the peripheral chlorosome antenna complex to the membrane-bound reaction center. The three-dimensional structure of this protein has been solved using protein crystallography to a resolution limit of 2.0 Å, with R_{work} and R_{free} values of 16.6 % and 19.9 % respectively. The structure is a trimer of three identical subunits related by a 3-fold symmetry axis. Each subunit has two beta sheets that surround 8 bacteriochlorophylls. The bacteriochlorophylls are all five coordinated, with the axial ligand being a histidine, serine, backbone carbonyl, or bound water molecule. The positions and orientations of most of the bacteriochlorophylls are well conserved in comparison to other FMO structures, but differences are apparent in the interactions with the surrounding protein. Unlike the other cofactors, the eighth bacteriochlorophyll has differences in its locations and the coordination of the central Mg. The implications of this structure on the ability of the FMO protein to perform energy transfer are discussed in terms of the experimental optical measurements.

Introduction

A diverse family of pigment-protein antenna complexes in photosynthetic organisms capture light and direct this energy to the

integral membrane reaction center complexes where it is converted into chemical energy in the form of charge separation for the eventual creation of energy rich compounds *(1)*. In green sulfur bacteria, light is absorbed by chlorosomes, which are large complexes attached to the cytoplasmic side of the inner cell membrane *(2–4)*. The light energy is transferred from the chlorosomes to a bacteriochlorophyll (BChl) *a*-containing protein termed the Fenna-Matthews-Olson (FMO) protein *(5)*. The FMO protein is a water-soluble protein but is embedded in the cytoplasmic membrane and serves as an energy transfer funnel between the chlorosome and the integral membrane protein called the reaction center, which is the site of the conversion of the light energy to electron transfer *(6)*.

Due to the key role that FMO plays in the transfer of energy from the chlorosome to the reaction center, the properties of FMO have been subjected to considerable study. The presence of several BChl cofactors has provided the opportunity to probe the electronic states of the FMO after light excitation using steady-state and transient optical spectroscopy *(7)*. However, the BChl cofactors are highly interacting, which makes assignment of the spectral features to individual cofactors problematic, especially since the energy transfer processes involve quantum effects *(8–11)*. Time-resolved 2D optical spectroscopy has provided the experimental means to probe the couplings between the BChl cofactors although interpretation at a molecular level requires assignment of the optical transitions *(12–14)*.

The FMO complex from *Prosthecochloris aestuarii* (*Ptc. aestuarii*) 2K was the first protein containing BChl to be crystallized *(15)* and have its three dimensional structure determined using X-ray diffraction *(16–17)*. Subsequently, the structure of the FMO complex from *Chlorobaculum tepidum* (*Cbl. tepidum*), previously named *Chlorobium tepidum*, was determined *(18–19)*. Both complexes are trimers, with three identical subunits that are related by a 3-fold symmetry axis. Each protein subunit was found to embed seven BChl cofactors although more recent structures have identified an eighth BChl cofactor that is present with a range of occupancies *(20–21)*. The structure of *Cbl. tepidum* is similar to that of *Ptc. aestuarii* 2K reflecting the significant sequence homology found for the FMO proteins *(22)*. To understand the relationship between the structure and energy transfer function, we have determined the three-dimensional structure of the FMO protein from a third organism, *Pelodictyon phaeum* (*Pld. phaeum*), that had been predicted to have significant structural differences compared to *Cbl. tepidum* and *Ptc. aestuarii* 2K based upon a comparison of the spectroscopic properties *(23)*. In this paper, the three-dimensional structure of the FMO from *Pld. phaeum* as determined using protein crystallography is described. The availability of this new structural model presents the opportunity to re-examine the structural aspects that give this complex the ability to perform energy transfer with unusual quantum contributions.

Cell growth and protein purification

The *Pld. phaeum* cells were grown anaerobically at room temperature with ~100 µE light intensity for 3-5 days in two 15 L sealed carboys. The FMO protein was isolated essentially as previously described *(18)*. The FMO protein was extracted from the membrane by Na_2CO_3, and collected as a supernatant after ultracentrifugation. The crude FMO extract was dialyzed against Tris/HCl buffer (pH 8.0) until the pH dropped to 8.0. The FMO protein was then purified by a combination of a Q Sepharose HP ion exchange column (GE Healthcare, USA) and an S-300 Sephacryl HR gel filtration column (GE Healthcare, USA) until the final $OD_{267}/OD_{371} < 0.56$. The protein was stored in 20 mM Tris/HCl pH 8.0 and 50 mM trisodium citrate prior to crystallization.

Crystallization

Crystals of the FMO protein were obtained using the hanging drop method. The protein sample concentration was poised at an absorbance of 9.0 at 810 nm, corresponding to a concentration of approximately 6 mg/mL, in a 50 mM disodium citrate buffer. The original conditions tested for crystallization were based upon the conditions used to crystallize the FMO from *Cbl. tepidum (18)*, which had the protein solution mixed with an equal volume of the reservoir containing 6% polyethylene glycol 4000, 20% 2-propanol, and 0.1 M sodium citrate (pH 5.6). The optimal conditions had a reservoir solution containing 0.1 mM 4-(2-hydroxyethyl)-1-pipersazineethanesulfonic acid (pH 7.5), 16% polyethylene glycol 2000 monomethyl ether with a volume of 0.4 mL. The protein drop had a

volume of 2 μL and was poised at 3 mg/mL with 0.05 mM 4-(2-hydroxyethyl)-1-pipersazineethanesulfonic acid (pH 7.5), 25 mM citrate and 4% polyethylene glycol 2000 monomethyl ether. The trays were kept at room temperature in the dark with blue-green crystals appearing in two to three days and reaching full size in two weeks. Prior to the diffraction measurements, the crystals were placed into a cryoprotectant consisting of 60% polyethylene glycol 400, 10% polyethylene glycol 2000 monomethyl ether and 10 mM citrate.

Data Collection and Refinement

The crystals of FMO from *Pld. phaeum* grew in the dark and as green hexagonal rods with a typical length of 0.1 mm and 0.05 mm width. The crystals belong to the hexagonal space group $P6_3$ with unit cell dimensions of a = b = 84.0 Å and c = 115.7 Å along with α = β = 90° and γ = 120° with one protein subunit per asymmetric unit (Table 4.1). Diffraction data were measured from a single crystal at the Brookhaven National Laboratory on the NSLS-X126 beamline using an ADSC detector at a wavelength of 1.081 Å. A full diffraction data set was measured to a resolution limit of 1.99 %, integrated using MOSFLM *(24)*, and scaled with SCALA *(25)*. The initial phases were determined using the molecular replacement method using PHENIX *(26)* with the 3EOJ model of the FMO from *Ptc. aestuarii* 2K *(21)*. A unique orientation and position were obtained. Manual model building for all structures was performed in Coot *(27)* with iterative rounds of refinement being performed using PHENIX.

Procheck *(28)* and Sfcheck *(29)* were applied to verify the completed structures. The completed model has R_{work} and R_{free} values of 16.6 % and 19.9 % respectively. All of the amino acids are evident within the electron density except for the first four amino acids at the N-terminus and two residues 209 and 210 that are part of a loop between two "-strands. Two prolines, 41 and 324, are in a cis conformation. The Ramachandran plot showed 93.3 % in the preferred region and 6.7 % in the allowed regions and no outliers. The average temperature factors for the protein and water molecules are 21.0 and 37.8 Å2 respectively yielding an overall value of 22.7 Å2. The root mean square deviations of bond distances and angles are 0.007 Å and 3.0 °, respectively. Figures were made using Pymol *(30)*.

Results

Three-dimensional structure of the FMO protein from Pld. phaeum

The FMO protein from *Pld. phaeum* consists of three identical subunits with each subunit consisting of 362 amino acids that form a number of long and short β-strands and several α-helices (Figure 4.1). The two large β-sheets form a ‘taco shell’ surrounding the eight BChl cofactors. β-strands 1 and 2 run anti-parallel to each other and start the largest β-sheet. Adjacent to strand 2 are anti-parallel β-strands 11 - 12 and 3-4 and adjacent to strand 1 are strands 13-17. The second β-sheet wraps around behind the first sheet and is formed by β-strands 5-10. The open section of the two large β-sheets is closed by the presence of 6 α-helices and connecting loops. Helix 1 is situated behind the largest β-

sheet and is exposed at the surface of the subunit. Three identical subunits are arranged as a trimer with each subunit related by a crystallographic 3-fold axis of symmetry (Figure 4.2). The trimer is approximately cylindrical in shape with an 8 nm diameter and a 5 nm height. The trimer is positioned in the cell membrane with the 3-fold symmetry axis approximately perpendicular to the membrane surface, with one face oriented towards the chlorosome and the other face towards the membrane *(6, 31)*. This orientation places the more hydrophobic surface of the trimer interacting with the baseplate of the chlorosome and the more hydrophilic surface partially embedded into the phospholipid head groups of the cell membrane and adjacent to the reaction center. The resulting alignment of α-helices 5 and 6 towards the membrane has been proposed to facilitate energy transfer to the reaction center *(10)*.

The three interacting subunits have a number of contacts between each subunit that hold the trimer together, primarily involving β-strand 8 and α-helices 1, 2 and 3. A large number of protein-protein interactions are found at the interfaces between the subunits that contribute to the stability of the trimer (Figure 4.3). The predominant amino acids are polar and charged residues Asp, Asn, Glu, Arg, Lys and Ser, many of which form salt bridge contacts between the subunits. Most of these interactions are well conserved, for example in one region of the interface, Arg 127 of α-helix 1 forms a bridge with Asp *103 and Asp *153 that is in α-helix 2 of the neighboring monomer. Asn 128 also forms a salt bridge contact with

Asp *103. In other regions, Arg 138 of β-strand 8 forms a bridge with Glu *178 of α-helix 3. Tyr 119-Glu *178, where * represents the neighboring monomer, is one such polar interaction. Also, Arg 195 is found in a loop region and forms a bridge with Glu *178 of "-helix 3 and Asp *174 also from α-helix 3. While these interactions are generally conserved, there is a unique salt bridge that occurs in *Pld. phaeum* between Asp 123 and Glu *156.

Each subunit surrounds seven BChl molecules, identified as BChl 1 through 7, with an eighth BChl, identified as BChl 8, on the interface with the neighboring protein subunit of the trimer (Figure 4.4). The BChl molecules are arranged in an asymmetric fashion. Considering only BChl 1 through 7, the closest distance between any two BChls within any subunit, as measured by the Mg to Mg distance, is 11 Å and the farthest distance is 30 Å. None of the BChls from one subunit are near any from the neighboring subunit, with the closest distance being 25 Å. In contrast, BChl 8 is far from the other seven BChls within each subunit, with distances ranging between 20 and 40 Á, with BChl 6 being the closest, but has a close distance of 12 Å to BChl 1 of the neighboring subunit.

BChls 1 through 7 are all five-coordinated, to either a histidine, a backbone carbonyl, or a bound water molecule. His 106 located in β-strand 7 coordinates BChl 1. BChl 2 is coordinated by a water molecule in close proximity to the side chain of Asn 75. His 294 and His 293 located in α-helix 7 coordinate BChl 3 and 7 respectively. The long loop connecting

β-strands 12 and 13 contains His 286 that serves as the axial ligand for BChl 4 and BChl 6 is coordinated to His141 of β-strand 8. The backbone carbonyl of Leu 238 of α-helix 5 ligates BChl 5. For the FMO from *Pld. phaeum*, BChl 8 is five-coordinated to the side chain of Ser 164 but, unlike the other seven BChls, BChl 8 has many interactions with amino acid residues from both subunits with notable differences found among the FMO structures as discussed below.

Discussion

The function of FMO is to transfer light energy from the chlorosomes to the reaction center where energy conversion occurs. The structure of FMO from *Pld. phaeum* has been solved at a resolution limit of 2.0 Å. The FMO protein is a trimer with each subunit forming a ‘taco shell’ arrangement of two β sheets arranged with α helices filling the open end of the shell. There are numerous interactions between neighboring subunits that stabilize the trimer, which is the biologically relevant structure. The two β-sheets surround a total of eight BChl molecules that are arranged with various separation distances without any symmetrical pattern. Seven of the BChls are buried within the protein and five-coordinated. The eighth BChl lies on the edge of the subunit at the interface formed with the neighboring subunit of the trimer.

The backbone structure of the FMO from *Pld. phaeum* closely resembles those of the previously solved structures of the FMO protein from *Ptc. aestuarii* 2K and *Cbl. tepidum (16–21)*. The structures from the

three species can be closely overlaid showing all secondary structural features conserved, with an average rms deviation of 0.5 Å, with only minor differences for the connecting loops. This close match of the backbone reflects a strong sequence homology observed among the FMO proteins. An alignment of 17 FMO protein sequences, including *Pld. phaeum*, *Ptc. aestuarii* 2K and *Cbl. tepidum (22)*, shows that most amino acid residues are conserved, with the notable exception of an aerobic phototrophic bacterium, *Candidatus* Chloracidobacterium thermophilum although in all species the critical amino acid residues forming the binding sites for the BChls are identical. Among the FMO from green sulfur bacteria, the sequences from *Pld. phaeum* and *Ptc. aestuarii* 2K are very closely conserved with both showing differences compared to *Cbl. tepidum*.

In the alignment of the FMO from green sulfur bacteria, the longest continuous conserved length of amino acids is 14 residues with other stretches of 10, 8 and 7 residues also occurring. There are multiple regions of 2 to 6 amino acids that are fully conserved throughout the sequences. Many of these conserved sections are clustered close to the coordinating residues for the BChl molecules (Figure 4.5). The longest conserved stretches consisting of at least 7 amino acids are clustered around BChls 3 and 6. BChl 6 has two short stretches of conserved residues (numbering for *Pld. phaeum*) that consist of residues 110—116 and 136—143. BChl 3 has three different stretches of conserved residues

that are in close proximity to it. The longest stretch has 14 residues numbering 254-267. The two smaller stretches include residues 290—299 and 346—352. It is interesting but not surprising that these two BChls have the longest conserved sections of amino acids. BChl 3 is believed to be the lowest energy pigment that transfers energy from FMO to the reaction center and BCh 6 is believed to be one of the high-energy pigments as deduced from spectral fitting studies *(8, 10, 32–34)*. Thus, the strong conservation of amino acid residues surrounding BChl 3 and 6 reflects the key roles that these two cofactors play in the energy transfer function of the FMO protein.

This conservation of amino acid residues is also seen in the protein-protein interactions that are found in the interface between neighboring subunits of the trimer. Comparing the FMO proteins from *Pld. phaeum*, *Ptc. aestuarii* 2K and *Cbl. tepidum*, the types of interactions are the same in all three structures with few notable differences (in comparing equivalent residues, the residue numbers for *Pld. phaeum* are shifted by four residues compared to *Ptc. aestuarii* 2K and *Cbl. tepidum*). For example, in one region the same salt bridges are present in all three structures involving amino acids from two different subunits, Asp 103, Asp 107, Asp 127, Arg 131, and Asp 157 (Figure 4.3). The conservation of these interactions presumably is due to the trimeric nature of FMO in all species.

Comparison of Protein – BChl interactions in the three FMO structures

BChl 1 has its Mg coordinated with the same histidine axial ligand, namely His 106 in *Pld. phaeum* and 110 in *Cbl. tepidum* and *Ptc. aestuarii* 2K, but four differences are evident in the interactions with amino acid residues lying within 10 Å of the metal center (Figure 4.6). Within van der Waals distance of the macrocycle is Val 99 in *Pld. phaeum* and *Cbl. tepidum* (Val 103) but *Ptc. aestuarii* 2K has Leu 103 in that equivalent location. Both *Pld. phaeum* and *Ptc. aestuarii* 2K have a nearby Phe, namely Phe 161 and Phe 165 respectively, but the interaction with an aromatic residue is lost with the presence of Thr 165 in *Cbl. tepidum*. Also a serine is present for both *Pld. phaeum* and *Cbl. tepidum* at amino acid residues 217 and 220 respectively compared to Thr 221 for *Ptc. aestuarii* 2K. At 10 Å, a difference is also found with *Pld. phaeum* having Lys 107 but *Ptc. aestuarii* 2K has Ser 111 and *Cbl. tepidum* has Thr 111.

A comparison of BChl 2 from the three species shows that there are a number of differences within a 10 Å environment (not shown). This BChl is located at a protein-protein interface of the trimer and interacts with several residues from the adjacent subunit that are highly conserved. Unlike the other BChls, a bound water molecule rather than the protein coordinates the central Mg. Different amino acids are adjacent to ring B in each structure, *Pld. phaeum* (Leu 37), *Cbl. tepidum* (Ser 41) and *Ptc. aestuarii* 2K (Ile 41). Next to this amino acid residue is an Ala in *Pld. phaeum* and *Cbl. tepidum*, positions 42 and 46 respectively, but *Ptc. aestuarii* 2K has Thr 46. Near ring A, *Pld. phaeum* and *Ptc. aestuarii* 2K

have Ile 66 and Ile 70 respectively, but *Cbl. tepidum* has Phe 70 in that location. Close to ring C is a Val in both *Pld. phaeum* and *Cbl. tepidum*, Val 99 and Val 103 respectively, while Leu 103 is found in *Ptc. aestuarii* 2K. A difference further removed from the BChl is Thr 73 and Thr 77 in *Pld. phaeum* and *Cbl. tepidum*, respectively, but Val 77 in *Ptc. aestuarii* 2K.

The region surrounding BChl 3 is highly conserved across all three species of bacteria for over 20 amino acids within a 10 Å distance from the Mg, including the histidine coordinating the central Mg, namely His 294, 298, and 297 in *Pld. phaeum*, *Ptc. Aestuarii* 2K, and *Cbl. tepidum* respectively (Figure 4.7). One minor difference is that both *Pld. phaeum* and *Ptc. aestuarii* 2K have an Ala at residues 35 and 39 respectively, but *Cbl. tepidum* has Pro 39. Also, *Cbl. tepidum* has a polar residue, Ser 41 rather than an aliphatic residue, namely Leu 37 and Ile 41, as found in *Pld. phaeum* and *Ptc. aestuarii* 2K respectively. This conservation of the binding site (Figure 4.5) is consistent with the hypothesis that the protein interactions need to be maintained in order for BChl 3 to serve its role of being the site where the exciton energy collects within the FMO protein before the energy is transferred to the reaction center *(32–34)*.

The environment around BChl 4 is also highly conserved with roughly 24 amino acids in close proximity, including the coordinating histidine residue, namely His 286, 290, and 289 in *Pld. phaeum*, *Ptc. Aestuarii* 2K, and *Cbl. tepidum* respectively (not shown). Three differences

are observed at a distance of about 10 Å from the central magnesium atom. Both *Pld. phaeum* and *Ptc. aestuarii* 2K have Leu 269 and Leu 273, respectively while the corresponding residue is Phe 272 in *Cbl. tepidum*. A second difference is that *Pld. phaeum* has Ile 25 where *Cbl. tepidum* and *Ptc. aestuarii* 2K both have Val 29 in that position. The third difference has *Cbl. tepidum* differing again from *Pld. phaeum* and *Ptc. aestuarii* 2K, Ala 34 in *Cbl. tepidum* and Gly 27 and Gly 31 in *Pld. phaeum* and *Ptc. aestuarii* 2K respectively.

There are a number of differences involving the interactions of BChl 5 with the surrounding protein although the coordination of the central Mg to the backbone oxygen of Leu 238, or equivalently Leu 242 and Leu 241 in *Ptc. aestuarii* 2K and *Cbl. tepidum* respectively, is present in all structures (Figure 4.8). Of the 27 amino acids that lie at the 10 Å distance from the central Mg of the BChl, only six are closer than 8 Å and two of those six are not conserved. Near ring B, both *Pld. phaeum* and *Ptc. aestuarii* 2K have an aromatic residue, namely Phe 62 and Phe 66 respectively, but *Cbl. tepidum* has Ile 66. Near ring D is the difference that *Cbl. tepidum* has Phe 272 and *Pld. phaeum* and *Ptc. aestuarii* 2K have the aliphatic residue Leu 269 and Leu 273, respectively. In addition, there are several smaller changes. The first minor difference is that *Pld. phaeum* and *Ptc. aestuarii* 2K have Ile 48 and Ile 52 respectively, while *Cbl. tepidum* has Leu 52. The second is that *Pld. phaeum* and *Ptc. aestuarii* 2K are alike with Ala 50 and Ala 54 but *Cbl. tepidum* has Val 54. The third is

that *Pld. phaeum* and *Ptc. aestuarii* 2K are identical with Val 241 and Val 245 and *Cbl. tepidum* has Ser 244. The final minor difference is that *Pld. phaeum* and *Ptc. aestuarii* 2K both have an Ile at 250 and 254 respectively, while *Cbl. tepidum* differs with Val 253 in that position.

The protein environments of BChls 6 and 7 are highly homologous with most amino acid residues being identical, including the coordinating histidines, namely residues 141 and 293 in *Pld. phaeum*, 145 and 297 in *Ptc. Aestuarii* 2K, and 145 and 296 in *Cbl. tepidum* (not shown). One exception for the region of BChl 6 is near ring A where Phe 62 and Phe 66 are present in *Pld. phaeum* and *Ptc. aestuarii* 2K respectively but *Cbl. tepidum* has Ile 66. For BChl 7, Ile is found near ring A in *Pld. phaeum* and *Cbl. tepidum*, namely positions 189 and 192 respectively, while Ile 193 is present in *Ptc. aestuarii* 2K.

Comparison of position and orientation of BChl 8 in the three FMO structures

The three-dimensional structures of the FMO proteins from *Cbl. tepidum* and *Ptc. aestuarii* 2K originally had seven BChls whose positions were well identified in the electron density calculated at resolution limits of 1.9 – 2.2 Å *(16–19)*. In an effort to crystallize the FMO-reaction center complex, new conditions were identified that yielded the structure of the FMO protein at a resolution limit of 2.4 Å *(20)*. Despite the lower resolution limit, a new feature in the electron density maps was the clear presence of an eighth BChl at the interface region between subunits in the trimer. This BChl is five coordinated to the backbone carbonyl of Tyr 124. By

comparison, the electron density for the eighth BChl was very incomplete in the other data sets with the electron density in that region being unidentifiable *(21)*. Presumably, BChl 8 is labile and in the course of purifying the FMO protein this cofactor can be lost. For the FMO protein from *Ptc. aestuarii* 2K, the electron density in that region was modeled as arising from the protein in two different conformations, one with the site occupied and the other with the cofactor missing *(21)*. In those models the unoccupied conformation has some large changes, for example Tyr 124 and seven bound water molecules occupy the position normally occupied by BChl 8. From this density, it is difficult to assign the precise location and orientation of the BChl, and the model has the cofactor located based upon the model of the FMO from *Cbl. tepidum*. With this positioning, BChl 8 has two axial ligands provided by the side chain of Ser 168 in addition to the backbone carbonyl of Asp 123.

The stabilization of BChl 8 presumably is influenced by a number of factors in the protein preparation and crystallization procedures *(18, 20)*. For example, while the resuspension methods for the cell pellets were similar, the purification of the FMO protein used an initial pH of 9.0 followed by sonication compared to the RC purification protocol that utilized a pH of 7.3 followed by cell disruption with a French press. In addition, a high concentration of detergent was used to extract the FMO-RC complex from the membrane and a detergent was included in all the following purification steps, while the FMO purification rarely involves

using detergent. Moreover, the crystallization of the purified FMO protein made use of 10-20 mM Tris-HCl pH 8.0, 12% polyethylene glycol 600 or 16 % (w/v) polyethylene glycol 4000, 50-100 mM sodium citrate and 10-20 % (v/v) 2-proponal. Several contrasting differences are apparent in the crystallization of the RC-FMO complex. The crystallization was initiated with the RC-FMO complex, but the RC aggregated and released the FMO from the complex. The crystallization conditions contained a different precipitation agent, 0.2 M magnesium chloride, a different pH, 0.1 M 4-(2-hydroxyethyl)-1-pipersazineethanesulfonic acid (pH 7.5), and a different agent, 2.6 M hexanediol.

This region of the FMO protein from *Pld. phaeum* had a significant amount of unassigned electron density for maps calculated using a refined model with BChl 8 not present in the model. Immediately adjacent to this region is the aromatic amino acid residue, Phe 161. Placing the BChl at a location identical to BChl 8 from *Cbl. tepidum* is not possible due to a pronounced steric clash with Phe 161. This hindrance is not observed in *Cbl. tepidum*, for which the corresponding residue is the much smaller amino acid Thr 165. While the corresponding residue in *Ptc. aestuarii* 2K is also a Phe, that model has the same placement of BChl 8 only in the unoccupied model with the aromatic side chain swung out away from BChl 8 in the occupied model. As a check on the placement of Phe 161 in FMO from *Pld. phaeum*, the aromatic side chain was deleted from the structure and the resulting electron density map showed strong electron density in

the modeled position near BChl 8, which is the major rotamer orientation, and no density in any alternate orientation. Starting with the electron density map calculated without BChl 8, the position and orientation of the BChl cofactor was moved to be in the electron density while avoiding any van der Waals interactions with the surrounding amino acid residues. After consideration of several positions and orientations followed by refinement, a unique optimized position and orientation was found that resulted in structure with a favorable R_{work} and R_{free} of 16.6 and 19.9 % respectively. In this refined model, the macrocycle and substitutents of BChl 8 are very well covered by the electron density, including the beginning of the (C-C-CO-O-phytyl) chain on ring D (Figure 4.9). The site is fully occupied with a formal calculation based upon the electron density yielding a value of 87 %. The surface-exposed region, namely rings B and C, do have a pronounced increase in disorder compared to the buried regions, as measured by the temperature factors that have average values in those regions of 70 and 40 $Å^2$ respectively.

In this refined position, the BChl 8 is five coordinated with the axial ligand being the side chain of Ser 164 from the adjacent subunit of the trimer (Figure 4.10). This aspect of the structure significantly differs from the FMO from *Cbl. tepidum* that has coordination to the backbone carbonyl of Tyr 123. In *Ptc. aestuarii* 2K, the central Mg is six coordinated, to the backbone carbonyl of residue 123 as well as the side chain of Ser 168. In *Pld. phaeum*, the Mg is much farther at 6.5 Å from this carbonyl (of

the equivalent Tyr 119) and although a bound water molecule is nearby it is too far to coordinate the Mg. Thus, the BChl 8 of *Pld. phaeum* is not six coordinated as predicted previously *(21)* and assignment of type 1 and 2 spectra based only on the protein sequence in this region is not valid. In addition to the displacement of the FMO from *Pld. phaeum* compared to *Cbl. tepidum* and *Ptc. aestuarii* 2K, the macrocycle is also rotated by approximately 120° resulting in the rings A and C of *Pld. phaeum* being in positions roughly similar to rings C and D in *Cbl. tepidum* and *Ptc. aestuarii* 2K. Thus, both the position and orientation of BChl 8 in *Pld. phaeum* is strikingly different compared to the nearly identical models found in *Cbl. tepidum* and *Ptc. aestuarii* 2K.

In addition to the coordination and positioning of BChl 8, comparing the FMO structures of the three species to each other shows that there are multiple differences in the amino acids within the 10 Å environment. Five of these differences involve interactions with the neighboring subunit and two substitutions occur on the parent subunit. As discussed above, the most substantial difference is the presence of Phe of 161 in *Pld. phaeum* compared to Thr 165 in *Cbl. tepidum*. Another change again shows that *Cbl. tepidum* is different with Lys 167 and both *Pld. phaeum* and *Ptc. aestuarii* 2K have Gln 163 and Gln 166, respectively. Two other differences are *Pld. phaeum* and *Ptc. aestuarii* 2K having Met, at 117 and 121 respectively, and Tyr, at 118 and 122 respectively, compared to Leu 124 and Phe 125 in *Cbl. tepidum*.

Implications for optical spectroscopy and energy transfer

The FMO proteins have a characteristic optical absorption spectrum at room temperature, with a Q_Y peak at 808-810 nm, a Q_X peak at 600-602 nm, and a Soret band at 370-371 nm. Correspondingly, the fluorescence emission has a peak at 814-816 nm. However, at 77 K, both the absorption and circular dichroism spectra show distinct components, with the FMO protein from *Pld. phaeum* having a number of distinct spectral features compared to the FMO proteins from *Ptc. aestuarii* 2K, *Cbl. tepidum*, *Chlorobium limicola*, *Chlorobium vibrioforme*, and *Chlorobium phaeovibrioides (21, 23)*. For *Pld. phaeum*, three peaks are evident in the spectrum in the Q_Y region centered at 805 nm, 814 nm, and 824 nm with a shoulder at 795 nm. The intensity of the band at 806 nm was the largest among these five spectra. For the other four species, the relative intensities differ and the peaks are slightly shifted relative to the FMO from *Pld. phaeum*. For example, the spectrum of *Cbl. tepidum* has peaks at 805 nm, 814 nm, and 824 nm but the ratio of the amplitudes of the 815 peak compared to the 806 nm peak is less than one in contrast to a ratio greater than one for *Pld. phaeum*.

The CD spectrum of *Pld. phaeum* also shows differences when compared to the spectra of FMO from the other species *(23)*. At room temperature, FMO from *Pld. phaeum* shows a maximum at 797 nm and two minima at 780 nm and 821 nm. Small differences are evident in the position of these features in the other spectra. For example, these three

features are also found in *Cbl. tepidum* but the maximum occurs at 796 nm and the minima occur at 780 nm and 824 nm. The largest discernable difference between the two species is the absence of a shoulder at 808 nm in the spectrum of *Pld. phaeum*. These spectral differences are more pronounced when the temperature is lowered. The spectrum from *Pld. phaeum* has one positive peak at 802 nm and two minima at 790 nm and 821 nm. There are also two shoulders that are found at 810 nm and 817 nm. The low temperature CD spectrum from *Cbl. tepidum* has two positive peaks at 800 nm and 813 nm, and two minima at 807 nm and 825 nm. There are no shoulders seen in the low temperature CD spectrum of *Cbl. tepidum*.

Detailed theoretical analyses have been performed on understanding the energetics and energy transfer properties of the FMO proteins *(7)*. Key to these functional properties is the extensive electronic couplings among the BChl molecules. The spectral features of the low temperature absorption and circular dichroism spectra have been modeled in terms of these excitonic interactions among the BChl pigments. Thus, the differences in the spectral features of the FMO protein from *Pld. phaeum* compared to the other species reflects differences in the couplings that could arise due to a number of structural factors. For BChls 1-7, the position, orientation, and protein interactions of the three structures are largely similar; in particular the coordination is identical in all three cases. Some differences are apparent in the protein environment,

although the impact of these differences on the spectral properties is not established. The planarity of the macrocycles can play a critical role in the electronic structure as discussed elsewhere *(18)*. Thus, the spectral differences of *Pld. phaeum* compared to FMO from the other species should reflect the effect of such structural differences.

The theoretical studies of the spectra and energy transfer properties of the FMO have focused on individual subunits containing seven BChl molecules and consequently seven exciton levels *(7)* although a hole-burning study was interpreted in terms of eight exciton components *(35)*. While BChls 1-7 have close interactions within each subunit of the trimer, the closest distance between any of these BChl from two different subunits is over 23 Å. Due to this large distance, the BChls of each subunit are modeled as being functionally independent of each other. The presence of BChl 8 alters this interpretation, as this cofactor bridges the neighboring subunits of the trimer and is close to BChl 1 of the neighboring subunit (Figure 4.11). While many different factors influence the coupling *(36)*, since one of the primary factors is the separation distance, BChl 8 should be coupled with BChl 1 and influence the spectral features associated with BChl 1 and 2. If BChl 8 is indeed excitonically coupled to BChl 1, then the difference in the spectral features for FMO from different species may predominately arise due to differences in the position of BChl 8, with an increase of 2.5 Å in the distance between BChl 1 and 8 for *Pld. phaeum* compared to the distances found in *Cbl. tepidum*

and *Ptc. aestuarii* 2K. Such an impact is also suggested by two-dimensional spectroscopic measurements of the FMO from *Pld. phaeum* that show disagreement between the calculated and observed couplings *(13)*. In particular, theoretical calculations suggest that BChl 1 has a strong influence on excitons corresponding to the optical transitions at 790 and 809 nm and so the addition of couplings involving BChl 8 would impact this optical region as found in the circular dichroism spectra.

In summary, the three-dimensional structure of FMO from *Pld. phaeum* has been determined to a resolution limit of 2.0 Å. A strong conservation is evident for many structural features, including the arrangement of BChls 1–7 and the protein region surrounding BChl 3. However, pronounced differences are also apparent, in particular the position, orientation, and coordination of BChl 8. A comparison of the structural aspects should account for the observed spectroscopic similarities and differences, especially when the contribution of BChl 8 is included. Together these FMO structures provide a structural platform for understanding the quantum effects and other features involving the energy transfer function of this BChl-protein complex.

Table 4.1. Crystallographic data summary of FMO from *Pelodictyon phaeum*[1]

Space group	$P6_3$
Resolution limits (Å)	48.0 – 1.99 (2.05 – 1.99)
Unit Cell (Å)	a = b = 84.0, c = 115.8
Angle (°)	α = β = 90, γ = 120
Data Collection	
Beamline	NSLS-X126
Wavelength (Å)	1.0809
Reflections:	
Total	230,676
Unique	29,967
Multiplicity	7.7
I/σ (I)	2.0 (1.2)
R_{Merge} (%)[2]	0.186 (0.58)
Completeness (%)	94.5 %
Refinement	
R_{work} (%)/R_{free} (%)[3]	16.6/19.9
Number in asymmetric unit	1
Average B-factor (Å^2)	22.7
Residues modeled (#)	4 - 208, 211 - 362
Number of ligand/ion atoms	0
Number of water molecules	363
RMSD Bond length (Å)	0.007
RMSD Bond angle (°)	3.027
Ramachandran plots	
Preferred (%)	93.3
Allowed (%)	6.8
Outlier (%)	0

[1] *Numbers in the parentheses are for the outer shell*

[2] $R_{merge} = \Sigma_{hkl}\Sigma_j |I_j(hkl) - [I(hkl)]| \Sigma_{hkl}\Sigma_j[I(hkl)]$ *where* $I_j(hkl)$ *is the intensity reflection, and I(hkl) is the mean intensity of symmetry related h, k, l*

[3] $R_{work} = \Sigma| |F_{obs}| - |F_{cal}|| / \Sigma |F_{obs}|$, R_{free} *was calculated by setting aside 5% of the reflection data*

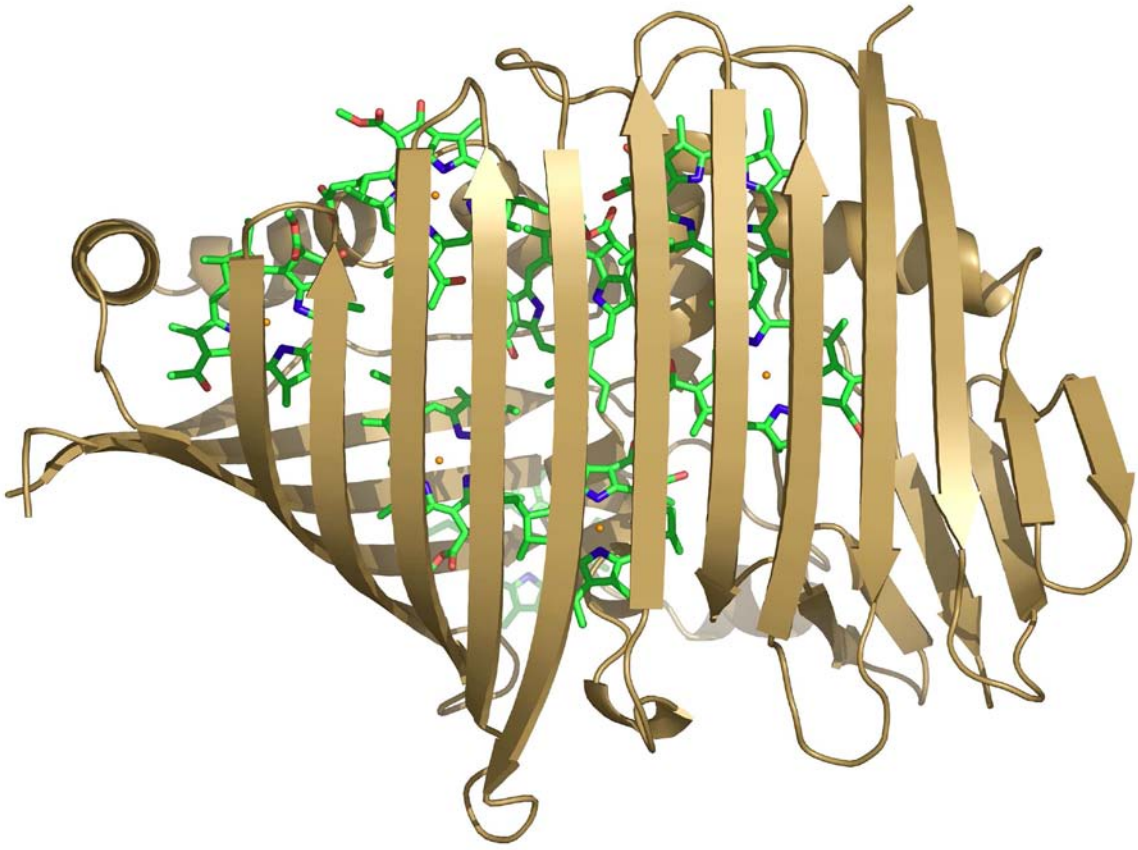

Figure 4.1. The three-dimensional structure of an individual subunit of FMO from Pld. Phaeum. The protein subunit, which is shown as a ribbon diagram (sand), consists largely of #-strands surrounding eight BChl cofactors (atom type). The ‘taco shell’ fold of the FMO monomer encases seven of the BChl cofactors within the folded protein while the eighth BChl cofactor is more peripheral.

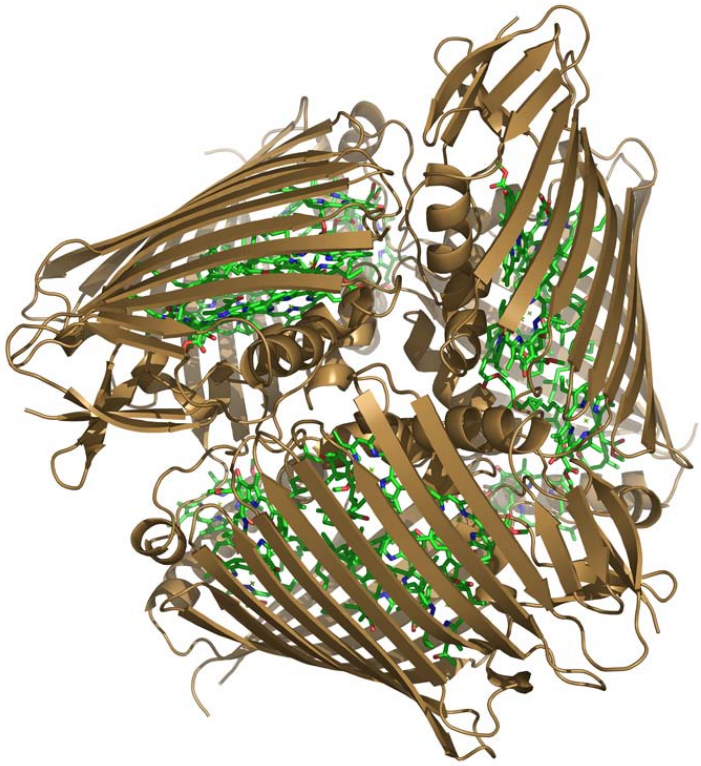

Figure 4.2. The three-dimensional structure of FMO from Pld. phaeum. In the cell, the protein is a trimer with the three identical subunits related by a crystallographic 3-fold symmetry axis approximately vertical to the paper. The protein subunits (sand) surround eight BChls cofactors (atom type). The α-helices close off the fold and provide important protein-protein contacts to form the trimer. BChl 8 is located at the interface between two subunits.

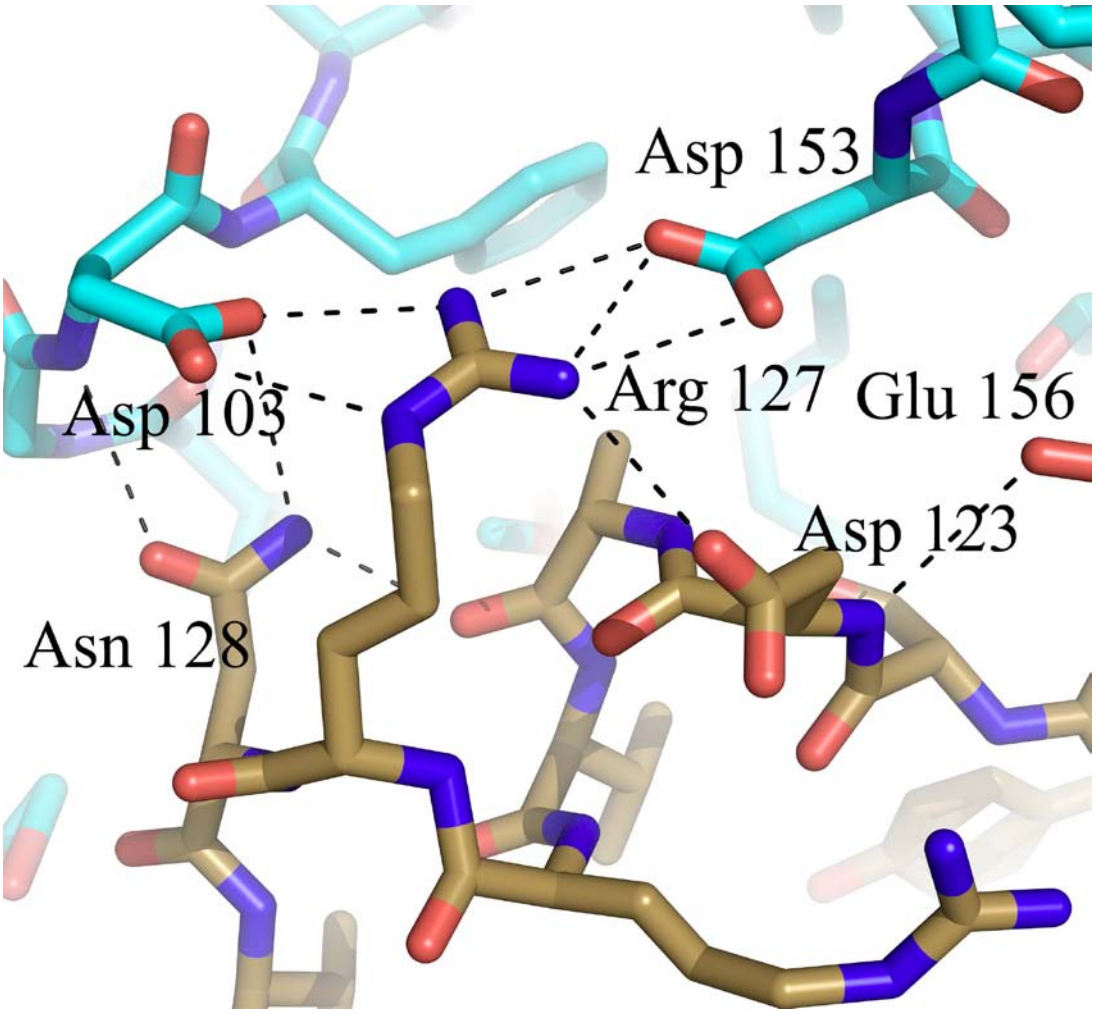

Figure 4.3. The interface between subunits of the trimer of FMO from Pld. phaeum. The interface between subunits is primarily stabilized by salt bridges such as those involving Asp 103, Arg 127, Asn 128, and Asp 153 for *Pld. phaeum* (carbons color coded sand or cyan depending upon subunit, oxygen red, nitrogen blue). Equivalent interactions are found involving Asp 107, Asp 127, Asn 128, Arg 131, and Asp 157 in *Cbl. tepidum* and Asp 107, Asp 127, Arg 131, Asn 132 and Arg 157 in *Ptc. aestuarii* 2K. Not all interactions are conserved, for example, the salt bridge between Asp 123 and Glu 156 is found only in *Pld. phaeum.*127, Arg 131, Asn 132, and Arg 157 in *Ptc. aestuarii* 2K. Not all interactions are conserved, for example, the salt bridge between Asp 123 and Glu 156 is found only in *Pld. phaeum.*

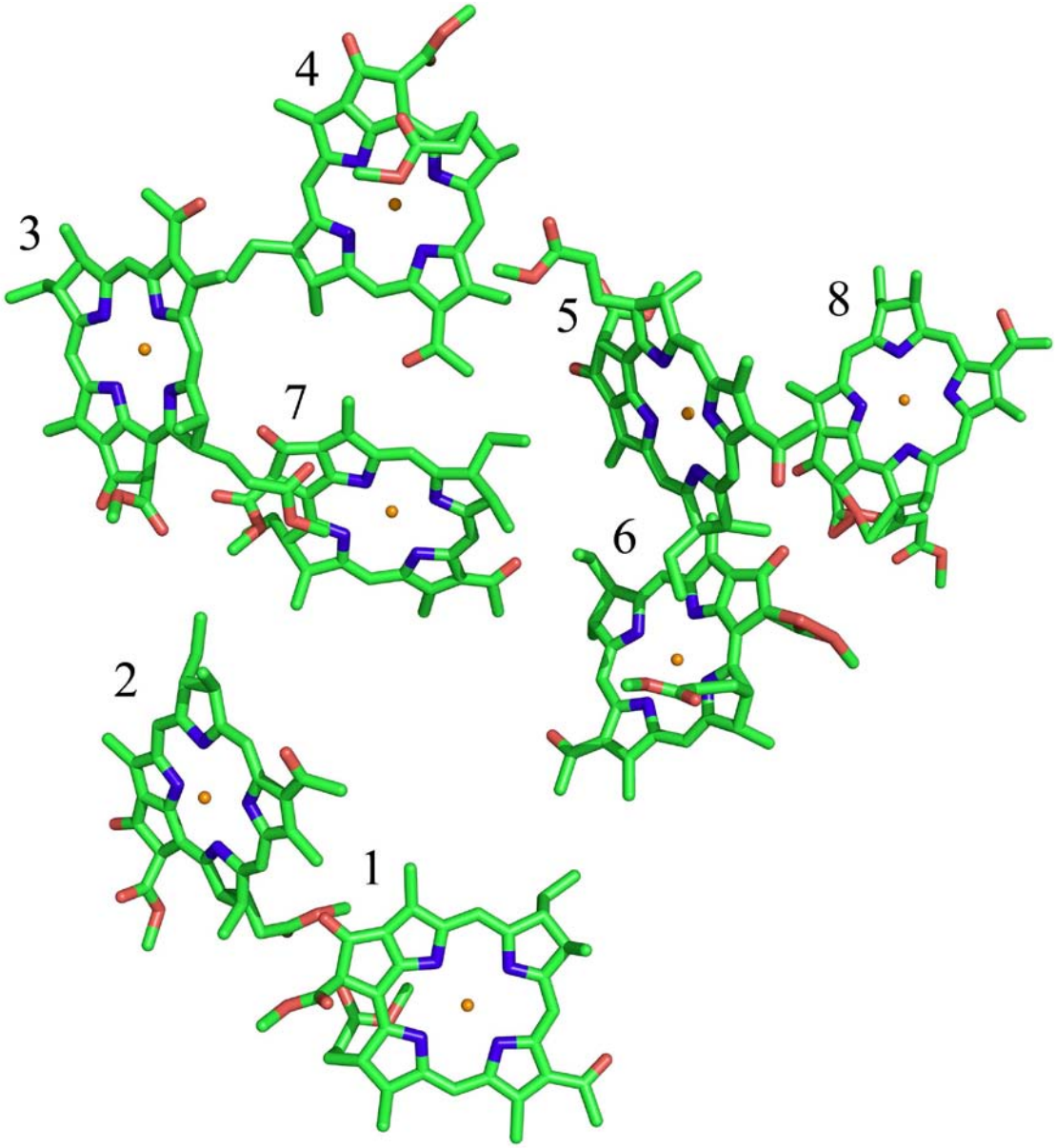

Figure 4.4. The three-dimensional structure of the BChl cofactors of each subunit of FMO from Pld. phaeum. The BChls (atom type) are arranged in an asymmetric fashion and shown without the phytyl chains for clarity. Each BChl is numbered as described in the text.

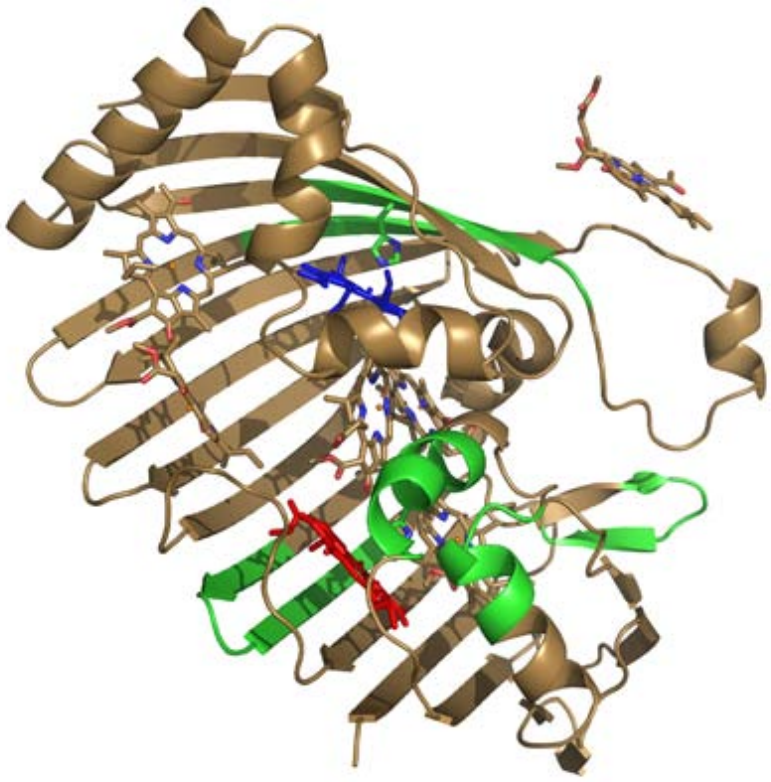

Figure 4.5. Strongly conserved regions for FMO proteins from different species. Shown is the structure of the FMO from Pld. phaeum with the regions of conserved amino acids for FMO proteins from green sulfur bacteria highlighted in green. These conserved regions are clustered around BChl 3 (red), which is thought to be the lowest energy pigment that transfers energy to the reaction center, and BChl 6 (blue), which is believed to be one of the highest energy
pigments that is in close proximity to the chlorosome

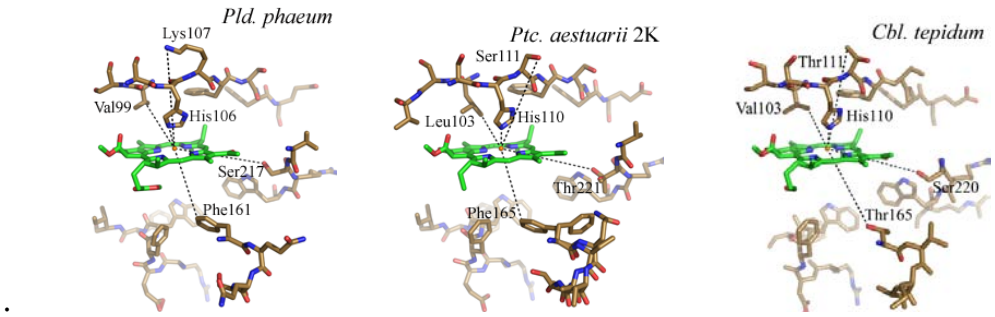

Figure 4.6. Comparison of the structure of BChl 1 and nearby amino acid residues. All three structures have a His as an axial ligand for the Mg of the BChl, namely 106 in *Pld. phaeum* and 110 in *Ptc. aestuarii* and *Cbl. tepidum* (color coded by atom type with green and sand for carbon in BChl 1 and the protein, respectively). A comparison of the differences shows that *Pld. phaeum* and *Ptc. aestuarii* have Phe 161 and Phe 165 respectively but *Cbl. tepidum* has Thr 165 in that location. Two conservative differences are Val 99 and Ser 217 in *Pld. phaeum* and *Cbl. tepidum* (Val 103 and Ser 220) but *Ptc. aestuarii* 2K has Leu 103 and Thr 221. In addition, the ionizable amino acid residue Lys 107 is not present in *Ptc. aestuarii* 2K or *Cbl. tepidum*, which have Ser 111 and Thr 111 respectively.

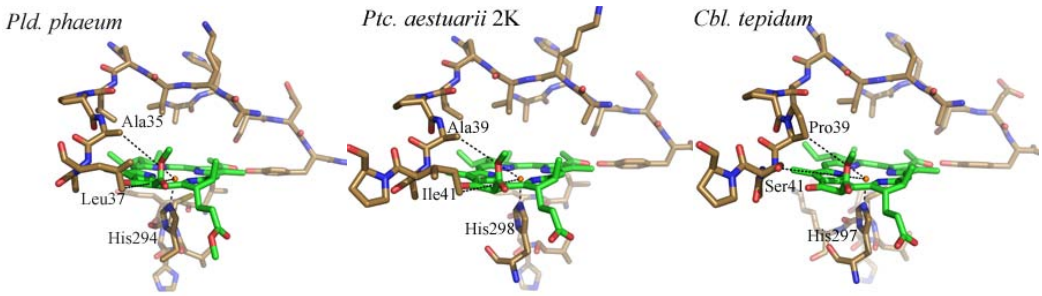

Figure 4.7. Comparison of the structure of BChl 3 and nearby amino acid residues. All three structures have a His as an axial ligand for the Mg of the BChl, namely 294, 298, and 297 in *Pld. phaeum, Ptc. aestuarii,* and *Cbl. tepidum* respectively (color coded by atom type with green and sand for carbon in BChl 3 and the protein, respectively). The binding site is very well conserved, with only two minor differences involving aliphatic residues, namely Ala in *Pld. phaeum* and *Ptc. aestuarii* at residues 35 and 39 respectively, is Pro 39 in *Cbl. tepidum* and Leu 37 in *Pld. phaeum* is equivalent to Ile 41 in *Ptc. aestuarii* and Ser 41 in *Cbl. tepidum*.

Figure 4.8. Comparison of the structure of BChl 5 and nearby amino acid residues. All three structures have a backbone carbonyl as an axial ligand for the Mg of the BChl, namely 238, 242, and 241 in *Pld. phaeum, Ptc. aestuarii*, and *Cbl. tepidum* respectively (color coded by atom type with green and sand for carbon in BChl 5 and the protein, respectively). The binding site for BChl 5 shows a number of differences among the three species. *Pld. phaeum and Ptc. aestuarii* 2K have a phenylalanine residue, Phe 62 and Phe 66 respectively, but *Cbl. tepidum* differs with Ile 66. A second difference is that *Pld. phaeum* and *Ptc. aestuarii* 2K both have Leu 269 and Leu 273 respectively, where *Cbl. tepidum* has Phe 272. Another difference is that *Pld. phaeum* and *Ptc. aestuarii* have a valine at residues 241 and 245 respectively but *Cbl. tepidum* contains Ser 244. Other minor differences include *Pld. phaeum* having two isoleucine residues, Ile 48 and Ile 250, that are also found in *Ptc. aestuarii* 2K, namely Ile 52 and Ile 254, with *Cbl. Tepidum* differing with Leu 52 and Val 253 at the equivalent locations

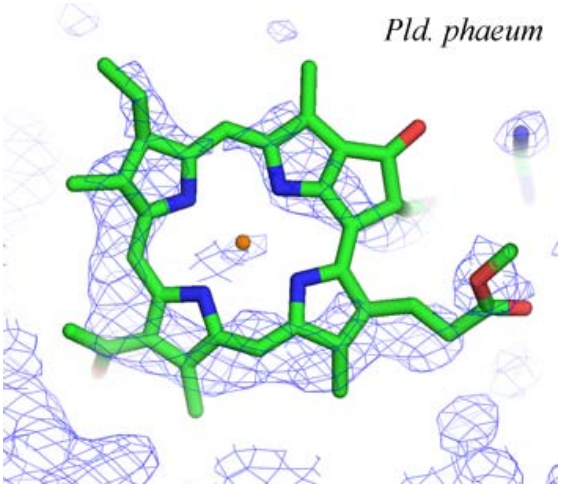

Figure 4.9. Electron density map for BChl 8 of FMO from Pld. phaeum. Shown is the refined position of BChl 8 (atom type) and the 2FoFc electron density map contoured at 1& (blue mesh).

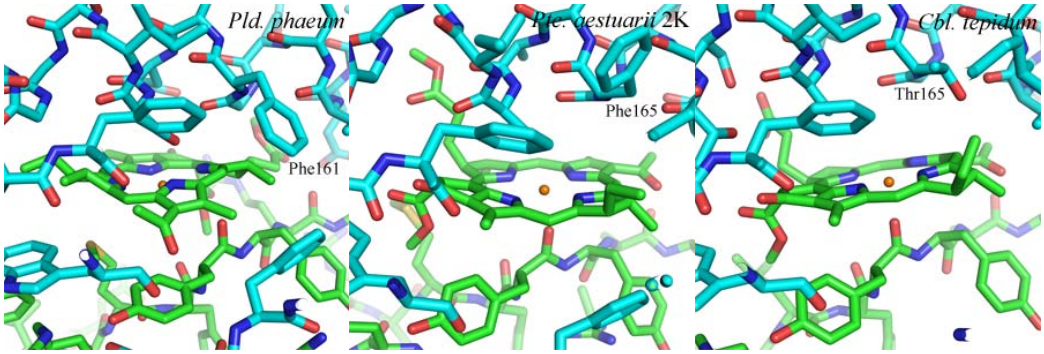

Figure 4.10. Comparison of the structure of BChl 8 and nearby amino acid residues. Unlike the other BChls, the position and orientation of BChl 8 is not conserved. The coordination of the Mg differs in each structure, with BChl 8 is five coordinated to Ser 164 from the adjacent subunit of the trimer in *Pld. phaeum,* coordinated to the backbone carbonyl of Tyr 123 in *Cbl. tepidum,*
and six coordinated to the backbone carbonyl of Tyr 123 and Ser 168 in *Ptc. aestuarii* 2K.

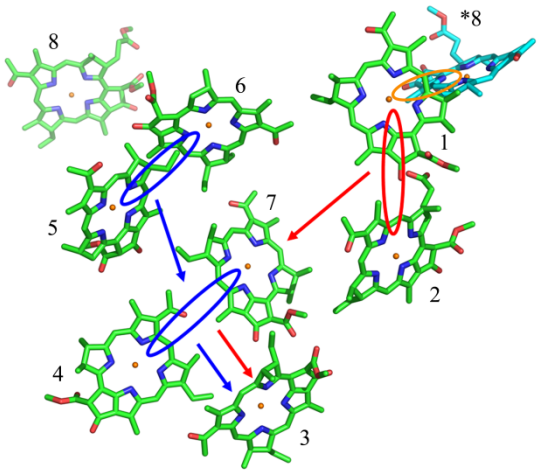

Figure 4.11. Cofactor arrangement in FMO from Pld. phaeum with exciton couplings. Shown are BChls 1–8 (colored by atom type with green carbon) as well as a BChl 8 from a neighboring subunit (BChl *8) of the trimer (colored by atom type with cyan carbon). The calculated couplings between BChls are illustrated by ellipses (red and blue) with the direction of energy flow shown by arrows (red and blue). The distance between BChl *8 and BChl 1 is comparable
to those found between other coupled BChls leading to the prediction of an additional coupling as illustrated by an ellipse between BChl *8 and BChl 1 (orange) that has not been included in any of the exciton models. Figure adapted from Brixner and coworkers (*12*).

References

1. Blankenship, R. E. (2002) Molecular Mechanisms of Photosynthesis, Wiley-Blackwell, Malden MA.

2. Staehelin, L. A., Golecki, J. R., and Drews, G. (1980) Supermolecular organization of chlorosomes (Chlorobium vesicles) and their membrane attachment sites in Chlorobium limicola. *Biochim. Biophys. Acta 589*, 30–45.

3. Psencik, J., Ikonen, T. P., Laurinmäki, P. Merckel, M. C., Butcher, S. J., Serimaa, R. E., and Tuma, R. (2004) Lamella organization of pigments in chlorosomes, the light harvesting complexes of green photosynthetic bacteria. *Biophys. J. 87*, 1165–1172.

4. Oostergetela, G. T., Reus, M., Chew, A.G. M., Bryant, D. A., Boekema, E. J., and Holzwarth, A. R. (2007) Long-range organization by bacteriochlorophyll in chlorosomes of Chlorobium tepidum investigated by cryo-electron microscopy. *FEBS Lett. 581*, 5435–5439.

5. Olson, J. M. (2004) The FMO protein. *Photosynth. Res. 80*, 181–187.

6. Wen, J., Zhang, H., Gross, M. L., and Blankenship, R. E. (2009) Membrane orientation of the FMO antenna protein from Chlorobaculum tepidum as determined by mass spectroscopy based footprinting. *Proc. Natl. Acad. Sci. U. S. A. 106*, 6134–6139.

7. Milder, M. T. W., Bruggemann, B., van Grondelle, R., and Herek, J. L. (2010) Revisiting the optical properties of the FMO protein. *Photsynth. Res. 104*, 257–274.

8. Adolphs, J., and Renger, T. (2006) How proteins trigger excitation energy transfer in the FMO complex of green sulfur bacteria. *Biophys. J. 91*, 2778–2797.

9. Engel, G. S., Calhoun, T. R., Read, E. L., Ahn, T. K., Man&al, T., Cheng, Y. C., Blankenship, R. E., and Fleming, G. R. (2007) Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems. *Nature 446*, 782–786.

10. Müh, F., Madjet, M. E. A., Adolphs, J., Abdurahman, A., Rabenstein, B., Ishikita, H., Knapp, E. W., and Renger, T. (2007) "-Helices direct excitation energy flow in the Fenna-Matthews-Olson protein. *Proc. Natl. Acad. Sci. U. S. A. 104*,16862–168673.

11. Mohseni, M., Rebentrost, P., Lloyd, S., an Aspuru-Guzik, A. (2008) Environment-assisted quantum walks in energy transfer of photosynthetic complexes. *J. Chem. Physics 129*, 174106.

12. Brixner, T., Stenger, J., Vaswani, H. M., Cho, M., Blankenship, R. E., and Fleming, G. R. (2005) Two-dimensional spectroscopy of electronic couplings in photosynthesis. *Nature 434*, 625–628.

13. Read, E. L., Engel, G. S., Calhoun, T. R., Man&al, T., Ahn, T. K., Blankenship, R. E., and Fleming, G. R. (2007) Cross-peak-specific two-dimensional electronic spectroscopy. *Proc. Natl. Acad. Sci. U. S. A 104*, 14203–14208.

14. Panitchayangkoon, G., Hayes, D., Fransted, K. A., Caram, J. R., Harel, E., Wen, J., Blankenship, R. E., and Engel, G. S. (2010) Long-lived quantum coherence in photosynthetic complexes at physiological temperature. *Proc. Natl. Acad. Sci. U. S. A. 107*, 12766–12770.

15. Olson, J. M. (1978) Bacteriochlorophyll a-proteins from green bacteria. *in The Photosynthetic Bacteria* (Clayton, R. K., and Sistrom, W. R. Eds) pp 161–178, Plenum Press, NY.

16. Matthews, B. W., Fenna, R. E., Bolognesi, M. C., and Schmid, M. F. (1979) Structure of a bacteriochlorophyll a-protein from the green photosynthetic bacterium Prosthecochloris aestuarii. *J. Mol. Biol. 131*, 259–285.

17. Tronrud, D. E., Schmid, M. F., and Matthews, B. W. (1986) Structure and X-ray amino acid sequence of a bacteriochlorophyll a protein from Prosthecochloris aestuarii at 1.9 Å resolution. *J. Mol. Biol. 188*, 443–454.

18. Li, Y. F., Zhou, W., Blankenship, R. E., and Allen, J. P. (1997) Crystal structure of the bacteriochlorophyll a protein from Chlorobium tepidum. *J. Mol. Biol. 271*, 456–471.

19. Camara-Artigas, A., Blankenship, R. E., and Allen, J. P. (2003) The structure of the FMO protein from Chlorobium tepidum at 2.2 Å resolution. *Photosynth. Res. 75*, 49–55.

20. BenShem, A., Frolow, F., and Nelson, N. (2004) Evolution of photosystem I-from symmetry through pseudosymmetry to asymmetry. *FEBS Lett. 564*, 274–280.

21. Tronrud, D. E., Wen, J., Gay, L., and Blankenship, R. E. (2009) The structural basis for the difference in absorbance spectra for the FMO antenna protein from various green sulfur bacteria. *Photosynth. Res. 100*, 79–87.

22. Tsukatani, Y., Wen, J., Blankenship, R. E., and Bryant, D. A. (2010) Characterization of the FMO proein from the aerobic chlorophototroph, Candidatus Chloracidobacterium thermophilum. *Photosynth. Res. 104*, 201–209.

23. Hu, D. (2001) Investigation of the Fenna-Matthews-Olson protein from photosynthetic green sulfur bacteria. Ph. D. thesis, Arizona State University, Tempe AZ.

24. Leslie, A. G. W. (1999) Integration of macromolecular diffraction data. *Acta Crystallogr. D 55*, 1696–1702.

25. Collaborative Computational Project, Number 4 (1994) The CCP4 suite-programs for protein crystallography. *Acta Crystallogr. D 50*, 760–763.

26. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Mariarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D58*, 1948–1954.

27. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta. Crystallogr. D 60*, 2126–2132.

28. Laskowski, R. A., MacArthur, M. W., Moss, D. S., Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst. 26*, 283–291.

29. Vaguine, A. A., Richelle, J., and Wodak, S. J. (1999) SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr. D55*, 191–205.

30. DeLano, W.L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, Palo Alto, CA, USA.

31. Melkozernov, A. N., Olson, J. M., Li, Y. F., Allen, J. P. and Blankenship, R. E. (1998) Orientation and Excitonic Interactions of the Fenna-Matthews-Olson Protein in Membranes of the Green

Sulfur Bacterium Chlorobium tepidum. *Photosyn. Res. 56*: 315–328.

32. Louwe, R. J. W., Vrieze, J., Aartsma, T. J., and Hoff, A. J. (1997) Toward an integral interpretation of the optical steady-state spectra of the FMO-complex of Prosthecochloris aestuarii. 1. An investigation with linear-dichroic absorbance-detected magnetic resonance. *J. Phys. Chem. 101*,11273–11279.

33. Wendling, M., Przyjalgowski, M. A., Gülen, D., Vulto, S. I. E., Aartsma, T. J., van Grondelle, R., and Amerongen, H. (2002) The quantitative relationship between structure and polarized spectroscopy in the FMO complex of Prosthecochloris aestuarii: refining experiments and simulations. *Photosynth Res 71*, 99–123.

34. Vulto, S. I. E., Neerken, S., Louwe, R. J. W., de Baat, M. A., Amesz, J., and Aartsma, T. J. (1998) Excited-state structure and dynamics in FMO antenna complexes from photosynthetic green sulfur bacteria. *J. Phys. Chem. 102*,10630–10635.

35. Johnson, S. and Small, G. (1991) Excited-state structure and energy-transfer dynamics of the bacteriochlorophyll a antenna complex from Prosthecochloris aestuarii. *J. Phys. Chem. 95*, 471–479.

36. Renger, T. (2010) Theory of excitation energy transfer: from structure to function. *Photosynth. Res. 102*, 471–485.

Chapter 5

Conclusion and Future work

Future Work on SMNWT

The SMNWT protein has proven to be an extremely difficult protein to crystallize and collect X-ray data on. Although the method for production and isolation of SMNWT and SMND7 is effective, many crystallization trials resulted in aggregated protein that was not stable in the crystallization buffer. A future strategy to combat the difficulty in the crystallization is to conduct a cocrystallization trial with the Gemin2 protein. It has been shown that the Gemin2 protein makes SMNWT more stable and this added stability may aid in obtaining a high quality crystal that will produce diffraction data. Another future goal is to obtain a structure that includes all the Gemin proteins while in the SMN complex. This may shed light on the mechanism of SMN in Spinal Muscular Atrophy.

There is no cure to date for the Spinal Muscular Atrophy. If a crystal structure can be obtained for both the SMNWT and the SMND7 forms, it would become possible to attempt drug targeting studies. If the SMND7 form can be targeted with a drug and have a gain of function effect, the disease symptoms could possibly be much less severe or eliminated.

Future Work on FMO

The Fenna Matthews Olson protein from three different organisms has been crystallized with the presence of the 8th bacteriochlorophyll-*a*

molecule and is now clearly assigned in the structure. This new structure will also allow detailed simulations with all eight of the BChls of an individual monomer and the 24 BChls of the entire trimer. When the spectra that was obtained from experimental methods and the simulation spectra are compared, it should show if the energy transfer within FMO includes all three monomers or if the energy transfer is confined to the individual monomers of the protein. The current simulations of FMO do not include the 8^{th} BChl or include transfer of energy from one monomer to another.

Complete list of References

1. www.fsma.org accessed 2010

2. Rochette, C. F., Gilbert, N., and Simard, L. R. (2001) *SMN* gene duplication and the emergenece of the *SMN2* gene occureed in distinct hominids: *SMN2* is unique to *Homo sapiens. Hum. Genet. 108*, 255-266.

3. Schrank, B., Gotz, R., Gunnersen, J. M., Ure, J. M., Toyka, K. V., Smith, A. G., and Sendtner, M. (1997) Inactivation of the survival motor neuron gene, a candidate gene for human spinal muscular atrophy, leads to massive cell death in early mouse embryos. *Proc. Natl. Acad. Sci. U. S. A. 94***,** 9920-9925.

4. Hsieh-Li, H. M., Chang, J. G., Jong, Y. J., Wu, N. M., Tsai, C. H., and Li, H. (2000) A mouse model for spinal muscular atrophy. *Nat. Genet. 24,* 66-70.

5. Brzustowicz, L. M., Lehner, T., Castilla, L. H., Penchaszadeh, G. K., Wilhelmsen, K. C., Daniels, R., Davies, K. E., Lepper, M., Ziter, F., Wood, D. , Dubowitz, V., Zerres, K., Hausmanowa-Petrusewicz, I., Ott, J., Munsat, T. L., & Gilliam, T. C. (1990) Genetic mapping of chronic childhood-onset spinal muscular atropy to chromosome 5q11.2-13.3. *Nature 344,* 540-541.

6. Melki, J., Abdelhak, S., Sheth, P., Bachelot, M. F., Burlet, P., Marcadet, A., Aicardi, J., Barois, A., Carrier, J. P., Fardeau, D., Ponsot, G., Billette, T., Angelini,C., Barbosa, C., Ferriere, G., Lanzil, G., Ottolini, A., Babron, M. C., Cohen, D., Hanauer, A., Clerget-Darpoux, F., Lathrop, M., Munnich, A., and Frezal, J. (1990) Gene for chronic proximal spinal muscular atrophies maps to chromosome 5q. *Nature 344,* 767-768.

7. Gilliam, T. C., Brzustowicz, L. M., Castilla, L. H., Lehner, T., Penchaszadeh, G. K., Daniels, R. J., Byth, B. C., Knowles, J., Hislop, J. E., Shapira, Y., Dubowitz, V., Munsat, T.L., Ott, J., and Davies, K. E. (1990) Genetic homogeneity between acute and chromic forms of spinal muscular atrophy. *Nature 345***,** 823-825.

8. Lefebvre, S., Buglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., Le Pasilier, D., Frezai, J., Cohen, D., Weissenbach, J., Munnich, A., and Melki, J. (1995) Identification and Characterization of a Spinal Muscular Atrophy-Determining Gene. *Cell 80,*155-165.

9. www.Uniprot.org accessed 2010

10. Lefebrvre, S., Burlet, P., Liu, Q., Bertrandy, S., Clermont, O., Munnich, A., Dreyfuss, G., and Melki, J. (1997) Correlation between severity and SMN protein level in spinal muscular atrophy. *Nat. Genet. 16,* 265-269.

11. Lorson, C. L., Hahnen. E., Androphy, E. J., and Wirth, B. (1999) A single nucleotide in the *SMN* gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl. Acad. Sci. U. S. A. 96,* 6307-6311.

12. Kashima, T., and Manley, J. L. (2003) A negative element in *SMN2* exon 7 inhibits splicing in spinal muscular atrophy. *Nature genet. 34*, 460-463.

13. Lorson, C. L., and Androphy, E. J. (2000) An exonic enhancer is required for inclusion of an essential exon in the SMN-determining gene *SMN. Hum. Mol. Genet. 9,* 259-265.

14. Alias, L., Bernal, S., Fuentes-Prior, P., Barceló, M. J., Also, E., Martínez-Hernández, R., Rodríguez-Alvarez, F. J. , Martín, Y., Aller, E., Grau, E., Peciña, A., Antiñolo, G., Galán, E., Roas, A. L., Fernández-Burriel, M., Borrego, S., Millán, J. M., Hernández-Chico, C., Baiget, M., and Tizzano, E. F. (2009) Mutation update of spinal muscular atrophy in Spain: molecular characterization of 745 unrelated patients and identification of four novel mutation in the *SMN1* gene. *Hum. Genet. 125*, 29-39.

15. Burghes, A. H. M., and Beattie, C. E. (2009) Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick? *Nat. Rev. Neurosci. 10*, 597-609.

16. Setola, V., Terao, M., Locatelli, D., Bassanini, S., Garattini, E., and Battaglia, G. (2007) Axonal-SMN (a-SMN), a protein isoform of the survival motor neuron gene, is specifically involved in axonogenesis. *Proc. Natl. Acad. Sci. U. S. A*. *140,* 1959-1964.

17. Liu, Q., and Dreyfuss, G. (1996) A Novel nuclear structure containing the survival of motor neurons protein. *EMBO J. 15,* 3555-3565.

18. Pellizzoni, L. (2007) Chaperoning ribonucleoprotein biogenesis in health and disease. *EMBO Rep.8*, 340-345.

19. Gubitz, A.K., Feng, W., and Dreyfuss, G. (2004) The SMN complex. *Exp. Cell Res. 296,* 51-56.

20. Pellizzoni, L., Yong, J., and Dreyfuss, G. (2002) Essential Role for the SMN Complex in the Specificity of snRNP Assembly. *Science 298*, 1775-1779.

21. Meister, G., Buhler, D., Laggerbauer, B., Zobawa, M., Lottspeich, F., and Fischer, U. (2000) Characterization of a nuclear 20S complex containing the survival of motor neurons (SMN) protein and a specific subset of spliceosomal Sm proteins. *Hum. Mol. Genet. 9,* 1977-1986.

22. Otter, S., Grimmler, M., Neuenkirchen, N., Chari, A., Sickmann, A., and Fischer, U. (2007) A Comprehensive Interaction Map of the Human Survival of Motor Neuron (SMN) Complex. *J. Biol. Chem. 282,* 5825-5833.

23. Carissimi, C., Saieva, L., Baccon, J., Chiarella, P., Maiolica, A., Sawyer, A., Rappsilber, J., and Pellizzoni, L. (2006) Gemin8 is a Novel Component of the Survival Motor Neuron complex and Functions in Small Nuclear Ribonucleoprotein Assembly. *J. Biol. Chem. 281,* 8126-8134.

24. Ogawa, C., Usui, K., Aoki, M., Ito, R., Itoh, M., Kai, C., Kanamori-Katayama, M., Hayashizaki, Y., and Suzuki, H. (2007) Gemin2 Plays an Important role in Stabilizing the Survival of Motor Neuron Complex. *J. Biol. Chem. 282,* 11122-11134.

25. Charroux, B., Pellizzoni, L., Perkinson, R. A., Yong, J., Shevchenko, A., Mann, M., and Dreyfuss, G. (2000) A Novel Component of the Smn Complex That Is Found in Both Gems and Nucleoli. *J. Cell Biol. 148*. 1177-1186.

26. Charroux, B., Pellizzoni, L., Perkinson, R. A., Shevchenko, A., Mann, M., and Dreyfuss, G. (1999) Gemin3: A Novel DEAD Box Protein that Interacts with SMN, the Spinal Muscular Atrophy Gene Product, and Is a Component of Gems. *J. Cell Biol. 147,* 1181-1193.

27. Battle, D. J., Lau, C-K., Wan, L., Deng, H., Lotti, F., and Dreyfuss, G. (2006) The Gemin5 Protein of the SMN Complex Identifies snRNAs. *Mol. Cell 23,* 273-279.

28. Ma, Y., Dostie, J., Dreyfuss, G., and Van Duyne, G. D. (2005) The Gemin6-Gemin7 Heterodimer from the Survival of Motor Neurons Complex Has an Sm Protein-like Structure. *Structure 13*, 883-892.

29. Carissimi, C., Saieva, L., Gabanella, F., and Pellizzoni, L. (2006) Gemin8 is required for the Architecture and Function of the Survival Motor Neuron Complex. *J. Biol. Chem. 281,* 37009-37016.

30. Raker, V. A., Plessel, G., and Lührmann, R. (1996) The snRNP core assembly pathway: identification of stable core protein heteromeric coplexes and an snRNP subcore particle *in vitro*. *EMBO J. 15*, 2256-2269.

31. Raker, V. A., Hartmuth, K., Kastner, B., and Lührmann, R. (1999) Spliceosomal U snRNP core Assembly: Sm Proteins Assemble onto an Sm Site RNA Nonanucleotide in a Specific and Thermodynamically Stable Manner. *Mol. Cell Biol.19*, 6554-6565.

32. Paushkin, S., Gubitz, A.K., Massenet, S., and Dreyfuss, G. (2002) The SMN complex, an assemblysome of ribonucleoproteins. *Curr. Opin. Cell Biol. 14,* 305-312.

33. Pellizzoni, L., Kataoka, N., Charroux, B., and Dreyfuss, G. (1998) A Novel Runction for SMN, the spinal Muscular Atophy Disease Gene Product, in Pre-mRNA splicing. *Cell 95,* 615-624.

34. Meister, G., Buhler, D., Pillai, R., Lottspeich, F., and Fischer, U. (2001) A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. *Nat. Cell Biol. 3,* 945-949.

35. Meister, G., and Fischer, U. (2002) Assisted RNP assembly: SMN and PRMT5 complexes cooperate in the formation of spliceosomal UsnRNPs. *EMBO J. 21,* 5853-5863.

36. Yong, J., Pellizzoni, L., and Dreyfuss, G. (2002) Sequence-specific interaction of U1 snRNA with the SMN complex. *EMBO J. 21*, 1188-1196.

37. Golembe, T. J., Yong, J., Battle, D. J., Feng, W., Wan, L., and Dreyfuss, G. (2005) Lymphotropic *Herpesvirus simiri* Uses the SMN Complex To Assemble Sm Cores on Its Small RNAs. *Mol. Cel. Biol. 25*, 602-611.

38. Golembe, T. J., Yong, J., and Dreyfuss, G. (2005) Specific Sequence Features, Recognized by the SMN Complex Identify snRNAs and Determine Their Fate as snRNPs. *Mol. Cel. Biol. 25,* 10989-11004.

39. Pellizzoni, L., Charroux, B., and Dreyfuss, G. (1999) SMN mutants of spinal muscular atrophy patients are defective in binding to snRNP proteins. *Proc. Natl. Acad. Sci. U. S. A. 96,* 11167-11172.

40. Bertrandy, S., Burlet, P., Clermont, O., Huber, C., Fondrat, C., Thierry-Mieg, D., Munnich, A., and Lefebvre, S. (1999) The RNA-binding

properties of SMN: deletion analysis of the zebrafish orthologue defines domains conserved in evolution. *Hum. Mol. Genet. 8,* 775-782.

41. Lorson, C.L., and Androphy, E.J. (1998) The domain encoded by exon 2 of the survival motor neuron protein mediates nucleic acid binding. *Hum. Mol. Genet. 7,* 1269-1275.

42. Young, P. J., Man, N. T., Lorson, C.L., Le, T. T., Androphy, E. J., Burghes, A. H. M., and Morris, G. E. (2000) The exon 2b region of the spinal muscular atrophy protein, SMN, is involved in self-association and SIP1 binding. *Hum. Mol. Genet. 9,* 2869-2877.

43. Lorson, C. L., Strasswimmer, J., Yao, J-M., Baleja, J.D., Hahnen, E., Wirth, B., Le, T., Burghes, A.H.M., and Androphy, E.J. (1998) SMN Oligomerization defect correlates with spinal muscular atrophy severity. *Nat. Genet. 19,* 63-66.

44. Liu, Q., Fischer, U., Wang, F., and Dreyfuss, G. (1997) The Spinal Muscular Atrophy Disease Gene Product, SMN, and Its Associated Protein SIP1 Are in a Complex with Spliceosomal snRNP Proteins. *Cell 90,* 1013-1021.

45. Selenko, P., Sprangers, R., Stier, G., Buhler, D., Fischer, U., and Sattler, M. (2001) SMN Tudor domain structure and its interaction with the Sm proteins. *Nat. Struct. Biol. 8,* 27-31.

46. Sprangers, R., Groves, M. R., Sinning, I., and Sattler, M. (2003) High-resolution X-ray and NMR Structures of the SMN Tudor Domain: Conformational Variation in the Binding Site for Symmetrically Dimethylated Arginine Residues. *J. Mol. Biol. 327,* 507-520.

47. Bühler, D., Raker, V., Luhrmann, R., and Fischer, U. (1999) Essential role for the tudor domain of SMN in spliceosomal U snRNP assembly: implications for spinal muscular atrophy. *Hum. Mol. Genet. 8,* 2351-2357.

48. Frugier, T., Tiziano, F. D., Cifuentes-Diaz, C., Miniou, P., Roblot, N., Dierich, A., Le Meur, M., and Melki J. (2000) Nuclear targeting defect of SMN lacking the C-terminus in a mouse model of spinal muscular atrophy. *Hum. Mol. Genet. 9,* 849-858.

49. Lorson, C. L., Rindt, H., and Shababi, M. (2010) Spinal muscular atrophy : mechanisms and therapeutic strategies. *Hum. Mol. Genet. 19*, R111-R118.

50. Eggert, C., Chari, A., Laggerbauer,B., and Fischer, U. (2006) Spinal muscular atrophy: the RNP connection. *TRENDS Mol. Med. 12,* 113-121.

51. Gabanell, F., Butchbach, M. E. R., Saieva, L., Carissimi, C., Burghes, A. H. M., and Pellizzoni, L. (2007) Ribonucleoprotein Assembly Defects Correlate with Spinal Muscular Atrophy Severity and Preferentially Affect a Subset of Spliceosomal snRNPs. *PLoS One 2,* e921.

52. Rossoll, W., Jablonka, S., Andreassi, C., Kroning A-K., Karle, K., Monani, U. R., and Sendtner, M. (2003) Smn, the spinal muscular atrophy-determining gene product, modulates axon growth and localization of β-actin mRNA in growth cones of motoneurons. *J. Cell Bio. 163*, 801-812.

53. McWhorter, M. L., Monani, U. R., Burghes, A. H. M., and Beattie, C. E. (2003) Knockdown of the survival motor neuron (Smn) protein in zebrafish causes defects in motor axon outgrowth and pathfinding. *J Cell Biol 162,* 919-931.

54. Carrel, T. L., McWhorter, M. L., Workman, E., Zhang, H., Wolstencroft, E. C., Lorson, C., Bassell, G. J., Burghes, A. H. M., and Beattie, C. E. (2006) Survival Motor Neuron Function in Motor Axons is Independent of Functions Required for Small Nuclear Ribonucleoprotin Biogenesis. *J. Neurosci. 26,* 11014-11022.

55. Fan, L., and Simard, L. R. (2002) Survival motor neuron (SMN) protein: role in neurite outgrowth and neuromuscular maturation during neuronal differentiation and development. *Hum. Mol. Genet. 11,* 1605-1614.

56. Blankenship, R. B. (2002) Molecular Mechanisms of Photosynthesis *London: Blackwell Science*

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

58. Matthews, B. W., Fenna, R. E., Bolognesi, M. C., and Schmid, M. F. (1979) Structure of a bacteriochlorophll *a*-protein from the green photosynthetic bacterium *Prosthecochloris aestuarii. J. Mol. Biol. 131*, 259-285.

59. Tsukatani, Y., Wen, J., Blankenship, R. E., and Bryant, D. A. (2010) Characterization of the FMO protein from the aerobic chlorophotroph, *Candidatus* Chloracidobacterium thermophilum. *Photosynth Res. 104,* 201-209.

60. Ganapathy, S., Oostergetel, G. T., Wawrzyniak, P. K., Reus, M., Chew, A. G. M., Buda, F., Boekema, D. A., Bryant, D. A., Holzwarth, A. R., and de Groot, H. J. M. (2009) Alternating *syn-anti* bacteriochlorophylls form concentric helical nanotubes in chlorosomes. *Proc. Natl. Acad. Sci. U. S. A. 106*, 8525-8530.

61. Frigaard, N-U., Chew, A. G. M., Julia, H. L., Maresca, J. A., and Bryant, D. A. (2003) *Chlorobium tepidum*: insights into the structure, physiology, and metabolism of a green sulfur bacterium derived from the complete genome sequence. *Photosyn. Res. 78*, 93-117.

62. Li, Y. F., Zhou, W., Blankenship, R. E., and Allen, J. P. (1997) Crystal structure of the bacteriochlorophyll *a* protein from *Chlorobium tepidum*. *J. Mol. Biol. 271*, 456-471.

63. Wen, J., Zhang, H., Gross, M. L., and Blankenship, R. E. (2009) Membrane orientation of the FMO antenna protein from *Chlorobaculum tepidum* as determined by mass spectroscopy based footprinting. *Proc. Natl. Acad. Sci. U. S. A. 106,* 6134—6139.

64. Brixner, T., Stenger, J., Vaswani, H. M., Cho, M., Blankenship, R. E., and Fleming, G. R. (2005) Two-dimensional spectroscopy of electronic couplings in photosynthesis. *Nature 434*, 625-628.

65. Johnson, S. and Small, G. (1991) Excited-state structure and energy-transfer dynamics of the bacteriochlorophyll a antenna complex from *Prosthecochloris aestuarii*. *J. Phys. Chem. 95*, 471—479.

66. Louwe, R. J. W., Vrieze, J., Aartsma, T. J., and Hoff, A. J. (1997) Toward an integral interpretation of the optical steady-state spectra of the FMO-complex of *Prosthecochloris aestuarii.* 1. An investigation with linear-dichroic absorbance-detected magnetic resonance. *J. Phys. Chem. 101,* 11273—11279.

67. Förster, T. (1960) Transfer Mechanisms of Electronic Excitation Energy. *Rad. Res.Supplement. 2*, 326-339.

68. Selvin, P. R. (1995) Fluorescence Resonance Energy Transfer. *Methods Enzymol. 246.* 300-344.

69. Jong, S., Newton, M. D., and Silbey, R. J. (2004) Multichromophoric Förster Resonance Energy Transfer. *Phys. Rev. Lett. 92,* 218301.

70. van Amerongen, H. , van Grondelle, R. , and Valkunas, L. (2000) Photosynthetic Excitons. *World Scientific, London.*
71. Joachimiak, A. (2009) High-throughput crystallography for structural genomics. *Curr. Opin. Struct. Biol. 19*. 573-584.

72. Hochuli, E., Döbeli, H., and Schacher, A. (1987) New Metal Chelate Adsorbent Selective for Proteins and Peptides Containing Neighbouring Histidine Residues. *J. Chromotography 411*, 177-184.

73. Smith, D. B., and Johnson, K. S. (1988) Single-step purification of polypeptides expressed in *Escherichia* coli as fusions with glutathione S-transferase. *Gene 67*, 31-40.

74. Stofko-Hahn, R. E., Carr, D. W., and Scott, J. D. (1992) A single step purification for recombinant proteins Characterization of a microtubule associcated protein (MAP 2) fragment which associates with the type II cAMP-dependent protein kinase. *FEBS Lett. 302*, 274-278.

75. Brizzard, B. L., Chubet, R. G., and Vizard, D. L. (1994) Immunoaffinity purification of FLAG epitope-tagged bacterial alkaline phosphatase using a novel monoclonal antibody and peptide elution. *Biotechniques 16*, 730-735.

76. Carrio, M. M., and Villaverde, A. (2001) Protein aggregation as bacterial inclusion bodies is reversible. *FEBS Lett. 489*, 29-33.

77. Carrio, M. M., and Villaverde, A. (2002) Construction and deconstruction of bacterial inclusion bodies. *J. Biotechnol. 96*, 3-12.

78. Clark, E. D. B., (1998) Refolding of recombinant proteins. *Curr. Opin. Biotechnol. 9*, 157-163.

79. Fahnert, B., Lile, H., and Neubauer, P. (2004) Inclusion Bodies: Formation and Utilisation. *Adv. Biochem. Eng. Biotechnol. 89*, 93-142.

80. Kane, J. F., and Hartley, D. L. (1988) Formation of recombinant protein inclusion bodies in *Escherichia coli. Trends Biotechnol*. *6*, 95-101.

81. Prouty, W. F., Karnovsky, M. J., and Goldberg, A. L. (1975) Degradation of abnormal proteins in *Escherichia coli.* Formation of protein inclusions in cells exposed to amino acid analogs. *J. Biol. Chem. 250*, 1112-1122.

82. Baneyx, F., and Mujacic, M. (2004) Recombinant protein folding and misfolding in *Escherichia coli. Nature Biotechnol. 22*,1399-1408.

83. Martínez-Alonso, M., González-Montalbán, N., García-Fruitós, E., and Villaverde, A. (2009) Learning about protein solubility from bacterial inclusion bodies. *Microbial Cell Factories 8*, 4.

84. Singh, S. M., and Panda, A. K. (2005) Solubilization and refolding of bacterial inclusion body proteins. *J. Biosci. Bioeng*. *99*, 303-310.

85. Speed, M. A., Wang, D. I. C., and King, J. (1996) Specific aggregation of partially folded polypeptide chains: The molecular basis of inclusion body composition. *Nature. Biotechnol. 14*, 1283-1287.

86. Fink, A. L. (1998) Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Folding & Design 3*, R9-R23.

87. Luheshi, L. M., Crowther, D. C., and Dobson, C. M. (2008) Proteins misfolding and disease: from the test tube to the organism. *Curr. Opin. Chem. Biol. 12*, 25-31.

88. King, J., Haase-Pettingell, C., Robinson, A. S., Speed, M., and Mitraki, A. (1996) Thermolabile folding intermediates: inclusion body precursors and chaperonin substrates. *FASEB J.10*, 57-66.

89. Oberg, K., Chrunyk, B. A., Wetzel, R., and Fink, A. L. (1994) Native-like Secondary Structure in Inerleukin-1.beta. Inclusion Bodies Attenuated Total Reflectance FTIR. *Biochemistry 33*, 2628-2634.

90. Teschke, C. M., and King, J. (1993) Folding of the phage P22 coat protein in vitro. *J. Biol. Chem. 32*, 10839-10847.

91. Georgiou, G., Valax, P., Ostermeier, M., and Horowitz, P. M. (1994) Folding and aggregation of TEM β-lactamase: Analogies with the formation of inclusion bodies in *Escherichia coli*. *Protein Sci. 3*, 1953-1960.

92. Georgiou, G., and Valax, P. (1999) Isolating inclusion bodies from bacteria. *Methods Enzymol. 309*, 48-58.

93. Maachupalli-Reddy, J., Kelley, B. D., and Clark, E.D.B. (1997) Effect of Inclusion Body Contaminants on the Oxidation Renaturation of Egg White Lysozyme. *Biotechnol. Prog. 13*, 144-150.

94. Babbitt, P. C., West, B. L., Buechter, D. D., Kuntz, I. D. and Kenyon, G. L. (1990) Removal of a Proteolytic Activity Associated with Aggregates Formed from Expression of Creatine Kinase in *Escherichia coli* Leads to Improved Recovery of Active Enzyme. *Nat. Biotechnol.* 8, 945-945.

95. Betton, J.-M., Sasson, N., Hofnung, M., and Laurent, M. (1998) Degradatikon *versus* Aggregation of Misfolded Maltose-binding Protein in the Periplasm of *Escherichia coli*. *J. Biol. Chem. 273*, 8897-8902.

96. Cubarsí, R., Carrió, M. M., and Villaverde, A. (2001) *In Situ* Proteolytic Digestion of Inclusion Body Polypeptides Occurs as a Cascade Process. *Biochem. Biophys. Res. Commun. 282*, 436-441.

97. Corchero, J. L., Viaplana, E., Benito, A., and Villaverde, A. (1996) The position of the heterologous domain can influence the solubility and proteolysis of β-galactosidase fusion proteins in *E. coli*. *J. Biotechnol.* **48**, 191-200.

98. Carbonell, X., and Villaverde, A. (2002) Protein aggregated into bacterial inclusion bodies does not result in protection from proteolytic digestion. *Biotechnol. Lett. 24*, 1939-1944.

99. Middelberg, A.P.J. (2002) Preparative protein refolding. *Trends Biotechnol. 20*. 437-443.

100. Tsumoto, K., Ejima, D., Kumagai, I., and Arakawa, T. (2003) Practical considerations in refolding proteins from inclusion bodies. *Protein Expression Purif. 28*. 1-8.

101. Villaverde, A., and Carrió, M. M. (2003) Protein aggregation in recombinant bacteria: biological role of inclusion bodies. *Biotechnol. Lett. 25*. 1385-1395.

102. Fischer, B., Sumner, I., and Goodenough, P. (1993) Isolation, renaturation, and formation of disulfide bonds of eukryotic proteins expressed in *Escherichia coli* as inclusion bodies. *41.* 3-13.

103. Valax, P., and Georgiou, G. (1993) Molecular Characterization of β-Lactamase Inclusion Bodies Produced in *Escherichia coli*. 1. Composition. *Biotechnol. Prog*. *9*, 539-547.

104. Khan, R. H., Rao, K. B. C., Eshwari, A. N. S., Totey, S. M., and Panda, A. K. (1998) Solubilization of Recombinant Ovine Growth Hormone with Retention of Native-like Secondary Structure and Its Refolding from the the Inclusion Bodies of *Escherichia coli. Biotechnol. Prog. 14*, 722-728.

105. Levinthal, C. (1968) Are the pathways to protein folding. *J. Chim. Phys. 65.* 44.

106. Daggett, V., and Fersht, A. R. (2003) Is there a unifying mechanism for protein folding? *Trends Biochem. Sci. 28.* 18-25.

107. Sela, M., White, Jr. F. H., and Anfinsen, C. B. (1957) Reductive Cleaveage of Disulfide Bridges in Ribonuclease. *Science 125.* 691-92.

108. Dill, K. A., and Chan, H. S. (1997) From Levinthal to pathways to funnels. *Nat. Struct. Biol. 4.* 10-19.

109. Brockwell, D. J., and Radford, S. E. (2007) Intermediates: ubiquitous species on folding energy landscapes? *Curr. Opin. Struct. Biol. 17*. 30-37.

110. Radford, S. E., and Dobson, C. M. (1999) From computer simulations to human disease: emerging themes in protein folding. *Cell 97.* 291-298.

111. Dobson, C. M., Šali, A., and Karplus, M. (1998) Protein Folding: A Perspecitive from Theory and Experiment. *Angewandte Chemie 37.* 868-893.

112. Kramer, G., Ramachandiran, V., and Hardesty, B. (2001) Cotranslational folding – omnia mea mecum porto? *Int. J. Biochem. Cell Biol. 33.* 541-553.

113. Hartl, F. U., and Hayer-Hartl, M. (2009) Converging concepts of protein folding *in vitro* and *in vivo. Nat. Struct. Mol. Biol. 16.* 574-581.

114. Ignatova, Z., and Gierasch, L. M. (2004) Monitoring protein stability and aggregation *in vivo* by real-time fluorescent labeling. *Proc. Natl. Acad. Sci. U. S. A. 101.* 523-528.

115. Ellis, J. (1987) Proteins as molecular chaperones. *Nature 328.* 378-379.

116. Valejo, L. F., and Rinas, U. (2004) Optimized procedure for renaturation of recombinant human bone morphogenetic protein-2 at high protein concentration. *Biotechnol. Bioeng*. *85*, 601-609.

117. Dobson, C. M. (2003) Protein folding and misfolding. *Nature 426*, 884-890.

118. Ugwu, S. O., and Apte, S. P. (2004) The effect of Buffers on Protein Conformational Stability. *Pharm. Technol. March*, 86-113.

119. McPherson, A. (2004) Introduction to protein crystallization. *Methods 34.* 254-265.

120. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature 181*, 662-666.

121. Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, J. D., Westbrook, J. D., and Zardecki, C. (2002) The Protein Data Bank. *Acta. Cryst. D. 58*. 899-907.

122. Bolanos-Garcia, V. M., and Chayen, N. E. (2009) New directions in conventional methods of protein crystallization. *Prog. Biophy. Mol. Biol. 101.* 3-12.

123. Caylor, C. L., Dobrianov, I., Lemay, S. G., Kimmer, C., Kriminski, S., Finkelstein, K. D., Zipfel, W., Webb, W. W., Thomas, B. R., Chernov, A. A. and Thorne, R. E. (1999) Macromolecular impurities and disorder in protein crystals. *Proteins: Struct., Funct., Bioinf. 36*. 270-281.

124. Chernov, A. A. (1999) Estimates of internal stress and related mosaicity in solution grown crystals: proteins. *J Cryst. Growth 196*. 524-534.

125. Bergfors, T. (2003) Seeds to crystals. *J. Struct. Biol.142.* 66-76.

126. Saridakis, E., and Chayen, N. E. (2009) Towards a ‘universal’ nucleant for protein crystallization. *Trends Biotechnol. 27.* 99-106.

127. Porath, J., Carlsson, J., Olsson, I., and Belfrage, G. (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature 258*, 598-599.

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

129. Allen, J. P., Feher, G., Yeates, T. O., Komiya, H., and Rees, D. C. (1987) Structure of the reaction center from *Rhodobacter* sphaeroides R-26: the protein subunits. Proc*. Natl. Acad. Sci. USA 84*, 5730-5734.

130. Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C., and Shore, V. C. (1960) Structure of Myoglobin: A Three-Dimensional Fourier synthesis at 2Å. Resolution. *Nature 185*, 422-427.

131. Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G., and North, A. C. T. (1960) Structure of Haemoglobin: A Three-Dimensional Fourier Synthesis at 5.5Å. Resolution, Obtained by X-Ray Analysis. *Nature 185*, 416-422.

132. Zouni, A., Witt, H. T., Kern, J., Fromme, P., Krauss, N., Saenger, W., and Orth, P. (2001) Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8Å resolution. *Nature 49*, 739-743.

133. Kamiya, N., and Shen, J. R. (2003) Crystal structure of oxygen-evolving photosystem II from *Thermosynechococcus vulcanus* at 3.7Å resolution. *Proc. Natl. Acad. Sci. USA 100*, 98-103.

134. Ferreira, K. N., Iverson, T. M., Maghlaoui, K., Barber, J., and Iwata, S. (2004) Architecture of the photosynthetic oxygen-evolving center. *Science 303*, 1831-1838.

135. Loll, B., Kern, J., Saenger, W., Zouni, A., and Biesiadka, J. (2005) Towards complete cofactor arrangement in the 3.0Å resolution structure of photosystem II. *Nature 438*, 1040-1044.

136. Stowell, M. H. B., McPhillips, T. M., Rees, D. C., Soltis, S. M., Abresch, E., and Feher, G. (1997) Light-induced structural changes in photosynthetic reaction center: implications for mechanism of electron-proton transfer. *Science 276*, 812-816.

137. Baxter, R. H., Seagle, B. L., Ponomarenko, N., and Norris, J. R. (2005) Cryogenic structure of the photosynthetic reaction center of *Blastochloris* viridis in the light and dark. Acta *Crstallogr D 61*, 605-612.

138. Feher, G., Allen, J. P., Okamura, M. Y., and Rees, D. C. (1989) Structure and function of bacterial photosynthetic reaction centres. *Nature 339*, 111-116.

139. Hunter, N., Daldal, F., Thurnauer, M., and Beatty, J. T. (eds) (2008) *Springer-Verlag Publishers* Dordrecht, the Netherlands

140. Deisenhofer, J., Epp, O., Sinning, I., and Michel, H. (1995) Crystallographic refinement at 2.3Å Resolution and Refined Model of the Photosynthetic Reaction Centre from *Rhodopseudomonas viridis. J. Mol. Biol. 246*, 429-457.

141. Chang, C. H., El-Kabbani, O., Tiede, D., Norris, J., and Schiffer, M. (1991) Structure of the membrane-bound protein photosynthetic reaction center from *Rhodobacter sphaeroides. Biochemistry 30*, 5352-5360.

142. Ermler, U., Fritzsch, G., Buchanan, S. K., and Michel, H. (1994) Structure of the photosynthetic reaction centre from *Rhodobacter sphaeroides* at 2.65Å resolution: cofactors and protein-cofactor interactions. *Structure 2*, 925-936.

143. McAuley, K. E., Fyfe, P. K., Ridge, J. P., Isaacs, N. W., Cogdell, R. J., and Jones, M. R. (1999) Structural details of an interaction between cardiolipin and an integral membrane protein. *Proc. Natl. Acad. Sci. USA* **96**, 14706-14711.

144. Camara-Artigas, A., Brune, D., and Allen, J. P. (2002) Interactions between lipids and bacterial reaction centers determined by protein crystallography. *Proc. Natl. Acad. Sci. USA. 99*, 11055-11060.

145. Debus, R. J., Feher, G., and Okamura, M. Y. (1985) LM complex of reaction centers from *Rhodopseudomonas sphaeroides* R-26: characterization and reconstitution with the H subunit. *Biochemistry 24*, 2488-2500.

146. Cheng, Y. S., Brantner, C. A., Tsapin, A., and Collins, M. L. P. (2000) Role of the H protein in assembly of the photochemical reaction center and intracytoplasmic membrane in *Rhodospirillum rubrum*. *J. Bacteriol. 182*, 1200-1207.

147. Tehrani, A., Prince, R. C., and Beatty, J. T. (2003) Effects of Photosynthetic Reaction Center H Protein Domain Mutations on Photosythetic Properties and Reaction Center Assembly in *Rhodobacter sphaeroides. Biochemistry 42*, 8919-8928.

148. Lupo, D., and Ghosh, R. (2004) The reaction center H subunit is not required for high levels of light-harvesting complex 1 in *Rhodospirillum rubrum* mutants. *J. Bacteriol 186*, 5585-5595.

149. Axelrod, H. L., and Okamura, M. Y. (2005) The structure and function of the cytochrome c 2: reaction center electron transfer complex from *Rhodobacter sphaeroides. Photosynth. Res 85*, 101-114.

150. Axelrod, H. L., Abresch, E. C., Okamura, M. Y., Yeh, A. P., Rees, D. C., and Feher, G. (2002) X-ray structure determination of the cytochrome c2: reaction electron transfer complex from *Rhodobacter sphaeroides*. *J. Mol. Biol. 319,* 501-515.

151. Miyashita, O., Okamura, M. Y., and Onuchic, J. M. (2005) Interprotein electron transfer from cytochrome c2 to photosynthetic reaction center: Tunneling across an aqueous interface. *Proc. Natl. Acad. Sci. USA 102*, 3558-3563.

152. Wydrzynski, T. J., and Satoh, K. (eds) (2005) Photosystem II: The Light-Driven Water: Plastoquinone Oxidoreductase. *Springer-Verlag Publishers*, Dordrecht, the Netherlands

153. Ben-Shem, A., Frolow, F., and Nelson, N. (2003) Crystal structure of plant photosystem I. *Nature 426*, 630-635.

154. Ben-Shem, A., Frolow, F., and Nelson, N. (2004) Evolution of photosystem I-from symmetry through pseudosymmetry to asymmetry. *FEBS Lett 564*, 274-280.

155. Raszewski, G., Diner, B. A , Schlodder, E., and Renger, T. (2008) Spectroscopic properties of reaction center pigments in photosystem II core complexes: revision of the multimer model. *Biophys. J. 95,* 105-119.

156. Yano, J., Kern, J., Sauer, K., Latimer, M. J., Pushkar, Y., Biesiadka, J., Loll, B., Saenger, W., Messinger, J., Zouni, A., and Yachandra, V. K. (2006) Where water is oxidized to dioxygen: structure of the photosynthetic Mn4Ca cluster. *Science 314*, 821-825.

157. Goldbeck, J. (ed) (2006) *Springer-Verlag Publishers* Dordrecht, the Netherlands

158. Krauss, N., Hinrics, W., Witt, I., Fromme, P., Pritzkow, W., Dauter, Z., Betzel, C., Wilson, K. S., Witt, H. T., and Saenger, W. (1993) Three-

dimensional structure of system I of photosynthesis at 6Å resolution. *Nature 361*, 326-331.

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

160. Krauss, N. (2008) In: Fromme, P. (ed) *Wiley-Blackwell*, 23-64

161. Amunts, A., Drory, O., and Nelson, N. (2007) The structure of a plant photosystem I supercomplex at 3.4Å resolution. *Nature 447*, 58-63.

162. Cogdell, R. J., Gall, A., and Köhler, J. (2006) The architecture and function of the light-harvesting apparatus of purple bacteria: from single molecules to *in vivo* membranes. *Q. Rev. Biophys. 39*, 227-324.

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

164. Papiz, M. Z., Prince, S. M., Howard, T., Cogdell, R. J., and Isaacs, N. W. (2003) The Structure and thermal Motion of the B800-850 LH2 Complex from *Rps. Aacidophila* at 2.0Å Resolution and 100K: New Structural Features and Functionally Relevant Motions. *J. Mol. Biol. 326*, 1523-1538.

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

166. Karrasch, S., Bullog, P. A., and Ghosh, R. (1995) The 8.5Å projection map of the light-harvesting compolex I from *Rhodospirillum rubrum* reveals a ring composed of 16 subunits. *EMBO J.14,* 631-638.

167. Roszak, A. W., Howard, T. D., Southall, J., Gardiner, A. T., Law, C. J., Isaacs, N. W., and Cogdell, R. J. (2003) Crystal structure of the RC-LH1 core complex from *Rhodopseudomonas palustris Science 302*, 1969-1972.

168. Tronrud, D. E., and Matthews, B. W. (1993) In: Norris, J. , Deisenhofer, J. (eds) *Academic Press,* New York 13-21

169. Camara-Artigas, A., Blankenship, R. E., and Allen, J. P. (2003) The structure of the FMO protein from *Chlorobium* tepidum at 2.2Å resolution. *Photosynth. Res. 75*, 49-55.

170. Read, E. L., Schlau-Cohen, G. S., Engel, G. S., Wen, J., Blankenship, R. E., and Fleming, G. R. (2008) Visualization of excitonic structure in the Fenna-Matthews-Olson photosynthetic complex by polarization-dependent two-dimensional electronic spectroscopy. *Biophys. J. 95*, 847-856.

171. Blankenship, R. E. (2002) Molecular Mechanisms of Photosynthesis, Wiley-Blackwell, Malden MA.

172. Staehelin, L. A., Golecki, J. R., and Drews, G. (1980) Supermolecular organization of chlorosomes (Chlorobium vesicles) and their membrane attachment sites in Chlorobium limicola. *Biochim. Biophys. Acta 589*, 30–45.

173. Psencik, J., Ikonen, T. P., Laurinmäki, P. Merckel, M. C., Butcher, S. J., Serimaa, R. E., and Tuma, R. (2004) Lamella organization of pigments in chlorosomes, the light harvesting complexes of green photosynthetic bacteria. *Biophys. J. 87*, 1165–1172.

174. Oostergetela, G. T., Reus, M., Chew, A.G. M., Bryant, D. A., Boekema, E. J., and Holzwarth, A. R. (2007) Long-range organization by bacteriochlorophyll in chlorosomes of Chlorobium tepidum investigated by cryo-electron microscopy. *FEBS Lett. 581*, 5435–5439.

175. Olson, J. M. (2004) The FMO protein. *Photosynth. Res. 80*, 181–187.

176. Milder, M. T. W., Bruggemann, B., van Grondelle, R., and Herek, J. L. (2010) Revisiting the optical properties of the FMO protein. *Photsynth. Res. 104*, 257–274.

177. Adolphs, J., and Renger, T. (2006) How proteins trigger excitation energy transfer in the FMO complex of green sulfur bacteria. *Biophys. J. 91*, 2778–2797.

178. Engel, G. S., Calhoun, T. R., Read, E. L., Ahn, T. K., Man&al, T., Cheng, Y. C., Blankenship, R. E., and Fleming, G. R. (2007) Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems. *Nature 446*, 782–786.

179. Müh, F., Madjet, M. E. A., Adolphs, J., Abdurahman, A., Rabenstein, B., Ishikita, H., Knapp, E. W., and Renger, T. (2007) "-Helices direct excitation energy flow in the Fenna-Matthews-Olson protein. *Proc. Natl. Acad. Sci. U. S. A. 104*,16862–168673.

180. Mohseni, M., Rebentrost, P., Lloyd, S., an Aspuru-Guzik, A. (2008) Environment-assisted quantum walks in energy transfer of photosynthetic complexes. *J. Chem. Physics 129*, 174106.

181. Read, E. L., Engel, G. S., Calhoun, T. R., Man&al, T., Ahn, T. K., Blankenship, R. E., and Fleming, G. R. (2007) Cross-peak-specific two-dimensional electronic spectroscopy. *Proc. Natl. Acad. Sci. U. S. A 104*, 14203–14208.

182. Panitchayangkoon, G., Hayes, D., Fransted, K. A., Caram, J. R., Harel, E., Wen, J., Blankenship, R. E., and Engel, G. S. (2010) Long-lived quantum coherence in photosynthetic complexes at physiological temperature. *Proc. Natl. Acad. Sci. U. S. A. 107*, 12766–12770.

183. Olson, J. M. (1978) Bacteriochlorophyll a-proteins from green bacteria. *in The Photosynthetic Bacteria* (Clayton, R. K., and Sistrom, W. R. Eds) pp 161–178, Plenum Press, NY.

184. Tronrud, D. E., Schmid, M. F., and Matthews, B. W. (1986) Structure and X-ray amino acid sequence of a bacteriochlorophyll a protein from Prosthecochloris aestuarii at 1.9 Å resolution. *J. Mol. Biol. 188*, 443–454.

185. BenShem, A., Frolow, F., and Nelson, N. (2004) Evolution of photosystem I-from symmetry through pseudosymmetry to asymmetry. *FEBS Lett. 564*, 274–280.

186. Tronrud, D. E., Wen, J., Gay, L., and Blankenship, R. E. (2009) The structural basis for the difference in absorbance spectra for the FMO antenna protein from various green sulfur bacteria. *Photosynth. Res. 100*, 79–87.

187. Hu, D. (2001) Investigation of the Fenna-Matthews-Olson protein from photosynthetic green sulfur bacteria. Ph. D. thesis, Arizona State University, Tempe AZ.

188. Leslie, A. G. W. (1999) Integration of macromolecular diffraction data. *Acta Crystallogr. D 55*, 1696–1702.

189. Collaborative Computational Project, Number 4 (1994) The CCP4 suite-programs for protein crystallography. *Acta Crystallogr. D 50*, 760–763.

190. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Mariarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D58*, 1948–1954.

191. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta. Crystallogr. D 60*, 2126–2132.

192. Laskowski, R. A., MacArthur, M. W., Moss, D. S., Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst. 26*, 283–291.

193. Vaguine, A. A., Richelle, J., and Wodak, S. J. (1999) SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr. D55*, 191–205.

194. DeLano, W.L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, Palo Alto, CA, USA.

195. Melkozernov, A. N., Olson, J. M., Li, Y. F., Allen, J. P. and Blankenship, R. E. (1998) Orientation and Excitonic Interactions of the Fenna-Matthews-Olson Protein in Membranes of the Green Sulfur Bacterium Chlorobium tepidum. *Photosyn. Res. 56*: 315–328.

196. Wendling, M., Przyjalgowski, M. A., Gülen, D., Vulto, S. I. E., Aartsma, T. J., van Grondelle, R., and Amerongen, H. (2002) The quantitative relationship between structure and polarized spectroscopy in the FMO complex of Prosthecochloris aestuarii: refining experiments and simulations. *Photosynth Res 71*, 99–123.

197. Vulto, S. I. E., Neerken, S., Louwe, R. J. W., de Baat, M. A., Amesz, J., and Aartsma, T. J. (1998) Excited-state structure and dynamics in FMO antenna complexes from photosynthetic green sulfur bacteria. *J. Phys. Chem. 102*,10630–10635.

198. Renger, T. (2010) Theory of excitation energy transfer: from structure to function. *Photosynth. Res. 102*, 471–485.