Studies on the Three-dimensional Structures of Proteins

Using X-ray Crystallography

by

Chenda O. Seng

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

Approved November 2013 by the
Graduate Supervisory Committee:

James P. Allen, Chair
Rebekka Wachter
Mark Hayes

ARIZONA STATE UNIVERSITY

December 2013

ABSTRACT

X-ray diffraction is the technique of choice to determine the three-dimensional structures of proteins. In this study it has been applied to solve the structure of the survival motor neuron (SMN) proteins, the Fenna-Mathews-Olson (FMO) from *Pelodictyon phaeum* (*Pld. phaeum*) protein, and the synthetic ATP binding protein DX.

Spinal muscular atrophy (SMA) is an autosomal recessive genetic disease resulting in muscle atrophy and paralysis via degeneration of motor neurons in the spinal cord. In this work, we used X-ray diffraction technique to solve the structures of the three variant of the of SMN protein, namely SMN 1-4, SMN-WT, and SMN-Δ7. The SMN 1-4, SMN-WT, and SMN-Δ7 crystals were diffracted to 2.7 Å, 5.5 Å and 3.0 Å, respectively. The three-dimensional structures of the three SMN proteins have been solved.

The FMO protein from *Pld. phaeum* is a water soluble protein that is embedded in the cytoplasmic membrane and serves as an energy transfer funnel between the chlorosome and the reaction center. The FMO crystal diffracted to 1.99Å resolution and the three-dimensional structure has been solved.

In previous studies, double mutant, DX, protein was purified and crystallized in the presence of ATP (Simmons et al., 2010; Smith et al. 2007). DX is a synthetic ATP binding protein which resulting from a random selection of DNA library. In this study, DX protein was purified and crystallized without the presence of ATP to investigate the conformational change in DX structure. The crystals of DX were diffracted to 2.5 Å and the three-dimensional structure of DX has been solved.

DEDICATION

Dedicated to my parents for giving me the love and support to succeed at my goals. I would never be where I am today without their encouragements.

ACKNOWLEDGMENTS

I wish to thank Dr. James P. Allen and Dr. JoAnn C. Williams for given me the great opportunity as an undergraduate and graduate researcher to participate in research. I would like to thank Dr. Allen for given me a chance to learn about many paths in research and allowed me to work freely and develop the strategies as a graduate student. Finally, I wish to thank families and friends for their supports.

TABLE OF CONTENTS

Page

LIST OF TABLES viii

LIST OF FIGURES ix

CHAPTER

1 An overview of x-ray crystallograpy 1

Introduction 1

Basic concept of crystallization 2

Conditions affecting crystal growth 3

Basic concept of X-ray diffraction 5

The unit cell 8

Data collection and analysis 11

References 14

2 INTRODUCTION TO SPINAL MUSCULAR ATROPHY 16

Introduction 16

Classification of SMA 16

SMN protein 18

References 24

3 THREE-DIMENSIONAL STRUCTURE OF SURVIAL MOTOR NEURON (SMN) PROTEIN : SMN 1-4 27

Introduction 27

Materials and methods 28

SMN 1-4 protein expression and purification 28

Results 30
Crystallization 30
Data collection and analysis 31
Refinement 32
Three-dimensional structure of SMN 1-4 33
Discussion 35
Comparison of SMN Tudor domains 35
Comparison of SMN Tudor domains to other proteins 36
References 40
4 THREE-DIMENSIONAL STRUCTUREs OF SURVIVAL MOTOR NEURON (SMN) PROTEINS: SMN-WT AND SMN-Δ7 43
Introduction 43
Materials and methods 45
SMN-WT and SMN-Δ7 protein expression and purification 45
Size exclusion chromatography (SEC) 45
Results and discussion 45
Protein crystallization 45
Data collection 47
Data analysis 48
Refinement 49
Three-dimensional structure of SMN- Δ7 50
Three-dimensional structure of SMN-WT 52
References 57

5 The three-dimensional STRUCTURE OF FMO PROTEIN
FROM Pelodictyon phaeum 59
Introduction 59
Materials and methods 61
Protein expression and purification 61
Results 62
Crystallization 62
Data collection and refinement 62
Three-dimensional structure of the FMO protein from *Pld. phaeum* 64
Discussion 69
Comparison of protein-BChl interactions in the three FMO structures . 73
Comparison of position and orientation of *BChl* 8 in the three
FMO structures 79
Implications for optical spectroscopy and energy transfer 80
References 86
6 THREE-DIMENSIONAL STRUCTURE OF Synthetic ATP
BINDING PROTEIN (DX) 90
Introduction 90
Results and discussion 91
Protein crystallization 91
Data collection and analysis 92
Structure refinement 93

The three-dimensional structure of DX protein 94

References..... 96

REFERENCES..... 97

LIST OF TABLES

Table		Page
2.1	Classification of SMA	17
3. 1	Folding buffers	30
3. 2	Summary of X-ray diffraction data and refinement of SMN 1-4	32
3. 3	The organization of the residue ranges of each domain	34
4. 1	Summary of X-ray diffraction data and refinement of SMN-WT and SMN-Δ7	50
4. 2	Summary of secondary structure and domain arrangment of SMN-Δ7	52
4. 3	Summary of secondary structure and domain arrangment of SMN-WT	55
5. 1	Crystallographic data summary of FMO from *Pld. phaeum*	64
6. 1	Summary of X-ray diffraction data and refinement of DX	94

LIST OF FIGURES

Figure		Page
1. 1	Diffraction from one dimensional array of molecules	7
1. 2	Diffraction from crystallographic planes	8
1. 3	Bravais lattices	10
2. 1	Genetic basis of SMA	18
2. 2	The known roles of the SMN protein	19
2. 3	Schematic representation of the SMN complex	20
2. 4	Structure of the SMN Tudor domain	21
2. 5	Partial SMN structure	22
2. 6	The three-dimensional structure of MBP-YG box dimer	23
3. 1	SMN 1-4 sequence	28
3. 2	Three-dimensional structure of SMN 1-4 protein	33
3. 3	Subunit A color-coded according to the encoded exon	34
3. 4	Domains of subunit A of SMN 1-4	35
3. 5	Comparison of the Tudor domains	36
3. 6	The comparison of SMN Tudor domain to Sm structures	37
3. 7	Comparison of the three domains of SMN to several other protein components	38
3. 8	Several missense mutations have been found in exons 1-4 in SMA patients	39
4. 1	Localization of the nucleotides	43
4. 2	SMN-WT and SMN-Δ7 sequences	44

4. 3 Image of SMN-Δ7 protein crystals 46

4. 4 Image of X-ray diffraction of SMN-Δ7 48

4. 5 The three-dimensional structure of SMN-Δ7 51

4. 6 Domains of SMN-Δ7 52

4. 7 The three-dimensional structure of SMN-WT 53

4. 8 The monomer of SMN-WT 54

4. 9 Comparison of SMN-WT domains 55

4. 10 Comparison of C-terminal region of SMN-WT and SMN-Δ7 56

5. 1 The three-dimensional structure of FMO from *Pld. phaeum* 66

5. 2 The interface between subunits of the trimer of FMO from *Pld. phaeum* ... 68

5. 3 The three-dimensional structure of BChl cofactors of each subunit of FMO from *Pld. phaeum* 69

5. 4 Strongly conserved regions for FMO proteins from different species 72

5. 5 Comparison of the structure of BChl 1 and nearby amino acid residues 75

5. 6 Comparison of the structure of BChl 5 and nearby amino acid residues 78

5. 7 Superposition of the old model and the new model of BChl 8 with the omit electron density map 80

5. 8 Absorption and CD spectra of five FMO proteins 82

5. 9 Cofactor arrangement in FMO from *Pld. phaeum* with exciton couplings .. 84

6. 1 Three-dimensional structure of 18-19 and DX proteins 91

6. 2 Image of DX protein crystals 92

6. 3 Image of X-ray diffraction of DX protein 93

6. 4 Three-dimensional structure of DX protein 95

Chapter 1

AN OVERVIEW OF X-RAY CRYSTALLOGRAPHY

Introduction

In this dissertation, X-ray diffraction is used to solve the three-dimensional structure of three different proteins, survival motor neuron (SMN) protein associated with spinal muscular atrophy (SMA), Fenna-Mathews-Olsen (FMO) a photosynthetic protein from *Pelodictyon phaeum* (*Pld. phaeum*), and the synthetic ATP-binding DX protein.

This dissertation is comprised of six chapters. In Chapter 1, an overview of X-ray crystallography, the basic concept of X-ray diffraction is discussed based partially on a published paper (Allen et al., 2009). This chapter will present the concept of the X-ray crystallography technique used for solving three-dimensional structures of proteins. Chapter 2, introduction to spinal muscular atrophy (SMA), covers the basic understanding of SMA, comparison of the SMN protein to related structures of other proteins, and the roles of SMN in the cell. In Chapter 3, three-dimensional structure of SMN 1-4, the three-dimensional structure of SMN 1-4 is presented, including a comparison to the previously published SMN Tudor domain (Selenko et al., 2001; Sprangers et al., 2003). The model of SMN 1-4 was used for molecular replacement to solve the structure of SMN-WT and SMN-Δ7. In Chapter 4, three-dimensional structures of survival motor neuron (SMN) proteins, SMN-WT and SMN-Δ7, the three-dimensional structures of SMN-WT and SMN-Δ7 are described. The three-dimensional structure of the FMO protein from *Pld. phaeum* is described in Chapter 5 and the implication for energy transfer is discussed, which is covering material published in Photosynthesis Research (Larson et al., 2011). Also discussed are a modification of the structure as

presented in Tronrud and Allen, 2012. The three-dimensional structure of synthetic ATP binding protein (DX) will be discussed in Chapter 6, including protein crystallization, X-ray diffraction, data analysis and three-dimensional structure.

Multiple authors contributed to the published article (Allen et al., 2009) that serves as the basis of this chapter, and my contribution to this work was incorporation of X-ray diffraction figures and related text (Figures 1.3-1.5).

Basic concept of crystallization

X-ray crystallography is an excellent tool for visualizing protein structure at the atomic level, which enriches our understanding of protein function. To date, the Protein Data Bank (PDB) contains approximately 82,000-protein structures determined by X-ray crystallography. X-ray diffraction crystallography is one of the most common methods for obtaining detail structure of protein and their complexes such as protein-protein, cofactors and ligands. In addition to X-ray diffraction, many proteins of fewer than 200 residues have been solved by nuclear magnetic resonance (NMR) in which provides a model of the molecule in solution, rather than in the crystalline state.

The most common method used for protein crystallization is vapor diffusion in which a droplet with the lower concentration of crystallization components (protein and precipitant) is equilibrated against a reservoir solution of higher concentration in a closed container. It can be done by the hanging drop or sitting drop methods in crystallization plates of 24 or 96 wells. Hanging drop vapor diffusion uses a very small volume of the protein solution, 2 µl to 10 µl. Sometimes, the volume of the protein used can be increased up to 20 µl, depending on the surface tension of the protein solution. When the surface tension of the protein solution is very low, the sitting drop technique can be

substituted. In both cases, the reservoir should be filled with a precipitant solution with a volume of approximately 500 µl to 1 ml.

Conditions affecting crystal growth

There are many factors that influence the crystallization of proteins, such as purity of the protein, protein concentration, choice of buffers, pH, temperature and precipitants. Crystallization trials are repeated multiple times to obtain good quality crystals. The most important factor in crystallization is the purity of the protein, as impure protein will not yield suitable crystals. The impurity of the protein solution results in a decrease of crystal quality, because crystals are not well packed in a crystal lattice form; in other words, it is disordered. When the protein solution is not pure, micro-crystals are sometimes observed and maximum growth cannot be reached. Different batches of protein will have different purities, which affects crystal growth.

Depending on types of buffer and the protein concentration, the quality of crystals can be varied. Slight changes in protein concentration or buffer affect the size and quality of the crystals (Rhodes, 2006). HEPES (4,2-hydroxyethyl-1-piperazineethansulfonic acid) or Tris (Tris hydroxymehtyl aminomethane) are the most common buffers used in crystallization. The pH is the most important determinant of protein solubility because proteins accumulate different charges at a different pH values (Bergfords, 1999). By changing of pH, it alters the ionization state of surface amino acid residues, which allows protein to contact and form crystals. Besides the buffer and pH, temperature is also another important parameter in crystallization process since it affects protein stability and solubility, which leads to crystal growth.

The concentration of precipitants plays an important role in the protein solubility (Papanikolau et al., 2004). Many precipitants can affect the protein solubility such as salts, polyethylene glycol (PEG), and organic solvents. The ionic strength of salt enhances the hydrophobic effects of the protein. Depending upon the nature of the particular cation and anion of the salts, it can maintain the solubility of the protein in the solution in which crystals may form. The salt ions can participate in specific interactions with the protein molecules in the solution and result in crystal formation (Durbin and Feher, 1996). Depending on the nature of protein, ammonium sulfate, $(NH_4)_2SO_4$, can also be used for protein crystallization. Control of pH should be taken into consideration when the $(NH_4)_2SO_4$ salt is added, since the sulfate ions will change the pH of the protein solution. The isoelectric point, pI, of the protein can also be used as a starting point for the choice of pH and precipitants (Bergfords, 1999). At a certain pH, which is defined as the pI, the numbers of positive charges and negative charges on a protein are equal and the protein is electrically neutral (Rosenberg, 2004).

When all conditions are set for the initial screen, possible additives should be tested. Possible additives are metal ions, chelators, low-molecular-weight organic compounds, detergents, and ligands (Wood, 1990). These additives have been found to improve the quality of the crystal growth in some cases.

The process of nucleation is analogous to the process of condensation of droplets (Durbin and Feher, 1996). The nucleation could be homogenous or heterogeneous. Homogenous nucleation involves only association of the molecule itself. Favorable conditions and the formation of an amorphous aggregate are hard to achieve for protein crystallization, because protein behavior changes in different environments.

Heterogeneous nucleation is involved in association of the protein to other substances. Heterogeneous nucleation is indicated when crystals tend to form on dust particles, fibers, or other visible objects. Therefore, prior filtration may significantly reduce the number of crystals (Durbin and Feher, 1996). When objects contact with protein molecules, the nucleation forms at the lower supersaturation level. Therefore, it poisons the surface of the crystal. Excessive crystal nucleation and growth rates often result in poor diffraction quality (Jovine, 2000).

Basic concept of X-ray diffraction

X-ray diffraction has been used for 100 years to solve the structures of molecules. The first structures of proteins, myoglobin and hemoglobin, were solved over 50 years ago (Kendrew et al., 1960; Perutz et al., 1960). Diffraction is the technique of choice for the determination of the three-dimensional structures of proteins. The basic concept of diffraction can be understood by considering a simple array of molecules that are uniformly spaced in a line (Figure 1.1). 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

$$\Delta = a \sin \theta \qquad (1.1)$$

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

$$n\lambda = a \sin \theta, \qquad (1.2)$$

where n is an integer.

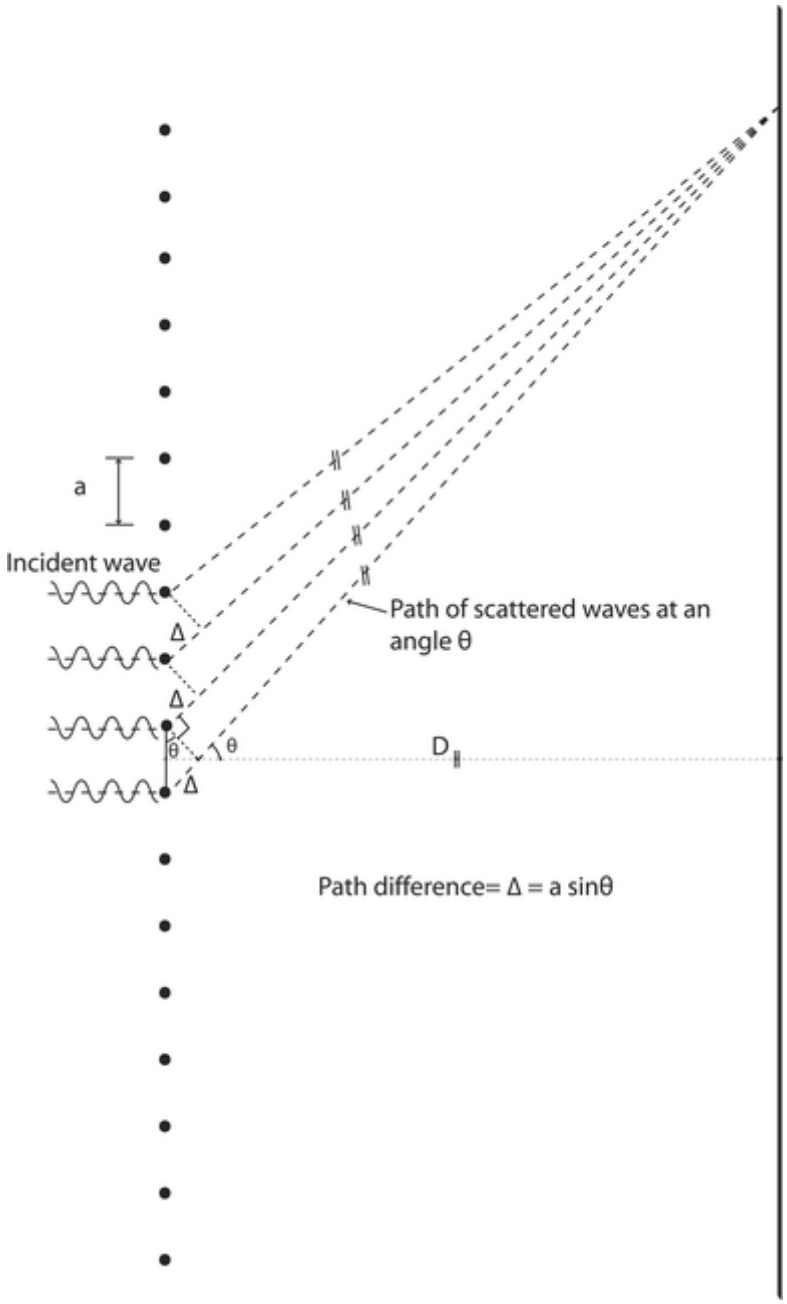

Figure 1.1. Diffraction from a one-dimensional array of molecules. Each molecule is spaced a distance a apart from the neighboring molecule. The incident wave is considered to scatter from each molecule to a point at a large distance D, with the scattered waves having different path lengths. At an angle θ, the path length difference between neighboring molecules is given by a sin θ. When the path length difference is equal to a multiple of the wavelength, the waves add constructively and a diffraction peak is observed. For simplicity, only four of the scattered waves are shown in the figure (Allen et al., 2009).

For diffraction from a physical crystal, the scattering in three dimensions must be considered, although the analysis is essentially the same (Figure 1.2). 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:

$$n\lambda = 2a \sin \theta, \qquad (1.3)$$

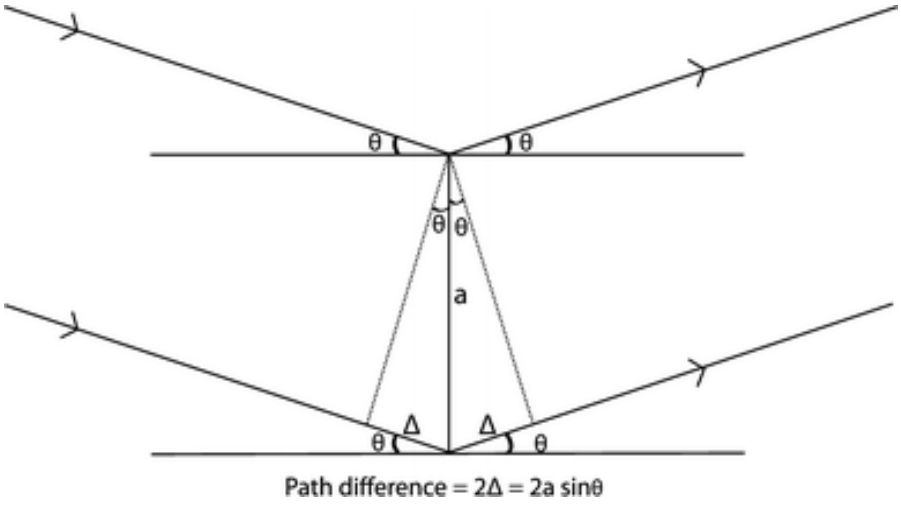

Figure 1.2. Diffraction from crystallographic planes. The crystal is considered to consist of a series of planes that can each scatter the incident wave. As was true for the one-dimensional array, the difference between the scattered waves is the difference in the path length between the neighboring planes, which equals 2a sin θ. When this path difference equals a multiple of the wavelength, the waves add constructively and a diffraction peak is observed (Allen et al., 2009).

The unit cell

All crystals, including those 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

great number of unit cells stacked next to each other. As originally studied in the 1800s 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 1.3). In some cases, the symmetry is simple. For example, cubic crystals have a fourfold symmetry along the directions of the edges and twofold symmetry along the diagonals. In other cases, the symmetry is more complexes; for example, unit cells can have a 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 improved compared to crystals with simple rotation axes.

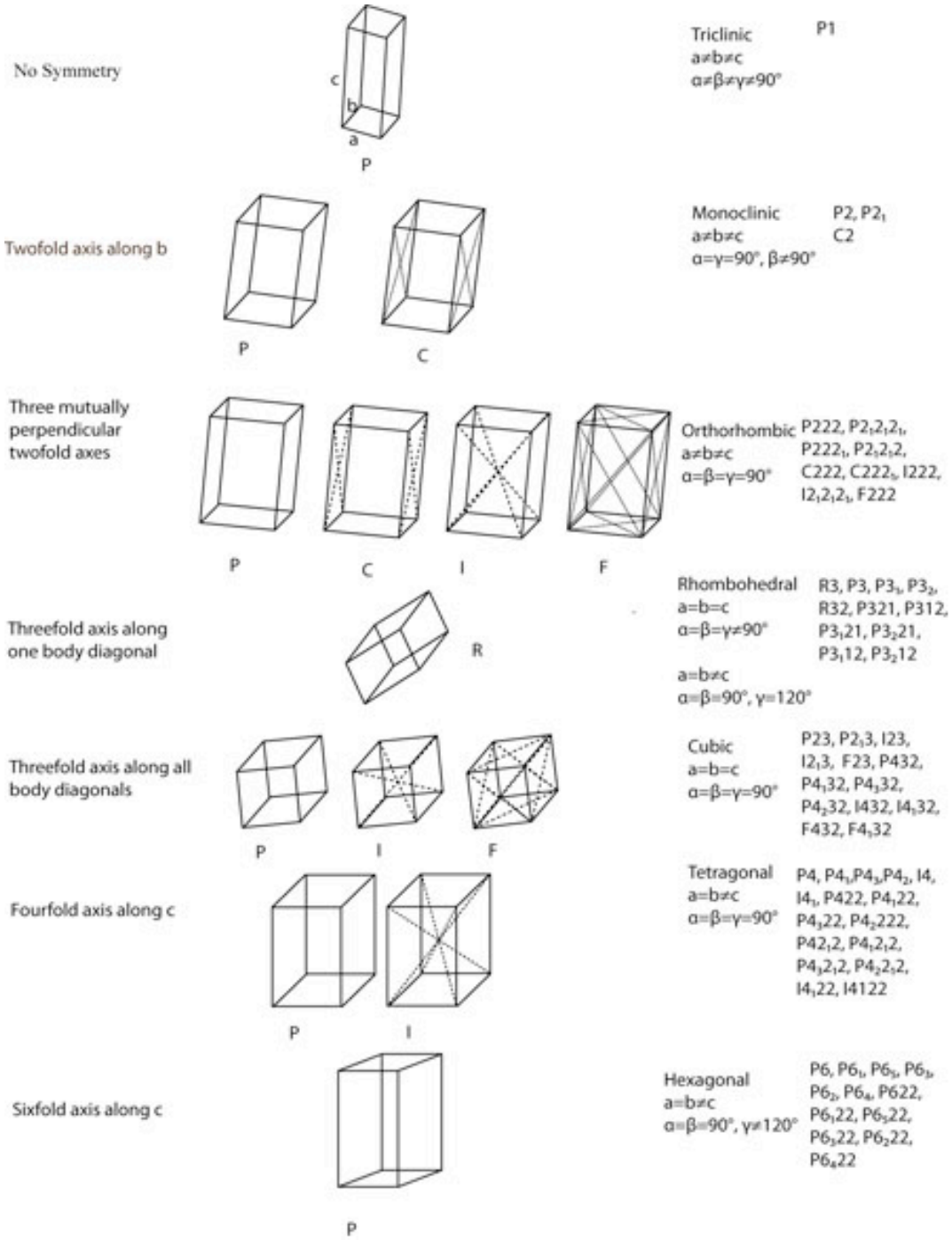

Figure 1.3. Bravais lattices. Shown are the seven types of unit cells and the allowed unique lattices. The relationships among the three characteristic cell lengths, a, b, and c are shown as well as for the three angles, α, β, and γ. Also indicated are the symmetries for each cell type. Each cell type can have primitive (P), face-centered (C), body-centered (I), or all face-centered (F) cells. The dashed lines show that for these different cell types, there can be additional lattice points. The body-centered cells have an additional lattice point in the center (Allen et al., 2009).

Data collection and analysis

Once crystals have been obtained, X-ray diffraction measurements are performed using either an in-house facility 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 images. The quality of the data is usually described by the resolution limit, which 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. In addition to the resolution limit, there are other measures of the quality, with the crystals 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 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 it is 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 amplitude and phase, however, 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*

$$\rho(x,y,z) = \sum_{hkl} F_{hkl}\ e^{-2\pi i\,(hx+ky+lz)} \qquad (1.4)$$

Notice that the summation yields not the position of the nuclei but rather $\rho(x,y,z)$, 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.

Several experimental approaches can be used to solve the phase problem. The first approach, multiple anomalous dispersion, MAD, 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. In the second approach MIR (multiple isomorphous replacement) a metal is incorporated into the protein and the diffraction is compared to the protein without the metal. In this case, measurements of the diffraction from the crystals with different bound metals are needed to give accurate estimations of the phases. The last approach, molecular replacement, MR, can be utilized to solve the structures. In this case, there must be an existing structure that is highly homologous to the unknown structure. Essentially, the homologous structure is virtually 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. 1.4. The outcome is a map of the electron density that is usually contoured at a level significantly above the background. Fitting of the maps involves identifying the atoms that give rise to each region of the density. While this is often manually done on a graphics terminal by a crystallographer, increasingly, the analysis of electron density maps is being performed directly by sophisticated programs. Once the native structure has been determined, structural modifications by mutagenesis or biochemical techniques can be rapidly identified provided the structural differences are small.

References

Allen, J.P., Seng, C., Larson, C. (2009). Structures of proteins and cofactors: X-ray crystallography. Photosynth Res. *102,* 231–240.

Bergfords, T. M. (1999). Protein Crystallization: Techniques, Strategies, and Tips (La Jolla, California: International University Line)

Durbin, S.D., and Feher, G. (1996). Protein crystallization. Annu. Rev. Phys. Chem. *47*, 171-204.

Drenth, J. (2007) Principles of Protein X-Ray Crystallography. 3rd edition. (New York: Springer)

Jovine, L. (2000). A simple technique to control macromolecular crystal nucleation efficiently using a standard vapour-diffusion set up. J. Appl. Cryst. *33*, 988-989.

Larson, C.R., Seng, C.O., Lauman, L., Matthies, H.J., Wen, J., Blankenship, R.E., Allen, J. P. (2011). The three-dimensional structure of the FMO protein from *Pelodictyon phaeum* and the implications for energy transfer. Photosynth Res. *107,* 139–150.

Kendrew, J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.G., Davies, D. R., Phillips, D.C., Shore, V.C. (1960). Structure of myoglobin. Nature *185*, 422–427.

Papanikolau, Y., Kotsifaki, D., Fadouloglou, V.E., Gazi, A.D., Glykos, N. M., Cesareni, G., Kokkinidis, M. (2004). Ionic strength reducers: an efficient approach to protein purification and crystallization. Application to two Rop variants. Acta. Cryst. *D60*, 1334-1337.

Perutz, M.F., Rossmann, M.G., Cullis, A. F., Muirhead, H., Will, G., North, A.C.T. (1960). Structure of haemoglobin. Nature *185*, 416–422.

Rhodes, G. (2006). Crystallography Made Crystal Clear (Elsevier Academic Press)

Selenko, P., Sprangers, R., Stier, G., Buhler, D., Fischer, U., Sattler, M. (2001). SMN tudor domain structure and its interaction with the Sm proteins. Nature *8*, 27–31.

Sprangers, R., Groves, M.R., Sinning, I., and Sattler, M. (2003). High-resolution X-ray and NMR structure of the SMN Tudor domain: Conformational variation in the binding site for symmetrically dimethylated arginine residues. J. Mol. Biol. *327,* 507-520.

Tronrud, D.E., and Allen, J.P. (2012). Reinterpretation of the electron density at the site of the eighth bacteriochlorophyll in the FMO protein from *Pelodictyon phaeum*, Photosynth Res. *112*, 71–74.

Wood, S.P. (1990). Purification for Crystallography. In Protein Purification Application: A Practical Approach, E.L.V Harris and S. Angal, ed. (Oxford, Oxford University Press) pp. 45–58.

Chapter 2

INTRODUCTION TO SPINAL MUSCULAR ATROPHY

Introduction

Spinal muscular atrophy, SMA, is an autosomal recessive disease characterized by degenerative disorder of anterior horn cells of the spinal cord (Lefebvre et al., 1995, Wirth, 2000). SMA causes child death in 1 in 6,000 to 10,000 live-births (Lefebvre et al., 1995; Lorson et al., 1999) and is the second leading cause in the death of young children after cystic fibrosis (Wirth, 2000). There are three common types of SMA ranging from severe to mild forms (Alia et al., 2009) dependent upon the number of copies of the survival motor neuron (*SMN1*) gene. The survival of motor neuron (SMN) gene is located at position q13 of the chromosome 5 (Young et al., 2003). There are two copies of the *SMN* gene, *SMN1* (telomeric copy) and *SMN2* (centromic copy), which are almost identical, except a single nucleotide change from C to T in the *SMN2* gene that causes an alternation of exon 7 splicing resulting in SMA (Lefebvre et al., 1995; Wang and Dreyfuss, 2004; Young et al., 2003). This exon skipping event leads to a lower cellular functionality; however, it only affects the anterior horn cells of the spinal cord. Approximately 95% of SMA patients are found to have the *SMN 1* gene absent (Burghes, 1997; Clermont et al., 1997, Fischer et al., 1999; Lefebvre et al., 1995; Young et al., 2003). Depending upon the arrangement of the SMN genes, SMA can be ranged from acute to mild forms.

Classification of SMA

In 1991, the International SMA Consortium classified three common clinical types of SMA (Table 2.1) (Munsat, 1991; Nicole et al., 2002; Wirth, 2000). The absence

of the *SMN1* gene causes the severity of SMA. In SMA type I patients, *SMN1* gene are lost or mutation while the *SMN2* gene remains. Type I, namely Werdnig-Hoffmann disease, the most lethal form of SMA. It is characterized by severe, generalized muscle weakness and hypotonia at birth or within the next 6 months (Nicole et al., 2002). Type I patients are unable to sit, stand or walk, and many have respiratory problems. Usually, death occurs within two years of age by respiratory failure (Lefebvre et al., 1995). The homozygous deletion of the *SMN1* gene with an extra copy of the *SMN2* gene appears in a majority of type II SMA, which is an intermediate form. Type II patients are able to sit but never gain the ability to stand or walk without assistance (Nicole et al., 2002). The life expectancy of those with type II SMA is between two and four years. In a majority of type III SMA patients, two extra copies of the *SMN2* gene are present. Type III, also known as Kugelberg-Welander disease, is the less severe form of SMA. Patients carrying type III have proximal muscle weakness (Lefebvre et al., 1995; Nicole et al., 2002). Symptoms may appear between two and 17 years of age. Patients are able to walk during the early stages, but require the support of a wheelchair during adulthood. Death may occur in early adulthood for those with type III SMA.

Table 2.1. Classification of SMA

Type	Age of onset	Function	Age of occurrence/death
I	at birth-6 months	never sit	< 2
II	< 18 months	able to sit, never stand	> 2
III	> 18 months	stand alone	Adult

Table adapted from Munsat, 1991

As shown in Figure 2.1, the transcription of *SMN1* is mainly produced as the full-length form whereas *SMN2* is transcribed into both the full-length and the truncated form lacking exon 7, SMN-Δ7. The truncated form is found most abundant in SMA patient and

correlated to the disease severity (Young et al., 2003). The *SMN2* gene produces mostly unstable proteins, which are degraded rapidly and expressed in very small amount of stable proteins.

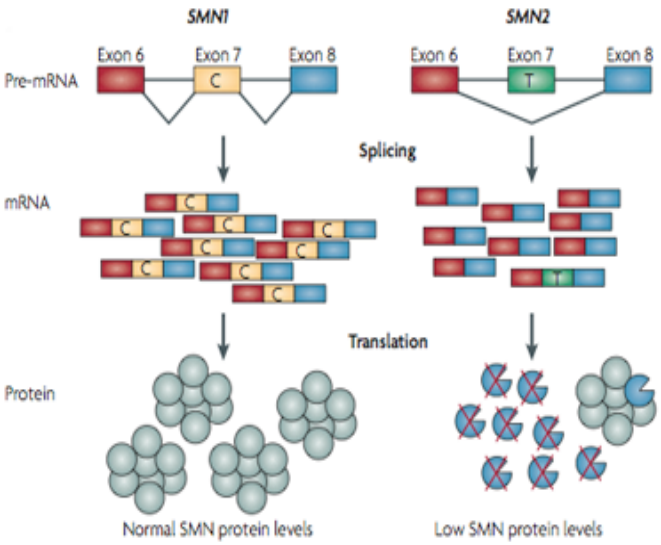

Figure 2.1. Genetic basis of SMA. The *SMN1* and *SMN2* are identical except the single-nucleotide change in exon 7 (C to T). *SMN1* is produced a normal protein whereas *SMN2* is produced low SMN protein and truncated form lacking exon 7, leading to unstable protein that is degraded rapidly (Burghes and Beattice, 2009).

SMN protein

The SMN gene encodes a protein of 294 amino acids that has a molecular weight of 38 kDa (Malatesta et al., 2004). SMN forms high molecular weight complexes in the cytoplasm and nucleus of the cell (Pellizzoni et al., 1999) (Figure 2.2). In the cytoplasm, SMN interacts with Sm proteins, which assist in the formation of the Sm core and helps the assembly of the small nuclear ribonucleoproteins (snRNPs) for nuclear transport (Zhang et al., 2011). Sm proteins are termed because of their reactivity with autoantibodies of the Sm serotype from patients with systemic lupus erythematosus (MacKenzie and Gendon, 2001). With lesser amounts of SMN present within the cell, the

Sm core is not formed resulting in limited of snRNP processing and failure to assemble the snRNPs (Gabanella et al., 2007; Pellizzoni et al., 1999). In the nucleus, SMN complex plays an important role in the pre-mRNA splicing. Once the uridine (U) small nuclear RNA (snRNA), U1, U2, U4, U5 and U6 enter the nucleus, the SMN protein assembled the snRNPs for the pre-RNA splicing. Significantly, the impairment of SMN function in Sm core formation causes decreasing levels of a subset of snRNPs (Gabanella et al., 2007). In SMA patients, the SMN-Δ7 protein is found to have less effective binding with the snRNPs complex than the SMN-WT.

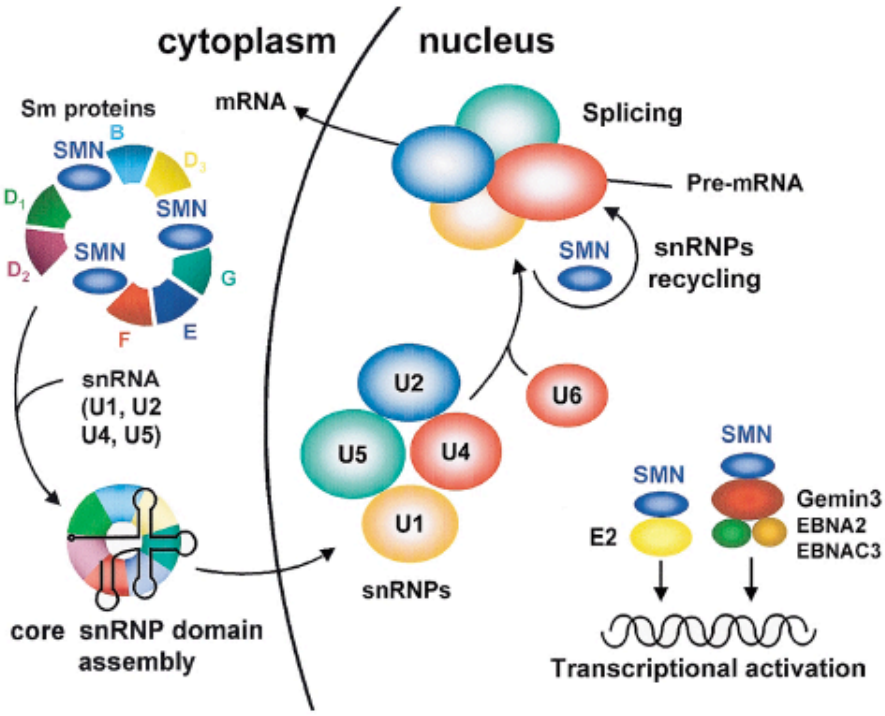

Figure 2.2. The known roles of the SMN protein: cytoplasmic assembly of the core snRNP complex; nuclear regeneration/reactivation of snRNPs involved in pre-mRNA splicing; and direct and indirect transcriptional activation (MacKenzie and Gendon, 2001).

SMN interacts with the Gemin proteins forming the SMN complex and highly enriched in structures termed gems (Gemini of coiled bodies) (Gubitz et al., 2004). SMN

has been found to directly interact with Gemin2, Gemin3, Gemin5, and Gemin7, and indirectly interact with Gemin4 and Gemin6. Association of Gemin4 and Gemin6 with the SMN protein is through the association of Gemin3 and Gemin7, respectively (Figure 2.3). Gemin2, which was formerly known as SMN interacting protein 1 (SIP1), was identified as a core component of the SMN complex (Ogawa et al., 2007; Wang and Dreyfuss, 2001). Recent studies have found that Gemin2 plays an essential role in snRNPs assembly through the stabilization of the SMN oligomer/complex via novel self-interaction (Ogawa et al., 2007; Shpargel and Matera, 2005). Gemin2 and Gemin3 degrade when the amount of SMN is reduced. Furthermore, another study found that the SMN complex lacking Gemin5 fails to promote uridine-rich (U) snRNP assembly (Otter et al., 2007).

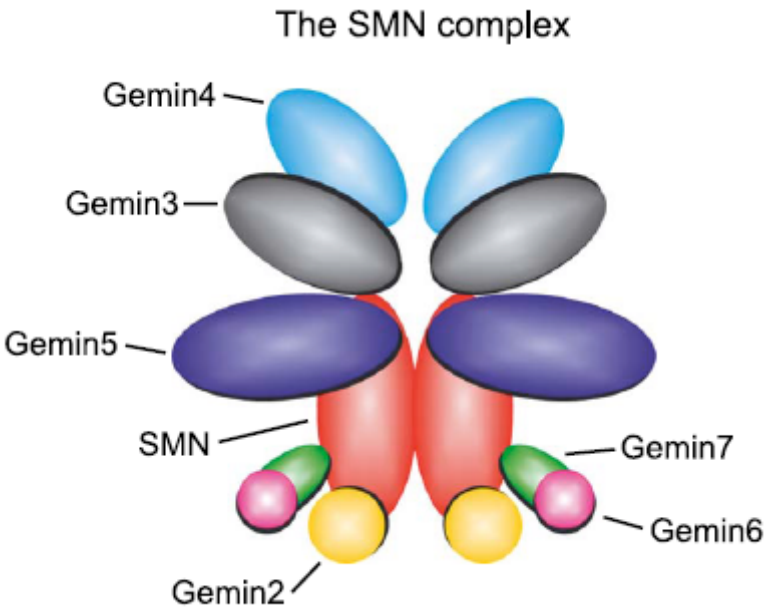

Figure 2.3. Schematic representation of the SMN complex. For simplicity, the SMN complex is illustrated in dimeric form (Gubitz et al., 2004).

The overall three-dimensional structure of SMN protein remains unknown with the exception of the Tudor domain, a short helix that binds to Gemin2, and a tyrosine-glycine-rich (YG box) region near the C-terminal region. The structure of the Tudor domain was solved by using X-ray crystallography (PDB:1MHN) (Sprangers et al., 2003) and NMR (PDB:1G5V) (Selenko et al., 2001). The SMN Tudor domain consists of 5 strongly bent anti-parallel β-barrels (Figure 2.4). This domain consists of amino acid residues 92-144 of the SMN protein and is largely encoded by exon 3 of the SMN gene (Buhler et al., 1999; Selenko et al., 2001; Sprangers et al., 2003).

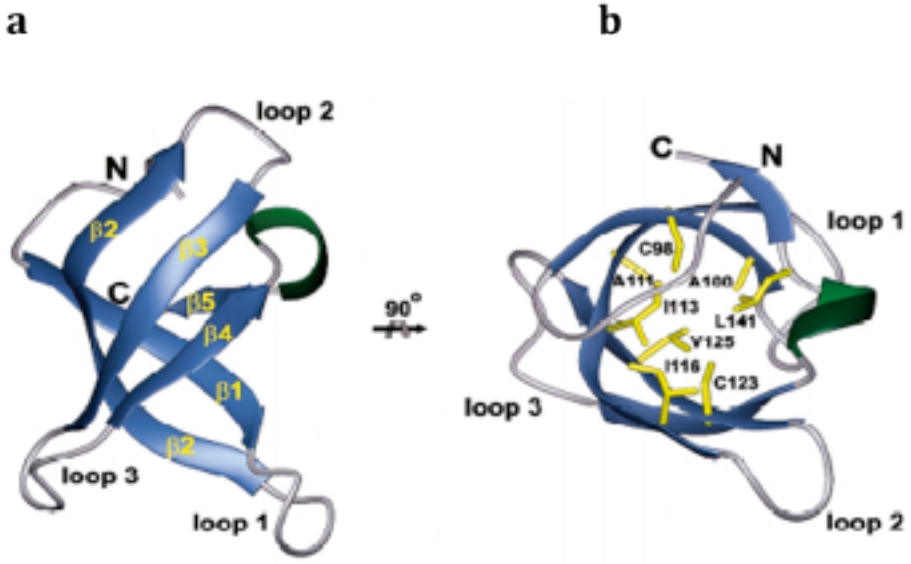

Figure 2.4. Structure of the SMN Tudor domain. a) The five strands form a barrel-like fold that is lined at the bottom by the long curved β2 strand and closed by anti-parallel interaction between β1 and Leu 12 in the short β5 strand. Strands β1-4 are connected by short turns, while strands β4 and β5 are linked by a helical turn. b) The formation of a hydrophobic core stabilizes the structure by highly conserved amino acid residues (Cys 98, Ala 100, Ala 111, Ile 113, Ile 116, Cys 123, Val 125, and Leu 141) (Selenko et al., 2001).

The Tudor domain facilitates SMN interactions in complexes with Sm proteins and snRNPs. The glutamic acid residue, position 134 (E134) of the β4 strand of the

Tudor domain interacts with the Sm protein at the C-terminus. A SMA-causing point mutation, glutamic acid to lysine at the position 134 (E134K), within the SMN Tudor domain prevents Sm binding (Selenko et al., 2001; Sprangers et al., 2003). The mutation does not specifically alter the Sm binding site; however, it changes the overall structure of the SMN (Buhler et al., 1999). The mutation of this single amino acid residue results in a change in the charge distribution affecting the electrostatic interactions with the positively-charged C-terminal Sm tails (Selenko et al., 2001; Sprangers et al., 2003).

Recently, the SMN_{Ge2BD} was crystallized as fusion product with Gemin2 and Sm pentamer complex (Zhang et al., 2011). The three-dimensional structure of the Gemin2-SMN_{Ge2BD}-Sm pentamer complex has been solved using X-ray (PDB:3S6N) (Zhang et al., 2011) and NMR (PDB:2LEH) (Sarachan et al., 2012). SMN_{Ge2BD} is an N-terminal region of SMN, the residue range 37 to 51. This small section of the SMN structure forms a helix that interacts with Gemin2 (Figure 2.5).

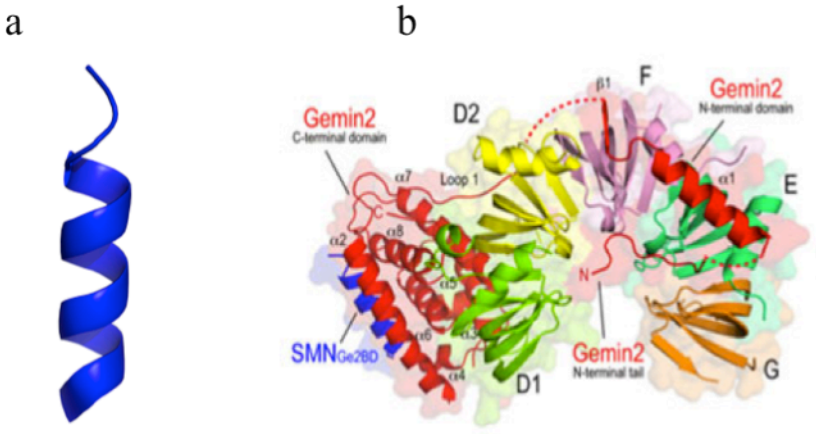

Figure 2.5. Partial SMN structure. a). Helix of N-terminal region, residues 37 to 51, of SMN that bind to Gemin2. b). SMN (blue) interaction with Gemin2 with five Sm proteins (Zhang et al., 2011).

The tyrosine-glycine rich, YG, region of SMN was crystallized as a fusion product with the maltose binding protein, MBP (Martin et al., 2012). The three-dimensional structure of the C-terminal YG region, containing residues ranging from 263 to 294 has been solved (Martin et al., 2012). The YG region forms a helix and interacts with each other to form a dimer (Figure 2.6).

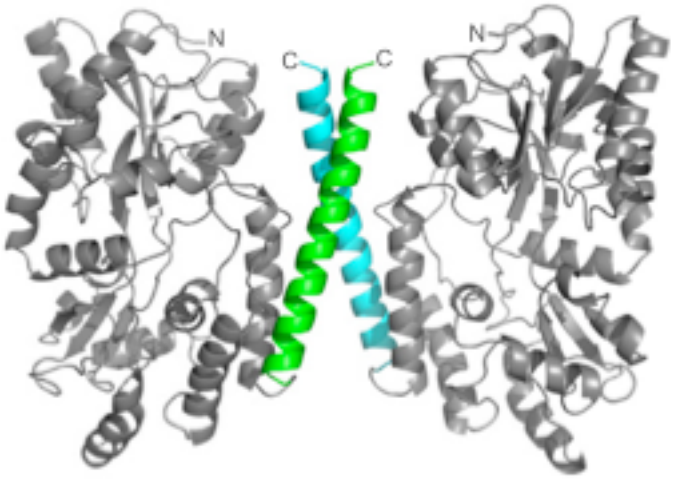

Figure 2.6. The three-dimensional structure of Maltose binding protein-tyrosine-glycine, MBP-YG, box dimer. MBP is shown in gray and the YG box monomer of SMN is shown in green and turquoise (Martin et al., 2012).

References

Buhler, D., Raker, V., Luhrmann, R., 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.

Burghes, A.H. (1997). When is a deletion not a deletion? When it is converted. Am. J. Hum. Genet. *61,* 9–15.

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

Burglen, L., Lefebvre, S., Clemont, O., Burlet, P., Viollet, L., Cruaud, C., Munnich, O., Melki, J. (1996). Structure and organization of the human survival motor neuron (SMN) gene. Gemonics. *32*, 479–482.

Gabanella, F., Butchbach, M.E., Saieva, L., Carissima, C., Burghes, A.H. M., Pellizoni, L. (2007). Ribonucleoprotein assembly defects correlate with spinal muscular atrophy severity and preferentially affect a subset of spliceosomal snRNPs. PloS ONE. *e921,* 1–12.

Gubitz, A.K., Feng, W., Dreyfuss, G. (2004). The SMN complex. Experimental Cell Research *296*, 51–56.

Lefebvre, S., Burglen, L., Reboullet, S., Clemon, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., Le Pasller, D., Frezal, J., Cohen, D., Weissenbach, J., Munnich, A., Melki, J. (1995). Identification and characterization of a spinal muscular atrophy-determining gene. Cell *80*, 155–165.

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

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

MacKenzie, A., and Gendon, N. (2001). Tudor reign. Nature Struct. Biol. *8*, 13–15.

Malatesta, M., Scassellati, C., Meister, G., Plottner, O., Buhler, D., Sowa, G., Martin, T., Keidel, E., Fischer, U., Fakan, S. (2004). Ultrastructural characterization of a nuclear domain highly enriched in survival of motor neuron (SMN) protein. Expirm. Cell Research. *292*, 312–321.

Martin, R., Gupta, K., Ninan, N.S., Perry, K., Van Duyane, G.D. (2012). The survival motor neuron protein forms soluble glycine zipper oligomers. Struct. *20,* 1–11.

Munsat, T.L. (1991). Workshop report: International SMA collaboration. Neuromuscular Disord*. 1*, 81.

Nicole, S., Diaz, C.C., Frugier, F., Melki, J. (2002). Spinal muscular atrophy: recent advances and future prospects. Muscle Nerve. *26*, 4–13.

Ogawa, C., Usui, K., Aoki, M., Ito, F., Itoh, M., Kai, C., Kanamori-Katayama, M., Hayashizaki, Y., Suzuki, H. (2007). Gemin2 plays and important role in stabilizing the survival of motor neuron complex. J. Biol. Chem*. 282*, 11122–11134.

Otter, S., Grimmler, M., Neuenkirchen, N., Chari, A., Sickmann, A., Fischer, U. (2007). A comprehensive interaction map of the human survival of motor neuron (SMN) complex. J. Biol. Chem*. 282*, 5825–5833.

Pellizzoni, L., Charroux, B., Dreyfuss, G. (1999). SMN mutants of spinal muscular atrophy patients are defective in binding to snRNP proteins. PNAS *94*, 11167–11172.

Sarachan, K.L., Valentine, K.G., Gupta, K., Moorman, V.R., Gledhill Jr., J.M., Bernens, M., Tommos, C., Wand, A.J., Duyne, D.V. (2012). Solution structure of the SMN-Gemin2 complex. Biochem. J. *445,* 361–370.

Selenko, P., Sprangers, R., Stier, G., Buhler, D., Fischer, U., Sattler, M. (2001). SMN tudor domain structure and its interaction with the Sm proteins. Nature *8*, 27–31.

Shpargel, K.B., and Matera, A.G. (2005). Gemin proteins are required for efficient assembly of Sm-class ribonucleoproteins. PNAS *102,* 17372–17377.

Sprangers, R., Groves, M.R., Sinning, I., and Sattler, M. (2003). High-resolution X-ray and NMR structure of the SMN Tudor domain: Conformational variation in the binding site for symmetrically dimethylated arginine residues. J. Mol. Biol*. 327,* 507–520.

Wang, J., and Dreyfuss, G. (2001). Characterization of function domains of the SMN protein *in vivo*. J. Biol. Chem. *276,* 45387–45393.

Wirth, B. (2000). An update of the mutation spectrum of the survival motor neuron gene (SMN1) in autosomal recessive spinal muscular atrophy (SMA). Human Mutation. *15*, 228–237.

Yong, J., Wan, L., Dreyfuss, G. (2004). Why do cells need an assembly machine for RNA-protein complexes? Trends in Cell Biol. *15,* 226–232.

Young, P. J., Francis, J. W., Lince, D., Coon, K., Androphy, E. J., Lorson, C. L. (2003). The Ewing's sarcoma protein interacts with the Tudor domain of the survival motor neuron protein. Mol. Brain Research. *119,* 37–49.

Zhang, R., So, B.R., Li, P., Young, J., Glisovic, T., Wan, L., Dreyfuss, G. (2011). Structure of a key intermediate of the SMN complex reveals Gemin2's crucial function in snRNP assembly. Cell *146,* 384–395.

Chapter 3

THE THREE-DIMENSIONAL STRUCTURE OF SURVIVAL MOTOR NEURON (SMN) PROTEIN: SMN 1-4

Introduction

It has been shown that SMA patients contain mostly the SMN-Δ7 protein with a small amount of the SMN-WT protein (previously explained in Chapter 2). The SMN-WT proteins are degraded rapidly in the cell whereas SMN-Δ7 proteins still remain. The SMN-Δ7 protein cannot compensate the loss SMN-WT, which results in loss of function and affects the cellular function in the spinal cord found in SMA patients. The SMN-WT protein was known to have a self-association (Burhler et al., 1999; Burghes and Beattie, 2009) of which the truncated form would reduce the SMN-SMN interaction. To investigate this, we constructed a truncated form of SMN, which encoded exon 1-4, namely SMN 1-4, was created to reduce the SMN-SMN association. The SMN 1-4 construct consists of 209 amino acid residues (Figure 3.1) and this construct was created to assist in purification, crystallization and determination of the three-dimensional structure of SMN-WT and SMN-Δ7.

We have determined the three-dimensional structure of the SMN 1-4 at a resolution limit of 2.7 Å with R_{work} and R_{free} value of 27.96% and 30.91%, respectively.

```
1    MRGSHHHHHH GMASMAMSSG GSGGGVPEQE DSVLFRRGTG QSDDSDIWDD 50

51   TALIKAYDKA VASFKHALKN GDICETSGKP KTTPKRKPAK KNKSQKKNTA 100

101  ASLQQWKVGD KCSAIWSEDG CIYPATIASI DFKRETCVVV YTGYGNREEQ 150

151  NLSDLLSPIC EVANNIEQNA QENENESQVS TDESENSRSP GNKSDNIKPK 200

201  SAPWNSFLPP PPPMPGPRLG PGK                              223
```

Figure 3.1 SMN 1-4 sequence. The fusion His-tag at the N-terminal of the sequence is shown in blue.

My contribution to this Chapter was solving the three-dimensional structure of SMN 1-4 that was obtained using crystallographic data from Dr. Craig Magee. The three-dimensional structure of SMN 1-4 was used for molecular replacement to solve the three-dimensional structures of SMN-WT and SMN-Δ7. All SMN clones were obtained from the laboratory of Chistian L. Lorson at the University of Missouri-Columbia, in Columbia, Missouri.

Materials and methods

SMN 1-4 protein expression and purification

The methods of SMN 1-4 protein expression, purification, and crystallization were adapted from Dr. Craig L. Magee (Magee, 2006). The SMN 1-4 gene was cloned into the pRSET vector with His_6 tag at the N-terminal region and overexpressed in *Escherichia coli, E.coli,* BL21 (DE3) cells. The cells were grown in super optimal broth (SOB) media at 37 ºC and induced with 1 mM isopropyl β-thiogalactoside (IPTG) when OD_{600} reached 0.6 - 0.9. Following an additional 4 hours of growth, bacteria cells were harvested by gravimetric centrifugation at 4ºC and 6,000 rpm for 10 minutes, and stored at -20 ºC. The cell pellet was approximately 10-15 grams per liter of growth media.

Then cell pellet was re-suspended in 10 ml of cold phosphate buffer saline (PBS) containing 140 mM NaCl, 2.7 mM potassium chloride (KCl), 10 mM sodium phosphate dibasic (Na_2HPO_4), 1.8 mM potassium phosphate monobasic (KH_2PO_4), pH 7.3, at 4 ºC overnight. Cells were lysed by sonication in the presence of 0.2 mg/ml lysozyme, 25 μg/ml DNase(I), 1 mM magnesium chloride ($MgCl_2$), and 1% v/v of Triton X-100 detergent. The lysate was clarified by centrifugation at 4 ºC and 13,000 rpm for 30 minutes in a KA-21.5 rotor (Kompsin) using a RC2-B automatic refrigerated centrifuge (Sorvall).

The inclusion body (IB) was re-suspended overnight at 4ºC in denaturing buffer containing 8 M urea, 0.095 M Na_2HPO_4, 0.005 M NaH_2PO_4, 0.5 M NaCl, 1 mM imidazole, 5 mM β-mercaptoethanol (BME) at pH 8.0 to solubilize SMN 1-4. Then the sample was clarified by centrifugation at 4ºC and 13,000 rpm for 30 minutes. The supernatant containing denature SMN 1-4 protein was purified by Ni affinity chromatography using fast flow Ni sepharose (Bioscience). The column was washed three times column volume with the denaturing buffer. The protein was eluted with elution buffer containing 8 M urea, 0.095 M Na_2HPO_4, 0.005 M NaH_2PO_4, 0.5 M NaCl, 250 mM imidazole, 5 mM BME at pH 8.0.

A purified SMN 1-4 was dialyzed against folding buffer A (Table 3.1) at 4ºC overnight using a 10 K MWCO SpectraPor 2 dialysis membrane. The SMN 1-4 proteins were concentrated using a 10 K MWCO Amicon ultrafiltration device (Millipore). The protein concentration was determined by UV/Vis spectroscopy at 280 nm using molar extinction coefficients, $\varepsilon = 28{,}480\ cm^{-1}\ M^{-1}$.

The SMN 1-4 protein was re-folded on a fast flow Ni sepharose (Bioscience) Omnifit FPLC column equipped with Pharmacia LKB FPLC (Fine Pressure Liquid Chromatography) system. The flow rate was set at 1 ml/min. The column was washed with buffer A (Table 3.1) to remove any aggregation or unbound SMN 1-4 protein. Then, the SMN 1-4 protein was folded in buffer B and eluted with buffer C (Table 3.1). The re-folded SMN 1-4 protein was dialyzed using a 10 K MWCO SpectraPor 2 dialysis membrane in dialysis buffer containing 0.095 M Na_2HPO_4, 0.005 M NaH_2PO_4, pH 8.0, 0.25 M NaCl, 1 mM BME, at 4ºC overnight. The folded SMN 1-4 protein was concentrated using a 10 K MWCO Microcon ultra-filtration device (Millipore). The protein concentration was determined by UV/Vis spectroscopy at 280 nm using molar extinction coefficients of ε= 28,480 cm^{-1} M^{-1}. The yield of pure folded SMN 1-4 protein obtained from a 1 L culture was approximately 4-6 mg. The SMN 1-4 protein was concentrated to approximately 15-20 mg/ml for crystallization.

Table 3.1. Buffers used for re-folding at pH 8.0

	Buffer A	Buffer B	Buffer C
Urea	6 M		
Sucrose		0.5 M	
Na_2HPO_4/NaH_2PO_4	0.095/0.005 M	0.095/0.005 M	0.095/0.005 M
NaCl	0.5 M	0.5 M	0.5 M
Imidazole	1 mM	1 mM	250 mM
BME	1 mM	10 mM	1 mM

Results

Crystallization

The SMN 1-4 protein was crystallized by the vapor diffusion technique using a hanging drop. Optimized protein crystallization conditions were resulted from the initial screens from Hampton search (HR1 and HR2). Briefly, droplets were prepared as

following 1 µl of the SMN 1-4 protein at 15 mg/ml were mixed with an equivalent volume of reservoir solution containing 50% v/v PEG 4,000, 100 mM Tris-HCl and 300 mM NaCl at pH 8.0. The SMN 1-4 protein crystals appeared within 3 to 4 weeks with dimensions approximately 50 to 100 microns x 20 microns x 30 microns. The SMN 1-4 protein crystallization and data collection were done by Dr. Craig L. Magee.

Data collection and analysis

Protein crystal measurements were performed at Arizona State University X-ray diffraction facility using the Rigaku R-AXIS IV^{++} detector at a wavelength of 1.54 Å. Diffraction data were indexed, refined, and integrated using crystal clear software 1.3.5 (Rigaku-MSC) and scaled using SCALA (CCP4). Data analysis Phaser 1.3.1 was used for the molecular replacement based on the SMN Tudor domain (PDB ID: 1MHN) as a search model (Selenko et al., 2001; Sprangers et al., 2003). SMN 1-4 crystals diffracted to a resolution limit of 2.7 Å with the space group $P2_1$ with cell dimensions a=55.05 Å, b=49.38 Å, c=68.23 Å along with angles α= 90°, β=120° and γ=90°. Total and unique reflections were 53,732 and 4,201, respectively. R_{merge} was 15.80% with 100% completeness. Data are summarized in Table 3.2.

Table 3.2. Summary of X-ray diffraction data and refinement of SMN 1-4

Space group	$P2_1$	
Resolution limits (Å)	2.7	
Unit Cell (Å)	a= 55.05, b=49.38, c= 68.23	
Angle (°)	α=90, β=120, γ=90	
Reflections:		
Total	53,732	
Unique	4,201	
Multiplicity	4.32	(4.03)
I/σ (I)	6.4	(3.5)
R_{Merge}(%)	15.80	(35.40)
Completeness (%)	100	(100)
Refinement		
R_{work} (%)/R_{free} (%)	27.96 /30.91	
Number in asymmetric unit	1	
RMSD Bond length (Å)	0.013	
RMSD Bond angle (°)	2.402	
Ramachandran plots		
Preferred (%)	78.60	
Allowed (%)	16.10	
Outlier (%)	5.40	

R-factor =Σ| $|F_{obs}|$ -$|F_{cal}||/ \Sigma |F_{obs}|$, Free R-factor was calculated with setting aside 5% of the reflection data. $R_{merge} = \Sigma_{hkl}\Sigma_j |I_j(hkl) – [I(hkl)] | \Sigma_{hkl}\Sigma_j[I(hkl)]$ where $I_j(hkl)$is the intensity reflection, and I(hkl)is the mean intensity of symmetry related h, k, l. Number in the parenthesis is the outer shell value.

Refinement

The SMN 1-4 structure was manually built using COOT (Crystallographic Object-Oriented Toolkit) (Emsley and Cowtant, 2004) and refined using PHENIX (Adam et al., 2010). The residues 1-88 and 148-209 of SMN 1-4, which are not part of the Tudor domain (residues 89-147), have been built in SMN 1-4.

The completed SMN 1-4 protein structure had R_{work} and R_{free} values of 27.96% and 30.91%, respectively (Table 3.2). Ramachandran plot values of the structure were 78.60%, 16.10% and 5.40% in the preferred, allowed and outlier, respectively. Root mean square deviation (RMSD) bond length and bond angle were 0.013 Å and 2.402º,

respectively. Procheck and Sfcheck were also used to verify the protein structure with the diffraction data and structural model (Laskowski et al., 1993 and Vaguine et al., 1999).

Three-dimensional structure of SMN 1-4

The SMN 1-4 structure consists of two subunits, A (green) and B (yellow), in the asymmetric unit (Figure 3.2). The A and B subunits are approximately related by a 2-fold symmetry axis. Loop regions of each subunit of SMN 1-4 expose the outer part of the structure (Figure 3.2). These loop regions of the two subunits are exposed to the surface area of the structure whereas the β-strands of the two subunits are facing each other.

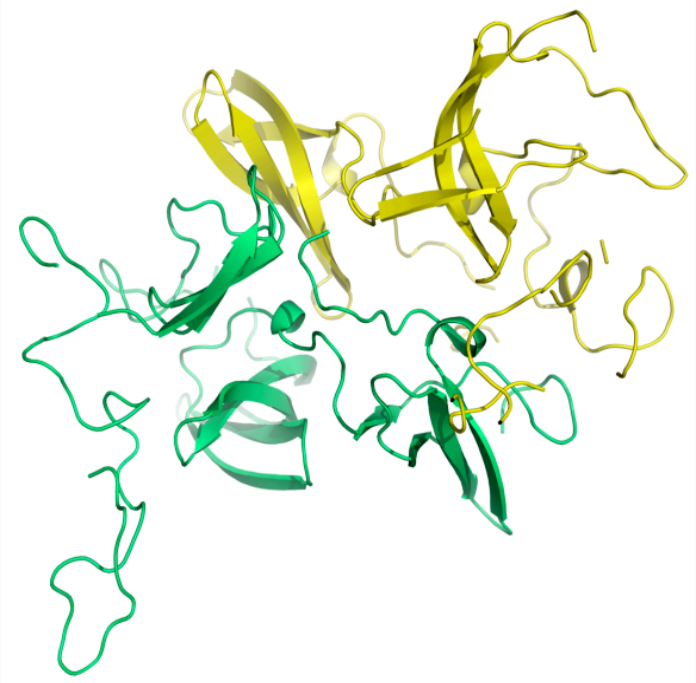

Figure 3.2. Three-dimensional structure of the SMN 1-4 protein showing two subunits, identified as the A (lime green) and B (yellow) subunits.

Each subunit consists of five encoded exons. The encoded exon, 1, 2a, 2b, 3, and 4 is color-coded in yellow, orange, green, red and purple, respectively (Figure 3.3).

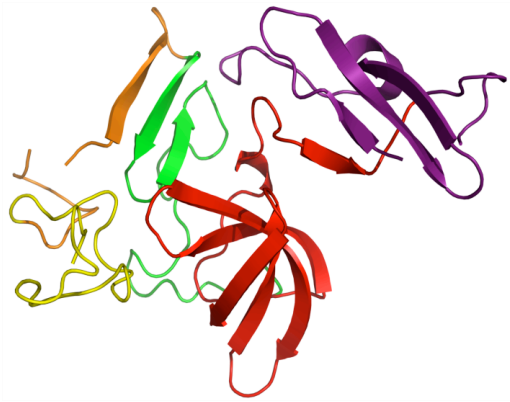

Figure 3.3. Subunit A color-coded according to the encoded exon: exon 1, 2a, 2b, 3, and 4 are in yellow, orange, green, red and purple, respectively.

Each subunit consists of three domains, A_1, A_2, and A_3 with residue ranges 41-67, 89-147, and 151-201 respectively (Table 3.3). A_1 consists of three β-strands whereas A_2 and A_3 both consist of four long bent β-strands and a short helix (Figure 3.4). All domains have similar β-strand folding. The three β-strands of domain A_1 form a β-sheet with a similar fold as domain A_2.

Table 3.3. The organization of the residue ranges of each domain

Domain	Residue range	Secondary Structure
A_1	41-67	Three antiparallel β-sheets
A_2	89-147	Four antiparallel β-sheets and a short α-helix
A_3	151-201	Two long and two short antiparallel β-sheets and a short α-helix

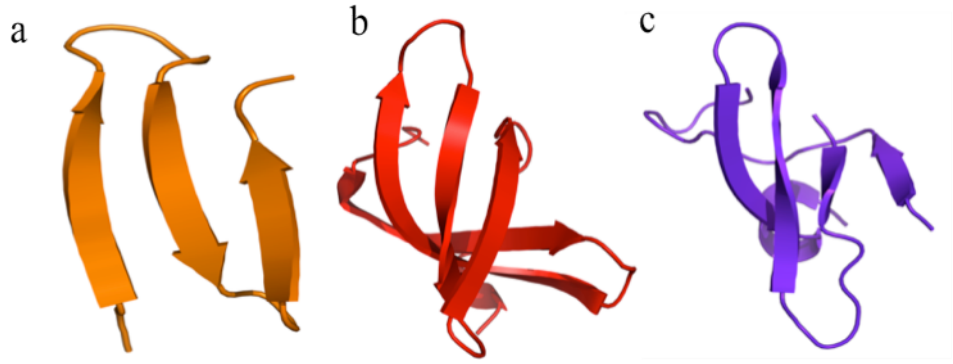

Figure 3.4. Domains of subunit A of SMN 1-4. Each subunit consists of three domains, which is the A subunit are termed A_1 (orange), A_2 (red), and A_3 (purple) and shown separately for comparison. The three domains all have Sm-like folds: a) The A_1 domain contains three anti-parallel β-strands, b) The A_2 domain (the Tudor Domain) consists of four anti-parallel β-strands and a short helix, and c) The A_3 domain contains four anti-parallel β-strands and one short α-helix.

Discussion

Comparison of SMN Tudor domains

Subunits A_2 and B_2 (Figure 3.5a,b) have been determined independently from the SMN Tudor domain (Figure 3.5c) that had been solved by Selenko et al., 2001 and Sprangers et al., 2003. All three domains have similar folding and consist of four β-strands. However, a small difference from all three Tudor domains is that domain B_2 has short β-strands which is caused by the disorder in the B subunit of the SMN 1-4 due to the crystallographic contacts.

The three domains are very similar as expected, largely consisting of four β-strands. Small differences are found in the length of the β-strands, which presumably reflect the presence of the flanking protein regions in the SMN 1-4 structure and a disorder in this region of the B chain due to crystallographic contacts.

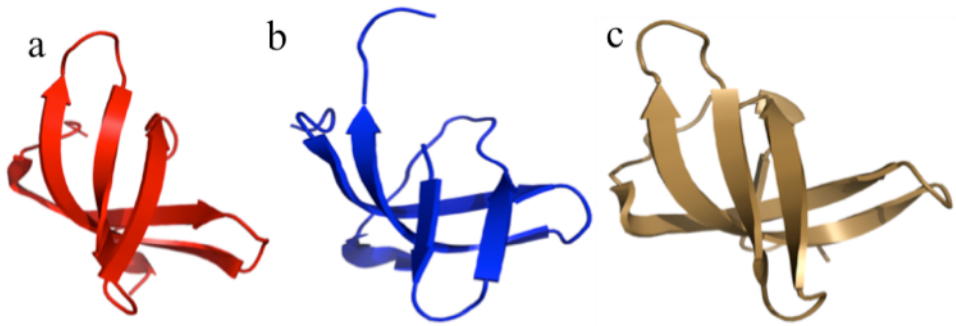

Figure 3.5. Comparison of the Tudor domains in: a). Subunit A_2 (red), b). B_2 (blue) (solved by this work), and as solved as an independent construction in c). (sand) by Sprangers and colleagues (2003).

Comparison of SMN Tudor domain to other proteins

The SMN protein is known to express ubiquitously in the cell and has an essential role in RNA processing. The SMN complex assists in the formation of the seven-core Sm protein in the cytoplasm (Zhang et al., 2011). Subunit A_2 of SMN 1-4 has a similar fold as the Sm protein except for the presence of an α-helix at the N-terminus or C-terminus of the Sm protein (Figure 3.6). The A_2 domain consists of four long β-strands and a short α-helix at the C-terminal region whereas Sm D_1 (Figure 3.6b) has one long β-strand and five short β-strands and a helix-turn-helix at the C-terminal region. Even though the number of β-strands in Sm D_1 is more than in the A_2 domain, both SMN and Sm D_1 shown to have the same fold, which is bent like a shell. The Sm D_2 (Figure 3.6c) also has similar fold as the A_2 domain except the long α-helix at the N-terminal region found in Sm D_2 whereas the A_2 domain has only a short α-helix at the C-terminal region.

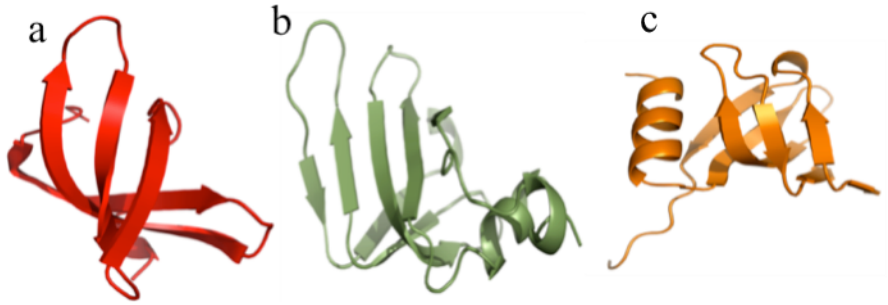

Figure 3.6. The comparison of a). SMN Tudor domain (red) to Sm structures b). SmD_1 (green) and c. SmD_2 (orange) (Kambach et al., 1999).

The SMN protein has interactions with the Gemin proteins (Gubitz et al., 2004). Gemin6 and Gemin7 are largely β-proteins with Sm-like folds consisting of several β strands. In addition to the central β-sheet arrangement, the proteins have one or two helices near their N or C terminal regions. The SMN A_2 domain (Figure 3.7a) has a similar β-strands fold to Gemin6 and Gemin7 (Figure 3.7b,c). However, the A_2 domain has a short α-helix at the C-terminus whereas Gemin6 has a helix at the C-terminus and a helix-turn-helix at the N-terminus and Gemin7 has a long α-helix at the N-terminal region. Gemin6 consists of five β-strands and Gemin7 has three long and two short β-strands.

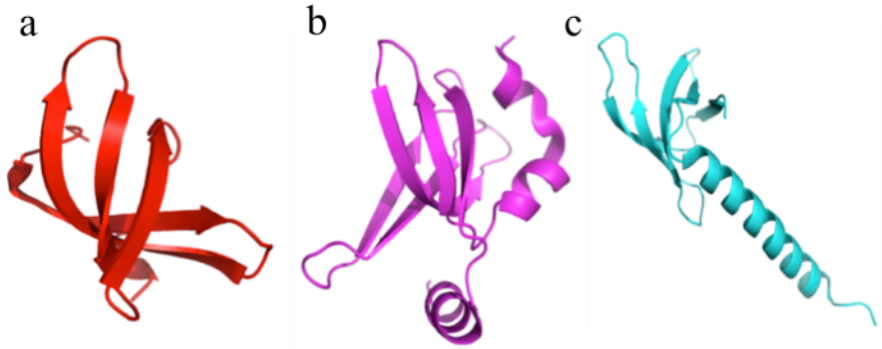

Figure 3.7. Comparison of the three domains of SMN to several other protein components shows a significant structural homology. Shown are the structures of : a) The SMN Tudor domain (red) (this work) and b) Gemin6 (magenta), and c) Gemin7 (cyan), (Ma et al., 2005).

In SMA patients, a number of missense mutations are found in exons 1 thru 4 (Sun et al., 2005). The mutation sites are not localized on one side of the structure, but spread throughout the structure of the SMN 1-4 (Figure 3.8). In some cases, the mutations occur for residues found on the surface of the proteins, for example Asp30, Asp44, and Gly95, and presumably the mutations interfere with binding to other proteins. Mutation of Gly95 and Ala111 result in reduce of binding of SMN to Sm protein and mutation of Asp30 and Asp 44 mutations perhaps affect the self oligomerization that lead to the loss of the binding ability Gemin2 (Liu et al., 1997; Sun et al., 2005). Gemin2 has strong binding with the SMN protein and is known to play an essential role in snRNPs assembly through the stabilization of the SMN complex via self-oligomerization (Otter et.al, 2007; Ogawa et al., 2007). For the A2G mutation, this amino acid residue is part of a loop that interacts with other regions and the mutation may destabilize this region of the protein.

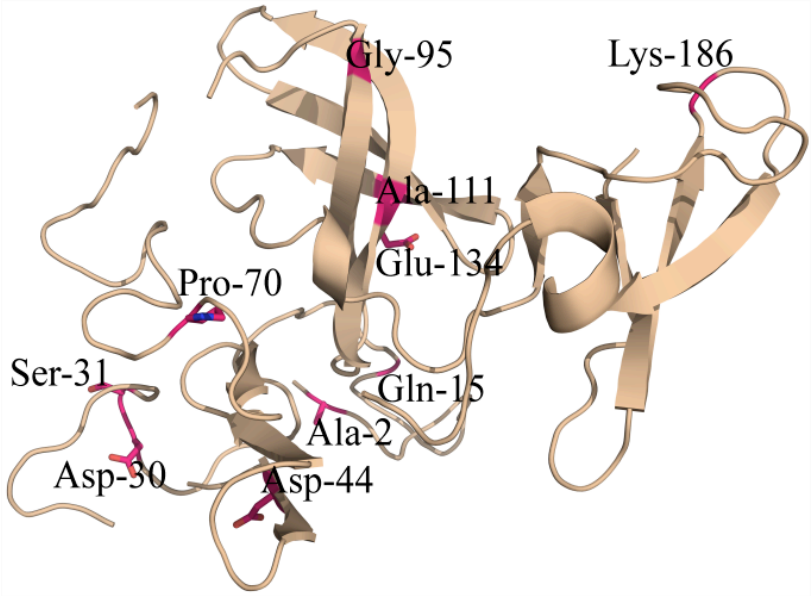

Figure 3.8. Missense mutations have been found in exons 1-4 in SMA patients (Sun et al., 2005).

The three-dimensional structure of the SMN 1-4 has been solved, but to gain an understanding at the molecular level of what leads to SMA requires the determination of the structures of SMN-WT and SMN-Δ7. The SMN 1-4 is 2/3 of the SMN-Δ7 and the final refinement of the SMN 1-4 model was sufficient for molecular replacement to solve the three-dimensional structure of SMN-Δ7 and SMN-WT. It will be described in Chapter 4.

References

Adams, P.D., Afonine, P.D., Bunkóczi, G., Chen, V.B., Davis, I.W., Echoo ls, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R.; Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger T.C., Zwart, P.H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta. Cryst. *D66,* 213–221.

Alias, L., Bernal, S., Fuentes-Prior, P., Barcelo, M. J., Also, E., Martinez-Hernandez, R., Rodrigue-Alvarez, R., Martin, Y., Aller, E., Grau, E., Pecina, A., Antinolo, G., Galan, E., Rosa, A. L., Fernandez-Burriel, M., Borrego, S., Millan, J. M., Hernandez-Chico, C., Baiget, M., Tizzano, E. F. (2009). Mutation update of spinal muscular atrophy in Spain: molecular characterization of 745 unrelated patients and identification of four mutations in the *SMN1* gene. Hum. Genet. *125,* 29–39.

Buhler, D., Raker, V., Luhrmann, R., 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.

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

Clermont, O., Burlet, P., Cruad, C., Bertrandy, S., Milki, J., Munnich, A., Lefvbre, S. (1997). Mutation analysis of the SMN gene in undeleted SMA patients. *Am. J.* Hum. Genet*. 61,* A329

Collaborative computational project, Number 4 (CCP4) (1994). "The CCP4 Suite: Programs for Protein Crystallography". Acta Cryst. *D50,* 760–763.

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

Gubitz, A. K., Feng, W., Dreyfuss, G. (2004). The SMN complex. Expr. Cell Res. *296,* 51–56.

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.

Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., Le Paslier, D., Frezal, J., Cohen, D., Weissenbach, J., Munnich, A., Melki, J. (1995). Identification and characterization of a spinal muscular atrophy-determining gene. Cell *80,* 155–165.

Liu, Q., Fischer, U., Wang, F., Dreyfuss, G. (1997). The spinal muscular atrophy disease gene product, SMN, and its associated protein SIPI are in a complex with spliceosomal snRNP proteins. Cell *90,* 1013–1021.

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

Kambach, C., Walke, S., Young, R., Avis, J. M., de la Fortelle, E., Raker, V.A., Luhrmann, R., Li, J., and Nagai, K. (1999). Crystal structures of two Sm protein complexes and their implications for the assemblyof the spliceosomal snRNPs. Cell 96, 375–387.

Ma, Y., Dostie, J., Dreyfuss, G., Van Duyne, G. D. (2005). The Gemin6-Gemin7 heterodimer from the survival of motor neurons complex has an Sm protein-like structure. Struct. *13,* 883–892.

Magee, L.C. (2006). Survival of motor neuron protein: structural studies of the SMN protein, a protein link to spinal muscular atrophy. Ph.D. Thesis, Arizona State University, Tempe, AZ.

Ogawa, C., Usui, K., Aoki, M., Ito, F., Itoh, M., Kai, C., Kanamori-katayama, M., Hayashizaki, Y., Suzuki, H. (2007). Gemin2 plays an important role in stabilizing the survival of motor neuron complex. J. Biol. Chem. *282,* 11122–11134.

Otter, S., Grimmler, M., Neuenkirchen, N., Chari, A., Sickmann, A., Fischer, U. (2007). A comprehensive interaction map of the human survival of motor neuron (SMN) complex. J. Biol. Chem. *282,* 5825–5833.

Parsons, D.W., McAndrew, P.E., Iannaccone, S.T., Mendell, R.J., Burghes, A.H. M., and Prior, T.W. (1998). Intragenic telSMN mutations: Frequency, distribution, evidence of a founder effect, and modification of the spinal muscular atrophy phenotype by cenSMN copy number. Am. J. Hum. Genet. *63,* 1712–1723.

Selenko, P., Sprangers, R., Stier, G., Buhler, D., Fischer, U., Sattler, M. (2001). SMN tudor domain structure and its interaction with the Sm proteins. Nature *8*, 27–31.

Sprangers, R., Groes, M.R., Sinning, I., and Sattler, M. (2003). High-resolution X-ray and NMR structure of the SMN Tudor domain: Conformational variation in the binding site for symmetrically dimethylated arginine residues. J. Mol. Biol. *327,* 507–520.

Sun, Y., Grimmler, M., Schwarzer, V., Schoenen, F., Fischer, U., and Wirth, B. (2005). Molecular and function analysis of intragenic *SMN1* mutations in patients with spinal muscular atrophy. Hum. Mutation *25,* 64–71.

Vaguine, A.A., Richelle, J., 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. Cryst. *D55,* 191–205.

Wirth, B., Herz, M., Wetter, A., Moskau, S., Hahnen, E., Rudnik-Schoneborn, S., Wienker, T., and Zerres, K. (1999). Quantitative analysis of survival motor neuron copies: Identification of subtle *SMN1* gene mutations in patients with spinal muscular atrophy, genotype-phenotype correlation, and implications for genetic counseling. Am. J. Hum. Genet. *64,* 1340–1356.

Wirth, B. (2000). An update of the mutation spectrum of the survival motor neuron gene (*SMN1)* in autosomal recessive spinal muscular atrophy (SMA). Hum. Mutation. *15,* 228–237.

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. Human Mol. Genet. *9,* 2869–2877.

Zhang, R., So, B.R., Li, P., Yong, J., Glisovic, T., Wan, L., Dreyfuss, G. (2011). Structure of a key intermediate of the SMN complex reveals Gemin2's crucial function in snRNP assembly. Cell *146,* 384–395

Chapter 4

THREE-DIMENSIONAL STRUCTURES OF SURVIVAL MOTOR NEURON (SMN) PROTEINS: SMN-WT AND SMN-Δ7

Introduction

The *SMN1* and the *SMN2* genes consist of nine exons (1, 2a, 2b, 3, 4, 5, 6, 7, 8) and both genes are almost identical, except for five different nucleotides, which are located from intron 6 to exon 8 (Figure 4.1). A single nucleotide change from C to T in exon 7, between the *SMN1* gene and the *SMN2* gene causes the alternate splicing in exon 7 in the *SMN2* gene forming a truncated phenotype called SMN-Δ7 (Lorson et al., 1998; Lorson and Androphy, 2000). SMA patients have predominantly resulting the SMN-Δ7 form of protein rather than the full-length form, SMN-WT.

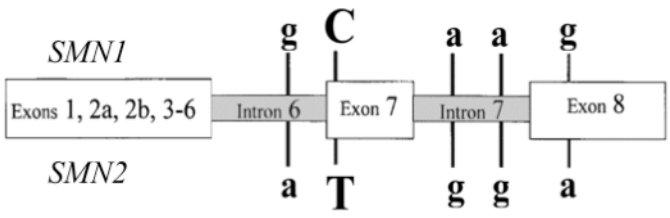

Figure 4.1. Localization of the nucleotides by which *SMN1* can be distinguished from *SMN2* (Wirth, 2000).

The SMN-WT protein consists of 294 amino acid residues and SMN-Δ7 consists of 282 amino acid residues (Figure 4.2). Both SMN-WT and SMN-Δ7 have been crystallized and the structures have been solved. The three-dimensional structure of SMN 1-4, described in Chapter 3, has been used as a search model to solve the structure of both SMN-WT and SMN-Δ7. The residues, 205 to 294 of SMN-WT and 205 to 282 of SMN-

Δ7, which are not part of the SMN 1-4 model, have been built in SMN-WT and SMN-Δ7, respectively. The completed three-dimensional structure of the SMN protein assists in understanding the cause of SMA and role of SMN in the cell.

SMN-WT

```
1    MRGSHHHHHH GMASMAMSSG GSGGGVPEQE DSVLFRRGTG QSDDSDIWDD 50
51   TALIKAYDKA VASFKHALKN GDICETSGKP KTTPKRKPAK KNKSQKKNTA 100
101  ASLQQWKVGD KCSAIWSEDG CIYPATIASI DFKRETCVVV YTGYGNREEQ 150
151  NLSDLLSPIC EVANNIEQNA QENENESQVS TDESENSRSP GNKSDNIKPK 200
201  SAPWNSFLPP PPPMPGPRLG PGKPGLKFNG PPPPPPPPPP HLLSCWLPPF 250
251  PSGPPIIPPP PPICPDSLDD ADALGSMLIS WYMSGYHTGY YMGFRQNQKE 300
301  GRCSHSLN                                              308
```

SMN-Δ7

```
1    MRGSHHHHHH GMASMAMSSG GSGGGVPEQE DSVLFRRGTG QSDDSDIWDD 50
51   TALIKAYDKA VASFKHALKN GDICETSGKP KTTPKRKPAK KNKSQKKNTA 100
101  ASLQQWKVGD KCSAIWSEDG CIYPATIASI DFKRETCVVV YTGYGNREEQ 150
151  NLSDLLSPIC EVANNIEQNA QENENESQVS TDESENSRSP GNKSDNIKPK 200
201  SAPWNSFLPP PPPMPGPRLG PGKPGLKFNG PPPPPPPPPP HLLSCWLPPF 250
251  PSGPPIIPPP PPICPDSLDD ADALGSMLIS WYMSGYHTGY YMEMLA     296
```

Figure 4.2. SMN-WT and SMN-Δ7 sequences. The SMN-WT sequence consists of a fusion tag (Blue) at the N-terminus and full length SMN and exon 7 shown in green. The SMN-Δ7 sequence consists of a fusion tag at the N-terminus, SMN lacking the region encoded by exon 7 and four amino acid residues from exon 8 shown in red.

My contributions to this chapter are the expression, purification, folding, crystallization, and solving structure of SMN-Δ7 using molecular replacement from the three-dimensional structure of SMN 1-4. Once the three-dimensional structure of SMN-Δ7 was completed, its structure was used for molecular replacement to solve the three-

dimensional structure of SMN-WT, using the crystallographic data from Dr. Craig Magee.

Materials and Methods

SMN-WT and SMN-Δ7 protein expression and purification

SMN-WT and SMN-Δ7 gene was cloned into an *E.coli* pRSET BL21 (DE3) over-expression system with a His_6 fusion tag at the N-terminal region. The SMN-WT and SMN-Δ7 cell culture, cell lysis, purification and folding was followed the SMN 1-4 procedure as describe in Chapter 3.

Size exclusion chromatography (SEC)

The folded SMN-WT and SMN-Δ7 in a 0.1 M Tris-HCl pH 8.0, 250 mM NaCl was purified using a gel filtration chromatography Speharose CL-6B (GE healthcare). The purified SMN-WT was concentrated using a 10 K microcon ultra-filtration device (Millipore). The SMN-WT and SMN-Δ7 protein concentration was determined by UV/Vis spectroscopy at 280 nm using molar extinction coefficients, $\varepsilon = 45{,}340\ cm^{-1}\ M^{-1}$ and $\varepsilon = 45{,}220\ cm^{-1}\ M^{-1}$, respectively. The yield of purified folded SMN-WT and SMN-Δ7 protein obtained from a 1 L culture was approximately 2-3 mg and 4-6 mg, respectively.

Results and Discussion

Protein crystallization

The hanging drop vapor diffusion method was used to crystallize SMN-WT. The SMN-WT crystallization initially was screened with Hampton Research grid screen 1 and grid screen 2 (HR1 and HR2), and the SMN 1-4 crystallization conditions. The initial conditions with SMN 1-4 crystallization conditions was not yield any crystals; however,

SMN-WT crystals were obtained using condition #25 of the Hampton Research grid screen 2 (HR2). The SMN-WT protein was concentrated to approximately 30 mg/ml in 0.1 M Tris-HCl pH 8.0, 250 mM NaCl. The protein crystallization drop with a volume 1 µl was combined with an equivalent volume of reservoir solution containing 1.8 M ammonium sulfate, 0.1 M MES pH 6.5 and 0.01 M cobalt chloride.

The SMN-Δ7 protein was crystallized using the hanging drop vapor diffusion technique. The initial conditions were based upon the SMN-WT crystallization. The initial conditions yielded micro crystals and after optimizing the conditions, larger crystals were produced. The SMN-Δ7 protein was concentrated to approximately 36 mg/ml in 0.1 M Tris-HCl pH 8.0, 250 mM NaCl. The protein crystallization drop with a volume 2 µl was combined with an 1.8 µl of reservoir solution containing 0.1 M Tris-HCl pH 9.0, 300 mM NaCl, 40% v/v PEG 4,000 and 0.2 µl of 10% w/v 1,2,3-heptanetriol. Small hexagonal crystals appeared within 3-5 days and grew to full size 50 µm x 70 µm x 25µm within four weeks (Figure 4.3).

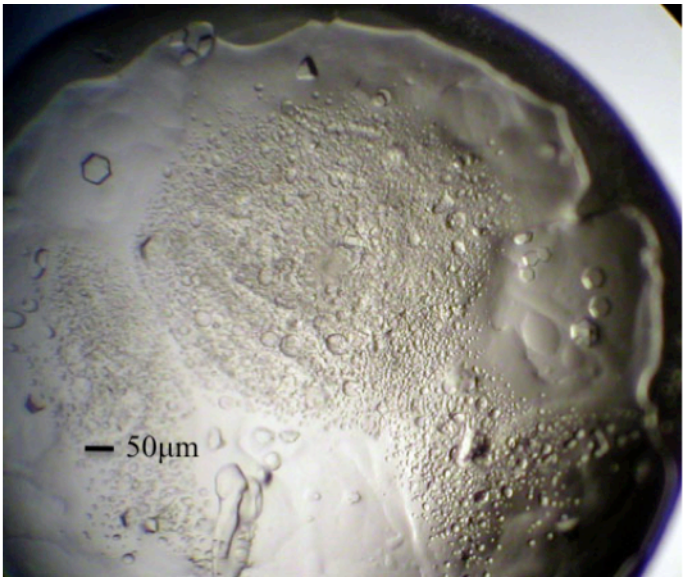

Figure 4.3 Image of SMN-Δ7 protein crystals.

Data collection

The SMN-WT crystals were frozen in liquid nitrogen and sent to the Advanced Light Source (ALS) beamline 5.0.1 at the Lawrence Berkeley National Laboratory. The SMN-WT crystal data collection was performed by Meitian Wang at ALS and found to diffract to a resolution limit of 5.5 Å. The SMN-WT diffraction data was indexed, refined, and integrated using software package CrystalClear 1.3.5 (Rigaku-MSC) and scaled using SCALA (CCP4). The SMN-WT crystals belonged to the space group $C2_1$ with cell dimensions a = 137.04 Å, b=169.82 Å, c=108.78 Å along with angles α=90°, β=128.54°, γ=90°. The SMN-WT data contained total and unique reflection of 24,090 and 7,946, respectively. Data contained I/σ (I) and R_{merge} value of 4.6 and 14.5 %, respectively, with 61.3% completeness.

The SMN-Δ7 crystals were frozen in liquid nitrogen and sent to the Brookhaven National Laboratory (BNL). The SMN-Δ7 crystal data collection was performed at the BNL on the NSLS-X126 beamline using an ADSC detector at a wavelength of 1.081 Å by Chad Simmons. A full diffraction data set was measured to a resolution limit of 3.0 Å (Figure 4.4). The SMN-Δ7 diffraction data was indexed, refined, and integrated using MOSFLM (Leslie, 1999) and scaled using SCALA (CCP4 1994). The SMN-Δ7 crystal belonged to the space group $C2_1$ with unit cell dimensions a = 107.10 Å, b=62.29 Å c=57.07 Å along with an angles α=90°, β=95.07°, γ=90°. The SMN-Δ7 data contained total and unique reflection of 22,532 and 7,286, respectively. Diffraction data contained I/σ (I) and R_{merge} value of 4.1 and 35.2 %, respectively, with 96.7% completeness.

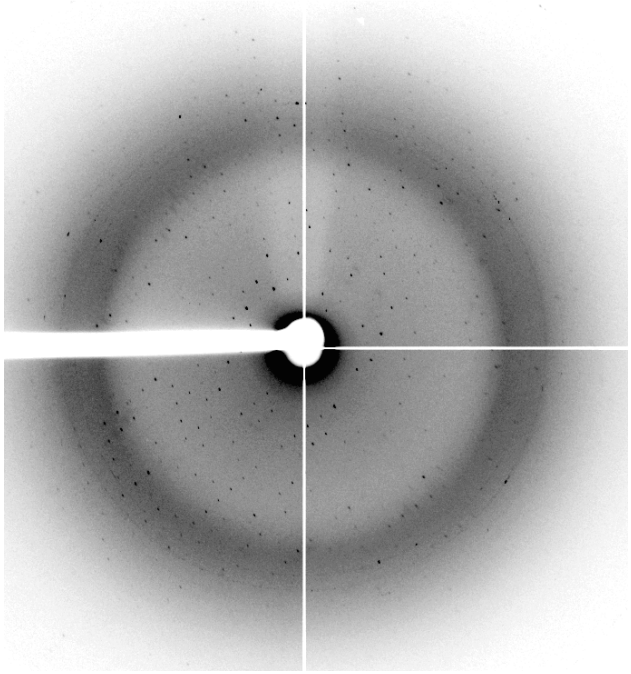

Figure 4.4. Image of X-ray diffraction of SMN-Δ7.

Data analysis

The SMN-14 structure (Chapter 3) was used as a search model for the SMN-Δ7 structure determination since the SMN 1-4 structure is 2/3 of the SMN-Δ7. The initial phases were determined using molecular replacement method using PHENIX (Adams et al., 2010). After processing, an electron density map was generated and the SMN-Δ7 structure resulted one subunit in asymmetric unit. The SMN-Δ7 structure was manually built using COOT based upon the electron density map with alternative rounds of refinement using PHENIX (Adams et al., 2010).

Due to the limitation of the SMN-WT resolution limit, 5.5 Å, SMN-Δ7 was built first. Then the complete SMN-Δ7 structure was used as a search model for the molecular replacement using Phaser (McCoy et al., 2007) for SMN-WT structure. The data processing shown that the SMN-WT obtained three subunits in asymmetric unit and electron density map generated for the model building. The SMN-WT structure was

manually built using COOT based upon the electron density map with alternative refinement using PHENIX (Adams et al., 2010).

Refinement

Procheck and Sfcheck programs were also used to verify the protein structures with the diffraction data and structural model (Laskowski et al., 1993 and Vaguine et al., 1999). The complete model of SMN-Δ7 agree with the electron density map as shown by the R_{work} and R_{free} values of 38.90 % and 42.08 %, respectively. The Ramachandran plots of SMN-Δ7 model had 97.60 % and 16.10 % in the preferred region and allowed region, respectively, with no outlier. The root mean squared deviations of bond distances and angles are 0.004 Å and 1.102°, respectively.

The complete model of SMN-WT has R_{work} and R_{free} values of 30.68% and 32.89%, respectively. Due to the limited resolution, only the backbone of SMN-WT has been reliably built in the electron density.

Pymol has been used for structural model viewing and the creation of most of the figures (DeLano, 2002). The summary of results of SMN-Δ7 and SMN-WT are shown in Table 4.1.

Table 4.1. Summary of X-ray diffraction data and refinement of SMN-WT and SMN-Δ7

	SMN-WT	SMN-Δ7
Space group	$C2_1$	$C2_1$
Resolution limits (Å)	5.5	3.0
Unit Cell (Å)	a = 137.04, b=169.82, c=108.78	a = 107.10, b=62.29, c=57.07
Angle (°)	α=90, β=128.54, γ=90	α=90, β=95.07, γ=90
Reflections:		
Total	24,090	22,532
Unique	7,946	7,286
Multiplicity	3.0	3.1 (3.0)
I/σ (I)	4.6	4.1 (2.1)
R_{Merge} (%)	14.5	35.2 (0.54)
Completeness (%)	61.3	96.7 (98.30)
Refinement		
R_{work} (%)/R_{free} (%)	30.68 /32.89	38.90/42.08
Number in asymmetric unit	3	1
RMSD Bond length (Å)	NA	0.004
RMSD Bond angle (°)	NA	1.102
Ramachandran plots		
Preferred (%)	NA	97.60
Allowed (%)	NA	2.30
Outlier (%)	NA	0.00

R-factor =Σ| $|F_{obs}|$ -$|F_{cal}||/ \Sigma |F_{obs}|$, Free R-factor was calculated with setting aside 5% of the reflection data. $R_{merge} = \Sigma_{hkl}\Sigma_j |I_j(hkl) – [I(hkl)] | \Sigma_{hkl}\Sigma_j[I(hkl)]$ where $I_j(hkl)$is the intensity reflection, and I(hkl)is the mean intensity of symmetry related h, k, l. Number in the parenthesis is the outer shell value.

Three-dimensional structure of SMN-Δ7

The three-dimensional structure of SMN-Δ7 forms a monomer with 282 amino acid residues and folded into β-strands and α-helices (Figure 4.5). All the β-strands form an antiparallel β-sheet and β-folded. The α-helices are found in both the C-terminal and N-terminal region. The N-terminal α-helices, α_1 and α_2, are located after three of antiparallel β-strands, β_1-β_3, and the other two α-helices, α_3 and α_4, are located at the end of the C-terminal region. The secondary structure and domain arrangements of the SMN-Δ7 are summarized in Table 4.2.

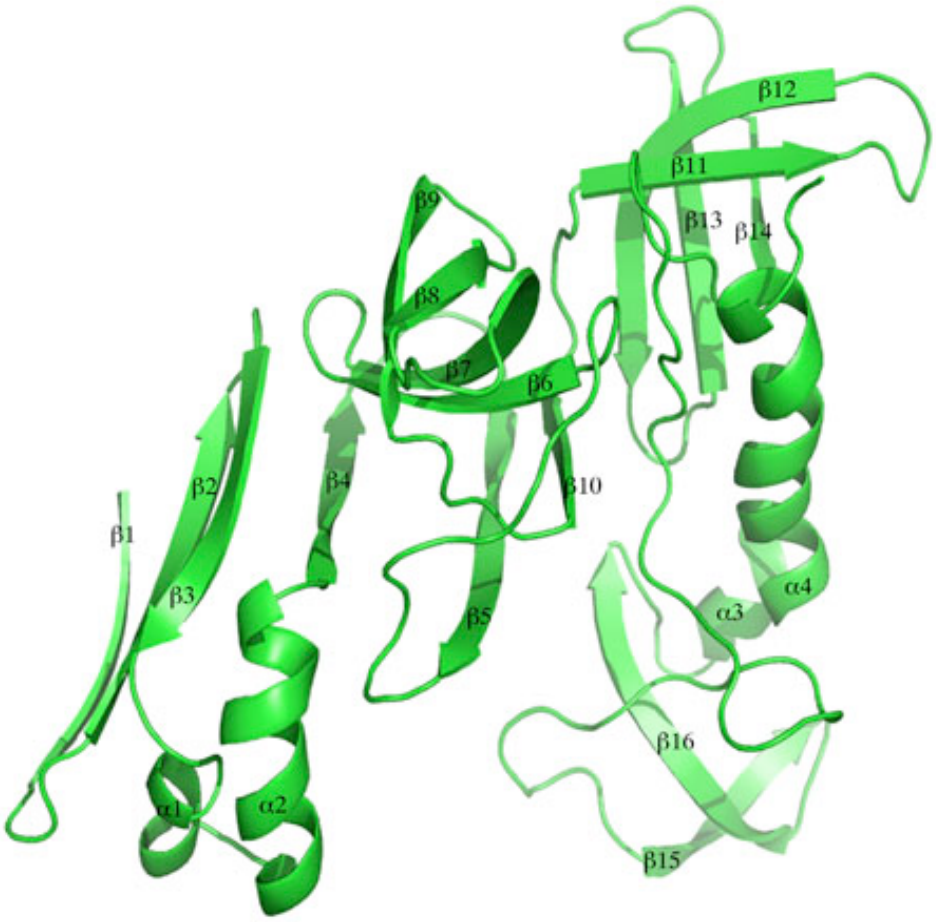

Figure 4.5. Three-dimensional structure of SMN-Δ7 consists of 16 β-strands and 4 α-helices.

The three-dimensional structure of SMN-Δ7 consists of four domains (Figure 4.6). Domain 1 consists of 5 antiparallel β-strands, β_1-β_5 (Figure 4.6a) and two α-helices, α_1 and α_2, whereas domain 2 (Tudor domain) consists of 5 antiparallel β-strands, β_6-β_{10} (Figure 4.5b). Domain 3 consists of 4 antiparallel β-strands, β_{11}-β_{14} (Figure 4.6c). Domain 4 consists of two bent antiparallel β-strands, β_{15}-β_{16} and one long, α_3, and short α-helix, α_4 (Figure 4.6d). Domain 2 and 3 are shown to have similar fold and it is shown as the mirror image of each other (Figure 4.5). Domain 1, 2, and 3 mostly form β-strands whereas domain 4 consists of two long β-strands, and two α-helices.

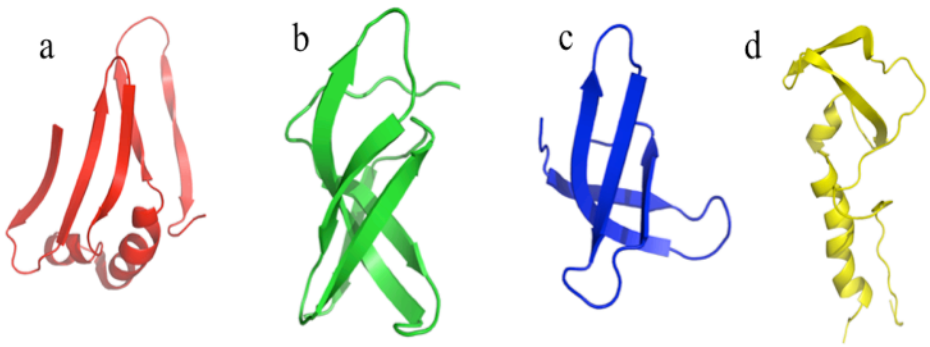

Figure 4.6. Domains of SMN-Δ7. a) Domain 1 shown in red. b) Domain 2 shown in green. c) Domain 3 shown in blue and d) Domain 4 shown in yellow.

Table 4.2 Summary of secondary structure and domain arrangement of SMN-Δ7

Residues	Secondary Structure	Domain
8-14	β_1	1
19-26	β_2	1
29-36	β_3	1
40-45	α_1	1
48-62	α_2	1
64-70	β_4	1
75-83	β_5	1
97-102	β_6	2 (Tudor domain)
106-117	β_7	2 (Tudor domain)
122-127	β_8	2 (Tudor domain)
131-137	β_9	2 (Tudor domain)
142-148	β_{10}	2 (Tudor domain)
153-158	β_{11}	3
164-174	β_{12}	3
178-184	β_{13}	3
190-194	β_{14}	3
230-237	β_{15}	4
240-250	β_{16}	4
255-260	α_3	4
263-278	α_4	4

Three-dimensional structure of SMN-WT

The three-dimensional structure of SMN-WT consists of three identical proteins with each protein containing 294 amino acid residues. The three proteins A, B and C are

shown in green, blue and red respectively (Figure 4.7). Protein A and C forms as a mirror image to each other and protein B locates in the middle approximate 1/3 of protein A and C. The α-helices of the protein A and C are exposed to the surface region whereas the α-helices of protein B are buried within the two proteins, A and C.

The α-helices of the protein A and C are exposed to the surface region whereas the α-helices of protein B are buried within the two proteins, A and C. All the β-strands form an antiparallel β-sheet and β-folded and buried inside the structure whereas the α-helices are exposed to the surface of the structure.

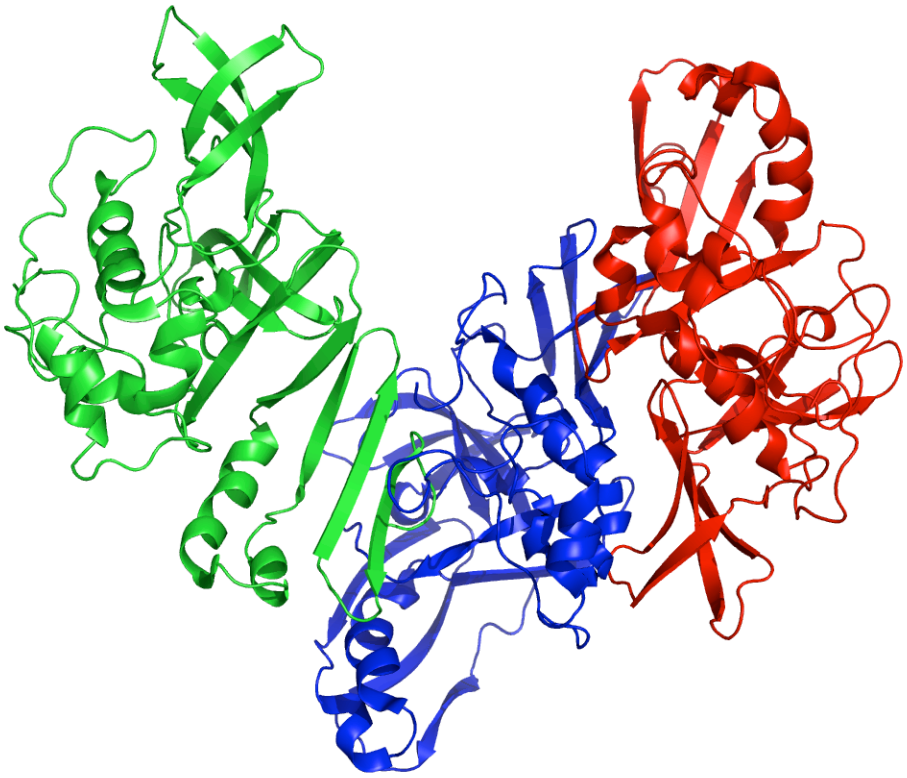

Figure 4.7. The three-dimensional structure of SMN-WT. The SMN-WT shows three identical subunits A, B and C, which are shown in green, blue and red, respectively.

The α-helices are positioned at the N-terminal and C-terminal region of the structure (Figure 4.8). The N-terminal α-helices, α_1 and α_2, are located after three of antiparallel β-strands, β_1, β_2, and β_3, and the other two α-helices, α_3 and α_4, and helix-turn-helix are located at the end of the C-terminal region.

Figure 4.8. The monomer of SMN-WT.

Each protein consists of four domains containing β-strands and several α-helices. Domain 1, β_1-β_5 and α_1-α_2, forms four antiparallel β-strands and a short α-helix (Figure 4.9a).. Domain 2 forms four antiparallel β-strands, β_6-β_9, and a short turn (Figure 4.9b). Domain 3 forms a short turn and four antiparallel β-strands, β_{10}-β_{12} (Figure 4.9c) and

domain 4 forms short α-helix, α_3, helix-turn-helix and long α-helix, α_4 (Figure 4.9d). The secondary structure and domain arrangements of the SMN-WT are summarized in Table 4.3.

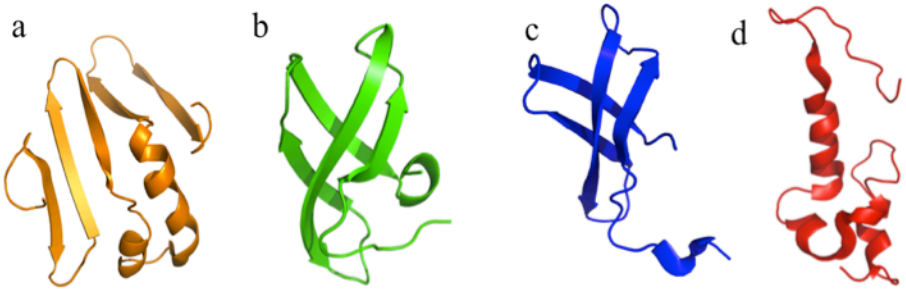

Figure 4.9. Comparison of SMN-WT domains. a. Domain 1, b. domain 2, c. domain 3 and d. domain 4 are shown in orange, green, blue and red, respectively.

Table 4.3 Summary of secondary structure and domain arrangement of SMN-WT

Residues	Secondary Structure	Domain
8-14	β_1	1
19-26	β_2	1
29-36	β_3	1
40-45	α_1	1
48-62	α_2	1
64-70	β_4	1
75-83	β_5	1
97-102	β_6	2 (Tudor domain)
106-117	β_7	2 (Tudor domain)
122-127	β_8	2 (Tudor domain)
131-136	β_9	2 (Tudor domain)
138-141	Short turn	2 (Tudor domain)
145-147	Short turn	3
153-158	β_{10}	3
164-173	β_{11}	3
178-184	β_{12}	3
188-194	β_{13}	3
233-240	α_3	4
250-260	Helix-turn-helix	4
263-278	α_4	4

Significant difference of SMN-WT and SMN-Δ7 is that the SMN-Δ7 form single subunit and SMN-WT forms a trimer in asymmetric unit. Besides, the SMN-Δ7 has the two β-stands, β_{15} and β_{16}, the short loop at the C-terminus of SMN-Δ7 (Figure 4.10a). The three-dimensional structure of the SMN-WT protein has the same general fold as the SMN-Δ7, except the short turn in domain 2 and domain 3 shown in Figure 4.9b and 4.9c, respectively, helix-turn-helix and the long extended loop at the C-terminal region (residues 278-294) (Figure 4.10b). These differences may possibly cause the defect in oligomerization (Buhler et al., 1999; Burghes and Beattie, 2009; Martin et al., 2012) and disrupt the assembly of Sm proteins, which lead to reduce the snRNP assembly.

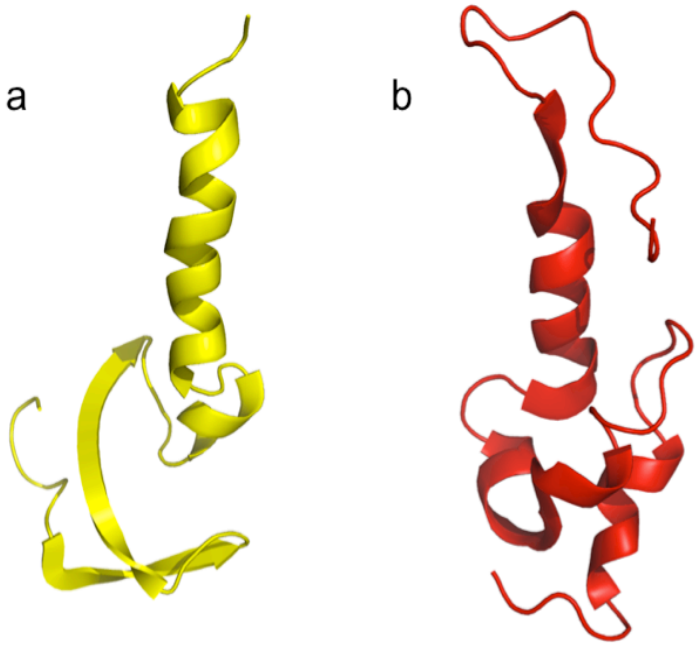

Figure 4.10. Comparison of C-terminal region of SMN-Δ7and SMN-WT. a) SMN-Δ7 is shown to have short loop (yellow) and b) SMN-WT is shown long loop (red).

References

Adams, P.D., Afonine, P.D., Bunkóczi, G., Chen, V.B., Davis, I.W., Echoo ls, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R.; Read, R.J., Richardson, D.C.; Richardson, J.S.; Terwilliger T.C.; Zwart, P.H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta. Cryst. *D66,* 213–221.

Buhler, D., Raker, V., Luhrmann, R., 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.

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

Collaborative computational project, Number 4 (CCP4) (1994). "The CCP4 Suite: Programs for Protein Crystallography". Acta Cryst. *D50,* 760–763.

DeLano, W.L. The PYMOL Molecular Graphics System. DeLano Scientific, Palo Alto, CA, USA, 2002.

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

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.

Leslie, A.G.W. (1999). Integration of macromolecular diffraction data. Acta Crystallogr. *D55*, 1696–1702.

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

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

Martin, R., Gupta, K., Ninan, N.S., Perry, K., Van Duyane, G.D. (2012). The survival motor neuron protein forms soluble glycine zipper oligomers. Struct. *20,* 1–11.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J. (2007). Phaser crystallographic software. J. Appl. Cryst. *40*, 658-674.

Selenko, P., Sprangers, R., Stier, G., Buhler, D., Fischer, U., Sattler, M. (2001). SMN tudor domain structure and its interaction with the Sm proteins. Nature *8*, 27–31.

Sprangers, R., Groves, M.R., Sinning, I., and Sattler, M. (2003). High-resolution X-ray and NMR structure of the SMN Tudor domain: Conformational variation in the binding site for symmetrically dimethylated arginine residues. J. Mol. Biol. *327,* 507–520.

Vaguine, A.A., Richelle, J., 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. Cryst. *D55,* 191–205.

Wirth, B. (2000). An update of the mutation spectrum of the survival motor neuron gene (*SMN1)* in autosomal recessive spinal muscular atrophy (SMA). Hum. Mutation. *15,* 228–237.

Chapter 5

THE THREE-DIMENSIONAL STRUCTURE OF FMO PROTEIN FROM *Pelodictyon phaeum*

Introduction

The Fenna-Mathews-Olson protein belongs to a diverse family of pigment–protein antenna complexes in some photosynthetic organisms that capture light and direct this energy to the integral membrane reaction center complexes where it is converted into chemical energy to generate energy rich compounds (Blankenship, 2002). In green sulfur bacteria, light is absorbed by chlorosomes, which are large complexes attached to the cytoplasmic side of the inner cell membrane (Oostergetel et al., 2007; Psencik et al., 2004; Staehelin et al., 1980). The light energy is transferred from the chlorosome to a bacteriochlorophyll (BChl) *a*-containing FMO protein (Olson, 2004). 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 into a charge-separated state (Wen et al., 2009).

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 (Milder et al., 2010). 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 (Adolphs and Renger, 2006; Engel et al., 2007; Gülen, 1996; Iseri and Gülen, 1999; Louwe et al., 1997; Lu and Pearlstein, 1993; Mohseni et al., 2008; Müh et al., 2007; Vulto et al., 1998a, b, 1999; Wendling et al., 2002). 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 use of a exciton Hamiltonian (Brixner et al., 2005; Panitchayangkoon et al., 2010; Read et al., 2007).

The FMO complex from *Prosthecochloris aestuarii* (*Ptc. aestuarii*) 2K was the first protein containing BChl to be crystallized (Olson, 1978) and have its three-dimensional structure determined using X-ray diffraction (Matthews et al., 1979; Tronrud et al., 1986). Subsequently, the structure of the FMO complex from *Chlorobaculum tepidum* (*Cbl. tepidum*), previously named *Chlorobium tepidum*, was determined (Camara-Artigas et al., 2003; Li et al., 1997). Both complexes are trimers with three identical subunits that are related by a threefold 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 (Ben-Shem et al., 2004; Tronrud et al., 2009). The structure of *Cbl. tepidum* is similar to that of *Ptc. aestuarii* 2K reflecting the significant sequence homology found for the FMO proteins (Tsukatani et al., 2010). 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*), which had been predicted to have significant structural differences compared to *Cbl. tepidum* and *Ptc. aestuarii* 2K based

upon a comparison of the spectroscopic properties (Hu, 2001). In this chapter, 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.

My contribution to this chapter is solving the three-dimensional structure of the FMO protein from *Pld phaeum* in which crystallographic data were obtained from other authors in the published paper (Larson et al., 2011). The purified FMO protein was obtained from the laboratory of Robert Blankenship at Washington University at St. Louis, Missouri.

Materials and methods

Protein expression and 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 (Li et al., 1997). 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.

Results

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* (Li et al., 1997), which had the protein solution mixed with an equal volume of the reservoir containing 6% polyethylene glycol (PEG) 4,000, 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% PEG 2,000 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% PEG 2,000 monomethyl ether. The trays were kept at room temperature in the dark with blue-green crystals appearing in 2–3 days and reaching full size in 2 weeks. Prior to the diffraction measurements, the crystals were placed into a cryoprotectant consisting of 60% PEG 400, 10% PEG 2,000 monomethyl ether, and 10 mM citrate.

Data collection and refinement

Green hexagonal crystals of FMO from *Pld. phaeum* grew with a typical length of 0.1 mm and 0.05 mm width. The crystals belonged to the hexagonal space group $P6_3$ with unit cell dimensions of $a = b = 84.0$ Å and $c = 115.8$ Å along with angles $\alpha = \beta = 90°$ and $\gamma = 120°$ with one protein subunit per asymmetric unit (Table 5.1). 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 Å, integrated using MOSFLM (Leslie, 1999), and scaled with SCALA (CCP4 1994). The initial phases were determined using the molecular replacement method using PHENIX (Adams et al., 2002) with the 3EOJ model of the FMO from *Ptc. aestuarii* 2 K (Tronrud et al., 2009). A unique orientation and position were obtained. Manual model building for all structures was performed in Coot (Emsley and Cowtan, 2004) with iterative rounds of refinement being performed using PHENIX. Procheck (Laskowski et al., 1993) and Sfcheck (Vaguine et al., 1999) 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 β-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 without outliers. The average temperature factors for the protein and water molecules are 21.0 and 37.8 $Å^2$, respectively, yielding an overall value of 22.7 $Å^2$. The root mean square deviations of bond distances and angles are 0.007 Å and 3.027°, respectively. Figures were made using Pymol (DeLano, 2002).

Table 5.1. Crystallographic data summary of FMO from *Pelodictyon phaeum*[a]

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	29,966
Multiplicity	7.7
I/σ (I)	2.0 (1.2)
R_{Merge} (%)[b]	0.186 (0.58)
Completeness (%)	94.5 %
Refinement	
R_{work} (%)/R_{free} (%)[c]	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

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

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

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

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 5.1). 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.

The three identical subunits are arranged as a trimer with each subunit related by a crystallographic threefold axis of symmetry. 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 threefold symmetry axis approximately perpendicular to the membrane surface, with one face oriented toward the chlorosome and the other face toward the membrane (Melkozernov et al., 1998; Wen et al., 2009). 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 toward the membrane has been proposed to facilitate energy transfer to the reaction center (Müh et al., 2007).

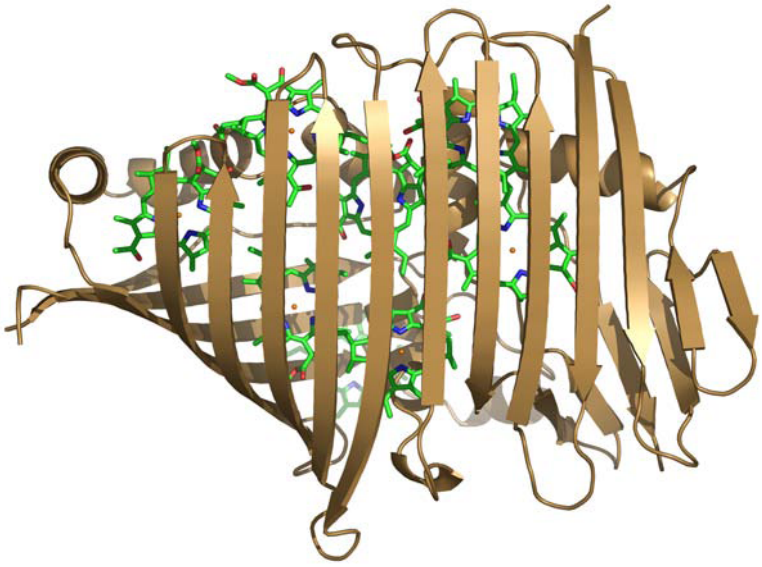

Figure 5.1. The three-dimensional structure of FMO from *Pld. phaeum*. In the cell, the protein is a trimer with the three identical subunits related by a crystallographic threefold symmetry axis approximately vertical to the plane. The protein subunits (*sand*) surround eight BChls cofactors (*atom 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. Coordinates shown are 3OEG for *Pld. phaeum.*

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 5.2). 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, Arg127 of α-helix 1 forms a bridge with Asp*103 and Asp*153, where * represents the neighboring monomer, 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. Tyr119-Glu*178 is one such polar interaction. Also, Arg195 is found in a loop region and

forms a bridge with Glu*178 of α-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 Asp123 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 5.3). The BChl molecules are arranged in an asymmetric fashion. Considering only BChl 1–7, 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 Å, with BChl 6 being the closest, but has a close distance of 12 Å to BChl 1 of the neighboring subunit.

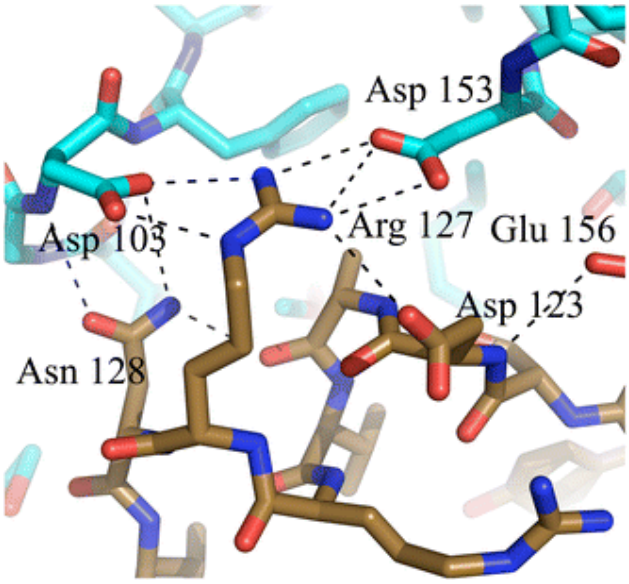

Figure 5.2. The interface between subunits of the trimer of FMO from *Pld. phaeum*. The interface between subunits is primarily stabilized by salt bridges such as those involving Asp103, Arg127, Asn128, and Asp153 for *Pld. phaeum* (carbons *color coded sand* or *cyan* depending upon subunit, oxygen *red*, nitrogen *blue*). Equivalent interactions are found involving Asp107, Asp127, Asn128, Arg131, and Asp157 in *Cbl. tepidum* and Asp107, Asp127, Arg131, Asn132, and Arg157 in *Ptc. aestuarii* 2K. Not all interactions are conserved, for example, the salt bridge between Asp123 and Glu156 is found only in *Pld. phaeum*. Coordinates shown are 3OEG for *Pld. phaeum*.

BChls 1–7 are all five-coordinated, to either a histidine, a backbone carbonyl, or a bound water molecule. His106 located in β-strand 7 coordinates BChl 1. BChl 2 is coordinated by a water molecule in close proximity to the side chain of Asn75. His294 and His293 located in α-helix 7 coordinate BChl 3 and 7, respectively. The long loop connecting β-strands 12 and 13 contains His286 that serves as the axial ligand for BChl 4 and BChl 6 is coordinated to His141 of β-strand 8. The backbone carbonyl of Leu238 of α-helix 5 ligates BChl 5. For the FMO from *Pld. phaeum*, BChl 8 is five-coordinated to the side chain of Ser164 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.

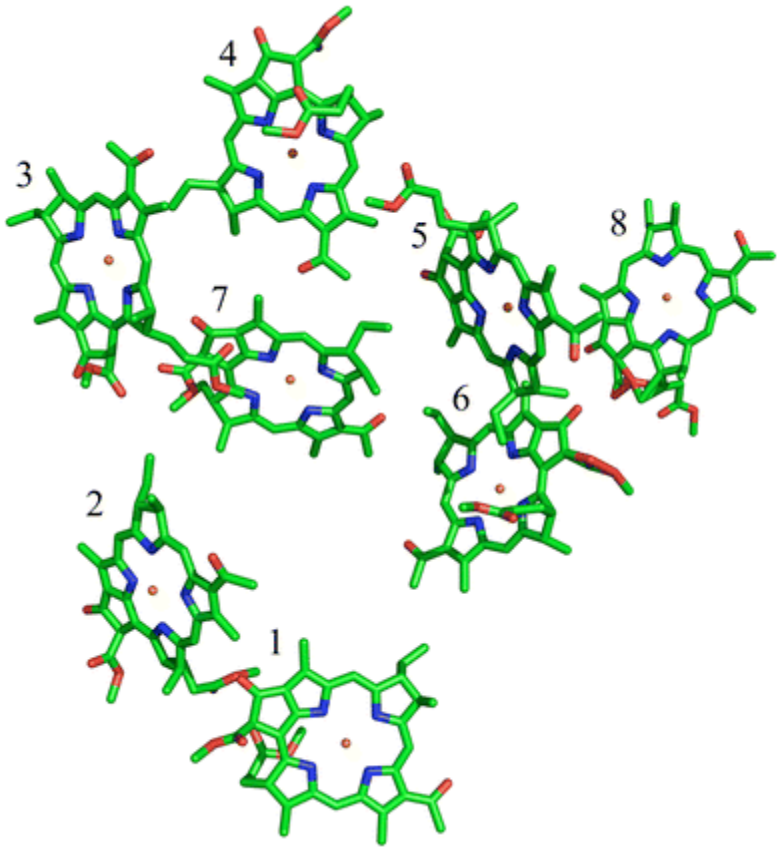

Figure 5.3. The three-dimensional structure of the BChl cofactors of each subunit of FMO from *Pld. phaeum.* The BChls (*atom type*) are arranged in an asymmetric fashion and shown without the phytyl chains for clarity. Each BChl is *numbered* as described in the text. Coordinates shown are 3OEG for *Pld. phaeum* (this work). Small sphere represent the Mg.

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* 2 K and *Cbl. tepidum* (Ben-Shem et al., 2004; Camara-Artigas et al., 2003; Li et al., 1997; Matthews et al., 1979; Tronrud et al., 1986; Tronrud et al., 2009). 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 2 K* and *Cbl. tepidum* (Tsukatani et al., 2010), 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 2–6 amino acids that are fully conserved throughout the sequences. Many of these conserved sections are clustered

close to the coordinating residues for BChls 3 and 6 (Figure 5.4). 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 and structure-based calculations (Adolphs and Renger, 2006; Louwe et al., 1997; Müh et al., 2007; Vulto et al., 1998a, b; Wendling et al., 2002;). 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.

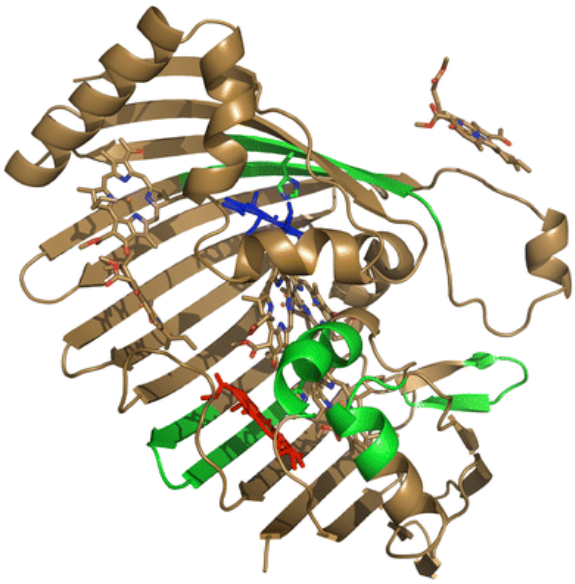

Figure 5.4. Strongly conserved regions for FMO proteins from different species. Shown is the structure of the FMO from *Pld. phaeum* with the regions of conserved amino acids for FMO proteins from green sulfur bacteria highlighted in *green*. These conserved regions are clustered around BChl 3 (*red*), which is thought to be the lowest energy pigment that transfers energy to the reaction center, and BChl 6 (*blue*), and believed to be one of the highest energy pigments that is in close proximity to the chlorosome. Coordinates shown are 3OEG for *Pld. phaeum* (this work).

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, Asp103, Asp107, Asp127, Arg131, and Asp157 (Figure 5.2). 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

The availability of the high-resolution FMO structures from *Pld. phaeum* (this work), *Cbl. tepidum* and *Ptc. aestuarii* 2K (Tronrud et al., 2009) provides the opportunity to investigate the conservation of structural features of the BChl environments. BChl 1 has its Mg coordinated with the same histidine axial ligand, namely, His106 *in Pld. phaeum* and His110 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 5.5). Within van der Waals distance of the macrocycle is Val99 in *Pld. phaeum* and *Cbl. tepidum* (Val 103) but *Ptc. aestuarii* 2 K has Leu103 in that equivalent location. Both *Pld. phaeum* and *Ptc. aestuarii* 2 K have a nearby Phe, namely, Phe161 and Phe165, respectively, but the interaction with an aromatic residue is lost with the presence of Thr165 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 Thr221 for *Ptc. aestuarii* 2 K. At 10 Å, a difference is also found with *Pld. phaeum* having Lys 107 but *Ptc. aestuarii* 2K has Ser111 and *Cbl. tepidum* has Thr111.

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. In all structures, the bound water ligand is hydrogen-bonded to Asn, namely, Asn75 in *Pld. phaeum* and *Cbl. tepidum* and Asn79 in *Ptc. Aestuarii* 2K. However, the number of water molecules hydrogen-bonded to the water ligand is two in the FMO structure from *Pld. phaeum*, three in *Ptc. aestuarii* 2K

with two and three reported for *Cbl. tepidum* (Ben-Shem et al., 2004; Tronrud et al., 2009). The impact of this variation of hydrogen bonds to the water ligand on the electronic structure of BChl 2 is not known.

Different amino acids are adjacent to ring B of BChl 2 in each structure, *Pld. phaeum* (Leu37), *Cbl. tepidum* (Ser 41) and *Ptc. aestuarii* 2K (Ile41). 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 Thr46. Near ring A, *Pld. phaeum* and *Ptc. aestuarii* 2K have Ile66 and Ile70, respectively, but *Cbl. tepidum* has Phe70 in that location. Close to ring C is a Val in both *Pld. phaeum* and *Cbl. tepidum*, Val99 and Val103, respectively, while Leu103 is found in *Ptc. aestuarii* 2K. A difference further apart from the BChl is Thr73 and Thr77 in *Pld. phaeum* and *Cbl. tepidum*, respectively, but Val77 in *Ptc. aestuarii* 2K.

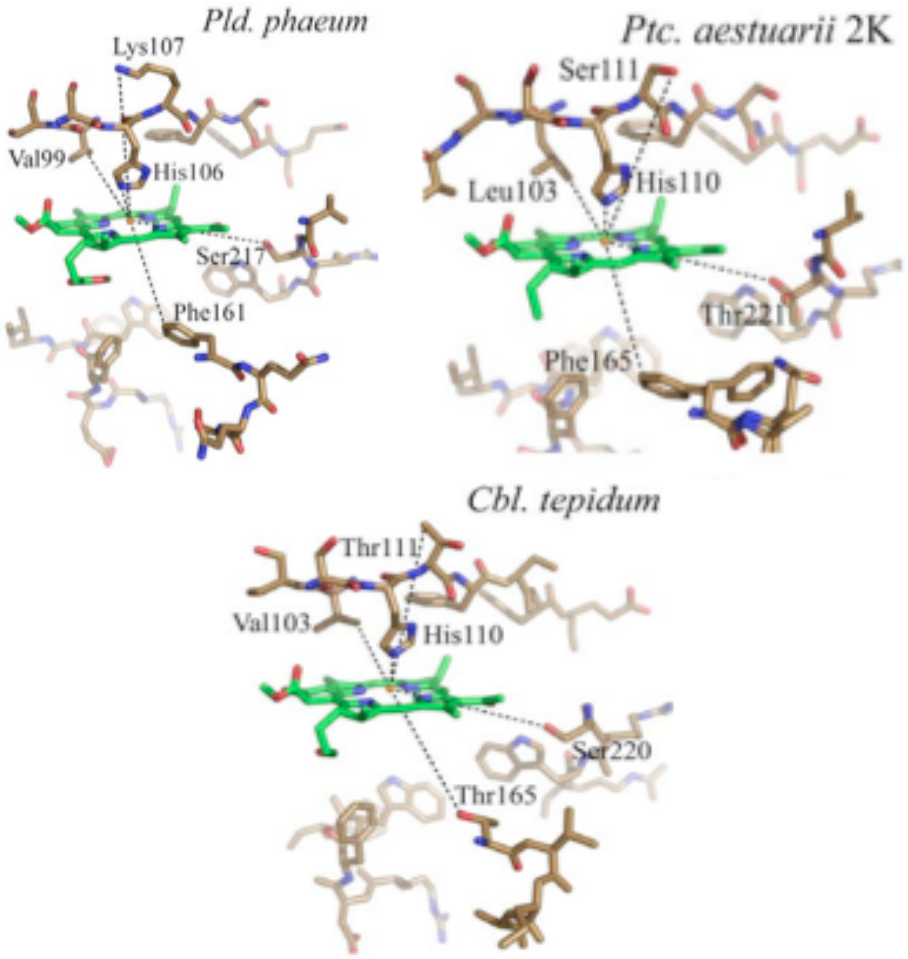

Figure 5.5. Comparison of the structure of BChl 1 and nearby amino acid residues. All three structures have a His as an axial ligand for the Mg of the BChl, namely, 106 in *Pld. phaeum* and 110 in *Ptc. aestuarii* 2K and *Cbl. tepidum* (*color coded* by atom type with *green* and *sand* for carbon in BChl 1 and the protein, respectively). It shows that *Pld. phaeum* and *Ptc. aestuarii* 2K have Phe161 and Phe165, respectively, but *Cbl. tepidum* has Thr165 instead in that location. Two conservative differences are Val99 and Ser217 in *Pld. phaeum* and *Cbl. tepidum* (Val103 and Ser220) but *Ptc. aestuarii* 2K has Leu103 and Thr221. In addition, the ionizable amino acid residue Lys107 is not present in *Ptc. aestuarii* 2 K or *Cbl. tepidum*, which have Ser111 and Thr111, respectively. Coordinates shown are 3OEG for *Pld. phaeum* (this work), 3EOJ for *Ptc. aestuarii* 2K (Tronrud et al., 2009), and 3ENI for *Cbl. tepidum* (Tronrud et al., 2009).

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, His294, 298, and 297 in *Pld. phaeum*, *Ptc. aestuarii* 2K, and *Cbl. tepidum*, respectively (not shown). 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 Pro39. Also, *Cbl. tepidum* has a polar residue, Ser 41 rather than an aliphatic residue, namely, Leu37 and Ile 41, as found in *Pld. phaeum* and *Ptc. aestuarii* 2K, respectively. This conservation of the binding site (Figure 5.4) 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 (Louwe et al., 1997; Vulto et al., 1998a; Wendling et al., 2002). The environment around BChl 4 is also highly conserved with roughly 24 amino acids in close proximity, including the coordinating histidine residue, namely, His286, 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 Mg atom. Both *Pld. phaeum* and *Ptc. aestuarii* 2K have Leu269 and Leu273, respectively, while the corresponding residue is Phe272 in *Cbl. tepidum*. A second difference is that *Pld. phaeum* has Ile25 where *Cbl. tepidum* and *Ptc. aestuarii* 2K both have Val29 in that position. The third difference has *Cbl. tepidum* differing again from *Pld. phaeum* and *Ptc. aestuarii* 2K, Ala34 in *Cbl. tepidum* and Gly27 and Gly31 in *Pld. phaeum* and *P. 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 Leu238 in *Pld. phaeum*, or equivalently Leu242 and Leu241 in *Ptc. aestuarii* 2K and *Cbl. tepidum*, respectively, is present in all structures (Figure 5.6). Of the 27 amino acid

residues 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, Phe62 and Phe66, respectively, but *Cbl. tepidum* has Ile66. Near ring D is the difference that *Cbl. tepidum* has Phe272 and *Pld. phaeum* and *Ptc. aestuarii* 2K have the aliphatic residues Leu269 and Leu273, respectively. In addition, there are several smaller changes. The first minor difference is that *Pld. phaeum* and *Ptc. aestuarii* 2K have Ile48 and Ile52, respectively, while *Cbl. tepidum* has Leu52. The second is that *Pld. phaeum* and *Ptc. aestuarii* 2K are alike with Ala50 and Ala54 but *Cbl. tepidum* has Val54. The third is that *Pld. phaeum* and *Ptc. aestuarii* 2K are identical with Val241 and Val245 and *Cbl. tepidum* has Ser244. The final minor difference is that *Pld. phaeum* and *Ptc. aestuarii* 2 K both have an Ile at 250 and 254, respectively, while *Cbl. tepidum* differs with Val253 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.

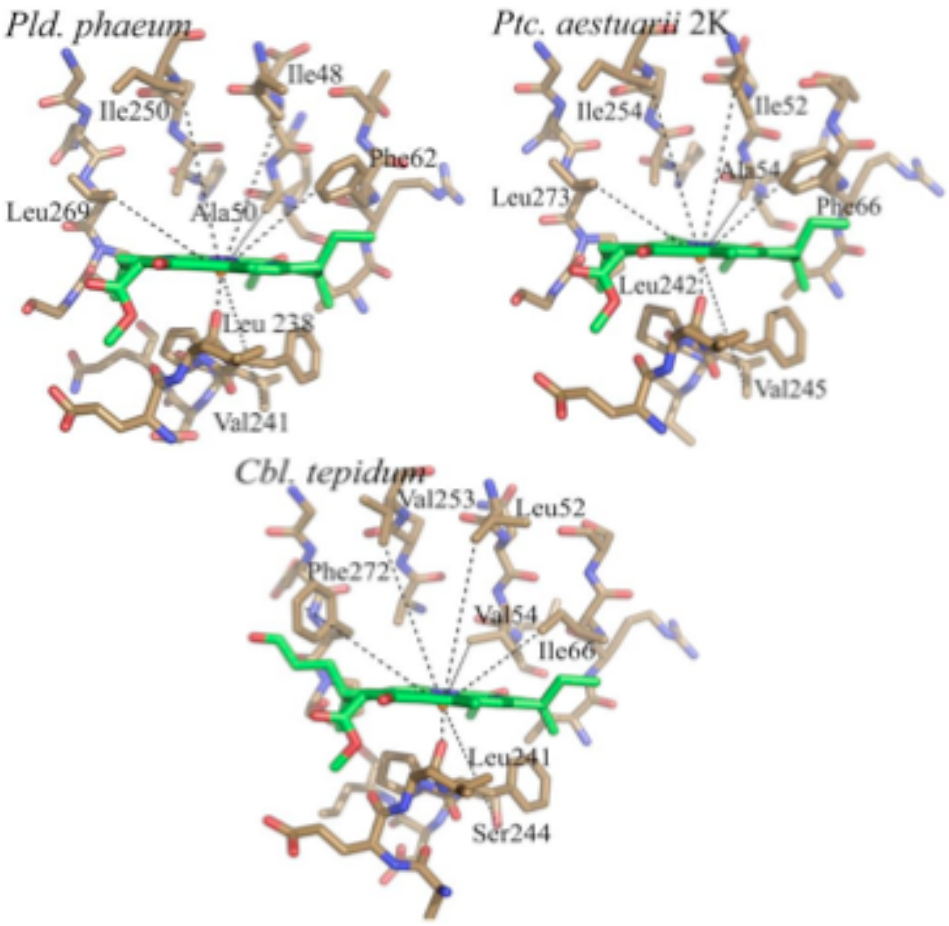

Figure 5.6. Comparison of the structure of BChl 5 and nearby amino acid residues. All three structures have a backbone carbonyl as an axial ligand for the Mg of the BChl, namely, 238, 242, and 241 in *Pld. phaeum*, *Ptc. aestuarii*, and *Cbl. tepidum*, respectively (*color coded* by atom type with *green* 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, Phe62 and Phe66, respectively, but *Cbl. tepidum* differs with Ile66. A second difference is that *Pld. phaeum* and *Ptc. aestuarii* 2K have Leu269 and Leu 273, respectively, where *Cbl. tepidum* has Phe 272. Another difference is that *Pld. phaeum* and *Ptc. aestuarii* 2K 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, Ile48 and Ile250, that are also found in *Ptc. aestuarii* 2K, namely, Ile52 and Ile254, with *Cbl. tepidum* differing with Leu52 and Val253 at the equivalent locations. Coordinates shown are 3OEG for *Pld. phaeum* (this work), 3EOJ for *Ptc. aestuarii* 2K (Tronrud et al., 2009), and 3ENI for *Cbl. tepidum* (Tronrud et al., 2009).

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

The three-dimensional structural models 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 (Camara-Artigas et al., 2003; Li et al., 1997; Matthews et al., 1979; Tronrud et al., 1986). 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 (Ben-Shem et al., 2004). Despite the higher 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 has an axial coordinate to the backbone carbonyl of Tyr124. 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 (Tronrud et al., 2009). 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 (Tronrud et al., 2009). In those models the unoccupied conformation has some large changes, for example Tyr124 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 Ser168 in addition to the backbone carbonyl of Asp123.

The region for BChl 8 has weak electron density and has been re-interpreted by Tronrud and Allen (Tronrud and Allen, 2012). This new model of FMO from *Pld. phaeum* was refined using Buster program (Brigcogne et al., 2009) and refit with Coot, which improved R_{work} and R_{free} values to 16.21% and 19.02%, respectively (Tronrud and Allen, 2012). The new model is identical to the previous, except the region of BChl 8 is fitted by a fragment of polyethylene glycol (PEG) and two water molecules (Figure 5.7) (Tronrud and Allen, 2012). There remains an unassigned region of electron density, which could be the BChl 8 molecule with low occupancy.

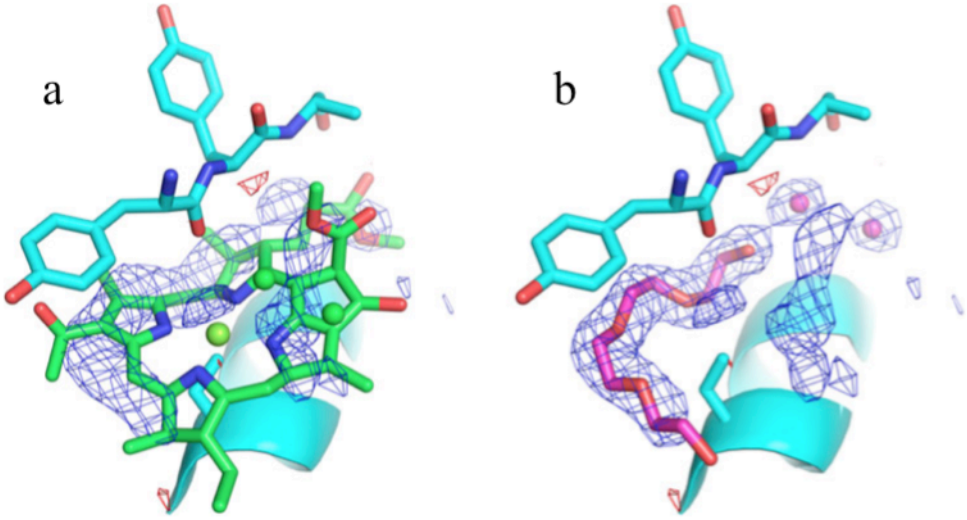

Figure 5.7. Superposition of the old model and the new model of BChl 8 with the omit electron density map. a. The Bchl 8 and two water molecules from the published model (3OEG). b. The new model (3VDI) with its PEG fragment and two water molecules occupying the strongest density (Tronrud and Allen, 2012).

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, both the absorption at 77K and room temperature 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* (Hu, 2001; Tronrud et al., 2009). For *Pld. phaeum*, three peaks are evident in the spectrum in the Q_Y region centered at 805, 814, and 824 nm with a shoulder at 795 nm (Figure 5.8a). The band at 805 nm has the highest amplitude among these five species. 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 805 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 (Figure 5.8b) (Hu, 2001). At room temperature, FMO from *Pld. phaeum* shows a maximum at 797 nm and two minima at 780 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 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 and 821 nm. There are also two shoulders that are found at 810 and 817 nm. The low temperature CD spectrum from *Cbl. tepidum* has two positive peaks at 800 and 813 nm, and two minima at 807 and 825 nm. There are no shoulders seen in the low temperature CD spectrum of *Cbl. tepidum.*

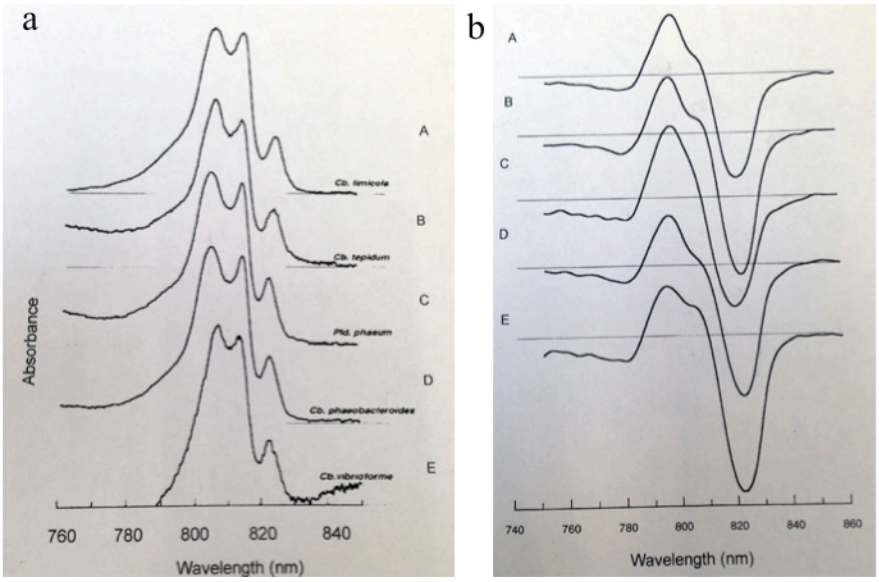

Figure 5.8. Absorption and CD spectra of five FMO proteins. a). Absorption spectra at of five FMO proteins at 77K. b). CD spectra of five FMO proteins at room temperature. A) *Chlorobium limicola,* B) *Chlorobium tepidum*, C) *Pelodictyon phaeum,* D) *Chorobium phaeobacteriods,* E) *Chorobium vibroforme* (Hu, 2001).

Detailed theoretical analyses have been performed to better understand the energetics and energy transfer properties of the FMO proteins (Milder et al., 2010). 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 reflect 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 (Li et al., 1997). In addition, long-range Coulomb interactions may be important in tuning the site energies of the BChls resulting in energetic differences despite similar local environments. Thus, the spectral differences of *Pld. phaeum* compared to FMO from the other species should reflect the effect of both local and long-range 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 (Milder et al., 2010) although a hole-burning study was interpreted in terms of eight exciton components (Johnson and Small, 1991). 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. The presence of BChl 8 alters the spectral interpretation, as this cofactor bridges the neighboring subunits of the trimer and is close to BChl 1 of the neighboring subunit (Figure 5.9). While many different factors influence the coupling (Renger, 2010), 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. 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.

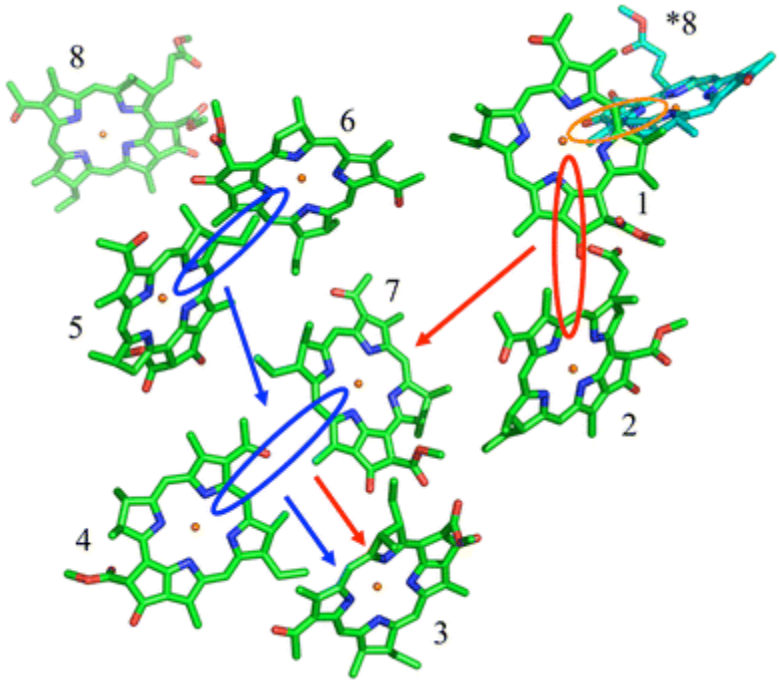

Figure 5.9. Cofactor arrangement in FMO from *Pld. phaeum* with exciton couplings. Shown are BChls 1–8 (colored by *atom type* with *green* carbon) as well as a BChl 8 from a neighboring subunit (BChl *8) of the trimer (colored by atom type with *cyan* carbon). The calculated couplings between BChls 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. Coordinates shown are from PDB:3OEG for *Pld. phaeum* (this study). Figured adapted from Brixner et al., 2005.

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, the position, orientation, and coordination of BChl 8

could only be placed based upon the other FMO structure due to weak electron density. 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.

References

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

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

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

Blankenship, R.E. (2002). Molecular mechanisms of photosynthesis; Wiley-Blackwell: Malden MA.

Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Smart, O.S., Vonrhein, C., Womack, T.O. Buster, version 2.8.0; Global Phasing Ltd., Cambridge, 2009.

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

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

Collaborative Computational Project, Number 4. (1994). The CCP4 suite-programs for protein crystallography. Acta Crystallogr. *D50*, 760–763.

DeLano, W.L. The PYMOL Molecular Graphics System. DeLano Scientific, Palo Alto, CA, USA, 2002.

Emsley, P., Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. *D60*, 2126–2132.

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

Gülen, D. (1996). Interpretation of the excited-state structure of the Fenna–Matthews–Olson pigment protein complex of *Prosthecochloris aestuarii* based on the simultaneous simulation of the 4 K absorption, linear dichroism, and singlet-triplet absorption spectra: a possible excitonic explanation? J. Phys. Chem. *100*, 17683–17689.

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

Iseri, E.I., Gülen, D. (1999). Electronic excited states and excitation transfer kinetics in the Fenna-Matthews-Olson protein of the photosynthetic bacterium *Prosthecochloris aestuarii* at low temperatures. Eur. Biophys. J. *28*, 243–253

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

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. Crystallogr. *26*, 283–291.

Leslie, A.G.W. (1999). Integration of macromolecular diffraction data. Acta Crystallogr. *D55*, 1696–1702.

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

Louwe, R.J.W., Vrieze, J., Aartsma, T.J., 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. *B101*, 11273–11279.

Lu, X., Pearlstein, R. (1993). Simulations of *Prosthecochloris* bacteriochlorophyll *a* protein optical spectra improved by parametric computer search. Photochem. Photobiol. *57*, 86–91.

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

Melkozernov, A.N., Olson, J.M., Li, Y.F., Allen, J.P.; Blankenship, R.E. (1998). Orientation and excitonic interactions of the Fenna–Matthews–Olson protein in membranes of the green sulfur bacterium *Chlorobium tepidum*. Photosynth Res. *56*, 315–328

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

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

Müh, F., Madjet, M.E.A., Adolphs, J., Abdurahman, A., Rabenstein, B., Ishikita, H., Knapp, E.W., Renger, T. (2007). α-Helices direct excitation energy flow in the Fenna–Matthews–Olson protein. Proc. Natl. Acad. Sci. USA *104*, 16862–16867.

Olson, J.M. (1978). Bacteriochlorophyll *a*-proteins from green bacteria. In: Clayton RK, Sistrom WR (eds) The photosynthetic bacteria. Plenum Press: New York, 161–178.

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

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

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

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

Read, E.L., Engel, G.S., Calhoun, T.R., Mančal, T., Ahn, T.K., Blankenship, R.E., Fleming, G.R. (2007). Cross-peak-specific two-dimensional electronic spectroscopy. Proc. Natl. Acad. Sci. USA *104*, 14203–14208.

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

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

Tronrud, D.E., Schmid, M.F., 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.

Tronrud, D.E., Wen, J., Gay, L., 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.

Tronrud, D.E., and Allen, J.P. (2012). Reinterpretation of the electron density at the site of the eighth bacteriochlorophyll in the FMO protein from *Pelodictyon phaeum*, Photosynth Res. *112*, 71-74.

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

Vaguine, A.A., Richelle, J., 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.

Vulto, S.I.E., de Baat, M.A., Louwe, R.J.W., Permentier, H.P., Neef, T., Miller, M., van Amerongen, H., Aartsma, T.J. (1998a). Exciton simulations of optical spectra of the FMO complex from the green sulfur bacterium *Chlorobium tepidum* at 6 K. J. Phys. Chem. B *102*, 9577–9582.

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

Vulto, S.I.E., de Baat, M.A., Neerken, S., Nowak, F.R., van Amerongen, H., Amesz, J., Aartsma, T.J. (1999). Excited state dynamics in FMO antenna complexes from photosynthetic green sulfur bacteria: A kinetic model. J. Phys. Chem. B *103*, 8153–8161.

Wen, J., Zhang, H., Gross, M.L., 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. USA *106*, 6134–6139.

Wendling, M., Przyjalgowski, M.A., Gülen, D., Vulto, S.I.E., Aartsma, T.J., van Grondelle, R., 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.

Chapter 6

THREE-DIMENSIONAL STRUCTURE OF SYNTHETIC ATP BINDING PROTEIN (DX)

Introduction

Besides solving the three-dimensional structures of SMN proteins and FMO protein, the three-dimensional structure of DX protein was also determined. The protein DX is a synthetic ATP binding protein resulting from *in vitro* selection process to mimic the enzymatic activity to the natural enzyme. The protein was synthesized and its biochemical properties were characterized and compared to the known biological ATP binding protein. Synthetic proteins have been designed with random selection from a DNA library. A protein pool containing a $6x10^{12}$ purified non-redundant random proteins each containing 80 amino acid residues was generated from a mRNA library of $4x10^{14}$ independent random sequences with specific constructs to avoid stop codons and frame shift mutation (Keefe and Szostack, 2001). After 18 round of random selections the first *de novo* protein, namely 18-19, with a high affinity ATP binding (k_d =100 nM) was selected from this pool (Keefe and Szostack, 2001).

The three-dimensional structure of protein 18-19 has been solved using X-ray crystallography (Surdo et al., 2004 and Smith et al., 2007). The structure revealed a novel α/β fold with ATP and a Zn metal cofactor to stabilize the protein fold that has not been seen in nature. To improve the stability of 18-19 protein, amino acid residue number 32 was mutated from glutamine to aspartic acid (N32D) and amino acid residue number 45 was mutated from aspartic acid to valine (D45V) yielding the double mutants, DX, protein (Figure 6.1). Surprisingly, the three-dimensional structure of the DX protein

shares the same folding features as protein 18-19 with ATP and Zn cofactor (Smith et al., 2007) (Figure 6.1).

To identify the conformational changes in the DX protein in the absence of ATP, it was not introduced during the purification process. The DX protein was crystallized and structure was solve using X-ray diffraction. The crystallization and three-dimensional structure of the DX protein are described in this chapter.

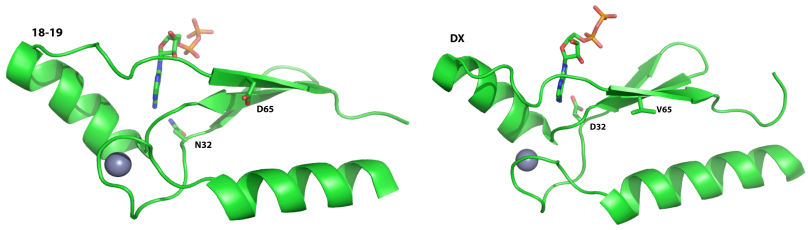

Figure 6.1. Three-dimensional structure of 18-19 and DX proteins (Simmons et al., 2010).

In this chapter, a purified DX protein was obtained from the laboratory Dr. John Chaput at the Biodesign Institute at Arizona State University and my contribution is the crystallization, data collection, data analysis, and solving the three-dimensional structure of DX.

Results and discussion

Protein crystallization

The DX protein was crystallized using the hanging drop vapor diffusion method at room temperature. The protein concentration was approximately 10 mg/ml in 25 mM Na_2HPO_4 pH 8.5 and 200 mM NaCl. The DX protein crystallization condition was optimized base on the crystallization condition from previous publication (Simmons et

al., 2010). A protein drop with a volume of 1 µl was combined with an equivalent volume of reservoir solution containing 150 - 200 mM sodium phosphate (pH 8.5), 500 mM sodium citrate, 400 mM sodium chloride and 10 - 25 % v/v PEG 400, without the presence of the ATP. The DX crystals have grown within 4 days to 250 microns x 75-100 microns x 75-100 microns (Figure 6.2).

Figure 6.2. Image of DX protein crystals.

Data collection and analysis

Crystal data was collected at the Arizona State University X-ray diffraction facility using Rigaku RAXIS IV++ detector at a wavelength of 1.54 Å. The diffraction data were indexed, refined, integrated and scaled using the HKL 2000 package (Otwinowski and Minor, 1997). The DX protein crystals diffracted to a resolution limit of 2.50 Å (Figure 6.3) with the space group $P3_221$, which also share the same space group as the previously solved structure (Smith et al., 2007; Simmons et al., 2009; Simmons et al., 2010). The cell dimensions were a=b= 72.19 Å, c= 54.82 Å with angles $\alpha=\beta= 90°$, $\gamma =120°$. Total observation and unique reflections were 29,575 and 5,845 respectively. R_{merge} was 8.5 % with 99.1% completeness. Data are summarized in Table 6.1.

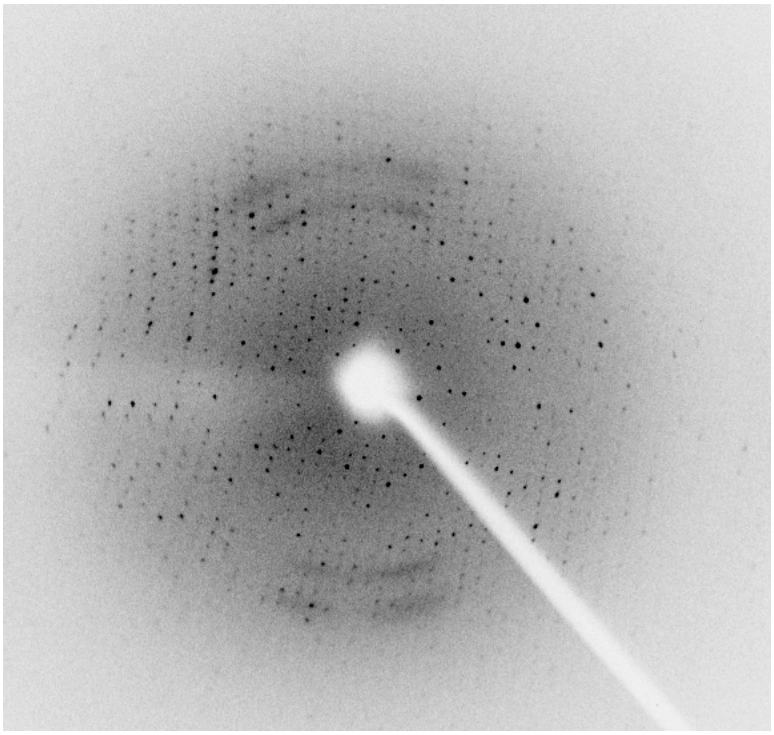

Figure 6.3. Image of X-ray diffraction of DX protein.

Structure refinement

The DX structure was refined using PHENIX (Adams et al., 2010) with the previously solved model (PDB ID: 3DGM) (Simmons et al., 2009). The complete DX structure had R_{work} and R_{free} of 17.31 % and 21.52 %, respectively (Table 6.1). Ramachandran plot values of the structure were 98.51 % and 1.49% in preferred and allowed regions, respectively, with no outlier. RMSD bond length and bond angle were 0.003 Å and 0.931°, respectively.

Table 6.1. Summary of X-ray data and refinement of DX

Data Collection	
Wavelength (Å)	1.54
Exposure time (s)	300
Oscillation range (°)	0.5
Cell dimensions (Å)	a= b= 72.19, c= 54.82
Angle (°)	$\alpha=\beta=90$, $\gamma=120$
Space group	$P3_221$
Resolution (Å)	41.22 - 2.50
Total Observations	29,575
Unique reflections	5,845
Multiplicity	5.0 (4.5)
I/σ	9.7
R_{merge} (%)	8.5
Completeness (%)	99.1 (96.0)
Refinement	
R_{work} (%)	17.31
R_{free} (%)	21.52
RMSD	
Bond length (Å)	0.003
Bond angle (°)	0.931
Ramachandran plot (%)	
Favored	98.51
Allowed	1.49
Outlier	0.00

R-factor =Σ| $|F_{obs}|$ -$|F_{cal}||/ \Sigma |F_{obs}|$, Free R-factor was calculated with setting aside 5% of the reflection data. $R_{merge} = \Sigma_{hkl}\Sigma_j |I_j(hkl) – [I(hkl)]| \Sigma_{hkl}\Sigma_j[I(hkl)]$ where $I_j(hkl)$is the intensity reflection, and I(hkl)is the mean intensity of symmetry related h, k, l.

The three-dimensional structure of DX protein

The DX protein structure has the same fold, α-helices and β-strands as previously reported for the DX structure (Figure 6.1) (Smith et al., 2007; Simmons et al., 2009; Simmons et al., 2010). Present are a Zn metal ion coordinated with Cys residues, a PEG molecule (Figure 6.4a).

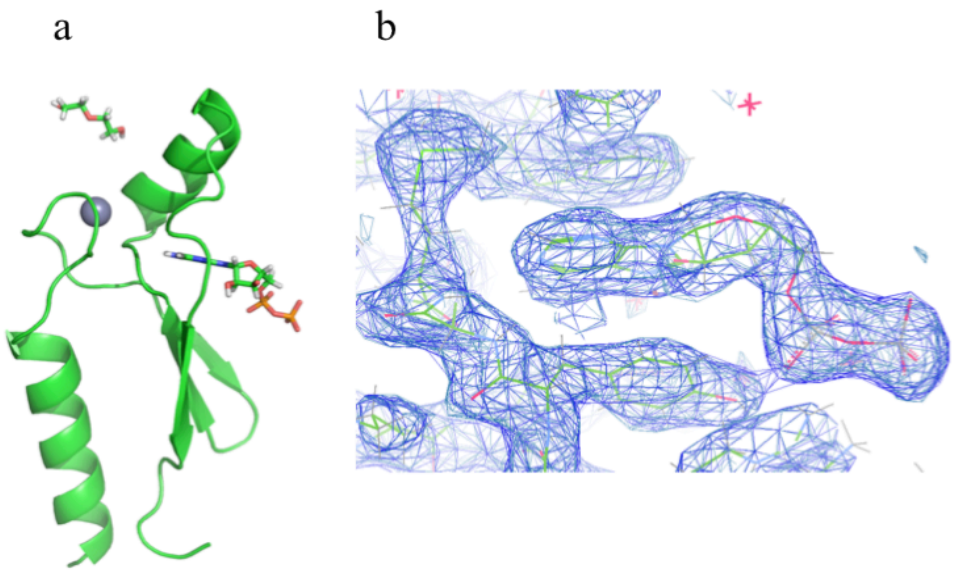

Figure 6.4. Three-dimensional structure of DX protein. a). DX structure shown in green with the presence of Zn (purple sphere), ATP (right) and small molecule PEG (top). The DX consists of three beta strands and two alpha helices. ATP interacts with DX in between bent β-strand and one of the α-helices. b). $2F_o$-F_c electron density map of ATP.

The crystallographic data revealed strong electron density of the ATP at the binding site. Although ATP was not introduced during purification, a small amount of ATP was generated during protein expression and carried through the purification process. The presence of ATP shown in the DX structure (Figure 6.4) (this work) strongly suggested that it is challenging to obtain a free ATP in DX protein sample. Therefore, it is unfeasible to conclude the role of ATP in conformation changes in DX protein from this work. To overcome this problem, it would be necessary to chemically synthesize the DX protein to avoid ATP inferring due to protein overexpressed in *E. coli* system. Then crystallization and X-ray diffraction of DX could be pursued for the conformation change.

References:

Adams, P.D., Afonine, P.D., Bunkóczi, G., Chen, V.B., I. W. Davis, I.W., Echoo ls, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger T.C., Zwart, P.H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta. Cryst. *D66,* 213–221.

Keefe, A.D., and Szostak, J.W. (2001). Functional proteins from a random-sequence library. Nature. *410,* 715–718.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Simmons, C.R., Stomel, J.M., McConnell, M.D., Smith, D.A., Watkin, J.L., Allen, J.P., and Chaput, J.C. (2009). A synthesis protein selected for ligand binding affinity mediates ATP hydrolysis. ACS. Chem. Biol. *4,* 649–658.

Simmons, C.R., Magee, C.L., Smith, D.A., Lauman, L., Chaput, J.C., and Allen, J.P. (2010). Three-dimensional structure reveal multiple ADP/ATP binding modes for a synthetic class of artificial proteins. Biochem. *49,* 8689–8699.

Smith, M.D., Rosenow, M.A., Wang, M., Allen, J.P., Szostak, J.W., and Chaput, J.C. (2007). Structural insights into the evolution of a non-biological protein: important surface residues in protein fold optimization. PloS One. *e467,* 1–11.

Surdo, P.L., Walsh, M.A., and Sollazzo, M. (2004). A novel ADP- and zinc- binding fold from function-directed *in vitro* evolution. Brief communication. *11,* 382–383.

REFERENCES

CHAPER 1

Allen, J.P., Seng, C., Larson, C. (2009). Structures of proteins and cofactors: X-ray crystallography. Photosynth Res. *102,* 231–240.

Bergfords, T. M. (1999). Protein Crystallization: Techniques, Strategies, and Tips (La Jolla, California: International University Line)

Durbin, S.D., and Feher, G. (1996). Protein crystallization. Annu. Rev. Phys. Chem. *47*, 171-204.

Drenth, J. (2007) Principles of Protein X-Ray Crystallography. 3rd edition. (New York: Springer)

Jovine, L. (2000). A simple technique to control macromolecular crystal nucleation efficiently using a standard vapour-diffusion set up. J. Appl. Cryst. *33*, 988-989.

Larson, C.R., Seng, C.O., Lauman, L., Matthies, H.J., Wen, J., Blankenship, R.E., Allen, J. P. (2011). The three-dimensional structure of the FMO protein from *Pelodictyon phaeum* and the implications for energy transfer. Photosynth Res. *107,* 139–150.

Kendrew, J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.G., Davies, D. R., Phillips, D.C., Shore, V.C. (1960). Structure of myoglobin. Nature *185*, 422–427.

Papanikolau, Y., Kotsifaki, D., Fadouloglou, V.E., Gazi, A.D., Glykos, N. M., Cesareni, G., Kokkinidis, M. (2004). Ionic strength reducers: an efficient approach to protein purification and crystallization. Application to two Rop variants. Acta. Cryst. *D60*, 1334-1337.

Perutz, M.F., Rossmann, M.G., Cullis, A. F., Muirhead, H., Will, G., North, A.C.T. (1960). Structure of haemoglobin. Nature *185*, 416–422.

Rhodes, G. (2006). Crystallography Made Crystal Clear (Elsevier Academic Press)

Selenko, P., Sprangers, R., Stier, G., Buhler, D., Fischer, U., Sattler, M. (2001). SMN tudor domain structure and its interaction with the Sm proteins. Nature *8*, 27–31.

Sprangers, R., Groves, M.R., Sinning, I., and Sattler, M. (2003). High-resolution X-ray and NMR structure of the SMN Tudor domain: Conformational variation in the binding site for symmetrically dimethylated arginine residues. J. Mol. Biol. *327,* 507-520.

Tronrud, D.E., and Allen, J.P. (2012). Reinterpretation of the electron density at the site of the eighth bacteriochlorophyll in the FMO protein from *Pelodictyon phaeum*, Photosynth Res. *112*, 71–74.

Wood, S.P. (1990). Purification for Crystallography. In Protein Purification Application: A Practical Approach, E.L.V Harris and S. Angal, ed. (Oxford, Oxford University Press) pp. 45–58.

CHAPTER 2

Buhler, D., Raker, V., Luhrmann, R., 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.

Burghes, A.H. (1997). When is a deletion not a deletion? When it is converted. Am. J. Hum. Genet. *61,* 9–15.

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

Burglen, L., Lefebvre, S., Clemont, O., Burlet, P., Viollet, L., Cruaud, C., Munnich, O., Melki, J. (1996). Structure and organization of the human survival motor neuron (SMN) gene. Gemonics. *32*, 479–482.

Gabanella, F., Butchbach, M.E., Saieva, L., Carissima, C., Burghes, A.H. M., Pellizoni, L. (2007). Ribonucleoprotein assembly defects correlate with spinal muscular atrophy severity and preferentially affect a subset of spliceosomal snRNPs. PloS ONE. *e921,* 1–12.

Gubitz, A.K., Feng, W., Dreyfuss, G. (2004). The SMN complex. Experimental Cell Research *296*, 51–56.

Lefebvre, S., Burglen, L., Reboullet, S., Clemon, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., Le Pasller, D., Frezal, J., Cohen, D., Weissenbach, J., Munnich, A., Melki, J. (1995). Identification and characterization of a spinal muscular atrophy-determining gene. Cell *80*, 155–165.

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

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

MacKenzie, A., and Gendon, N. (2001). Tudor reign. Nature Struct. Biol. *8*, 13–15.

Malatesta, M., Scassellati, C., Meister, G., Plottner, O., Buhler, D., Sowa, G., Martin, T., Keidel, E., Fischer, U., Fakan, S. (2004). Ultrastructural characterization of a nuclear domain highly enriched in survival of motor neuron (SMN) protein. Expirm. Cell Research. *292*, 312–321.

Martin, R., Gupta, K., Ninan, N.S., Perry, K., Van Duyane, G.D. (2012). The survival motor neuron protein forms soluble glycine zipper oligomers. Struct. *20,* 1–11.

Munsat, T.L. (1991). Workshop report: International SMA collaboration. Neuromuscular Disord. *1*, 81.

Nicole, S., Diaz, C.C., Frugier, F., Melki, J. (2002). Spinal muscular atrophy: recent advances and future prospects. Muscle Nerve. *26*, 4–13.

Ogawa, C., Usui, K., Aoki, M., Ito, F., Itoh, M., Kai, C., Kanamori-Katayama, M., Hayashizaki, Y., Suzuki, H. (2007). Gemin2 plays and important role in stabilizing the survival of motor neuron complex. J. Biol. Chem. *282*, 11122–11134.

Otter, S., Grimmler, M., Neuenkirchen, N., Chari, A., Sickmann, A., Fischer, U. (2007). A comprehensive interaction map of the human survival of motor neuron (SMN) complex. J. Biol. Chem. *282*, 5825–5833.

Pellizzoni, L., Charroux, B., Dreyfuss, G. (1999). SMN mutants of spinal muscular atrophy patients are defective in binding to snRNP proteins. PNAS *94*, 11167–11172.

Sarachan, K.L., Valentine, K.G., Gupta, K., Moorman, V.R., Gledhill Jr., J.M., Bernens, M., Tommos, C., Wand, A.J., Duyne, D.V. (2012). Solution structure of the SMN-Gemin2 complex. Biochem. J. *445,* 361–370.

Selenko, P., Sprangers, R., Stier, G., Buhler, D., Fischer, U., Sattler, M. (2001). SMN tudor domain structure and its interaction with the Sm proteins. Nature *8*, 27–31.

Shpargel, K.B., and Matera, A.G. (2005). Gemin proteins are required for efficient assembly of Sm-class ribonucleoproteins. PNAS *102,* 17372–17377.

Sprangers, R., Groves, M.R., Sinning, I., and Sattler, M. (2003). High-resolution X-ray and NMR structure of the SMN Tudor domain: Conformational variation in the binding site for symmetrically dimethylated arginine residues. J. Mol. Biol. *327,* 507–520.

Wang, J., and Dreyfuss, G. (2001). Characterization of function domains of the SMN protein *in vivo*. J. Biol. Chem. *276,* 45387–45393.

Wirth, B. (2000). An update of the mutation spectrum of the survival motor neuron gene (SMN1) in autosomal recessive spinal muscular atrophy (SMA). Human Mutation. *15*, 228–237.

Yong, J., Wan, L., Dreyfuss, G. (2004). Why do cells need an assembly machine for RNA-protein complexes? Trends in Cell Biol. *15,* 226–232.

Young, P. J., Francis, J. W., Lince, D., Coon, K., Androphy, E. J., Lorson, C. L. (2003). The Ewing's sarcoma protein interacts with the Tudor domain of the survival motor neuron protein. Mol. Brain Research*. 119,* 37–49.

Zhang, R., So, B.R., Li, P., Young, J., Glisovic, T., Wan, L., Dreyfuss, G. (2011). Structure of a key intermediate of the SMN complex reveals Gemin2's crucial function in snRNP assembly. Cell *146,* 384–395.

CHAPTER 3

Adams, P.D., Afonine, P.D., Bunkóczi, G., Chen, V.B., Davis, I.W., Echoo ls, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R.; Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger T.C., Zwart, P.H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta. Cryst. *D66,* 213–221.

Alias, L., Bernal, S., Fuentes-Prior, P., Barcelo, M. J., Also, E., Martinez-Hernandez, R., Rodrigue-Alvarez, R., Martin, Y., Aller, E., Grau, E., Pecina, A., Antinolo, G., Galan, E., Rosa, A. L., Fernandez-Burriel, M., Borrego, S., Millan, J. M., Hernandez-Chico, C., Baiget, M., Tizzano, E. F. (2009). Mutation update of spinal muscular atrophy in Spain: molecular characterization of 745 unrelated patients and identification of four mutations in the *SMN1* gene. Hum. Genet. *125,* 29–39.

Buhler, D., Raker, V., Luhrmann, R., 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.

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

Clermont, O., Burlet, P., Cruad, C., Bertrandy, S., Milki, J., Munnich, A., Lefvbre, S. (1997). Mutation analysis of the SMN gene in undeleted SMA patients. *Am. J.* Hum. Genet*. 61,* A329

Collaborative computational project, Number 4 (CCP4) (1994). "The CCP4 Suite: Programs for Protein Crystallography". Acta Cryst. *D50,* 760–763.

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

Gubitz, A. K., Feng, W., Dreyfuss, G. (2004). The SMN complex. Expr. Cell Res. *296,* 51–56.

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.

Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., Le Paslier, D., Frezal, J., Cohen, D., Weissenbach, J., Munnich, A., Melki, J. (1995). Identification and characterization of a spinal muscular atrophy-determining gene. Cell *80,* 155–165.

Liu, Q., Fischer, U., Wang, F., Dreyfuss, G. (1997). The spinal muscular atrophy disease gene product, SMN, and its associated protein SIPI are in a complex with spliceosomal snRNP proteins. Cell *90,* 1013–1021.

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

Kambach, C., Walke, S., Young, R., Avis, J. M., de la Fortelle, E., Raker, V.A., Luhrmann, R., Li, J., and Nagai, K. (1999). Crystal structures of two Sm protein complexes and their implications for the assemblyof the spliceosomal snRNPs. Cell 96, 375–387.

Ma, Y., Dostie, J., Dreyfuss, G., Van Duyne, G. D. (2005). The Gemin6-Gemin7 heterodimer from the survival of motor neurons complex has an Sm protein-like structure. Struct. *13,* 883–892.

Magee, L.C. (2006). Survival of motor neuron protein: structural studies of the SMN protein, a protein link to spinal muscular atrophy. Ph.D. Thesis, Arizona State University, Tempe, AZ.

Ogawa, C., Usui, K., Aoki, M., Ito, F., Itoh, M., Kai, C., Kanamori-katayama, M., Hayashizaki, Y., Suzuki, H. (2007). Gemin2 plays an important role in stabilizing the survival of motor neuron complex. J. Biol. Chem. *282,* 11122–11134.

Otter, S., Grimmler, M., Neuenkirchen, N., Chari, A., Sickmann, A., Fischer, U. (2007). A comprehensive interaction map of the human survival of motor neuron (SMN) complex. J. Biol. Chem. *282,* 5825–5833.

Parsons, D.W., McAndrew, P.E., Iannaccone, S.T., Mendell, R.J., Burghes, A.H. M., and Prior, T.W. (1998). Intragenic telSMN mutations: Frequency, distribution, evidence of a founder effect, and modification of the spinal muscular atrophy phenotype by cenSMN copy number. Am. J. Hum. Genet. *63,* 1712–1723.

Selenko, P., Sprangers, R., Stier, G., Buhler, D., Fischer, U., Sattler, M. (2001). SMN tudor domain structure and its interaction with the Sm proteins. Nature *8*, 27–31.

Sprangers, R., Groes, M.R., Sinning, I., and Sattler, M. (2003). High-resolution X-ray and NMR structure of the SMN Tudor domain: Conformational variation in the binding site for symmetrically dimethylated arginine residues. J. Mol. Biol. *327,* 507–520.

Sun, Y., Grimmler, M., Schwarzer, V., Schoenen, F., Fischer, U., and Wirth, B. (2005). Molecular and function analysis of intragenic *SMN1* mutations in patients with spinal muscular atrophy. Hum. Mutation *25,* 64–71.

Vaguine, A.A., Richelle, J., 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. Cryst. *D55,* 191–205.

Wirth, B., Herz, M., Wetter, A., Moskau, S., Hahnen, E., Rudnik-Schoneborn, S., Wienker, T., and Zerres, K. (1999). Quantitative analysis of survival motor neuron copies: Identification of subtle *SMN1* gene mutations in patients with spinal muscular atrophy, genotype-phenotype correlation, and implications for genetic counseling. Am. J. Hum. Genet. *64,* 1340–1356.

Wirth, B. (2000). An update of the mutation spectrum of the survival motor neuron gene (*SMN1)* in autosomal recessive spinal muscular atrophy (SMA). Hum. Mutation. *15,* 228–237.

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. Human Mol. Genet. *9,* 2869–2877.

Zhang, R., So, B.R., Li, P., Yong, J., Glisovic, T., Wan, L., Dreyfuss, G. (2011). Structure of a key intermediate of the SMN complex reveals Gemin2's crucial function in snRNP assembly. Cell *146,* 384–395.

CHAPTER 4

Adams, P.D., Afonine, P.D., Bunkóczi, G., Chen, V.B., Davis, I.W., Echoo ls, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R.; Read, R.J., Richardson, D.C.; Richardson, J.S.; Terwilliger T.C.; Zwart, P.H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta. Cryst. *D66,* 213–221.

Buhler, D., Raker, V., Luhrmann, R., 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.

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

Collaborative computational project, Number 4 (CCP4) (1994). "The CCP4 Suite: Programs for Protein Crystallography". Acta Cryst. *D50,* 760–763.

DeLano, W.L. The PYMOL Molecular Graphics System. DeLano Scientific, Palo Alto, CA, USA, 2002.

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

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.

Leslie, A.G.W. (1999). Integration of macromolecular diffraction data. Acta Crystallogr. *D55*, 1696–1702.

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

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

Martin, R., Gupta, K., Ninan, N.S., Perry, K., Van Duyane, G.D. (2012). The survival motor neuron protein forms soluble glycine zipper oligomers. Struct. *20,* 1–11.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J. (2007). Phaser crystallographic software. J. Appl. Cryst. *40*, 658-674.

Selenko, P., Sprangers, R., Stier, G., Buhler, D., Fischer, U., Sattler, M. (2001). SMN tudor domain structure and its interaction with the Sm proteins. Nature *8*, 27–31.

Sprangers, R., Groves, M.R., Sinning, I., and Sattler, M. (2003). High-resolution X-ray and NMR structure of the SMN Tudor domain: Conformational variation in the binding site for symmetrically dimethylated arginine residues. J. Mol. Biol. *327,* 507–520.

Vaguine, A.A., Richelle, J., 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. Cryst. *D55,* 191–205.

Wirth, B. (2000). An update of the mutation spectrum of the survival motor neuron gene (*SMN1)* in autosomal recessive spinal muscular atrophy (SMA). Hum. Mutation. *15,* 228–237.

CHAPTER 5

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

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

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

Blankenship, R.E. (2002). Molecular mechanisms of photosynthesis; Wiley-Blackwell: Malden MA.

Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Smart, O.S., Vonrhein, C., Womack, T.O. Buster, version 2.8.0; Global Phasing Ltd., Cambridge, 2009.

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

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

Collaborative Computational Project, Number 4. (1994). The CCP4 suite-programs for protein crystallography. Acta Crystallogr. *D50*, 760–763.

DeLano, W.L. The PYMOL Molecular Graphics System. DeLano Scientific, Palo Alto, CA, USA, 2002.

Emsley, P., Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. *D60*, 2126–2132.

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

Gülen, D. (1996). Interpretation of the excited-state structure of the Fenna–Matthews–Olson pigment protein complex of *Prosthecochloris aestuarii* based on the simultaneous simulation of the 4 K absorption, linear dichroism, and singlet-triplet absorption spectra: a possible excitonic explanation? J. Phys. Chem. *100*, 17683–17689.

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

Iseri, E.I., Gülen, D. (1999). Electronic excited states and excitation transfer kinetics in the Fenna-Matthews-Olson protein of the photosynthetic bacterium *Prosthecochloris aestuarii* at low temperatures. Eur. Biophys. J. *28*, 243–253

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

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. Crystallogr. *26*, 283–291.

Leslie, A.G.W. (1999). Integration of macromolecular diffraction data. Acta Crystallogr. *D55*, 1696–1702.

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

Louwe, R.J.W., Vrieze, J., Aartsma, T.J., 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. *B101*, 11273–11279.

Lu, X., Pearlstein, R. (1993). Simulations of *Prosthecochloris* bacteriochlorophyll *a* protein optical spectra improved by parametric computer search. Photochem. Photobiol. *57*, 86–91.

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

Melkozernov, A.N., Olson, J.M., Li, Y.F., Allen, J.P.; Blankenship, R.E. (1998). Orientation and excitonic interactions of the Fenna–Matthews–Olson protein in membranes of the green sulfur bacterium *Chlorobium tepidum*. Photosynth Res. *56*, 315–328

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

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

Müh, F., Madjet, M.E.A., Adolphs, J., Abdurahman, A., Rabenstein, B., Ishikita, H., Knapp, E.W., Renger, T. (2007). α-Helices direct excitation energy flow in the Fenna–Matthews–Olson protein. Proc. Natl. Acad. Sci. USA *104*, 16862–16867.

Olson, J.M. (1978). Bacteriochlorophyll *a*-proteins from green bacteria. In: Clayton RK, Sistrom WR (eds) The photosynthetic bacteria. Plenum Press: New York, 161–178.

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

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

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

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

Read, E.L., Engel, G.S., Calhoun, T.R., Mančal, T., Ahn, T.K., Blankenship, R.E., Fleming, G.R. (2007). Cross-peak-specific two-dimensional electronic spectroscopy. Proc. Natl. Acad. Sci. USA *104*, 14203–14208.

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

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

Tronrud, D.E., Schmid, M.F., 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.

Tronrud, D.E., Wen, J., Gay, L., 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.

Tronrud, D.E., and Allen, J.P. (2012). Reinterpretation of the electron density at the site of the eighth bacteriochlorophyll in the FMO protein from *Pelodictyon phaeum*, Photosynth Res. *112*, 71-74.

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

Vaguine, A.A., Richelle, J., 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.

Vulto, S.I.E., de Baat, M.A., Louwe, R.J.W., Permentier, H.P., Neef, T., Miller, M., van Amerongen, H., Aartsma, T.J. (1998a). Exciton simulations of optical spectra of the FMO complex from the green sulfur bacterium *Chlorobium tepidum* at 6 K. J. Phys. Chem. B *102*, 9577–9582.

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

Vulto, S.I.E., de Baat, M.A., Neerken, S., Nowak, F.R., van Amerongen, H., Amesz, J., Aartsma, T.J. (1999). Excited state dynamics in FMO antenna complexes from photosynthetic green sulfur bacteria: A kinetic model. J. Phys. Chem. B *103*, 8153–8161.

Wen, J., Zhang, H., Gross, M.L., 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. USA *106*, 6134–6139.

Wendling, M., Przyjalgowski, M.A., Gülen, D., Vulto, S.I.E., Aartsma, T.J., van Grondelle, R., 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.

CHAPTER 6

Adams, P.D., Afonine, P.D., Bunkóczi, G., Chen, V.B., I. W. Davis, I.W., Echoo ls, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger T.C., Zwart, P.H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta. Cryst. *D66,* 213–221.

Keefe, A.D., and Szostak, J.W. (2001). Functional proteins from a random-sequence library. Nature. *410,* 715–718.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Simmons, C.R., Stomel, J.M., McConnell, M.D., Smith, D.A., Watkin, J.L., Allen, J.P., and Chaput, J.C. (2009). A synthesis protein selected for ligand binding affinity mediates ATP hydrolysis. ACS. Chem. Biol. *4,* 649–658.

Simmons, C.R., Magee, C.L., Smith, D.A., Lauman, L., Chaput, J.C., and Allen, J.P. (2010). Three-dimensional structure reveal multiple ADP/ATP binding modes for a synthetic class of artificial proteins. Biochem. *49,* 8689–8699.

Smith, M.D., Rosenow, M.A., Wang, M., Allen, J.P., Szostak, J.W., and Chaput, J.C. (2007). Structural insights into the evolution of a non-biological protein: important surface residues in protein fold optimization. PloS One. *e467,* 1–11.

Surdo, P.L., Walsh, M.A., and Sollazzo, M. (2004). A novel ADP- and zinc- binding fold from function-directed *in vitro* evolution. Brief communication. *11,* 382–383.