Predicting the Structure of the Light-Harvesting Complex II of *Rhodospirillum molischianum*

Xiche Hu, Dong Xu, Kenneth Hamer and Klaus Schulten
Theoretical Biophysics
Beckman Institute
University of Illinois at Urbana–Champaign, Urbana, IL 61801

Juergen Koepke and Hartmut Michel
Max-Planck-institut für Biochemie
Abteilung Molekulare Membranbiologie
6000 Frankfurt, Germany

Abstract We attempted to predict through computer modeling the structure of the light-harvesting complex II (LH-II) of *Rhodospirillum molischianum*, before the impending publication of the structure of a homologous protein solved by means of X-ray diffraction. The protein studied is an integral membrane protein of sixteen independent polypeptides, eight α-apoproteins and eight β-apoproteins, which aggregate and bind to 24 bacteriochlorophyll a's and twelve lycopenes. Available diffraction data of a crystal of the protein, which could not be phased due to a lack of heavy metal derivatives, served to test the predicted structure, guiding the search. In order to determine the secondary structure, hydropathy analysis was performed to identify the putative transmembrane segments and multiple sequence alignment propensity analyses were used to pinpoint the exact sites of the 20 residue long transmembrane segment and the four residue long terminal sequence at both ends, which were independently verified and improved by homology modeling. A consensus assignment for the secondary structure was derived from a combination of all the prediction methods used. Three-dimensional structures for the α- and the β-apoprotein were built by comparative modeling. The resulting tertiary structures are combined, using X-PLOR, into an $\alpha\beta$ dimer pair with bacteriochlorophyll a's attached under constraints provided by site directed mutagenesis and spectral data. The $\alpha\beta$ dimer pairs were then aggregated into a quaternary structure through further molecular dynamics simulations and energy minimization. The structure of LH-II so determined is an octamer of $\alpha\beta$ heterodimers forming a ring with a diameter of 70 Å.

Keywords: integral membrane protein; light-harvesting complex; purple bacteria; protein folding; protein structure; sequence analysis; molecular replacement.

The photosynthetic apparatus of purple bacteria consist of two types of protein-pigment complexes: the reaction centers and the light-harvesting complexes. The main function of the light-harvesting complexes is to collect and transfer light energy to the reaction centers where it is trapped in a charge separated state. In most purple bacteria, the photosynthetic membranes contain two types of light-harvesting complexes: the B870 complex [light harvesting complex I (LH-I)] which is directly bound to the photosynthetic reaction centers; and the B800-850 complex, also known as light harvesting complex II (LH-II) which is not directly associated with the reaction centers, but transfers energy to the reaction centers via LH-I (Zuber, 1985).

Studies of the mechanism of the photosynthetic reaction center have benefited greatly from the solution of its three-dimensional structure (Deisenhofer & Michel, 1989). However, structural information about light-harvesting complexes is still limited to spectroscopic and biochemical characterization (Hawthornthwaite & Cogdell, 1991, Sundstrom & van Grondelle, 1991, Zuber & Brunisholz, 1991). The LH-II complex of *Rs. molischianum* has been crystallized and X-ray diffraction data have been collected up to 2.4 Å resolution (Koepke & Michel, unpublished; Michel, 1991). To resolve a structure from measured diffraction intensities requires knowledge of phases which is unobtainable from a single diffraction experiment. Conventionally, the phase problem is solved by means of the multiple isomorphous replacement method. An alternative solution to the phase problem is to phase the structure by using a homologous structure in a procedure called molecular replacement (Lattman, 1985, Rossmann, 1972). In this method, a homologous probe structure is fit into the unit cell of the unknown structure and used to generate an initial phasing model for the unknown structure. Currently there exists no homologous structure to LH-II of *Rs. molischianum*. Although several types of light-harvesting complexes have been crystallized and characterized by X-ray crystallography and electron microscopy (Cogdell & Hawthornthwaite, 1993, Germeroth *et al.*, 1993, Karrasch *et al.*, 1995, Welte *et al.*, 1985), no structure has yet been published[1].

[1] This situation will soon be changed by the expected publication of a structure for

We attempted to predict the structure of LH-II of *Rs. molischianum*. The results achieved will demonstrate to what extent current methods are capable of predicting the structure of membrane proteins.

In anticipation of the publication of the crystal structure of LH-II of *Rps. acidophila*, we present here (1) the predicted secondary structures for the α- and the β-apoprotein of *Rs. molischianum*; (2) the predicted quaternary structure for the aggregated LH-II complex of *Rs. molischianum* and (3) the procedures adopted which are all based on methods developed in other laboratories. We report this work in the spirit of the 1994 Asilomar Conference, [2] i.e., the predicted structure should be published before crystallographic data are available to truly evaluate the merit of predictions (Barton & Russell, 1993, Benner *et al.*, 1992).

All light-harvesting complexes display a remarkable similarity in the way they are constructed (Zuber, 1985, Zuber & Brunisholz, 1991). The basic structural unit is a heterodimer of small α- and β-apoproteins containing non-covalently bound pigments (bacteriochlorophyll and carotenoids). These heterodimers aggregate to a large complex, functioning as light harvesting antennae. The size of the aggregate depends on the type of light-harvesting complex and varies from species to species, ranging from a putative hexamer for LH-II of *Rb. sphaeroides* (Boonstra *et al.*, 1993) to a hexadecamer for LH-I of *Rs. rubrum* (Karrasch *et al.*, 1995). Various models have been proposed for the light-harvesting complex of purple bacteria (Olsen & Hunter, 1994, Zuber, 1986, Zuber & Brunisholz, 1991). The majority of these models were concerned with secondary structural features and the topology of the heterodimers. However, no atomic level modeling of the aggregated complex has been attempted before. Our goal was to build a model structure for LH-II of *Rs. molischianum* to test it against available, unphased diffraction data and to use it in the future as a probe structure in the framework of the molecular replacement method to phase the diffraction data. The task of the prediction is divided into four stages:

LH-II from *Rps. acidophila* by Cogdell et al., who have successfully solved the phase problem for their structure by conventional means

[2] Critical assessment of techniques for protein structure prediction, Asilomar, California, December 4-8, 1994.

(1) predict the secondary structure of the α- and the β-apoproteins from their amino acid sequences; (2) build the tertiary structure of the α- and the β-apoprotein by comparative modeling; (3) form an aggregated complex (quaternary structure) of the α- and β-apoproteins by means of molecular dynamics simulations and energy minimization under the constraint of biochemical and spectral data and the predicted secondary structural features; (4) perform a molecular replacement test using the predicted structure as a probe structure and determine R values. A flowchart of the entire procedure is provided in Figure 1.

The molecular dynamics simulations and energy minimization described in the following sections were carried out using the program X-PLOR (Brünger, 1992). All the simulation protocols were programmed with the versatile X-PLOR script language. An integration time step of 1 fs was chosen in the Verlet algorithm. Our simulation of LH-II placed the protein in a vacuum. The parameters and charges used for the system were, respectively, the CHARMm all-atom parameter file *parallh22x.pro* and the CHARMm all-atom partial charge file *topallh22x.pro* (Brooks et al., 1983; MacKerell & Karplus, unpublished) except for BChla. The partial charges and parameters for BChla were taken from those used in (Treutlein *et al.*, 1988) for BChlb except for slight modifications to accommodate BChla. A cut-off distance of 12 Å for non-bonded interactions and a dielectric constant $\epsilon = 1$ were employed.

Discussion

The LH-II complex of *Rs. molischianum* consists of two BChlα-binding apoproteins α and β with the following sequences:

α: `SNPKDDYKIWLVINPSTWLPVIWIVATVVAIAVHAAVLAAPGSNWIALGAAKSAAK`
β: `AERSLSGLTEEEAIAVHDQFKTTFSAFIILAAVAHVLVWVWKPWF`

The smallest compositional unit of LH-II contains a pair of α- and β-apoproteins, three BChla and 1.5 lycopene molecules (Germeroth *et al.*, 1993). It has been determined by sedimentation equilibrium experiments that the native LH-II complex is an octamer of such $\alpha\beta$ units (Kleinekofort

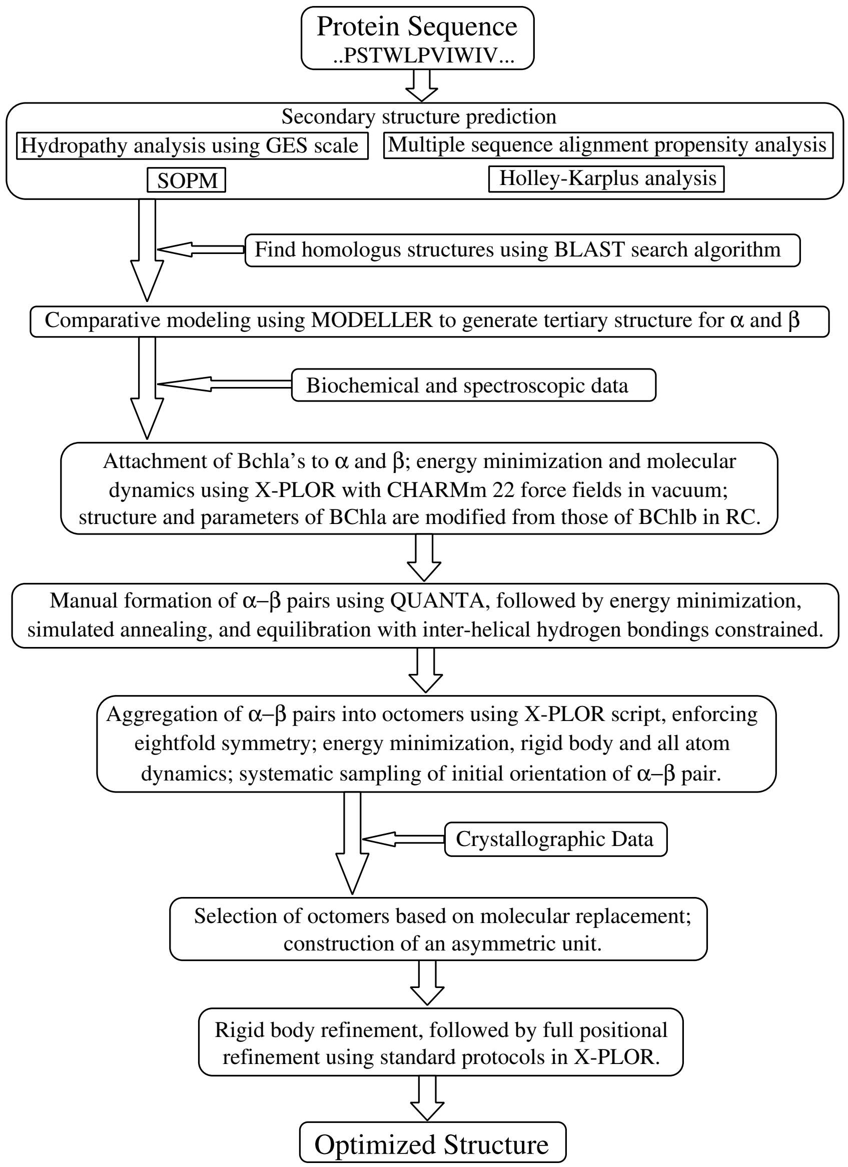

Figure 1: Flowchart outlining the entire structure prediction procedure.

et al., 1992). The space group for the crystal is $P42_12$ with cell dimensions of $92 \times 92 \times 209$ Å.

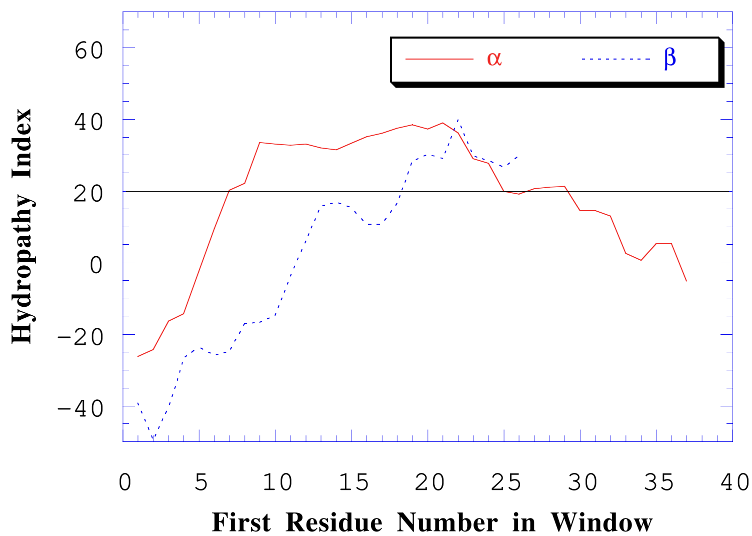

Figure 2: Hydropathy plot of the α- and β-apoproteins of the LH-II complex based on the GES hydrophobicity scale (see text) with a window size of 20 amino acids. The abscissa corresponds to the first residue number of a window of length 20.

Since the LH-II complex is an integral membrane protein, we performed a hydropathy analysis to identify the putative transmembrane segments (Kyte & Doolittle, 1982, White, 1994). The transmembrane segments of apoproteins are usually forced to adopt an α-helical conformation due to constraints of the hydrophobic core of the membrane (Engelman *et al.*, 1986). Hydropathy analysis assumes that transmembrane segments are comprised mainly of hydrophobic residues because of the low solubility of polar side chains in nonpolar lipid bilayers. It requires about 20 amino acids (in an a helical conformation) to span the hydrocarbon regions of fluid bilayers that are typically 30 Å thick. Shown in Figure 2 is a hydropathy plot for the α-

and β-apoproteins based on the GES hydrophobicity scale (Engelman *et al.*, 1986) with a window size of 20 amino acids. The GES scale is derived from the free energy cost for transferring amino acids from the interior of a membrane to its water surroundings. On such a scale, a peak of 20 kcal/mol or higher identifies a transmembrane segment (Engelman *et al.*, 1986). Figure 2 clearly shows that a transmembrane segment exists for both the α- and the β-apoprotein. The highest peak occurs at a first residue number of 21 and 22 for the α- and the β-apoprotein, respectively. The transmembrane segments are thus identified as α-Val-21:α-Alα-40 and β-Thr-22:β-Trp-41. It is worth mentioning that, although hydropathy analysis has been successfully applied to identifying transmembrane helices, the method can not determine the ends of helices precisely (Michel *et al.*, 1986). The accuracy of the prediction relies on a proper choice of the window size and the cut-off value for identifying membrane spanning residues which have been calibrated against known structures of membrane proteins (Cornette *et al.*, 1987, Esposti *et al.*, 1990, Eisenberg, 1984, Engelman *et al.*, 1986, Jahnig, 1990, Kyte & Doolittle, 1982, White, 1994). At present, the number of membrane protein structures known are still too few to generate a valid calibration. In addition, transmembrane helices are not always 20 residues long. All the transmembrane helices in photosynthetic reaction center of *Rps. viridis*, for example, are 24-30 residues long (Michel *et al.*, 1986). In our approach, we utilized hydropathy analysis solely to identify the hydrophobic core of transmembrane helices. We note that hydropathy analyses with other two widely used hydrophobicity scales, i.e., Kyte & Doolittle (Kyte & Doolittle, 1982) and Eisenberg consensus (Eisenberg *et al.*, 1982) scales, generated essentially the same hydrophobic core as the GES scale, for the α- and the β-apoprotein [see Figures S1 and S2 in the supplementary material].

In addition to the hydropathy analysis, we have also carried out a multiple sequence alignment propensity analysis using the method of Persson and Argos (Persson & Argos, 1994) which combines two sets of propensity values (one for the middle, hydrophobic portion and one for the terminal region of the transmembrane span) to determine the transmembrane segments from multiply aligned amino acid sequences. A novel aspect of this method is the use of evolutionary information in the form of multiple se-

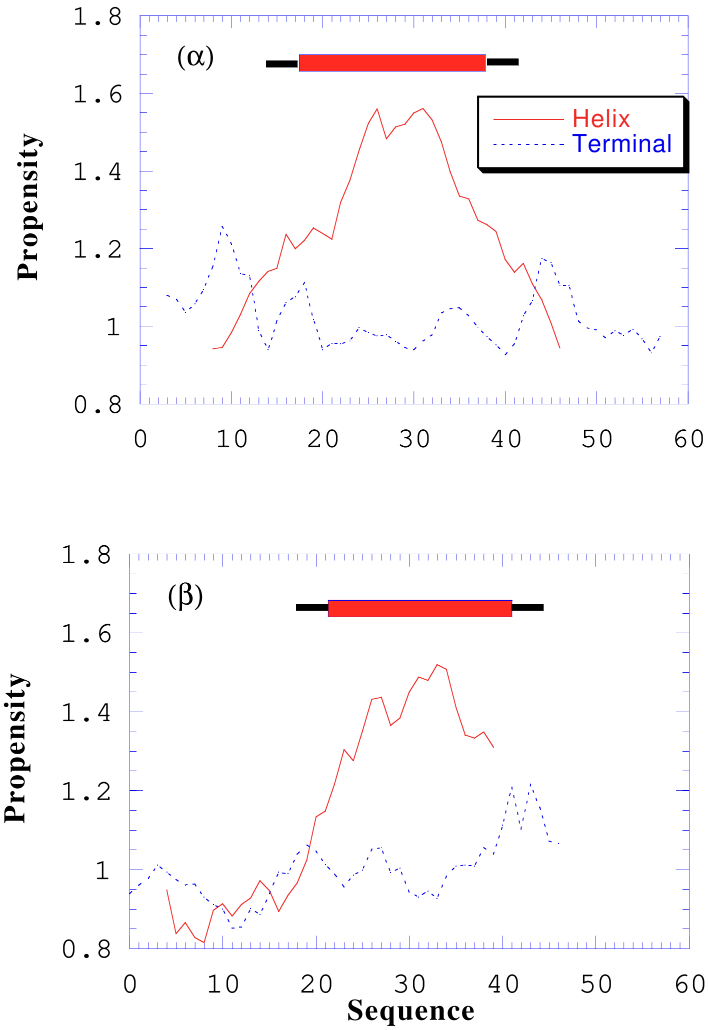

Figure 3: Propensity profile and assignment of 20 residue long transmembrane span (thick bar) and four residue long helical terminal (thin bar) for the α- and β-apoproteins based on the multiple sequence alignment propensity analysis method of Persson and Argos(Persson & Argos, 1994). Solid line - propensity for transmembrane span; dashed line - propensity for terminal region.

quence alignments as input in place of a single sequence. The method was shown to be more successful than predictions based on a single sequence alone. A total of twelve homologous sequences of LH-II and LH-I complexes have been aligned (see Table S1 in the supplementary material) and analyzed. As shown in Figure 3, the transmembrane segment determined with this method spans from Trp-18 to Val-37 for the α-apoprotein and from Thr-22 to Trp-41 for the β-apoprotein. The hydropathy analysis identifies the region of the transmembrane segment. Multiple sequence alignment propensity analysis further pinpoints the most probable site of the 20 residue long transmembrane segment and the four residue long terminal sequence at both ends (see Figure 3).

To confirm the above secondary structure prediction, we performed more sequence analyses with other available secondary structure prediction methods including SOPM (self optimized prediction method) and Holley-Karplus analysis (Geourjon & Deleage, 1994, Holley & Karplus, 1991) [see Tables S2 and S3 in the supplementary material]. SOPM takes structure classes into account and iteratively optimized prediction parameters to increase prediction quality. The Holley-Karplus prediction is an information-based neural network approach. The results are consistent with the assignment of the transmembrane segment, for both the α- and the β-apoprotein, as derived above. Both analyses demonstrate that residues in the putative transmembrane segments have a high tendency to form an α-helix.

The secondary structure assignments were further verified and improved by homology modeling. Although there exists no structure which is highly homologous to either the α- or the β-apoprotein as a whole, we have found structures in the PDB which are homologous to multiple fragments of the α- and the β-apoprotein. Figure 4 lists some of the homologous fragments to the α- and the β-apoprotein which resulted from a PDB BLAST search (Altschul *et al.*, 1990). [see Table S4 in the supplementary material for details of PDB BLAST search results]. It is worth mentioning that a homology with 26% identity and 50% positive exists between a segment of the α-apoprotein [α-Leu-11 to α-Alα-40] and the transmembrane helix D of the M chain in the photosynthetic reaction center from *Rb. sphaeroides* [M chain, residues 196:225] (Michel *et al.*, 1986). To establish the statistical significance of this alignment, we performed a statistical analysis with the BESTFIT program in the GCG package (Devereux *et al.*, 1984). Using a gap generating penalty of 3.0 and a gap extension penalty of 0.1, the BESTFIT program generated the same alignment as shown in Figure 4 with a quality of 19.2. The average quality for 100 randomized alignments, in which the query sequence was randomly permuted (shuffled), was 12.7 with a standard deviation of 1.3. That gives rise to a Z-score of 5, which indicates a "possibly significant" alignment according to (Doolittle, 1981, Doolittle, 1986, Lipman & Pearson, 1985, Pearson, 1990). Perhaps a more convincing support for this alignment is the fact that structurally, both proteins exist as an α-helical transmembrane segment, and functionally, both proteins contain bacteriochlorophyll-binding residues. Also, two short segments of the β-apoprotein are highly homologous to two corresponding segments in the L chain of the photosynthetic reaction center from *Rb. sphaeroides* [see Figure 4]. Reaction center L chain sequence 122:133 AFAILAYLTLVL is located in the center of the transmembrane helix C (Michel et al., 1986), which exhibits a close correspondence to our secondary structure assignment of the transmembrane segment for the β-apoprotein. Sequence WVKLPWW is near the C-terminal of the reaction center L chain, and so is the corresponding sequence WVWKPWF in the β-apoprotein.

A secondary structure analysis was performed for all the homologous proteins listed in Figure 4 with the method of Kabsch and Sander which defines secondary structure of a protein by means of pattern recognition of hydrogen-bonded and geometrical features (Kabsch & Sander, 1983). The resulting secondary structure assignment for all the homologous fragments is listed in Table S4 along with PDB BLAST search results. As shown in Table S4, all the homologous fragments located within the putative transmembrane region, except Eco Ri Endonuclease, exist in an α-helical conformation.

We not only used homology modeling to verify the secondary structure assignment, but also used it to improve our secondary structure assignment. In case that no clear-cut secondary structure assignment can be made, the secondary structural features of the homologous structure was employed to establish the secondary structure identity of the α- and β-apoproteins.

Figure 4: Alignment of Homologous Sequences. 1R1E: Eco Ri Endonuclease (E.C.3.1.21.4) complex with TCGCGAATTCGCG; 2RCR: Photosynthetic reaction center from Rhodobacter sphaeroides; 1ACB: Alpha-chymotrypsin (E.C.3.4.21.1) complex with Eglin C; 1CPC: C-Phycocyanin; 1TYP: Trypanothione reductase (E.C.1.6.4.8); 1AAM: Aspartate aminotransferase (E.C.2.6.1.1) mutant; 1GRA: Glutathione reductase (E.C.1.6.4.2) (oxidized) complex; 256B: Cytochrome b562 (oxidized); 1EPS: 5-enol-pyruvyl-3-phosphate synthase (E.C.2.5.1.9); 2BBQ: Thymidylate synthase (E.C.2.1.1.45) complex.

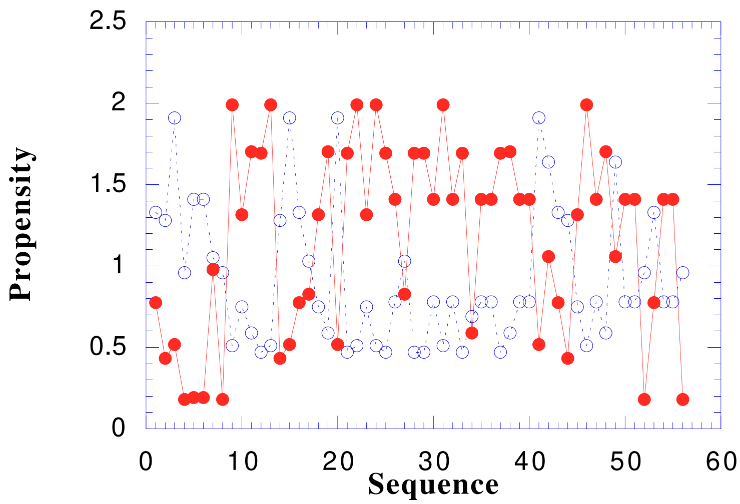

Figure 5: Propensity profile of transmembrane helix and reverse-turn for the α-apoprotein. Solid circles - transmembrane helix propensity base on Persson and Argos scale (Persson & Argos, 1994); open circles - reverse-turn propensity by Levitt(Levitt, 1978).

Specifically, the N and C termini of the transmembrane helix for the α-apoprotein were set to Ser-16 and Alα-41, respectively, in analogy to the homologous transmembrane helix D of the reaction center M subunit and in consideration of the known fact that all residues in the NPS (residues 14:16 in the α-apoprotein) and the PGSN (residues 41:44 in the α-apoprotein) segments have a high propensity for a reverse turn [see Figure 5 and (Levitt, 1978). This assignment is essentially consistent with the result of a multiple sequence alignment propensity analysis for the α-apoprotein. Similarly, the C terminus of the transmembrane helix for the β-apoprotein was set to Lys-42 in analogy to the homologous reaction center L subunit sequence 266:272 WVKLPWW and in consideration of the proline residue. The N terminus of the transmembrane helix for the β-apoprotein was determined to be Glu-10 in analogy to the highly homologous glutathione reductase sequence GLTEDEAI. This assignment of the secondary structure at the N terminus is consistent with both SOPM and Holley-Karplus predictions.

The final secondary structure assignment for both the α- and the β-apoprotein is listed in Figure 6. It is a consensus assignment derived from a combination of all the prediction methods used. It should be pointed out that in addition to a transmembrane helix, an interfacial helix of ten residues [α-Ile-46 to α-Alα-55] has also been identified at the C-terminal of the α-apoprotein.

This assignment is supported by three observations: (1) residues in the sequence IALGAAKSAA have a high propensity to form an α-helix as is evident from SOPM and Holley-Karplus analyses [see Tables S2 and S3] and other propensity analyses we have performed; (2) the homologous fragment KALIAAKNKA from trypanothione reductase, as listed in Figure 4, is an α-helix; and (3), as shown in Figure 7, the segment is highly amphipathic and suitable for a location at the interfacial region. The Trp and charged Lys residues face the lipid head group and all the hydrophobic residues face the interior of the membrane. This helical wheel representation of the amphipathic helix can be quantified in terms of hydrophobic moment as put forward by Eisenberg (Eisenberg, 1984, Eisenberg *et al.*, 1984, Eisenberg *et al.*, 1982).

A reverse-turn segment for the α-apoprotein, PGSN [residues 41:44], was

```
                                    α-apoprotein
Topology                TTTTMMMMMMMMMMMMMMMMMMMMTTTT    IIIIIIIII
Sequence  SNPKDDYKIWLVINPSTWLPVIWIVATVVAIAVHAAVLAAPGSNWIALGAAKSAAK
Secondary CCCCCCCCEEEEECCHHHHHHHHHHHHHHHHHHHHHHHHHCCCCHHHHHHHHHHHC
structure

                                    β-apoprotein
Topology                                  TTTTMMMMMMMMMMMMMMMMMMMMTTTT
Sequence             AERSLSGLTEEEAIAVHDQFKTTFSAFIILAAVAHVLVWVWKPWF
Secondary structure  CCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCC
```

Figure 6: Assignment of Secondary Structure and Topology. M - Transmembrane region; T - Terminal (interfacial) region; I - Interfacial helix.

assigned based on a propensity analysis with Levitt's scale (Levitt, 1978). Shown in Figure 5 is the propensity profile of the transmembrane helix and a reverse-turn for the α-apoprotein. Solid circles indicate the transmembrane helix propensity (for a single sequence) based on the Persson and Argos scale (Persson & Argos, 1994); open circles represent the reverse-turn propensity according to Levitt (Levitt, 1978). All four residues show a reverse-turn propensity higher than one. In summary, we have identified a [transmembrane helix - reverse-turn - interfacial helix] motif for the α-apoprotein. Such a motif has been observed also in other membrane proteins (Kühlbrandt *et al.*, 1994, White, 1994).

Based on the predicted secondary structural features, the tertiary structures for both the α- and the β-apoprotein were built by means of comparative modeling using the program MODELLER (Sali & Blundell, 1993). In comparative modeling, the homologous structure is used as a template for the unknown structure. In our case, homologous fragment structures, as aligned in Figure 4, were employed for such procedure. MODELLER (Sali & Blundell, 1993) obtains the three dimensional model of an unknown protein by satisfying spatial restraints in the form of probability density functions (pdfs) derived from the alignment of the unknown structure with one or more homologous structures. The pdfs restrain C_a - C_a distances, main-chain N-

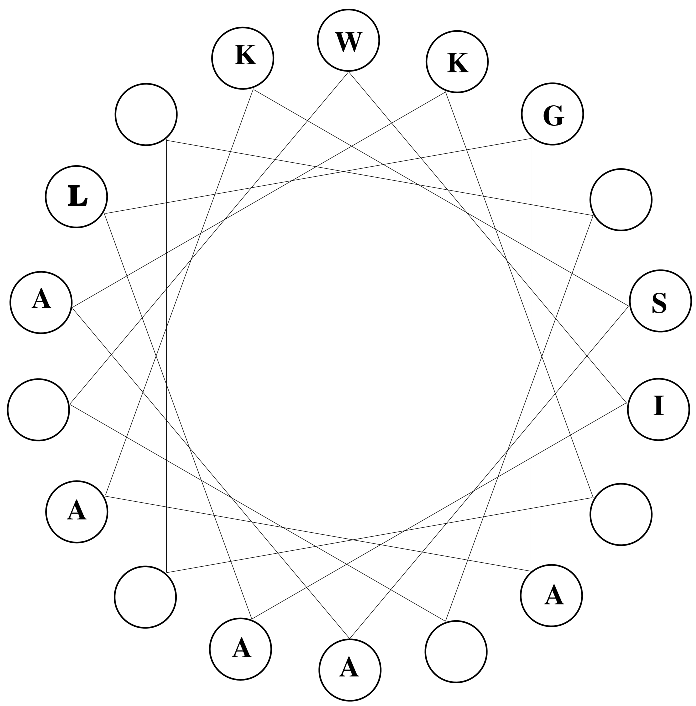

Figure 7: Helical wheel representation of the interfacial α-helix from α-Trp-45 to α-Lys-56.

O distances, main-chain and side-chain dihedral angles. The optimization is carried out by the variable target method that applies the conjugated gradient algorithm to positions of all non-hydrogen atoms. In the present case, the transmembrane helical segments of the α-apoprotein and of the β-apoprotein, together with the interfacial helix of the α-apoprotein were built with the program MODELLER. To build the [transmembrane helix - reverse-turn - interfacial helix] motif for the α-apoprotein, the transmembrane helix and the interfacial helix are optimally linked with the reverse-turn fragment. In the normal practice of comparative modeling, the reverse-turn fragment is selected from a fragment library to generate the best fit between main secondary structure segments. Here, the reverse-turn fragment was built by superimposing two homologous fragments AVPGA and APGGN as listed in Figure 4. The rest of the terminal residues were also added using the corresponding homologous structures as templates. It is our obligation to point out that sequence homology between short ($\leq$ 5 residues) fragments may not be used as basis for structure identity. The terminal residues and the reverse turn segment were modeled this way simply because it is a better alternative than assuming an extended conformation. The α- and the β-apoproteins so constructed are each optimized by energy minimization with fixed protein backbone.

BChla is an integral part of the pigment-protein complex. Placement of BChla in the complex is essential to model building. Resonance Raman spectra demonstrated that all Mg^{2+} of all BChla molecules in the LH-II complex of *Rs. Molischianum* are 5-coordinate (Germeroth *et al.*, 1993). Multiple sequence alignment of light-harvesting complexes (see Table S1 in the supplementary material) shows three highly conserved histidine residues. Most likely, the histidines form the binding site for the three BChla's, a hypothesis supported by site-directed mutagenesis experiments on related systems (Bylina *et al.*, 1988, Crielaard *et al.*, 1994, Visschers *et al.*, 1994, Zuber, 1986). As can be seen from sequence alignment (Table S1), the motif α-Alα-30 ... α-His-34 ... α-Leu-38 is highly conserved (see Figure 8). It has been reported that the corresponding Ala (-4 residues of the conserved His) in LH-I of *Rb. capsulatus* can only be substituted with small residues, Gly, Ser and Cys (Bylina *et al.*, 1988), indicating the structural importance of "smallness" of the residue in the -4 position to His. A similar binding pocket β-Phe-27... β-Alα-31...β-His-35 is also highly conserved (see Figure 8) in the β-apoprotein. The conserved Phe residue in the -8 position to His is also important (Zuber & Brunisholz, 1991). However, there exists no such conserved binding pocket for β-His-17, making the assignment of β-His-17 as the binding site for B800 BChla less certain.

Residue conservation can not only provide clues about the helix-pigment interactions, but also provide clues about helix-lipid interactions. The lipid facing residues tend to be more evolutionarily variable than internal residues (Donnelly *et al.*, 1993, von Heijne & Manoil, 1990). The highly conserved residues, for the α-apoprotein (α-Alα-30 ... α-His-34 ... α-Leu-38) are four residues apart and thus aligned on the same side of the helical wheel, and so are the β-apoprotein (β-Phe-27... β-Alα-31...β-His-35). Taking into account the conserved residues in both the α- and the β-apoprotein, the interhelical arrangement, as shown in Figure 8, is the most probable one. The side of highly conserved residues, on both the α- and the β-apoprotein, face inward and the other side is in contact with lipid bilayers.

As we built the heterodimer, the following well-known facts were also taken into consideration: (1) based on the observation of cleavage of part of the N-terminal domains of the α- and β-apoproteins of LH-I of *Rs. rubrum* on the cytoplasmic side of the membrane by partial hydrolysis with protelytic enzymes (Brunisholz *et al.*, 1984), both α- and β-apoproteins should be oriented with their N-termini towards the cytoplasm; (2) strong circular dichroism (CD) signals suggest exciton interactions between pairs of BChla molecules (Zuber, 1993 and references therein); (3) linear dichroism (LD) data indicate that the B850 Bchla's are approximately perpendicular to the membrane and the B800 BChla is parallel to the membrane (Kramer *et al.*, 1984); (4) Fourier-Transform Resonance Raman spectroscopy and site-directed mutagenesis (Fowler *et al.*, 1994) of a related light-harvesting complex indicate that another highly conserved residue, α-Trp-45, is hydrogen-bonded to the 2-acetyl group of BChla. All these observations impose constraints on the structure: the $\alpha\beta$ heterodimer should be arranged interhelically with two B850 BChla binding histidine residues facing each other as shown in Figure 8; the two B850 BChla's are paired and oriented perpendicular to the membrane plane. Thus, the basic unit of the LH-II complex is configured with the B850 BChla pair sandwiched between two helices of the $\alpha\beta$ heterodimer. We will return to the consequence of hydrogen bond constraints in the following section.

The construction of the complete aggregated complex was based on the two stage model suggested in (Popot & Engelman, 1990) which assumes that the individual helices are formed prior to the formation of the helix bundle. We developed a protocol to aggregate the transmembrane helices into an octamer of eight $\alpha\beta$ subunits by means of molecular dynamics simulations and energy minimization under the constraints of experimental data (see Figure 1 for an outline of the entire procedure). Our procedure consists of three essential steps: In a first step, the α- and β-apoproteins were constructed by comparative modeling based on information obtained through homology and secondary structure analyses as described above. The optimized tertiary structures for the α- and the β-apoprotein were preserved in the subsequent simulations thereafter by applying for the helical backbone harmonic restraints to the distances between the *i*-th carbonyl and the *(i+4)*-th amide nitrogen and to the distances between the *i*-th and the *(i+4)*-th C_a; by applying in the turn region the harmonic restraint to the dihedral angles of the main chain. No restraint was applied to coiled terminal residues.

The harmonic potential

$$E(R) = \frac{K_B T}{2C_{ij}^2}(R - R_0)^2 \tag{1}$$

and the harmonic function

$$E(\phi) = C_{well}(\phi - \phi_0)^2 \tag{2}$$

were utilized for distance restraints and dihedral angle restraints (Brünger, 1992). In our procedure, the strengths of restraining potentials were set to relatively weak values by choosing C_{ij} = 0.2 Å and C_{well} = 20.0 kcal $mol^{-1}rad^{-2}$. The equilibrium distances (R_o) and dihedral angles (f_o) were extracted from the optimized structures of the α- and the β-apoprotein built by MODELLER. Two BChla's binding to β-His-17 and β-His-35 were manually attached to the β-apoprotein using the program QUANTA, and a third BChla was attached similarly to α-His-34. We employed the heavy atom coordinates of BChla from the crystal structure of BChlb in the photosynthetic reaction center of *Rps. viridis* (Deisenhofer *et al.*, 1985) and added explicit hydrogens using X-PLOR's function *hbuild*. The binding conformation between BChlb and His in the crystal structure of *Rps. viridis* was employed for the placement of the BChla's. The X-PLOR utility *patch* was used to build the ligand bond between magnesium and the nitrogen of His, and was followed by rigid body minimization between the apoprotein and the BChla's. Subsequently, energy minimization runs were performed with harmonic restrains, followed by three 1 ps molecular dynamics runs at consecutively increasing temperatures of 100, 200 and 300 K to equilibrate the system. The lycopenes were not included in the model structure.

In the second step, the complete octamer was constructed enforcing an eight-fold symmetry. A self rotational search (Lattman, 1985, Brünger, 1992, McRee, 1993) of the X-ray diffraction data indicated that, most likely, the LH-II complex possessed an eightfold symmetry. Thus, the task of constructing the octamer was reduced to building a protomer for the eightfold octamer. There are two possible choices for the protomer: the $\alpha\beta$ heterodimer or a neighboring $\alpha\beta$ pair as shown in Figure 9. Since we intended to perform an intermediate step optimization for the $\alpha\beta$ dimer, the neighboring $\alpha\beta$ pair was chosen as the protomer. The rationale behind our choice is that in the $\alpha\beta$ heterodimer ($\alpha_2\beta_1$) in Figure 9, the two helices separated by the B850 BChla pair are too far apart to generate any significant interactions to be optimized. In contrast, in the neighboring $\alpha\beta$ pair ($\alpha_1\beta_1$) in Figure 9, inter-unit hydrogen bonds exist. As mentioned above, the α-Trp-45 is hydrogen-bonded to the 2-acetyl group of BChla. However, there is no direct information about which of the two B850 BChla's is involved in such a hydrogen bond (Fowler *et al.*, 1994). Based on the optimized tertiary structure for both the α- and the β-apoprotein, we found that it is spatially more feasible to form an inter-unit hydrogen bond, i.e., a hydrogen bond between α-Trp-45 and 2-acetyl group of BChla binding to β-His-35. Since open hydrogen bonds are extremely unstable in a lipid environment, we also

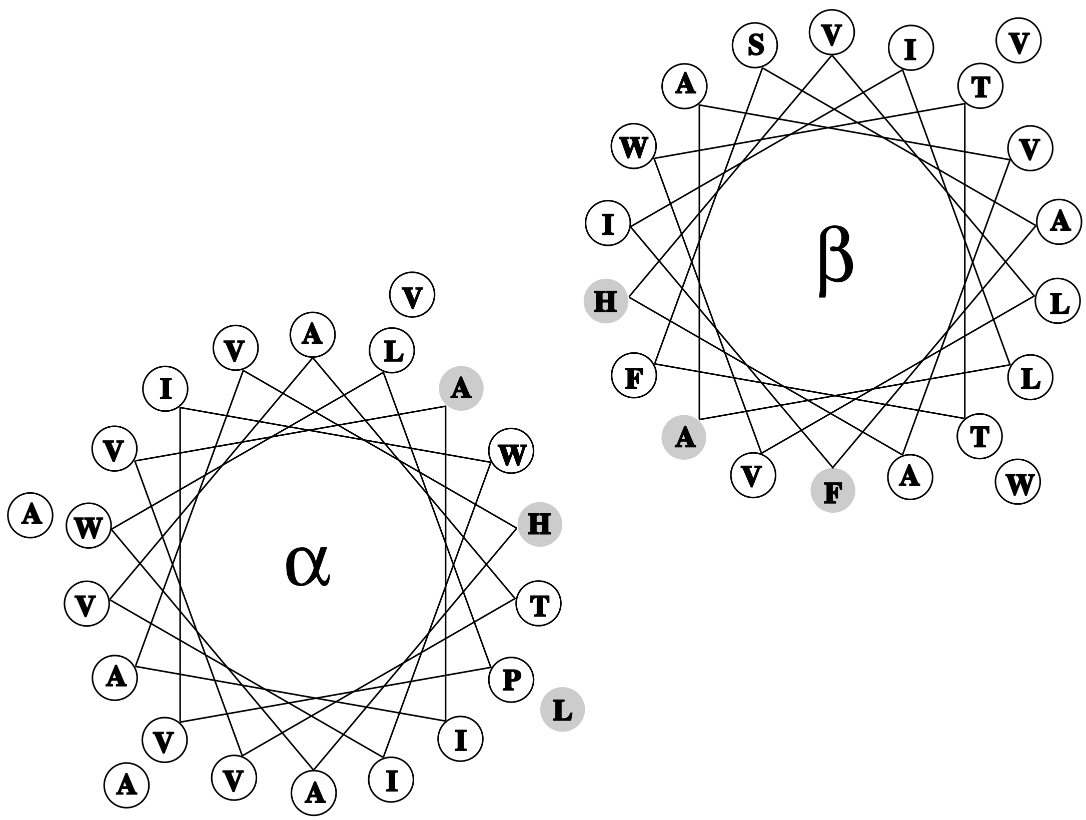

Figure 8: Proposed interhelical arrangement of the α- and β-apoproteins. The residues in shaded circles are highly conserved.

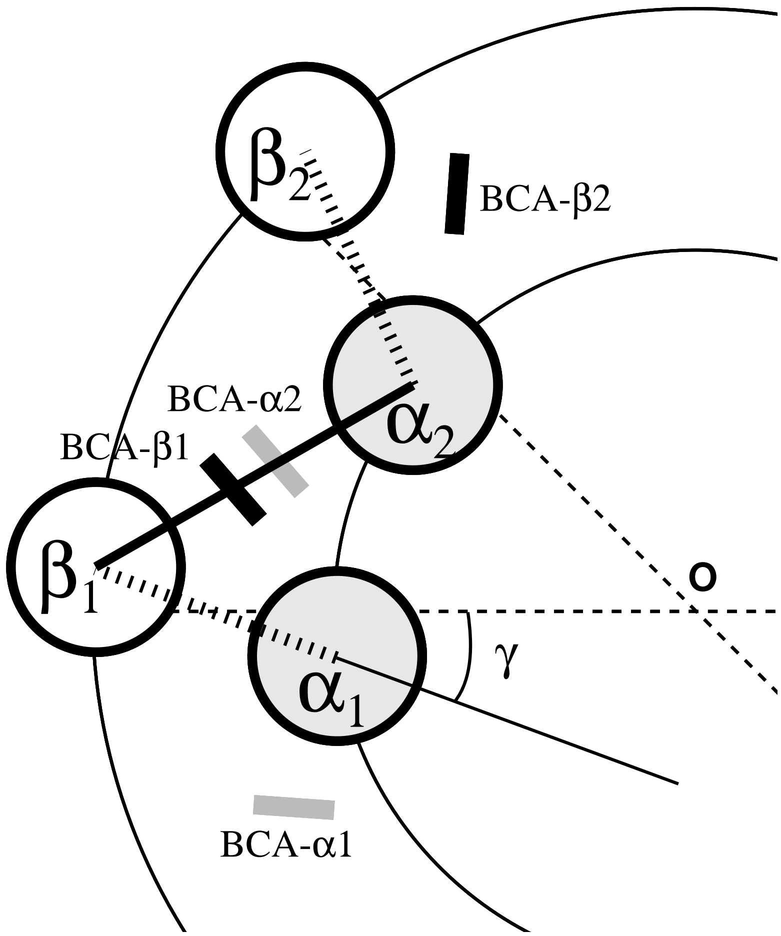

Figure 9: Definition of $\alpha\beta$ pair geometries. Heterodimer: $\alpha\beta$ dimer formed between β_1 and α_2, linked by a solid line; neighboring pair: $\alpha\beta$ dimer formed between β_1 and α_1, linked by a dashed line; γ is defined as the angle between the vector linking two helices and the radial vector from the center of the octamer to the center of the neighboring pair.

arranged to have β-Trp-41 hydrogen bonded to the 2-acetyl group of BChla binding to α-His-34. We used QUANTA to place the neighboring pair ($a_1 b_1$ as shown in Figure 9) and arranged the $\alpha\beta$ heterodimers in a configuration as shown in Figure 8 and established hydrogen bonds between α-Trp-45 and the 2-acetyl group of BChla attached to β-35 His, as well as between β-Trp-41 and the 2-acetyl group of BChla binding to α-His-34. We applied the "soft van der Waals" option in X-PLOR to minimize the energy content of the neighboring $\alpha\beta$ pair, followed by rigid body minimization with two helices and three BChla's as five rigid bodies. Thereafter, simulated annealing was applied by heating the system to 2000 K and slowly cooling it down. This procedure was followed by a 100 ps dynamics run at 300 K, obtaining an equilibrated structure of the neighboring $\alpha\beta$ pair.

In a third step, the equilibrated neighboring $\alpha\beta$ pairs with all three bound BChla's were aggregated into an octamer by means of long time molecular dynamics simulations followed by energy minimization with eight-fold symmetry and all the restraints as described above enforced throughout the entire simulation. An iterative protocol consisting of multiple cycles was employed to optimize the octamer structure. Each cycle started with a 200 step rigid body minimization (with the entire protomer as a rigid body), followed by a 2.5 ps rigid body dynamics run at 600 K, then a 5 ps molecular dynamics run at 300 K, ending with a 200 step Powell minimization. The radius of the octamer was monitored to detect convergence. Initially, the octamer was constructed with an outer diameter as large as 100 Å to avoid any close inter-unit contact. During the first cycle, a larger cutoff distance for non-bonded interactions was required (we chose 15 Å), since the initial inter-unit distances are large. The iterative process was terminated when the final radii at the end of two consecutive runs differed by less than 0.25 Å. At the end, the octamer was minimized again with the Powell algorithm until it converged to a energetically optimal configuration or until a limit of 700 cycles was reached. It took about 50-100 ps of equilibration (10 to 20 iterations) before the radii converged. Test runs with longer times indicated that the radius of the octamer began to fluctuate around its average value after 50 ps dynamics. This simulation protocol is based on energy minimization perturbed by successive dynamic equilibration, which can be viewed as a dynamical analogy to simulated annealing.

Protein folding is related to the problem of global minimization, the solution of which relies heavily on proper initial configurations. Starting from a heterodimer built within the constraints of all the biochemical and spectroscopic data dramatically reduces the phase space needed to be sampled, enhancing the chance of steering a structure into the basin of attraction of the global minimum. In the present case, a systematic sampling of initial configurations has been attempted by arranging two different helical assemblies and by varying the relative orientations (angle γ in Figure 9) of the vector linking two helices and the radial vector from the center of the octamer to the center of a neighboring pair. There are two alternatives to assemble the eightfold octamer ring: one may place the α-apoprotein inside and the β-apoprotein outside or the other way around. Since there exist no observations in favor of one or the other, both arrangements were considered in our modeling. Angle γ in Figure 9 had been sampled from -50^o to 50^o in steps of 10^o. Altogether, a total of 22 octamers were constructed and optimized. Twelve of the 22 optimized structures were rejected either because the size of the octamer was too large to fit the unit cell or because of a severe distortion of the pairing of B850 BChla's. The rest of the optimized structures were fed into the molecular replacement procedure. The structures were evaluated as described in the following section.

It is worth mentioning that the optimized octamers were not selected based on the criterion of energy minimization. Instead, the molecular replacement test was employed as the ultimate test of the correctness of the model. The predicted structures were then employed as a search model in the framework of the molecular replacement method as implemented in the program X-PLOR (Brünger, 1992). In the molecular replacement method, a six-dimensional search is required to find the best match between observed and calculated diffraction data. In practice, the method is implemented as a three-dimensional rotational search followed by a three-dimensional translational search (Lattman, 1985, Rossmann, 1972). We also performed a Patterson correlation refinement preceding the translational search to filter the peaks of the rotational search (Brünger, 1990). Four of the remaining ten structures were capable of probing the position of molecules in the crystal. The rotational search oriented the a helices of the octamers in a direction parallel to the c-axis of the crystal unit cell. The translational search further placed the octamers in the site of the four-fold axis of the crystal. The two optimal structures resulting from this search are provided in the electronic supplementary material in the X-PLOR PDB format. Both are ring structured octamers, one with β-apoproteins inside and α-apoproteins outside [labeled as octamer_1.pdb] and another with α-apoproteins inside and β-apoproteins outside [labeled as octamer_2.pdb]. With 6-12 Å resolution data, the R value after rigid body refinement for the former structure is 49.8% and decreases to 29.8% after positional refinement, and the values for the latter structure are 50.45% and 30.3%, respectively.

Shown in Figure 10(a) is one of the two best octamers resulting from the molecular replacement search. The octamer forms a ring with an outer diameter of about 70 Å with the α-apoprotein on the inside and the β-apoprotein on the outside. The 70 Å outer diameter includes the side chains of the β-apoprotein. On average, the radius of the eight β-apoproteins (measured from the center of the helices) outer ring is 33 Å; the radius of the eight α-apoproteins inner ring is 21 Å. The Mg to Mg separation in the $\alpha\beta$ heterodimer is about 8.8 Å. The Mg to Mg separation between adjacent heterodimers is 16.0 Å. The inner helices are tilted against the membrane plane normal by about 6^o and the outer helices are tilted by about 9^o. The resulting BChla's are optimally oriented to capture light coming from all directions and are close enough for efficient energy transfer. As depicted in Figure 10(b), the B850 BChla pair were sandwiched between two helices of the $\alpha\beta$ heterodimer. Such an arrangement has been suggested for the LH-I complex of *Rs. rubrum* to interpret the ring-shaped 8.5 Å resolution projection map (Karrasch *et al.*, 1995). It should be pointed out that our model shows the expected placement of the two B850 Bchla's perpendicular to the membrane, near the periplasmic side. However, the B800 BChla is not oriented parallel to the cytoplasmic side of the membrane as expected from LD data due to the lack of helix turn residues near β-His-17.

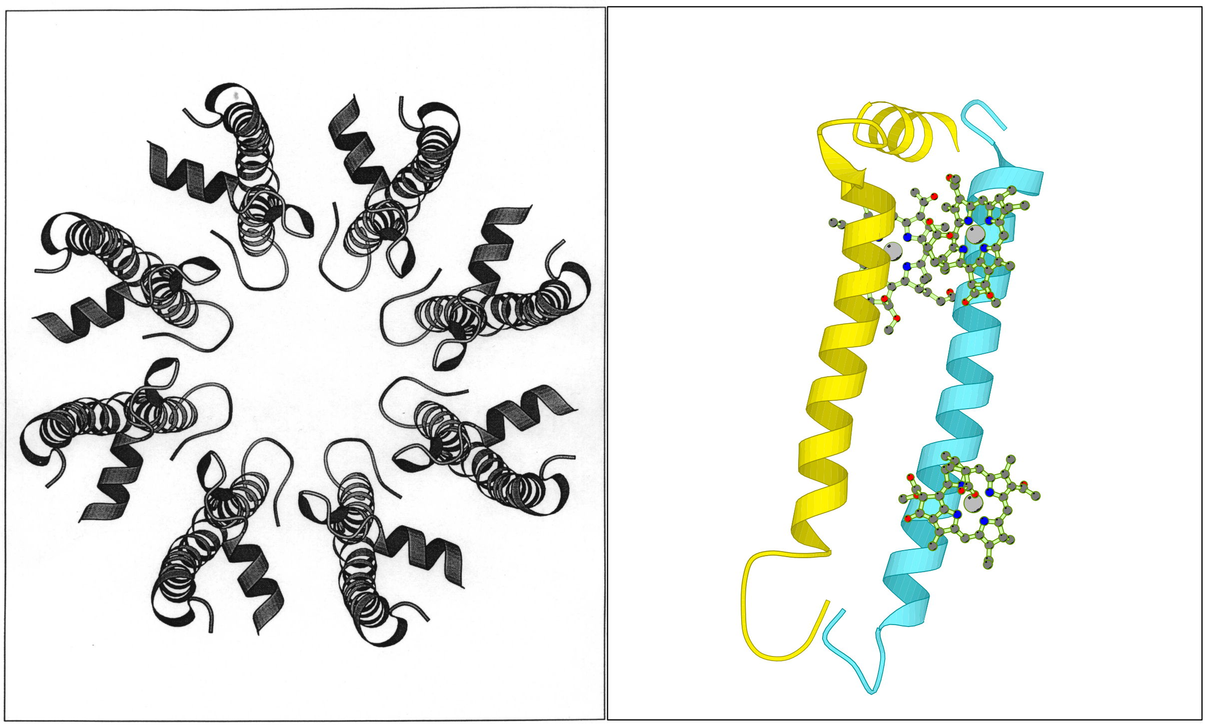

Figure 10: Schematic structure of the LH-II complex with coordinates taken from the file octamer_2.pdb in the supplementary material. (a) Octamer (top view with C terminus pointing upward) (b) $\alpha\beta$ heterodimer, i.e., the monomer "unit" of LH-II. Shown are the C_a trace and locations of three Bchla's binding to α-34 His, β-17 His and β-35 His. The C terminus is on top [produced with MOLSCRIPT(Kraulis, 1991).

Summary

In this article, we report our efforts in predicting structure of the LH-II complex of *Rs. molischianum*. Presently, prediction of tertiary structure is only of practical use when the structure of a homologous protein is already known (Hilbert *et al.*, 1993, Fasman, 1989). We illustrate how one can proceed to predict the structure of an integral membrane protein in case that a highly homologous structure is not available. Figure 1 outlines the structure prediction processes which we have adopted. Secondary structures were determined using an approach which combines hydropathy analyses to identify the putative transmembrane segments and multiple sequence alignment propensity analyses to pinpoint the transmembrane segment and the four residue long terminal sequence at both ends, which were independently verified and improved by homology modeling. Hydropathy analyses showed that it is energetically favorable to form a transmembrane segment for the α- and the β-apoprotein. Propensity analyses demonstrated that the putative transmembrane segments have a high tendency statistically to be an α-helical membrane spanner. The homologous fragments of the putative transmembrane core are mostly α-helices, which is in good agreement with the reported success of a secondary structure prediction algorithm based on the hypothesis that short (≥ 7 residues) homologous sequences of amino acids have identical secondary structure tendencies (Levin *et al.*, 1986). This similarity principle has also been used to determine the ends of transmembrane helices for the α- and the β-apoprotein. Our approach combining several prediction schemes is necessary since most secondary structure prediction methods are accurate only up to 60-70%. A consensus assignment for the secondary structure was derived from a combination of all the prediction methods used. We have identified a [transmembrane helix - reverse-turn - interfacial helix] motif for the α-apoprotein. Biochemical and spectroscopic data were gathered and analyzed to extract structural constraints. Tertiary structures for the α- and the β-apoprotein were built by comparative modeling with constraints derived from the predicted secondary structural features and the homologous fragments. The resulting tertiary structures were combined into an intermediate form, i.e. an $\alpha\beta$ dimer pair with BCh-

la's attached, under the constraints of biochemical and spectral data. We also took into account factors such as residue conservation, pigment-protein interactions and protein-lipid interactions. Molecular dynamics simulations and energy minimization were performed to form the aggregated LH-II complex. Available, unphased, X-ray diffraction data served to test ten predicted structures by molecular replacement method. The structure of LH-II so determined is an octamer of $\alpha\beta$ heterodimers forming a ring with a diameter of 70 Å.

Structure prediction of integral membrane proteins shows great promise and merits further investigation. While thousands of membrane protein sequences are available, there are few detailed three dimensional membrane protein structures known. Obvious, there is a great need to predict membrane protein structures, with a good opportunity for success. Due to constraints of the hydrophobic core of the membrane, the transmembrane segments of membrane proteins usually adopt two type of conformations: α-helix or β-sheet. The 2-D nature of a membrane provides a strong constrain upon the arrangement of the transmembrane segments that cross it. Most transmembrane helices are either perpendicular to the membrane plane or tilted against membrane plane normal by less than 30^o. It is our hope that the work reported here will be useful to structure prediction of other membrane proteins.

Note Added at Time of Revision

At the time this article is revised, the crystal structure of LH-II from *Rps. acidophila* determined by conventional multiple isomorphous replacement method by McDermott et al. had appeared in press (McDermott *et al.*, 1995). This permits us to compare our prediction to an observed structure, albeit not for the same protein.

As shown in Table S5, there exists a significant sequence homology between LH-II of *Rps. acidophila* strain 10050 and *Rs. molischianum* for both the α- and the β-apoprotein. Figure 11, compare the predicted secondary structure assignment for *Rs. molischianum* with the X-ray resolved secondary structure for *Rps. acidophila* (McDermott *et al.*, 1995). The pre-

Figure 11: Comparison of Secondary Structure. H - Helix; E - b Sheet; C - Coil; 3 - 3 turn helix; ? - Unresolved. Secondary structure of LH-II of Rps. acidophila extracted from (Mcdermott et al., 1995).

dicted secondary structure of LH-II of *Rs. molischianum* compares well with that of the *Rps. acidophila* crystal structure. We correctly identified the transmembrane helices for both the α- and the β-apoprotein. In particular, the predicted [transmembrane helix - reverse-turn - interfacial helix] motif for the α-apoprotein is observed in the crystal structure.

It should be pointed out that the aggregated LH-II complex of *Rs. molischianum* consists of eight $\alpha\beta$ heterodimers while that of *Rps. acidophila* consists of nine heterodimers. The overall helical assembly of the LH-II complex of *Rs. molischianum* described in the previous sections corresponds well with that of *Rps. acidophila*. The LH-II complex of *Rps. acidophila* is a ring-shaped aggregate of nine $\alpha\beta$ heterodimers with nine-fold symmetry. The transmembrane helices of nine α-apoprotein are packed side by side to form a hollow cylinder of radius 18 Å. The nine helical β-apoproteins are arranged radially with the α-apoproteins to form an outer cylinder of radius 34 Å. The α-apoprotein helices are parallel to the ninefold axis to within 2^o, and the β-apoprotein helices are inclined by 15^o to this axis (McDermott *et al.*, 1995).

The detailed three dimensional structures of the protomers, which con-

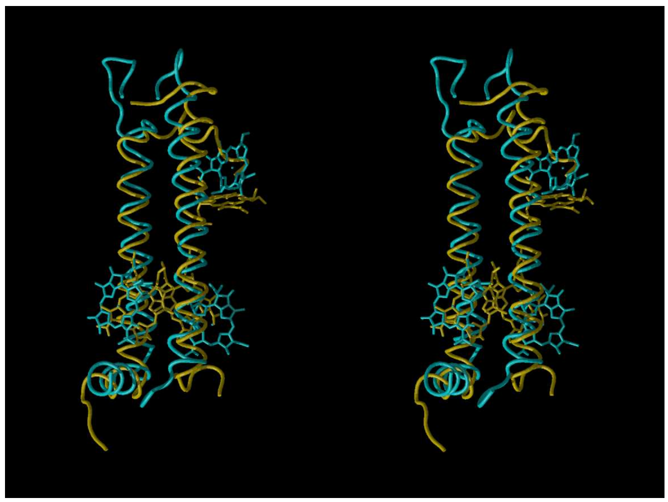

Figure 12: Stereos of the protomers from the predicted LH-II complex of *Rs. molischianum* (colored in cyan, coordinates taken from the file octamer_2.pdb in the supplementary material) and the X-ray resolved LH-II complex of *Rps. acidophila* (colored in yellow, coordinates courtesy of Cogdell and co-workers, permission sought for publication) [produced with VMD].

sists of a pair of the α- and the β-subunit, from the predicted LH-II complex of *Rs. molischianum* and the X-ray resolved LH-II complex of *Rps. acidophila* are compared in Figure 12. The two structures are aligned using the Kabsch least square fitting algorithm (Kabsch, 1976) as implemented in X-PLOR to obtain the best fit between transmembrane segments of both the α- and the β-apoprotein. The fit resulted in a rms deviation of 2.9 Å on the average for C_a atoms of all the transmembrane residues. This deviation arises from two factors: (1) the $\alpha\beta$ pair in the predicted structure is more closely packed than the $\alpha\beta$ pair in the crystal structure; (2) the β-apoprotein is dislocated relatively by about half a helical turn. Individually, an independent fit shows that the tertiary structures of both the α- and the β-apoprotein between two structures corresponds well to each other. For the α-apoprotein, the averaged rms deviation for C_a atoms of the transmembrane helical core is 1.1 Å. The [transmembrane helix - reverse-turn - interfacial helix] motives of the α-apoprotein from both structures exhibit a close similarity in both position and orientation. For the β-apoprotein, the averaged rms deviation for C_a atoms of the transmembrane helical core is 1.7 Å. The main discrepancy is the placement of BChla's. While our predicted structure places the B850 BChla pair between the $\alpha\beta$ heterodimer as shown in Figures 10(b) and 12, the X-ray structure shows that all B850 BChla's are arranged such that the chlorophylls in the pair overlap the chlorophylls of the two neighboring pairs and that a ring of overlapping BChla's is formed. As expected, the B800 BChla is oriented parallel to the membrane plane in the crystal structure. A surprise is the binding of the B800 BChla to the formyl group of the fMet residue at the N terminus of the α-apoprotein which, for the purpose, has to dive into the interior of the membrane by as much as 9 Å. As stated before, our difficulties in placing the B800 BChla are attributed to a lack of pertinent biochemical information. It should also be born in mind that although in the majority of LH-II, even LH-I complexes, there exists a corresponding Met residue at the N terminus of the α-apoprotein, the N terminus of the α-apoprotein of *Rs. molischianum* is a Ser residue. As a result, a slightly different conformation for *Rs. molischianum* is expected. The difference in the number of the constituting $\alpha\beta$ dimers (eight vs nine) is a clear indication of a difference as well. The extent of this difference will be clear once the structure of LH–II from *Rs. molischianum* has been determined by x-ray crystallography at 2.4 Å resolution.

Supplementary Material in Electronic Appendix

The Electronic Appendix contains a sequence alignment (Table_S1) of the α- and β-apoproteins of the LH-II complex of *Rs. molischianum* with α- and β-apoproteins of LH-II and LH-I complexes from other photosynthetic bacteria. Table_S2 and Table_S3 show SOPM and Holley-Karplus predictions of secondary structure, respectively. Table_S4 lists detailed PDB BLAST search results for both the α- and β-apoproteins. Table_S5 shows sequence alignment between LH-II of *Rs. molischianum* and *Rps. acidophila.* Figure_S1 and Figure_S2 plot hydropathy profiles based on the Kyte & Doolittle scale and the Eisenberg consensus scale, respectively. Two PDB files are provided: octamer_1.pdb is for an optimized LH-II complex with β-apoproteins inside and α-apoproteins outside and octamer_2.pdb is for an optimized LH-II complex with α-apoproteins inside and β-apoproteins outside.

Acknowledgments

This work was supported by the Carver Charitable Trust and the National Institute of Health (NIH grant P41RR05969). KS is very grateful to Prof. Richard J. Cogdell and co-workers for making the coordinates of X-ray resolved LH-II structure of *Rps. acidophila* available to us while this article was under revision. We would like to thank Prof. Antony Crofts and Dr. Zan Schulten for many helpful discussions.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Barton, G. & Russell, R. (1993). Pretein structure prediction. *Nature,* **361**.

Benner, S., Cohen, M., & Gerloff, D. (1992). Correct structure prediction? *Nature,* **359**.

Boonstra, A. F., Visschers, R. W., Calkoen, F., van Grondelle, R., van Bruggen, E. F., & Boekema, E. J. (1993). Structural characterization of the B800-850 and B875 light-harvesting antenna complexes from *Rhodobacter-Sphaeroides* by electron microscopy. *Biochimica et Biophysica Acta,* **1142**, 181.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., vid J. States, D., Swaminathan, S., & Karplus, M. (1983). CHARMm: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **4** (2), 187–217.

Brünger, A. T. (1990). Extension of molecular replacement: A new search strategy based on Patterson correlation refinement. *Acta Cryst.* **A46**, 46–57.

Brünger, A. T. (1992). *X-PLOR, Version 3.1, A System for X-ray Crystallography and NMR.* The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University.

Brunisholz, R., Wiemken, V., Suter, F., Bachofen, R., & Zuber, H. (1984). The light-harvesting polypeptides of rhodospirillum rubrum. ii. localisation of the amino-terminal regions of the light-harvesting polypeptides b 870-alpha and b 870-beta and the reaction-centre subunit l at the cytoplasmic side of the photosynthetic membrane of rhodospirillum rubrum g-9+. *Hoppe-Seylers Zeitschrift fur Physiologische Chemie,* **365**.

Bylina, E., Robles, S., & Youvan, D. (1988). Directed mutations affecting the putative bacteriochlorophyll-binding sites in the light-harvesting I antenna of *Rhodobacter capsulatus. Israel Journal of Chemistry,* **28**, 73–78.

Cogdell, R. & Hawthornthwaite, A. (1993). Preparation, purification, and crystallization of purple bacteria antenna complexes. In: *The Photosynthetic Reaction Center*, (Deisenhofer, J. & Norris, J., eds) p. 23, San Diego: Academic Press.

Cornette, J., Cease, K., Margalit, H., Spouge, J., Berzofsky, J., & DeLisi, C. (1987). Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. *Journal of Molecular Biology,* **195**, 659.

Crielaard, W., Visschers, R., Fowler, G., van Grondelle, R., Hellingwerf, K., & Hunter, C. (1994). Probing the B800 bacteriochlorophyll binding site of the accessory light-harvesting complex from *Rhodobacter sphaeroides* using site-directed mutants. I. Mutagenesis, effects on binding, function and electrochromic behaviour of its carotenoids. *Biochim. Biophys. Acta,* **1183**, 473–482.

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

Deisenhofer, J. & Michel, H. (1989). The photosynthetic reaction centre from the purple bacterium *Rhodopseudomonas viridis. EMBO J.* **8**, 2149–2170.

Devereux, J., Haeberli, P., & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the vax. *Nucleic Acids Research,* **12**, 387.

Donnelly, D., Overington, J. P., Ruffle, S. V., Nugent, J. H., & Blundell, T. L. (1993). Modeling α-helical transmembrane domains: the calculation and use of substitution tables for lipid-facing residues. *Protein Science,* **2**, 55.

Doolittle, R. (1981). Similar amino acid sequences: Chance or common ancestry. *Science,* **214**.

Doolittle, R. (1986). *Of URFS and ORFS: A primer on how to analyze derived amino acid sequences.* Mill Valley, CA:.

Eisenberg, D. (1984). Three-dimensional structure of membrane and surface proteins. *Annual Review of Biochemistry,* **53**, 595.

Eisenberg, D., Schwarz, E., Komaromy, M., & Wall, R. (1984). Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *Journal of Molecular Biology,* **179**, 125.

Eisenberg, D., Weiss, R., Terwilliger, T., & Wilcox, W. (1982). Hydrophobic moments and protein structure. *Faraday Symposia of the Chemical Society,* **17**, 109.

Engelman, D. M., Steitz, T. A., & Goldman, A. (1986). Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Ann. Rev. Biophys. Biophys. Chem.* **15**, 321–353.

Esposti, M. D., Crimi, M., & Venturoli, G. (1990). A critical evaluation of the hydropathy profile of membrane proteins. *European Journal of Biochemistry,* **190**.

Fasman, G. (1989). Protein conformational prediction. *Trends in Biochemical Sciences,* **14**, 295.

Fowler, G., Sockalingum, G., Robert, B., & Hunter, C. (1994). Blue shifts in bacteriochlorophyll absorbance correlate with changed hydrogen bonding patterns in light-harvesting 2 mutants of *Rhodobacter Sphaeroides* with alterations at α-Tyr-44 and α-Tyr-45. *Biochemical Journal,* **299**, 695.

Geourjon, C. & Deleage, G. (1994). SOPM: a self optimised prediction method for protein secondary structure prediction. *Protein Engineering,* **7**, 157.

Germeroth, L., Lottspeich, F., Robert, B., & Michel, H. (1993). Unexpected similarities of the B800-850 light-harvesting complex from *Rhodospirillum molischianum* to the B870 light-harvesting complexes from other purple photosynthetic bacteria. *Biochemistry,* **32**, 5615–5621.

Hawthornthwaite, A. M. & Cogdell, R. J. (1991). Bacteriochlorophyll binding proteins. In: *Chlorophylls*, (Scheer, H., ed) pp. 493–528, Boca Raton: CRC Press.

Hilbert, M., Bohm, G., & Jaenicke, R. (1993). Structural relationships of homologous proteins as a fundamental principle in homology modeling. *Proteins,* **17**.

Holley, L. H. & Karplus, M. (1991). Neural networks for protein structure prediction. *Methods in Enzymology,* **202**, 204.

Jahnig, F. (1990). Structure predictions of membrane proteins are not that bad. *Trends in Biochemical Sciences,* **15**.

Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. *Acta Cryst.* **A32**.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Bioploymers,* **22**, 2577.

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

Kleinekofort, W., Germeroth, L., van der Broek, J., Schubert, D., & Michel, H. (1992). The light-harvesting complex II (B800/850) from *Rhodospirillum molischianum* is an octamer. *Biochimica et Biophysica Acta,* **1140**, 102–104.

Kramer, H. J. M., van Grondelle, R., Hunter, C. N., Westerhuis, W. H. J., & Amesz, J. (1984). Pigment organization of the B800-850 antenna complex of *Rhodopseudomonas sphaeroides. Biochim. Biophys. Acta,* **765**, 156–165.

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

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

Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105.

Lattman, E. (1985). Diffraction methods for biological macromolecules. Use of the rotation and translation functions. *Methods in Enzymology,* **115**, 55–77.

Levin, J., Robson, B., & Garnier, J. (1986). An algorithm for secondary structure determination in proteins based on sequence similarity. *FEBS,* **205**.

Levitt, M. (1978). Conformational preference of amino acids in globular proteins. *Biochemistry,* **17**, 4277.

Lipman, D. & Pearson, W. (1985). Rapid and sensitive protein similarity searches. *Science,* **227**, 1435.

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

McRee, D. (1993). *Practical Protein Crystallography.* San Diego: Academic Press.

Michel, H. (1991). General and practical aspects of membrane protein crystallization. In: *Crystallization of Membrane Proteins*, (Michel, H., ed) pp. 74–88, Boca Raton, Florida: CRC Press.

Michel, H., Weyer, K. A., Gruenberg, H., Dunger, I., Oesterhelt, D., & Lottspeich, F. (1986). The 'light' and 'medium' subunits of the photosynthetic reaction centre from *rhodopseudomonas viridis*: isolation of genes, nucleotide and amino acid sequence. *EMBO J.* **5**, 1149.

Olsen, J. D. & Hunter, C. N. (1994). Protein structure modelling of the bacterial light-harvesting complex. *Photochem. Photobiol.* **60**, 521.

Pearson, W. (1990). Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods in Enzymology,* **183**, 63.

Persson, B. & Argos, P. (1994). Prediction of transmembrane segments in proteins utilising multiple sequence alignments. *Journal of Molecular Biology,* **237**, 182.

Popot, J. & Engelman, D. (1990). Membrane protein folding and oligomerization: the two-stage model. *Biochemistry,* **29**, 4031.

Rossmann, M., ed (1972). *The Molecular Replacement Method*. New York: Gordon and Breach.

Sali, A. & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology,* **234**, 779–815.

Sundstrom, V. & van Grondelle, R. (1991). Dynamics of excitation energy transfer in photosynthetic bacteria. In: *Chlorophylls*, (Scheer, H., ed) pp. 627–704, Boca Raton: CRC Press.

Treutlein, H., Schulten, K., Deisenhofer, J., Michel, H., Brünger, A., & Karplus, M. (1988). Molecular dynamics simulation of the primary processes in the photosynthetic reaction center of *Rhodopseudomonas viridis*.

In: *The Photosynthetic Bacterial Reaction Center: Structure and Dynamics*, (Breton, J. & Verméglio, A., eds) volume 149 of *NATO ASI Series A: Life Sciences* pp. 139–150. Plenum New York.

Visschers, R. W., Crielaard, W., Fowler, G. J., Hunter, C. N., & van Grondelle, R. (1994). Probing the B800 bacteriochlorophyll binding site of the accessory light-harvesting complex from *Rhodobacter sphaeroides* using site-directed mutants. II. a low temperature spectroscopy study of structural aspects of the pigment-protein conformation. *Biochim. Biophys. Acta,* **1183**, 483.

von Heijne, G. & Manoil, C. (1990). Membrane proteins: from sequence to structure. *Protein Engineering,* **4**, 109.

Welte, W., Wacker, T., Leis, M., Kreutz, W., Shiozawa, J., Gadon, N., & Drews, G. (1985). Crystallization of the photosynthetic light harvesting pigment-protein complex B800–850 of *Rhodopseudomonas capsulata*. *FEBS Lett.* **182**, 260.

White, S. H. (1994). Hydropathy plots and the prediction of membrane protein topology. In: *Membrane protein structure: experimental approaches*, (White, S. H., ed) , New York: Oxford University press.

Zuber, H. (1985). Structure and function of light-harvesting complexes and their polypeptides. *Photochem. Photobiol.* **42**, 821–825.

Zuber, H. (1986). Structure of light-harvesting antenna complexes of photosynthetic bacteria, cyanobacteria and red algae. *Trends Biochem. Sci.* **11**, 414–419.

Zuber, H. & Brunisholz, R. (1991). Structure and function of antenna polypeptides and chlorophyll-protein complexes: Principles and variability. In: *Chlorophylls*, (Scheer, H., ed) pp. 627–692, Boca Raton: CRC Press.