Prediction of the Structure of an Integral Membrane Protein–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 illustrate in this chapter how one proceeds to predict the structure of integral membrane proteins when a highly homologous structure is unknown. We focus here on the prediction of the structure of the light-harvesting complex II (LH–II) of *Rhodospirillum molischianum*, an integral membrane protein of 16 polypeptides aggregating and binding to 24 bacteriochlorophyll a's and 12 lycopenes. Hydropathy analysis was performed to identify the putative transmembrane segments, which were independently verified by multiple sequence alignment propensity analyses and homology modeling. A consensus assignment for secondary structure was derived from a combination of all the prediction methods used. Transmembrane helices were built by comparative modeling. The resulting tertiary structures were then aggregated into a quaternary structure through molecular dynamics simulations and energy minimization under constraints provided by site directed mutagenesis and FT Resonance Raman spectra, as well as conservation of residues. The structure of LH–II, so determined, was an octamer of $\alpha\beta$ heterodimers forming a ring with a diameter of 70 Å. We discuss how the resulting structure may be used to solve the phase problem in X-ray crystallography in a procedure called molecular replacement. We will also discuss the exciton structure which results from the circular arrangement of chlorophyls in LH-II.

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

Introduction

Membranes are complex structures containing bilayers of amphiphilic phospholipids with proteins either loosely associated on the surface (peripheral membrane proteins, e.g., phospholipase A_2) or embedded (integral membrane protein, e.g., bacteriorhodopsin). This chapter deals with the structure prediction of integral membrane proteins, defined as proteins having peptide chains with substantial tertiary structure within the nonpolar region of the lipid bilayer (Engelman, 1982). With the advent of recombinant DNA technology, determination of primary sequences of proteins is proceeding at a much faster pace than determination of atomic resolution protein structures. While thousands of membrane protein sequences are available, there are few detailed three dimensional membrane protein structures known. Available structures, at present, include the photosynthetic reaction centers from *Rps. viridis* (Deisenhofer *et al.*, 1985) and *Rb. sphaeroides* (Allen *et al.*, 1987), porins from *Rb. capsulatus* (Weiss *et al.*, 1991) and *E. coli* (Cowan *et al.*, 1992), bacteriorhodopsin from *H. halobium* (Henderson *et al.*, 1990), and a plant light harvesting complex (Kühlbrandt *et al.*, 1994). Since structural information of membrane proteins is vital to understanding of their cellular functions, a great effort has been made to predict membrane protein structures from their primary sequences (White, 1994, Popot *et al.*, 1994, von Heijne, 1994b, von Heijne, 1994a, von Heijne, 1992, Cramer *et al.*, 1992, Popot, 1993, Popot & de Vitry, 1990, Popot & Engelman, 1990, Jähnig, 1989, Tuffery *et al.*, 1994, Arkin *et al.*, 1994, Fasman, 1989a, Argos *et al.*, 1982).

Most structure prediction algorithms can be categorized into three main classes: statistical (Chou & Fasman, 1978, Garnier *et al.*, 1978, Levitt, 1978, Rao & Argos, 1986, Persson & Argos, 1994, Lohmann *et al.*, 1994, Holley & Karplus, 1989, Kuhn & Leigh, 1985), physico-chemical (Kyte & Doolittle, 1982, Argos *et al.*, 1982, Engelman *et al.*, 1986, Cornette *et al.*, 1987, Rees *et al.*, 1989, White, 1994, Eisenberg, 1984, von Heijne, 1988), and comparative (Zvelebil *et al.*, 1987, Blundell *et al.*, 1987, Cohen *et al.*, 1986, Presnell *et al.*, 1992, Busetta, 1986, Johnson *et al.*, 1994, Sali & Blundell, 1993, Rooman & Wodak, 1988). Presently, prediction of tertiary structure is only of practical use when the structure of a homologous protein is already known. Protein homology modeling typically involves the prediction of side-chain conformations in the modeled protein while assuming a main-chain trace taken from a known tertiary structure of a homologous protein. However, the tertiary structures of proteins have been successfully predicted when experimentally derived constraints are used in conjunction with heuristic methods (Ring & Cohen, 1993). In such a knowledge-based approach information, both from the three-dimensional structures of homologous proteins and from the general analysis of protein structure, is used to derive constraints for modeling a protein of known sequence, but unknown structure.

It is certainly not our intention, and nearly impractical, to give a complete review of this explosively developing field. Interested readers are referred to

the excellent review articles assembled in the authoritative books edited by Fasman (Fasman, 1989b) and Creighton (Creighton, 1992) and those listed above. The rest of this chapter will focus on our recent efforts in predicting the structure of the light-harvesting complex II (LH–II) of *Rs. molischianum*, an integral membrane protein of 16 polypeptides aggregating and binding to 24 bacteriochlorophyll-a (BChla) and 12 lycopenes.

Light Harvesting Complex

Photosynthetic organisms radically increase the efficiency and decrease the complexity of their energy gathering apparatus by surrounding the complex photosynthetic reaction centers with simple, pigment-rich protein aggregates known as light harvesting complexes or antenna complexes. The reaction center has a very rapid cycle time, on the order of 10^3 photons per second (van Grondelle & Sundstrom, 1988), and cannot independently collect enough photons to saturate itself. With the inclusion of the antenna complexes, the bacterium can collect and channel to each reaction center much more light energy.

In most purple bacteria there are two basic types of light harvesting complexes: the light harvesting complex I, or LH–I, is found directly surrounding the photosynthetic reaction centers, while the light harvesting complex II, or LH–II, surrounds the LH–I—reaction center aggregates. LH–I absorbs at longer wavelengths than LH–II, typically with a strong absorption band between 870 and 1015 nm, and is found in all types of purple photosynthetic bacteria (Zuber, 1993). LH–II is found in some species (notable exceptions are *Rs. rubrum* and *Rps. viridis*) of bacteria, and typically has one absorption band between 820 and 860 nm, and another around 800 nm (Zuber, 1985). For *Rs. molischianum*, the LH–II complex displays two peaks at 800 and 850 nm and is often referred to as the B800-850 complex. Since the photosynthetic reaction center absorbs in the deep infrared (960 nm for the *Rps. viridis* reaction center) there is a clear energetic hierarchy in the light-harvesting system, with the LH–II complex absorbing light at the highest energy, surrounding LH–I which absorbs at a lower energy, which in turn surrounds the reaction center which absorbs at the lowest energy. This arrangement naturally channels energy from the outer regions of the antenna complex to the reaction center.

Understanding of the mechanism of the photosynthetic reaction center has been greatly enhanced by the determination 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 (Rossmann, 1972, Lattman, 1985). 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. At the time when this project was initiated, there existed no homologous structure to LH–II of *Rs. molischianum*. We attempted to predict the structure of *Rs. molischianum* and intended to use the predicted structure as a probe structure in the molecular replacement method to resolve the 2.4 Å X-ray diffraction data into an atomic structure. We report here the predictions that we have made with emphasis on structure prediction methods. At the end, the current prediction will be compared with the recently published structure for LH–II from *Rps. acidophila* by McDermott et al., who have successfully solved the phase problem for their structure by conventional means (Mcdermott *et al.*, 1995).

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 two small polypeptides, commonly referred to as α and β apoproteins, both shorter than 60 amino acids, which non-covalently bind BChla and carotenoid molecules. These heterodimers aggregate into 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 are 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 is to build a model structure for LH–II of *Rs. molischianum* and to use it as a probe structure in the framework of the molecular replacement method. The ultimate correctness of the predicted structure can be tested by its ability to serve as a successful search model to resolve the X-ray diffraction data in terms of a consistent electron density profile to which an atomic structure can be configured.

Method

In practice, the task of prediction is divided into three stages: (1) predict the secondary structure of the α- and β-apoproteins from their amino acid sequences; (2) build the tertiary structures for the α- and β-apoproteins by comparative modeling; (3) fold the tertiary structures into an aggregated complex (quater-

nary structure) by means of molecular dynamics simulations and energy minimization under the constraints of experimental data and the predicted secondary structure features. Finally, the molecular replacement test was performed using the predicted structure as a probe structure to resolve the unknown crystal structure. A flowchart of the entire procedure is provided in Figure 1.

The molecular dynamics simulations and energy minimizations described in this chapter 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. The 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, A. MacKerell *et al.*, a) 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 $\varepsilon = 1$ were employed.

Prediction of Secondary Structure

The LH–II complex of *Rs. molischianum* consists of two BChla-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 *et al.*, 1992). The space group for the crystal is $P42_12$ with cell dimensions of $92 \times 92 \times 209$ Å.

Since the LH–II complex is an integral membrane protein, we performed 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 takes about 20 amino acids (in an α helix) 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-subunits 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. Figure 2 clearly shows

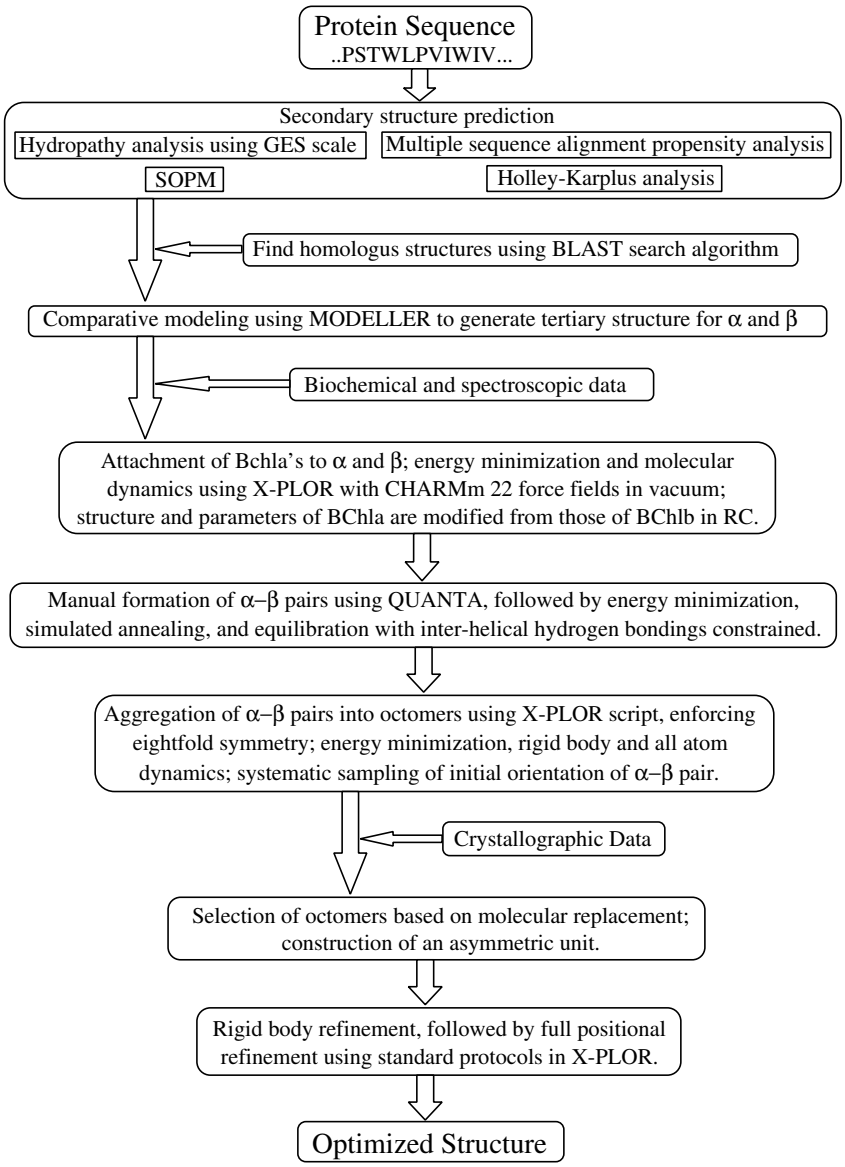

Figure 1: Flowchart outlining the entire structure prediction procedure.

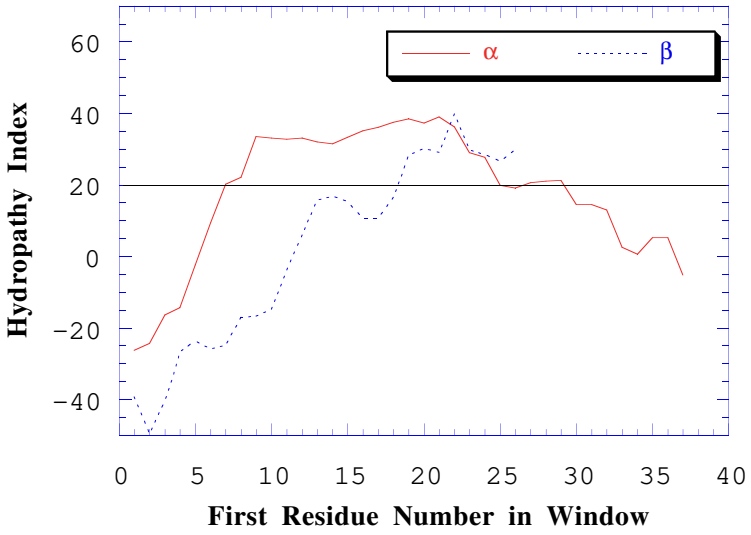

Figure 2: Hydropathy plot of α- and β-apoproteins of the LH–II complex based on the GES hydrophobicity scale (see text) with a window size of 20 amino acids.

that a transmembrane segment exists for both the α- and β-apoproteins. The highest peak occurred at a first residue number of 21 and 22 for the α- and the β-apoprotein respectively. The transmembrane segments were thus identified as α-Val-21:α-Ala-40 and β-Thr-22:β-Trp-41. 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.

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 sequence 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 12 homologous sequences of LH–II and LH–I complexes have been aligned (see Figure 3) and analyzed. As shown in Figure 4 , 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

Figure 3: Sequence alignment 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. Coding: Bold – Highly conserved; Shaded – Nearly conserved. Alignment done using program MACAW (Multiple Alignment Construction & Analysis Workbench) (Schuler *et al.*, 1991).

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.

To confirm the above secondary structure prediction, we performed more sequence analyses with various secondary structure prediction methods including SOPM (self optimized prediction method) and Holley-Karplus analysis (Geourjon & Deleage, 1994, Holley & Karplus, 1989, Holley & Karplus, 1991). 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. SOPM and Holley-Karplus predictions are listed in Tables I and II, respectively. The results are consistent with the assignment of transmembrane segments, for both the α- and β-apoprotein, derived above. Both analyses demonstrate that the 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 existed no structure which was highly homologous to either α- or β-apoprotein as a whole, we have found structures in the PDB (Protein Data Bank) which are homologous to multiple fragments of α- and β-apoproteins. Table III lists some of the homologous fragments to α- and β-apoproteins resulting from a PDB BLAST search (Altschul *et al.*, 1990). Detailed PDB BLAST search results are given in Table IV. A homology

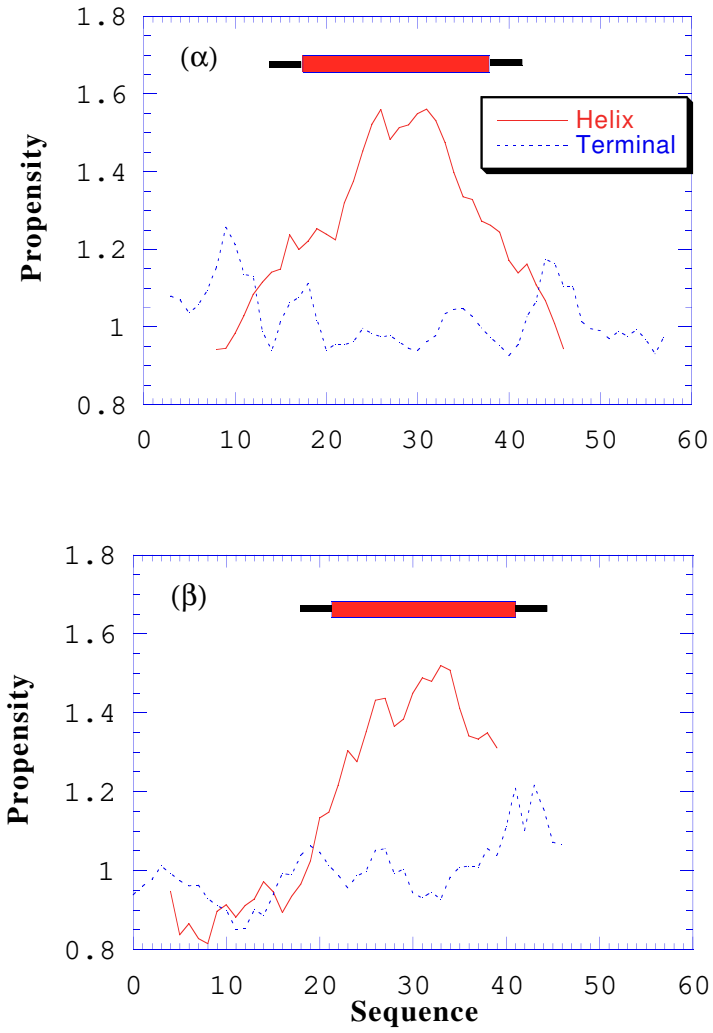

Figure 4: Propensity profile and assignment of 20 residue long transmembrane span (thick bar) and four residue long helical terminal (thin bar) for α- and β-apoproteins of the LH–II complex based on the multiple sequence alignment propensity analysis method of Persson and Argos. Solid line - transmembrane helix; dashed line - terminal region.

Table I. SOPM Prediction

$\alpha - apoprotein$

```
SNPKDDYKIWLVINPSTWLPVIWIVATVVAIAVHAAVLAAPGSNWIALGAAKSAAK
CCCCCCCCCEEEEECCCCCEEEHHHHHHHHHHHHHHHHHHHHCCCCHEEHHHHHHHHH
```

$\beta - apoprotein$

```
AERSLSGLTEEEAIAVHDQFKTTFSAFIILAAVAHVLVWVWKPWF
HHHHHCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
```

Self optimized prediction method (SOPM) for protein secondary structure prediction (Geourjon & Deleage, 1994).

Table II. Holley-Karplus Prediction

$\alpha - apoprotein$

```
SNPKDDYKIWLVINPSTWLPVIWIVATVVAIAVHAAVLAAPGSNWIALGAAKSAAK
-------eEEEEe-------hhHHHHHHHHHHHHHhhhhh-----hhHHHHHHHHh--
```

$\beta - apoprotein$

```
AERSLSGLTEEEAIAVHDQFKTTFSAFIILAAVAHVLVWVWKPWF
--------hHHHHHHHHHhh-hhhh-hHHHHHHhhhhHHH------
```

Neural network based informational approach for protein secondary structure prediction(Holley & Karplus, 1989, Holley & Karplus, 1991).

with 26% identity and 50% positive exists between a segment of the α apoprotein [α-Leu-11 to α-Ala-40] and the transmembrane helix D of the M subunit of the photosynthetic reaction center of *Rb. sphaeroides* [M subunit, 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 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 generates exactly the same alignment as shown in Table III with a quality of 19.2. The average quality for 100 randomized alignments in which the query sequence is randomly permuted (shuffled) is 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 (Lipman & Pearson, 1985, Pearson, 1990). Perhaps a more convincing support for this alignment is the fact that structurally, both proteins exist as α-helical transmembrane segment, and functionally, both proteins contain bacteriochlorophyll-binding residues. Also, two short segments of β-apoprotein are highly homologous to two corresponding segments in the L subunit of the photosynthetic reaction center of *Rb. sphaeroides* [see Table III]. Reaction center L subunit sequence 122:133 `AFAILAYLTLVL` is located in the center of the transmembrane helix C (Michel *et al.*, 1986), which corresponds well with our secondary structure assignment of the transmembrane segment for the β-apoprotein. Sequence

WVKLPWW near the C-terminal of the reaction center L subunit corresponds well with sequence WVWKPWF of the β-apoprotein of LH–II.

Table III. Alignment of Homologous Sequences

alpha		SNPKDDYKIWLVINPSTWLPVIWIVATVVAIAVHAAVLAAPGSNWIALGAAKSAAK
1R1E|E	97:106	---KDDYGEWRVV----------
2RCR|M	196:225	----------LFYNPFHGLSIAFLYGSALLFAMHGATILA----------
1ACB|E	22:26	----------AVPGS----------
1CPC|L	68:72	----------APGGN----------
1TYP|A	91:102	----------NWKALIAAKNKA-
beta		AERSLSGLTEEEAIAVHDQFKTTFSAFIILAAVAHVLVWVWKPWF
1AAM	349:365	---SFSGLTKEQVLRLREEF----------
1GRA	381:388	------GLTEDEAI----------
256B|A	85:97	----------KEAQAAAEQLKTT----------
2RCR|L	122:133	----------AFAILAYLTLVL--------
1EPS	22:36	----------KTVSNRALLLAALAH----------
2BBQ|A	52:61	----------LRSIIHELLW-----
2RCR|L	266:272	----------WVKLPWW

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.

Homology modeling can also be used to improve secondary structure assignment. In case that no clear-cut secondary structure assignment can be made, the secondary structural features of the homologous structure can be employed to establish the secondary structure identity of α- and β-apoproteins. Specifically, the N and C termini of the transmembrane helix for the α-apoprotein were set to Ser-16 and Ala-41 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 (residue 14:16) and PGSN (residue 41:44) segments have a high tendency to form reverse turn (Levitt, 1978). 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 Tyr-10 in analogy to the highly homologous glutathione reductase sequence GLTEDEAI. This assignment of the secondary structure at the N terminal is consistent with both SOPM and Holley-Karplus predictions.

The final secondary structure assignment for both the α- and β-apoproteins is listed in Table V. 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 10 residues [α-Ile-46 to α-Ala-55] has also been identified for the α-apoprotein at the C-terminal. This assignment is supported by the following observations: (1) residues in the sequence `IALGAAKSAA` have a high propensity to form an α-helix as evident from SOPM and Holley-Karplus analyses [see Tables I and II] and other propensity analyses which we have performed; (2) the homologous fragment `KALIAAKNKA` from trypanothione reductase, as listed in Table III, is an α-helix; and (3) as shown

Table IV(A). PDB BLAST Search[a] Results

α-apoprotein
(1) pdb|1R1E|E Identities = 6/10 (60%), Positives = 7/10 (70%)

`Query:  4 KDDYKIWLVI 13`
`KDDY W  V+`
`Sbjct:  97 KDDYGEWRVV 106`

(2) pdb|2RCR|M Identities = 8/30 (26%), Positives = 15/30 (50%)

`Query:  11 LVINPSTWLPVIWIVATVVAIAVHAAVLAA 40`
`L  NP  L + ++  + +  A+H A + A`
`Sbjct:  196 LFYNPFHGLSIAFLYGSALLFAMHGATILA 225`
`2nd str[b]:  TTS HHHHHHHHHHHHHHHHHHHHHHHHHT`

(3) pdb|1ACB|E Identities = 4/5 (80%), Positives = 4/5 (80%)

`Query:  39 AAPGS 44`
`A PGS`
`Sbjct:  22 AVPGS 26`

(4) pdb|1CPC|L Identities = 4/5 (80%), Positives = 4/5 (80%)

`Query:  40 APGSN 45`
`APG N`
`Sbjct:  68 APGGN 72`

(5) pdb|1TYP|A Identities = 8/12 (66%), Positives = 9/12 (75%)

`Query:  44 NWIALGAAKSAA 55`
`NW AL AAK+ A`
`Sbjct:  91 NWKALIAAKNKA 102`
`2nd str[b]:  -HHHHHHHHHHH`

a. Basic local alignment search tool (BLAST) search (Altschul *et al.*, 1990).
b. Secondary structure determined by DSSP (Kabsch & Sander, 1983). H: 4-helix (α-helix); E: extended strand, participates in β-ladder; T: H-bonded turn; S: bend.

Table IV(B). PDB BLAST Search Results

β-apoprotein

(1) pdb|1AAM| Identities = 6/16 (37%), Positives = 12/16 (75%)

```
Query:    4 SLSGLTEEEAIAVHDQ 19
            S SGLT+E+ + + ++
Sbjct:  349 SFSGLTKEQVLRLREE 364
2nd str:    EEE---TTTTTTSSSS
```

(2) pdb|1GRA| Identities = 7/8 (87%), Positives = 8/8 (100%)

```
Query:    7 GLTEEEAI 14
            GLTE+EAI
Sbjct:  381 GLTEDEAI 388
2nd str:    E--HHHHH
```

(3) pdb|256B|A Identities = 7/13 (53%), Positives = 9/13 (69%)

```
Query:   11 EEAIAVHDQFKTT 23
            +EA A  +Q KTT
Sbjct:   85 KEAQAAAEQLKTT 97
2nd str:    HHHHHHHTHHHHH
```

(4) pdb|2RCR|L Identities = 7/12 (58%), Positives = 8/12 (66%)

```
Query:   26 AFIILAAVAHVL 37
            AF ILA + VL
Sbjct:  122 AFAILAYLTLVL 133
2nd str:    HHHHHHHHHHHT
```

(5) pdb|1EPS| Identities = 7/15 (46%), Positives = 11/15 (73%)

```
Query:   21 KTTFSAFIILAAVAH 35
            KT  +  ++LAA+AH
Sbjct:   22 KTVSNRALLLAALAH 36
2nd str*:   HHHHHHHHHHHHHHH
```

(6) pdb|2BBQ|A Identities = 4/10 (40%), Positives = 7/10 (70%)

```
Query:   30 LAAVAHVLVW 39
            L ++ H L+W
Sbjct:   52 LRSIIHELLW 61
2nd str:    HHHHHHHHHH
```

(7) pdb|2RCR|L Identities = 4/7 (57%), Positives = 5/7 (71%)

```
Query:   39 WVWKPWF 45
            WV  PW+
Sbjct:  266 WVKLPWW 272
2nd str:    HHT-TTS
```

in Figure 5, the segment is highly amphiphilic and suitable for sitting 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 amphiphilic helix can be quantified in terms of hydrophobic moment (Eisenberg *et al.*, 1982, Eisenberg, 1984, Eisenberg *et al.*, 1984).

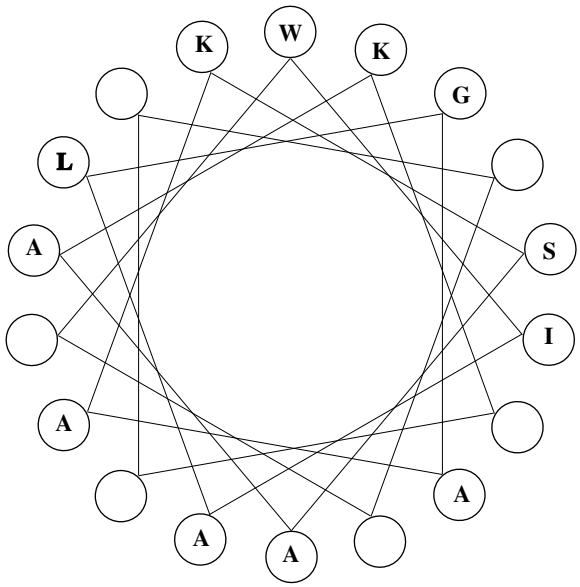

Figure 5: Helical wheel representation of the interfacial α-helix from α-Trp-45 to α-Lys-56.

A reverse-turn segment for the α-subunit, `PGSN`, was assigned based on a propensity analysis with Levitt's scale (Levitt, 1978). Shown in Figure 6 is the propensity profile of the transmembrane helix and 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 also been observed in other membrane proteins (White, 1994, Kühlbrandt *et al.*, 1994).

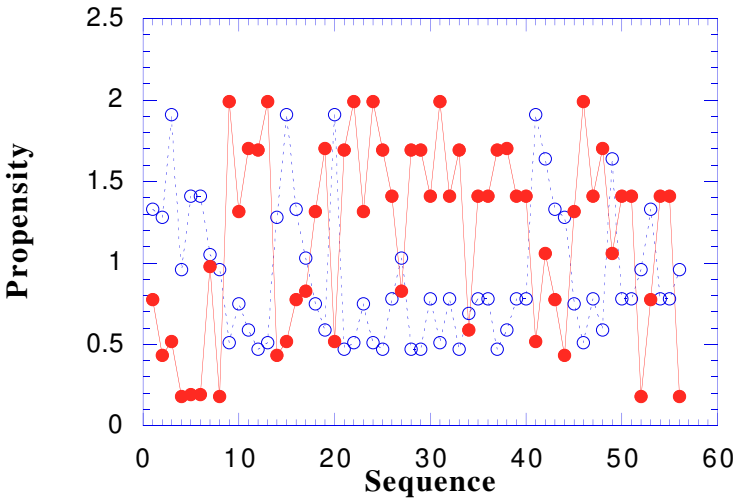

Figure 6: Propensity profile of transmembrane helix and reverse-turn for α-apoprotein. Solid circles – transmembrane helix propensity base on Persson and Argos scale (Persson & Argos, 1994); open circles – reverse-turn propensity by Levitt (Levitt, 1978).

Table V. Assignment of Secondary Structure and Topology

	$\alpha - apoprotein$
Topology	----------TTTTMMMMMMMMMMMMMMMMMMMMTTTT----IIIIIIIIII-
Sequence	SNPKDDYKIWLVINPSTWLPVIWIVATVVAIAVHAAVLAAPGSNWIALGAAKSAAK
Secondary structure	CCCCCCCCEEEEECCHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHC
	$\beta - apoprotein$
Topology	----------TTTTMMMMMMMMMMMMMMMMMMMMTTTT---
Sequence	AERSLSGLTEEEAIAVHDQFKTTFSAFIILAAVAHVLVWVWKPWF
Secondary structure	CCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCC

Note: M - Transmembrane region; T - Terminal (interfacial) region; I - Interfacial helix.

Construction of the α and β Apoproteins

Based on the predicted secondary structure feature, tertiary structures for both the α- and β-apoproteins 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 Table III, are employed in such a procedure. In MODELLER (Sali & Blundell, 1993), the three dimensional model of the unknown protein is obtained by satisfying spatial restraints in the form of probability density functions (pdfs) derived from the alignment of the unknown with one or more homologous structures. The pdfs restrain C_α-C_α distances, main-chain N-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 were 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 Table III. The rest of the terminal residues were also added using the corresponding homologous structures as templates. The α- and the β-apoproteins so constructed were each optimized by energy minimization with fixed protein backbone.

Placement of BChla's and Construction of $\alpha\beta$ Heterodimer

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 Figure 3) reveals three highly conserved His 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 (Figure 3), the motif α-Ala-30 ... α-His-34 ... α-Leu-38 is highly conserved (see Figure 7). 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-Ala-31...β-His-35 is also highly conserved (see Figure 7) in the β-apoprotein. The conserved Phe residue in the -8 position to His is also significant (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-Ala-30 ... α-H is-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-Ala-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 7, 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. It has been observed that many transmembrane α helices are

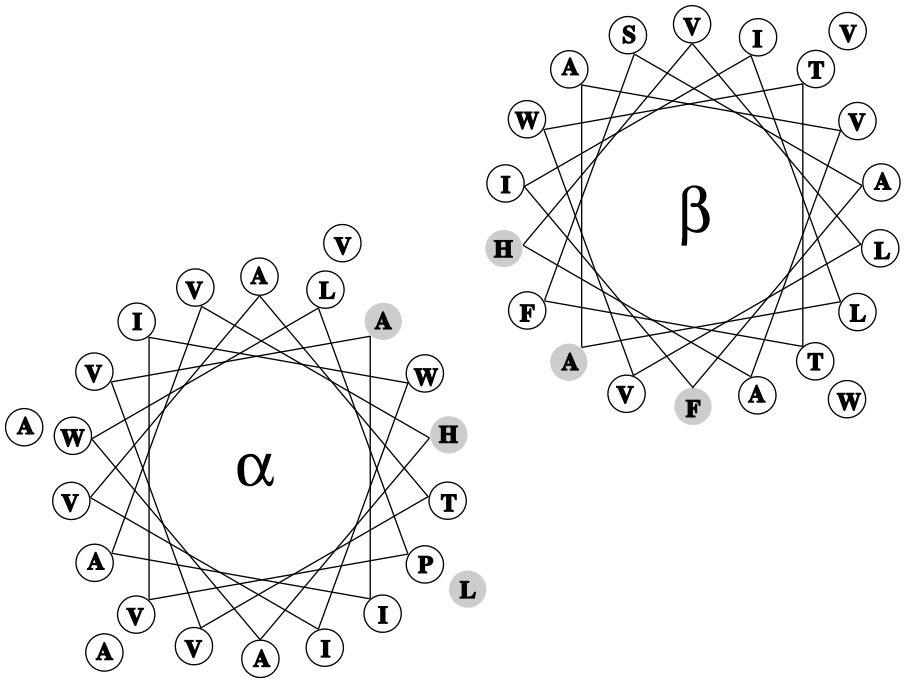

Figure 7: Proposed interhelical arrangement of α- and β-apoproteins. The residues in shaded circles are highly conserved.

amphipathic, with opposing polar and nonpolar faces oriented along the long axis of the helix (Rees *et al.*, 1989, von Heijne & Manoil, 1990, Segrest *et al.*, 1990). Energetically, the apolar surfaces of transmembrane helices provide a good interface to the lipid bilayers, the polar side of inward-facing residues tend to be in the interhelical region to eliminate possible contacts between polar residue and lipid (Rees *et al.*, 1989). In our case, transmembrane cores for both

the α- and the β-subunit are highly hydrophobic as shown in the helical wheel representation of Figure 7. Except for the BChla binding histidine residues, the core of both transmembrane segments consists of nearly all nonpolar residues. On one hand, this potentially useful rule for arranging interhelical packing is not applicable to the current case. On the other hand, it may be a good indication of a lack of interhelical contacts in the transmembrane core for the LH–II complex.

As we build 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-terminals 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 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: (1) the $\alpha\beta$ heterodimer should be arranged interhelically with two B850 BChla binding histidine residues facing each other as shown in Figure 7; (2) the two B850 BChla's are paired and oriented perpendicular to the membrane plane. Therefore, the basic unit of the LH–II complex was configured with the B850 BChla pair sandwiched between two helices of the $\alpha\beta$ heterodimer. We will come back to the consequence of hydrogen bond constraint in the following section.

Construction of the Complete Octamer

The construction of the complete aggregated complex is 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 $\alpha\beta$ heterodimers by means of molecular dynamics simulations and energy minimization under the constraints of experimental data (see Figure 1).

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 by applying (1) for 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_α;

by applying (2) in the turn region the harmonic restraints to the dihedral angles of the main chain. No restraint was applied to all coiled terminal residues. 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 coordinate 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 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 restraints, 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.

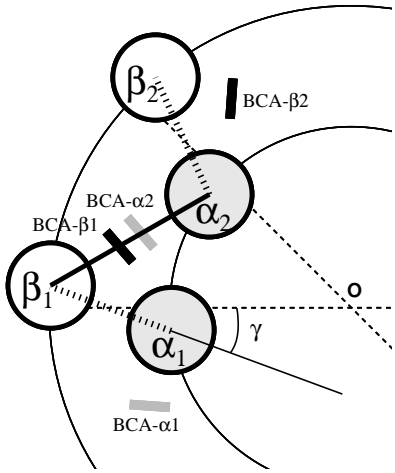

Figure 8: Definition of $\alpha\beta$ pair geometries. Heterodimer: $\alpha\beta$ dimer formed between β_1 and α_2, linked by a solid line; Neighboring pair: $\alpha\beta$ dimer formed between β_1 and α_1, linked by a dashed line; γ is defined as the angle between the vector linking two helices and the radial vector from the center of the octamer to the center of the neighboring pair

In the second step, the complete octamer was constructed enforcing an eight-fold symmetry. A self rotational search (Brünger, 1992, McRee, 1993, Lattman, 1985) of the X-ray diffraction data indicated that, most likely, the LH–II complex possesses an eightfold symmetry. Thus, the task of constructing the octamer was reduced to building a protomer for the eightfold octamer. There were two possible choices for the protomer: the $\alpha\beta$ heterodimer or a neighboring

$\alpha\beta$ pair as shown in Figure 8. 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 was that in the $\alpha\beta$ heterodimer ($\alpha_2\beta_1$ in Figure 8), the two helices separated by the B850 BChla pair are too far apart to generate any significant interaction to be optimized. In contrast, in the neighboring $\alpha\beta$ pair ($\alpha_1\beta_1$ in Figure 8), 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 α and β subunits,we found that it is more feasible spatially 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 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 (β_1 and α_1 as shown in Figure 8) so that: (1) the $\alpha\beta$ heterodimers are arranged in a configuration as shown in Figure 7; (2) hydrogen bonds were established between α-Trp-45 and the 2-acetyl group of BChla attached to β-35 His, and 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 neighboring $\alpha\beta$ pair, followed by rigid body minimization with two helices and three BChla's as five rigid bodies. Then, a 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 runs at 300 K, obtaining an equilibrated structure of the neighboring $\alpha\beta$ pair.

In a third step, the equilibrated neighboring $\alpha\beta$ pair with all three bound BChla's were combined into an octamer by means of long time molecular dynamics simulations and energy minimization with eightfold 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 large cutoff distance for non-bonded interactions was required (we chose 15 Å), since the initial inter-unit distances were 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 minimum energy configuration or until a limit of 700 cycles was reached. It normally 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. The underlying physical principle behind this simulation protocol is 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 to be sampled, enhancing the chance of placing the initial structure in the basin of attraction of the global minimum. A systematic sampling of initial configurations has also been attempted by varying the relative orientations (angle γ in Figure 8) of the vector linking two helices and the radial vector from the center of the octamer to the center of the neighboring pair. The best predicted structure is determined by the molecular replacement procedure as described in the following session. The structure that gives rise to the lowest R value is deemed the most optimal one.

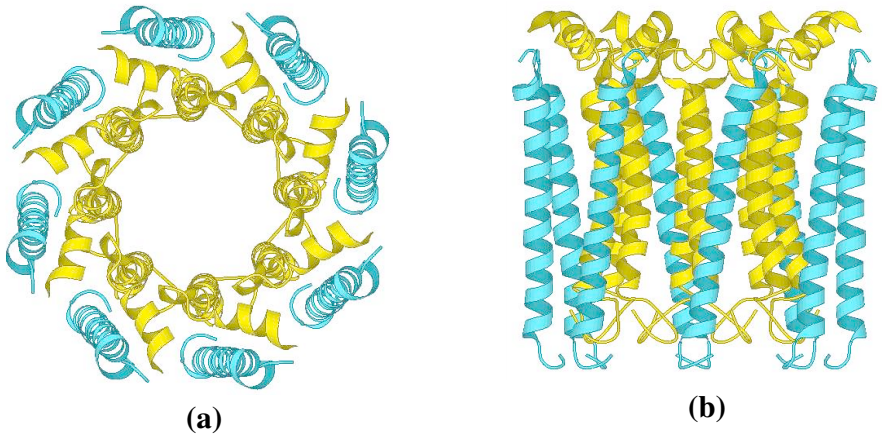

Figure 9: Schematic structure of the LH–II complex: (a) top view with C terminus pointing upward, showing a ring structure with 70 Å diameter, with α inside and β outside; (b) side view with C terminus on top. [produced with MOLSCRIPT (Kraulis, 1991)].

Shown in Figure 9 is one of the octamers we have optimized. The octamer forms a ring with an outer diameter of about 70 Å with α inside and β outside. This 70 Å outer diameter covers the side chains of the β-apoprotein. Typically, the radius of the eight β poproteins (measured from the center of the helices) varied between 27.8 and 32.8 Å; the radius of the eight α apoproteins varied between 15.1 and 21.0 Å. The Mg to Mg separation in the $\alpha\beta$ heterodimer was about 8.8 Å. The Mg to Mg separation between adjacent heterodimers was 16.0 Å. The inner helices were tilted against the membrane plane normal by about 6^o and the outer helices were tilted by about 9^o. The resulting BChla's are opti-

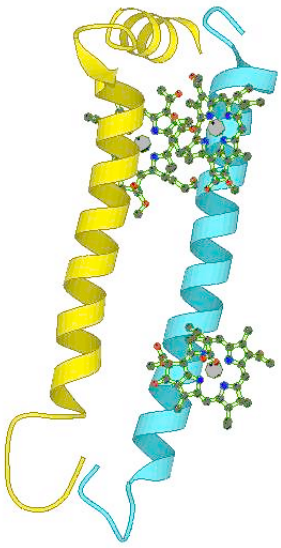

Figure 10: Schematic drawing of the $\alpha\beta$ heterodimer, i.e., the monomer "unit" of LH–II. Shown are the predicted secondary structure and locations of three Bchla's binding to α-34 His, β-17 His and β-35 His. The C terminus is on top.

mally oriented to capture light coming from all directions and are close enough for efficient energy transfer (see the following section on exciton properties). As depicted in Figure 10, the B850 BChla pair are sandwiched between two helices of the $\alpha\beta$ heterodimer. Such an arrangement has recently 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 a lack of helix turn residues near β-His-17. As stated above, the placement of BChla at β-His-17 is itself problematic. The reason we placed it there is that we did not have any better clues at the time when the work reported here was carried out.

Future Work

It is worth mentioning that the optimized octamers are not selected based on the criterion of energy minimization. Although the vacuum environment is similar to a lipid bilayer in terms of hydrophobicity, the peptide-lipid interactions are not explicitly included in the conformational energy. Instead, the molec-

ular replacement test is employed as the ultimate test of the correctness of the model. The predicted structures are presently 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 perform a Patterson correlation refinement preceding the translational search to filter the peaks of the rotational search (Brünger, 1990). The preliminary results have been encouraging as demonstrated by the model's ability to probe the position of molecules in the crystal. The rotational search orients the α helices of the octamers in a direction parallel to the c-axis of the crystal unit cell. The translational search further places the octamers in the site of the four-fold axis of the crystal. Currently, for our best model with 6-12 Å resolution data, the R value after rigid body refinement is 49.8% and decreases to 29.8% after positional refinement. Work is in progress at two fronts to resolve the X-ray diffraction data for the LH–II of *Rs. molischianum*: (1) a more systematic sampling of initial octamer configurations are undertaken; (2) a rigid body simulated annealing protocol is being developed and will be implemented in both the Patterson correlation refinement and the rigid body refinement procedures (Brünger, 1990, Brünger, 1992) in place of the rigid body minimizer to improve the convergence radius of the minimizer. It is our hope that the structure will be further refined to resolve the 2.4 Å diffraction data. If successful, this work will be the first using an *ab initio* predicted structure as a probe structure in the framework of molecular replacement to solve the phase problem in X-ray crystallography structure determination. The methodology developed may be useful for structure prediction of other integral membrane proteins.

Comparison with the Structure of LH–II of *Rps. acidophila*

This chapter is based on an article we recently submitted for publication in Protein Science (Hu et al., in press). [1] As this chapter was written, the crystal structure of LH–II from *Rps. acidophila* determined by conventional multiple isomorphous replacement method was published by McDermott et al. (Mcdermott *et al.*, 1995). This permits us to compare our prediction to an observed structure, albeit not for the same protein. There exists significant sequence homology between the α-subunit from LH–II of *Rps. acidophila* strain 10050 and *Rs. molischianum*. For the β-subunit, the two sequences are homologous, but to a lesser extent. In Table VI, a comparison between the predicted secondary

[1] Modified versions of this article have been submitted to *Protein Science*.

structure assignment for *Rs. molischianum* and the X-ray resolved secondary structure for *Rps. acidophila* (Mcdermott *et al.*, 1995) is given.

Table VI. Comparison of Secondary Structures

	$\alpha - apoprotein$
Rs. molischianum	SNPKDDYKIWLVINPSTWLPVIWIVATVVAIAVHAAVLAAPGSNWIALGAAKSAAK
Prediction[a]	CCCCCCCCEEEEECCHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHC
Rps. acidophila	---MNQGKIWTVVNPAIGIPALLGSVTVIAILVHLAILSHTTWFPAYWQGGVKKAA
X-ray[b]	---C333333333HHHHHHHHHHHHHHHHHHHHHHHHHHHCCCHHHHHHH???????

	$\beta - apoprotein$
Rs. molischianum	AERSLSGLTEEEAIAVHDQFKTTFSAFIILAAVAHVLVWVWKPWF-
Prediction	CCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCC-
Rps. acidophila	-----ATLTAEQSEELHKYVIDGTRVFLGLALVAHFLAFSATPWLH
X-ray	-----CCCCCChhhhHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCC

a: H - Helix; E - β-Sheet; C - Coil; 3 - 3 Turn Helix; ? - Unresolved.
b: The secondary structures for LH–II of *Rps. acidophila* are extracted from (Mcdermott *et al.*, 1995).

The predicted secondary structure of LH–II of *Rs. molischianum* compares well with that of the *Rps. acidophila* crystal structure. In particular, the predicted [transmembrane helix – reverse-turn – interfacial helix] motif for the α-apoprotein was observed in the crystal structure.

For comparison, the overall structure of the aggregated LH–II complex of *Rps. acidophila* is briefly outlined below. For a complete description, interested readers are referred to the original publication (Mcdermott *et al.*, 1995). 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 two B850 BChla's are bonded to the two conserved histidine residues at the periplasmic side, sandwiched between the $\alpha\beta$ heterodimer and oriented perpendicular to the membrane plane. The B800 BChla is oriented parallel to the membrane plane, and bonded to the formyl group of the f-Met residue at the N terminus of the α-apoprotein. Overall, the B850 BChla's form almost a symmetrical ring, with an Mg to Mg distance of 8.7 Å within the $\alpha\beta$ heterodimer and 9.7 Å between the adjacent heterodimers. 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 section corresponds well with that of *Rps. acidophila*. The biggest discrepancy lies at the placement of

BChla's. While our predicted structure place the B850 BChla pair unevenly, the X-ray structure shows that all B850 BChla's are arranged nearly symmetrically in a ring conformation. 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 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 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.

Exciton Properties

The ring-like structure of the light harvesting system with its approximate eight-fold symmetry poses the question in how far this architecture of the protein is an optimal choice for its function, namely, to absorb sunlight and transfer the energy absorbed to the photosynthetic reaction center. We want to demonstrate here that the architecture, which realizes a cyclic arrangement of chlorophyls, is indeed optimal, yielding an excitonic structure which can serve very well for the energy transfer. For this purpose we idealize the arrangement of the chlorophyls to one in which a 2N-fold symmetry axes exists for the 2N chromophores. Adjacent chromophores are oriented to each other in close analogy to the arrangement of the special pair in the photosynthetic reaction center of *Rps. viridis* (Deisenhofer *et al.*, 1985) for which the coupling energies are well known (see e.g., [Eccles et al.,1988]).

The proximity of the chlorophyls leads to a coupling of their excited states such that the stationary excited states are actually linear combinations of excited states of the individual chromophores, denoted by j where $j = 1, 2, \ldots 2N$. The relevant excited state of the chlorophyls is the so-called Q_y state. The chlorophyls in LH–II of *Rps. acidophila* and *Rs. molischianum* are positioned such that the transition dipole moments $\vec{D}_j = \langle \text{ground}|\vec{\mathcal{D}}|\text{Chl}_j^*\rangle$ are oriented in the plane of the membrane (Kramer *et al.*, 1984, Zuber, 1993). Here Chl_j denotes the Q_y excited state of the j-th chlorophyl and $\vec{\mathcal{D}}$ denotes the dipole moment operator. The transition dipole moments of neighboring cholophyls are approximately anti-parallel. If one assumes otherwise a perfect 2N symmetry, the transition dipole moments are given by the Cartesian vectors

$$\vec{D}_j \;=\; D_0 \begin{pmatrix} \cos\phi_j \\ \sin\phi_j \\ 0 \end{pmatrix}, \qquad \phi_j \;=\; \frac{\pi j(N+1)}{N}\,, \quad j \;=\; 1,2,\ldots 2N\,. \tag{1}$$

The system of 2N chlorophyl has 2N states in which a single chlorophyl is in the excited Q_y state. These states are formally written

$$|j\rangle \;=\; |\mathrm{Chl}_1\ \mathrm{Chl}_2 \cdots \mathrm{Chl}_j^* \cdots \mathrm{Chl}_{2N}\rangle\,, \quad j \;=\; 1,2,\ldots 2N\,. \tag{2}$$

The states $|j\rangle$ are coupled which each other through Coulomb interactions between the chlorophyls. At present, we neglect all couplings, except those between adjacent chlorophyls. From studies of the excitonic structure of the photosynthetic reaction center this coupling is well known. It holds for the Hamiltonian matrix in the adopted notation

$$\langle j|\hat{H}|j\pm 1\rangle \;=\; V_o \tag{3}$$

where V_o is positive and approximately equal to 500 cm^{-1} in case of the special pair of the photosynthetic reaction center (Eccles *et al.*, 1988). The diagonal elements of the Hamiltonian are

$$\langle j|\hat{H}|j\rangle \;=\; E_0 \tag{4}$$

where E_o is the excitation energy of the Q_y states of the individual chlorophyls. All other matrix elements vanish in the stated approximation, i.e.,

$$\langle j|\hat{H}|k\rangle \;=\; 0 \quad \text{for } k \;\neq\; j,\, j\pm 1. \tag{5}$$

For the eigenstates of the Hamiltonian holds

$$\hat{H}\,\widetilde{|n\rangle} \;=\; \epsilon_n\,\widetilde{|n\rangle} \tag{6}$$

where

$$\widetilde{|n\rangle} \;=\; \frac{1}{\sqrt{2N}} \sum_{j=1}^{2N} e^{ijn\pi/N}\,|j\rangle \tag{7}$$

and

$$\epsilon_n \;=\; E_o \;+\; 2\,V_o\,\cos\frac{\pi n}{N}\,. \tag{8}$$

In these expressions n are integers in the range $n \;=\; -N+1, -N+2, \ldots, N$. The state of lowest energy, for the present case with $V_o \;>\; 0$, is

$$\epsilon \;=\; \epsilon_N \;=\; E_o \;-\; 2\,V_o\,. \tag{9}$$

The states $\widetilde{|n\rangle}$ represent superpositions of the excited states of the individual chlorophyls and are referred to as excitons. Some of these exciton states carry

oscillator strength, i.e., they are capable of absorbing and emitting photons. This property is governed by the transition dipole moments

$$\langle \text{ground}|\vec{\mathcal{D}}|\widetilde{n\rangle} \;=\; \frac{1}{\sqrt{2N}} \sum_{j=1}^{2N} e^{ijn\pi/N}\, \vec{D}_j \tag{10}$$

Using (1) one can derive

$$|\langle \text{ground}|\vec{\mathcal{D}}|\widetilde{n\rangle}|^2 \;=\; N\, D_o^2\, (\, \delta_{n,-N+1} \;+\; \delta_{n,N-1}\,)\,. \tag{11}$$

Hence, only two excitonic states carry oscillator strength. These states have identical energies

$$\epsilon_\pm \;=\; E_o \;-\; 2\,V_o\, \cos\frac{\pi}{N} \tag{12}$$

which is for large enough N approximately

$$\epsilon_\pm \;\approx\; \epsilon \;+\; 2\,V_o \left(\frac{\pi}{N}\right)^2 \,. \tag{13}$$

We can summarize our findings stating that the coupling of the excited Q_y states in LH–II, in the case of a $2N$-fold symmetry axes for the arrangement of chlorophyls, leads to an exciton spectrum with lowest energy ϵ as given by (9) carrying zero oscillator strength, two degenerate excited states of approximate energy $\epsilon + 2V_o(\pi/N)^2$, carrying all the oscillator strength and $2N-3$ further excited states of higher energy, with maximum energy $\epsilon + 2V_o$, none of which carries oscillator strength. Assuming a coupling strength $V_o = 400$ cm^{-1}, i.e., 20 percent less than the value for the special pairs in photosynthetic reaction centers, one can estimate

$$\epsilon_\pm \;\approx\; \epsilon \;+\; 0.1\text{eV} \left(\frac{\pi}{N}\right)^2 \,. \tag{14}$$

such that, for $N = 8$, holds $\epsilon_\pm \approx \epsilon + 0.015$ eV. This implies that the energy separation between the lowest energy ϵ and the energy $\epsilon_\pm$ of the optically allowed exciton states is less than the thermal energy at room temperature. As a result, one can expect that in thermal equilibrium electronic excitations of LH–II are strongly allowed and, hence, can transfer their energy efficiently towards the reaction center.

It is of interest to speculate how this optimal situation could be interrupted, resulting in less than optimal energy transfer as may be desirable in case of intense light. For this purpose the protein could move the chlorophyls pairwise together, such that the central Mg^{2+} ions lie as close as possible to each other, given a certain separation of the chlorophyl planes. In this case the exciton

Hamiltonian would become (specifying only non-vanishing matrix elements)

$$\hat{H} = \begin{pmatrix} E_o & V_1 & & & & & V_2 \\ V_1 & E_o & V_2 & & & \cdots & \\ & V_2 & E_o & V_1 & & \cdots & \\ & & V_1 & E_o & V_2 & & \\ & & & V_2 & E_o & & \\ & \vdots & \vdots & & & \ddots & \\ V_2 & & & & & & E_o \end{pmatrix} \tag{15}$$

In this case, the matrix elements V_1 would be large and negative and the matrix elements V_2 would be small and positive. The resulting exciton spectrum would consist of a narrow band of optically forbidden exciton states of energy $E_o + V_1 - V_2 < E < E_o + V_1 + V_2$, separated by a wide energy gap $\Delta \approx 4\,V_1$ which could easily exceed thermal energies from a band of optically allowed exciton states of energy $E_o - V_1 - V_2 < E < E_o - V_1 + V_2$. In this case the optically allowed states would not be thermally accessible. Energy transfer would involve optically forbidden exciton states and, hence, would be less efficient, resulting in internal conversion of the absorbed photon energy in the LH–II proteins.

Acknowledgments

This work was supported by the Carver Charitable Trust and the National Institute of Health (NIH grant P41RR05969). We would like to thank Antony Crofts and Zan Schulten for many helpful discussions. James Schnitzer participated in this work during the early phase of the project.

References

A. MacKerell *et al.* (a). (unpublished).

Allen, J., Yeates, T., Komiya, H., & Rees, D. (1987). Structure of the reaction center from *Rhodobacter sphaeroides* R-26: the protein subunits. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6162.

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.

Argos, P., Rao, J., & Hargrave, P. (1982). Structural prediction of membrane-bound proteins. *European Journal of Biochemistry,* **128**, 565.

Arkin, I., Adams, P., MacKenzie, K., Lemmon, M., Brünger, A., & Engelman, D. (1994). Structural organization of the pentameric transmembrane α-helices of phospholamban, a cardiac ion channel. *Embo Journal,* **13**, 4757.

Blundell, T. L., Sibanda, B. L., Sternberg, M. J., & Thornton, J. M. (1987). Knowledge-based prediction of protein structures and the design of novel molecules. *Nature,* **326**, 347.

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. A., 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 B870-α and B870-β 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**, 689.

Busetta, B. (1986). Examination of folding patterns for predicting protein topologies. *Biochimica et Biophysica Acta,* **870**, 327.

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.

Chou, P. & Fasman, G. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Advances in Enzymology and Related Areas of Molecular Biology,* **47**, 45.

Cohen, F. E., Abarbanel, R. M., Kuntz, I. D., & Fletterick, R. J. (1986). Turn prediction in proteins using a pattern-matching approach. *Biochemistry,* **25**, 266.

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.

Cowan, S. W., Schirmer, T., Rummel, G., Steiert, M., Ghosh, R., Pauptit, R. A., Jansonius, J. N., & Rosenbusch, J. P. (1992). Crystal structures explain functional properties of 2 *E. coli* porins. *Nature,* **358** (6389), 727–733.

Cramer, W., Engelman, D., Heijne, G. V., & Rees, D. (1992). Forces involved in the assembly and stabilization of membrane proteins. *Faseb Journal,* **6**, 3397.

Creighton, T., ed (1992). *Protein folding*. New York: W.H. Freeman.

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.

Deisenhofer, J., Epp, O., Mikki, 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.

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.

Eccles, J., Honig, B., & Schulten, K. (1988). Spectroscopic determinants in the reaction center of *Rhodo -pseudomonas viridis*. *Biophys. J.* **53**, 137–144.

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. (1982). An implication of the structure of bacteriorhodopsin: globular membrane proteins are stabilized by polar interactions. *Biophys. J.* **37**, 187.

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.

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

Fasman, G., ed (1989b). *Prediction of protein structure and the principles of protein conformation*. New York: Plenum.

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.

Garnier, J., Osguthorpe, D., & Robson, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *Journal of Molecular Biology,* **120**, 97.

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.

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

Holley, L. H. & Karplus, M. (1989). Protein secondary structure prediction with a neural network. *Proc. Natl. Acad. Sci. USA,* **86**, 152–156.

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

Hu, X., Xu, D., Hamer, K., Schulten, K., Koepke, J., & Michel, H. (a). Predicting the structure of the light-harvesting complex II of *Rhodospirillum molischianum*. Protein Science (in press).

Jähnig, F. (1989). Structure prediction for membrane proteins. In: *Prediction of protein structure and the principles of protein conformation*, (Fasman, G., ed) p. 707, New York: Plenum.

Johnson, M., Srinivasan, N., Sowdhamini, R., & Blundell, T. (1994). Knowledge-based protein modeling. *Critical Reviews in Biochemistry and Molecular Biology,* **29**, 1.

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.

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.

Koepke, J. & Michel, H. (a). (unpublished).

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.

Kuhn, L. & Leigh, J. (1985). A statistical technique for predicting membrane protein structure. *Biochimica et Biophysica Acta,* **828**, 351.

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.

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.

Lohmann, R., Schneider, G., Behrens, D., & Wrede, P. (1994). A neural network model for the prediction of membrane-spanning amino acid sequences. *Protein Science,* **3**, 1597.

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

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) p. 74, 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. (1993). Integral membrane protein structure - transmembrane α-helices as autonomous folding domains. *Current Opinion In Structural Biology,* **3**, 532.

Popot, J. & de Vitry, C. (1990). On the microassembly of integral membrane proteins. *Annual Review of Biophysics and Biophysical Chemistry,* **19**, 369.

Popot, J., de Vitry, C., & Atteia, A. (1994). Folding and assembly of integral membrane proteins: An introduction. In: *Membrane protein structure: experimental approaches*, (White, S., ed) p. 41, New York: Oxford University press.

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

Presnell, S. R., Cohen, B. I., & Cohen, F. E. (1992). A segment-based approach to protein secondary structure prediction. *Biochemistry,* **31**, 983.

Rao, J. M. & Argos, P. (1986). A conformational preference parameter to predict helices in integral membrane proteins. *Biochimica et Biophysica Acta,* **869**, 197.

Rees, D., DeAntonio, L., & Eisenberg, D. (1989). Hydrophobic organization of membrane proteins. *Science,* **245**, 510.

Ring, C. & Cohen, F. (1993). Modeling protein structures: construction and their applications. *Faseb Journal,* **7**, 783.

Rooman, M. & Wodak, S. (1988). Identification of predictive sequence motifs limited by protein structure data base size. *Nature,* **335**, 45.

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.

Schuler, G., Altschul, S., & Lipman, D. (1991). A workbench for multiple alignment construction and analysis. *Proteins: Structure, Function, and Genetics,* **9**, 180.

Segrest, J., Loof, H. D., Dohlman, J., Brouillette, C., & Anantharamaiah, G. (1990). Amphipathic helix motif: Classes and properties. *Proteins, Struct. Funct. Genet.* **8**, 103.

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.

Tuffery, P., Etchebest, C., Popot, J., & Lavery, R. (1994). Prediction of the positioning of the seven transmembrane α-helices of bacteriorhodopsin. a molecular simulation study. *Journal of Molecular Biology,* **236**, 1105.

van Grondelle, R. & Sundstrom, V. (1988). Excitation energy transfer in photosynthesis. In: *Photosynthetic Light–Harvesting Systems*, (Scheer, H., ed) pp. 403–438, Berlin, New York: Walter de Gruyter & Co.

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. (1988). Transcending the impenetrable: how proteins come to terms with membranes. *Biochimica et Biophysica Acta,* **947**, 307.

von Heijne, G. (1992). Membrane protein structure prediction—hydrophobicity analysis and the positive-inside rule. *Journal of Molecular Biology,* **225**, 487.

von Heijne, G. (1994a). Decoding the signals of membrane protein sequence. In: *Membrane protein structure: experimental approaches*, (White, S., ed) p. 27, New York: Oxford University press.

von Heijne, G. (1994b). Membrane proteins: from sequence to structure. *Annual Review of Biophysics and Biomolecular Structure,* **23**, 167.

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

Weiss, M., Kreusch, A., Nestel, U., Welte, W., Weckesser, J., & Schulz, G. (1991). The structure of porin from *Rhodobacter capsulatus* at 1.8 Å resolution. *FEBS Lett.* **280**, 379.

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.

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

Zuber, H. (1993). Structural features of photosynthetic light-harvesting systems. In: *The Photosynthetic Reaction Center*, (Deisenhofer, J. & Norris, J. R., eds) p. 43, San Diego: Academic Press.

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.

Zvelebil, M. J., Barton, G. J., Taylor, W. R., & Sternberg, M. J. (1987). Prediction of protein secondary structure and active sites using the alignment of homologous sequences. *J. Mol. Biol.* **195**, 957.