DIMACS Series in Discrete Mathematics
and Theoretical Computer Science
Volume **00**, 19xx

Knowledge Based Structure Prediction of the Light-Harvesting Complex II of *Rhodospirillum molischianum*

Xiche Hu, Dong Xu, Kenneth Hamer, Klaus Schulten, Juergen Koepke, and Hartmut Michel

Abstract. We illustrate in this article how one proceeds to predict the structure of integral membrane proteins using a combined approach in which molecular dynamics simulations and energy minimization are performed based on structural information from conventional structure prediction methods and experimental constraints derived from biochemical and spectroscopical data. 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. Multiple sequence alignment propensity analyses further pinpointed 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 secondary structure was derived from a combination of all the prediction methods used. The three-dimensional structures for the α- and the β-apoprotein were built by comparative modeling. The resulting tertiary structures were combined into an $\alpha\beta$ dimer pair with bacteriochlorophyll a's attached under constraints provided by site directed mutagenesis and FT Resonance Raman spectra, as well as by conservation of residues. The $\alpha\beta$ dimer pairs were then aggregated into a quaternary structure through molecular dynamics simulations and energy minimization. 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.

Introduction

One of the outstanding problems in structural biology is protein folding, i.e., given the linear sequence of amino acids that constitute a protein, what would be

1991 *Mathematics Subject Classification.* Primary 70F10,92-08; Secondary 92C05,92D20.

Key words and phrases. light-harvesting complex; purple bacteria; protein folding; protein structure; sequence analysis.

This work was supported by the Carver Charitable Trust and the National Institute of Health (NIH grant P41RR05969).

its native, or folded, state in three-dimensional space. The native conformation of a protein corresponds to a global minimum in free energy space. Mathematically, the problem can be formulated as global minimum search in multidimensional space. The solution to this problem has proven difficult because in theory it belongs to a class known as NP complete problems, and in practice the objective function contains too many local minima [**GJ79, CSW94, NVS91**]. As a result, current structure prediction methods are mostly empirical in nature [**Fas89a**]. We illustrate in this article how one proceeds to predict the structure of integral membrane proteins using a combined approach in which molecular dynamics simulations and energy minimization are performed based on structural information from conventional structure prediction methods and experimental constraints derived from biochemical and spectroscopical data.

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* [**DEM⁺85**] and *Rb. sphaeroides* [**AYKR87**], porins from *Rb. capsulatus* [**WKN⁺91**] and *E. coli* [**CSR⁺92**], bacteriorhodopsin from *H. halobium* [**HBC⁺90**], and a plant light harvesting complex [**KWF94**]. 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 [**Whi94, PdVA94, vH94b, vH94a, vH92, CEHR92, Pop93, PdV90, PE90, J8̈9, TEPL94, AAM⁺94, Fas89b, ARH82**]. Most structure prediction algorithms can be categorized into three main classes: statistical [**CF78, GOR78, Lev78, RA86, PA94, LSBW94, HK89, KL85**], physicochemical [**KD82, ARH82, ESG86, CCM⁺87, RDE89, Whi94, Eis84, vH88**], and comparative [**ZBTS87, BSST87, CAKF86, PCC92, Bus86, JSSB94, SB93, RW88**]. 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 [**RC93**]. 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. The rest of this article 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. While the reaction center has a very rapid cycle time, on the order of 10^3 photons per second [**vGS88**], it 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 [**Zub93**]. 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 [**Zub85**]. 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.

Studies of the mechanism of the photosynthetic reaction center have been greatly enhanced by the determination of its three-dimensional structure [**DM89**]. However, structural information about light-harvesting complexes is still limited to spectroscopic and biochemical characterization [**HC91, SvG91, ZB91**]. The LH–II complex of *Rs. molischianum* has been crystallized and X-ray diffraction data have been collected up to 2.4 Å resolution [**KM, Mic91**]. 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 [**Ros72, Lat85**]. 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 (1) the predicted secondary structure for the α- and β-polypeptides; and (2) the predicted quaternary structure for the aggregated LH-II complex of *Rs. molischianum* as well as the 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 [**MPF$^+$95**].

All light-harvesting complexes display a remarkable similarity in the way they are constructed [**Zub85, ZB91**]. The basic structural unit is a heterodimer of two small polypeptides, commonly referred to as α and β, both shorter than 60 amino acids, which non-covalently bind BChla and carotenoid molecules. 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* [**BVC+93**] to a hexadecamer for LH–I of *Rs. rubrum* [**KBG95**].

Various models have been proposed for the light-harvesting complex of purple bacteria [**OH94, Zub86, ZB91**]. 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 α- and β-polypeptides from their amino acid sequences; (2) build the tertiary structures for α- and β-polypeptides by comparative modeling; (3) fold the tertiary structures into an aggregated complex (quaternary 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 article were carried out using the program X-PLOR [**Brün92**]. 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` [**BBO+83, A.**] except for BChla. The partial charges and parameters for BChla were taken from those used in [**TSD+88**] 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 polypeptides 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 polypeptides, three BChla and 1.5 lycopene molecules [**GLRM93**]. It has been determined by sedimentation equilibrium experiments that the native LH–II complex is an octamer of such $\alpha\beta$ units [**KGvdB+92**]. 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 [**KD82, Whi94**]. The transmembrane segments of polypeptides are usually forced to adopt an α-helical conformation due to constraints of the hydrophobic core of the membrane [**ESG86**].

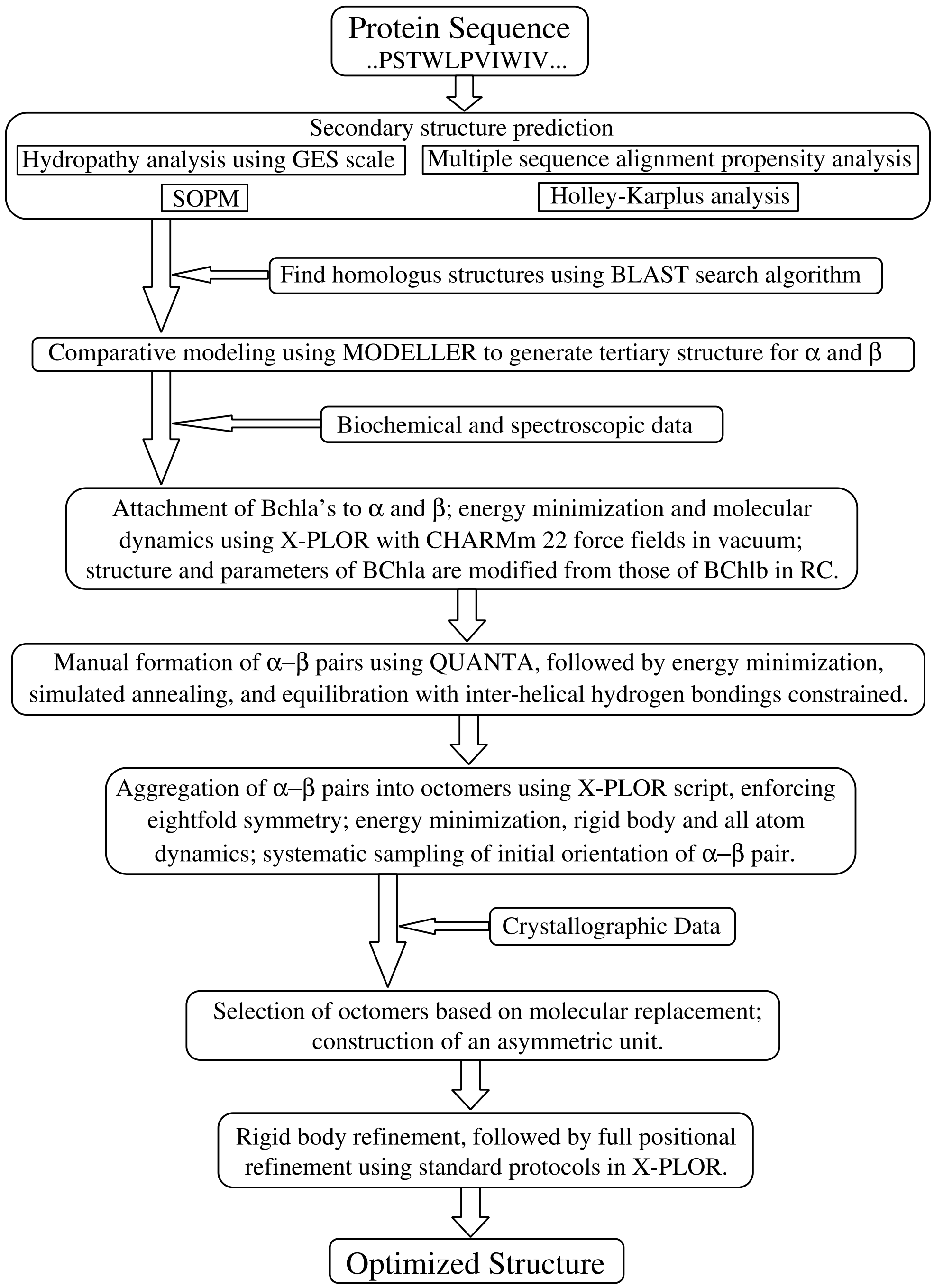

FIGURE 1. Flowchart outlining the entire structure prediction procedure.

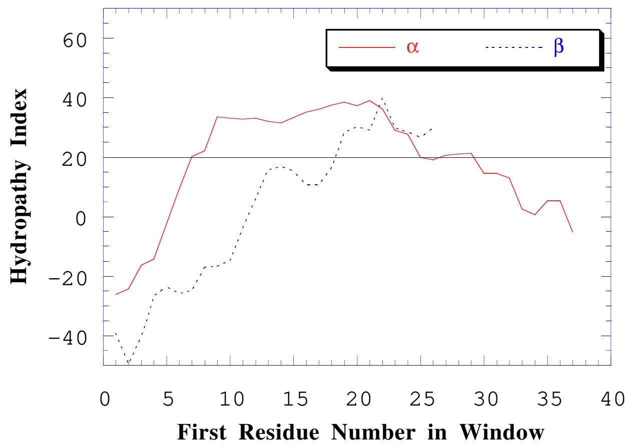

FIGURE 2. Hydropathy plot of α- and β-polypeptides of the LH–II complex based on the GES hydrophobicity scale (see text) with a window size of 20 amino acids.

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 [**ESG86**] 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 that a transmembrane segment exists for both α- and β-polypeptides. The highest peak occurs at a first residue number of 21 and 22 for the α- and the β-polypeptide respectively. The transmembrane segments are thus identified as α-Val-21:α-Ala-40 and β-Thr-22:β-Trp-41. Hydropathy analyses with two other widely used hydrophobicity scales, i.e., Kyte & Doolittle [**KD82**] and Eisenberg consensus [**EWTW82**] scales, generate essentially the same hydrophobic core for the α- and the β-polypeptide 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 [**PA94**] 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 by this method spans from Trp-18 to Val-37 for the α-polypeptide and from Thr-22 to Trp-41 for the β-polypeptide. The hydropathy analysis identifies the region of the transmembrane segment. Multiple sequence

```
                                      α-polypeptide
Rs. Molischianum B800-850             snpkddYKIWLVINPSTWLPVIWIVATVVAIAVHAAVLAAPGSNWIALGAAKsaak----
Rc. gelatinosus DSM 149 B800-850      -----mWRIWRLFDPMRAMVAQAVFLLGLAVLIHLMLLGTNKFNWLDGAKKApvasa---
Rs. rubrum B870                       -----mWRIWQLFDPRQALVGLATFLFVLALLIHFILLSTERFNWLEGASTQpvqts---
Rb. capsulatus B870                   --mskfYKIWLVFDPRRVFVAQGVFLFLLAVLIHLILLSTPAFNWLTVATAKhgyvaaaq
Rb. sphaeroides B870                  --mskfYKIWMIFDPRRVFVAQGVFLFLLAVMIHLILLSTPSYNWLEISAAKynrvavae
Rps. marina B880                      -----mWKVWLLFDPRRTLVALFTFLFVLALLIHFILLSTDRFNWMQGAPTApaqts---
Rp. viridis B1015                     eyrtasWKLWLILDPRRVLTALFVYLTVIALLIHFGLLSTDRLNWWEFQRGLpkaa----

Rs. Molischianum B800-850             snpkddYKIWLVINPSTWLPVIWIVATVVAIAVHAAVLAAPGSNWIALGAAKsaak----
Rp. viridis B1015                     eyrtasWKLWLILDPRRVLTALFVYLTVIALLIHFGLLSTDRLNWWEFQRGLpkaa----
Rc. gelatinosus DSM 149 B800-850      ---mnQGKVWRVVKPTVGVPVYLGAVAVTALILHGGLLAKTDWFGaywnggkkaaaa---
Rb. sphaeroides  2.4.1 B800-850       ---mtNGKIWLVVKPTVGVPLFLSAAVIASVVIHAAVLTTTTWLPayyqgsaavaae---
Rb. capsulatus B800-850               ---mnNAKIWTVVKPSTGIPLILGAVAVAALIVHAGLLTNTTWFAnywngnpmatvvava
Rps. acidophila Ac70.50 B800-850      ---mnQGKIWTVVNPSVGLPLLLGSVTVIAILVHAAVLSHTTWFPaywqgglkkaa----
Rps. palustris 2.6.1 B800-850         ---mnQGRIWTVVNPGVGLPLLLGSVTVIAILVHYAVLSNTTWFPkywngatvaapaaa-

                                      β-polypeptide
Rs. Molischianum B800-850             ----AERSLSGLTEEEAIAVHDQFKTTFSAFIILAAVAHVLVWVWKPWf--------
Rc. gelatinosus DSM 149 B800-850      --aeRKGSISGLTDDEAQEFHKFWVQGFVGFTAVAVVAHFLVWVWRPWl--------
Rs. rubrum B870                       --evKQESLSGITEGEAKEFHKIFTSSILVFFGVAAFAHLLVWIWRPWvpgpngys-
Rb. capsulatus B870                   -adkNDLSFTGLTDEQAQELHAVYMSGLSAFIAVAVLAHLAVMIWRPWf--------
Rb. sphaeroides B870                  -adkSDLGYTGLTDEQAQELHSVYMSGLWPFSAVAIVAHLAVYIWRPWf--------
Rps. marina B880                      aeidRPVSLSGLTEGEAREFHGVFMTSFMVFIAVAIVAHILAWMWRPWipgpegia-
Rp. viridis B1015                     --adLKPSLTGLTEEEAKEFHGIFVTSTVLYLATAVIVHYLVWTAKPWiapipkgwv
Rc. gelatinosus DSM 149 B800-850      addaNKVWPSGLTTAEAEELQKGLVDGTRIFGVIAVLAHILAYAYTPWlh-------
Rb. sphaeroides  2.4.1 B800-850       -ddlNKVWPSGLTVAEAEEVHKQLILGTRVFGGMALIAHFLAAAATPWlg-------
Rb. capsulatus B800-850               -mtdDKAGPSGLSLKEAEEIHSYLIDGTRVFGAMALVAHILSAIATPWlg-------
Rps. acidophila Ac70.50 B800-850      --adDVKGLTGLTAAESEELHKHVIDGTRVFFVIAIFAHVLAFAFSPWlh-------
Rps. palustris 2.6.1 B800-850         addpNKVWPTGLTIAESEELHKHVIDGSRIFVAIAIVAHFLAYVYSPWlh-------
```

FIGURE 3. Sequence alignment of the α and β-polypeptides of the LH–II complex of *Rs. molischianum* with α- and β-polypeptides of LH–II and LH–I complexes from other photosynthetic bacteria. Color coding: Blue – Highly conserved; Magenta – Nearly conserved. Alignment done using program MACAW (Multiple Alignment Construction & Analysis Workbench) [**SAL91**].

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.

Table I. SOPM and Holley-Karplus Predictions

	$\alpha - polypeptide$
sequence	SNPKDDYKIWLVINPSTWLPVIWIVATVVAIAVHAAVLAAPGSNWIALGAAKSAAK
SOPM[a]	CCCCCCCCEEEEEECCCCCEEEHHHHHHHHHHHHHHHHHHHHCCCCHEEHHHHHHHHH
Holley-Karplus[b]	-------eEEEEe-------hhHHHHHHHHHHHHHhhhhh-----hhHHHHHHHh--
	$\beta - polypeptide$
sequence	AERSLSGLTEEEAIAVHDQFKTTFSAFIILAAVAHVLVWVWKPWF
SOPM	HHHHHCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Holley-Karplus	--------hHHHHHHHHhh-hhhh-hHHHHHHhhhhHHH------

a: Self optimized prediction method (SOPM) for protein secondary structure prediction [**GD94**]. b: Neural network based informational approach for protein secondary structure prediction[**HK91**].

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 [**GD94, HK89, HK91**]. SOPM takes structure classes into account and iteratively optimized prediction parameters to increase prediction quality. The Holley-Karplus

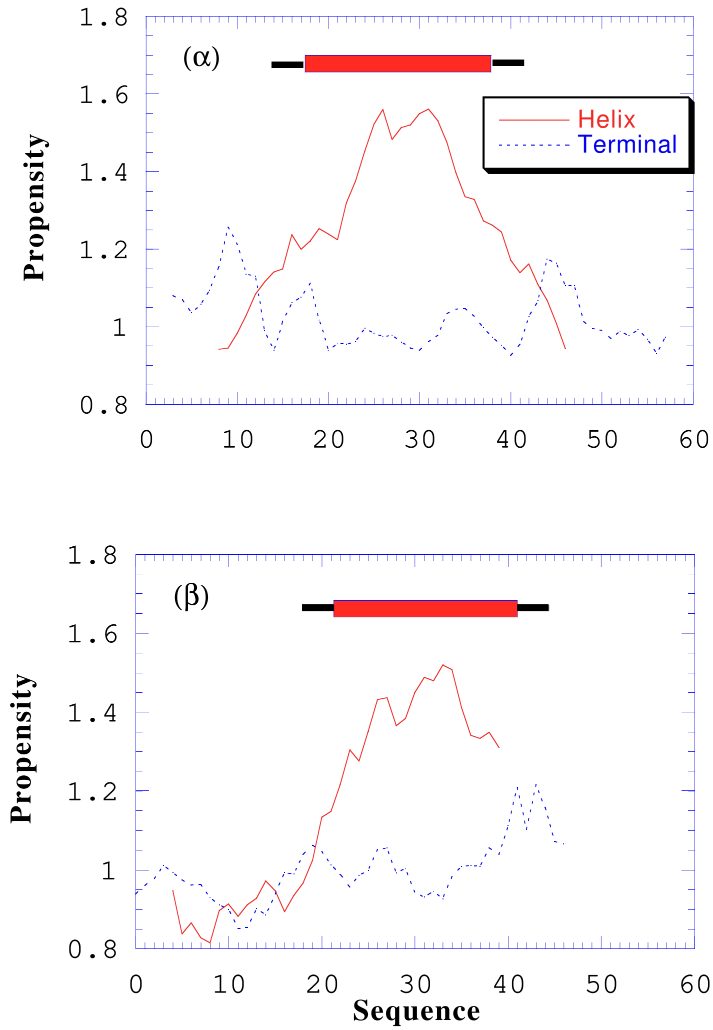

Figure 4. Propensity profile and assignment of 20 residue long transmembrane span (thick bar) and four residue long helical terminal (thin bar) for α- and β-polypeptides of the LH–II complex based on the multiple sequence alignment propensity analysis method of Persson and Argos. Solid line - transmembrane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

prediction is an information-based neural network approach. SOPM and Holley-Karplus predictions are listed in Table I. 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 exists no structure which is highly homologous to either α- or β-apoproteins 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 II lists some of the homologous fragments to α- and β-apoproteins resulting from a PDB BLAST search [**AGM$^+$90**]. Detailed PDB BLAST search results are given in Table III. A homology with 26% identity and 50% positive exists between a segment of the α-apoprotein [α-Leu-11 to α-Ala-40] and the transmembrane helix D of the M subunit of the photosynthetic reaction center of *Rb. sphaeroides* [M subunit, residues 196:225] [**MWG$^+$86**]. To establish the statistical significance of this alignment, we performed a statistical analysis with the BESTFIT program in the GCG package [**DHS84**]. Using a gap generating penalty of 3.0 and a gap extension penalty of 0.1, the BESTFIT program generated exactly the same alignment as shown in Table II with a quality of 19.2. The average quality for 100 randomized alignments in which the query sequence is randomly permuted (shuffled) is 12.7 with a standard deviation of 1.3. That gave rise to a Z-score of 5, which indicated a "possibly significant" alignment according to [**LP85, Pea90**]. Perhaps a more convincing support for this alignment is the fact that structurally, both proteins exist as α-helical transmembrane segment, and functionally, both proteins contain bacteriochlorophyll-binding residues. Also, two short segments of β-apoprotein are highly homologous to two corresponding segments in the L subunit of the photosynthetic reaction center of *Rb. sphaeroides* [see Table II]. Reaction center L subunit sequence 122:133 `AFAILAYLTLVL` is located in the center of the transmembrane helix C [**MWG$^+$86**], which exhibits a close correspondence with our secondary structure assignment of the transmembrane segment for the β-apoprotein. Sequence `WVKLPWW` near the C-terminal end of the reaction center L subunit shows a close correspondence with sequence `WVWKPWF` of the β-apoprotein of LH–II.

Table II. Alignment of Homologous Sequences

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

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.

A secondary structure analysis was performed for all the homologous proteins listed in Table II with the method of Kabsch and Sander which defines the secondary structure of a protein by means of pattern recognition of hydrogen-bonded and geometrical features [**KS83**]. The resulting secondary structure assignment for all the homologous fragments are listed in Table III along with PDB BLAST search results. As shown in Table III, all the homologous fragments located within the putative transmembrane region, except Eco Ri Endonuclease, exist in an α-helical conformation.

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

α-polypeptide

(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 [**AGM$^+$90**].
b. Secondary structure determined by DSSP [**KS83**]. H: 4-helix (α-helix); E: extended strand, participates in β-ladder; T: H-bonded turn; S: bend.

Table III(B). PDB BLAST Search Results

β-polypeptide

(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
```

Homology modeling was also 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 were employed to establish the secondary structure identity of α- and β-polypeptides. Specifically, the N and

C termini of the transmembrane helix for the α-polypeptide 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 a reverse turn [**Lev78**]. 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 terminus is consistent with both SOPM and Holley-Karplus predictions.

The final secondary structure assignment for both α- and β-polypeptides is listed in Table IV. 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 other propensity analyses we have performed; (2) the homologous fragment `KALIAAKNKA` from trypanothione reductase, as listed in Table II, is an α-helix; and (3) as shown 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 an amphiphilic helix can be quantified in terms of the hydrophobic moment [**EWTW82, Eis84, ESKW84**].

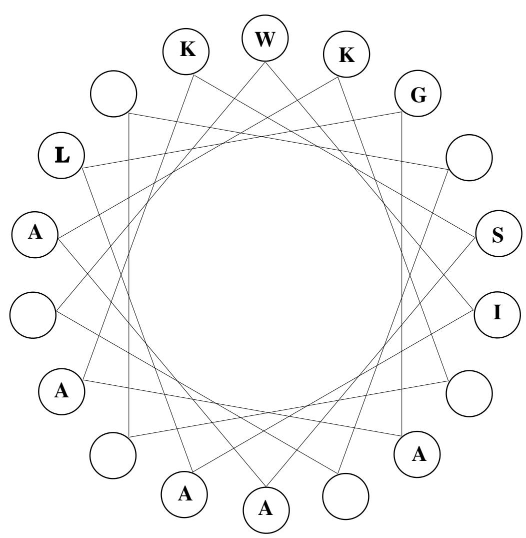

FIGURE 5. Helical wheel representation of the interfacial α-helix from α-Trp-45 to α-Lys-56.

A reverse-turn segment for the α-subunit, `PGSN`, is assigned based on a propensity analysis with Levitt's scale [**Lev78**]. Shown in Figure 6 is the propensity profile

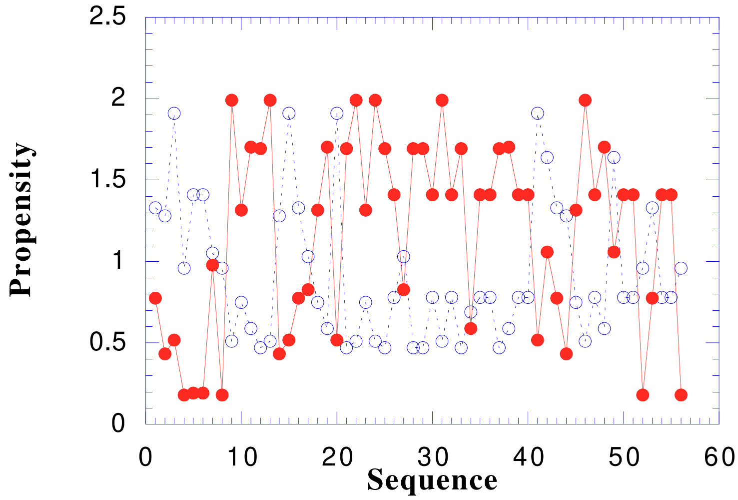

FIGURE 6. Propensity profile of transmembrane helix and reverse-turn for α-polypeptide. Solid circles – transmembrane helix propensity base on Persson and Argos scale [**PA94**]; open circles – reverse-turn propensity by Levitt [**Lev78**].

of the transmembrane helix and reverse-turn for the α-polypeptide. Solid circles indicate the transmembrane helix propensity (for a single sequence) based on the Persson and Argos scale [**PA94**]; open circles represent the reverse-turn propensity according to Levitt [**Lev78**]. 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-polypeptide. Such a motif has also been observed in other membrane proteins [**Whi94, KWF94**].

Table IV. Assignment of Secondary Structure and Topology

	α − polypeptide
Topology	`-------------TTTTMMMMMMMMMMMMMMMMMMMMTTTT----IIIIIIIIII-`
Sequence	`SNPKDDYKIWLVINPSTWLPVIWIVATVVAIAVHAAVLAAPGSNWIALGAAKSAAK`
Secondary structure	`CCCCCCCCEEEEEECCHHHHHHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHC`
	β − polypeptide
Topology	`--------------TTTTMMMMMMMMMMMMMMMMMMMMTTTT---`
Sequence	`AERSLSGLTEEEAIAVHDQFKTTFSAFIILAAVAHVLVWVWKPWF`
Secondary structure	`CCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCC`

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

Construction of α and β Apoproteins

Based on the predicted secondary structure feature, tertiary structures for both the α- and β-polypeptides were built by means of comparative modeling using the program MODELLER [**SB93**]. 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 II, were employed in such a procedure. In MODELLER [**SB93**], 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 structure 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. Specifically, two transmembrane helices (one for the α-apoprotein and another for the β-apoprotein) plus the interfacial helix of the α-polypeptide were built with the program MODELLER. To build the [transmembrane helix – reverse-turn – interfacial helix] motif for the α-polypeptide, 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 II. 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 backbones.

Placement of BChla's and Construction of the $\alpha\beta$ 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 [**GLRM93**]. Multiple sequence alignment of light-harvesting complexes (see Figure 3) reveals 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 [**BRY88, CVF$^+$94, VCF$^+$94, Zub86**]. 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 [**BRY88**], 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-polypeptide. The conserved Phe residue in the -8 position to His is also significant [**ZB91**]. 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 [**DOR$^+$93, vHM90**]. The highly conserved residues, for the α-polypeptide (α-Ala-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-polypeptide (β-Phe-27... β-Ala-31...β-His-35). Taking into account the conserved residues in both the α- and the β-polypeptide, 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-polypeptide, face inward and the other side is in contact with lipid bilayers.

It has been observed that many transmembrane α helices are amphipathic, with opposing polar and nonpolar faces oriented along the long axis of the helix [**RDE89,

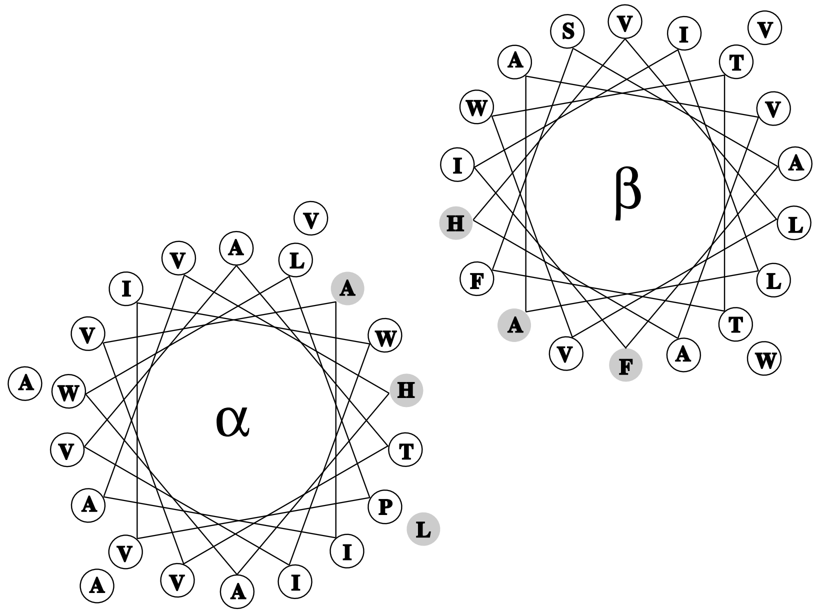

FIGURE 7. Proposed interhelical arrangement of α- and β-polypeptides. The residues in shaded circles are highly conserved.

vHM90, SLD$^+$90]. 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 [**RDE89**]. In our case, transmembrane cores for both the α- and 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-polypeptides of LH–I of *Rs. rubrum* on the cytoplasmic side of the membrane by partial hydrolysis with proteolytic enzymes [**BWS$^+$84**], both α- and β-polypeptides should be oriented with their N-terminals towards the cytoplasm; (2) strong circular dichroism (CD) signals suggest exciton interactions between pairs of BChla molecules [**Zub93**]; (3) linear dichroism data indicate that the B850 BChla's are approximately perpendicular to the membrane and the B800 BChla is parallel to the membrane [**KvGH$^+$84**]; (4) Fourier-transform Resonance Raman spectroscopy and site-directed mutagenesis [**FSRH94**] 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. To put it simply: (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 is 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 was based on the two stage model suggested in [**PE90**] 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 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_α; by applying in the turn region the harmonic restraints to the dihedral angles of the main chain. No restraint was applied to 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* [**DEM$^+$85**] 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 polypeptide 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 ^{o}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 [**Brün92, McR93, Lat85**] of the X-ray diffraction data indicated that, most likely, the LH–II complex possesses a pseudo-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 the 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 [**FSRH94**]. 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 the 2-acetyl group of BChla binding to β-His-35. Since open

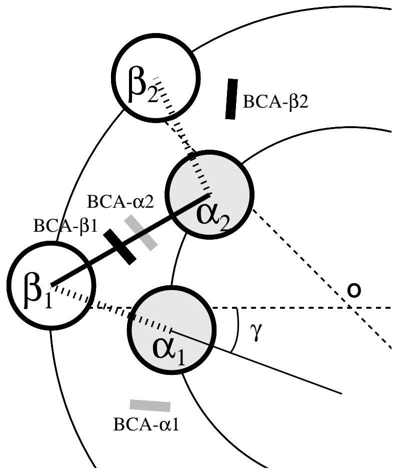

FIGURE 8. Definition of $\alpha\beta$ pair geometries. Heterodimer: $\alpha\beta$ dimer formed between β_1 and α_2, linked by a solid line; Neighboring pair: $\alpha\beta$ dimer formed between β_1 and α_1, linked by a dashed line; γ is defined as the angle between the vector linking two helices and the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

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 the 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 ^{o}K and slowly cooling it down. This procedure was followed by a 100 ps molecular dynamics run at 300 ^{o}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 ^{o}K, then a 5 ps molecular dynamics run at 300 ^{o}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 begins 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 practical solution of which still 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 (1) arranging two kinds of helical assembly; (2) varying the relative orientations (angle γ in Figure 8) of the vector linking two helices and the radial vector from the center of octamer to the center of neighboring pair. There are two options to assemble the eightfold octamer ring in terms of spatial positions of the α- and the β-apoproteins. One is α-polypeptide inside and β-polypeptide outside, and another is β-polypeptide inside and α-polypeptide outside. Since there exist no direct biochemical data in favor of one of the two, both arrangements were considered in our modeling. Angle γ was sampled from -50^o to 50^o with a step size of 10^o. Thus, 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 dimension or because of severe distortion of pairing of B850 BChla's. The rest of the optimized structures were fed to the molecular replacement procedure. The best predicted structure was determined by the molecular replacement procedure as described in the following session. The structure that gave rise to the lowest R value was deemed the most optimal one.

Figure 9 shows one of the octamers we optimized. The octamer forms a ring with an outer diameter of about 70 Å with α-polypeptide inside and β-polypeptide outside. This 70 Å outer diameter covers the side chains of the β-apoprotein. Typically, the radius of the eight β-apoproteins (measured from the center of the helices) varied between 27.8 Å and 32.8 Å; the radius of the eight α-apoproteins varied between 15.1 Å and 21.0 Å. The Mg^{2+} to Mg^{2+} separation in the $\alpha\beta$ heterodimer was about 8.8 Å. The Mg^{2+} to Mg^{2+} 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 optimally oriented to capture light coming from all directions and are close enough for efficient energy transfer. 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 [**KBG95**]. 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. As stated above, the placement of BChla at β-His-17 is itself problematic. The reason we placed it

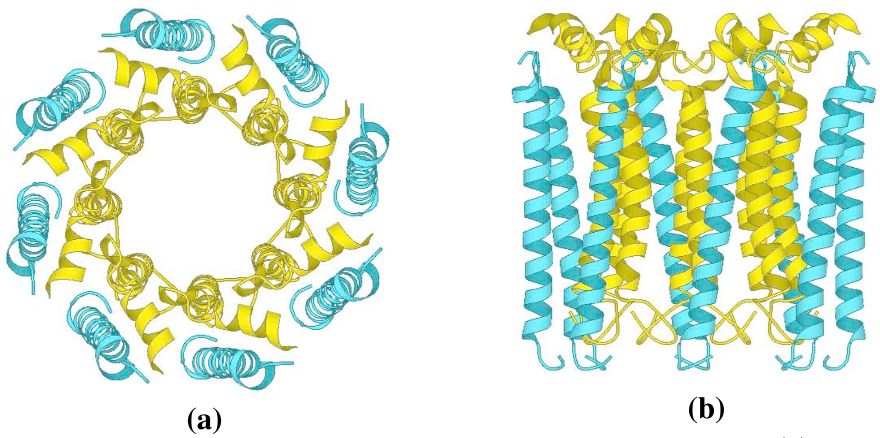

FIGURE 9. Schematic structure of the LH–II complex: (a) top view with C terminal pointing upward, showing a ring structure with 70 Å diameter, with α-polypeptide inside and β-polypeptide outside; (b) side view with C terminal on top. [produced with MOLSCRIPT [**Kra91**]].

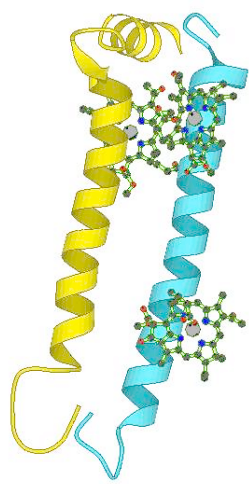

FIGURE 10. Schematic drawing of the $\alpha\beta$ heterodimer, i.e., the monomer "unit" of LH–II. Shown are the predicted secondary structure and locations of three Bchla's binding to α-34 His, β-17 His and β-35 His. The C terminal is on top.

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 were 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 molecular replacement test was 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 [**Brün92**]. 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 [**Lat85, Ros72**]. We also performed a Patterson correlation refinement preceding the translational search to filter the peaks of the rotational search [**Brün90**]. 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 oriented the α 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. 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 on 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 will be 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 [**Brün90, Brün92**] 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 demonstration of use of 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 article is based on an article we recently submitted for publication in Protein Science [**HXH$^+$**].[1]

As this article was written, the crystal structure of LH–II from *Rps. acidophila* determined by conventional multiple isomorphous replacement method was published by McDermott et al. [**MPF$^+$95**]. There exists significant sequence homology between the α-subunit from LH–II of *Rps. acidophila* strain 10050 and *Rs. molischianum*. For the β-subunit, the two sequences are homologous, but to a lesser extent. In Table V, a comparison between the predicted secondary structure assignment for *Rs. molischianum* and the X-ray resolved secondary structure for *Rps. acidophila* [**MPF$^+$95**] is given.

[1] Modified versions of this article have been submitted to *Protein Science* and will also appear as a chapter in "Membrane Structure and Dynamics", K. M. Merz and B. Roux, Eds., (Birkhäuser, Cambridge, to be published).

Table V. Comparison of Secondary Structures

	$\alpha - polypeptide$
Rs. molischianum	SNPKDDYKIWLVINPSTWLPVIWIVATVVAIAVHAAVLAAPGSNWIALGAAKSAAK
Prediction	CCCCCCCCEEEEEECCHHHHHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHC
Rps. Acidophila	---MNQGKIWTVVNPAIGIPALLGSVTVIAILVHLAILSHTTWFPAYWQGGVKKAA
X-ray	---C33333333HHHHHHHHHHHHHHHHHHHHHHHHHHHCCCHHHHHHH???????
	$\beta - polypeptide$
Rs. molischianum	AERSLSGLTEEEAIAVHDQFKTTFSAFIILAAVAHVLVWVWKPWF
Prediction	CCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCC
Rps. Acidophila	-----ATLTAEQSEELHKYVIDGTRVFLGLALVAHFLAFSATPWLH
X-ray	-----CCCCCChhhhHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCC

Note: H - Helix; E - β-Sheet; C - Coil; 3 - 3 Turn Helix; ? - Unresolved.

Given the fact that substantial sequence homology exists between LH–II of *Rs. molischianum* and that of *Rps. acidophila*, the predicted secondary structure of LH–II of *Rs. molischianum* compares well with that of the *Rps. acidophila* crystal structure. Interestingly, the predicted [transmembrane helix – reverse-turn – interfacial helix] motif for the α-polypeptide 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 [**MPF$^+$95**]. 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" [**MPF$^+$95**]. 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^{2+} to Mg^{2+} 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 in the placement of BChla's. While our predicted structure places 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. Perhaps the biggest 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, and 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 2.4 Å resolution X-ray diffraction data for LH–II of *Rs. molischianum* are resolved.

References

[A.] A. MacKerell *et al.*, (unpublished).

[AAM+94] I.T. Arkin, P.D. Adams, K.R. MacKenzie, M.A. Lemmon, A.T. Brünger, and D.M. Engelman, *Structural organization of the pentameric transmembrane α-helices of phospholamban, a cardiac ion channel*, Embo Journal **13** (1994), 4757.

[AGM+90] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, *Basic local alignment search tool*, J. Mol. Biol. **215** (1990), 403–410.

[ARH82] P. Argos, J.K. Rao, and P.A. Hargrave, *Structural prediction of membrane-bound proteins.*, European Journal of Biochemistry **128** (1982), 565.

[AYKR87] J.P. Allen, T.O. Yeates, H. Komiya, and D.C. Rees, *Structure of the reaction center from Rhodobacter sphaeroides R-26: the protein subunits*, Proc. Natl. Acad. Sci. U.S.A. **84** (1987), 6162.

[BBO+83] Bernard R. Brooks, Robert E. Bruccoleri, Barry D. Olafson, Da vid J. States, S. Swaminathan, and Martin Karplus, *CHARMm: a program for macromolecular energy, minimization, and dynamics calculations*, J. Comp. Chem. **4** (1983), no. 2, 187–217.

[Brün90] Axel T. Brünger, *Extension of molecular replacement: A new search strategy based on Patterson correlation refinement*, Acta Cryst. **A46** (1990), 46–57.

[Brün92] Axel T. Brünger, *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, 1992.

[BRY88] E.J. Bylina, S.J. Robles, and D.C. Youvan, *Directed mutations affecting the putative bacteriochlorophyll-binding sites in the light-harvesting I antenna of* Rhodobacter capsulatus, Israel Journal of Chemistry **28** (1988), 73.

[BSST87] T. L. Blundell, B. L. Sibanda, M. J. Sternberg, and J. M. Thornton, *Knowledge-based prediction of protein structures and the design of novel molecules*, Nature **326** (1987), 347.

[Bus86] B. Busetta, *Examination of folding patterns for predicting protein topologies.*, Biochimica et Biophysica Acta **870** (1986), 327.

[BVC+93] A. F. Boonstra, R. W. Visschers, F. Calkoen, R. van Grondelle, E. F. van Bruggen, and E. J. Boekema, *Structural characterization of the B800-850 and B875 light-harvesting antenna complexes from Rhodobacter-Sphaeroides by electron microscopy.*, Biochimica et Biophysica Acta **1142** (1993), 181.

[BWS+84] R. A. Brunisholz, V. Wiemken, F. Suter, R. Bachofen, and H. Zuber, *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** (1984), 689.

[CAKF86] F. E. Cohen, R. M. Abarbanel, I. D. Kuntz, and R. J. Fletterick, *Turn prediction in proteins using a pattern-matching approach.*, Biochemistry **25** (1986), 266.

[CCM+87] J.L. Cornette, K.B. Cease, H. Margalit, J.L. Spouge, J.A. Berzofsky, and C. DeLisi, *Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins*, Journal of Molecular Biology **195** (1987), 659.

[CEHR92] W.A. Cramer, D.M. Engelman, G. Von Heijne, and D.C. Rees, *Forces involved in the assembly and stabilization of membrane proteins*, Faseb Journal **6** (1992), 3397.

[CF78] P.Y. Chou and G.D. Fasman, *Prediction of the secondary structure of proteins from their amino acid sequence.*, Advances in Enzymology and Related Areas of Molecular Biology **47** (1978), 45.

[CSR+92] S. W. Cowan, T. Schirmer, G. Rummel, M. Steiert, R. Ghosh, R. A. Pauptit, J. N. Jansonius, and J. P. Rosenbusch, *Crystal structures explain functional properties of 2 E. coli porins*, Nature **358** (1992), no. 6389, 727–733.

[CSW94] T. Coleman, D. Shalloway, and Z. Wu, *A parallel build-up algorithms for global energy minimizations of molecular clusters using effective energy simulated annealing*, J. of Global Optimization **4** (1994), 171.

[CVF+94] W. Crielaard, R.W. Visschers, G.J. Fowler, R. van Grondelle, K.J. Hellingwerf, and C.N. Hunter, *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** (1994), 473.

[DEM+85] J. Deisenhofer, O. Epp, K. Mikki, R. Huber, and H. Michel, *Structure of the protein subunits in the photosynthetic reaction center of* Rhodopseudomonas viridis *at 3 Å resolution*, Nature **318** (1985), 618–624.

[DHS84] J. Devereux, P. Haeberli, and O. Smithies, *A comprehensive set of sequence analysis programs for the vax*, Nucleic Acids Research **12** (1984), 387.

[DM89] J. Deisenhofer and H. Michel, *The photosynthetic reaction centre from the purple bacterium Rhodopseudomonas viridis*, EMBO J. **8** (1989), 2149.

[DOR+93] D. Donnelly, J. P. Overington, S. V. Ruffle, J. H. Nugent, and T. L. Blundell, *Modeling α-helical transmembrane domains: the calculation and use of substitution tables for lipid-facing residues*, Protein Science **2** (1993), 55.

[Eis84] D. Eisenberg, *Three-dimensional structure of membrane and surface proteins*, Annual Review of Biochemistry **53** (1984), 595.

[ESG86] D. M. Engelman, T. A. Steitz, and A. Goldman, *Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins*, Ann. Rev. Biophys. Biophys. Chem. **15** (1986), 321–353.

[ESKW84] D. Eisenberg, E. Schwarz, M. Komaromy, and R. Wall, *Analysis of membrane and surface protein sequences with the hydrophobic moment plot*, Journal of Molecular Biology **179** (1984), 125.

[EWTW82] D. Eisenberg, R.M. Weiss, T.C. Terwilliger, and W. Wilcox, *Hydrophobic moments and protein structure*, Faraday Symposia of the Chemical Society **17** (1982), 109.

[Fas89a] G.D. Fasman (ed.), *Prediction of protein structure and the principles of protein conformation*, Plenum, New York, 1989.

[Fas89b] G.D. Fasman, *Protein conformational prediction*, Trends in Biochemical Sciences **14** (1989), 295.

[FSRH94] G.J.S. Fowler, G.D. Sockalingum, B. Robert, and C.N. Hunter, *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** (1994), 695.

[GD94] C. Geourjon and G. Deleage, *SOPM: a self optimised prediction method for protein secondary structure prediction.*, Protein Engineering **7** (1994), 157.

[GJ79] M.R. Garey and D.S. Johnson, *Computers and intractability: A guide to the theory of np-completeness*, Freeman, San Francisco, 1979.

[GLRM93] L. Germeroth, F. Lottspeich, B. Robert, and H. Michel, *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** (1993), 5615–5621.

[GOR78] J. Garnier, D.J. Osguthorpe, and B. Robson, *Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins*, Journal of Molecular Biology **120** (1978), 97.

[HBC+90] R. Henderson, J. M. Baldwin, T. A. Ceska, F. Zemlin, E. Beckmann, and K. H. Downing, *Model for the structure of Bacteriorhodopsin based on high-resolution electron cryo-microscopy*, J. Mol. Biol. **213** (1990), 899–929.

[HC91] A. M. Hawthornthwaite and R. J. Cogdell, *Bacteriochlorophyll binding proteins*, Chlorophylls (Boca Raton) (H. Scheer, ed.), CRC Press, 1991, pp. 493–528.

[HK89] L. H. Holley and M. Karplus, *Protein secondary structure prediction with a neural network*, Proc. Natl. Acad. Sci. USA **86** (1989), 152–156.

[HK91] L. H. Holley and M. Karplus, *Neural networks for protein structure prediction*, Methods in Enzymology **202** (1991), 204.

[HXH+] Xiche Hu, Dong Xu, Kenneth Hamer, Klaus Schulten, Juergen Koepke, and Hartmut Michel, *Predicting the structure of the light-harvesting complex II of Rhodospirillum molischianum*, Protein Science (in press).

[J8̈9] F. Jähnig, *Structure prediction for membrane proteins*, Prediction of protein structure and the principles of protein conformation (New York) (G.D. Fasman, ed.), Plenum, 1989, p. 707.

[JSSB94] M.S. Johnson, N. Srinivasan, R. Sowdhamini, and T.L. Blundell, *Knowledge-based protein modeling*, Critical Reviews in Biochemistry and Molecular Biology **29** (1994), 1.

[KBG95] S. Karrasch, P.A. Bullough, and R. Ghosh, *8.5 Å projection map of the light-harvesting complex I from* Rhodospirillum rubrum *reveals a ring composed of 16 subunits*, EMBO J. **14** (1995), 631.

[KD82] J. Kyte and R. F. Doolittle, *A simple method for displaying the hydropathic character of a protein*, J. Mol. Biol. **157** (1982), 105.

[KGvdB+92] W. Kleinekofort, L. Germeroth, J. van der Broek, D. Schubert, and H. Michel, *The light-harvesting complex II (B800/850) from Rhodospirillum molischianum is an octamer*, Biochimica et Biophysica Acta **1140** (1992), 102–104.

[KL85] L.A. Kuhn and J.S.J. Leigh, *A statistical technique for predicting membrane protein structure*, Biochimica et Biophysica Acta **828** (1985), 351.

[KM] J. Koepke and H. Michel, (unpublished).

[Kra91] P. Kraulis, *MOLSCRIPT—a program to produce both detailed and schematic plots of protein structures*, J. Appl. Cryst. **24** (1991), 946–950.

[KS83] W. Kabsch and C. Sander, *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features*, Bioploymers **22** (1983), 2577.

[KvGH+84] H. J. M. Kramer, R. van Grondelle, C. N. Hunter, W. H. J. Westerhuis, and J. Amesz, *Pigment organization of the B800-850 antenna complex of Rhodopseudomonas sphaeroides*, Biochim. Biophys. Acta **765** (1984), 156–165.

[KWF94] W. Kühlbrandt, D.-N. Wang, and Y. Fujiyoshi, *Atomic model of plant light-harvesting complex by electron crystallography*, Nature **367** (1994), 614.

[Lat85] E. Lattman, *Diffraction methods for biological macromolecules. Use of the rotation and translation functions.*, Methods in Enzymology **115** (1985), 55.

[Lev78] M. Levitt, *Conformational preference of amino acids in globular proteins*, Biochemistry **17** (1978), 4277.

[LP85] D.J. Lipman and W.R. Pearson, *Rapid and sensitive protein similarity searches*, Science **227** (1985), 1435.

[LSBW94] R. Lohmann, G. Schneider, D. Behrens, and P. Wrede, *A neural network model for the prediction of membrane-spanning amino acid sequences*, Protein Science **3** (1994), 1597.

[McR93] D.E. McRee, *Practical protein crystallography*, Academic Press, San Diego, 1993.

[Mic91] H. Michel, *General and practical aspects of membrane protein crystallization*, Crystallization of membrane proteins (Boca Raton, Florida) (H. Michel, ed.), CRC Press, 1991, p. 74.

[MPF+95] G. Mcdermott, S.M. Prince, A.A. Freer, A.M. Hawthornthwalte-Lawless, M.Z. Paplz, R.J. Cogdell, and N.W. Isaacs, *Crystal structure of an integral membrane light-harvesting complex from photosynthetic bacteria*, Nature **374** (1995), 517.

[MWG+86] H. Michel, K. A. Weyer, H. Gruenberg, I. Dunger, D. Oesterhelt, and F. Lottspeich, *The 'light' and 'medium' subunits of the photosynthetic reaction centre from Rhodopseudomonas viridis: Isolation of genes, nucleotide and amino acid sequence*, EMBO J. **5** (1986), 1149.

[NVS91] A. Nayeem, J. Vila, and H.A. Scheraga, *A comparative study of the simulated-annealing and monte carlo-with-minimization approaches to the minimum-energy structure of polypeptides: [met]-enkephalin.*, J. of Computational Chemistry **12** (1991), 594.

[OH94] J. D. Olsen and C. N. Hunter, *Protein structure modelling of the bacterial light-harvesting complex*, Photochem. Photobiol. **60** (1994), 521.

[PA94] B. Persson and P. Argos, *Prediction of transmembrane segments in proteins utilising multiple sequence alignments*, Journal of Molecular Biology **237** (1994), 182.

[PCC92] S. R. Presnell, B. I. Cohen, and F. E. Cohen, *A segment-based approach to protein secondary structure prediction*, Biochemistry **31** (1992), 983.

[PdV90] J. Popot and C. de Vitry, *On the microassembly of integral membrane proteins*, Annual Review of Biophysics and Biophysical Chemistry **19** (1990), 369.

[PdVA94] J.L. Popot, C. de Vitry, and A. Atteia, *Folding and assembly of integral membrane proteins: An introduction*, Membrane protein structure: experimental approaches (New York) (S.H. White, ed.), Oxford University press, 1994, p. 41.

[PE90] J.L. Popot and D.M. Engelman, *Membrane protein folding and oligomerization: the two-stage model*, Biochemistry **29** (1990), 4031.

[Pea90] W.R. Pearson, *Rapid and sensitive sequence comparison with FASTP and FASTA*, Methods in Enzymology **183** (1990), 63.

[Pop93] J.L. Popot, *Integral membrane protein structure - transmembrane α-helices as autonomous folding domains.*, Current Opinion In Structural Biology **3** (1993), 532.

[RA86] J.K. Mohana Rao and P. Argos, *A conformational preference parameter to predict helices in integral membrane proteins*, Biochimica et Biophysica Acta **869** (1986), 197.

[RC93] C.S. Ring and F.E. Cohen, *Modeling protein structures: construction and their applications*, Faseb Journal **7** (1993), 783.

[RDE89] D.C. Rees, L. DeAntonio, and D. Eisenberg, *Hydrophobic organization of membrane proteins*, Science **245** (1989), 510.

[Ros72] M.G. Rossmann (ed.), *The molecular replacement method*, Gordon and Breach, New York, 1972.

[RW88] M.J. Rooman and S.J. Wodak, *Identification of predictive sequence motifs limited by protein structure data base size*, Nature **335** (1988), 45.

[SAL91] G.D. Schuler, S.F. Altschul, and D.J. Lipman, *A workbench for multiple alignment construction and analysis*, Proteins: Structure, Function, and Genetics **9** (1991), 180.

[SB93] A. Sali and T. L. Blundell, *Comparative protein modelling by satisfaction of spatial restraints*, Journal of Molecular Biology **234** (1993), 779.

[SLD$^+$90] J.P. Segrest, H. De Loof, J.G. Dohlman, C.G. Brouillette, and G.M. Anantharamaiah, *Amphipathic helix motif: Classes and properties*, Proteins, Struct. Funct. Genet. **8** (1990), 103.

[SvG91] V. Sundstrom and R. van Grondelle, *Dynamics of excitation energy transfer in photosynthetic bacteria*, Chlorophylls (Boca Raton) (H. Scheer, ed.), CRC Press, 1991, pp. 627–704.

[TEPL94] P. Tuffery, C. Etchebest, J.L. Popot, and R. Lavery, *Prediction of the positioning of the seven transmembrane α-helices of bacteriorhodopsin. a molecular simulation study*, Journal of Molecular Biology **236** (1994), 1105.

[TSD$^+$88] Herbert Treutlein, Klaus Schulten, J. Deisenhofer, H. Michel, Axel Brünger, and Martin Karplus, *Molecular dynamics simulation of the primary processes in the photosynthetic reaction center of* Rhodopseudomonas viridis, The Photosynthetic Bacterial Reaction Center: Structure and Dynamics (New York) (J. Breton and A. Verméglio, eds.), NATO ASI Series A: Life Sciences, vol. 149, Plenum, New York, 1988, pp. 139–150.

[VCF$^+$94] R. W. Visschers, W. Crielaard, G. J. Fowler, C. N. Hunter, and R. van Grondelle, *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** (1994), 483.

[vGS88] R. van Grondelle and V. Sundstrom, *Excitation energy transfer in photosynthesis*, Photosynthetic Light–Harvesting Systems (Berlin, New York) (H. Scheer, ed.), Walter de Gruyter & Co., 1988, pp. 403–438.

[vH88] G. von Heijne, *Transcending the impenetrable: how proteins come to terms with membranes*, Biochimica et Biophysica Acta **947** (1988), 307.

[vH92] G. von Heijne, *Membrane protein structure prediction—hydrophobicity analysis and the positive-inside rule*, Journal of Molecular Biology **225** (1992), 487.

[vH94a] G. von Heijne, *Decoding the signals of membrane protein sequence*, Membrane protein structure: experimental approaches (New York) (S.H. White, ed.), Oxford University press, 1994, p. 27.

[vH94b] G. von Heijne, *Membrane proteins: from sequence to structure.*, Annual Review of Biophysics and Biomolecular Structure **23** (1994), 167.

[vHM90] G. von Heijne and C. Manoil, *Membrane proteins: from sequence to structure*, Protein Engineering **4** (1990), 109.

[Whi94] S. H. White, *Hydropathy plots and the prediction of membrane protein topology*, Membrane protein structure: experimental approaches (New York) (S. H. White, ed.), Oxford University press, 1994.

[WKN+91] M.S. Weiss, A. Kreusch, U. Nestel, W. Welte, J. Weckesser, and G.E. Schulz, *The structure of porin from* Rhodobacter capsulatus *at 1.8 Å resolution*, FEBS Lett. **280** (1991), 379.

[ZB91] H. Zuber and R.A. Brunisholz, *Structure and function of antenna polypeptides and chlorophyll-protein complexes: Principles and variability*, Chlorophylls (Boca Raton) (H. Scheer, ed.), CRC Press, 1991, pp. 627–692.

[ZBTS87] M. J. Zvelebil, G. J. Barton, W. R. Taylor, and M. J. Sternberg, *Prediction of protein secondary structure and active sites using the alignment of homologous sequences*, J. Mol. Biol. **195** (1987), 957.

[Zub85] H. Zuber, *Structure and function of light-harvesting complexes and their polypeptides*, Photochem. Photobiol. **42** (1985), 821.

[Zub86] H. Zuber, *Structure of light-harvesting antenna complexes of photosynthetic bacteria, cyanobacteria and red algae*, Trends Biochem. Sci. **11** (1986), 414.

[Zub93] H. Zuber, *Structural features of photosynthetic light-harvesting systems*, The Photosynthetic Reaction Center (San Diego) (J. Deisenhofer and J. R. Norris, eds.), Academic Press, 1993, p. 43.

Theoretical Biophysics, Beckman Institute, University of Illinois at Urbana—Champaign, Urbana, IL 61801

E-mail address: `xhu@ks.uiuc.edu`
E-mail address: `xu@ks.uiuc.edu`
E-mail address: `khamer@ks.uiuc.edu`
E-mail address: `kschulte@ks.uiuc.edu`

Max-Planck-institut für Biochemie, Abteilung Molekulare Membranbiologie, 6000 Frankfurt, Germany

E-mail address: `koepke@max.mpibp.uni-frankfurt.de`