Partial site-specific assignment of a uniformly $^{13}$C, $^{15}$N enriched membrane protein, light-harvesting complex 1 (LH1), by solid state NMR

Lei Huang a,b, Ann E. McDermott b,*

a Key Laboratory of Molecular Engineering of Polymers of the Ministry of Education, Department of Macromolecular Science, Fudan University, Shanghai 200433, China
b Columbia University, Department of Chemistry, 3000 Broadway, New York, NY 10027, USA

Received 22 May 2007; received in revised form 2 January 2008; accepted 3 January 2008

Available online 26 January 2008

Abstract

Partial site-specific assignments are reported for the solid state NMR spectra of light-harvesting complex 1, a 160 kDa integral membrane protein. The assignments were derived from 600 MHz $^{15}$N–$^{13}$CO–$^{13}$C$\alpha$ and $^{15}$N–$^{13}$C$\alpha$–$^{13}$CX correlation spectra, using uniformly $^{13}$C, $^{15}$N enriched hydrated material, in an intact and precipitated form. Sequential assignments were verified using characteristic $^{15}$N–$^{13}$C$\alpha$–$^{13}$CB side chain chemical shifts observed in 3D experiments. Tertiary contacts found in 2D DARR spectra of the selectively $^{13}$C enriched sample provided further confirmatory evidence for the assignments. The assignments include the region of the Histidine ligands binding the Bacteriochlorophyll chromophore. The chemical shifts of C$\alpha$ and C$\beta$ resonances indicated the presence of typical $\alpha$-helical secondary structure, consistent with previous studies.

© 2008 Elsevier B.V. All rights reserved.

Keywords: Light-harvesting complex 1; Solid state NMR; Membrane protein; Sequential assignment

1. Introduction

Plants, algae and bacteria convert light energy to chemical energy photosynthetically: photons are absorbed by the light-harvesting complex (LHC) whose excitation energy is then transferred to the reaction center (RC), leading to a charge separation across the photosynthetic membrane. Of the many photosynthetic systems, those of purple bacteria have been the best characterized, and their membrane proteins were among the first to be described at atomic resolution [1–4]. Most purple bacteria have two types of LHCs, light-harvesting complex 1 (LH1) and light-harvesting complex 2 (LH2), which are replaced in some bacteria under low-light conditions by light-harvesting complex 3 (LH3) [5]. The LH2 complex from Rhodopseudomonas (Rps.) acidophila provided the first high-resolution crystal structure of purple bacterial LHC [6,7]. The diffraction data showed two concentric cylinders of helical protein subunits that enclose the pigment molecules. Nine helical $\alpha$ subunits are packed side by side to form an inner cylinder of radius 18 Å. The nine helical $\beta$ subunits are arranged near each of the $\alpha$ subunits.
to form an outer cylinder with a radius of 34 Å. Three Bchl rings are present for each αβ dimer in LH2, as contrasted with the two found in LH1. Recently, the crystal structure of the LH1–RC from *Rps. palustris* (Fig. 1b) has been characterized by X-ray diffraction and resolved to 4.8 Å [8]. At this resolution, the electron density map provided structural information, such as the positions and angles of the helices, but atomic or molecular level structural information, such as positions of pigments (carotenoids and Bchls), could not be reliably derived. The data show a reaction center surrounded by a double LH1 ring, 16 transmembrane α subunits on the inside ring and 15 transmembrane β subunit on the outside ring, arranged in an approximately elliptical form. In general the peptides can be viewed as forming an αβ dimer, but the repetition of αβ subunits is not believed to be perfectly 16 fold symmetric, because in one location there is single α helix located in the inner ring, without a counterpart in the outer ring; in other words the outer ring is not really closed. In contrast, the previous EM studies on LH1, 1108

although substantial solution NMR studies have been carried out on single polypeptides from LHCs, the integral complex is a formidable size for solution NMR. Solid state NMR becomes a promising alternative when a protein’s size approaches the limit of the current solution NMR techniques. In contrast to solution NMR, where a long history exists for the structural study of soluble proteins, solid state NMR of biological macromolecules is still in an early stage of development. It is encouraging that a few proteins [18,19], have been solved to high resolution by solid state NMR. Significant progress has been made recently instituted by a helical membrane protein PufX, which also plays key roles in the quinol/quinone exchange between RC and cytochrome *bc1* complex [13,14]. To clarify the LH1–RC arrangement and provide atomic level details, further evidence on the structure at higher resolution is necessary.

Three solution NMR structures of an individual LH1 α or β polypeptide have been deposited in the Protein Data Bank [15,16]. These three studies were performed using protein samples dissolved in organic solvents (a mixture of chloroform and methanol), without chromophores present. The structure of the LH1 β polypeptide (PDB file: 1DX7) of *Rb. sphaeroides* determined by solution NMR spectroscopy comprised two α helical regions joined by a flexible four-residue linker [15]. This structure was later confirmed by a solution NMR study in Zwittergent micelles (PDB file: 1JO5) [17]. The LH1 α (PDB file: 1XRD) polypeptide of wild-type purple photosynthetic bacterium *Rs. rubrum* had a short helical segment at its N terminus, followed by a five-residue β turn and a long 32 amino acid helix spanning the transmembrane domain [16]. The β polypeptide (PDB file: 1WRG) of the *Rs. rubrum* had a single helix of 32 amino acids in the transmembrane region with short turns on the two termini [16].

Although substantial solution NMR studies have been carried out on single polypeptides from LHCs, the integral complex is a formidable size for solution NMR. Solid state NMR becomes a promising alternative when a protein’s size approaches the limit of the current solution NMR techniques. In contrast to solution NMR, where a long history exists for the structural study of soluble proteins, solid state NMR of biological macromolecules is still in an early stage of development. It is encouraging that a few proteins [18,19], have been solved to high resolution by solid state NMR. Significant progress has been made recently in...
exploring assignment strategies and structure determination methodologies for proteins in the solid state. Nearly complete and partial protein sequence assignments have been reported for dimeric Crh [20], BPTI [21], Thioredoxin [22], GB1 [23], Mastoparan-X [24] and others. Also, an interesting indirect $^1H-^1H$ distance detection method resulted in the determination of Kaliotoxin structure, although at a lower resolution than direct $^{13}C$, $^{15}N$ detection methods [25].

Following on from the encouraging work on soluble proteins, several solid state NMR studies explored the possibility of studying insoluble membrane proteins or protein aggregates, an area where significant additional challenges and significant scientific gains can be expected. Initial residual type assignments have been carried out for colicin la channel [26] and outer-membrane protein G [27]. The full assignment of coat protein PfI was recently accomplished [28]. LH2 complexes have a similar molecular arrangement as LH1, in that there is significant sequence homology, and both form two concentric rings, with different size. The integral membrane protein LH2 of Rps. acidophila is an ($\alpha\beta)_5$ nonamer [6] of 94 residues $\alpha$ helical asymmetric unit. With four different site-specific biosynthetically isopropyl enriched LH2 samples, 76 residues have been assigned based on $^{13}C-^{13}C$ spin diffusion and band selective $^{13}C-^{15}N$ correlation experiments [29,30].

In this paper, we employed two 3D $^{15}N-^{13}C-^{13}C$ correlation spectra on a uniformly $^{13}C$, $^{15}N$ enriched LH1 sample to achieve a partial specific assignment. Characteristic secondary chemical shifts were observed for the assigned residues.

2. Material and methods

2.1. Protein NMR sample preparation

Uniformly $^{13}C$, $^{15}N$ enriched LH1 cells of Rb. sphaeroides were over-expressed from a genomic deletion strain of Rhodobacter sphaeroides (SK102), which was modified to eliminate genes of LH2, RC and PuX, as described previously [17]. LH1 protein expressed with this method will form a monomeric αβ dimer or the αβ···dimer, corresponding to the association of Bchl from the αβ dimer and the αβ monomer, respectively. These data are consistent with limited dissociation of intact LH1 rings either during spectroscopy, or more likely during the process of redissolving in detergent. The dissociated sample only accounted for a low percentage in both the before-NMR and the after-NMR sample (less than 5% of sample exists as dissociated forms, $\alpha\beta$ dimer or the $\alpha\beta\beta$ monomer, and the majority of the sample forms an intact LH1 ring). The fine structure and resolution in 1D and 2D NMR spectra provided another indicator of sample stability. Identical 1D and 2D $^{13}C$ spectra were observed before and after lengthy multidimensional NMR experiments, which further confirmed the sample stability.

Singly $^{13}C$ enriched LH1 sample was prepared in the same way as the uniformly $^{13}C$, $^{15}N$ enriched LH1 sample, except that the cells were grown on a medium with $2-^{13}C$ glycerol as the sole source of carbon. When cells were grown on $2-^{13}C$ glycerol, the C bound to carbon in the protein are fully labeled while carbons on the side chains could be fully or selectively labeled, as described previously for E. Coli [18,26,35].

2.2. NMR spectroscopy: sequential assignment strategy and experimental conditions

Partial sequential assignments were achieved using two types of 3D experiments, the polarization transfer pathways of which are indicated in Fig. 2. An NCO/C cx experiments established sequential resonance correlations (Fig. 2a). Polarization was first transferred from nitrogen ($^15N_2$) to the preceding carbonyl ($^13CO_i$) using a selective CP pulse; then homonuclear $^13CO_i-^{13}CO_{i-1}$ transfers were accomplished using the RFDR sequence. The analogous NCO/C cx experiments established intra-residue resonance correlations (Fig. 2b). Here a DARR sequence was used to accomplish a more broadband homonuclear $^{13}CO_i-^{13}CO_{i-1}$ and $^{13}CO_i-^{13}CO_{i+1}$ transfers [21]. $^{13}CO$ resonance pairs were identified in the N$_2$CO, CO intra-residue 3D spectrum and the $N_2$CO/COCO sequential 3D spectra, providing hypothetical sequential readings. Such assignments were extended backwards and forwards along the protein.

Three dimensional (3D) $^{15}N-^{13}C-^{13}C$ and two-dimensional (2D) $^{13}C-^{13}C$ correlation spectra of uniformly $^{13}C$, $^{15}N$ enriched samples were acquired on a Varian Chemagists InfinityPlus 600 MHz spectrometer. The instrument operated at Larmor frequencies of 599.3 MHz for $^1H$, 150.7 MHz for $^{13}C$, and 60.7 MHz for $^{15}N$, using triple resonance Magic Angle Spinning (MAS) probes with rotor sizes of either 3.2 mm or 4 mm. The sample rotation frequency was 13 to 13.5 kHz. The temperature was −30 °C to −35 °C for 3D experiments, and −20 °C for 2D experiments. Temperature was measured at the end of the variable temperature gas flow line; the sample temperature is expected to be higher than that reading due to sample spinning and proton decoupling. 2D $^{13}C-^{13}C$ spectra of selectively $^{13}C$ enriched sample were acquired on a Bruker 750 MHz spectrometer at a temperature of −20 °C.

In 2D DARR [36] experiments, $^{13}C-^{13}C$ cross polarization (CP) was accomplished with a proton field strength of 75.8 kHz and a $^{13}C$ field strength, ramping down from 58.3 kHz to 48.5 kHz during a 1 ms mixing time. Proton irradiation with a field strength of 13.5 kHz was applied during the 11.8 ms
Fig. 2. Pulse sequences and polarization transfer pathways for 3D NCOCA and NCACX experiments: (a) DCP–RFDR pulse sequence for NCOCA spectra, (b) DCP–DARR pulse sequence for NCACX spectra. Red arrows in the peptide sequences indicate heteronuclear inter- (NCOCA spectra) or intra- (NCACX spectra) N–C polarization transfer. Blue arrows indicate intra-residue homonuclear C–C polarization transfer.
DARR mixing period for the uniformly $^{13}$C, $^{15}$N enriched sample. 1024 points were collected in the direct and indirect dimensions with spectral widths of 67 kHz in both dimensions. Experiments on the selectively $^{13}$C enriched samples utilized DARR mixing times of 50, 300 and 500 ms, with proton irradiation of 14 kHz during the DARR mixing. 512 data points were acquired in both dimensions. In the two NCO$^{13}$C and NCOC$^{13}$X 3D experiments [37], the $^{15}$N–$^{13}$CO peaks of each sample were correlated with $^{13}$C–$^{13}$CO for the NCOC$^{13}$X experiments. A2D$^{13}$C transfer was achieved with a $^{15}$N field strength of 33.7 kHz and an average $^{13}$C field strength of 20.2 kHz with a downward ramp of 3 kHz while the carbon excitation was set at 176.43 ppm and the contact time was 5 ms. The $^{15}$N–$^{13}$CO transfer in 3D NCOC$^{13}$X experiments was achieved with a $^{15}$N field strength of 34.5 kHz and an average $^{13}$C field strength of 24.5 kHz with a downward ramping of 3.7 kHz while the carbon excitation was set at 60.4 ppm and the contact time was 3 ms. Homonuclear $^{15}$N–$^{13}$C transfer in NCOC$^{13}$X experiment was accomplished using the TFDR element with a mixing time of 1.18 ms. In Fig. 2(b), the homonuclear $^{13}$C–$^{13}$C transfer in NCOC$^{13}$X experiments was achieved using the DARR sequence element with a mixing time of 11.1 ms and proton irradiation of 13.5 kHz. Both $^{15}$N–$^{13}$CO and $^{13}$C–$^{13}$C correlations were observed using the DARR sequence element in 3D NCOC$^{13}$X experiments. TPPM proton decoupling with a field strength of 75 kHz was applied during the data acquisition. The time proportional phase incrementation (TPPI) method was utilized for phase-sensitive detection. In NCO$^{13}$X experiments, 48 data points were collected on both indirect dimensions, and the sampling of direct dimension was 1024 points. The spectral widths were 66.7 kHz, 3.4 kHz and 3.4 kHz for the $^{13}$CO, $^{15}$N and $^{13}$CO dimensions, respectively. In NCO$^{13}$X experiments, 48 and 96 data points were collected on the indirect $^{15}$N and $^{13}$CO dimensions, respectively, and the 1024 points were taken in the direct dimension. The spectral widths were 66.7 kHz, 3.4 kHz and 6.8 kHz for $^{13}$CX, $^{15}$N and $^{13}$CO dimensions, respectively.

The NMR data were processed with NMRPipe [42] and analyzed with Sparky [43]. Cosine-bell apodization functions were applied for all three dimensions of NCO$^{13}$X experiments. For NCOC$^{13}$X experiments, a Gaussian apodization function with 50 Hz line broadening was applied for the direct-detected dimension, and cosine-bell apodization functions were applied for the two indirect dimensions. For 2D $^{13}$C–$^{13}$C data, a Gaussian apodization function with 60 Hz line broadening was applied for the direct-detected dimension, and a cosine-bell apodization function was applied for the indirect dimension.

The carbon dimension was referenced externally to DSS, using a solid adamantane methylene peak at 40.26 ppm [44]. The nitrogen dimension was referenced externally to 25 °C liquid ammonium using the ammonium chloride peak at 32.53 ppm.

2.3. Optimization of NMR sample conditions:

The precipitation conditions used for the NMR sample can affect protein stability and NMR spectral quality. Since elevated ionic strength in the sample could cause sample heating during NMR experiments, lower salt concentrations are preferred when optimizing the precipitation condition. Different precipitants, salts, and buffers were screened and three conditions were found to produce stable protein precipitations at the lowest salt concentration: (1) 15% PEG400 (polyethylene glycol 400), 50 mM MgCl$_2$, and 25 mM HEPES pH7.5 buffer; (2) 10% PEG6000, 20 mM MgCl$_2$, and 25 mM HEPES buffer pH7.5 and (3) 35% MPD at the same buffer and salt concentration as in (2) (the concentrations of precipitant agents refer to the conditions before mixing the precipitant solution with the protein solution at a 1:1 ratio).

![Fig. 3. 2D DARR spectrum of uniformly $^{13}$C, $^{15}$N enriched LH1 with 11.8 ms mixing time. (a) Full spectrum, (b) expanded CO–CX cross peak region, (c) expanded CO–CX (aliphatic) cross peak region (CX generally refers to any neighboring carbon). In (c), assigned $\alpha$–C$\beta$–CO cross peaks unambiguously indicate an $\alpha$ helical secondary structure based on a $^{13}$C chemical shift index [45]. For example, in the case of Ala and Val (in the square boxes), their peaks represent typical $\alpha$ helical chemical shifts of Ala and Val. In (b), some assigned C$\beta$–CO peaks are correlated with C$\alpha$–C$\beta$ peaks of (c). Also noted in (b), several Ala residues have well resolved CO peaks.](image-url)
To compare the NMR spectral quality for samples under the above three precipitation conditions, two individual peaks were selected from 2D $^{13}$C--$^{13}$C DARR spectra, and their maximum linewidths at half-height were measured to indicate the resolution. Crosspeaks corresponding to $\beta$Ala37 $\beta$CO and $\alpha$CO were selected. The $^{13}$CO chemical shift (183.5 ppm) of the selected Ala is shifted downfield as compared with the rest of 11 Ala and therefore cross peaks

Fig. 4. Amino acid sequences of LH1 $\alpha\beta$ polypeptides (access numbers in SWISS-PROT are P02949 and P02951). One Bchl molecule is bound to each $\alpha$ subunit and one to each $\beta$ subunit. The Bchl binding sites (Histidine sidechains) are labeled in green. Predicted transmembrane regions are underlined, and bold italic letters represent the sequentially assigned residues reported here.

Fig. 5. Strip plots illustrating peaks assigned in the $\alpha$ subunit from $\alpha$Arg14 to $\alpha$Gln20, and in the $\beta$ subunit from $\beta$Val33 to $\beta$Ala40. Seven strips containing NC$\alpha$C$\beta$ peaks are included in the plots to illustrate the basis of the sequential assignments. They are $\alpha$V16 and $\beta$V33, $\beta$A34, $\beta$V36, $\beta$A37, $\beta$L39 and $\beta$A40. Each strip was extracted at the indicated $^{15}$N chemical shift from the corresponding 3D experiments. The vertical axes are $\alpha$C chemical shifts and the horizontal axes represent CO and $\beta$C chemical shifts. The N chemical shifts where the planes are truncated are listed below the CO and $\beta$C chemical shifts. The strip width is 2 ppm.
involving the $^{13}$CO of Ala57 were easily identified. Linewidths for this CO peak are comparatively less affected by the $^1$H decoupling field strength because the lack of directly bonded $^1$H. Experimental parameters for three protein samples are listed in Table 1; equivalent $^1$H–$^{13}$C CP conditions were applied in these three 2D DARR experiments.

The downfield Ala C=O peak was observed with good signal-to-noise ratios in the spectra of the PEG400 and PEG6000 samples, but was not observable in the spectra of MPD-precipitated samples. The signal intensities of the Ala C=O peaks in the MPD spectra were also weaker than those in PEG spectra. It might be concluded that the PEG precipitated samples show a higher spectral sensitivity than MPD sample overall, and the comparison of the $^{13}$CO linewidths suggests that the spectra taken on the PEG400 precipitated sample have the best spectral quality. The assignments reported here were all based on PEG400 precipitated samples.

3. Results and discussion

3.1. Sequential assignments

The majority of the α and β peptides in the membrane protein LH1 have α helical structures. Correspondingly, the dispersion of the chemical shifts is poor and the homonuclear two-dimensional $^{13}$C spectra were poorly resolved (Fig. 3). Nevertheless, based on two selective 3D experiments correlating backbone $^{15}$N resonances to the resonances of neighboring $^{13}$C spin systems, about 15% of the amino acids were sequentially assigned (Fig. 4).

Representative sequential assignments are displayed using strip plots prepared in Sparky [43] (Fig. 5). 2D planes, cut from the 3D spectra (NCOCA and NCACX) at a given $^{15}$N resonance of interest (value indicated at bottom), display the $^{13}$C CO shifts along the horizontal axes and the $^{13}$Cα along the vertical axes. The solid lines guide the eye by connecting the same $^{13}$CO–$^{13}$Cα cross peak through sequential correlations ($^{13}$N–$^{13}$C–$^{13}$Cα) in the NCOCA experiment to the intra-residue correlations ($^{13}$N–$^{13}$C–$^{13}$CO) in the NCACX experiment. Fig. 5a displays a sequential walk from Arg14 to Gln20 of the α subunit. The assignment of the α subunit is interrupted by missing glycine (αG21) and proline (αP13) signals in the 3D spectra. Fig. 5b displays a sequential walk from Val33 to Ala40 of the β subunit. The congestion of these spectra in the valine (βV41) and alanine (βA40) spectral regions prevents further correlations along the amino acid sequence of the β chain. Chemical shifts determined for the sequentially assigned amino acids are listed in Table 2. Some assignments were additionally confirmed using high-resolution DARR experiments of selectively labeled samples; these are highlighted in red. In carrying out the assignments, allowing carrying out the assignments, acceptable deviation of chemical shifts (comparing one spectrum to another) of the assigned peaks was less than 0.3 ppm.

All sequential assignments are located in the transmembrane regions of LH1. Those residues are more conformationally rigid and gave stronger signals compared with those located in the flexible parts of the peptide. The fact that the transmembrane residues were preferentially detected relative to the solvent exposed residues is reminiscent of a previous study on the LH2 complex [46], where all detected signals were also reported to be from the transmembrane residues.

In our 3D NCACX spectra, only about 50 $^{15}$N$^{13}$C$^{13}$CO peaks can be identified out of 106 amino acids that comprise the αβ dimer in LH1 sequence. Even when considering missing signals of proline and glycine, whose cross peaks are particularly weak in these studies, are discounted, the observed peaks only account for half of the sequence. The most likely explanation for the missing signals is the poor dispersion of the spectra, although another explanation for the missing peaks could be conformational exchange. We prefer the former hypothesis based on preliminary data on this system and analogous studies on a closely related system. Studies of uniformly isotopically enriched LH2, closely related to LH1 in both its sequence homology and ring structure, using high field SSNMR allowed assignment of only 16 out of 94 residues [46] whereas use of a variety of selective isotopic enrichment schemes [29,30] allowed essentially a full sequence assignment by reducing spectral congestion. Analogously, 2D CC spectra of 2-13C labeled LH1 (Fig. 6) also showed well resolved peaks and led to the assignments of αHis32, αVal29, βHis20, and βSer21, all of which could not be defined in spectra taken on uniformly labeled samples. We may conclude that the limited number of assignments is caused by the spectral overlap of a uniformly labeled protein sample. Therefore, a selectively labeled LH1 sample is likely to provide better-dispersed spectra and a more complete sequence assignment. In the long run, use of higher magnetic field strengths will be very important, given the narrow line widths. Moreover, better excitation for dynamic residues and for prolines which lack an amide proton, will help to extend the assignments significantly.

3.2. Verification of sequential assignment

Sequential assignments were verified using characteristic Cβ chemical shifts of $^{15}$N$^{13}$C$^{13}$Cβ cross peaks in the intra-residue

Table 2

<table>
<thead>
<tr>
<th>α subunit</th>
<th>N</th>
<th>CO</th>
<th>Cα</th>
<th>Cβ</th>
<th>β subunit</th>
<th>N</th>
<th>CO</th>
<th>Cα</th>
<th>Cβ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg14</td>
<td>118.7</td>
<td>178.6</td>
<td>57.5</td>
<td></td>
<td>Val33</td>
<td>120.9</td>
<td>176.6</td>
<td>67.0</td>
<td>30.8</td>
</tr>
<tr>
<td>Arg15</td>
<td>117.8</td>
<td>177.2</td>
<td>59.6</td>
<td></td>
<td>Ala34</td>
<td>121.3</td>
<td>178.9</td>
<td>55.4</td>
<td>17.8</td>
</tr>
<tr>
<td>Val16</td>
<td>117.7</td>
<td>178.2</td>
<td>66.6</td>
<td>31.4</td>
<td>Ile35</td>
<td>113.3</td>
<td>176.9</td>
<td>61.9</td>
<td></td>
</tr>
<tr>
<td>Phe17</td>
<td>114.4</td>
<td>177.4</td>
<td>56.7</td>
<td></td>
<td>Val36</td>
<td>121.3</td>
<td>176.2</td>
<td>67.3</td>
<td>30.8</td>
</tr>
<tr>
<td>Val18</td>
<td>120.4</td>
<td>176.8</td>
<td>67.0</td>
<td></td>
<td>Ala37</td>
<td>120.9</td>
<td>183.5</td>
<td>55.1</td>
<td>18.9</td>
</tr>
<tr>
<td>Ala19</td>
<td>119.3</td>
<td>179.7</td>
<td>55.9</td>
<td></td>
<td>His38</td>
<td>123.0</td>
<td>178.3</td>
<td>58.7</td>
<td></td>
</tr>
<tr>
<td>Gln20</td>
<td>118.3</td>
<td></td>
<td></td>
<td></td>
<td>Leu39</td>
<td>121.3</td>
<td>176.8</td>
<td>57.9</td>
<td>41.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ala40</td>
<td>120.0</td>
<td>181.6</td>
<td>54.5</td>
<td>18.3</td>
</tr>
</tbody>
</table>

Assignments that are highlighted with red color in the table are confirmed by the high-resolution DARR experiments on selectively labeled protein.
correlation 3D NCoCX spectra. Amino acids whose $^{15}$N$^{13}$C$^{13}$C$\alpha$ peaks were assigned in the 3D spectra were Val16 of the $\alpha$ chain and Val33, Val36, Ala37, Leu39, and Ala40 of the $\beta$ chain. Those $^{15}$N$^{13}$C$^{13}$C$\beta$ peaks are also illustrated in the strip plots of Fig. 5, positioned after the $^{15}$N$^{13}$C$^{13}$CO strips of the same amino acid, and are included in the chemical shift lists of Table 2. Amino acid sidechain resonances are typically diagnostic for the residue type and therefore are very useful in verification of the sequential walk in a congested spectrum. For example, identifications of the $C\beta$ shift for Val16 of the $\alpha$ subunit and Val33, Val36 of the $\beta$ subunit unambiguously confirmed that those peaks belonged to Val, not Ile.

In addition, significant confirmation of the assignment was obtained using sequential contacts identified in the DARR spectra of the selectively $^{13}$C enriched sample using relatively long mixing times (300 and 500 ms). Peaks found in the long mixing time (300 and 500 ms) spectra that lacked corresponding peaks in the short mixing time spectra (below 50 ms) were assumed to be multiple bond contacts. For example, many C$\alpha$–C$\gamma$ cross peaks can be identified and those from aromatic residues were particularly useful in the assignments. In Fig. 6(a), exhibiting the 50 ms DARR spectrum, only three intra-residual histidine cross peaks can be observed, but in Fig. 6(b1), the 300 ms DARR spectrum, the sequential $\beta$A37C$\alpha$–H38C$\gamma$ contact can be identified. In Fig. 6, the verified sequential assignments are highlighted with connecting lines; they are also listed in Table 3. To distinguish among three Val–Ala pairs, differently colored lines are used, starting from lower left of the spectrum, in the VCo–Ac region. Black lines connect $\beta$V36–A37, red lines connect $\beta$V33–A34 and green lines connect $\beta$V41–A40.

### Table 3

<table>
<thead>
<tr>
<th>Observed contacts</th>
<th>Chemical shifts (ppm)</th>
<th>Observed contacts</th>
<th>Chemical shifts (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$R15C–V16C$\beta$</td>
<td>(59, 7, 31.4)</td>
<td>$\beta$A37C–H38C$\alpha$</td>
<td>(55.2, 58.7)</td>
</tr>
<tr>
<td>$\beta$S21C–H20C$\gamma$</td>
<td>(59.8, 128.5)</td>
<td>$\beta$A37C–H38C$\gamma$</td>
<td>(55.2, 128.5)</td>
</tr>
<tr>
<td>$\beta$V33C–A34C$\alpha$</td>
<td>(67.0, 55.3)</td>
<td>$\beta$H38C–L39C$\beta$</td>
<td>(58.9, 41.1)</td>
</tr>
<tr>
<td>$\beta$V33C–A34C$\gamma$</td>
<td>(31.0, 55.2)</td>
<td>$\beta$H38C–L39C$\gamma$</td>
<td>(58.7, 26.4)</td>
</tr>
<tr>
<td>$\beta$A34C–I35C$\alpha$</td>
<td>(55.3, 61.8)</td>
<td>$\beta$L39C–$\beta$A40C$\alpha$</td>
<td>(41.2, 54.8)</td>
</tr>
<tr>
<td>$\beta$L35C–V36C$\beta$</td>
<td>(61.8, 31.5)</td>
<td>$\beta$L40C–$\beta$A41C$\beta$</td>
<td>(54.8, 30.4)</td>
</tr>
<tr>
<td>$\beta$V36C–A37C$\alpha$</td>
<td>(31.6, 55.3)</td>
<td>$\beta$A40C–V41C$\beta$</td>
<td>(67.0, 54.9)</td>
</tr>
<tr>
<td>$\beta$V36C–A37C$\gamma$</td>
<td>(67.3, 55.3)</td>
<td>$\beta$A40C–V41C$\gamma$</td>
<td>(67.0, 54.9)</td>
</tr>
</tbody>
</table>
assigned to every nucleus, and there are no double or multiple chemical shifts observed for any resonance. This proves that amino acids of LH1 sequence have equivalent conformation in the repeating α or β polypeptides of LH1 complex. For an elliptical open ring that forms a dimeric S-shape, double or multiple chemical shifts would be observed for resonances of the same amino acid. This result supports that the LH1 complex used for NMR spectra forms a monomeric circular ring structure.

3.3. Characteristic α-helical secondary chemical shifts and dihedral angles

Chemical shifts of the backbone Cα and CO atoms, and to a lesser extent N and the sidechain Cβ are indicators of a protein secondary structure. When compared with the statistical chemical shifts in BMRB (Biological Magnetic Resonance Bank) and the studies of correlation between chemical shifts and secondary structures [45], the chemical shifts assigned in Fig. 3 clearly indicate an α helical secondary structure for LH1, e.g. all carbonyl chemical shifts are above 173 ppm and AlaCα (55 ppm), AlaCβ (17 ppm), ValCα (66 ppm), ValCβ (21 ppm), LeuCα (58 ppm), LeuCβ (40 ppm) etc. all have typical α helical chemical shifts.

The dihedral φ, ϕ angles of assigned LH1 sequences predicted by the TALOS program [47] also indicated that the assigned sequences have α-helical secondary structures. In Fig. 7, the predicted dihedral angles are plotted. The 13C chemical shifts used as input to TALOS have been adjusted by −0.12 ppm because the chemical shifts used in TALOS are referenced to TSP, whereas our table uses a DSS reference.

Structural studies of membrane proteins are particularly challenging due to the difficulty of crystallizing and even due to the difficulty of solubilizing them with detergents that retain native structure. Few membrane proteins have been resolved to high resolution. The successful partial assignment of the integral membrane protein LH1, alongside the previous structural determination of small soluble proteins using solid state NMR methods [18,19] proves that solid state NMR is a promising tool for determining membrane protein structures at moderate magnetic field strength on an easily prepared uniformly labeled protein sample. LH1 not only is a target for structure determination, but because of its thermal and chemical stability it also provides an excellent system to investigate modern solid state NMR techniques. LH1 samples prove to be stable across a wide range of temperatures and under different precipitation conditions.

It will be important to obtain information about the remainder of the protein. The NMR experiments performed at lower temperature might help to improve the signal intensities of resonances from flexible regions of protein, more likely, as elaborated above, different types of selectively labeled samples will simplify the analysis and allow identification of more residues.

Two previous solution NMR studies on LH1 β polypeptides of *Rb. sphaeroides* [15,17] indicated that each β subunit formed two α-helical regions joined by a four-residue linker. It is a concern however that the position of the linker is slightly different in two studies; it is shifted by two residues comparing these two studies. In their studies, the single peptide was investigated in the absence of Bchl. Studies on LH2 β polypeptides of *Rs. molischianum* [7,48], *Rps. acidophila* [7] and on LH1 α or β polypeptides of *Rs. rubrum* [16] all indicated a single continuous α helix structure. Though we could not identify the linker region from our assignments on LH1 α/β peptide, so our studies do not resolve this dispute, it is important to take seriously the proposal of the linker. Because for LH1 β polypeptide of *Rb. sphaeroides*, the helical segment between 27Leu and 44Trp is embedded in the membrane and the other helical segment spanning from N-terminal to 26Gly is outside the membrane. The residues connecting above two helical segments will likely bend along the surface of the membrane and take a more flexible conformation, comparing with the residues of the α helical conformation.

Assignments of the central regions of LH1 in the membrane have a particular functional significance. In LH1, Bchl molecules are associated with each α or β subunit through direct coordination of Mg²⁺ (in the pigment) and the imidazole sidechain of histidine residues on the peptide, including αHis32 and βHis38 of the LH1 sequence. The assignment of two histidines (αHis32 and βHis38) in Fig. 6 will eventually help to clarify the interaction between protein and pigments. Thus these studies open the door for future functional studies of the system.

Acknowledgements

We thank Dr. Craig Schenck for providing the chromatophores, and Dr. Mark Girvin, Paul Sorgen, Ray Krueger-Koplin, and Sue Krueger-Koplin (Albert Einstein College of Medicine) for providing the purification protocol and for many useful discussions. We also thank Dr. Tatyana Igunenova for her help in preparing the sample and pulse sequences. This work was financially supported by the National Science Foundation (NSF0316248).


