Journal of Biomolecular NMR (2005) 31: 279–293
DOI 10.1007/s10858-005-1604-8

Article

Residual backbone and side-chain ^{13}C and ^{15}N resonance assignments of the intrinsic transmembrane light-harvesting 2 protein complex by solid-state Magic Angle Spinning NMR spectroscopy

A. J. van Gammeren, F. B. Hulsbergen, J. G. Hollander & H. J. M. de Groot*
Leiden Institute of Chemistry, Gorlaeus Laboratoria, Leiden University, P.O. Box 9502, 2300, RA Leiden, The Netherlands

Received 15 October 2004; Accepted 27 January 2005

Key words: biosynthesis, pattern isotope labeling, transmembrane protein complex, sequential assignment, solid-state magic angle spinning NMR

Abstract

This study reports the sequence specific chemical shifts assignments for 76 residues of the 94 residues containing monomeric unit of the photosynthetic light-harvesting 2 transmembrane protein complex from *Rhodopseudomonas acidophila* strain 10050, using Magic Angle Spinning (MAS) NMR in combination with extensive and selective biosynthetic isotope labeling methods. The sequence specific chemical shifts assignment is an essential step for structure determination by MAS NMR. Assignments have been performed on the basis of 2-dimensional proton-driven spin diffusion ^{13}C–^{13}C correlation experiments with mixing times of 20 and 500 ms and band selective ^{13}C–^{15}N correlation spectroscopy on a series of site-specific biosynthetically labeled samples. The decreased line width and the reduced number of correlation signals of the selectively labeled samples with respect to the uniformly labeled samples enable to resolve the narrowly distributed correlation signals of the backbone carbons and nitrogens involved in the long α-helical transmembrane segments. Inter-space correlations between nearby residues and between residues and the labeled BChl *a* cofactors, provided by the ^{13}C–^{13}C correlation experiments using a 500 ms spin diffusion period, are used to arrive at sequence specific chemical shift assignments for many residues in the protein complex. In this way it is demonstrated that MAS NMR methods combined with site-specific biosynthetic isotope labeling can be used for sequence specific assignment of the NMR response of transmembrane proteins.

Abbreviations: LH2 – light-harvesting 2 protein; PDSD – proton driven spin diffusion.

Introduction

Insoluble and non-crystalline proteins, like most transmembrane proteins and protein aggregates, play a central role in intercellular communication and mediation of biological processes. This important class of proteins is involved in many biologically important functions, which can be understood thoroughly only when their detailed structures and interactions with the microscopic environment are known. Currently, only a few examples of transmembrane proteins with known crystal structure exist, which are mainly bacterial and plant photoreceptors (Koepke et al., 1996; Prince et al., 1997; Kühlbrandt 2001). Another example is rhodopsin, the G-protein coupled visual membrane photoreceptor, present in vertebrates,

*To whom correspondence should be addressed. E-mail: ssnmr@chem.leidenuniv.nl

including the human eye (Palczewski et al., 2000). Generally, transmembrane proteins are not easily prepared in a crystalline or water-soluble form, because the lipids associated with these proteins impede crystallization needed for X-ray crystallography and quench the rapid reorientation in solution that is a prerequisite for solution NMR spectroscopy. In contrast, for solid-state NMR spectroscopy local order and a homogeneous environment in the protein sample are sufficient. Hence, solid-state NMR has the potential of becoming a leading technology for structural investigations of transmembrane proteins.

In the last decade, Magic Angle Spinning (MAS) NMR has developed rapidly to resolve structures of microcrystalline peptide and protein samples at atomic resolution (Straus et al., 1998; Hong 1999; Detken et al., 2001; Pauli et al., 2001; Castellani et al., 2002; Böckmann et al., 2003). These protein samples have β-sheet motifs containing polar residues. This leads to a favorable chemical shift dispersion and good spectral resolution, facilitating the chemical shift assignment of these preparations. In contrast, transmembrane proteins generally consist of α-helical segments containing many residues with aliphatic side chains, which are constrained all in virtually the same secondary structure, yielding a very narrow chemical shift dispersion. This complicates the sequence specific assignment of the many narrowly distributed and overlapping correlations.

On the basis of data analysis in many genome projects, it has been suggested that the relatively small membrane spanning proteins, which span the membrane 1–2 times, are most abundant of all transmembrane proteins in organisms (Wallin and von Heijne, 1998; Krogh et al., 2001; Simon et al., 2001). These relative small transmembrane proteins contain ~30–70 residues. In this contribution, the 94 residues containing monomeric unit of the light-harvesting 2 (LH2) transmembrane protein complex from the anaerobic *Rhodopseudomonas* (*Rps.*) *acidophila* strain 10050 purple bacterium has been chosen as a model for the study of transmembrane proteins by MAS NMR. The monomer is a complex of two α-helical transmembrane segments, which are the α-subunit and β-subunit and the primary sequence is shown in Figure 1. The LH2 complex is a good model for solid-state NMR technology development, since the high abundance of this protein in a photosynthetic bacterium makes it convenient to introduce ^{13}C and ^{15}N isotope labels and the X-ray structure is known.(McDermott et al., 1995; Prince et al., 1997). In addition, the LH2 system is a protein complex that contains cofactors. Nine of these monomeric complexes form a concentric ring with 9-fold symmetry with a molecular weight of ~130 kDa. The rings form a mesoscopic assembly of many complexes, through which the light harvesting is realized. With NMR primary, secondary, tertiary and quaternary structure can be probed, as well as protein-cofactor arrangement and interactions.

At an early stage, the LH2 transmembrane protein complex has been used to explore the range and resolution of the MAS NMR in the study of transmembrane proteins (Egorova-Zachernyuk et al., 2001). In the present study, biosynthetic ^{13}C isotope pattern labeling methods are used to overcome the spectral crowding, leading to the first sequence specific assignment of a real transmembrane protein by solid-state NMR. This complements previous studies on small peptides reconstituted in lipid bilayers.(Opella et al., 2002; Marassi and Opella, 2003; Fujiwara et al., 2004; Opella and Marassi, 2004). A set of LH2 protein samples has been prepared, consisting of selectively labeled LH2 complexes obtained by a controlled growth of the bacteria in the presence of [1,2,3,4–^{13}C]-succinic acid, [1,4–^{13}C]-succinic acid, [2,3–^{13}C]-succinic acid or a mixture of uniformly labeled amino acids (van Gammeren et al., 2004). Referring to the isotopically labeled nutrient source in the expression medium these preparations are denoted as the U-LH2, 1,2,3,4-LH2, 1,4-LH2,

α: MNQGK^{5}IWTVV^{10}NPAIG^{15}IPALL^{20}GSVTV^{25}IAILV^{30}HLAIL^{35}SHTTW^{40}FPAYW^{45}QGGVK^{50}KAA

β: ATLTA^{5}EQSEE^{10}LHKYV^{15}IDGTR^{20}VFLGL^{25}ALVAH^{30}FLAFS^{35}ATPWL^{40}H

Figure 1. The primary structure of the α- and β-subunits of the LH2 complex from *Rps. acidophila* strain 10050. Residues in the α-helical segments are underlined. The N-carboxyl-αM1, αH31 and αH30 residues are coordinated to the central Mg-ions of the BChl *a* cofactors.

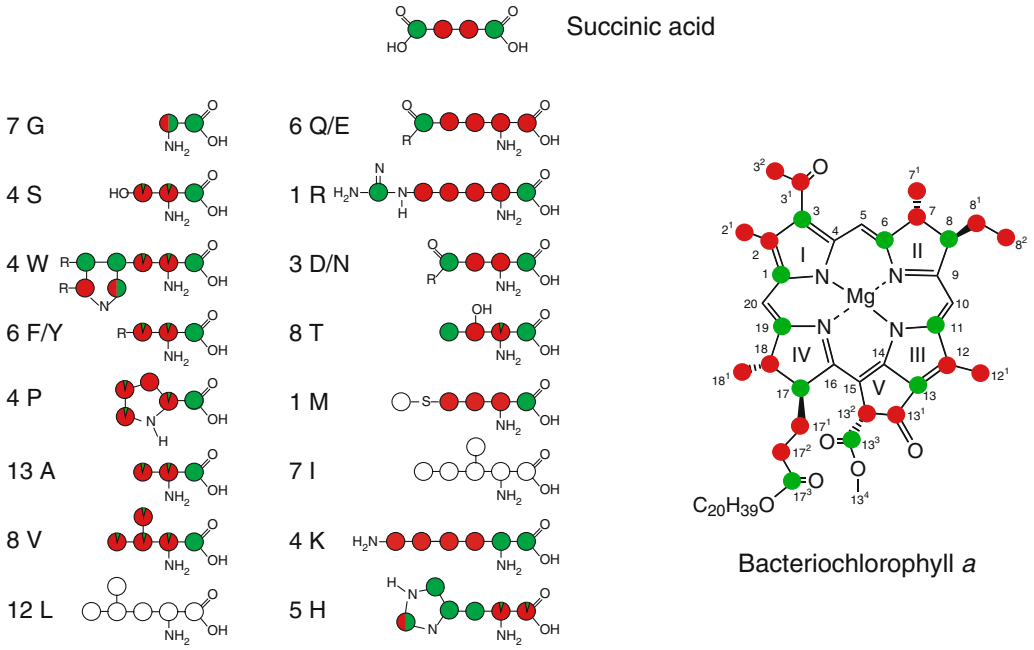


Figure 2. Labeling patterns of amino acids and Bacteriochlorophyll cofactors in the LH2 transmembrane protein complex obtained from *Rps. acidophila* strain 10050 expressed in a medium containing [2,3−^{13}C]-succinic acid (red) or [1,4−^{13}C]-succinic acid (green). I and L are isotope labeled by the amino acid mixture nutrient source.

2,3-LH2 and AA-LH2 sample, respectively. The labeling patterns of all amino acid residues and for the Bacteriochlorophyll *a* (BChl *a*) cofactors in 1,2,3,4-LH2, 1,4-LH2, 2,3-LH2 are shown in Figure 2 (van Gammeren et al., 2004). The biological growth conditions to prepare the AA-LH2 sample are different from the growth conditions for samples prepared from succinic acid. For succinic acid labeled samples, an excess of succinic acid was used to enhance the uptake of this nutrient source, while for the AA-LH2 sample preparation an excess of labeled amino acid mixture was used to enhance the uptake from the amino acid nutrient source. The scrambling pattern in the succinic acid labeled LH2 samples is different from the scrambling in the AA-LH2 sample. For the AA-LH2 sample, I and L are uniformly labeled, while also minor fractions of the A, G, V and P residues are labeled. By using both the samples labeled by succinic acid and the AA-LH2 sample, it is possible to arrive at a sequence specific assignment for 76 residues of the 94 in the LH2 complex, which is the largest membrane protein complex to date used for assignment studies with MAS NMR.

Materials and methods

For the NMR experiments, about 10 mg of U-LH2, 1,2,3,4-LH2, 1,4-LH2, 2,3-LH2 or AA-LH2 protein sample was transferred into a 4.0-mm CRAMPS rotor. On these samples, 2D homonuclear ^{13}C–^{13}C correlation spectra were recorded using PDSD MAS NMR spectroscopy on a Bruker AV750 spectrometer equipped with a double channel CP-MAS probe head and using a ^{13}C radio frequency of 188 MHz. The proton $\pi/2$ pulse was set to 3.1 μs, corresponding with a nutation frequency of 80.6 kHz. ^{13}C B_1 field strengths of 50 kHz corresponding with a cross polarization time of 2.0 ms were applied during a 100 to 50% ramped CP sequence (Metz et al., 1994). In the PDSD experiment, two-pulse phase modulation (TPPM) decoupling was applied during the t_1 and t_2 periods (Bennett et al., 1998). A mixing time of 50 ms ($PDSD_{50}$) was used to transfer the magnetization into the side chains, providing intra-residue correlations. Proton driven spin diffusion experiments with a long mixing time of 500 ms ($PDSD_{500}$) have been applied to collect long-distance ^{13}C–^{13}C inter-residue correlations

(Castellani et al., 2003). 2D heteronuclear ^{13}C–^{15}N correlation spectra were recorded with the same spectrometer using a triple resonance CP-MAS probe head and band selective NC magnetization transfer. ^{15}N polarization was created with a 80–100% ramped amplitude CP matching and a contact time of 2.0 ms. During ^{15}N evolution TPPM decoupling with a r.f. field strength of 81 kHz was used (Metz et al., 1994; Bennett et al., 1998). To create a band-selective SPECIFIC-CP transfer from the nitrogen nuclei to either the CO or Cα, the carrier frequency was placed at 175 or 50 ppm, respectively (Baldus et al., 1998). During the dipolar contact time of 3.0 ms for the NCO transfer and 3.5 ms for the NCA transfer, a weak r.f. field of 22.5 kHz for ^{15}N was employed. The r.f. field strengths for ^{13}C were 14 and 31 kHz for the NCA and NCO band selective CP, respectively. Band selectivity was achieved using adiabatic amplitude modulations on the ^{15}N channel (Hediger et al., 1995; Baldus et al., 1996). During the NC transfer off-resonance continuous wave decoupling was applied at the ^{1}H frequency.

For experiments involving homonuclear spin diffusion transfer, to obtain multiple correlations of multiple carbons with one nitrogen in a macromolecular network, a spin diffusion transfer period was included prior to the acquisition in t_2. The $\pi/2$ pulses before and after the spin diffusion period were applied with a r.f. field of 45 kHz on the ^{13}C-channel. The spin diffusion period was 20 ms for the NCA(CO)CX and 30 ms for the NCOCACX transfer, where CX stands for any carbon atom. In both experiments, TPPM decoupling was applied during the t_1 and t_2 periods (Bennett et al., 1998). All samples were cooled to 253 K, and the MAS spinning frequency $\omega_R/2\pi$ was 8.5 kHz (± 2 Hz). The ^{13}CO resonance of U-[^{13}C,^{15}N]-Tyrosine·HCl at 172.1 ppm was used as an external reference for the calibration of the isotropic ^{13}C chemical shifts.

The MAS NMR data were processed with XWINNMR software (Bruker) and subsequently analyzed using the program SPARKY (T. D. Goddard and D.G. Kneller, University of San Francisco)

Results and discussion

By applying PDSD and band selective SPECIFIC CP NMR methods on the 2,3-LH2, 1,2,3,4-LH2 and AA-LH2 preparations, an unambiguous sequence specific chemical shift assignment for backbone carbons and backbone nitrogens for the majority of the residues can be obtained. While in the past inter-residue correlations have been used to resolve 3D structure, here they are used predominantly for a sequence specific assignment of the LH2 transmembrane protein complex (Castellani et al., 2002; Castellani et al., 2003). The chemical shift assignments for residues from the α- and β-subunit are summarized in Tables 1 and Table 2, respectively. In the first step of the assignment procedure, the $PDSD_{50}$ spectrum of the U-LH2 in Figure 3 was analyzed to identify the characteristic ^{13}C–^{13}C correlation patterns of side chains of individual residues. For a major fraction of the residues characteristic chemical shift patterns could be observed. For 4P, 8T and 4W residues, which occur both inside and outside the α-helix part, the number of spin systems in the spectra corresponds with the number of residues in the protein sequence. For the 13A, 12L and 9V residues, abundantly present in the aliphatic α-helical segments of the transmembrane LH2 protein, the ^{13}C–^{13}C correlations strongly overlap. For these residues it is difficult to resolve the spin systems corresponding with each residue. For instance, in the $PDSD_{50}$ spectrum three H spin systems and one K spin system have been observed, while H and K residues occur five and four times in the protein sequence, respectively. Hence two H spin systems and three K are either not detected or not resolved. Two of the five H residues and three of the four K residues are located in the flexible loops and termini of the protein outside the rigid α-helical part (Figure 1). Since βK13 is embedded in the rigid part of the α-helical segment of the β-subunit, the spin system of the K residue has been assigned to βK13. αH37 and βH41 are in the flexible part of the protein and most likely their correlation signals are quenched by dynamics. The three observable H spin systems are attributed to αH31, βH12 and βH30. In two of these H spin systems the Cα and Cβ responses are shifted upfield. They are assigned to αH31 and βH30, since these residues coordinate to the central Mg ion in the conjugated macrocycles of the BChl cofactors, which induce an upfield ring current shift (van Gammeren et al., 2004; Alia et al., 2001). The third set has been assigned to βH12, which is in the rigid α-helical segment of the β-subunit.

Table 1. Solid-state ^{15}N and ^{13}C chemical shift assignments for the α-subunit of the LH2 TM protein complex

Residue	^{15}N	^{13}CO	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	$^{13}C^{\gamma}$	$^{13}C^{\delta}$	$^{13}C^{\varepsilon}$	$^{15}N^{\varepsilon}$	$^{15}N^{\pi}$
αM1	87.6	169.8	55.3	31.2	27.8				
αN2	123.5	172.4	46.8	34.9					
[αQ3]	112.8		53.7						
[αG4]	104.7	170.9	42.6						
αK5	113.6								
[αI6][a]		174.0	57.6	36.5	14.1/25.3	9.0			
αW7	[121.0]	172.0	51.4	24.1	109.1	125.5	137.7	135.6	
αT8	110.2	174.4	56.7	64.1	17.8				
αV9									
αV10	120.5	170.3	57.2	29.2	17.7/19.0				
αN11	125.0	173.6	48.0	37.6	176.3				
αP12	144.8	172.9	61.5	28.6	22.9	47.5			
αA13	115.3	177.1	51.5	14.7					
[αI14][a]	119.7	174.6	61.1	33.6	14.5/22.6	12.0			
αG15	105.0	170.1	43.9						
αI16	117.3	171.4	62.9	30.2	24.0/12.7	8.8			
αP17	131.7	173.9	62.5	28.2	25.4	45.5			
[αA18][b]			51.3	14.9					
[αL19][b]				37.2	22.0	20.0/18.5			
αL20	120.2	174.5	49.6	41.5	23.6	23.6/18.6			
αG21	108.8	172.1	44.3						
αS22	118.8	171.5	59.8	60.7					
αV23	120.6	173.3	63.9	29.1	18.3/19.1				
αT24	115.0		64.1	65.2	16.7				
αV25	122.9	175.3	62.7		17.6/19.1				
αI26	124.0	170.4	58.4	33.4	25.2/12.3	10.45			
αA27	120.8	177.2	51.5	15.3					
αI28	119.3	172.4	63.3	34.3	26.3/13.0	9.0			
αL29	120.5		54.3	37.5	22.6	17.7/19.1			
αV30	118.3	173.5	63.2	26.9	15.1/17.4				
αH31	112.4	173.0	58.5	23.4	125.1	118.2	132.9		168.1
[αL32][b]		173.7	53.8	38.0	23.2	16.5/21.6			
[αA33][b]	120.4	176.4	51.6	13.2					
αI34	121.4	172.9	62.4	33.7	25.3/13.3	10.6			
αL35	119.0		54.4	37.9	22.7	17.8/20.6			
αS36	114.9	173.1	58.3	59.8					
αH37	not observed								
αT38	105.0	170.6	56.1	70.6	19.0				
αT39	107.2	169.0	56.0	64.1	17.1				
αW40	[118.6]	173.7	53.5	25.9	108.9	119.7	134.9	125.2	
αF41	122.4	169.9	54.9	30.8	134.7				
αP42	135.3	174.5	63.0	28.0	25.0	46.1			
αA43	120.8	178.8	50.4	14.9					
αY44	124.1	177.2	51.6	44.2					
[αW45][a]	118.6	173.6	54.0	25.8	105.3	126.1	136.3	133.7	
αQ46–αA53	Not observed								

These chemical shifts have been deposited with the BioMagResBank (BMRB). Tentative assignments are given in brackets.
[a]αW45/βW39 and αI6/αI14 datasets can be interchanged.
[b]αA18/αL19, αL32/αA33 and βL25/βA26/βL2 spin systems are assigned from the $PDSD_{500}$ of the AALH2 sample. Because of the frequent AL or LA combination in the sequence, a sequence specific assignment is difficult.

Table 2. Solid-state ^{15}N and ^{13}C chemical shift assignments for the β-subunit of the LH2 TM protein complex

Residue	^{15}N	^{13}CO	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	$^{13}C^{\gamma}$	$^{13}C^{\delta}$	$^{13}C^{\varepsilon}$	$^{15}N^{\varepsilon}$	$^{15}N^{\pi}$
[βA1]	77.1	171.2	47.5	16.3					
βT2	108.9	170.3	58.8	66.0	17.1				
βL3	115.9	177.1	54.4	38.3	24.1	18.4/22.1			
βT4	106.1	170.6	58.7	66.8	16.8				
βA5	118.0		51.5	14.3					
βE6	118.1	176.1	54.1	25.9	30.5				
βQ7	120.9	174.0	48.2	21.8	26.8				
βS8	110.2	173.6	57.3	59.6					
βE9									
βE10	117.2	178.3	55.1	25.2	30.8				
βL11									
βH12	119.4	173.8	56.0	27.6	132.9	113.5	133.7		166.4
[βK13]	114.5	170.7	50.4	31.0	23.4	27.8	48.4		
βY14									
βV15	114.2		62.1	27.7	17.1/19.0				
βI16	123.6	174.8	57.6	31.1	22.1/12.9	2.7			
βD17	115.6	173.6	50.0	37.6					
βG18	105.5	170.8	44.2						
βT19	106.4	176.0	58.0	65.8	17.5				
βR20[c]		171.4	56.9	26.8	25.5	41.1		81.2	
βV21									
βF22	120.6		54.4	36.2					
βL23		174.5	53.9	37.4	23.9	18.8/22.9			
βG24	106.3	171.9	44.0						
[βL25][b]	119.2		54.0	37.4	23.6	19.5/22.1			
[βA26][b]	120.8		50.9	13.9					
[βL27][b]			54.1	37.5					
βV28	120.2	173.1	63.7	27.4	17.6/20.5				
βA29	120.3	176.1	51.4	14.4					
βH30	113.1	173.2	57.8	23.4	125.2	118.8	133.5		169.0
βF31	120.7	173.5	59.6	36.0	136.2				
[βL32]	117.4		53.7	38.3	22.6	19.1/20.4			
[βA33]	120.51		51.5	15.5					
βF34	121.5	175.8	58.3	34.9	134.7				
βS35	105.0		55.9	59.7					
βA36	120.5	173.8	50.5	16.8					
βT37	108.6	170.5	57.3	67.4	17.3				
βP38	126.5		58.9	28.4	25.2	46.6			
[βW39][a]	120.53		55.0	25.0	108.5	123.0	135.4	128.2	
βL40–βH41	Not observed								

These chemical shifts have been deposited with the BioMagResBank (BMRB).
Tentative assignments are given in brackets.
[a]αW45/βW39 and αI6/αI14 data sets can be interchanged.
[b]αA18/αL19, αL32/αA33 and βL25/βA26/βL2 spin systems are assigned from the $PDSD_{500}$ of the AALH2 sample. Because of the frequent AL or LA combination in the sequence, a sequence specific assignment is difficult.
[c] The chemical shifts for C^{z}, NH^{1}, NH^{2} of βR20 are 156.2, 76.2 and 70.1 ppm, respectively.

By using the 1,2,3,4-LH2, 2,3-LH2 and AA-LH2 samples, the overlap in the $PDSD_{50}$ data of the U-LH2 is reduced. According to Figure 2, the majority of the Cα and Cβ carbons of the residues synthesized from succinic acid are ^{13}C labeled by using [2,3−^{13}C]-succinic acid. The aliphatic carbons

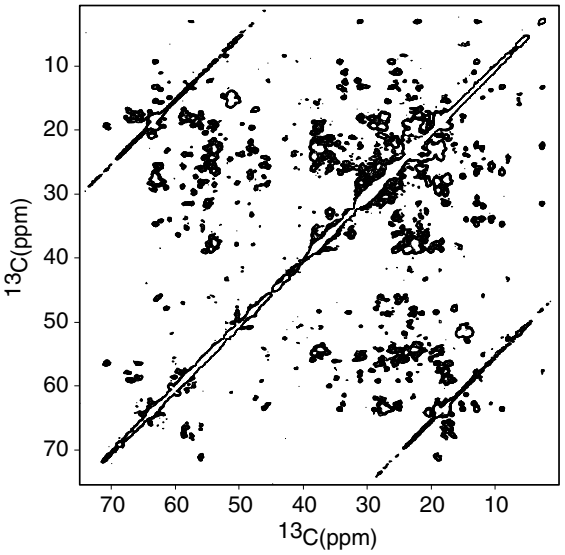


Figure 3. The aliphatic part of the $PDSD_{50}$ ^{13}C–^{13}C correlation spectrum of the U-LH2.

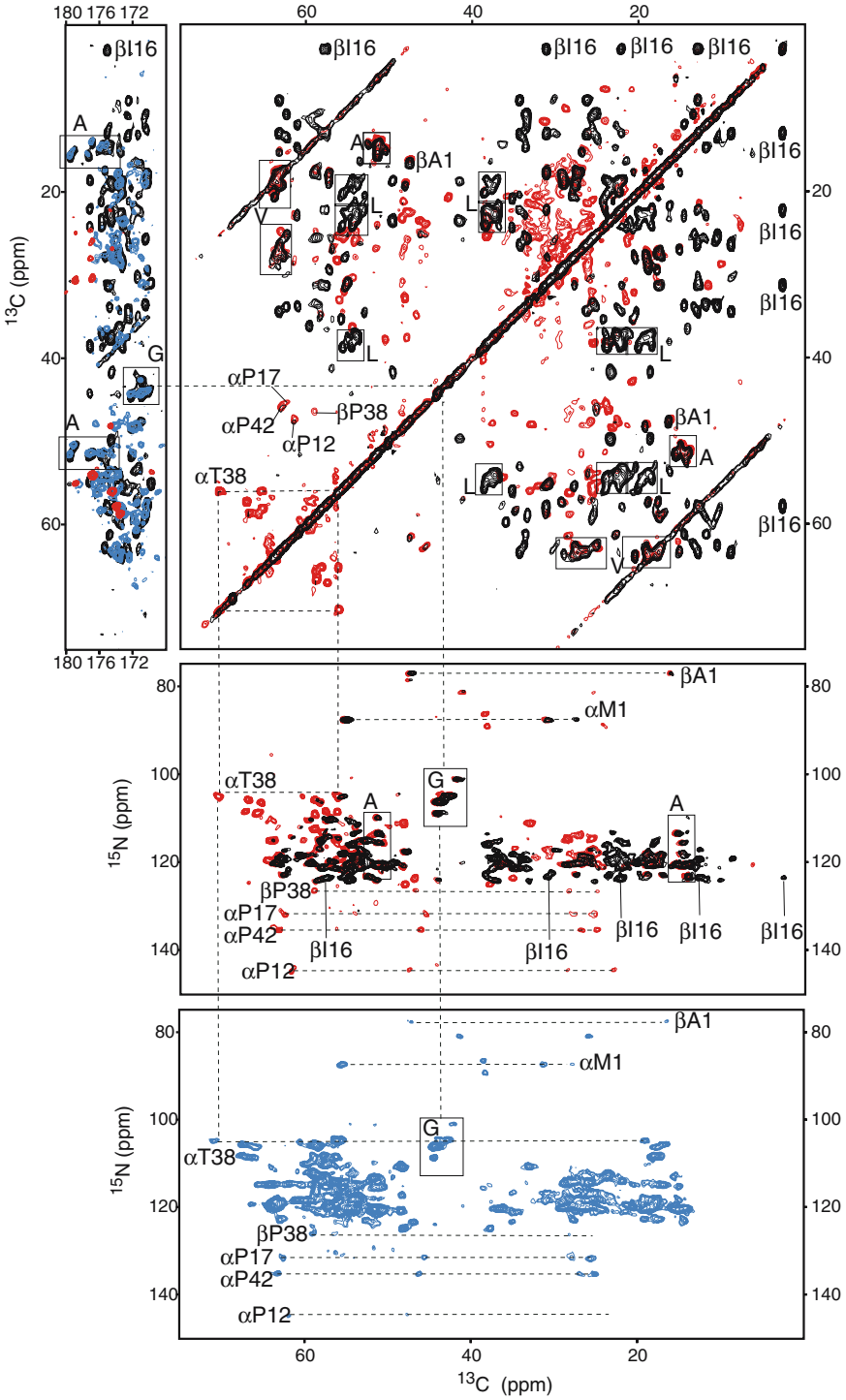
of the H residues are not labeled in the 2,3-LH2 sample and the Cα-Cβ signals of this residue are assigned from the $PDSD_{50}$ dataset collected from 1,2,3,4-LH2. For the other residues, labeled by succinic acid, the $PDSD_{50}$ spectrum of 2,3-LH2 has been used to assign Cα-Cβ correlations. For detection of the I and L residues, the AA-LH2 was used. Due to the uptake of labels from the amino acid mixture, a minor fraction of the A, G, V and P residues is also labeled in this sample, and the AA-sample has been used to improve the assignments for these residues.

In the upper right panel in Figure 4, both the red coded $PDSD_{50}$ spectrum of 2,3-LH2 and the black coded $PDSD_{50}$ spectrum of AA-LH2 are shown. Compared to the $PDSD_{50}$ spectrum of the U-LH2 in Figure 3, the number of correlations is reduced, while the resolution in the spectra is improved. For backbone carbons, there is a significant reduction of J-couplings, which contributes to the resolution improvement. For most side-chain carbons the reduction of the scalar coupling is only small, which provides only a marginal improvement of the resolution in the spectra. The data from the pattern labeled samples shows also a reduced number of correlations in the aliphatic region between 0 and 75 ppm, since the I and L residues, which represent ~20% of the total protein, are not enriched when growing from the labeled succinic acid.

By aligning the $PDSD_{50}$ spectrum of 2,3-LH2 with the ^{15}N–^{13}C NCACX and NCACOCX correlation spectra of various pattern labeled samples, all Cα–Cβ correlations in the spectrum can be assigned to specific residues in the protein sequence. The eight T and four S residues are rapidly identified due to the relative downfield shifts of both T/S(Cα) and T/S(Cβ) chemical shifts. To illustrate the assignment procedure, correlation responses of the αT38 spin system are indicated in Figure 4. The Cα–Cβ correlations of T and S residues are well resolved between 56 and 71 ppm, close to the diagonal. The T(Cα–Cγ) and T(Cβ–Cγ) correlations, observed in the spectrum of the U-LH2 are quenched in the spectrum of 2,3-LH2, which shows that T(Cγ) is not labeled. The $PDSD_{50}$ spectra also show unique P(Cα–Cδ) correlations, which is important for the assignment of the P residues, shown in Figure 4.

As an example of the responses of the I residues from the AA-LH2 sample, the signals from βI16, located at the upfield limit of the aliphatic spectrum are indicated. In this way the characteristic multiple-bond correlation signal pattern for this type of residue is illustrated. Corresponding carbon correlation networks are observed for αI6, αI14, αI16, αI26, αI28 and αI34. In contrast, the correlations of the L residues strongly overlap and are difficult to resolve. The A and V residues are observed in both the 2,3-LH2 and AA-LH2 dataset (Figure 4). In the U-LH2 dataset V(Cα–Cβ) correlations overlap with I (Cα–Cγ1) and can be resolved by using the spectra from the pattern labeled samples.

The CO–Cα–Cβ correlation area for the 2,3-LH2 shows correlation signals for E, Q and H residues, which are the only residues that have both a ^{13}CO and a $^{13}C\alpha$ backbone carbon. To observe the CO–Cα–Cβ correlations for other residues synthesized from succinic acid, the blue coded spectrum of the 1,2,3,4-LH2 in the upper left panel of Figure 4 can be used, since in the 1,2,3,4-LH2 both the CO and the Cα are labeled. For the assignment of the chemical shifts for the I, L and V residues in this spectrum, in particular for the crowded CO–Cα–Cβ correlation area, the simplification due to the pattern labeling is essential. Assignments for G residues, which have no aliphatic side chains, are obtained starting from the G(Cα) response at ~40–43 ppm in the CO–Cα correlation region, where five out of seven G



residues are identified. Two residues, αG47 and αG48, are in the mobile part near the C-terminus of the α-subunit and are broadened beyond the limit of detection.

The ^{13}C chemical shift correlation patterns in the $PDSD_{50}$ spectra can be aligned with the ^{15}N–^{13}Cα–^{13}Cβ–^{13}Cγ–(13Cδ) (NCACX) correlation spectra to resolve the correlations with the

Figure 4. In the upper panels two regions from homonuclear ^{13}C–^{13}C PDSD50 correlation spectra collected from 2,3-LH2 (red) and AA-LH2 (black) are shown. The region shown in the upper left panel contains cross peaks involving the aliphatic carbons and carbonyl carbons, while the upper right panel shows correlations between aliphatic carbons, present in the side chains of the amino acids. In the left part, a few responses are observed for 2,3-LH2, belonging to H, Q and E residues. The responses from AA-LH2 in the carbonyl area are from I, L, A, G and V residues. The blue coded spectrum in the carbonyl region comprises carbonyl responses from 1,2,3,4-LH2. In the upper right panel the aliphatic responses are shown. The dashed lines indicate correlations involving the αT38 and four P residues for the 2,3-LH2, and correlations involving βI16 for the AA-LH2. The A residues that are labeled via both nutrient sources are also indicated. In the middle pane, the aliphatic region of the NCACX spectra of 2,3-LH2 (red) and AA-LH2 (black) are shown. The data are aligned with the PDSD50 spectrum and correlations involving αT38, βI16 and the 4P residues are indicated with dashed lines for the two different samples. The responses of the G residues are indicated with a rectangular box. The NCA signals are aligned with the carbonyl area of the PDSD50 spectrum. Finally, in the lower panel the NCACX spectrum of the 1,2,3,4-LH2 sample is shown.

^{15}N backbone signals. Previously, it has been shown that the ^{15}N–^{13}C correlation spectra of U-LH2 are too crowded to extract the individual chemical shifts (Egorova-Zachernyuk et al., 2001). The middle panel of Figure 4 presents the NCACX spectra of both 2,3-LH2 and AA-LH2, while the NCACX spectrum of 1,2,3,4-LH2 is presented in the lower panel of Figure 4. It is clear that the use of two pattern-labeled samples separates many responses that overlap in the data collected from the U-LH2 (Figure 3). Starting with the red coded NCACX spectrum of 2,3-LH2, intra-residual ^{15}N–$^{13}C\alpha$–$^{13}C\beta$–(Cγ) correlation networks are resolved and can be assigned by alignment with the correlation sets collected from the PDSD spectrum of this sample. For instance, αT38(Cγ) clearly shows up in the $PDSD_{50}$ spectrum of 1,2,3,4-LH2, but not in the $PDSD_{50}$ spectrum of 2,3-LH2, because the Cγ is unlabeled (Figure 2). The ^{15}N chemical shifts of the αP12, αP17, αP42 and βP38, i.e. all 4P residues of the LH2 complex, resonate downfield with respect to the ^{15}N response of other residues. ^{15}N responses for G residues are clearly resolved in the NCACX data from their N–Cα correlations that are relatively upfield on both the ^{13}C and ^{15}N chemical shift scales, and align well with the G(CO–Cα) responses in the $PDSD_{50}$ spectrum.

The NCACX spectrum of AA-LH2 in the central panel of Figure 4 shows ^{15}N–^{13}C correlations from labeled I and L residues. In addition, V, A and G residues are observed due to the partial incorporation of labels from the amino acid mixture nutrient source. The correlations involving these aliphatic residues in the α-helical part predominantly show up with ^{15}N chemical shifts of ~120 ppm. The ^{15}N shifts of the I residues have been assigned on the basis of the ^{15}N–Cδ correlations between 2 and 18 ppm ^{13}C shift. For example, the βI16 response in the NCACX spectrum is well in line with the corresponding correlation pattern in the $PDSD_{50}$ spectrum shown in the upper panel of Figure 4. The upfield shifted Cδ response of βI16 at 2.5 ppm is due to the short distance of 3.7 Å between this nucleus and the aromatic macrocycle of the B800 BChl cofactor, which produces a significant ring current shift (Papiz et al., 2003). For αM1 at the N-terminus of the α-subunit, a correlation set is found at 87.6 ppm ^{15}N shift. The carboxyl-αM1-N-terminus is coordinated to the central Mg-ion of the B800 cofactor, which contributes to the rigidity of the N-terminal loop of the α-subunit. This accounts for the observation of many residues outside the α-helical segments in our data.

The majority of the ^{15}N–^{13}C correlations of the V residues coincide with ^{15}N–^{13}C signals from I and L residues. In the NCACX data of the 2,3-LH2 sample the signals from the V residues are clearly resolved since the responses of the I and L residues are not present. A residues can be identified in the $PDSD_{50}$ spectra of both the AA-LH2 and the 2,3-LH2. The Cα–Cβ correlation signals of the A residues strongly overlap in the $PDSD_{50}$ spectra in the upper panel of Figure 4. A(Cα) shifts are observed between 47 and 53 ppm, while A(Cβ) signals are observed between 14 and 16 ppm. To distinguish between the various A residues the CO–Cα–Cβ correlation area is used, since the dispersion of the CO responses is largest. In the NCACX spectra, A residues can be distinguished conveniently because the A(Cα) signals are shifted upfield, while the ^{15}N backbone signals are between 117 and 124 ppm. One exception is the ^{15}N response at 77.2 ppm, which correlates with Cα and Cβ responses at 47.7 and 16.6 ppm, respectively. These correlations have been assigned tentatively to the βA1, which is at the N-terminus of the β-subunit.

For a sequence specific assignment of the spin systems inter-residual magnetization transfer from the backbone N_i to the $C_{i-1}O$ carbon was produced using the band selective SPECIFIC CP method (Baldus et al., 1998). The N_i–CO_{i-1} cross polarization, where the magnetization is transferred into the side chain of a previous residue, yields $N_iCO_{i-1}CA_{i-1}CX_{i-1}$ (NCOCACX) correlation spectra that can be aligned to both the $PDSD_{50}$ and the NCACX spectra to identify the correlating residues. Figure 5 shows the NCOCACX correlation spectra of 2,3-LH2, AA-LH2 and 1,2,3,4-LH2 in red, black and blue, respectively. For observation of the residues in the aliphatic part of the NCOCACX spectra, both the CO_i and the $C\alpha_i$ positions should be ^{13}C labeled and covalently connected with $^{15}N_{i+1}$. Such labeling patterns occur for H, Q and E residues in the 2,3-LH2. The correlations of the 2,3-LH2 in the NCOCACX spectrum can be aligned with the red coded correlations in the CO region of the PDSD spectrum in the upper left panel of Figure 4. In the NCOCACX spectrum of 2,3-LH2 in Figure 5a, 2 out of 5 H residues are observed, while in the PDSD spectrum 3 out of 5 H-residues are detected. The protein sequence shows that αH31 is followed by αL32, which is unlabeled in this sample. The 2 inter-residual $H(C\alpha_i$–$N_{i+1})$ correlations in the 2,3-LH2 NCOCACX spectrum can now be assigned unambiguously to βH30/βF31 and βH12/βK13, while the third H correlation set, which is only present in the $PDSD_{50}$ spectrum, can be assigned to αH30. This assignment is well in line with data collected from LH2 samples where only H residues are labeled (Alia et al., 2001). Two spin systems for the Q and E responses in this spectrum are assigned to βQ7 and βE10. βQ7 can be assigned due to the correlation with the nitrogen of βS8. Since βE10 is followed by the unlabeled βL11 in the 2,3LH2, a correlation with the βE10 is not detected in the NCACOCX spectrum of this sample. This leads to the specific assignment of the βE10 responses indicated in the upper panel of Figure 5.

The black NCOCACX spectrum of the AA-LH2 in Figure 5a shows strong I_i–I_{i+1}, L_i–L_{i+1}, A_i–A_{i+1}, V_i–V_{i+1} and G_i–G_{i+1} correlations. Like in the NCACX spectrum, most ^{15}N–^{13}C correlations in the NCOCACX spectrum of AA-LH2 have ^{15}N

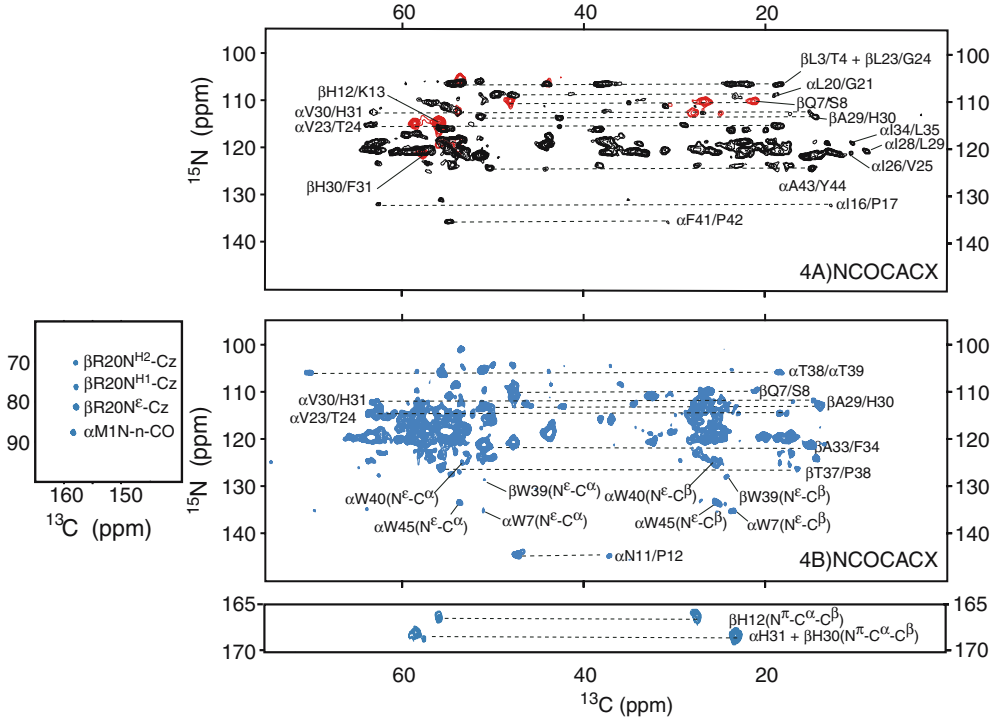


Figure 5. The aliphatic region of the NCOCACX spectra of 2,3-LH2 (red), AA-LH2 (black) and 1,2,3,4-LH2 (blue). Several sequence specific assignments are indicated in the figure.

chemical shifts close to 120 ppm that are difficult to resolve. Some of the correlation sets that are not resolved in the NCACX spectrum can be assigned using the upfield or downfield shifted ^{15}N response of the next residue. Examples are αL20/G21, αV23/T24, αV30/H31, αI16/P17, αF41/P42, βA29/H30, βL3/T4 and βL23/G24 (Figure 5a). Inter-residue $^{15}N_{i+1}$–ICδ_i correlations are observed between 2 and 18 ppm ^{13}C shift for αI24/L35, αI28/L29 and αI26/LV25. They can be assigned by aligning the spectrum to the NCACX spectrum of the AA-LH2 in the middle panel of Figure 4.

To observe inter-residual correlations of other residues synthesized by the succinic acid nutrient source, the NCOCACX spectrum of 1,2,3,4-LH2 is used. Residues synthesized from [1,2,3,4–^{13}C] succinic acid are labeled on both the Cα and the CO position. The NCOCACX spectrum of 1,2,3,4-LH2 in Figure 5b shows a strong reduction of correlation signals with respect to the NCACX spectrum in the lower panel of Figure 4c. Since the I and L residues are unlabeled in the 1,2,3,4-LH2, the I_i–$I_{i\pm1}$ and L_i–$L_{i\pm1}$ responses are not present in the NCOCACX spectrum. This reduces the spectral crowding by ~40% with respect to the NCOCACX spectrum of U-LH2. Hence ^{13}C correlations sets that are observed in the NCACX spectrum of the 1,2,3,4-LH2 and are not observed in the NCOCACX spectrum, can be assigned to residues that are followed by a I or L residue in the sequence. Labeled residues that are not followed by a L or I in the protein sequence give rise to correlations in the NCOCACX spectra of 1,2,3,4-LH2 and AA-LH2. Clear examples are the αV30/H31 and βA29/H30 correlations. These two examples also demonstrate how two nearby correlations of the αH31(Cα/Cβ) and βH30(Cα/Cβ) in the $PDSD_{50}$ spectra can be distinguished with help of the correlation to the adjacent residue, leading to an unambiguous assignment. The inter-residue correlations of αT38/T39, βT37/P38 and αN11/P12 are only present in the NCOCACX data collected from the 1,2,3,4-LH2, since both the Cα and the CO carbons of these residues are labeled in this sample.

The NCOCACX of the 1,2,3,4-LH2 sample also reveal intra-residue correlations between side chain nitrogens and carbons of the W, R and H residues. The ^{15}N chemical shifts from the side chain nitrogens are significantly different from the backbone ^{15}N shifts and can be conveniently used for the assignment of residues, facilitating also the sequence specific assignment. For instance, in Figure 5 the Nε–Cα–Cβ correlations for αW7, αW40, αW45 and βW39 are detected. In addition, a set of three intense correlations are observed for the Cz-carbon of βR20 at 156.2 ppm with Nε, NH1 and NH2 at 81.2, 76.2 and 70.1 ppm. Also three sets of Nπ–Cγ–Cδ–Cβ–Cα correlations at ^{15}N = 168.1, 166.4 and 169.0 ppm are observed for αH31, βH12 and βH30, respectively.

To collect more inter-residue correlations, $PDSD_{500}$ data were collected from the 2,3-LH2 and AA-LH2 samples. The reduction of ^{13}C labeling in these samples strongly suppresses relayed spin diffusion. This enables the magnetization transfer over longer distances than for uniformly labeled samples and provides the possibility to detect correlations between remote carbons from different residues. In contrast to the band selective magnetization transfer in the NCACOCX method, which yields selectively $i/i-1$ correlations, in the $PDSD_{500}$ method the magnetization is transferred to any nearby ^{13}C nuclei. The $PDSD_{500}$ spectra obtained from the 2,3-LH2 and AA-LH2 are shown in Figures 6 and 7, respectively. Using the intra-residue assignments of the $PDSD_{50}$ spectrum as a starting point, the long-range correlations in the $PDSD_{500}$ spectrum can be assigned to combinations of residues that are unique in the protein sequence. In this way, the long-range correlations contribute to a sequence specific assignment of the protein NMR response.

In the $PDSD_{500}$ spectrum of the 2,3-LH2 in Figure 6, a clear example of long-range inter-residue correlations is provided for αThr38. The αThr38 is in the middle of a short loop, which connects the two α-helical segments in the α-subunit. In addition to the αThr38(Cα–Cβ) correlation at 56.1–70.6 ppm, the Cβ signal of this residue at 70.6 ppm correlates with a unique set of signals, which can be assigned unambiguously to αP42(Cδ), αA43(Cα), αW40(Cα), αW40(Cβ), αS36(Cα) and αT39(Cα). The cross-peaks of the αH37 ^{13}C responses with αT38(Cβ) are attenuated in the $PDSD_{500}$ spectrum. This is attributed to the mobility of αH37, and is in line with the absence of its signal in the $PDSD_{50}$ spectra. A second example of long-range interactions is βF22. In the protein sequence, βF22 is located between labeled βV21 and unlabeled βL23. In line with the labeling pattern, only the βF22/V21 nearest neighbor

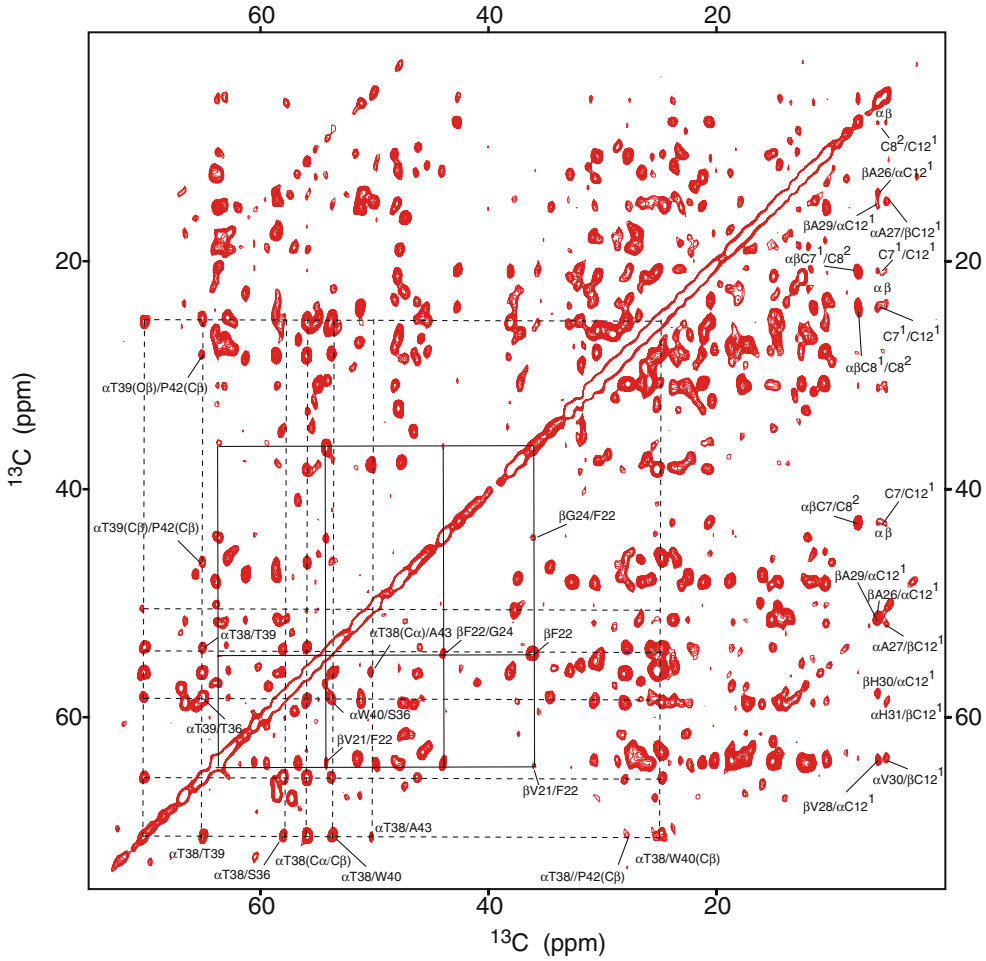


Figure 6. Contour plot of the PDSD500 spectrum of 2,3LH2. The figure shows inter-residue correlations that are of interest for the sequence specific assignment of the protein. The solid and dashed lines indicate two examples of correlation patterns, which are described in detail in the text.

signals are detected. In addition, βF22/βG24 correlations are weakly observed. Since [2,3–^{13}C]-succinic acid is an important precursor for the biosynthesis of BChls in photosynthetic bacteria, the BChls in the 2,3-LH2 are also pattern labeled by following the biosynthetic [2,3–^{13}C]-succinic acid or [1,4–^{13}C]-succinic acid incorporation pathway (Schulten et al., 2002). For 2,3-LH2 most of the peripheral groups of the BChls that interact with the protein matrix are labeled and the $PDSD_{500}$ spectrum of 2,3-LH2 in Figure 6 clearly shows αC12^1/βV28/βA29/βH30 and βC12^1/αA27/αV30/αH31 intermolecular correlations. Long-distance correlations within the BChl carbon framework are observed between αC12/C12^1/C8^1/C8^2/C7/C7^1 and αC12/C12^1/C8^1/C8^2/C7/C7^1. The correlations between the cofactor and the protein are well in line with the X-ray structure of the LH2. This shows that αC12^1 and βC12^1 are in the vicinity of the H residues that coordinate to the Mg ion of the βB850 cofactors, βH30 and αH31, respectively. The assignment of the αC12^1 and the βC12^1 responses from the cofactor–protein correlations confirm the selective assignments of the BChl cofactor signals in a previous study (van Gammeren et al., 2004). The αH31/βB850 and βH30/αB850 protein–cofactor interactions in the $PDSD_{500}$ spectra are interactions between a subunit and a cofactor of an adjacent subunit and constrain the quaternary structure of the complex. It is encouraging that cofactor–protein interactions are readily observed in the LH2 complex.

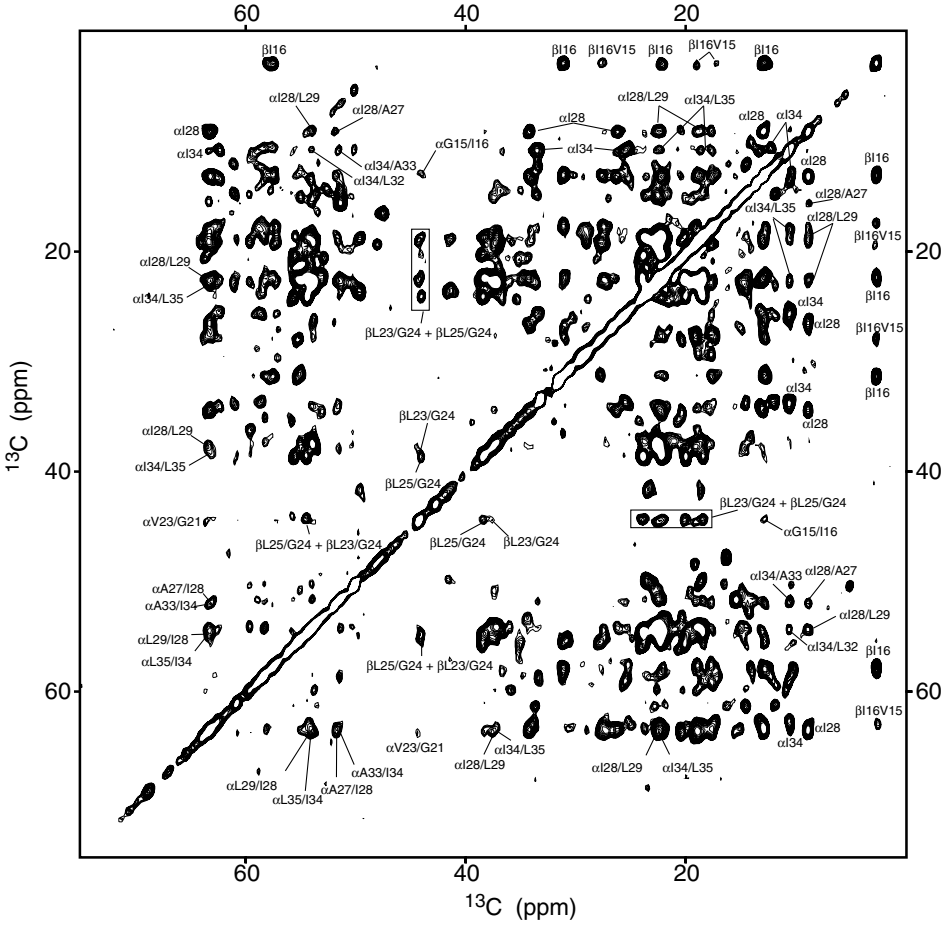


Figure 7. Contour plot of the PDSD500 spectrum of AA-LH2. Clear examples of inter-residue correlations described in text are indicated for αG15–I16, αI28–A27–L28, αI34–A33–L35, βI16–V15 and βG24–L23–L25.

This is promising for the investigation of many other cofactor or ligand binding transmembrane proteins.

The additional correlations in the $PDSD_{500}$ spectrum of the AA-LH2 compared to the $PDSD_{50}$ dataset collected from the same sample can be attributed to long-range correlations between remote I, L, G, A, and V residues. A clear example in the spectrum is provided for the βI16 in Figure 7, where βI16(Cδ) correlates with the ^{13}C responses of the βV15, which is the only labeled residue in the vicinity of βI16. Also the αI26/A27, αI28/A27/L29 and αI34/A33/L35 give rise to long range correlation signals. These long-range correlations can be related to unique combinations in the protein sequence and assigned unambiguously. For the αG15/I16 αG21/αL19 and βG24/βL23/βL25 long-range correlations in the $PDSD_{500}$ spectrum between I or L and G help to resolve signals from these aliphatic residues, which are badly resolved in the $PDSD_{500}$ spectrum due to strong overlap. Finally, signals of the A residues that strongly overlap in the $PDSD_{50}$ spectra of 2,3-LH2 and AA-LH2 are also conveniently resolved in the $PDSD_{500}$ by remote correlations with the A(Cα) and A(Cβ) carbons.

Most of the long-range correlations in both $PDSD_{500}$ spectra in Figure 6 and 7 are between residues that are within one α-helical turn. The strongest long-range correlations are observed for adjacent residues (i,$i\pm1$) in the protein sequence, while the intensities of the correlation signals for i,$i\pm2$ and i,$i\pm3$ are much weaker. The inter-helix distance in the LH2 complex of ~9.0–10 Å is too large for the observation of inter-helical correlations between polypeptide backbone atoms. This

contrasts with β-strand proteins. Here inter-strand correlations are rapidly observed since the distances between two adjacent β-strands are in the range of 4.6–5.4 Å (Castellani et al., 2002, 2003). It is known that interhelical side-chain side-chain contacts can be as short as 4.5 Å. Carbons involved in these contacts are predominantly aliphatic and resonate between 10 and 20 ppm. Since there are also many interresidue correlations, it is difficult to discriminate the long-range from the sequential correlations in this small and crowded part of the correlation spectrum with many correlations that are close to the diagonal.

Conclusions

MAS NMR spectroscopy has been applied on pattern labeled LH2 samples, which have been prepared biosynthetically by using selectively labeled succinic acid or a mixture of free amino acids. The selective and extensive labeling approach yields a reduced number of correlations in the 2D MAS NMR spectra. In addition, it increases the spectral resolution. In particular, for the backbone carbons, the J-couplings in the samples are reduced significantly. It is demonstrated that the use of pattern labeled samples subdivides the many overlapping correlations of the uniformly labeled transmembrane protein samples into datasets with resolved correlations and sequence specific assignments are obtained. The dilute isotope enrichment also paves the way for long-distance transfer of the magnetization, enabling the detection of inter-residue correlations that are used for the sequence specific assignment and for obtaining information about the structure of the protein complex.

The solid state assignment of the backbone carbons from the LH2 complex is nearly complete, albeit that the signals of the residues α47–53 and β40–41 at the two C-termini are not detected in the MAS NMR. This indicates that the C-termini are either flexible or disordered in the protein samples. In contrast, N-termini are clearly observed, showing that the N-termini are ordered uniformly. The order for the αN-terminus is possibly a result of the non-covalent bond between the N-carboxyl-αM1 and the cofactor. The βN-terminus, consisting of only the first four residues, is very short which may explain why it is not significantly disordered. The structure of the rigid part is unique, since no doubling of the resonances is observed. Additional correlation spectra recorded at lower temperature or with a different dipolar correlation sequence may resolve some additional chemical shifts from flexible side chains and from residues in the C-termini for a more detailed assignment.

Previous work on the LH2 complex presented some tentative assignments based on statistical chemical shift information, not based on correlations with sequence specific assigned residues (Egorova-Zachernyuk et al., 2001). In our current study, we have developed pattern labeled samples, which lead to an accurate and unambiguous assignment that is different from the earlier work. This shows that it is not yet possible to arrive at an assignment based on chemical shift information alone. In a preliminary step, the computer program TALOS has been applied to the data in Tables 1 and 2 for the prediction of the secondary structural elements for amino acids in LH2. The correspondence between the predicted structures and the high-resolution X-ray structure was poor. This corroborates that working from chemical shift data alone is insufficient at present.

In addition, unique correlations are observed between carbons from the BChl *a* cofactor and the LH2 protein. These provide structural information on the active site and on the organization of the subunits that interact with the cofactor, i.e. they represent constraints regarding the quaternary structure of the protein. In conclusion, the sequence specific assignment of the LH2 complex in this study shows that it is possible to assign MAS NMR data from long α-helical segments in transmembrane proteins, and paves the way for structure determination of this important class of biological constructs.

Acknowledgements

This research was supported in part by demonstration Project B104-CT97-2101 of the commission of the European Communities. HJMdG is a recipient of a PIONIER award of the chemical sciences section of the Netherlands Organization for scientific research (NWO). C. Erkelens and F. Lefeber are gratefully acknowledged for their support with the NMR experiments. We thank

P. Gast from the Leiden Institute of Physics for the use of the protein preparation facility.

References

Alia Matysik, J., Soede-Huijbregts, C., Baldus, M., Raap, J., Lugtenburg, J., Gast, P., Gorkom, H.J.van, Hoff, A.J. and Groot, H.J.M.de (2001) *J. Am. Chem. Soc.*, **123**(20), 4803–4809.

Baldus, M., Geurts, D.G., Hediger, S. and Meier, B.H. (1996) *J. Magn. Reson. Ser. A*, **118**(1), 140–144.

Baldus, M., Petkova, A.T., Herzfeld, J. and Griffin, R.G. (1998) *Mol. Phys.*, **95**(6), 1197–1207.

Bennett, A.E., Rienstra, C.M., Griffiths, J.M., Zhen, W.G., Lansbury, P.T. and Griffin, R.G. (1998) *J. Chem. Phys.*, **108**(22), 9463–9479.

Böckmann, A., Lange, A., Galinier, A., Luca, S., Giraud, N., Juy, M., Heise, H., Montserret, R., Penin, F. and Baldus, M. (2003) *J. Biomol. NMR*, **27**(4), 323–339.

Castellani, F., Rossum, A.van, Diehl, B., Rehbein, K.Oschkinat, H. (2003) *Biochemistry*, **42**(39), 11476–11483.

Castellani, F., Rossum, B.van, Diehl, A., Schubert, M., Rehbein, K. and Oschkinat, H. (2002) *Nature*, **420**(6911), 98–102.

Detken, A., Hardy, E.H., Ernst, M., Kainosho, M., Kawakami, T., Aimoto, S. and Meier, B.H. (2001) *J. Biomol. NMR*, **20**(3), 203–221.

Egorova-Zachernyuk, T.A., Hollander, J., Fraser, N., Gast, P., Hoff, A.J., Cogdell, R.J., Groot, H.J.M.de and Baldus, M. (2001) *J. Biomol. NMR*, **19**(3), 243–253.

Fujiwara, T., Todokoro, Y., Yanagishita, H., Tawarayama, M., Kohno, T., Wakamatsu, K. and Akutsu, H. (2004) *J. Biomol. NMR*, **28**(4), 311–325.

Hediger, S., Meier, B.H. and Ernst, R.R. (1995) *Chem. Phys. Lett.*, **240**(5–6), 449–456.

Hong, M. (1999) *Biophys. J.*, **76**(1), A392–A392.

Koepke, J., Hu, X.C., Muenke, C., Schulten, K. and Michel, H. (1996) *Structure*, **4**(5), 581–597.

Krogh, A., Larsson, B., Heijne, G.von and Sonnhammer, E.E.L. (2001) *J. Mol. Biol.*, **305**(3), 567–580.

Kühlbrandt, W. (2001) *Nature*, **411**(6840), 896–899.

Marassi, F.M. and Opella, S.J. (2003) *Protein Sci.*, **12**(3), 403–411.

McDermott, G., Prince, S.M., Freer, A.A., Hawthornthwaite-Lawless, A.M., Papiz, M.Z., Cogdell, R.J. and Isaacs, N.W. (1995) *Nature*, **374**(6522), 517–521.

Metz, G., Wu, X.L. and Smith, S.O. (1994) *J. Magn. Reson. Ser. A*, **110**(2), 219–227.

Opella, S.J. and Marassi, F.M. (2004) *Chem. Rev.*, **104**(8), 3587–3606.

Opella, S.J., Nevzorov, A., Mesleh, M.F. and Marassi, F.M. (2002) *Biochem. Cell Biol.*, **80**(5), 597–604.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M. and Miyano, M. (2000) *Science*, **289**(5480), 739–745.

Papiz, M.Z., Prince, S.M., Howard, T., Cogdell, R.J. and Isaacs, N.W. (2003) *J. Mol. Biol.*, **326**(5), 1523–1538.

Pauli, J., Baldus, M., Rossum, B.van, Groot, H.J.M.de and Oschkinat, H. (2001) *Chembiochem*, **2**(4), 272–281.

Prince, S.M., Papiz, M.Z., Freer, A.A., McDermott, G., Hawthornthwaite-Lawless, A.M., Cogdell, R.J. and Isaacs, N.W. (1997) *J. Mol. Biol.*, **268**(2), 412–423.

Schulten, E.A.M., Matysik, J., Alia Kiihne, S., Raap, J., Lugtenburg, J., Gast, P., Hoff, A.J. and Groot, H.J.M.de (2002) *Biochemistry*, **41**(27), 8708–8717.

Simon, I., Fiser, A. and Tusnady, G.E. (2001) *BBA-Protein Struct. M.*, **1549**(2), 123–136.

Straus, S.K., Bremi, T. and Ernst, R.R. (1998) *J. Biomol. NMR*, **12**(1), 39–50.

van Gammeren, A.J., Hulsbergen, F.B., Hollander, J.G. and de Groot, H.J.M. (2004) *J. Biomol. NMR*, **30**(4), 267–274.

van Gammeren, A.J., Buda, F., Hulsbergen, F.B., Kiihne, S., Hollander, J.G., Egorova-Zachernyuk, T.A., Fraser, N.J., Cogdell, R.J. and de Groot, H.J.M. (2004) *J. Am. Chem. Soc.* in press.

Wallin, E. and Heijne, G.von (1998) *Protein Sci.*, **7**(4), 1029–1038.