Spin-label EPR on Disordered and Amyloid Proteins

Hashemi Shabestari, M.

CHAPTER 6

EXPLORING THE STRUCTURE OF THE N-TERMINUS OF THE PLANT ANTENNA PROTEIN CP29

Recently, the crystal structure of the light-harvesting protein CP29 of plants was determined using X-ray crystallography (Pan et al., *Nat.Struct.Mol.Biol.* **2011,** 18, 309-315). However, the structure of the unusually long N-terminal domain of this membrane protein (about 100 amino-acid residues) remained elusive. We have studied the N-terminus with the use of electron paramagnetic resonance, EPR. Our study involves two approaches: exploring the mobility of the spin label with continuous wave EPR and distance determination with a pulsed EPR method. We demonstrate that the N-terminus of CP29 is relatively structured and five regions have been recognized that differ considerably in their dynamics. Two regions are relatively immobile and one of them shows alpha-helical character and is in contact with the bulk of the protein. This immobile part is flanked by highly dynamic and rather unstructured regions (loops) and we speculate that they may be important for the interaction with other light-harvesting proteins. A small part at the end of the N-terminus around residue 4 appears to be immobilized, presumably because it attaches non-covalently to the protein. This section of the N-terminus is close to a phosphorylation site (Thr-6) in related proteins (Fristedt et al., *PLoS One*. **2011**, 6, e24565), for example those encoded by the Lhcb4.2 gene. Phosphorylation might influence the interaction with other antenna complexes, thereby regulating the supramolecular organization in the thylakoid membrane.

Maryam Hashemi Shabestari, Cor J.A.M. Wolfs, Ruud B. Spruijt, Herbert van Amerongen, Martina Huber.

6.1 Introduction

In higher plants, photosynthesis is performed by two multiprotein complexes, photosystem (PS) I and II, which are composed of a core complex and an outer antenna part [1,2]. In PSII, the outer antenna system is built from a large number of complexes out of which the most relevant here are LHCII, the major antenna subunit, and three minor complexes, CP24, CP26, and CP29 [1,3]. Amongst the minor complexes, CP29 is the largest one and believed to collect, conduct, and possibly quench electronic excitation energy in photosynthesis [1,3-12]. Moreover, CP29 is essential for the proper assembly of PSII and its absence leads to a large drop in the average rate of excitation trapping in the reaction centers of PSII [10]. The membrane protein CP29 has five main membrane helices, helices A-E, three of which are trans-membrane and a long N-terminal domain with about 100 amino-acid residues (figure 6.1). The N-terminus contains a phosphorylation site which might be involved in the events occurring under high-light conditions [7] and its phosphorylation leads to a conformational change that can be detected with the use of circular dichroism [13]. The crystal structure of CP29 has recently been solved [14], with the exception of the N-terminal part, which lacked well-defined electron density. Sequence alignment suggests that the related antenna protein LHCII has a shorter N-terminus, in particular, it lacks the residues 58 to 97 that are present in the N-terminus of CP29 [15]. Therefore the better-characterized LHCII cannot be used as a model for the N-terminus of CP29.

Previous studies targeting the structure of the N-terminus of CP29 by FRET [8,10] and EPR [9] indicate that the N-terminus is folded back onto the protein and suggest the presence of some structural heterogeneity in the N-terminal part. In the present study, we employ spin-label EPR to learn more about the structure and dynamics of the N-terminus. Our study involves two approaches. Singly spin-labeled variants were used to study the mobility of the spin label by continuous-wave (cw) EPR and doubly spin-labeled variants were used to determine distances by double electron-electron resonance (DEER), a pulsed EPR method. To this aim, amino-acid residues were mutated one-by-one to a cysteine, spin labeled with a nitroxide spin label (i.e., site-directed spin labeling, SDSL), and reconstituted with the chlorophylls and carotenoids to form the holoprotein. In all cases this holoprotein was investigated. The present study demonstrates that the N-terminus of CP29 is relatively structured and consists of at least five different regions that differ in their secondary structure.

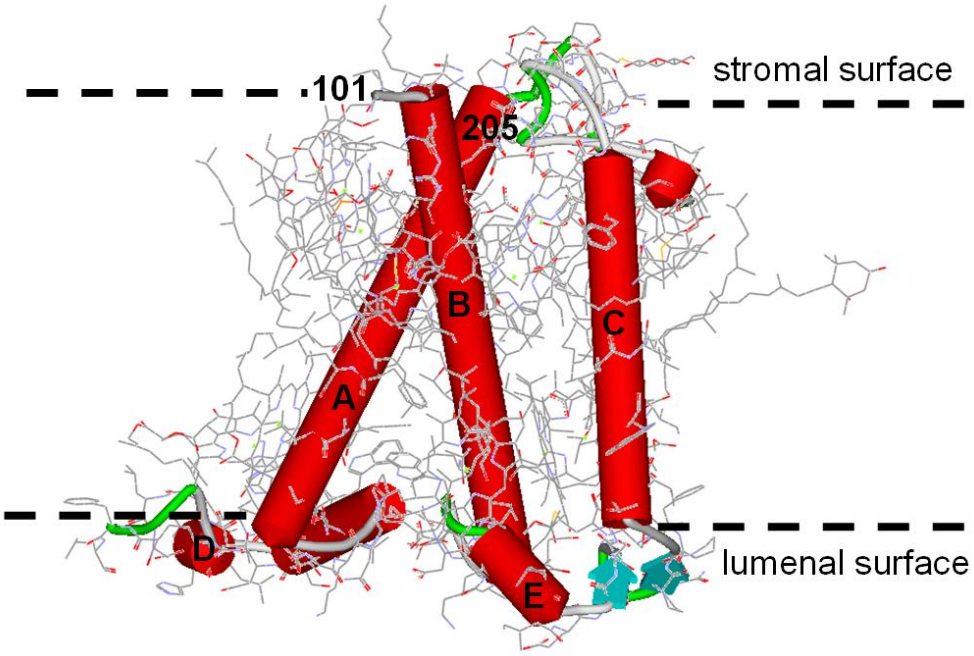

Figure 6.1 Schematic structural model of CP29 based on the X-ray structure of CP29. The five main helices (A-E) are shown in red. The N-terminus, consisting of residues 1-100, is not defined in the structure. Residue 101 is localized at the stromal surface of the protein, at the start of the helix B.

6.2 Materials and methods

6.2.1 Mutagenesis, labeling, and pigment reconstitution

The construction and isolation of overexpressed CP29 apoprotein from Lhcb4.1 cDNA of *Arabidopsis thaliana* (*A. Thaliana*) (from *Arabidopsis* Biological Resource Center DNA Stock Center) were performed as reported before [8,9]. The naturally occurring cysteine (position 108) was replaced by alanine. The mutant protein (C108A) was used to estimate the amount of nonspecific spin labeling [8,9]. Single cysteine mutants were introduced at 55 different positions in the N-terminus of the CP29 apoprotein using this template, resulting in the following mutants: G4C, G6C, A10C, A11C, S15C, A16C, T19C, V20C, T21C, T22C, P29C, G30C, A31C, I32C, S33C, G39C, S40C, L41C, V42C, G43C, G46C, F50C, G51C, L52C, G53C, A56C, E57C, Y58C, L59C, Q60C, F61C, S65C, Q68C, N69C, L70C, A71C, N73C, L74C, A75C, G76C, G80C, T81C, T83C, E84C, A85C, A86C, A88C, S90C, T91C, P92C, F93C, Q94C, S97C, G101C, C108C.

In the DEER experiments six different double cysteine mutants were investigated: G4C/S97C, G4C/G101C, G4C/A205C, A56C/S65C, A56C/T81C, and A56C/S97C.

Pigment isolation, labeling, and reconstitution of CP29 pigment complexes were performed as described before [8,9]. Solutions of the spin-labeled CP29 samples were washed and concentrated in sucrose-free β-D-maltoside (DM) buffer (0.03 % W/V + 10 mM Na_2HPO_4, pH 7.6) just before the EPR measurements. Integrity of the holoprotein samples was checked by fluorescence excitation and emission measurements, showing the complete absence of free chlorophylls and carotenoids in all preparations.

6.2.2 Continuous-wave EPR measurements

The room temperature EPR measurements were performed on an X-band Bruker Elexsys E-500 EPR system (Bruker, Rheinstetten, Germany) equipped with a super-high-Q cavity ER 4122SHQE in combination with a SuperX X-band microwave bridge type ER 049X. Temperature was controlled with a quartz variable-temperature Dewar inset (Eurotherm, Leesburg, VA). Samples were transferred to 50 µl capillaries and placed in a standard 4-mm quartz EPR tube. Spectra were recorded with 10 mT scan width, a microwave power of 5 mW, a modulation amplitude of 0.1 mT, and a scan time of 82 s at 279 K. Up to 150 scans were recorded to improve the signal/noise ratio [9].
The measurements at 80 K were performed using an Elexsys E680 spectrometer (Bruker, Rheinstetten, Germany). A rectangular cavity, equipped with a helium gas-flow cryostat (Oxford Instruments, United Kingdom) with an ITC502 temperature controller (Oxford Instruments, United Kingdom) was used. For the measurements in frozen solution, 3 mm outer diameter quartz tubes were used. To obtain a frozen glass 20 % glycerol was added to the samples before freezing them in liquid nitrogen. The frozen samples were inserted in the pre-cooled helium gas-flow cryostat. The EPR spectra were recorded using a modulation amplitude of 0.2 mT, a modulation frequency of 100 kHz, and a microwave attenuation of 0.159 mW. Typical accumulation time was 40 minutes.

6.2.3 Simulation of the cw EPR spectra

Information on the mobility of the spin label is encoded in the line-shape of the EPR spectra. In case of a non-restricted spin label, the lines are narrow, whereas for a restricted spin label the lines are broad. A spectrum consists of different components, corresponding to fractions of the samples in which spin labels have different mobility. Each component in a spectrum is defined by a rotation correlation time, τ_r, of the spin label. Simulation enabled us to define the τ_r as well as the amount of each component. The spectra were simulated using Matlab and the EasySpin package [16]. For all components the following parameters were used: g = [2.00906, 2.00687, 2.003] [17] and $A_{xx} = A_{yy} = 13$ MHz. The value used for A_{zz} was

similar for the fast and medium component but different for the slow component, as discussed before [17]. Over-modulation effects were taken into account in EasySpin. The quality of the agreement between the experimental and the simulated lines was checked by visual inspection.

6.2.4 Assessment of the cw EPR spectra

A quick method to define the mobility of the spin label, rather than performing the relatively time-consuming simulations, was to measure amplitude ratios directly from the spectra [18,19]. Here we used the ratio of the mobile (R) component to the strongly immobilized (L) component (R/L). The contribution of these two components was obtained by selecting specific positions in the EPR spectra B_{0L} (the magnetic field value at the outermost left peak = 329.4 ± 0.1 mT) and B_{0R} (the magnetic field value at the second peak from the left = 330.5 ± 0.1 mT) at which one component has a large amplitude and the other a small one. The ratio of the amplitudes of the two selected spectral positions (R/L) gives an indication of the mobility. The larger the ratio, the higher the mobility is. Certain periodicities in the mobility of sequential residues indicate secondary-structure elements: a periodicity of 2.0 a β-sheet and a periodicity of 3.6 a regular α-helix.
Another measure of the mobility was obtained using the inverse of the central line width (ΔB^{-1}, mT^{-1}) [19,20] and the inverse of the second moment ($<\Delta B^2>^{-1}$, $<mT^2>^{-1}$) [21-23]. The ΔB^{-1} parameter [19,24] is the higher the more mobile the spin label is. A more detailed analysis [21-23] defines the main topographic regions of a protein in a plot of the ΔB^{-1} vs. the $<\Delta B^2>^{-1}$: loop/surface sites, loop/contact sites, helix/surface sites, and helix/buried sites. We refer to this plot as the Hubbell plot (see Results).

6.2.5 Parameters to estimate the length of protein regions

The distance between two successive C_α atoms, $C_{\alpha i}$- $C_{\alpha i+1}$, in a protein in the random-coil conformation is reported to be 0.38 nm, assuming that all amide groups are restricted to the trans conformation [25]. So if the N-terminus of the CP29 (about 100 amino-acid residues) was in a random coil conformation, it would have a length of about 38 nm. For comparison, the largest dimension of the stromal surface of the CP29 protein is about 3.2 nm (distance between the farthest two residues e.g. 164-188 or 88-180), which is about 1/10 of the maximum total length of the N-terminus. So the N-terminus is not likely to be in an extended conformation. The distance separating each turn of a helix in the direction of the cylindrical axis is 0.54 nm, which results in a distance of 0.15 nm/residue for the length of the α-helix. For loops, two extreme situations are considered, a fat and a thin loop, which are maximally extended horizontally and vertically with respect to the stromal surface, respectively. Assuming a minimal turn diameter of 0.8 nm, the horizontally

extended loop has a height of 0.8 nm covering at least two residues at both ends, which leaves for a loop of n_L residues, n_L- 4 residues for the width of the horizontal loop. For the vertically extended loop, the height and the width of a loop of n_L residues would be ½ * n_L- 2 and 0.8 nm, respectively. From these numbers the length of different sections of the protein can be estimated (see results).

6.2.6 Pulsed EPR measurements

The DEER measurements were performed at X-band on an Elexsys E680 spectrometer (Bruker, Rheinstetten, Germany) at frequencies of about 9.3 GHz using a 3 mm split-ring resonator. The temperature was kept at 40 K with helium gas in a CF935 (Oxford Instruments, United Kingdom) cryostat with an ITC502 temperature controller (Oxford Instruments, United Kingdom). Samples were prepared in 3 mm outer diameter quartz tubes and were frozen in liquid nitrogen before insertion into a pre-cooled helium gas-flow cryostat. The pump and observer frequencies were separated by 70 MHz and adjusted as reported before [26]. The power of the pump-pulse was adjusted to invert the echo maximally [27]. The lengths of the pulses at the observer frequency were 16 and 32 ns for the $\pi/2$- and π-pulses. The length of the pump pulse was 12 ns. All DEER measurements were performed as two-dimensional experiments, in order to suppress the proton modulation [27]. The DEER time traces were measured for ten different τ_1-values spaced by 8 ns starting at τ_1 = 200 ns. The typical accumulation time per sample was 16 hours. The DEER data were analyzed using the "DeerAnalysis" program 2011 [28], which is available from www.epr.ethz.ch/software/index. After background correction of the data, the distance distribution was determined by Tikhonov regularization with an optimum regularization parameter determined by the L curve criterion [27,28]. Significance of features at long distances was checked with the validation tool in "DeerAnalysis".

6.3 Results

6.3.1 Continuous-wave EPR

The cw EPR spectra of the reconstituted holoproteins in the detergent-micelle solution are shown in figure 6.2. The spectra differ substantially from each other. Some of them have particularly broad lines (e.g. the spectra of S33C and G101C). Others have sharp lines (e.g. the spectra of A10C and A86C). As seen from the residue numbers, broader lines occur throughout the N-terminus.

More detailed information is obtained by spectral simulation, performed for selected spectra. For each spectrum (figure 6.3) at least two components are required that differ in rotation-correlation time τ_r (table 6.1). The broad-line spectra of the spin label at residues 4, 40, and 101 are simulated with three components.

Each spectrum has a substantial slow component (at least 18 %) with a τ_r longer than 50 ns, and maximally 5 % of a fast component (τ_r of about 1 ns). In the spectra of the spin label at residues 11 and 86, the slow component is absent, which indicates the more mobile character of these positions. The spectra have almost equal contributions of a fast component with a τ_r of about 1 ns, and a medium component with a τ_r of about 4 ns. Also, the spectrum of the spin label at residue 70 consists of two components, only in this case the amount of the medium component (τ_r of about 4 ns) is significantly larger than that of the fast component (τ_r of about 1 ns) (table 6.1).

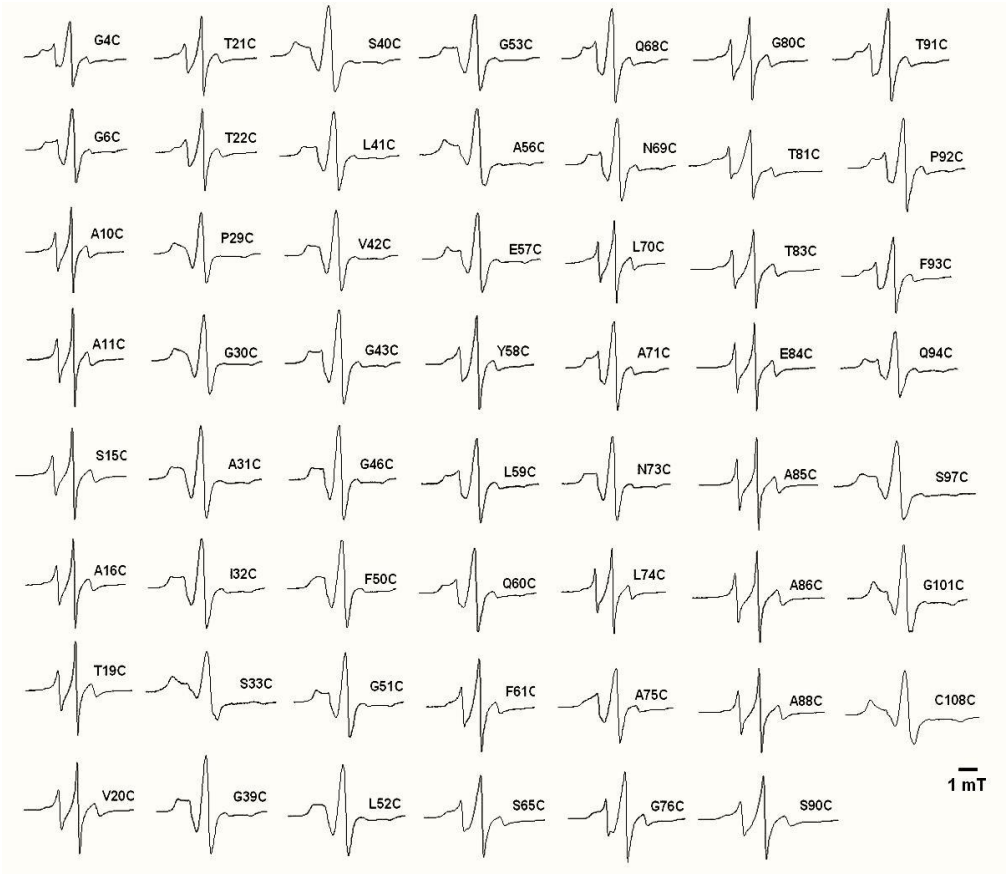

Figure 6.2 Room temperature EPR spectra of 55 singly labeled CP29 protein samples. In each spectrum the position of the nitroxide spin label is indicated.

Because of the multicomponent character of these spectra an easier approach, rather than the time-consuming simulation, is to compare the spectral intensities, R/L. The intensity at position R (figure 6.4 insert) derives from the fast and medium

components and at position L from the slow component. The higher the R/L ratio, the more mobile the spin label. The dependence of the R/L ratio is illustrated in figure 6.4 as a function of the residue number. It gives qualitative information, i.e., relative mobility of the stretch of protein investigated. Amongst the investigated residues (55 out of 110) about 40 % appeared to be immobile/ restricted, 35 % have intermediate mobility and only 25 % are highly mobile. Ultimately, also including the further analysis (see below), five distinct mobility regions are determined within the N-terminus: I: residues 10 to 22, II: residues 23 to 57, III: residues 58 to 81, IV: residues 82 to 91, and V: residues 92 and 108 (figure 6.4). Region V includes a stretch of residues (101-108) assigned to the transmembrane helix B, which extends from residue 101 onwards. The borders between regions are not precisely defined.

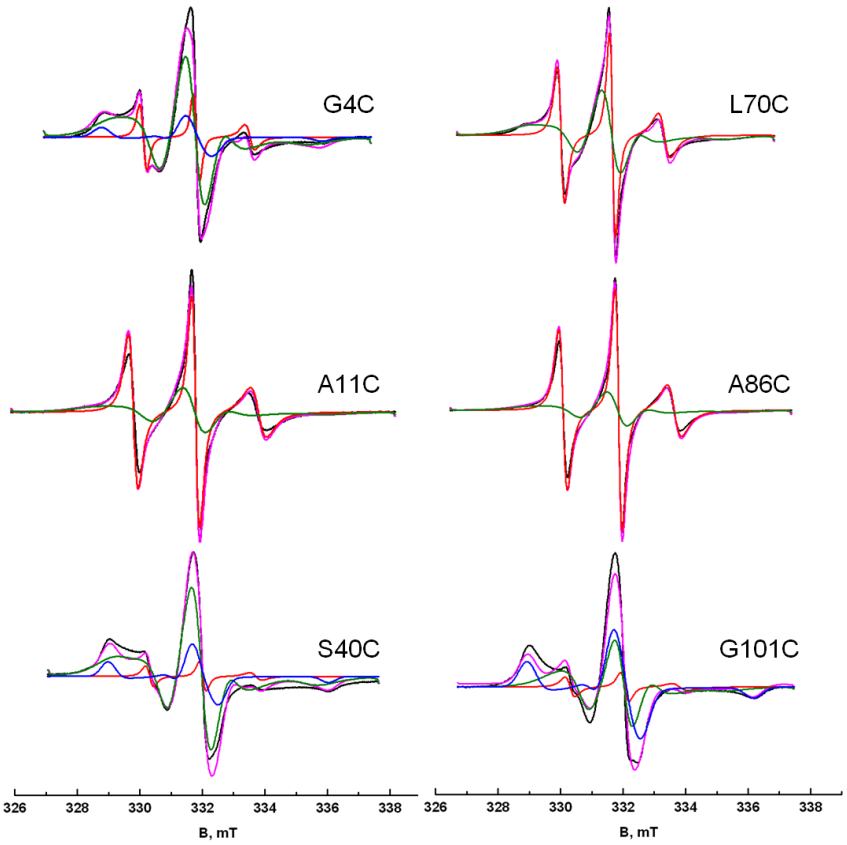

Figure 6.3 Room temperature EPR spectra of CP29 singly labeled at positions 4, 11, 40, 70, 86, and 101. The experimental spectra are shown in black. The simulated spectra are shown in magenta. The simulated spectra are adjusted to fit the intensity in the low field of the spectra to account for the asymmetry in the experimental spectra. Each simulated spectrum is a sum of multiple components. The fast, medium and slow components are shown in red, green and blue, respectively.

Table 6.1 The EPR parameters derived from the simulation of the EPR spectrum of singly labeled CP29 proteins. Given are: τ_r, rotation-correlation time, A_{zz}, hyperfine splitting along the z-direction, lw, the component line-width, and (%), the contribution of the given component to the total spectrum.

	fast				**medium**				**slow**			
residue	**τ_r (ns)**	**A_{zz} (MHz)**	**lw (mT)**	**%**	**τ_r (ns)**	**A_{zz} (MHz)**	**lw (mT)**	**%**	**τ_r (ns)**	**A_{zz} (MHz)**	**lw (mT)**	**%**
4	0.67 ± 0.01	109	0.14 ± 0.01	5	4.00 ± 0.10	109	0.32 ± 0.03	72	> 50	94	0.60 ± 0.05	18
11	1.05 ± 0.01	109	0.08 ± 0.01	47	4.00 ± 0.10	109	0.32 ± 0.03	53	-	-	-	-
40	0.83 ± 0.01	109	0.14 ± 0.01	2	4.30 ± 0.10	109	0.32 ± 0.03	78	> 50	94	0.60 ± 0.05	20
70	0.95 ± 0.01	109	0.08 ± 0.01	26	4.20 ± 0.10	109	0.32 ± 0.03	74	-	-	-	-
86	1.05 ± 0.01	109	0.08 ± 0.01	55	4.00 ± 0.10	109	0.32 ± 0.03	45	-	-	-	-
101	1.15 ± 0.01	109	0.14 ± 0.01	4	4.30 ± 0. 10	109	0.32 ± 0.03	40	> 50	94	0.60 ± 0.05	56

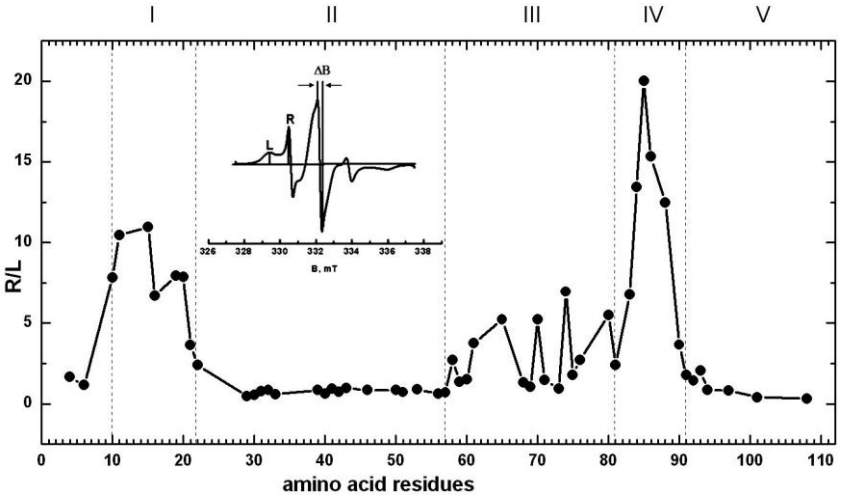

Figure 6.4 The R/L ratios versus the amino-acid residue number. The ΔB, L, and R peak indications are depicted on the inset spectrum. The line which connects the R/L values is just a guide to the eye. Five approximate regions (I-V) are defined according to the ratio of R/L, which is suggestive of different mobility regions within the N-terminus: I: residues 10 to 22, II: residues 23 to 57, III: residues 58 to 81, IV: residues 82 to 91 and V: residues 92 to108.

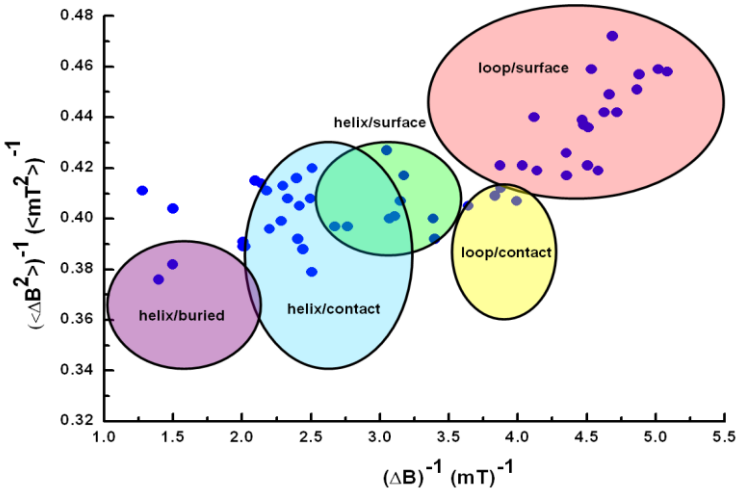

Figure 6.5 The values of the inverse second moment, $<\Delta B^2>^{-1}$, of the EPR lines, and of the inverse of the central line-widths, ΔB^{-1}, of the 55 singly labeled CP29 as determined from the EPR spectra (blue dots). The topological regions of a protein are indicated by ovals according to the Hubbell plot. The color-coded regions represent secondary structure in a protein: red: loop/surface site, yellow: loop/contact site, green: helix/surface site, blue: helix/contact site, and purple: helix/buried site.

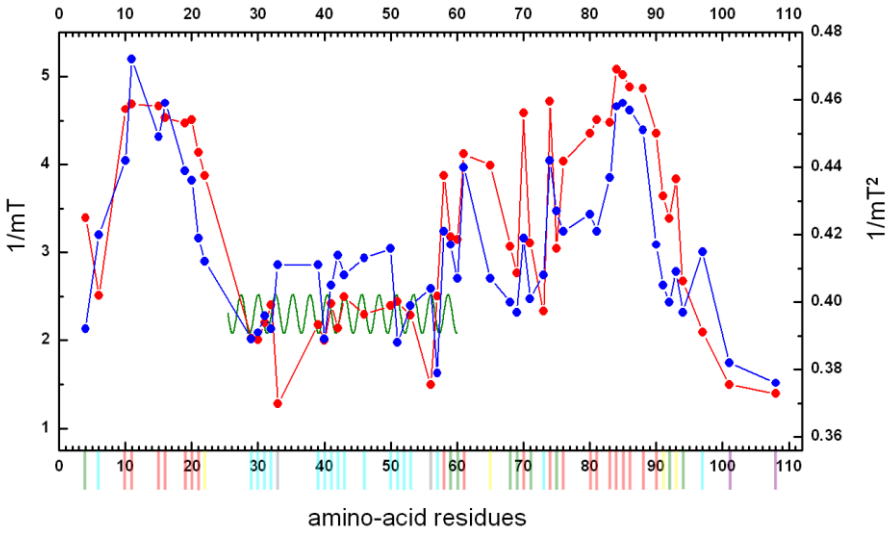

Figure 6.6 The $<\Delta B^2>^{-1}$ (blue dots) and the ΔB^{-1} (red dots) values versus the amino-acid residue number are given. For each amino-acid residue the color code, which is defined according to the Hubble plot (figure 6.5) is given along the horizontal axis. The superimposed solid-green line shows the 3.6 amino-acid periodicity of an alpha-helix.

More details can be obtained from the Hubbell plot [20-23]. It classifies mobility in terms of the location of the residue in the protein (topographic regions), using the inverse of the central line-width (ΔB^{-1}) [19,20] and the inverse of the second moment ($<\Delta B^2>^{-1}$) [20-23]. For the spin-labeled variants of CP29 these values are shown as blue dots in the Hubbell plot (figure 6.5). This plot also shows the topographic regions as colored areas [20,29]. Most points fall within these regions. To relate these points to the sequence, figure 6.6 shows the ΔB^{-1} and $<\Delta B^2>^{-1}$ parameters and the region color coding as a function of the sequence number of the residue. As shown in figure 6.6, the first high-mobility region I fits with the red area in the Hubbell plot, indicating a loop/surface region. The next region with low mobility II falls into the blue area in the Hubbell plot, indicating a helix/contact region. Next is region III, which spreads over the green, red, and yellow areas, indicating a helix/surface and a loop/contact region, respectively. The next high-mobility region IV is mainly in the loop/surface region of the Hubble plot. The last low-mobility region V coincides with the helix/buried region in the Hubbell plot, close to the transmembrane part of the protein.

We also checked for a possible periodicity in the mobility of sequential residues. While no part of the sequence exhibits a periodicity of two, indicative of a β-sheet, one stretch of residues (residues 39-46) exhibits a 3.6 periodicity (green line in figure 6.6), which suggests an α-helical stretch. The periodicity for residues 39 to 46 is most pronounced in the inverse of the central line-width. This proposed α-helical stretch is projected on an α-helical wheel diagram (figure 6.7), which reveals no clear hydrophobicity/ hydrophilicity pattern. Several residues on both sides of this helical region also fall on the green line, potentially extending the helix to residue 29 respectively 53. Clear evidence of periodicity requires data points from several successive residues and therefore, the periodicity between residues 29 to 38 and 47 to 53 cannot be confirmed. Since all residues between 29 and 53 are in the helix/contact region of the Hubbell plot, it is likely that the α-helix extends beyond the residues 39 to 46.

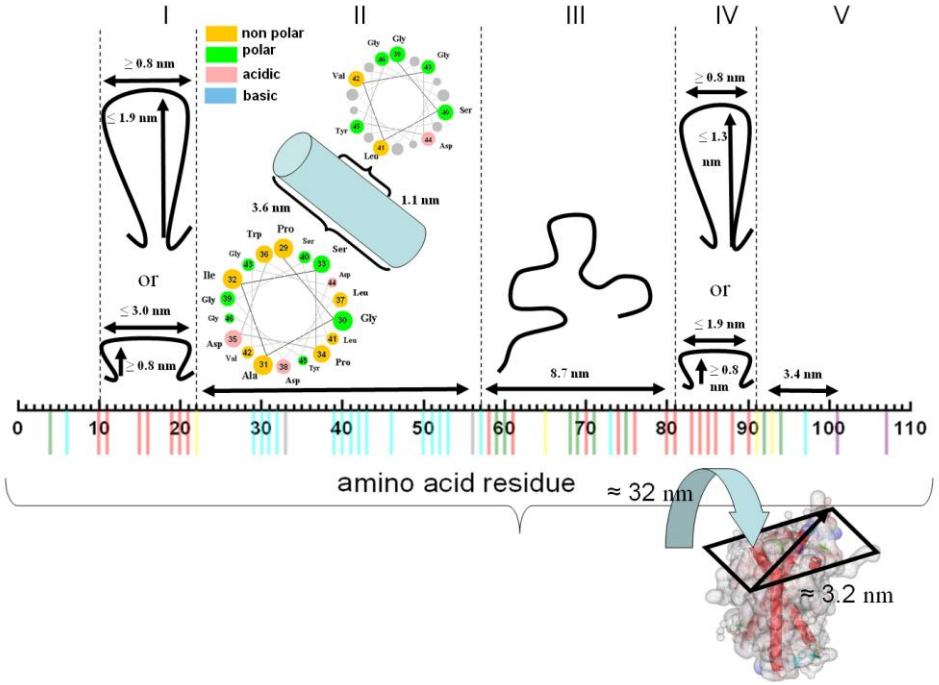

Figure 6.7 Interpretation of the mobility data obtained by EPR in the single mutants of CP29, giving geometrical constraints for the N-terminus. The residues of the N-terminus are shown horizontally, with colors from the Hubbell plot (figure 6.6). As a reference the diagonal of the stromal surface is 3.2 nm. The position of residues 101-110 should be close to the stromal surface and helix B. As an extended chain, residues 92-101 would cover a length of 3.4 nm (region V). Different models for the chain conformation result in the lengths shown for regions IV, III, and I (for details see text). Extreme loop dimensions, assuming minimum turn diameters of 8 Å and comparing long/narrow or wide/flat loops. Region II can be an α-helix. Helical-wheel diagram for region II with the corresponding amino-acid residues are depicted in colors. The color codes represent the type of amino-acid residues in the helical wheel. The numbers on each circle refer to the number of the residue within the N-terminus. Going down the helix, the size of the circles representing the residues in the helical wheel decreases. (http://cti.itc.Virginia.EDU/~cmg/Demo/wheel/-wheelApp.html)

6.3.2 Pulsed EPR

Six different intra-molecular distances within the N-terminus of the CP29 protein are assessed using pulsed EPR measurements. The DEER time traces show clear non-exponential decays, which differ for all double mutants (figure 6.8). The absence of visible modulation indicates multiple distances or broad distance distributions. Distance distributions are obtained by the Tikhonov regularization method (see materials and methods). The distance distributions are fitted to a sum of Gaussians, the parameters of which are given in table 6.2. Most of the distance

distributions are composed of multiple Gaussians (figure 6.8), suggesting that the N-terminus adopts several conformations. For all mutants but 56/65, the distributions have at least two components with a distance larger than 2 nm (table 6.2). Components with distances below 1.8 nm are disregarded because they cannot be reliably determined under our experimental conditions [30]. The presence of very short distances (below 1.5 nm) can be excluded from the absence of line broadening in cw EPR spectra [31] of the same mutants (figure 6.9).

Table 6.2 The distance parameters for doubly labeled CP29 proteins, which are derived from the analysis of DEER data. The DEER data are analyzed by means of Tikhonov regularization. Given are: <r>, distance in nm; S(r), the width of the distance distribution in nm, % is the contribution of each peak and the number of spins. The width of the distance distributions reflects the unknown conformation distribution of the protein domains, which are carrying the spin labels, and the conformation distribution of the spin labels themselves. To define the number of spins, we compare the experimental time traces to the time trace of a standard sample, which contains two spins.

mutant	**$\langle r_1 \rangle$ nm**	**S (r_1) nm**	**%**	**$\langle r_2 \rangle$ nm**	**S (r_2) nm**	**%**	**$\langle r_3 \rangle$ nm**	**S (r_3) nm**	**%**	**$\langle r_4 \rangle$ nm**	**S (r_4) nm**	**%**	**number of spins**
4/97	–	–	–	3.19	0.84	28	4.27	0.67	20	4.94	0.98	52	1.38
4/101	1.96	1.28	35	3.53	0.68	22	5.14	0.95	43	–	–	–	1.46
4/205	1.14	1.15	44	3.14	0.84	15	3.69	0.34	2	5.06	0.76	39	1.35
56/65	1.92	1.7	95	1.67	0.6	5	–	–	–	–	–	–	1.79
56/81	2.86	1.3	54	3.72	0.9	29	5.62	0.89	17	–	–	–	2.26
56/97	1.24	1.51	49	2.95	1.62	21	3.93	0.79	11	5.6	0.9	19	2.01

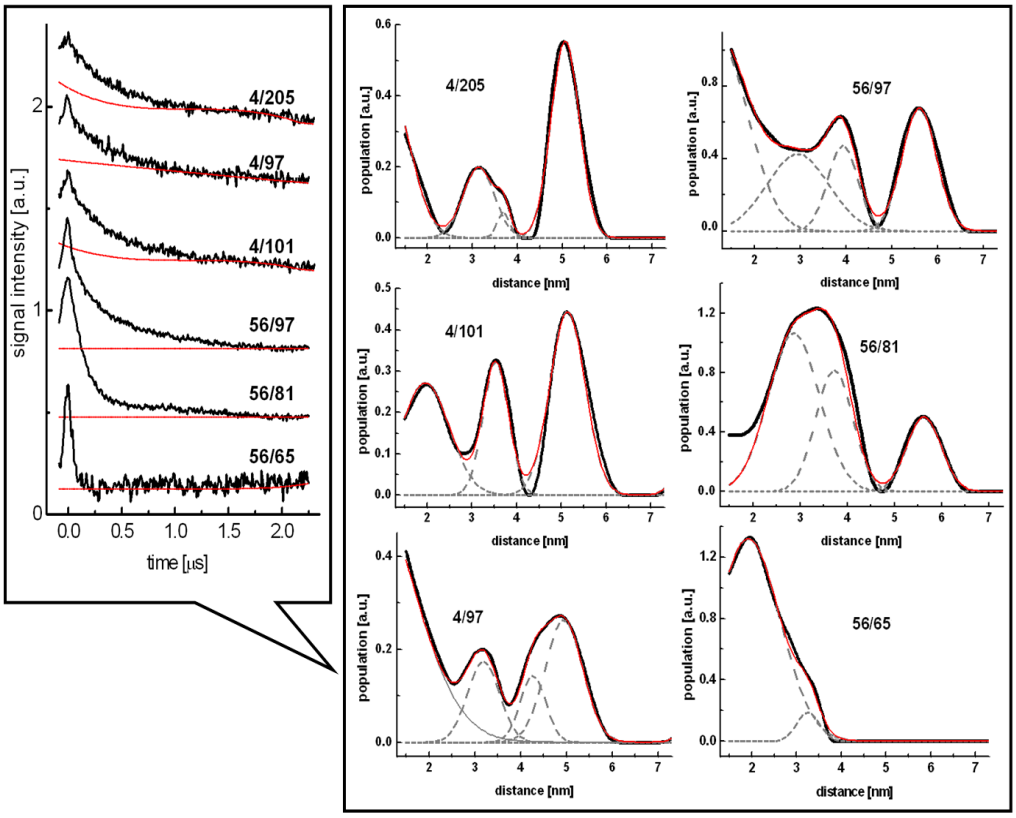

Figure 6.8 DEER data for mutants 4/97, 4/101, 4/205, 56/65, 56/97, and 56/81 are shown. Left: original DEER traces with the baseline used for correction. Right: individual distance distributions are given for each doubly labeled mutant with the corresponding Gaussian fits (performed using origin lab software). The experimental curve is shown in black and the best fit is shown in red.

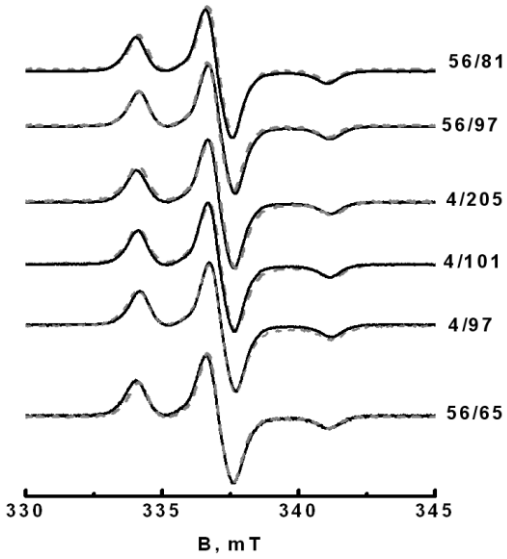

Figure 6.9 Continuous wave EPR spectra of the doubly-labeled mutants (solid lines) obtained at 80 K superimposed with the spectra of the respective singly labeled counterparts (dashed lines).

6.4 Localization of residue 4 of the N-terminus

The distances between residue 4 and residues 101 and 205, and the distance between residue 4 and the most-nearby chlorophyll as obtained from a time-resolved FRET study [8] in the CP29 protein has been used to obtain information on the position of the end (residue 4) of the N-terminus. The crystal structure of spinach CP29 [14] served as a template in a triangulation approach described in detail in appendix B. To take into account the multiple distances obtained with DEER, all geometrically possible combinations of these distances are considered. Eliminating positions located in the protein interior, four positions remain as possible locations for residue 4 (figure 6.10.b). These positions are all close to the stromal membrane surface. Three of these positions cluster in the area where helices B and C reach the protein surface, and one is located farther from the protein surface, close to the membrane surface.

Also, a hydrophobic groove at the stromal side of the protein had been discussed as a possible location for residue 4 [8]. We note that the distances from residue 205 to selected residues in that groove (R93, Q169, K170, Y173, P174, G175, G176, D179, A184, S185, K190, and E194) are all between 0.38 and 1.71 nm. Therefore, that particular location for residue 4 would not show up in our analysis because it

would yield distances between residue 4 and 205, which are outside the distance range experimentally accessible.

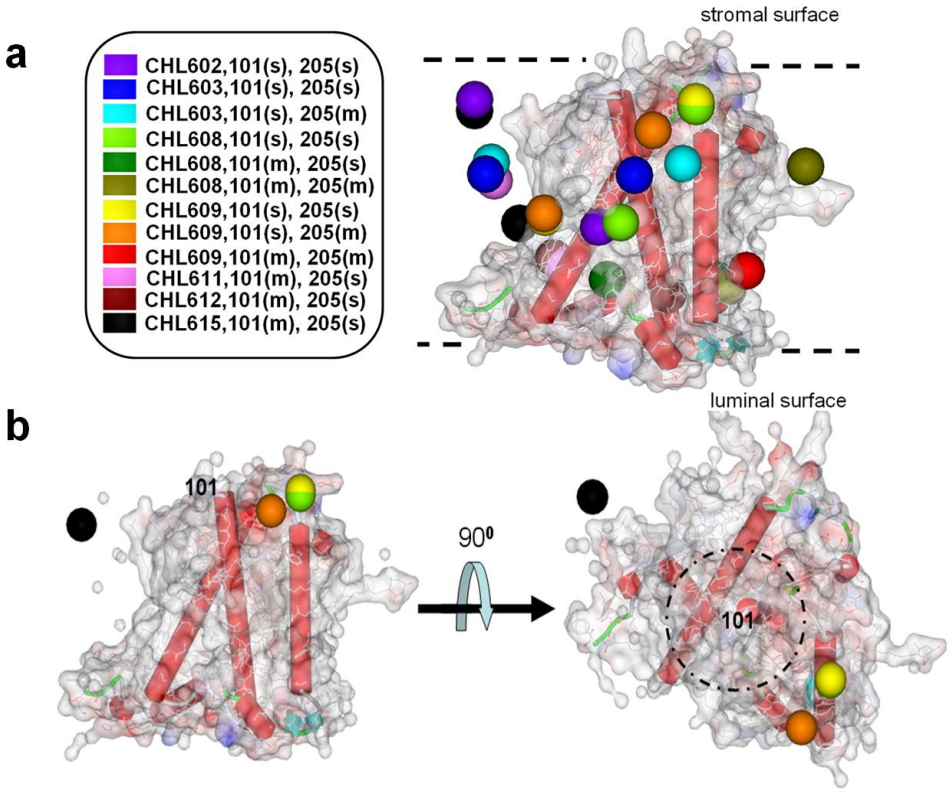

Figure 6.10 Possible locations of residue 4 shown as colored spheres, superimposed on the three-dimensional structure of the CP29 protein derived from the X-ray crystallography (orientation as in figure 6.1). a: All geometrically possible positions for residue 4. The positions are determined using distances from DEER and FRET (see text). The distance between the Mg-atom of all eight chlorophylls (see appendix B) to the C_α- atoms of residue 101 (88 in *Spinach*) and residues 205 (191 in *Spinach*) are measured using the coordinates from the X-ray structure [14]. The (s) and (m) stand for short and medium, which indicates the multiple distances obtained with DEER. Using a triangulation approach, 12 points of intersection are obtained. The intersection points are presented in different colors. For each color a three-section code is given. Each three-section code represents which chlorophyll (CHL) is used in combination with which two residues. The number given in front of CHL gives the number of the chlorophyll. b: Four positions remain possible for residue 4 after eliminating those positions that are located in the protein interior. The projection of the protein in the right hand side of the figure b is such that the yellow sphere eclipses the green sphere. The view along the membrane plane is given on the left and the view from the stromal side is given on the right. The dotted circle indicates the possible location of residue 97.

6.5 Discussion

We have investigated the N-terminus of CP29 spanning residues 1-110. This part of the protein is often considered to be flexible and only residues 101-110 are well defined in the X-ray structure [14]. Determining the structure of such a flexible and long region of a protein is challenging. As yet the reported algorithms for prediction of loop structures were applied for loops with 4 to 20 amino-acid-residues only [32].

Mobility of the spin label provides important markers for the structure. Here, we compare different methods of analysis of mobilities. Most rigorous is the simulation of the spectra, which reveals the multicomponent nature of the individual spectra. Absolute rotation-correlation times and the relative contribution of each component are obtained. Remarkably, spectra classified as strongly immobilized (residues 4 and 40) can have as little as 20 % of the slow component, so these spectra appear immobilized on account of a small ($\leq$ 5 %) population of the fast component rather than a large amount of the slow component. The multicomponent nature of the spectra could indicate different conformations of the N-terminus, however, spectra with multiple components have also been reported for single-conformation proteins [33]. Consequently, the multiple components are consistent with, but not proof for multiple conformations of the N-terminus. The interpretive value of the simulations is limited because the relation of the components and their τ_r values to the structural features of the protein is not yet understood.

The R/L method provides a qualitative measure of mobility. Comparison with the simulation shows that the spectral amplitude at the field position R derives mostly from the mobile components (the fast and the medium fraction in the simulation), whereas the amplitude at L derives from the least mobile spin labels (the slow fraction in the simulation). The R/L ratio is straightforward to determine, but because it depends on the composition of the spectrum it is difficult to apply consistently to other proteins. The analysis based on the Hubbell plot [19,20] refers to a method for which ample reference data is available, and which enables comparison to known reference protein positions. Using the R/L method three regions of low mobility are identified: the N-terminus itself (residues 4 and 6) and the regions labeled II, and V in figure 6.4. Comparison with the Hubbell values (figure 6.6) reveals that the mobility of the spin label in region II corresponds to those found in the helix contact regions, those at III to the outer surface of proteins, and those at V to the helix/buried regions.

The distance measurements show that the N-terminus is not extended. For the nine residues from 56-65, for which the fully extended chain would have a length of 3.4 nm a length of approximately 2 nm is found. This compactness is even more evident for the mutant 56-81, which has a major contribution of a distance of 2.9

nm and a minor one of 3.7 nm, whereas the fully extended length would be 9.5 nm. The sequentially more distant pair of 56-97 has a dominant distance around 3 nm. Another indication for compactness and close association of the N-terminus is the location of residue 4. In a significant fraction of the protein this residue is close to the protein surface.

The data point to multiple conformations of the N-terminus. This follows directly from the multiple distances found in DEER experiments (table 6.2) and is consistent with multi-component cw EPR spectra (table 6.1).

Figure 6.7 shows how the mobility information relates to structural features of the 100-amino-acid-residue N-terminus. Residue 101 is located at the stromal end of helix B and its position is known from the X-ray structure. Residue 97 is 4 residues away from residue 101, the first residue defined in the X-ray structure, so residue 97 must be in a radius of maximally 1.5 nm from the entrance of helix B on the stromal side of the protein (figure 6.10). In region IV, residues have a high mobility, which is suggestive of a loop region. Assuming standard parameters for turns and interresidue separation (see materials and methods) such a loop can have two extreme shapes: as a flat loop it covers 1.9 nm in width (figure 6.7), and as a long and thin loop it extends 1.3 nm into the stroma. Region III (residues 58 to 81) has a loop/contact- helix/contact character. So it must be more strongly attached to the protein surface than region IV. In region II residues 39 to 46 have a pronounced 3.6 periodicity. Given that the entire stretch of residues in region II lies in the helix/contact region of the Hubbell plot, it may well be that this helix extends from residues 29 to 53, corresponding to a total length of the helix of 3.6 nm. Finally, region I that is flexible and shows no periodicity likely is a loop. For region I a flat loop would cover 3.0 nm almost traversing the stromal surface of the protein and as a long and thin loop it would extend 1.9 nm into the stroma, which shows that this loop is longer than the loop in region IV.

The overall architecture reveals non-covalent attachment of several N-terminal residues (at least residues 4-6), of the middle part (residues 39-46) and of the end of the N-terminus (from residue 100 onwards). Two flexible regions (I and IV) flank a less flexible region. The N-terminal part of the protein is located on the stromal side of the protein, which suggests that the less mobile regions are embedded in the stromal surface of the protein.

The first conclusion is that the N-terminus, which in its fully extended form would have a length of 38 nm, is surprisingly structured. It loops back-and-forth on the stromal surface of the protein.

Considering figure 6.10, a location of residue 97 in the vicinity of the stromal side of helix B would yield distances between residues 97 and 4 that are consistent with

the proposed locations of residue 4 derived from distances involving residues 101, 205 and FRET data. Four possible locations of residue 4 are found, out of which three (yellow, green, and orange spheres in figure 6.10.b) cluster on the region where helices B and C reach the surface of the protein. The closeness of the three positions suggests one location. The fourth position (black sphere, figure 6.10.b) suggests a second location for residue 4. A third location of residue 4 in the hydrophobic groove cannot be excluded from the data. We further speculate that the first location corresponds to the largest population of N-terminus conformations.

The presence of multiple conformations of the N-terminus, which is concluded from the present EPR investigation is in agreement with the heterogeneity reported before [8,9]. Two loop regions are reported, a short one, and a longer one, which are either sticking out into the stroma [9] or lying flat on the hydrophobic core most of the time [8]. According to these EPR and FRET studies [8,9], residue 15 is part of the short loop, and residues 65, 82, and 90 belong to the long loop. Similar to those previous reports, we find two loops. We differentiate the reported long-loop region [8,9] further into two mobility sections (regions III and IV), rather than one continuous mobility section. So essentially the long loop of Kavalenka et al. [8] and Berghuis et al. [9] becomes our short loop, containing both phosphorylation sites [34]. Compared to the previous EPR study, around residue 65 we find several residues of lower mobility (residues 57, 59, 60, 68, 69, 71, 73, and 75) arguing that residues 58 to 81 do not form part of the loop, although residue 65 has a relatively high mobility. Therefore, in addition to the report of Kavalenka et al. [9], where 10 residues of the N-terminus were investigated, a lot of information can be resolved by characterizing 55 residues. The structure of residues 58-97 is particularly relevant, because these residues are only present in CP29, and not in LHCII, which otherwise is similar to CP29. This may help to explain the difference in function of CP29 and LHCII in the outer membrane antenna complex.

The present study demonstrates that the N-terminus of CP29 is relatively structured and is separated into different regions with strongly differing mobilities. Moreover, there are clear signs of heterogeneity, demonstrating that the N-terminus is not fixed in one particular structure but is able to adopt various conformations. The presence of phosphorylation sites that are affected by light conditions in this region of the protein, for example at position 81 and 83 [34], is a strong indication that the N-terminus is important in membrane organization. The flexibility of the N-terminus found in the present study thus may be functionally important, because it enables the N-terminus to switch between different conformations in order to let the system adapt to different environmental (light) conditions.

Reference List

[1] R.Croce, H.van Amerongen, *J.Photochem.Photobiol.B* **2011,** *104* 142-153.
[2] G.D.Scholes, G.R.Fleming, A.Olaya-Castro, G.R.van, *Nat.Chem.* **2011,** *3* 763-774.
[3] S.Caffarri, R.Kouril, S.Kereiche, E.J.Boekema, R.Croce, *EMBO J.* **2009,** *28* 3052-3063.
[4] S.Caffarri, K.Broess, R.Croce, A.H.van, *Biophys.J.* **2011,** *100* 2094-2103.
[5] T.K.Ahn, T.J.Avenson, M.Ballottari, Y.C.Cheng, K.K.Niyogi, R.Bassi, G.R.Fleming, *Science* **2008,** *320* 794-797.
[6] N.S.Ginsberg, J.A.Davis, M.Ballottari, Y.C.Cheng, R.Bassi, G.R.Fleming, *Proc.Natl.Acad.Sci.U.S.A* **2011,** *108* 3848-3853.
[7] S.Mauro, P.Dainese, R.Lannoye, R.Bassi, *Plant Physiology* **1997,** *115* 171-180.
[8] B.A.Berghuis, R.B.Spruijt, R.B.M.Koehorst, A.van Hoek, S.P.Laptenok, B.van Oort, H.van Amerongen, *European Biophysics Journal with Biophysics Letters* **2010,** *39* 631-638.
[9] A.A.Kavalenka, R.B.Spruijt, C.J.Wolfs, J.Strancar, R.Croce, M.A.Hemminga, A.H.van, *Biophys.J.* **2009,** *96* 3620-3628.
[10] B.van Oort, S.Murali, E.Wientjes, R.B.M.Koehorst, R.B.Spruijt, A.van Hoek, R.Croce, H.van Amerongen, *Chemical Physics* **2009,** *357* 113-119.
[11] M.Mozzo, F.Passarini, R.Bassi, A.H.van, R.Croce, *Biochim.Biophys.Acta* **2008,** *1777* 1263-1267.
[12] Y.E.Chen, Z.Y.Zhao, H.Y.Zhang, X.Y.Zeng, S.Yuan, *J.Exp.Bot.* **2013**.
[13] R.Croce, J.Breton, R.Bassi, *Biochemistry* **1996,** *35* 11142-11148.
[14] X.Pan, M.Li, T.Wan, L.Wang, C.Jia, Z.Hou, X.Zhao, J.Zhang, W.Chang, *Nat.Struct.Mol.Biol.* **2011,** *18* 309-315.
[15] S.Jansson, *Trends Plant Sci.* **1999,** *4* 236-240.
[16] S.Stoll, A.Schweiger, *Journal of Magnetic Resonance* **2006,** *178* 42-55.
[17] I.Sepkhanova, M.Drescher, N.J.Meeuwenoord, R.W.A.L.Limpens, R.I.Koning, D.V.Filippov, M.Huber, *Applied Magnetic Resonance* **2009,** *36* 209-222.
[18] Q.F.Ma, J.Hu, W.H.Wu, H.D.Liu, J.T.Du, Y.Fu, Y.W.Wu, P.Lei, Y.F.Zhao, Y.M.Li, *Biopolymers* **2006,** *83* 20-31.
[19] M.N.Oda, T.M.Forte, R.O.Ryan, J.C.Voss, *Nat.Struct.Biol.* **2003,** *10* 455-460.
[20] W.L.Hubbell, H.S.Mchaourab, C.Altenbach, M.A.Lietzow, *Structure* **1996,** *4* 779-783.
[21] W.L.Hubbell, C.Altenbach, *Current Opinion in Structural Biology* **1994,** *4* 566-573.
[22] H.S.Mchaourab, M.A.Lietzow, K.Hideg, W.L.Hubbell, *Biochemistry* **1996,** *35* 7692-7704.
[23] H.S.Mchaourab, T.Kalai, K.Hideg, W.L.Hubbell, *Biochemistry* **1999,** *38* 2947-2955.
[24] L.Columbus, W.L.Hubbell, *Trends in Biochemical Sciences* **2002,** *27* 288-295.
[25] W.G.Miller, C.V.Goebel, *Biochemistry* **1968,** *7* 3925-&.
[26] M.Drescher, G.Veldhuis, B.D.van Rooijen, S.Milikisyants, V.Subramaniam, M.Huber, *J.Am.Chem.Soc.* **2008,** *130* 7796-7797.
[27] G.Jeschke, *Chemphyschem* **2002,** *3* 927-932.
[28] G.Jeschke, V.Chechik, P.Ionita, A.Godt, H.Zimmermann, J.Banham, C.R.Timmel, D.Hilger, H.Jung, *Applied Magnetic Resonance* **2006,** *30* 473-498.
[29] J.M.Isas, R.Langen, H.T.Haigler, W.L.Hubbell, *Biochemistry* **2002,** *41* 1464-1473.
[30] G.Jeschke, Y.Polyhach, *Phys.Chem.Chem.Phys.* **2007,** *9* 1895-1910.
[31] M.Robotta, P.Braun, R.B.van, V.Subramaniam, M.Huber, M.Drescher, *Chemphyschem* **2011,** *12* 267-269.
[32] R.Subramaniam, T.Koppal, M.Green, S.Yatin, B.Jordan, J.Drake, D.A.Butterfield, *Neurochemical Research* **1998,** *23* 1403-1410.
[33] H.J.Steinhoff, *Frontiers in Bioscience* **2002,** *7* C97-C110.
[34] R.Fristedt, A.V.Vener, *PLoS One* **2011,** *6* e24565.