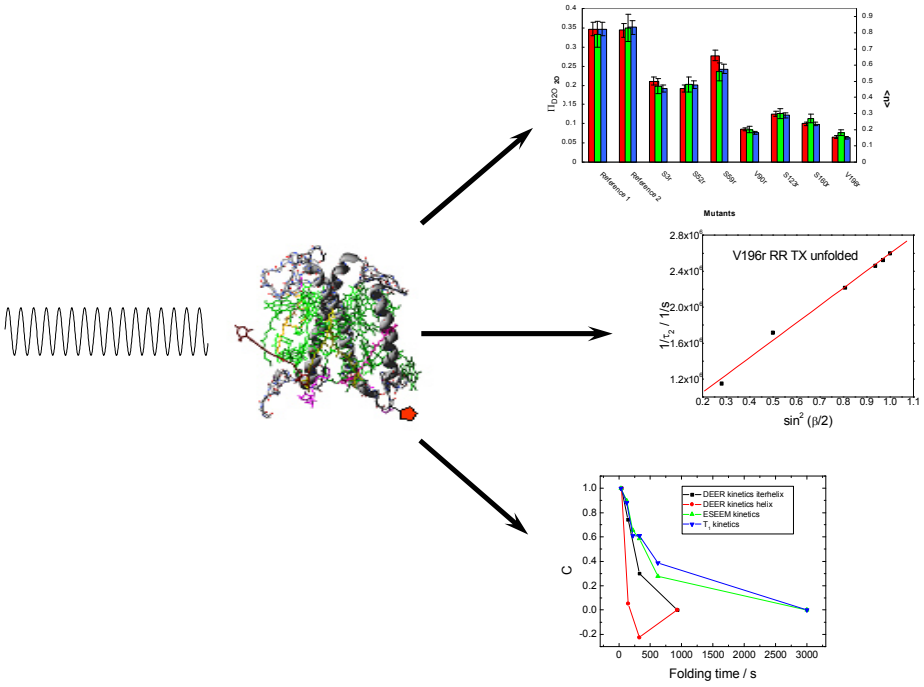
EPR Spectroscopic Investigation of Membrane Protein Structure and Folding on Light Harvesting Complex LHCIIb



Aleksei Viktorovich Volkov

D77 (Dissertation Universität Mainz)

EPR Spectroscopic Investigation of Membrane Protein Structure and Folding on Light Harvesting Complex LHCIIb

Dissertation
zur Erlangung des Grades
“Doktor der Naturwissenschaften”
in Promotionsfach Chemie

am Fachbereich Chemie, Pharmazie und Geowissenschaften
der Johannes Gutenberg-Universität
in Mainz

Aleksei Volkov

geboren in Tallinn

Mainz, den 23.06.2008

Measure what is measurable, and make measurable what is not so.

Galileo Galilei

To my wife

Contents

1. Introduction

Proteins are biological macromolecules that are indispensable for life on our planet. A special type of proteins, membrane proteins, define the functionality of the interface of living cells or cell organelles with their environment and play an important role in cell energetics.

Our knowledge on membrane protein structure is very scarce compared to the one on structure of soluble proteins **[Tamm 2005]**, although membrane proteins are assumed to comprise about 30% of protein reading frames **[Wallin 1998]**. Structure determination for proteins, and in particular membrane proteins, is a challenging task, despite the progress made in the last few years **[Arora 2001]**, **[Lacapere 2007]**. This challenge must be met as protein function can be understood in detail only if structure- and in many cases even structural dynamics- is known. Thus there is a continuing demand for new methods of structure determination that are applicable also to membrane proteins.

Protein folding is the process by which native three-dimensional structure is attained from a less organized state. In the case of soluble proteins this process is driven by the amino acid sequence. In the case of membrane proteins folding is also influenced by the interaction of the protein with the surrounding membrane or with a translocon **[Bowie 2005]**. This makes the understanding of folding for membrane proteins even more complicated than it already is for soluble proteins. Thus our knowledge on folding is also very scarce for membrane proteins compared to soluble proteins **[Tamm2005]**.

Although it is known which interactions determine protein folding, it is not quite clear how exactly this process happens. The question “How does the protein fold?” is not only of academic interest but also of high practical importance. It was shown that not only protein structure, but also unfavorable folding kinetics can be the reason for protein misfolding, and as a consequence cause diseases **[Cheng2008 and references there]**.

Membrane protein folding is assumed to proceed in two stages **[Popot 1990]**. In the first stage hydrophobic α-helices are established across the membrane and in the second stage they interact to form the functional tertiary structure. However there is evidence that the two-stage folding model may not fully represent the folding process and has to be extended **[Engelman 2003]**.

Having detailed experimental data on protein folding can further improve theoretical folding models. This in turn would help to understand the relation between amino acid sequence and three-dimensional structure and finally to design de-novo proteins **[Chen 2008]**. Furthermore protein folding is an example of a perfectly defined self-assembly process. Thus by understanding its principles, better control over the self-assembly process of synthetic materials can be expected.

These considerations show the importance of developing experimental techniques that can provide detailed information about the folding process.

As a model system the membrane protein pigment complex LHCIIb is used in this work. It is the main light harvesting complex of photosystem II of green plants. It consists of a membrane protein and several cofactors, such as chlorophyll a and b, carotenoids and lipids that are non-covalently bound to it. The protein in turn consists of a 232 amino acid polypeptide chain and features three transmembrane helices. **[Kühlbrandt 1994]**, **[Liu 2004]**, **[Standfuss 2005]**

As a crystal structure of the membrane protein complex, except for the N-Terminus of the protein, is available, LHCIIb is a suitable model system for development of structure determination methods. LHCIIb is involved in a number of regulatory processes, which are combined with conformational changes **[Nilsson 1997]**, **[Allen 2001]**, **[Ruban 2007]**. Methods for structure determination developed on LHCIIb in a non-crystalline state can shed light on these changes while a crystal structure is necessarily static. After establishing the techniques on known parts of the structure, the methods could be used to get new information about the unresolved N-terminus, known to be important for its regulatory function **[Hobe 1995], [Allen 2001]**.

LHCIIb has the astonishing property that it can self-assemble from the components in vitro on a time scale of a few minutes **[Plumley 1987]**, **[Paulsen 1990]**, **[Booth 1996]**, **[Reinsberg 2001]**, **[Horn 2002]**, **[Horn 2007]**. This makes LHCIIb also a nice model system to study the general principles of protein folding.

The method used in this work for the study of structure and folding of LHCIIb is EPR spectroscopy. Unlike standard methods used for protein studies, such as X-ray crystallography and NMR spectroscopy, EPR is not capable of obtaining a full set of atomic coordinates of a protein. However membrane proteins do not crystallize easily, which is necessary for X-ray studies and current size limitations of NMR techniques **[Tamm 2006]** exclude many proteins of interest. For EPR investigation no crystals are required and there is no limitation by the size of the protein. The increased sensitivity compared to NMR allows for reducing sample concentration if necessary. Electron cryomicroscopy provides an interesting alternative technique although it usually

requires at least two-dimensional crystals for obtaining highly resolved structures **[Henderson 2004]**. In principle near-atomic resolution can also be achieved with electron cryomicroscopy based on single-particle reconstruction techniques, as was recently demonstrated on a rotavirus particle **[Zhang 2008]**. It remains to be shown whether this technique can be extended to membrane proteins.

Furthermore there is a lack of approaches that provide reliable information on partially ordered structures and structural changes in membrane proteins. Such information is required to characterize folding intermediates and flexible domains that are involved in regulatory processes. Suitable approaches are difficult to develop on the basis of standard techniques as signal assignment to particular sites in the protein molecule fails for broad conformational ensembles for lack of resolution.

Site-directed spin labeling techniques, as they were used in this work, circumvent this problem as the signal originates exclusively from the labelled sites [**Hubbell 2000]**, **[Fanucci 2006]**. This is however also the main disadvantage of the EPR approach to diamagnetic proteins as this results in drastically decreased information content in comparison to the standard methods. Furthermore there is always a risk that covalently bound spin label might influence the native structure. All these considerations also apply to site-directed fluorescence labeling techniques **[Munishkina 2007], [Thóren 2004], [Heuck 2002]**, which have the advantage of higher sensitivity and can be applied even to single molecules *in vivo* **[Wennmalm 2007]**. Compared to such fluorescence labelling, spin labelling has the advantage that the nitroxide spin labels used for EPR are significantly smaller than usual chromophores and thus the probability of distortion of the protein structure by the label is smaller. Furthermore the weak coupling of spins to their environment allows for the separation of interactions by pulsed EPR experiments **[Schweiger 2001]**. Such separation of interactions improves reliability of the interpretation of signals and precision of their quantification.

In this work mostly pulse EPR techniques are used to develop the new approaches for membrane protein structure and folding characterization. To date the vast majority of site-directed spin labeling studies on membrane proteins has been performed with continuous-wave (CW) EPR approaches **[Hubbell 2000]**, **[Fanucci 2006]**. Pulse EPR techniques were applied only very recently and mostly for long-range distance measurements **[Jeschke 2004]**, **[Jeschke 2005]**, **[Xu 2006]**, **[Vamvouka 2008]**. This suggests that the potential of pulse methods for local structure determination is still not used completely. Therefore it is a further aim of this work to explore alternative pulse EPR techniques for studies on membrane proteins.

1. Introduction

Application of pulse EPR techniques to nitroxide-labelled samples requires measurement temperatures of 50 K, where the transversal relaxation times are still long enough to observe the signal. Folding kinetics can thus be studied only if the folding process can be initiated at a well defined time and stopped instantaneously by freezing the sample faster than on the typical time scale of folding. For LHCIIb folding is initiated by mixing of the apoprotein with pigments and lipids and proceeds on a time scale of several minutes that can be easily accessed by freeze-quench techniques.

The methods that are developed on this model system should be applicable to any self-assembling structure if site-directed spin labeling is possible and freeze-quench techniques can be used.

The structure of this work is the following:

In the theory section a brief overview of the EPR spectroscopy in general and the used EPR techniques in particular is given. Also LHCIIb structure and function as well as the protein folding phenomenon are explained.

Then the details of the highly non-trivial protein preparation, as well as the details of EPR experiments and the details of analysis of experimental data are shown.

In the results part first the potential of different EPR methods as well as the effects caused in EPR spectra by the environment of the nitroxide spin-label are investigated. Also distances between different protein positions are studied and compared with a rotamer library based simulation. Afterwards the structural changes of the protein between different micelle compositions are addressed. Different methods are then compared with each other, and advantages and disadvantages of each method are discussed.

The developed arsenal of methods is then applied to study the LHCIIb self-assembly process. As the methods used in this work provide local and global structural information, a rather detailed LHCIIb folding model can be constructed.

The perspectives of the developed approaches are discussed in the outlook section.

2. Theory

2.1. Introduction to EPR spectroscopy

Electron paramagnetic resonance (EPR) is a spectroscopic method for the determination of structure, spatial distribution and dynamics of paramagnetic species **[Schweiger 2001]**.

The paramagnetic species studied during this work are nitroxide radicals that serve as probes for structure and dynamics of selected sites in a membrane protein complex.

The unpaired electrons in these radicals with spin angular momentum **s** have a magnetic moment $\boldsymbol{\mu}$:

$$\boldsymbol{\mu} = -g\beta_e\mathbf{s} \qquad \textit{Eq. 1}$$

Here g is the electron g-value (g=2.0023193043737 for a free electron) and β_e= $9.27400899 \cdot 10^{-24}$ J/T being the Bohr magneton.

In an external magnetic field $\mathbf{B}_0$ with the strength B_0 the energy of the electron depends on the magnetic quantum number m_s with values $\pm$ 1/2 for an electron.

$$E = g\beta_e B_0 m_s \qquad \textit{Eq. 2}$$

The z axis of the spin system is defined as the $\mathbf{B}_0$ field direction. The sign of the magnetic quantum number depends on the orientation of the electron with respect to the external magnetic field.

The energy difference between the two spin states is equal to

$$\Delta E = g\beta_e B_0 \qquad \textit{Eq. 3}$$

It is possible to induce the transition between the two spin states by irradiating the system with electromagnetic radiation of frequency ω_s that will be further referred as the Larmor frequency.

$$\Delta E = g\beta_e B_0 = \hbar\omega_s$$

$$\omega_s = \frac{g\beta_e B_0}{\hbar} \qquad \textit{Eq. 4}$$

$\hbar$ is the Planck constant divided by 2π ($\hbar$ = 1.054571596·10^{-34} J·s). In the case of EPR the Larmor frequency is in the microwave range. During this work two types of frequencies were used: X-band (9 GHz) and W- band (90 GHz).

The most important information during the EPR experiment is the deviation of the sample frequency from the Larmor frequency of the free electron that is caused by the interaction of the spin label with the environment.

The resonance frequencies are described by the spin Hamiltonian of the system.

The Hamiltonian is a formal way of describing the energies of states of the spin system within the electronic ground state. The complete spin Hamiltonian for a single electron spin is shown in equation ***5*** **[Schweiger 2001]**:

$$H_0 = H_{EZ} + H_{ZFS} + H_{HF} + H_{NZ} + H_{NQ} + H_{NN} \qquad \textit{Eq. 5}$$

The terms in the equation ***5*** correspond to the following interactions:

H_{EZ}: Electron Zeeman interaction (interaction of the electron with the external magnetic field). This interaction was already treated in the previous section.

H_{ZFS}: Zero field splitting interaction (interaction of electron spins with each other in spin systems with group spins bigger than 1/2)

H_{HF}: Hyperfine coupling between electron spin and nuclear spin (interaction of electron spins with nuclear spins via electron density and through space)

H_{NZ}: Nuclear Zeeman interaction (interaction of nuclei with the external magnetic field)

H_{NQ}: Nuclear quadrupole interaction (interaction of a non spherical charge distribution of nuclei with spin ≥ 1, with the electric field gradient caused by the nuclei and electrons in the close vicinity)

H_{NN}: Spin-spin interactions between pairs of nuclear spins (through-space interactions between the nuclear spins)

The advantage of such a description of the spin system is that only the spin coordinates have to be taken in account, whereas the spatial degrees of freedom of the wave function are assumed to be constants of motion **[Schweiger 2001]**.

During this thesis only the electron Zeeman interaction, hyperfine interaction, as well as electron-electron interaction are considered. The Hamiltonians of these interactions are thus described more explicitly. They are given in units of angular frequencies, i. e. energies divided by $\hbar$.

Electron Zeeman interaction

This interaction is usually the dominant one for spin systems with spin S=1/2. The Hamiltonian for this interaction is given in equation ***6* [Schweiger 2001]**:

$$H_{EZ} = \beta_e \mathbf{B}_0^T \mathbf{g} \mathbf{S} / \hbar \qquad \textit{Eq. 6}$$

Here $\mathbf{B}_0^T$ is the transposed vector of the external magnetic field (the superscript T of the vector from now on denotes the transposed value), **g** is the electron g-tensor and **S** is the electron spin vector operator.

The g- tensor can be represented as the sum of the g_e value of the free electron times the unitary matrix and the correction term $\Delta\mathbf{g}$ which is a symmetric tensor resulting mostly from the spin orbit coupling and orbital Zeeman interactions.

Hyperfine interaction

This interaction is one of the most important sources of information in EPR, as the nuclear spin environment can be observed via the electron spins. The Hamiltonian for this interaction is given in equation ***7* [Schweiger 2001]**:

$$H_{HF} = \mathbf{S}^T \mathbf{A} \mathbf{I} \qquad \textit{Eq. 7}$$

The interaction tensor **A** can be presented as a sum of two terms (equation ***8*** **[Schweiger 2001]**)

$$\mathbf{A} = A_{iso}\mathbf{1} + \mathbf{A}_{dip} \qquad \textit{Eq. 8}$$

Here A_{iso} is an isotropic term that arises from the finite electron spin density at the position of the nucleus, which causes the coupling. The dipolar term $\mathbf{A}_{dip}$ results from the magnetic coupling between electron and nuclear spins through space. **I** is the nuclear spin vector operator.

Electron-electron interaction

If the interaction of electron spins with each other is strong, it is best described as a zero field interaction of a group spin. However, if the interaction between two unpaired electrons is weak (which was always the case in this study), they are better described by their individual spins $\mathbf{S}_1$ and $\mathbf{S}_2$, an exchange coupling tensor **J**, and a dipole- dipole coupling tensor **D**. The corresponding Hamiltonian for a two spin system, in the case of weak electron-electron interaction is described by equation ***9*** **[Schweiger 2001]**:

$$H_{TOTAL}=H_0(S_1)+ H_0(S_2)+H_{EXCH}+H_{DD} \quad \textit{Eq. 9}$$

$H_0(S_1)$ and $H_0(S_2)$ are Hamiltonians for each individual spin, H_{EXCH} is the exchange coupling Hamiltonian and H_{DD} is the dipole- dipole coupling Hamiltonian.

In principle the physics of electron- electron interaction is very similar to the one of the hyperfine interaction described above.

The exchange coupling between electrons takes place when the singly occupied molecular orbitals of two spins overlap to a significant extent, and as a consequence the electrons are exchanged between these orbitals. This process, for non-delocalized electrons, occurs at distances smaller than 1.5 nm.

Equation ***10*** gives the exchange coupling Hamiltonian **[Schweiger 2001]**:

$$H_{EXCH}= \mathbf{S}_1^{T}\mathbf{J}\mathbf{S}_2 \quad \textit{Eq. 10}$$

The dipole-dipole coupling between electrons takes place when the spin of one electron is sensing the magnetic moment of the other electron through space. It is described by the Hamiltonian in equation ***11*** **[Schweiger 2001]**.

$$H_{DD}= \mathbf{S}_1^{T}\mathbf{D}\mathbf{S}_2 \quad \textit{Eq. 11}$$

In the high field approximation and with neglect of g-tensor anisotropies the dipole-dipole coupling tensor in the principal axis system (coordinate system where the off diagonal elements of the tensor are zero) is given by the equation ***12***:

$$\mathbf{D}= \frac{\mu_0}{4\pi\hbar}\frac{g_1 g_2 \beta_e^2}{r_{12}^3}\begin{pmatrix} -1 & & \\ & -1 & \\ & & 2 \end{pmatrix} = \begin{pmatrix} -\omega_{DD} & & \\ & -\omega_{DD} & \\ & & 2\omega_{DD} \end{pmatrix} \qquad \textit{Eq. 12}$$

2.1.1. CW EPR spectroscopy

CW (continuous wave) EPR spectroscopy is a method, by which the allowed transitions of the spin system are being observed by changing the external magnetic field and keeping the microwave frequency constant.
In the following it will be explained how to predict and interpret the line positions of the nitroxide spectrum in liquid-state CW spectra from the Hamiltonian.
For nitroxides at low concentration the main contributions to the Hamiltonian are the electron Zeeman and the hyperfine interaction with the ^{14}N nucleus.
The interaction Hamiltonian for that is given in equation ***13***.

$$H_0 = \beta_e \mathbf{B}_0^T \mathbf{g} \mathbf{S} / \hbar + \mathbf{S}^T \mathbf{A} \mathbf{I} \qquad \textit{Eq. 13}$$

In the case of liquid-state spectra the **g** tensor averages to the isotropic g value, which is the average of the three principal tensor values **[Schweiger 2001]**. Also the dipolar coupling term of the hyperfine interaction vanishes. Taking the $\mathbf{B}_0$ field direction as z axis of the spin system the Hamiltonian obtains the following form:

$$H = \frac{\beta_e S_z g_{iso} B_0}{\hbar} + A_{iso} S_z I_z \qquad \textit{Eq. 14}$$

Solving the time independent Schrödinger equation with this Hamiltonian results in the energy eigenvalues in frequency units that are given in equation ***15***:

$$\omega(m_s, m_I) = \frac{\beta_e g_{iso} B_0 m_s}{\hbar} + A_{iso} m_s m_I \qquad \textit{Eq. 15}$$

Here m_I is the nuclear spin quantum number (-1, 0, +1 for the ^{14}N nuclei).
Equation ***15*** states, that every electron Zeeman energy level is split in three sublevels, caused by the hyperfine interaction as shown in Figure 1.
The presence of these energy levels can be detected by irradiating the system with microwave of a frequency matching the energy level difference, as already described in section 2.1. According to the selection rules the change of the electron quantum number during the experiment should be ±1. The nuclear quantum number should not change. Thus the frequency of transitions that is also directly seen in the spectrum is given by equation ***16***:

$$\omega = \beta_e g_{iso} B_0/\hbar + A_{iso} m_I \quad \textit{Eq. 16}$$

As the frequency of the transition depends on the field, in the experiment for technical reasons not the frequency is changed until the resonance condition is satisfied and the signal occurs, but the field, as already described in the beginning of this section.
The allowed transitions as well as the energy splitting in an external magnetic field of strength B_0 are shown in Figure 1:

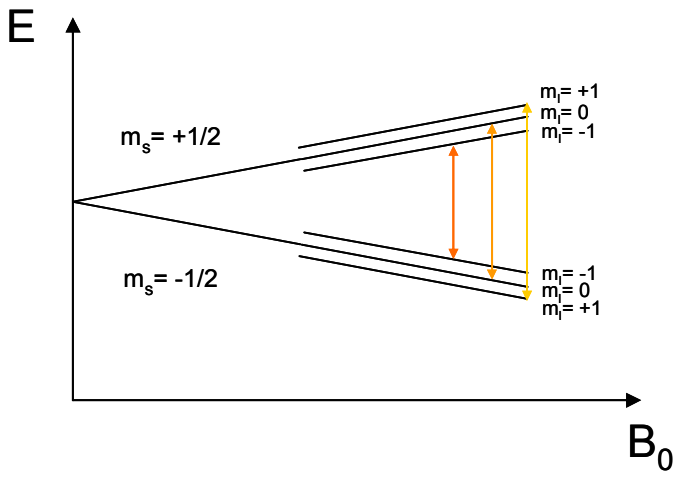


Figure 1: Energy level splittings in a B_0 field according to equation ***15***. The arrows represent the allowed transitions described by equation ***16***.

According to the considerations above, a nitroxide spectrum should consist of three lines. The separation between them is equal to A_{iso} as it is shown in the Figure 2a.

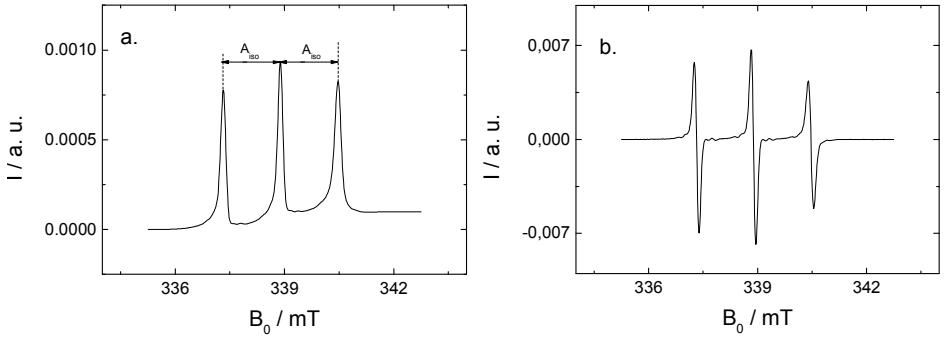


Figure 2: Typical liquid state CW EPR nitroxide spectrum. a. Absorption spectrum, b. Spectrum measured by field modulation. The absorption spectrum is the integrated CW EPR spectrum.

Figure 2 a shows experimental spectra that are in perfect agreement with expectations according to equation ***16***. Due to technical reasons the EPR spectra are usually recorded as a derivative of the absorption spectrum as shown in Figure 2 b. All the CW spectra in this work will be shown only in this differentiated mode.
In Figure 3 a typical solid state CW EPR spectrum of the same sample as in Figure 2 is shown:

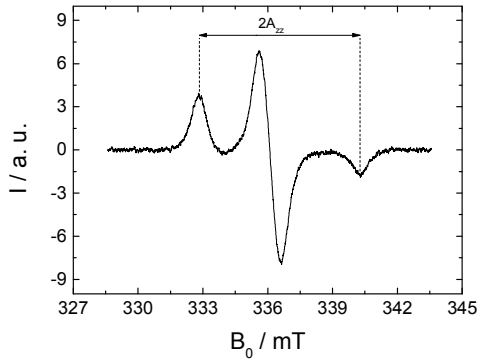


Figure 3: Solid state CW EPR nitroxide spectrum with definition of the $2A_{zz}$ parameter.

Figure 3 shows significant differences compared to Figure 2 b. The reason is that in the solid-state spectrum no averaging of the g tensor takes place, and the dipolar coupling tensor is also present. In order to get a mathematical description of the spectrum, the orientation dependence of the energy splittings of nitroxides has to be taken into

account. Examples for that can be found in **[Schweiger 2001]**. The $2A_{zz}$ parameter, that will be used later for determination of the polarity of the environment is also shown. It was already noted that A_{iso} is dependent on the electron spin density on the nucleus. When the electronic structure of the nitroxide bond is perturbed and the spin density is changed, for example by the change of solvent polarity, the coupling constant changes. The explanation for that is that one of the mesomeric structures according to which the electron is localized on the nitrogen shows charge separation in the N-O bond of the nitroxide, and thus should be stabilized in polar solvents. As the electron spin density in this mesomeric structure is localized in the p_π orbital of nitrogen that is aligned along the z axis of molecular coordinate system, A_{zz} should also increase. The changes can be observed in solution and in the solid state, as A_{iso} and **A** are related **[Owenius 2001]**.

These effects were utilized in this work to determine the polarity change of the spin label environment from the change of the hyperfine coupling.

2.1.1.1. CW progressive power saturation measurements

The CW progressive power saturation measurements differ significantly from simply recording the CW EPR spectrum as described above. In saturation measurements the intensity of the central nitroxide peak as a function of the microwave power is measured, and the shape of the obtained spectrum is disregarded. The microwave power where the intensity of the peak is reduced to half of its unsaturated value is called the saturation power.

This saturation power is directly proportional to the relaxation rate of the spin label, which increases due to collision with other paramagnetic species like O_2 or chromium (III) oxalate (CrOx). This increase is directly proportional to the collision frequency and thus is a measure of relative spin label accessibility to paramagnetic quenchers **[Altenbach 1989]**. As chromium oxalate has good water solubility **[Farahbakhsh 1992]** while oxygen is better soluble in unpolar solvents, and thus has good membrane solubility **[Wilhelm 1973]**, **[Wilhelm 1977]**, **[Windrem 1980]**, it is possible to probe the solvent or membrane accessibility of a spin label by this method **[Altenbach 1989]**.

2.1.2. Pulse EPR spectroscopy

In the CW spectrum the complete spin Hamiltonian is observed. This may be undesirable, as the spectra can get very complicated due to the many interactions present in the sample. Also sometimes it may be desirable to explicitly study one specific interaction in the presence of other large interactions, as it may be the only one carrying relevant information (like electron nuclear interaction in an ESEEM experiment or the electron electron dipolar interaction in a DEER experiment). The modern pulse EPR methods allow for doing this.

During the pulse experiment the electromagnetic field is not continuously fed in the system, but is applied only in certain intervals that are called the pulses.

The pulses change the spin states. In the classical picture this is explained as a change of the angle of the magnetization vector with respect to the external $\mathbf{B}_0$ field. The angle change depends on the $\mathbf{B}_1$ field amplitude of the applied pulse, as well as the length of the pulse. In the quantum mechanical picture a 90° ($\pi/2$) pulse creates spin coherence, whereas a 180° (π) pulse inverts magnetization.

The time between the pulses is called the evolution time. During this time the system evolves under the influence of the Hamiltonian.

To describe the time evolution of the spin, and thus the pulse experiment, the time-dependent Schrödinger equation for kets and bras given in equation ***17*** is used:

$$\frac{d}{dt}\left|\psi\right\rangle = -iH(t)\left|\psi\right\rangle$$

$$\frac{d}{dt}\left\langle\psi\right| = i\left\langle\psi\right|H(t)$$

Eq. 17

In the case of a typical EPR experiment, it is not the single spin we are dealing with, but an ensemble of electron spins that may be in different states. That is why it is not reasonable to use the wave function for its description.

The way to describe such an ensemble of quantum mechanical particles is via the density matrix.

The concept of density matrix can be derived from the wave function $\left|\psi\right\rangle$. The spin of an electron in the magnetic field can be observed in one of the two eigenstates denoted as α and β. These states are associated with the magnetic quantum number

m_s. The wave function for a spin can be represented as a superposition of the two states that are characterized by the wave functions $|\alpha\rangle$ and $|\beta\rangle$ with the complex coefficients c_α and c_β that satisfy the normalization condition and is given in the equation ***18***.

$$|\psi\rangle = c_\alpha|\alpha\rangle + c_\beta|\beta\rangle \qquad \textbf{Eq. 18}$$

The absolute value of the square of these coefficients gives the probability of finding the spin in a particular state.

A special case of a spin ensemble is the one where all the spins are in the same state. Thus coefficients c_α and c_β for the whole ensemble are the same, and the wave function describing this ensemble is just $|\psi\rangle^N$, with N being the number of spins. Such a spin state is called a pure state.

Any ensemble of spins can be described by superposition of n sub ensembles in pure states ($n \leq N$). These pure states are described by the wave functions $|\psi_i\rangle$ and probabilities p_i.

The ensemble average is defined in equation ***19***:

$$\overline{|c_\alpha|^2} = \sum_i p_i \left|c_\alpha^{(i)}\right|^2 \qquad \textbf{Eq. 19}$$

Ensemble averages of other coefficients are computed analogously.

The density matrix, σ for a spin, S=1/2, can be derived from these ensemble averages, and is given in equation ***20***:

$$\sigma = \begin{pmatrix} \overline{|c_\alpha|^2} & \overline{c_\alpha c_\beta^*} \\ \overline{c_\alpha^* c_\beta} & \overline{|c_\beta|^2} \end{pmatrix} \qquad \textbf{Eq. 20}$$

The difference between the diagonal elements is called polarization and is proportional to longitudinal magnetization. The off-diagonal elements quantify coherence. Their magnitude is proportional to the transverse magnetization.

The more general expression for the density matrix of any spin system is given in equation ***21***:

$$\sigma = \sum_{kl} \overline{c_l^* c_k} |l\rangle\langle k| \qquad \textbf{Eq. 21}$$

Here $|l\rangle$ and $|k\rangle$ are the states of the spin system.

The density matrix does not allow for getting the full information on the spin ensemble, i.e. all states of all sub ensembles however it allows to predict and interpret the outcome of the experiments. In other words it contains all the information of the ensemble that is experimentally accessible.

The time evolution of the density matrix is described by the Liouville-von Neumann equation (***Eq. 22***) that can be derived from the time dependent Schrödinger equation.

$$\frac{d}{dt}\sigma(t) = -i(H(t)\sigma(t) - \sigma(t)H(t)) \qquad \textbf{Eq. 22}$$

With the help of this equation it is possible to describe the spin ensemble at every time point in a pulse experiment.

In the special case of a time-independent Hamilton operator it is easy to integrate equation ***22*** to obtain the density matrix at the time point t.

$$\sigma(t)=\exp(-iHt)\sigma(0)\exp(iHt)= U(t)\sigma(0)U(t)^{\dagger} \qquad \textbf{Eq. 23}$$

The factor exp(-iHt)= U(t) in the equation ***23*** is called a propagator, as it propagates the density matrix in time.

If the Hamiltonian is time-dependent, the integration of the Liouville-von Neumann equation is not as straightforward as in the former case. Different approaches are however possible for example by using a Dyson time-ordering operator or subdividing the evolution of the spin system into sufficiently small intervals, during which the Hamiltonian is constant.

The expectation value of the observable O from the density matrix can be determined using the equation ***24***:

$$\langle O\rangle = \mathrm{tr}\{\hat{O}\sigma\} \quad \textbf{Eq. 24}$$

Here $\hat{O}$ is an operator for this observable.

Thus it is possible to determine the outcome of any pulse experiment at any time from the considerations described above.

Equation ***22*** neglects relaxation. In order to get a more general description of a quantum mechanical system, a relaxation superoperator Ξ has to be introduced. The result is given in equation ***25***.

$$\frac{d}{dt}\sigma(t) = -i(H(t)\sigma(t) - \sigma(t)H(t)) - \Xi(\sigma(t) - \sigma_{eq}) \quad \textit{Eq. 25}$$

Here σ_{eq} is the density operator for the system in thermal equilibrium. In the description of pulse sequences the relaxation effects are usually neglected. A more detailed description of relaxation phenomena is given in section 2.1.2.3.

In the following sections the most important pulse experiments used during this work will be described.

2.1.2.1 Electron spin echo (ESE) experiment

ESE spectroscopy is a pulse EPR method where the intensity of the electron Hahn echo is observed as a function of the field position. The reason for doing this is that due to the hardware limitations of EPR spectrometers it is not possible to excite the whole spectrum at once. The information obtained during this experiment is similar to the one from CW EPR, with the only difference that relaxation effects may additionally influence the spectrum. This additional information allows to obtain the orientation dependence of the relaxation **[Millhauser 1984]**, **[Rohrer 1996]** as different positions of the spectra represent different orientations, as well as information about spin label dynamics in the sample [**Dzuba 1996]**, **[Kirilina 2001]**, **[Kirilina 2004]**, **[Leporini 2003]**.

The pulse sequence of the ESE experiment as well as a typical nitroxide solid state EPR spectrum are given in Figure 4:

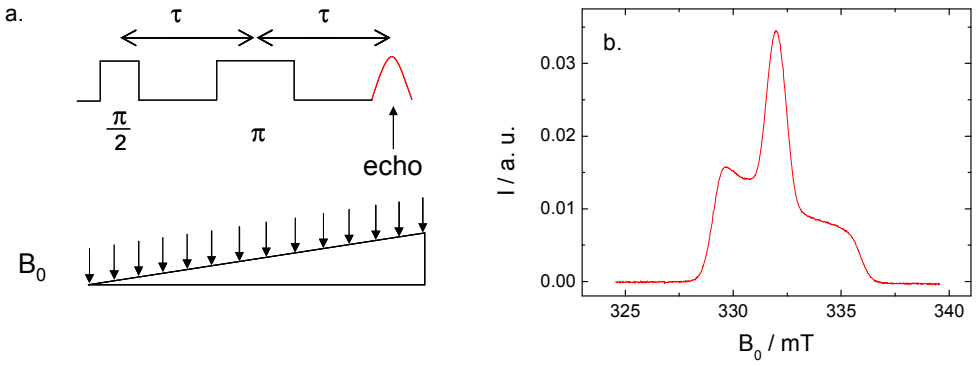


Figure 4: a. ESE pulse sequence. The arrows symbolize the detection points as a function of the field. b. Typical ESE spectrum.

During this work the ESE spectra were used to calibrate the resonance field position, obtain information about nitroxide dynamics in the protein and to check the protein folding quality.

2.1.2.2. ESEEM experiment

The ESEEM (Electron Spin Echo Envelope Modulation) effect is a phenomenon where the electron spin echo is modulated with the resonance frequencies of nuclear spins coupled to it. The physical interaction causing this effect is described by the hyperfine Hamiltonian introduced in section 2.1.

The explanation for the ESEEM effect is based upon the excitation of forbidden nuclear transitions with the microwave pulse due to mixing of levels with different nuclear magnetic quantum number by the part of the anisotropic hyperfine interaction that is perpendicular to the magnetic field. A detailed description of the ESEEM effect can be found in **[Berliner 1984]** and **[Schweiger 2001]**.

The three-pulse ESEEM experiment is a stimulated echo based pulse EPR method that allows to measure nuclear frequencies via exciting and observing only electron spins. Information about the distance and number of the nuclei coupled to the electron can be obtained.

The three-pulse ESEEM experiment proved to be powerful in solving problems of structural biology. Determination of the position of the N terminus of small proteins in the membrane **[Carmieli 2006]**, the location of a small protein in the membrane **[Salnikov 2006]** and water accessibility of a fatty acid in the protein binding pocket

[De Simeone 2007] were already reported as applications of ESEEM. Also work was done on profiling the water penetration in micelles **[Szajdzinska 1984]** and along the membrane **[Bartucci 2003]**, **[Noethig 2004]**, **[Erilov 2005]**, **[Bartucci 2006].**

The pulse sequence of the ESEEM experiment and a typical ESEEM signal are given in Figure 5:

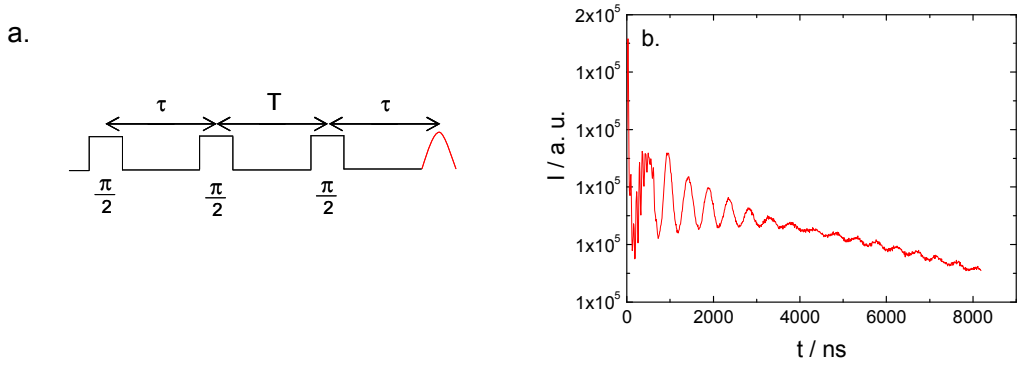


Figure 5: a. Three-pulse ESEEM sequence. b. Typical ESEEM signal with pronounced deuterium modulation.

The shape of the signal without the relaxation background is given by equation ***26***:

$$I(t)=1-4/3I(I+1)k[1/2((1-\cos(\omega_\alpha(T+\tau))(1-\cos\omega_\beta\tau)] \qquad \textit{Eq. 26}$$

Here I is the nuclear spin quantum number, ω_α and ω_β are the nuclear transition frequencies and k is the modulation depth parameter, which is proportional to the inverse sixth power of electron nuclear distance. Equation ***26*** shows what information can be extracted from ESEEM data. This analytical expression is best valid for small modulation depth **[Ponti 1997],** which was not always the case for the data measured during this work.

As the modulation depth is proportional to the inverse six power of the electron nuclear distance, different methods in the time and frequency domain of the signal were developed to extract it **[Ichikawa 1979]**, **[Schweiger 2001]**, **[Bartucci 2003],** [**Carmieli 2006**].

In this work the analysis of time-domain ESEEM data is performed with the program described in **[Hinderberger 2004]** that allows determining the number and distance to the coupled nuclei. Care is however needed, as the analytical equations used there to obtain the number and the distance to the coupled nuclei are best valid for small

modulation depth and thus some error in the obtained distance and number of coupled nuclei is possible.
Also new ways of analyzing the data in time and frequency domain that are model independent and thus applicable to any modulation depth are used in this work, and are described in detail in the sections 3.2.2.4. and 4.1.3.

2.1.2.3. Electron relaxation and experiments to measure relaxation

Relaxation is a process of recovery from a non-equilibrium state to equilibrium. In the case of EPR, due to the introduction of the external magnetic field along the z-direction, two relaxation processes are distinguished. Namely the recovery of spin polarization (longitudinal relaxation) with the characteristic time constant of this process called longitudinal relaxation time (T_1) and the loss of spin coherence (transversal relaxation) with the characteristic time constant of this process called transversal relaxation time (T_2).
T_1 describes the recovery of spin polarization to a steady value that is proportional to the magnetic susceptibility of the sample. This process can be traced back to the non-radiative energy transfer from the spin system to the internal degrees of freedom of the environment. Such transfer is possible due to coupling to the internal degrees of freedom of the environment. The spontaneous emission does not play a role at the frequencies studied, as the probability for this process scales with ω^3 and its rate is negligible for the frequencies relevant in EPR.
T_2 describes the loss of spin coherence perpendicular to the z-direction. It can be any process that takes the spins off resonance. The process does not change the energy of the system. The origin of T_2 is in the variation of the relative energy levels of the spin system. In solids T_2 is often not well defined and empirical parameters like T_m, that are associated with the decay of the primary echo, can be defined. In this work still the term T_2 is used to describe these processes in order to be consistent with the presented theory.
The T_1 and T_2 relaxation times are related to each other, as they both depend on the interaction of the spin system with the time-dependent stochastic magnetic or electric fields of the environment that usually originate from thermal motions. The molecular motion should have a time scale similar to the Larmor frequency. In the case of T_2 an additional motion at very long time scales influences the relaxation. That is why the relaxation times are usually not equal: the processes that influence them happen on different time scales.

As the time-dependent fields that cause the relaxation have a stochastic origin it is reasonable to characterize them via an autocorrelation function G(τ). The function G(τ) is defined as the average of the stochastically fluctuating B-field of the environment at time t times the field value at time t+τ as shown in equation ***27*** **[Carrington 1967]**

$$G(\tau) = \overline{B(t)B(t+\tau)} \quad \textit{Eq. 27}$$

Although the fields are random, and their individual average values are zero, the autocorrelation function has a finite value.
The autocorrelation function can be described by the exponential function shown in equation ***28***:

$$G(\tau) = \overline{B^2}\exp\left(-\frac{|\tau|}{\tau_c}\right) \qquad \textit{Eq. 28}$$

$\overline{B^2}$ is the mean square amplitude of the fluctuating field. τ_c is the correlation time of fluctuations (approximately how long it takes before field changes sign).
The spectral density function J(ω) is twice the Fourier transform of the autocorrelation function and is given in equation ***29***:

$$J(\omega) = 2\int_0^\infty G(\tau)\exp(i\omega t)d\tau \qquad \textit{Eq. 29}$$

For the exponentially decaying function it obtains the form given in equation ***30***:

$$J(\omega) = 2\overline{B^2}\frac{\tau_c}{1+\omega^2\tau_c^2} \qquad \textit{Eq. 30}$$

Under the assumption that the motions are fast enough to average out the part of the anisotropic interaction they modulate, T_1 is described by the formula presented in equation ***31*** **[Schweiger 2001]**:

$$T_1 = \frac{1}{\left(\frac{g\beta_e}{\hbar}\right)^2 \left(\overline{B_x^2} + \overline{B_y^2}\right) \frac{\tau_c}{1+\omega^2\tau_c^2}} \qquad \textit{Eq. 31}$$

$\overline{B_x^2}$ and $\overline{B_y^2}$ are the mean square amplitudes of the fluctuating field along the x and y directions. Equation ***31*** is very similar to the reciprocal of equation ***30***, showing the direct connection between stochastically fluctuating fields and relaxation effects. Equation ***31*** also shows that fields perpendicular to the z- axis are effective for T_1 relaxation.

Usually not one, but many processes influence relaxation. The overall relaxation rate is then the reciprocal sum of all contributions (equation ***32***)

$$\frac{1}{T_{overall}} = \sum_i \frac{1}{T_i} \qquad \textit{Eq. 32}$$

The processes that can cause a spin flip must be capable of absorbing the same amount of energy as the spin system looses by the transition. One way for the system to do that is to absorb a phonon (vibration quant) of the same frequency as the energy difference between the two spin levels. This type of relaxation is called direct process. Another way for the spin system to loose the energy is to absorb a phonon with a larger frequency than that of a spin state. The spin system goes then into a virtual level. The transition to the spin ground state is accompanied by the reemission of the phonon with a frequency that equals the initial frequency plus the frequency of the virtual level. This process is called the Raman process. From the considerations above it is seen that such kind of process is nearly independent on the resonance frequency. It may be more efficient than the direct process as phonon densities are much larger at the higher energies involved in the Raman process than at the low energies of electron spin transitions.

As was stressed earlier, in the case of T_2 relaxation no energy exchange with the environment takes place. The T_2 signal dephasing is typically associated with the double spin flip processes that destroy the correlations between the involved spins. As single spin flips also destroy correlations the formula for T_2 is given by the equation ***33*** (same approximation as in equation ***31***) **[Schweiger 2001]**:

$$T_2 = \frac{1}{\left(\frac{g\beta_e}{\hbar}\right)^2 \left\{\frac{1}{2}\left(\overline{B_x^2} + \overline{B_y^2}\right)\frac{\tau_c}{1+\omega^2\tau_c^2} + \overline{B_Z^2}\tau_C\right\}} \qquad \textit{Eq. 33}$$

The first term in the second factor of the denominator originates from the same spin flips that are responsible for longitudinal relaxation T_1 whereas the second part describes the double flipping process. Here $\overline{B_z^2}$ is the mean square amplitude of the fluctuating field along the z direction. This means, that the processes influencing T_2 relaxation have to fluctuate perpendicular to the xy plane.

The processes that influence the relaxation in nitroxides are discussed in combination with the experimental results in the section 4.1.4.

For a more comprehensive description of relaxation phenomena, the reader is referred to references **[Carrington 1967]**, **[Berliner 2000]**, **[Schweiger 2001]**.

The experiment used to measure the T_1 relaxation time is the inversion recovery experiment. The pulse sequence and the measured signal are shown in Figure 6:

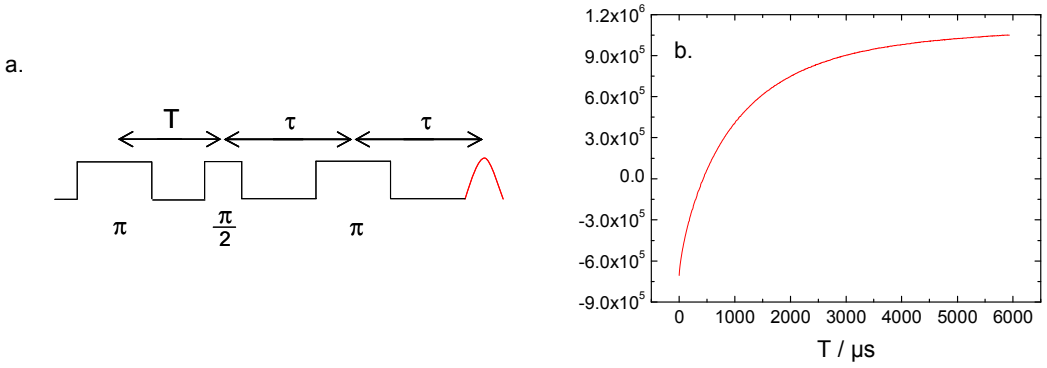


Figure 6: a. Inversion recovery pulse sequence. b. Obtained experimental signal.

The principle of the experiment is based on the inversion of the magnetization with a π pulse with subsequent signal detection via the Hahn echo. As the time T between the inversion pulse and detection echo is incremented, it is possible to observe the recovery of the spin system. The obtained curve is then fitted (in an ideal case) by equation ***34***:

$$I(t) = I_0\left[1 - 2\exp\left(\frac{T}{T_1}\right)\right] \qquad \textit{Eq. 34}$$

The experiment used to measure the T_2 relaxation time is the Hahn echo. The pulse sequence and the measured signal are shown in Figure 7:

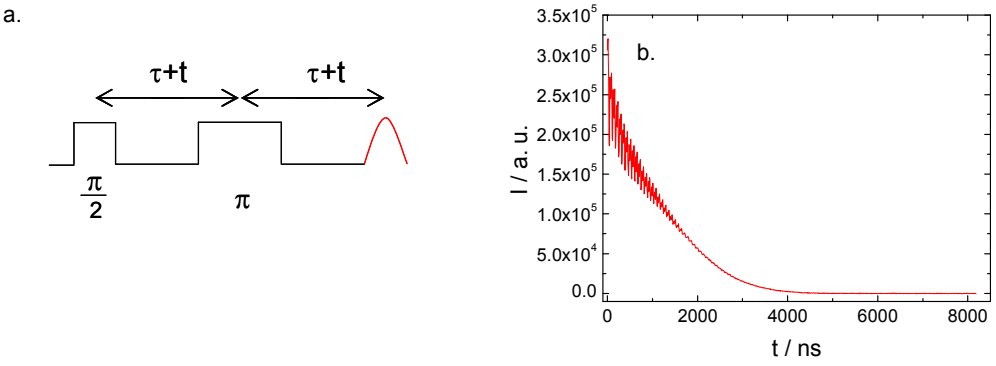


Figure 7: a. Hahn echo pulse sequence. b. Obtained experimental signal.

The experiment is based on incrementation of the interpulse delay and detection of the echo as a function of this delay. The non-random field inhomogeneities are refocused, whereas the stochastic processes cause the signal loss.

The ESEEM modulation on the decay of the echo, that can introduce errors in T_2 determination, is clearly seen. The curve from Figure 7 b. is fitted (in an ideal case) by equation ***35***:

$$I(t) = I_0 \cdot \exp\left(-\frac{2t}{T_2}\right) \quad \textit{Eq. 35}$$

Usually the measured relaxation curves could not be fitted by a single exponential, which is a hint for a distribution of relaxation times. The effective relaxation times are then being determined and the best fit values are taken. Exact determination is shown for every particular sample.

2.1.2.4. The DEER experiment as a tool for distance measurements

Double electron-electron resonance (DEER) is a pulse EPR method that allows for measuring the distance distribution between two electron spins in the range from 1.5 to 8 nm **[Jeschke 2002]**. The pulse sequence was introduced by **[Milov 1984 and references there]** and then improved by **[Pannier 2000]**.

There are alternative EPR methods for electron-electron distance determination like double-quantum EPR **[Borbat 1999]** and single-frequency methods like SIFTER **[Jeschke 2000]**. DEER is still considered to be the method of choice as it is technically easier to perform than double-quantum EPR **[Schweiger 2001]** and in comparison to SIFTER the data analysis procedure is easier to perform and more reliable **[Jeschke, G. Personal communication]**.

During the years DEER developed into a powerful method for determining the local structure in different sorts of complex materials, like copolymer structures **[Pannier 2001]**, nanocomposites **[Mao 2008]** and biomaterials **[Jeschke 2007 and references there]**, **[Schiemann 2007]**.

DEER does not only allow for the measurement of the interspin distance. Also the number of coupled spins can be determined from the modulation depth of the DEER signal **[Milov 1984], [Schweiger 2001], [Hilger 2005]**, **[Bode 2007]**. This additional information even allowed determining the oligomerisation state of a membrane protein **[Hilger 2005]**.

The application of DEER to study structural changes, as it was done in this work and in **[Ionita 2008]**, seems to be a very promising field of study.

The physics of the DEER experiment is based on the dipole-dipole Hamiltonian introduced in section 2.1. Based on the theoretical considerations introduced previously the DEER echo amplitude $I_{DEER}^{intramol}(t)$ under neglect of non-secular (not conserving Zeeman energy **[Antonov 1984]**) parts of electron-electron coupling and using the point-dipole approximation for electron spins is given by equation ***36***:

$$I_{DEER}^{intamol}(t) = \cos(\omega_{ee} t) \qquad \textit{Eq. 36}$$

$$\omega_{ee} = \omega_{dd} + J = \frac{\mu_0 g_a g_b \beta_e^2}{4\pi\hbar} \frac{1}{r^3} (3\cos^2\theta - 1) + J \qquad \textit{Eq. 37}$$

Here t is the time (see Figure 9 for details), ω_{dd} is the dipole-dipole coupling frequency, μ_0 is the permeability of vacuum ($4\pi \cdot 10^{-7}$ NA^{-2}), θ is the angle between the axis connecting the two electrons and the static field $\mathbf{B}_0$ and J is an exchange coupling

parameter that can be neglected at interspin distances larger than 1.5 nm in non-conducting matrices and in the absence of a fully conjugated pathway between radical centers.

In the case of a random distribution of spin pairs with respect to $\mathbf{B}_0$ and no orientation selection by the microwave pulses, the Fourier transform of $I_{DEER}^{intramol}(t)$ results in a Pake pattern **[Schweiger 2001]**. The different orientations contribute with weighting of $\sin\theta$ and thus the Pake pattern has the highest intensity at $\theta = \pi/2$ **[Jeschke 2002b]**. In the case of a real DEER experiment the obtained spectrum is not exactly the Pake pattern due to the limited excitation bandwidth of the microwave pulses. Still from the singularities of this spectrum the dipolar frequency and thus information about the interspin distance can be obtained using equation ***37*** **[Pannier 2000]**. A typical Fourier transformed (background corrected) dipolar spectrum obtained in a DEER experiment is shown in Figure 8:

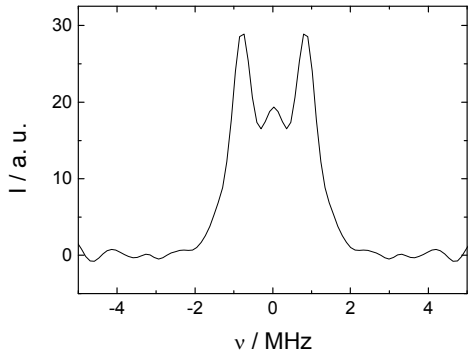


Figure 8: Typical Fourier transformed DEER spectrum. From the position of the singularities the electron-electron distance can be determined. In the case of more than one distance in the sample the Pake pattern consists of the sum of Pake patterns corresponding to different distances.

The necessity of a background correction arises from the following complication: in the case of the real samples there is also the contribution from the coupling to spins in other molecules that are homogeneously distributed in the sample. The contribution from these spins to the signal is given by equation ***38***:

$$I_{DEER}^{intermol}(t) = \exp(-\kappa c F_b t) \qquad \textit{Eq. 38}$$

Parameter κ is given by equation ***39***:

$$\kappa = \frac{2\pi\mu_0\beta_e^2 g_a g_b}{9\sqrt{3}\hbar} \qquad \textit{Eq. 39}$$

Here c is the volume spin concentration and F_b is the fraction of B spins excited by the pump pulse. **[Schweger 2001]**

The measured signal is a product of the intermolecular and intramolecular contribution given in equation ***40***:

$$I_{DEER}^{measured}(t) = I_{DEER}^{intramol}(t) \cdot I_{DEER}^{intermol}(t) \qquad \textit{Eq. 40}$$

In order to obtain only the intramolecular information the measured signal should be divided by $I_{DEER}^{intermol}(t)$. This procedure is performed routinely with the "DEERAnalysis2006" program, as will be shown in detail later. A good separation between these two signal components is possible only when the distances within the studied object are smaller than the distances between these objects **[Jeschke 2006]**.

In order to get a distance distribution from the DEER data more sophisticated analysis has to be performed than just the determination of the maximum of the Pake pattern. The obtained time-domain data $I_{DEER}^{intramol}(t)$ can be simulated with the analytical function given in tequation ***41***:

$$I_{DEER}^{intramol}(t) = 1 - \int_0^{\frac{\pi}{2}} \sin\theta\lambda(\theta)\left[1 - \cos(\omega_{ee}(\theta)t)\right] d\theta \qquad \textit{Eq. 41}$$

This function differs from the one shown in equation ***36***, as it also takes into account experimental imperfections. Partial excitation can be considered by the orientation dependent parameter $\lambda(\theta)$ that determines the modulation depth of the signal **[Schweiger 2001]**. In "DeerAnalysis2006" the data is fitted by a similar function as in equation ***41***, however the modulation depth is assumed to be orientation independent. Before the Tikhonov regularization procedure is performed the signal is renormalized, so that the dependence on the modulation depth disappears.

The determination of the distance distribution from the experimental DEER data is an ill-posed problem, meaning that similar dipolar evolution functions can produce strongly different distance distributions **[Jeschke 2006]**. The current approach for obtaining a distance distribution from DEER data is based on the Tikhonov regularization procedure, a mathematical method that searches for the best compromise between smoothness and least mean square deviation between the simulated curve and the measured dipolar data. The regularization parameter α is a measure for the smoothness of the fit. The optimal regularization parameter can be determined from the L-curve as described in **[Jeschke 2006]**. This approach is utilized in "DEERanalysis2006" program and was also used in data analysis during this work.
The pulse sequence **[Pannier 2000]** and the positions of the pulses concerning the ESE spectrum according to **[Jeschke 2006b]** are shown in Figure 9:

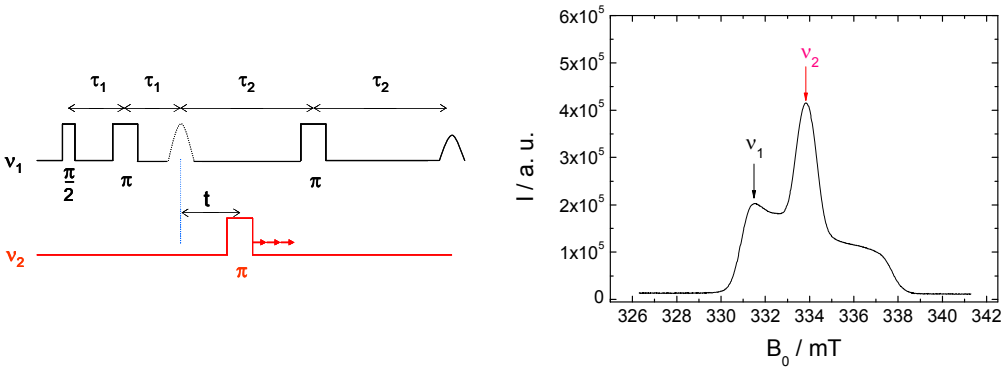


Figure 9: DEER pulse sequence on the left, and the positions of the pump pulse (red arrow) and observer pulse (black arrow) in the ESE spectrum on the right.

The exact experimental details for the DEER experiment as well as the data analysis procedure are given in section 3.2.

2.2. LHCII Structure and Function

LHCII is a complex involving one of the most abundant membrane proteins on earth. It is the main light harvesting complex of photosystem II. Its main function is to transport the light energy of the sun to photosystem II where it is converted to chemical energy.
LHCII consists of a membrane protein and several cofactors, such as chlorophyll a and b, carotenoids and lipids that are non-covalently bound to it. The chlorophylls allow the energy transport in form of excitons in the pigment protein complex. The carotenoids stabilize the protein pigment complex **[Plumley 1987]**, allow the light absorption in a broader spectral region **[Kühlbrandt 1994]**, **[Liu 2004]** and protect the complex from photooxidation **[Kühlbrandt 1994]**, **[Young 1996]**.
The protein in turn consists of a 232 amino acid polypeptide chain and features three transmembrane helices **[Kühlbrandt 1994]**, **[Liu 2004]**, **[Standfuss 2005]**. The basic structural and functional LHCII unit is a trimer **[Liu 2004]**.
The monomeric and trimeric structures of LHCII are visualized in Figure 10.

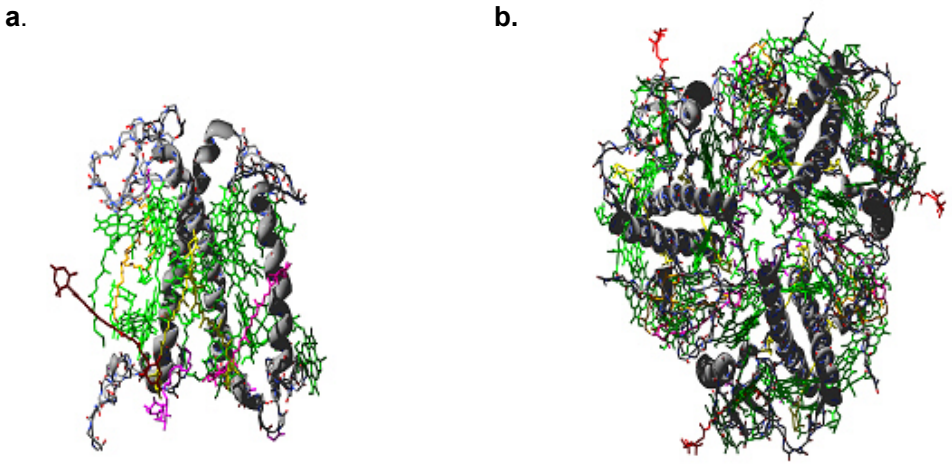


Figure 10: LHCIIb X-ray structure **[Standfuss 2005]**. a. Side view of the monomer. b. Top-view of the trimer.
The carbons of the protein backbone are colored gray, the nitrogens blue and the oxygens red. The secondary structure elements are also shown in gray. Chlorophyll a is shown in light green and chlorophyll b in dark green. The carotenoids are colored in yellow (lutein), red (neoxantin) and violet (violaxantin). The lipids are colored in orange (phosphatidyl glycerol, PG) and magenta (digalactosyldiacylglycerol, DGDG).

Apart from light harvesting, LHCII is also involved in a number of regulatory processes, which are combined with conformational changes **[Nilsson 1997]**, **[Allen 2001]**, **[Ruban 2007]**.

LHCII can also self-assemble from the components in vitro on a time scale of a few minutes **[Plumley 1987]**, **[Paulsen 1990]**, **[Booth 1996]**, **[Reinsberg 2001]**, **[Horn 2002]**, **[Horn 2007]**, which makes this protein a very convenient system for the study of the folding process of membrane proteins.

There are a few slightly different proteins in native LHCII complexes. In this work the LHCb1 protein was used. Its complex with pigments and lipids will be further referred to as LHCIIb.

2.3. Protein folding

Protein folding is the process of acquisition of the three-dimensional structure of a protein from a disordered state.

The main driving force of this process is considered to be the tendency of the polar amino acids to be solvent exposed, and of the unpolar amino acids to be shielded from the polar solvent. The details of this process are however much more complicated, as there are many additional forces, like disulfide bonds, ionic bonds, and hydrogen bonds that are responsible for the protein stability. As only the amino acid sequence is encoded in the DNA, it is astonishing that also the three- dimensional structure of proteins is uniquely determined by it.

If protein folding would be a totally random process, according to Levinthal the folding of a 100 amino acid long chain would take time significantly longer than the age of the universe. This contradicts the common observation that most of the proteins in the living cell are in the correct folded state. From that it is concluded that proteins fold upon certain folding pathways. The protein folding funnel concept proved to be useful in understanding of how proteins fold. **[Voet 2004]**

Membrane protein folding is more complicated than that of soluble proteins. While the fold of a soluble protein is normally completely defined by the amino acid sequence, in the case of a membrane protein the bilayer as well as the translocon also influence its folding **[Bowie 2005]**.

Our current knowledge on the folding process of membrane proteins is much scarcer than for soluble proteins. The reason for that is that the structure determination of membrane proteins, their handling and even their denaturation is difficult. The

additional lipid component makes the investigation of membrane protein folding even more complicated. **[Buchner 2005]**

A simplified idea of how membrane proteins fold is given by the two-stage folding model. According to it the first stage is the formation of independently stable transmembrane helices. The second stage is the interaction between these helices to form the tertiary fold of the polypeptide. **[Popot 1990]**

As the two-stage folding is an oversimplification of the real folding process **[Buchner 2005]** it is of interest to test whether, and to which extent this model is applicable to LHCIIb. This may eventually help to develop new membrane protein folding models and thus to improve our understanding of membrane protein folding.

The study of protein folding is of high importance, as unfolded or wrongly folded proteins are the reason for some diseases, with Alzheimer and Parkinson among them **[Santucci 2008], [Chen 2008]**.

The information obtained from studies of protein folding also can be used for better understanding of self assembly processes in general.

The phenomenon of self assembly is of enormous importance for supramolecular chemistry **[Steed 2000]**. A deeper understanding of self assembly is expected to be beneficial to many fields of chemistry **[Lindsey 1991]**.

3. Experimental Part

3.1. Protein preparation

3.1.1. LHCIIb monomers

Several mutant versions of the Lhcb1*2 (AB80) gene **[Cashmore 1984]** from pea (pisum sativum) that contain a single or double cysteine were constructed by replacing serine or valine at different protein sites. The DNA sequence was checked by sequencing (Genterprise). The mutation positions as well as schematic picture of LHCIIb are shown in Figure 11 a.

a. b.

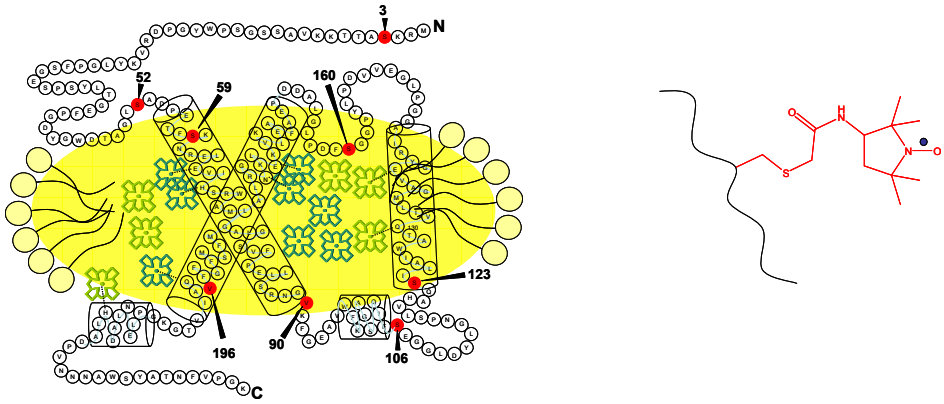


Figure 11: a. LHCIIb investigated in this work. Each circle with a letter represents an amino acid. The red circles with numbers show the mutation positions where the spin labels are attached. Cylinders represent the protein helices. The chlorophylls are schematically shown as green porphyrin macrocycle cartoons. Other pigments and lipids are not shown. The yellow oval around the protein represents the micelle as expected from the crystal structure. **[Kühlbrandt 1994]**, **[Liu 2004]**, **[Standfuss 2005]** The figure is adapted from **[Dockter 2005]**. b. Nitroxide spin label attached to inserted cysteine. The polypeptide chain is symbolized by a wavy line.

Protein overexpression in *Escherichia coli* was performed as described previously in **[Paulsen 1990]**. Some of the Lhcb1 mutants were available from former members of the Prof. Dr. Paulsen group **[Bender 2004]**.

The purified apoproteins were dissolved (1mg/ml) in an aqueous solution of 0.5 % LDS, 20 mM sodium phosphate (pH 7), and 2 mM tris-(2-cyanoethyl)phosphine (1 M in DMF) and were incubated 2 h at 37 °C. The proteins were then spin labelled on cysteine by adding 3-(2-Iodoacetamido)-2,2,5,5-tetramethyl-1-pyrrolidinyloxyl (PROXYL IAA) (10 mg/ml DMSO solution (Sigma) 20 fold molar excess over protein) and incubated overnight at 37 °C on a shaker. Labelling efficiency was determined to be at least 90 %. The structure of the spin labelled position is shown in Figure 11 b.

Spin labelled proteins were precipitated by adding 5 % trichloracetic acid at room temperature and immediate centrifugation (12000 g for 5 min at 4 °C). This step has to be performed fast, as reduction of the spin label can occur **[Kroll 1999]**. The protein pellet was washed 5 times with distilled water and dried for 15 min at ambient temperature.

The spin labelled protein was reconstituted with pigment extract from pea thylakoids **[Paulsen 1993]**, isolated as described in **[Paulsen 1990]**, to self-assemble into the biologically relevant light-harvesting complex LHCIIb.

In the following spin-labelled mutants are abbreviated in the form S52r for a serine (S) at position 52 that was mutated to a cysteine and then labelled by PROXYL IAA radical (r). In the case when the valine was mutated to cysteine, the “V” is written before the mutation position number. In the case of double mutants the mutation positions are given by the slash. For example V90r/V196r mutant means that the valines in positions 90 and 196 were replaced by the cysteines that were further spin labelled. In all mutants the native single cysteine at position 79 was replaced by serine. Thus the mutant without cystein and no further modifications is referred as C79S. The measurements performed on the different spin labelled mutants are assumed to be representative also for this protein part where the spin label is attached to, as the spin label is covalently bound and small. That is why by comparison of different properties of protein parts only the mutation positions are mentioned (eg. S3 and not S3r). As usual the numeration begins at the N- terminus.

The samples used during this thesis are of two different types. One type is a fully spectroscopically characterized LHCIIb without any unfolded or incorrectly folded protein purified by an ultracentrifugation procedure which will be called **UZ** sample. These samples were used for structural investigations and method development.

There will be two types of UZ samples discussed in this work. One type contains n-dodecyl-β-D-maltoside and will be referred to as **LM** samples. The other type of UZ samples contains Triton X 100 detergent and will be further called **TX** samples.
If the sample type is not specified, it is a UZ LM composition.
The **UZ LM** samples are prepared as follows:
Monomeric reconstituted LHCIIb was purified by ultracentrifugation on 0.1 to 1 M sucrose density gradients containing 0.1 % (w/v) n-Dodecyl-ß-D-maltoside and 5 mM Tris- HCl (pH 7.8). After spinning for 16 h at 230000 g at 4 °C, the band containing monomeric LHCIIb was collected and concentrated by Amicon ultra centrifugal filter units (30 kDa, Millipore) up to ca. 600 μM.
The samples for the saturation measurements were prepared in the same way, with the only difference that to some of them chromium (III) oxalate (CrOx) was added to a final concentration of 10 mM.
The prepared protein complexes were checked by fluorescence and CD spectroscopy, in order to exclude possible effects of structural influence of the mutations and spin labeling on the native structure.
The fluorescence measurements were performed on Fluoromax- 2 fluorescence spectrometer (ISA SPEX / Yobin Yvon) with Ministat Compatible Control (Huber Kältemaschinenbau) cooling system. The sample was introduced in the cell with the width of 5 mm. “Front Face” measurement technique was used. The temperature during the measurement was 20 °C.
The fluorescence spectra of all the used mutants showed a characteristic peak of chlorophyll a fluorescence at 680 nm for excitation of chlorophyll b at 470 nm, proving excitation transfer and thus the intactness of LHCIIb **[Carpentier 2004]**. The typical fluorescence spectrum of LHCIIb monomers is shown in Figure 12.

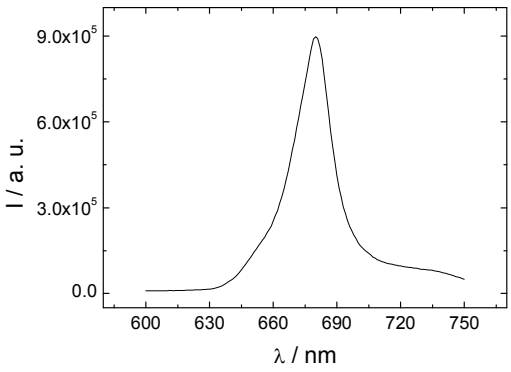


Figure 12: Fluorescence spectrum of an LHCIIb UZ LM sample.

The CD measurements were performed on a J-810-S (Jasco Labor- und Datentechnik GmbH) spectrometer. The wavelength range was between 400 and 750 nm. The scan speed was 50 nm/min. The temperature during the measurement was 4 °C.
Each CD spectrum was corrected by subtraction of the sucrose CD data from it.
The CD spectra were used as a fingerprint and showed characteristic peaks at 491 nm, 649 nm and 680 nm for all the used mutants (Figure 13) and thus prove the similarity of the obtained reconstituted mutant with native LHCII **[Hobe1994]**.

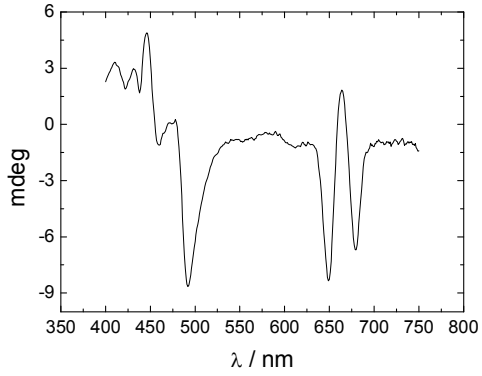


Figure 13: CD spectrum of LHCIIb UZ LM sample.

The **UZ TX** samples were prepared in exactly the same way as UZ LM samples till the purification step with ultracentrifugation. Monomeric LHCIIb was purified by ultracentrifugation on 0.1 to 1 M sucrose density gradients containing 0.05 % (w/v) Triton X 100 and 5 mM Tris- HCl (pH 7.8). After spinning for 16 h at 230000 g at 4 °C, the band containing monomeric LHCIIb was collected and concentrated by Amicon ultra centrifugal filter units (30 kDa, Millipore) up to ca. 600 μM. The so prepared samples were tested by fluorescence and CD spectroscopy as described previously. The fluorescence and CD spectra are shown in Figures 14 and 15.

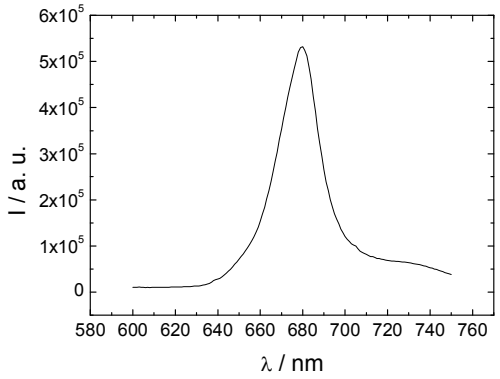


Figure 14: Fluorescence spectrum of LHCIIb UZ TX sample.

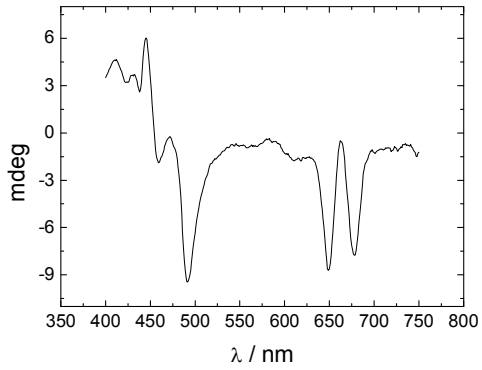


Figure 15: CD spectrum of LHCIIb UZ TX sample.

Some LHCIIb samples were containing a hexa histidiyl tag on the C- terminus. No influence of this tag on LHCIIb properties was observed in the laboratory of Professor Paulsen for years, and thus the protein samples with hexa histidiyl tags were treated and analysed in the same way.

3.1.1.1. LHCIIb trimers

The LHCIIb trimerization procedure is performed by loading the folded LHCIIb monomers, prepared as described above, with hexahistidyl tags on a chromatography column packed with material that allows for trimerization. The idea behind that, is that through the complexation of protein with the column material via the hexahistidyl rest, and following decomplexation the proteins come spatially so close to each other that the trimerisation can occur.

This column material consists of sepharose that is loaded with 2 volume equivalents of 0.3 M nickel chloride solution. The excess nickel was removed from the column with 3 volumes of 50 mM Tris- HCl buffer (pH 7.5). The column was then cooled to 4 °C. The column was then equilibrated with 2 volume equivalents of buffer consisting of 1 % OG, 0.1 M Tris- HCl (pH 9) and 12.5 % saccharose.

The LHCIIb monomers were loaded on the so prepared column where they attach by complexation of nickel by the hexahistydyl tags. The column was then filled with the same buffer as used for equilibration, and left closed for 30 minutes at 4 °C. Afterwards the column was washed one time with the equilibration buffer, and two times with a buffer consisting of 0.05 % Triton X 100, 0.01 M Tris- HCl (pH 7.5) and 0.1 mg/ml DPPG. Afterwards the created Protein-Nickel-Sepharose complex was dissociated and the LHCIIb trimers removed from the column by washing it with buffer containing 0.05 % Triton X 100, 10 mM Tris- HCl (pH 7.5), 0.3 M imidazol and 0.1 mg mg/ml DPPG with the volume corresponding to 5 column volumes.

The so prepared trimers were separated from monomers by ultracentrifugation on saccharose density gradients as described above. The only difference between the LM and TX samples is the centrifugation gradient composition as described above.

The obtained trimers were characterized by fluorescence and CD spectroscopy.

Characteristic trimer spectra in LM and TX micelles are shown in the Figures 16-19.

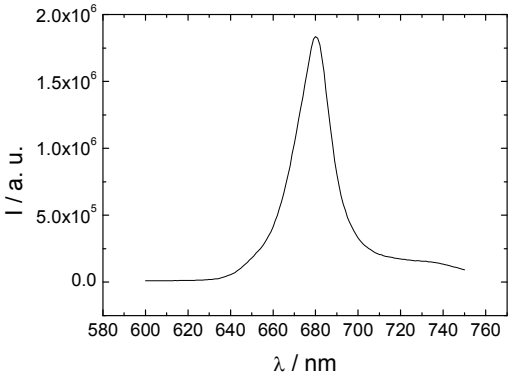


Figure 16: Fluorescence spectrum of a S59 UZ LM LHCIIb trimer sample.

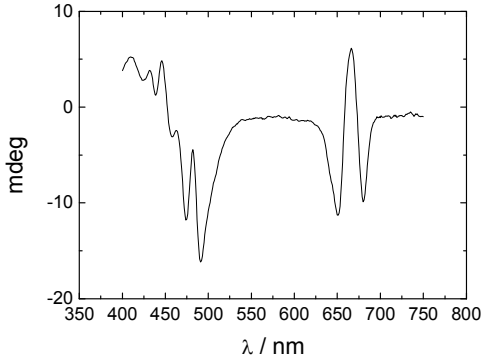


Figure 17: CD spectrum of a S59 UZ LM LHCIIb trimer sample.

The CD spectrum of LHCIIb UZ LM trimer looks very similar to the native LHCII trimer **[Hobe 1994]**.

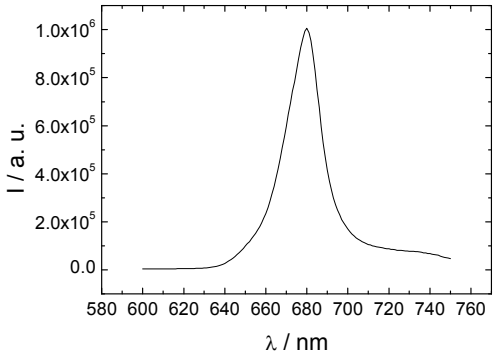


Figure 18: Fluorescence spectrum of a S59 UZ TX LHCIIb trimer sample.

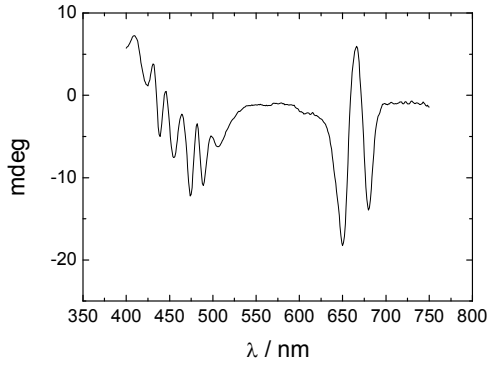


Figure 19: CD spectrum of a LHCIIb S59 UZ TX trimer sample.

The CD spectrum of LHCIIb UZ TX shows differences at small wavelengths compared to the native LHCII trimer **[Hobe 1994]**.

An EPR study of structural changes of LHCIIb between different micelle compositions and their correlation with CD and fluorescence spectra will be discussed in section 4.1.5.4.

Not all the mutants prepared during this work could be trimerized. The reason for that may be small changes in protein structure due to the spin label or positioning of the spin label in the interface between molecules in the trimer.

3.1.1.2. Deuterium exchange experiments

4 ml of the reconstituted purified LHCIIb monomer sample, prepared as described above, in buffer of pH 7.8 consisting of 0.275 M Sucrose, 0.1 % n-Dodecyl-ß-D-maltoside, 5 mM Tris- HCl (pH 7.8) was concentrated at 3000 g in a 30 kDa Amicon ultra centrifugal filter units (30 kDa, Millipore) at 4 °C until the sample volume reached 50 μl. The sample was then diluted to the volume of 250 μl with the same buffer as described above, however this time prepared with deuterated water. After that the sample was again concentrated to 50 μl. This procedure was repeated five times.
The necessary concentration was adjusted by dilution with the buffer.
The comparison of CD and fluorescence spectra of deuterated with non-deuterated samples did not show any difference (data not shown), meaning that the deuterium exchange procedure did not disturb the protein structure.

3.1.2. Reference samples

The reference samples were prepared as follows:
PROXYL IAA was dissolved in DMSO (20 mg/ml). The solution was then added to the buffer consisting of 0.275 M saccharose, 0.1 % n-Dodecyl-ß-D-maltoside 5 mM Tris-HCl (pH 7.8) to get 600 μM spin label concentration. This buffer simulates the one of LM sample after ultracentrifugation in a sucrose gradient. The spin label in buffer is then mixed with an equal volume of 80% glycerol. This sample is further referred to as **Reference 2**.
In the case of deuterated samples the buffer was made on the basis of deuterated water and deuterated glycerol was added. The reference sample contains some DMSO, which is not the case for the protein sample.
Also the 300 μM solution of PROXYL IAA in water (0.8 % DMSO) was measured. This sample is further referred to as **Reference 1**. This was done to separate the possible effects of micelles on the reference sample.

3.1.3. Cryoprotection of LHCIIb samples

All the LHCIIb samples used during this work were mixed 1:1 with 80 % glycerol as a cryoprotectant **[Jeschke 2005]**.
In the case of deuterated samples the protein was mixed 1:1 with glycerol d_8 Isotec (80 % solution in D_2O (Aldrich)).
After mixing with glycerol the samples were shock frozen in liquid nitrogen for further measurements.
The shock frozen samples should represent the protein state similar to the one in a fluid environment. It should be noted that once the sample was frozen it was newer thawed up again, as there is some evidence for structural changes during thawing.
Also the samples that were measured at room temperature were mixed 1:1 with 80 % glycerol in order to have a direct comparison with all the other samples.

3.1.4. LHCIIb samples for kinetic measurements

For kinetic measurements a different sample preparation technique was used. The samples prepared like that are further referred to as **RR TX**. During the kinetic measurements the unfolded protein sample, after being mixed with pigments and lipids as cofactors, is directly frozen in liquid nitrogen after a certain time, further referred to as folding time. The folding time is assumed to begin when the protein is mixed with the pigments **[Booth 1996]**, and ends when the sample is quenched in liquid nitrogen. For each folding time a sample specimen has to be prepared.
The consequence of such a preparation is that any unfolded or wrongly folded protein remains in the sample. That is why much effort was put in optimizing the conditions for reproducible protein folding, which required changes in composition for this type of samples.
The further discussion about the problems and reproducibility of RR TX protein samples for kinetic measurements can be found in section 4.2.1 and in **[Dockter 2008]**.
In the following the preparation of 100 µl of RR TX sample is described:
Protein solution: The mixture of 200 µg LHCIIb apoprotein in 25 µl 200 mM lithium-borate buffer with 25 % sucrose (pH 8.1) is heated to 100 °C for 2 minutes with 19.5 µl water and 5 µl 10% LDS. The mixture is cooled down and 0.5 µl of 1M β-mercaptoethanol is added.

Pigment solution: The solution of 3 mg of n-octyl-β-D-Glycoside, 5µl 10% Triton X 100, 1µl 1% deoxycholic acid sodium salt monohydrate, 13 µl (10 mg/ml) 1,2 dipalmitoyl-sn-glycero-3-(Phospho-rac-(1-glycerol)), 0.5 µl 1M ß-mercaptoethanol, 25 µl lithium borate-buffer with 25 % sucrose (pH 8.1) in 6.5 µl H_2O is prepared. To this mixture 21.4 µl of total pea leaf plant extract (20 mg/ml) in ethanol was added. The pigment extraction from pea thylakoids was performed differently than described in **[Paulsen 1990]**, as it was observed that for maximizing the protein folding yield the quality of the pigments is crucial and had to be optimized. Also in order to evaluate the folding kinetics experiments the concentration of all the reagents should not change in time. The pigments prepared by other procedures than the ones described in this section showed chlorophyll aggregation, which made them unsuitable for the investigation of folding kinetics.

The new pigment extraction method developed together with C. Dockter from University of Mainz is performed as follows:

The upper leaves of peas were gathered. Then the leaves (about 1 kilogram) were chopped in 1.5 liters of buffer consisting of 25 mM Tris- HCl (pH 7.0) 1 mM dithiotreitol, 330 mM sorbitol at 4 °C in a Waring blender (10 impulses of 10 seconds). The mixture was filtered in a centrifugal beaker that was cooled with ice, through gauze. The filtered rest was centrifuged at 4 °C at 18000 g in a Beckman cooling centrifuge with a JLA 10500 rotor and the supernatant was discarded. The pellet was resuspended in 500 ml distilled acetone. The acetone phase was centrifuged at 4 °C at 18000 g in a Beckman cooling centrifuge with a JLA 10500 rotor. The acetone phase was collected and evaporated under high vacuum conditions. The residue was again dissolved in acetone and evaporated under high vacuum. This step was repeated until all the water was removed. The remaining dry pellet was dissolved in 20 ml of acetone and analyzed concerning its constitution **[Dockter 2008]**.

The basic idea of the new extraction method is that the pigments are not purified via phase separation. It is assumed that during this step the native lipids which seem to be very important for the stability of pigment solution, can gather on the interface due to their amphiphilic nature, and are at least partially removed from the final pigment preparation, which is unwanted. Also the pigments prepared with the new method show better solubility in organic solvents, which is highly desirable for RR TX sample preparation.

The proof for increased quality of pigments prepared by this new technique is shown in **[Dockter 2008]**.

The thin layer chromatography analysis of the pigments obtained with the new extraction method shows in addition to standard lipids a large amount of lipids that were not present when using the extraction method described in **[Paulsen 1990]** (data not shown). The mass spectrometric analysis of thin layer chromatography fractions of the new lipids did not yet allow for getting further insight into their composition and structure. It might be of interest to further investigate the new lipid fractions regarding their structure with another analytical method, like NMR, in order to obtain their structure. As the effort for that is expected to be rather high and this question is beyond the scope of this thesis, such analysis was not further pursued in this work.

In order to fold the protein and assemble LHCIIb the protein solution is mixed with the pigment solution (Vortex). It is assumed that the folding starts from this moment **[Booth 1996]**. The pigments were prepared 30 seconds before they were mixed with the protein in order to prevent their aggregation **[Dockter 2008]**. Before quenching in liquid nitrogen the same volume of 80% glycerol is added to the sample. As the change of chemical composition might influence the protein folding kinetics, this step was performed directly before the quenching (ca. 30 seconds). The sample is then quenched in liquid nitrogen.

The RR TX protein samples with folding times of 15 minutes and more are referred to as “folded” if the folding time is not specified, as from the literature it is known that LHCIIb self assembly should be completed after this time **[Booth 1996]**, **[Horn 2002]**.

The unfolded samples were prepared under identical conditions as described above, with the only difference that no pea extract was added. Such samples will be further referred as “RR TX unfolded”.

The unfolded LDS samples had just the composition of the protein solution. Such kind of samples will be further referred as “LDS unfolded”.

The samples for kinetic measurements with deuterated buffer were prepared exactly in the same way as described above, with the only difference that all the buffers were prepared on the basis of deuterated water, and glycerol d_8 was used for cryoprotection.

3.1.5. Sample concentration

The concentration was determined photometricaly at 670 nm.
The concentration of UZ samples for X-band measurements was not always the same. For three-pulse ESEEM measurements it was kept at 300 μM as slight concentration effects were observed in the data. Slightly different concentrations were used for V196r (395 μM) and S52r samples (263 μM), but these are still assumed to be in the range where concentration effects are smaller than other experimental errors.
For UZ samples for CW progressive saturation measurements the concentration was also kept fixed at 300 μM.
The UZ sample concentration for CW EPR and relaxation measurements was not specifically adjusted. After preparation all samples were in a concentration range between 130 and 400 μM. No significant effects were observed in CW EPR spectra or relaxation data when reproducing experiments for the same mutant at different concentrations in this range. In acquisition of echo-detected EPR (ESE) spectra concentration varied from 150 to 300 μM. As the observed effects on line shape are relaxation induced **[Dzuba 1996]** and no effects on relaxation time were detected in this concentration range, it was assumed that this variation is not a significant source of errors.
The concentration of the UZ DEER samples was not specifically adjusted and varied in the range from 150 to 300 μM. No effects on the DEER spectra in the concentration range from 30 μM to 275 μM were observed.
The concentration of the RR TX samples was always 40 μM.
The sample concentration for W-band measurements was always 300 μM.

While sample preparation methods were optimized in close collaboration, all actual sample preparations as well as their characterization by gel electrophoresis, fluorescence and CD spectroscopy and concentration determination were done by Christoph Dockter at Mainz University, department of General Botany, unless otherwise noted.

3.2. EPR

3.2.1. Experimental procedure

3.2.1.1. X- Band CW measurements

The CW EPR spectra were measured on a Miniscope 200 spectrometer (Magnettech GmbH) with a TE102 rectangular resonator. The measurements were performed at 103 K under nitrogen cooling with a TC HO2 temperature controller (Magnettech).
Samples were loaded into 3 mm outer diameter home made quartz capillaries (50-100 μL volume), shock frozen in liquid nitrogen, and rapidly inserted into the resonator at 103 K.
It was checked that the typical microwave power of 10 μW applied during these measurements did not lead to saturation broadening. The modulation amplitude was set to 0.2 mT with a width of the central line of the nitroxide spectrum around 1 mT. The sweep width was 15 mT. Ten scans were averaged, with 4096 data points and a scan time of 60 s each.

3.2.1.2. X- Band CW progressive power saturation measurements

The CW progressive power saturation measurements were performed on a Bruker Elexsys EX 580 EPR spectrometer using a JagMar loop gap resonator. The samples were inserted into a TPX capillary and subjected to continuous N_2 or air flow at 295 K for at least 15 minutes before the measurement in order to remove the unwanted gas from the sample.
Continuous gas flow was upheld during measurements. Samples with chromium oxalate (CrOx) as a relaxation agent were measured in N_2 atmosphere.
The Bruker ER4111VT variable temperature unit was used for temperature control at 295 K. The nitroxide spectrum as a function of microwave power was then recorded at 295 K. The microwave power was increased from 23 dB to 8 dB in 3 dB steps. The modulation amplitude was set to 0.1 mT with a width of the central line of nitroxide spectrum at this temperature being about 0.3 mT in the unsaturated spectra. The sweep width was 7.5 mT, with 512 data points.

3.2.1.3. X-band pulse measurements

The X-band pulse EPR measurements were performed on a Bruker Elexsys EX 580 EPR spectrometer using a Bruker Flexline split-ring resonator ER 4118X_MS3. The resonator was overcoupled to Q ~ 100. All pulse measurements were performed at 50 K with liquid helium cooling using an Oxford CF935 cryostat with an Oxford ITC4 temperature controller.

Samples were loaded into 3 mm outer diameter home made quartz capillaries (50-100 μL volume), shock frozen in liquid nitrogen, and rapidly inserted into the resonator at 50 K.

EPR spectra were measured with field-swept echo detected EPR using a Hahn echo sequence $\pi/2$-τ-π-τ-*echo* **[Schweiger 2001]** and a 15 mT field sweep. The interpulse delay time τ was 200 ns and the pulse lengths were 16 ns for $\pi/2$ pulse and 32 ns for π pulse. The integration gate length was 200 ns.

The transversal relaxation data (T_2) was acquired with a Hahn echo sequence $\pi/2$-τ-π-τ-*echo* and applying a [(+x)-(-x)] phase cycle to the $\pi/2$ pulse. An initial interpulse delay τ of 200 ns was incremented by 8 ns. The integrated echo intensity was measured as a function of this increment with an integration gate of 32 ns length centered at the echo maximum. The pulse lengths were 16 ns for the $\pi/2$ pulse and 32 ns for the π pulse.

Three-pulse ESEEM measurements were performed with $\pi/2$-τ-$\pi/2$-T-$\pi/2$-τ-*echo* pulse sequence with a [(+x +x +x)-(+x +x +x)-(+x +x -x)+(+x +x -x)+(-x +x +x)-(-x +x +x)-(-x +x -x)+(-x +x -x)] phase cycle **[Schweiger 2001]**. The pulse length of the $\pi/2$ pulses was 16 ns and the interpulse delay τ= 344 ns, corresponding to a proton blind spot, was kept constant. Suppression of the proton modulations at the blind spot improves precision of the fitting of the deuterium modulation. The second interpulse delay time T with initial value of 80 ns was incremented in steps of 8 ns. The integration gate length was 32 ns.

The longitudinal relaxation data (T_1) were acquired with an inversion recovery pulse sequence π-T-$\pi/2$-τ-π-τ-*echo* with a [(+x)-(-x)] phase cycle applied to the $\pi/2$ pulse. The delay time T with an initial value of 2000 ns was incremented in steps of 5800 ns and the interpulse delay τ= 400 ns was kept constant. The pulse lengths were 52 ns for the $\pi/2$ pulse and 104 ns for the π pulse for the detection subsequence and 24 ns for the inversion π pulse. This combination of a hard inversion pulse with a soft echo detection is sufficient to eliminate contributions by spectral diffusion at temperatures of 40 K and higher **[Jäger 2008]**.

The integration gate length of 104 ns matched the longest pulse in the detection subsequence to maximize the signal-to-noise ratio **[Jeschke 2007b]**.
The inversion pulse was tuned by maximizing the negative echo at a delay time *T*=1000 ns.
Other flip angles and phases were generally adjusted with a $\pi/2$-τ-π-τ-*echo* pulse sequence by changing the amplitude and the phase of the pulse at τ= 400 ns, except for the ESE experiments, where the phase was tuned at τ= 200 ns.
In the ESE and T_2 relaxation measurements for the test of sample quality of RR TX samples the phase and amplitude of the signal were tuned at τ= 200 ns for double mutants. For single mutants tuning was performed as described previously. Any small differences in the signal due to the differing tuning times were assumed to be negligible by data comparison.
The field position for pulse measurements corresponded to the maximum of the nitroxide spectrum and was determined by the ESE experiment.

DEER experiments were performed with a $\pi/2(\nu_{observer})$-τ_1-$\pi(\nu_{observer})$-t`-$\pi(\nu_{pump})$-($\tau_1$ + $\tau_2$ – t`)-$\pi(\nu_{observer})$-t_2-*echo* pulse sequence with a [(+x)-(-x)] phase cycle applied to the $\pi/2$ pulse. The time t` was incremented in steps of 8 ns. The time $\tau_1$ =200 ns was kept constant. The $\tau_2$ was varying between 0.8 $\mu$s and 2.5 $\mu$s depending on the $T_2$ relaxation properties of the sample and distance that needed to be measured. The dipolar evolution time is defined as t= t`-τ_1. The data analysis was performed for t>0 ns. The pump frequency (ν_{pump}) was set to the center of the resonator dip that is also coincident with the maximum of the nitroxide EPR spectrum, whereas the observer frequency ($\nu_{observer}$) was 65 MHz higher and coincided with the low field local maximum of the spectrum (Figure 9). The observer pulse lengths were 32 ns for both $\pi/2$ and π pulses. The π pump pulse length was 12 ns. The width of the integrator gate was 32 ns. Proton modulation was averaged by adding traces at eight different τ_1 values, starting at τ_1= 200 ns and incrementing it by 8 ns.
The measurement time of DEER experiments was typically around 8 to 12 hours.
The repetition time between every pulse experiment was 6 ms in order for the system to relax completely. In the case of smaller repetition times changes in the modulation depth in DEER experiments as well as different shapes of ESE spectra were observed.
If not further specified than the measurements in X-band are meant.

3.2.1.4. W-band pulse measurements

The W-band pulse EPR measurements were performed on the Bruker Elexsys EX 680 EPR spectrometer using a W-band EN600-1021H TeraFlex ENDOR resonator. The resonator was not overcoupled. All pulse measurements were performed at 50 K with liquid helium cooling by an Oxford CF935 cryostat with an Oxford ITC4 temperature controller.

Samples were loaded with a syringe into 0.87 mm outer diameter and 0.7 mm inner diameter home-made quartz capillaries to the height of about 3 mm, and inserted into the resonator at 50 K.

EPR spectra were measured with field-swept echo-detected EPR using a Hahn echo sequence $\pi/2$-τ-π-τ-*echo* **[Schweiger 2001]** with 25 mT field sweep. The interpulse delay time τ was 240 ns and the pulse lengths were 48 ns for the $\pi/2$ pulse and 96 ns for the π pulse. The integrator gate width was 200 ns. The waiting time between the pulse sequences was 6 ms.

The flip angles and phases for the ESE experiment were adjusted with a $\pi/2$-τ-π-τ-*echo* pulse sequence by changing the amplitudes and the phases of the pulses at τ= 240 ns.

The longitudinal relaxation data (T_1) were acquired with a saturation recovery pulse sequence *saturation*-T-$\pi/2$-τ-π-τ-*echo* with a [(+x)-(-x)] phase cycle applied to the $\pi/2$ pulse. The saturation pulse has the same task as the inversion pulse in the inversion recovery pulse sequence. The delay time T with an initial value of 22000 ns was incremented in steps of 5800 ns and the interpulse delay τ= 400 ns was kept constant. The pulse lengths were 48 ns for the $\pi/2$ pulse and 96 ns for the π pulse for the detection subsequence and 20000 ns for the saturation pulse. The integrator gate width was 96 ns. The saturation pulse was set to the maximum power value. The experiment was performed on the field corresponding to the maximum of the nitroxide spectrum.

The flip angles and phases for the saturation recovery experiment were adjusted with a $\pi/2$-τ-π-τ-*echo* pulse sequence by changing the amplitudes and the phases of the pulses at τ= 400 ns. The waiting time between the pulse sequences was 15 ms.

During all the pulse experiments data was recorded in a quadrature detection mode (in phase and with 90 degree phase offset). This allows for the measurement of the real as well as imaginary part of the signal, which in turn allows for performing a phase correction of the obtained data.

3.2.2 Data analysis

3.2.2.1. X- Band CW measurements

The $2A_{zz}$ values were obtained from the CW data by fitting the minima and maxima of the nitroxide spectrum by fifth-order polynomials **[Panek 2006]** using the home-written MATLAB based "DynAnalysis" program. The error in these measurements is dominated by determination of these relatively broad extrema. However also other error sources like line broadening due to unresolved proton hyperfine coupling are possible.

The shown CW spectra were background corrected using a home written MATLAB program, by subtracting the first order polynomial fitted on the first and last 15 % of the spectral data points.

3.2.2.2. X- Band CW progressive power saturation measurements

The CW spectra were background corrected using a home written MATLAB program, by subtracting the first order polynomial fitted on the first and last 15 % of the spectral data points. The intensity of central nitroxide peak (A) was determined manually with Origin program (Microcal Software Inc.).

In saturation measurements the intensity dependence of the central nitroxide peak as a function of microwave power (P) is measured. This data was fitted with a home written MATLAB program to equation ***42*** **[Altenbach 1994]**

$$A = I\sqrt{P}\left[1+\left(2^{\frac{1}{\varepsilon}}-1\right)\frac{P}{P_{\frac{1}{2}}}\right] \quad \textit{Eq. 42}$$

The scaling factor I, the homogeneity coefficient ε and the power $P_{1/2}$, where the intensity of the nitroxide peak is reduced to half of its unsaturated value (saturation power), are adjustable parameters**.**

To factor out contributions of other relaxation mechanisms and technical parameters of the spectrometer and resonator a dimensionless accessibility parameter Π is calculated according to equation ***43*** **[Farahbakhsh 1992]**.

$$\Pi(\text{Quencher}) = \frac{P_{\frac{1}{2}}(\text{Quencher}) - P_{\frac{1}{2}}(N_2)}{\Delta H} \cdot \frac{\Delta H(\text{DPPH})}{P_{\frac{1}{2}}(\text{DPPH})} \quad \textit{Eq. 43}$$

Here $P_{1/2}$(DPPH) is the power, where the intensity of the 2,2-diphenyl-1-picrylhydrazyl (DPPH) peak is reduced to the half of its unsaturated value, ΔH(DPPH) is the line width of DPPH resonance, and ΔH was taken as the average line width of the central line of nitroxide spectrum with and without quencher. Accessibility parameters for O_2 were multiplied by 5, as the measurements were performed with air rather than pure oxygen.

3.2.2.3. Relaxation

The relaxation curves could not always be fitted by monoexponential decay functions. To compare relaxation behavior between the different mutants the effective relaxation times τ_1 for longitudinal and τ_2 for transversal relaxation are defined that serve to quantify the decay by a single number. Decay curves acquired by Hahn echo experiments were usually fitted over the maxima of the proton or deuterium modulation as described in **[Lindgren 1997]** to minimize the influence from destructive interference of nuclear modulations. For monoexponential fits τ_2 was taken directly from the fit. Otherwise it is defined as the time where echo intensity has decayed to 1/e of its initial value. For some samples the T_2 data was fitted by the second order exponential decay curve given in equation ***44*** with A, T_{2A}, B, T_{2B} and C being the fitting parameters. Compared to direct determination from the experimental data this procedure entails noise averaging **[Jäger 2008]**.

$$I(t) = A\exp(-\tau / T_{2A}) + B\exp(-\tau / T_{2B}) + C \quad \textit{Eq. 44}$$

The τ_2 parameter is computed from the parameters of the fit as the time where the biexponential function has decayed to 1/e of its initial value. As the expression I(t) for a biexponential decay cannot be solved analytically for the 1/e time, τ_2 was determined by numerically minimizing |I(t)-1/e| with a home-written Matlab program.
The transversal relaxation parameters relate to twice the interpulse delay τ so that the determined relaxation times or relaxation parameters should be multiplied with two.

Longitudinal relaxation data were fitted by the biexponential decay function shown in equation ***45***

$$I(T) = I_0(1 - 2[A\exp(-\tau / T_{1A}) + B\exp(-\tau / T_{1B})]) \quad \textit{Eq. 45}$$

Here I_0, A, T_{1A}, B and T_{1B} are the fitting parameters. The relaxation parameter τ_1 is defined as the time where a fraction of 1-1/e of the echo signal has recovered. This parameter is computed from the parameters of the fit as the time where the biexponential function has decayed to 1/e of its initial value. Compared to direct determination of τ_1 from the experimental data this procedure entails noise averaging **[Jäger 2008]** and depends less on the efficiency of echo inversion. For the 1/e time, τ_1 was determined by numerically minimizing |I(T)-1/e| with a home-written MATLAB program also used for τ_2 determination.
The relaxation time analysis for W-band measurements was performed in the same manner.

3.2.2.4. X- Band ESEEM

ESEEM data were analyzed with a home written MATLAB program kindly provided by Gunnar Jeschke. Phase, frequency and intensity of the deuterium modulation were determined from the spectrum, that was obtained by fitting a 11^{th} order polynomial background function to the primary data, subtracting this background function and dividing the difference by the background function, applying a Hamming window, zero-filling the data to four times the original length of the data set, Fourier transformation and computation of the magnitude spectrum. By virtue of the division step that corresponds to a deconvolution the modulation depth information is preserved in the spectrum. Alternatively, modulation depth was determined by deconvolution and baseline correction of the primary data as above and fitting an oscillation with Gaussian decay envelope and fixed frequency as well as variable amplitude, phase and Gaussian width to the time-domain data (see section 4.1.3. for details). As another alternative, primary time-domain data were analyzed in terms of the number and distance of closest approach of the deuterium nuclei using the spherical shell model **[Ichikava 1979]** as implemented in a home-written MATLAB program **[Hinderberger 2004]**. As already noted previously some error in the determined electron nuclear distances and number of nuclei due to the large modulation depth of some of the studied samples is possible.

3.2.2.5. X- Band DEER

The DEER measurements were analyzed using the "DeerAnalysis2006" program written by Gunnar Jeschke. The program allows easy manipulation of dipolar evolution data. With the help of this program also distance distributions can be calculated. The program is available free of charge from the internet **[http://www.mpip-mainz.mpg.de/~jeschke/distance.html]**.

The experimental data is being loaded in the program. Then the starting point and the final point (cut off time) of data analysis are defined. The final point is usually chosen so that no modulation coming from interpulse crossing is analyzed. Furthermore the starting point for the background fit is defined. All these points could be defined automatically or manually. If not further specified the starting point for data analysis is at 128 ns, the cut-off time is at 1400 ns for measurements with τ_2= 1500 ns, and at 2400 ns for τ_2= 2500 ns (note that these τ_2 parameters refer to the DEER pulse sequence as shown in the Figure 9, and have nothing in common with the relaxation parameters defined previously). These points were determined manually. The background correction step was usually performed automatically, as the optimum choice of this parameter varies between the data sets. The time point for background correction will be given explicitly for every DEER measurement. After the region of analysis was defined automatic phase correction is performed in order to reduce any effects due to a phase shift during the measurement.

The background subtraction step is very important, as the distance distribution obtained later is sensitive to the starting point of background decay subtraction and to the type of the function assumed for the background. The quality of the assumed decay shape, which was based on a three-dimensional homogeneous distribution of the spins in the sample, was controlled by subtracting the experimental curve determined by DEER measurement of the mixture of V90r and V196r single mutants from the experimental data, and comparison of the distance distributions determined by the two different background correction methods (for details see section 4.1.5.1.).

Because of the consideration above always the same parameters were used for the comparison of distance distribution and dipolar spectra in protein folding kinetics studies and for estimating distance changes as a function of different micelle composition. This was however not the case for comparison of different mutants with each other, as the best evaluation parameters differ depending on the mutation position.

After the background subtraction, which originates from intermolecular interactions, the data that contains only intramolecular contributions are fitted by a distance distribution using the Tikhonov regularization procedure. The Tikhonov regularization parameter α is determined from the L curve as explained in **[Jeschke 2006]**. The usual value for α in experiments described here is 10 unless further specified.

The transformation of the dipolar evolution function to a distance distribution is an ill-posed problem, which means it is very sensitive to the initial parameters **[Hansen 1992]**. That is why the signal-to-noise ratio of DEER spectra should be as good as possible in order to obtain a reliable distance distribution.

In a distance distribution the higher the peak at a certain distance, the more probable is this distance in the measured sample.

The Figure 20 summarizes the whole procedure of dipolar spectra analysis with the "DeerAnalysis2006" program.

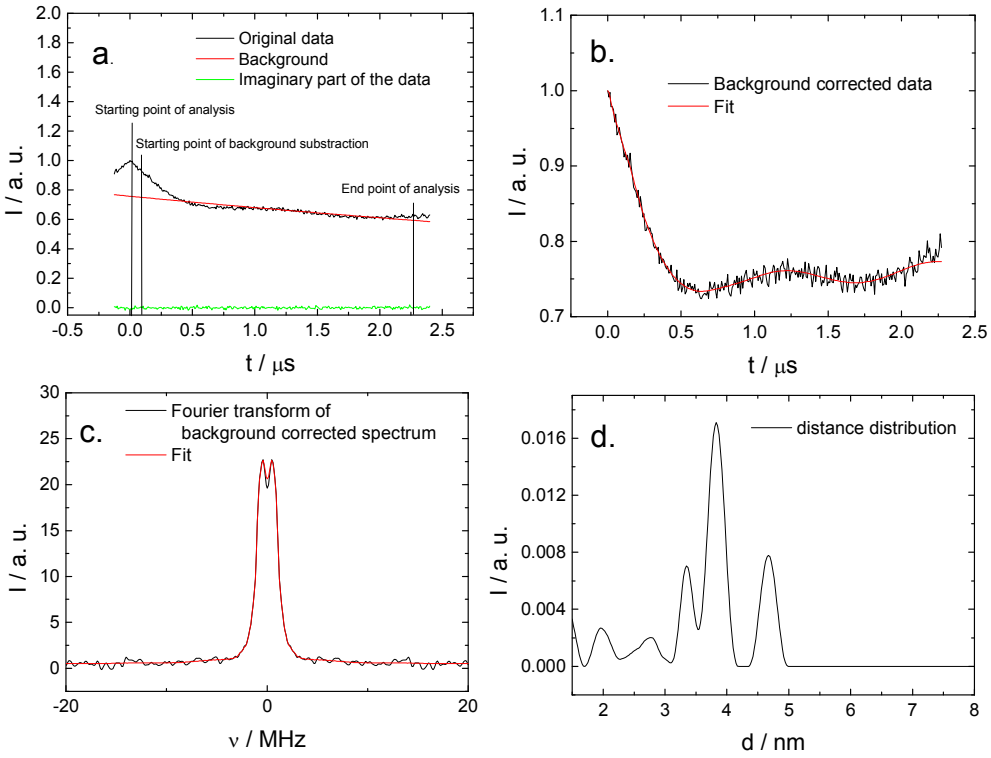


Figure 20: DEER spectrum analysis procedure for V90r/V196r UZ LM mutant.
a. original data. b. background corrected data c. Fourier transform of background corrected data. d. distance distribution obtained after Tikhonov regularization procedure.

The obtained distance distribution is then exported as an ASCII file and analyzed with the Origin program.

4. Results

4.1. LHCIIb structure investigation with EPR

This chapter shows what structural information about LHCIIb can be obtained with different EPR methods. The structure investigation on LHCIIb was always performed on UZ samples. They have a more defined composition than RR TX samples, and thus are more suitable for structural investigation and EPR method development. If not further specified, the UZ LM samples are used. After understanding the effects influencing the data, the presented methods are applied to the investigation of the folding and self-assembly process of LHCIIb.

4.1.1. CW EPR spectroscopy

As already described in theory section, the hyperfine splitting constant is sensitive to the solvent polarity.

The maximum extrema splitting $2A_{zz}$ (see Figure 3) in the solid state corresponds to twice the hyperfine coupling constant along the lobes of the p_π orbital on the ^{14}N nucleus of the nitroxide. Localization of the unpaired electron in this orbital corresponds to a charge-separated state, while localization in the p_π orbital on oxygen corresponds to a neutral state. Hence spin density in the ^{14}N orbital and thus A_{zz} (the half of $2A_{zz}$) increase with polarity of the environment **[Steinhoff 2000], [Owenius 2001]**, **[Kurad 2003]**.

Thus A_{zz} values can provide information about relative water accessibility of spin labels in membranes **[Griffith 1974], [Kurad 2003]**, **[Noethig 2004]**, **[Bartucci 2003]** and polarity of the spin label environment in proteins **[Steinhoff 2000]**

Figure 21 shows typical solid-state nitroxide spectra for mutants S52r and V196r and Reference 2 sample.

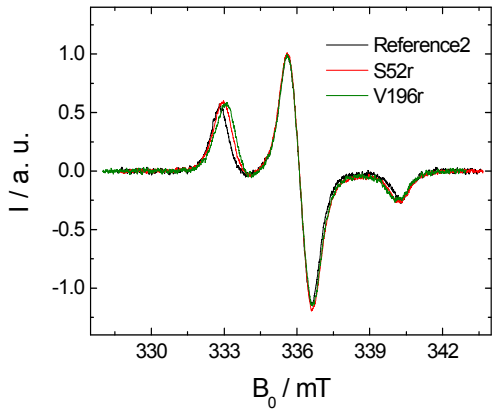


Figure 21: Typical solid state nitroxide spectra for some characteristic mutants and a reference sample. The spectra are intensity and field normalized.

From Figure 21 it is seen, that $2A_{zz}$ depends on mutation position. This allows for getting environmental information about the spin-labelled amino acid in LHCIIb. The A_{zz} values as a function of mutation position are summarized in Figure 22.

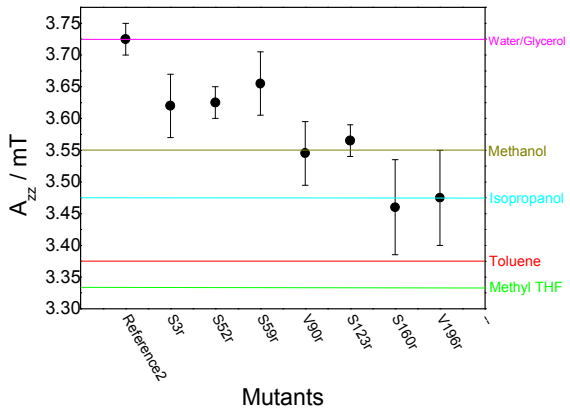


Figure 22: A_{zz} values for different LHCIIb mutants. Data from reference measurements of PROXYL IAA spin label in solvents of various polarity (measured by T. Bund) are shown as parallel lines.

As no difference between the A_{zz} values between deuterated and non-deuterated samples was observed, the average of the A_{zz} value was taken, except for samples S160r and V196r, where only one measurement was performed. The estimated errors

during the measurements are shown in the plot. The maximum error for PROXYL IAA measurements in different solvents is estimated to be 0.05 mT.

The Reference 2 sample should correspond to the most solvent accessible position and indeed its A_{zz} value is also the largest one. Among the mutants S3r, S52r and S59r form a group of residues with large A_{zz}, V90r, S123r a group with intermediate A_{zz} and S160r and V196r a group with low A_{zz}.

The semi-quantitative estimate of the polarity of the spin label environment can be obtained by comparison to A_{zz} of PROXYL IAA in solvents of different polarity (horizontal lines in Figure 22).

Although micelles are present in the Reference 2 sample, the nitroxide has an environment comparable to the one in a water/glycerol mixture without detergents. Mutants S3r, S52r, S59r have environments less polar than this mixture but more polar than methanol, V90r and S123r have an environment with a polarity comparable to the one of methanol, and mutants S160r and V196r have an environment with a polarity similar to the one of isopropanol.

The range of A_{zz} values observed on detergent-solubilized LHCIIb labelled with PROXYL IAA (3.45-3.65 mT) is similar to the one found for membrane-solubilized bacteriorhodopsin labelled with 1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl)methanethiosulfonate (3.50-3.675 mT) **[Steinhoff 2000]** but differs significantly from the one observed with DOXYL-labelled lipids in lipid bilayers (3.20-3.50 mT) **[Kurad 2003]**. This can be traced back to a polarity-independent difference of the spin density on nitrogen between the oxygen-free heterocycles like PROXYL and 1-oxyl-2,2,5,5-tetramethylpyrroline on the one hand and the oxygen containing heterocycle DOXYL on the other hand.

4.1.2. CW progressive power saturation measurements

As described previously saturation measurements can give information on accessibility of the nitroxide spin label to paramagnetic water soluble (CrOx) and membrane soluble (O_2) quenchers and thus on their relative position.

Figure 23 shows the intensity dependence of the central nitroxide peak measured as a function of microwave power for different mutants. In the Figure 23 data for Reference 2 sample, whose environment is highly polar according to the A_{zz} value, the S52r mutant, whose environment is highly polar, and the V196r mutant, whose environment is only weakly polar, are shown. Although differences in saturation behavior can be discerned they are less impressive than the differences in A_{zz}. As seen from the figure, the dependence of the intensity of the central nitroxide peak on microwave power is generally well fitted by equation ***42*** allowing to get the parameters for accessibility determination.

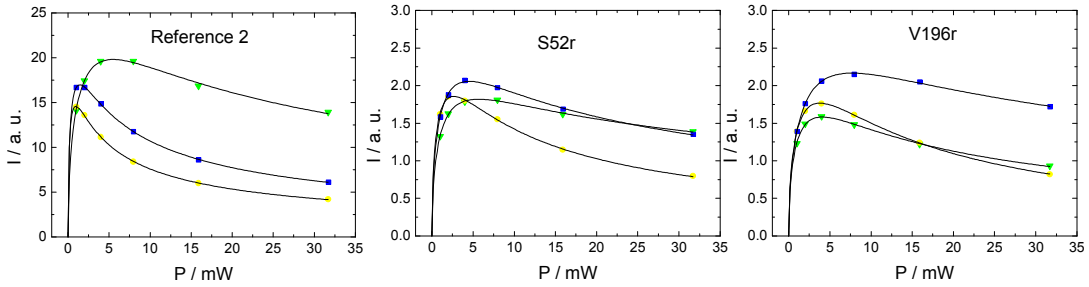


Figure 23: Progressive power saturation measurements on different LHCIIb mutants. Intensity of the central nitroxide peak as a function of microwave power is shown. ● N_2, ▼ CrOx, ■ air (20 % O_2). The solid lines are fits according to equation ***42***.

The accessibility parameters for oxygen and CrOx determined from the power saturation behavior according to equation ***43*** for different spin label positions are summarized in the Figure 24. The error of these measurements was estimated to be 1% for CrOx accessibility and 5% for oxygen accessibility. The main source of errors here is assumed to be the varying concentration of the quenchers due to the sample preparation.

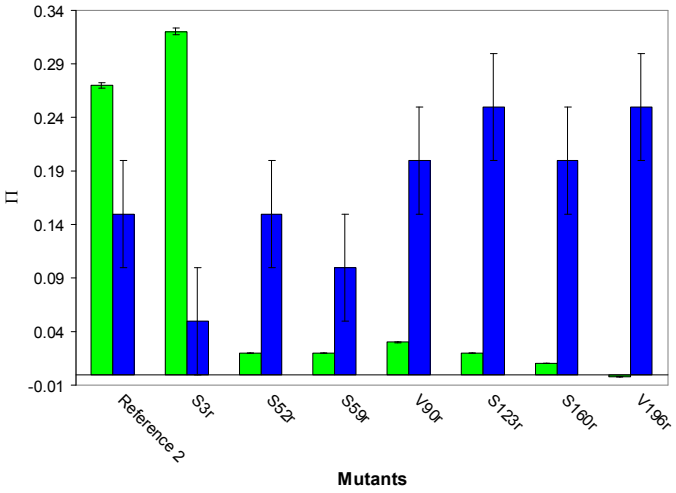


Figure 24: Accessibility parameters from CW progressive power saturation measurements for different LHCIIb mutants. CrOx (green bars), O_2 (blue bars). The determined oxygen accessibility was multiplied with 5, as air contains only 20% of oxygen.

Variation of the water accessibility parameter for CrOx as a relaxation agent is rather small among all the mutants, except for S3r, which exhibits CrOx accessibility similar to the one of unbound spin label in the buffer (Figure 24). This is another hint that position S3 as well as the Reference 2 sample are in the aqueous phase. The negative value for CrOx accessibility of the V196r mutant is due to experimental errors in determination of $P_{1/2}$ and is interpreted as non-accessibility to CrOx.

Variation of the oxygen accessibility parameter is generally small although significantly larger than the experimental error. Mutants V90r, S123r, S160r and V196r exhibit enhanced oxygen accessibility correlated with very low CrOx accessibility. Such behavior is usually interpreted as an exposure to a lipid environment, or, in this case, a location in the core of the micelle and thus exposure to detergent alkyl chains.

In the previous work done in our group similar high CrOx accessibility for the S3 position has already been observed **[Dockter 2005]**. However, due to the differing experimental conditions no detailed comparison between the data is possible.

4.1.3. ESEEM

The three-pulse ESEEM experiment can selectively measure the electron nuclear hyperfine coupling. The depth of the echo modulation and its decay in the ESEEM signal contains information about the amount and the distance of the nuclei coupled to the electron **[Ichikawa 1979]**, **[Schweiger 2001]**. Here three-pulse ESEEM is utilized for determination of the spin label position in a medium size transmembrane protein by probing its solvent accessibility.

The comparison of the deuterium modulation depth in ESEEM experiments between different mutants and a reference sample measured at the same experimental conditions is shown in Figure 25.

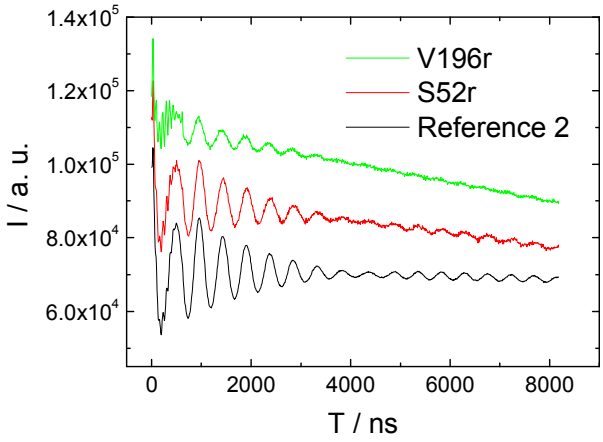


Figure 25: ESEEM time-domain data for different LHCIIb mutants and a reference sample.

Differences in modulation depth can be clearly seen as a function of spin label position. The interpretation of these measurements is that V196r is most probably buried in the hydrophobic parts of membrane or protein, whereas S52r and the Reference 2 sample are accessible to deuterated buffer. This figure proofs the possibility of using ESEEM for deuterium accessibility determination also for medium size transmembrane proteins.

The obtained ESEEM data was analyzed in terms of modulation depth and by fits of the spherical shell model **[Ichikawa 1979]**, **[Schweiger 2001]**.

In previous work ESEEM-based water accessibility parameters were defined in different ways. In a simple definition the intensity of the Fourier transformed deuterium ESEEM peak is normalized to the intensity of the proton peak [**Bartucci 2003**]. This

definition, however, depends on the assumption that the local concentration of non-water protons in contact with the spin label is invariable. This may not be the case for membrane proteins. Due to suppression effects for protons it also depends rather strongly on the choice of the first interpulse delay τ in the three-pulse ESEEM pulse sequence (see Figure 5 for details). A more robust definition is based on the modulation depth K, defined as the peak-to-peak amplitude between the first maximum and first minimum of the deuterium modulation [**Carmieli 2006**]. This parameter is a direct measure for the distance and concentration of deuterium nuclei. However, the definition by a visual fit that involves averaging of proton modulations leads to decreased precision compared to the parameter based on spectral intensity. Furthermore this parameter also depends on the choice of τ, although not as strongly as the proton-normalized deuterium peak intensity.

The deuterium modulation depth can be determined precisely in the following way. The primary ESEEM data are corrected for the unmodulated part by fitting a polynomial function B(T) [**Schweiger 2001**], subtracting B(T) from the primary data, and dividing the difference by B(T) to obtain a deconvoluted and normalized nuclear modulation function N(T). Fourier transformation of N(T) provides the complex ESEEM spectrum. The phase-independent magnitude spectrum is then obtained by computing the absolute value of complex spectral intensities. Typical magnitude ESEEM spectra for three characteristic mutation positions are shown in Figure 26.

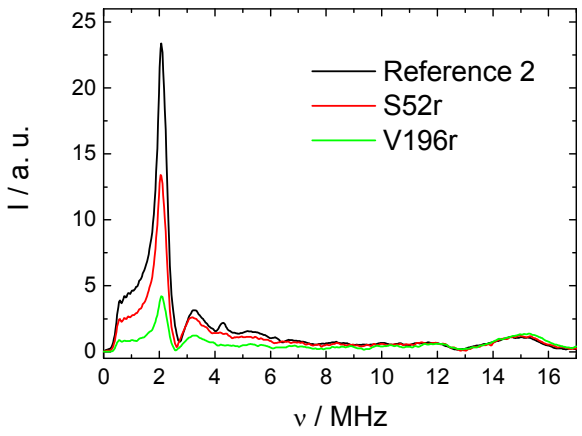


Figure 26: Fourier-transformed ESEEM data for different LHCIIb mutants. The intensity of the peak at around 2 MHz can be directly used as a deuterium accessibility parameter.

Due to proper normalization the amplitude $I(\nu_D)$ of the peak at the deuterium frequency ν_D in these spectra is proportional to the modulation depth as can be seen by comparison of Figure 25 with Figure 26.

From Figure 26 it is also seen that the ESEEM peak around 2 MHz has two components. The low intensity broad component is known to originate from D_2O molecules that are directly hydrogen bonded to the spin label. The high intensity narrow component reflects the D_2O molecules that are not hydrogen-bonded to spin label **[Erilov 2005], [Bartucci 2006]**. As it is known that the latter spectral component is linearly dependent on the deuterium concentration **[Erilov 2005 and references there]**, only this spectral part was used for further analysis.

As the phase of the echo and resonance frequency were explicitly controlled after every experiment and did not show significant changes, thus proving that there were no significant hardware errors during the experiment, the main error for these measurements is assumed to come from imperfections of the data analysis. For the given signal-to-noise ratio the error for these measurements is estimated to be 5%.

The described analysis method may provide different absolute values in different laboratories as it depends on the implementation of the Fourier transform algorithm and on the number of data points. For that reason the intensity of deuterium peak ($I(\nu_D)$) as shown in the Figure 26 is unlikely to be comparable between studies performed in different labs. A well defined numerical value can be obtained in the following way. From the real and imaginary peak intensities in the complex spectrum, $Re(\nu_D)$ and $Im(\nu_D)$, phase ϕ of the deuterium modulation can be estimated as $\phi=\text{atan}[Im(\nu_D)/ Re(\nu_D)]$. A damped harmonic oscillation function

$$D(t) = k_D \cos(2\pi\nu_D T + \phi)\exp(-T^2/\tau_0^2) \qquad \textit{Eq. 46}$$

is then least-square fitted to $N(T)$ by varying deuterium modulation depth k_D, damping constant τ_0 and phase ϕ while keeping ν_D fixed. The Gaussian damping function was found to provide good fits of the decay of the oscillation for all samples. Modulation depth k_D is directly related to the primary data and is thus independent of computational algorithms. The orientation average $\langle k \rangle$ of the modulation depth k in the theoretical description of three-pulse ESEEM [**Schweiger 2001**] is related through

$$k_D = \frac{\langle k \rangle}{2}[1-\cos(2\pi\nu_D\tau)] \qquad \textit{Eq. 47}$$

where time τ in the suppression factor is the delay between the first two pulses. To the modulation depth parameter as defined in [**Carmieli 2006**] k_D is related through $k_D \approx K/2$. The suppression factor $1-\cos(2\pi\nu_D\tau)$ can range between 0 and 2. To minimize the influence of proton modulations on the fit of the deuterium modulation, the measurements were performed at τ = 344 ns, which corresponds to the j = 5 blind spot of proton modulation (j= $\nu_H\tau$= 5). Compared to the choice of τ= 204 ns in **[Erilov 2005]**, which corresponds to the j= 3 proton blind spot, this value leads to stronger suppression of the broad wings due to hydrogen-bonded deuterons and thus to more stable fits of the modulation that stems exclusively from non- hydrogen-bonded deuterons. For τ= 344 ns the suppression factor has a value of 1.21, for τ= 204 ns it has a value of 1.89, and for the j= 4 blind spot τ= 272 ns it has a value of 1.92.

By fitting with fixed frequency ν_D, only deuterium nuclei contribute to k_D. Neglecting effects of the small deuterium quadrupole coupling on modulation depth and assuming sufficiently long distances the contribution of each individual deuteron is proportional to r^{-6} [**Ichikawa 1979**]. The modulation depth is thus dominated by deuterons at the distance of closest approach. However, the contribution of directly hydrogen-bonded deuterons, which leads to the low-intensity broad wings of the deuterium peaks in Figure 26 [**Erilov 2005**], [**Bartucci 2006**], is suppressed in k_D, as the Gaussian damping function guards against fitting of fast-decaying components of the modulation. The fitting procedure thus selects the narrow spectral component whose intensity and corresponding modulation depth are proportional to the concentration of D_2O molecules that are not hydrogen-bonded to the spin label **[Erilov 2005]**.

At such distances the modulation depth k_D at any orientation and thus also its orientation average $\langle k \rangle$ scales as ν_D^{-2} **[Schweiger 2001]**. A three-pulse ESEEM based water accessibility parameter Π_{D_2O} that is independent of the choice of interpulse delay τ and the exact static field B_0 for the measurement, is defined in equation 48

$$\Pi_{D_2O} = \frac{2k_D}{[1-\cos(2\pi\nu_D\tau)]} \cdot \left(\frac{\nu_D}{2\,\mathrm{MHz}}\right)^2 \qquad \textit{Eq. 48}$$

By normalization of the deuterium frequency to a standard value of 2 MHz it is ensured that at X-band frequencies with τ= 344 ns this accessibility parameter is similar to the parameter K defined in **[Carmieli 2006]**.

The main error for such an accessibility parameter determination is mostly caused by the noise. As the signal-to-noise ratio for all these experiments is good, the error in the deuterium accessibility parameter is estimated to be 5%.

According to ESEEM theory *k* depends on the number n and distances r of nuclei coupled to the electron spin. The influence of the number of nuclei and their distance of closest approach on the modulation depth can be separated by fitting the whole modulation by the spherical shell model [**Ichikawa 1979**]. The separation is possible since the damping of the modulation, quantified by empirical parameter τ_0, depends on the distance but not on the number of nuclei. This analysis provides a distance of closest approach of 0.35 nm for all LHCIIb mutants. Variations in modulation depth are thus exclusively due to variations in the average number n of deuterons at this distance. This number ranges between about 0.2 for the least accessible V90r and V196r mutants and 0.55 for the most accessible S59r mutant. Even larger values of about 0.8 are found for the reference samples with unbound spin labels. Note that these values may not be unique, as was pointed out in previous work [**Erilov 2006**], where longer distances of closest approach were assumed.

The error in deuterium accessibility using this analysis method is assumed to be higher than for those ones described above. The reason for that is that in order to get a reasonable fit of the data only a few deuterium modulations were taken for analysis. That is why the error of this evaluation method is assumed to be 10%. Furthermore systematic errors due to the rather high modulation depth are possible.

The ESEEM deuterium accessibility parameters determined by using different evaluation methods are summarized in the Figure 27.

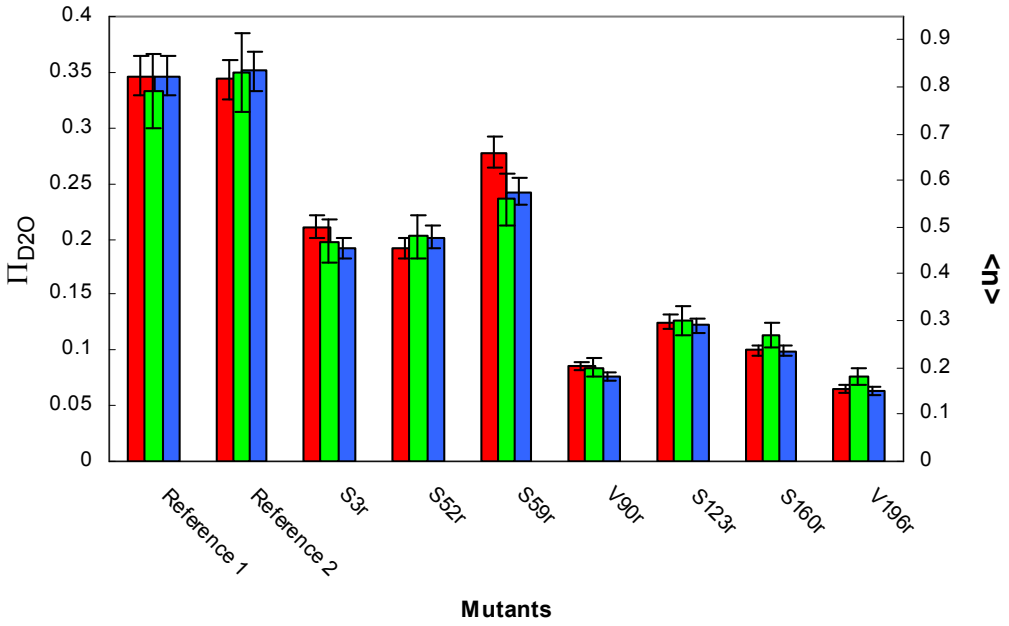


Figure 27: Water accessibility parameters obtained by different analysis procedures from deuterium three-pulse ESEEM data. For each sample different accessibility parameters are given: Π_{D_2O} (red bars, left vertical scale), the deuterium peak intensity of ESEEM spectra normalized on the ESEEM accessibility of Reference 1 sample (blue bars, arbitrary units), and the average number of deuterium nuclei from a spherical shell model simulation (green bars, right vertical scale).

As seen in Figure 27 the accessibility parameter Π_{D_2O}, the deuterium peak intensity I, and the average number <n> of deuterium nuclei correlate very well with each other. Among these parameters Π_{D_2O} is best suited for comparison of measurements on different proteins and in different laboratories. However, if the technique is applied at the same EPR frequency with the same interpulse delay τ and data sets with the same number of points are background-corrected and Fourier-transformed by the same programs, Π_{D_2O} is related to I by $\Pi_{D_2O} = \varsigma \cdot I$ with a constant factor ς. For that reason the evaluation of folding kinetics data is performed below via the deuterium peak intensity dependence.

Figure 27 again proofs that different LHCIIb mutants have different accessibility to deuterated buffer. Both reference samples show high deuterium accessibilities. From that it is concluded that the effect of micelles in the reference sample is not significant concerning accessibility. In other words, the probe resides in the aqueous phase rather

than in the core of the micelles. Mutants S3r, S52r and S59r seem to be highly solvent exposed, as they have deuterium accessibility comparable to the reference sample. The mutants S123r, S160r are at intermediate positions, and V90r and V196r are probably buried in the protein.

In contrast to the CrOx accessibility parameter Π_{D_2O} is significantly smaller for all sites in the protein, including site S3, than for the reference samples. Unlike with CrOx the slight accessibility of the V196r mutant can still be detected. In general the differences in water accessibility between different sites are better visible than with progressive saturation experiments.

4.1.4. Relaxation

Effective relaxation times were measured on different LHCIIb mutants. Since many stochastical processes can influence the relaxation in the solid state, the interpretation of the obtained data is not as straightforward as interpretation of CW EPR or ESEEM data. In this work it was tried to understand the effects influencing the relaxation of a nitroxide spin label in LHCIIb on the basis of empirical relaxation parameters.

These studies may be of importance, as the relaxation measurements are technically not as demanding as other pulse experiments, but the information that can be obtained from them is unique, as will be shown later. The measurements presented here can hopefully be used as a reference for further membrane protein studies with relaxation techniques.

X- Band

Figure 28 shows the characteristic curves determined by relaxation measurements as well as the fits for τ_2 determination as described in section 3.2.2.3. The parameter τ_1 was determined from biexponential fits as described in the section 3.2.2.3. The fits are indistinguishable from the data with bare eyes and are thus not shown. Determination of the effective transversal relaxation time τ_2, which coincides with the transversal relaxation time T_2 for stretched exponential and monoexponential decays, is more complicated. This is because the maxima of the nuclear modulation have to be fitted in modulated echo decays (see, e.g. **[Lindgren 1997]**) as shown in Figure 28. Furthermore for the two protonated reference samples, decays were strongly non-exponential. This can be caused by the additional relaxation mechanism caused by DMSO methyl protons that are present in these samples.

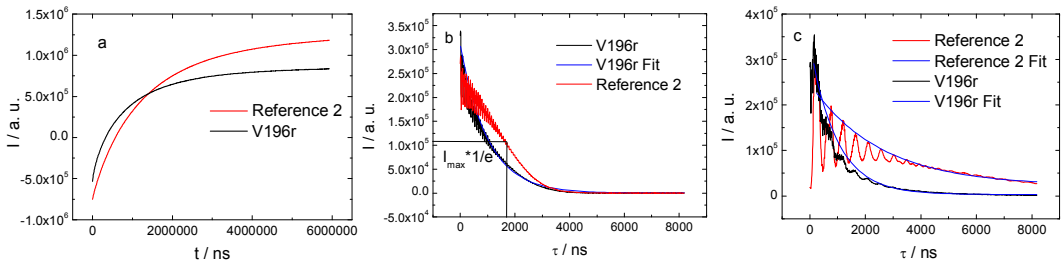


Figure 28: a: Longitudinal relaxation curves, b: Transversal relaxation curves, c: Transversal relaxation curves in deuterated solvent. The effective relaxation time determination is also shown.

The effective relaxation times determined from fitting the relaxation curves are summarized in Table 1.

Mutant	τ_1 / µs		τ_2 / ns	
	H_2O	D_2O	H_2O	D_2O
Reference 1	1331	1350	3200	4030
Reference 2	1352	1337	3408	5063
S3r	1070	1029	2178	3893
S52r	1054	1029	2185	3532
S59r	1140	1181	2369	5174
V90r	1021	978	2140	2525
S123r	1063	1027	2118	2488
S160r	871	871	1924	2468
V196r	862	705	2000	1923

Table 1: Longitudinal (τ_1) and transversal (τ_2) effective relaxation times of different LHCIIb mutants in different solvents.

The maximal error for τ_1 determination was estimated to be 10%. The error for τ_2 determination is estimated to be 40 ns by comparison of the obtained data on the same sample. Due to the different fitting procedures also a systematic error for the τ_2 relaxation times is possible.

It is known that the spin label environment strongly affects transversal relaxation time. The effects mostly contributing to T_2 are considered to be the local proton concentration and the methyl group rotation **[Lindgren 1997]**, **[Zecevic 1998]**, **[Huber 2001]**.

For the solvents containing no methyl group, there is a linear correlation between proton concentration in the solvent and relaxation rate **[Lindgren 1997]**. However in comparison to the soluble protein human carbonic anhydrase II, where labels at surface-exposed sites have much longer relaxation times than those at buried sites [**Huber 2001**], for detergent-solubilized LHCIIb only minor variations of τ_2 between different mutation positions are observed in protonated buffer.

In order to proof the effect of surrounding protons on T_2, relaxation measurements in deuterated solvents were performed. The change in τ_2 on deuteration of the buffer depends strongly on the labelled site in LHCIIb (Table 1) as it does in carbonic anhydrase II.

This change arises because transversal relaxation of electron spins in protonated samples in the low temperature limit is predominantly due to fluctuations of the hyperfine field at the electron imposed by the protons. These fluctuations in turn are induced by proton spin diffusion. For deuterons with an about seven times smaller magnetic moment the fluctuations are also much smaller. The contribution of this mechanism to the relaxation rate is given by **[Brown 1979]**, **[Schweiger 2001]**

$$\Delta\left(\frac{1}{\tau_2}\right)=\frac{0.37\mu_0\left(g\beta_e\right)^{1/2}\left(g_n\mu_n\right)^{3/2}\left[I(I+1)\right]^{1/4}}{4\pi\hbar}C \qquad \textit{Eq. 49}$$

where C is the number concentration of nuclear spins (protons) that induce this relaxation, g_n their nuclear g value (5.585694701), g the electron spin g value, μ_n the nuclear magneton ($5.05078324\cdot10^{-27}$ J/T) and I is the spin quantum number of the nuclei. This expression suggests that the contribution is about 13 times smaller for deuterons than for protons and still neglects additional slow-down of spin diffusion for quadrupole nuclei such as deuterium due to quadrupolar broadening of their resonance line. The contribution of deuterons can thus be neglected in the difference of relaxation rates in protonated and deuterated media and a local concentration of protons can be computed by equating $\Delta(1/\tau_2)$ in ***Eq. 49*** with this difference. This approach implicitly assumes that all exchangeable protons are fixed in space. While libration of protons in hydrogen bonds can be neglected on the relevant length scales, hyperfine field fluctuations due to rotation of methyl groups may be relevant. However, as long as the

methyl groups are protonated in both samples, as is the case in experiments with deuterated water, their contribution to transversal relaxation cancels in the difference $\Delta(1/\tau_2)$ and can be neglected.

According to considerations above the samples that are stronger exposed to water should show larger $\Delta(1/\tau_2)$. This allows an alternative way of determination of solvent exposure of the spin label.

The $\Delta(1/\tau_2)$= $(1/\tau_2(H_2O)- 1/\tau_2 (D_2O))$ values for different mutants are presented in the Table 2.

Mutant	**$\Delta(1/\tau_2) \cdot 10^4$ / 1/ns**
Reference 1	0.64
Reference 2	0.96
S3r	2.02
S52r	1.62
S59r	2.29
V90r	0.71
S123r	0.70
S160r	1.15
V196r	-0.20

Table 2: Effective transversal relaxation rate differences for different mutation positions.

A large value of relaxation rate difference means a strong change of the relaxation time when deuterium buffer is added and thus indicates a solvent accessible residue.

The negative value for the V196r mutant probably comes from experimental errors.

The difference in relaxation rates between protonated and deuterated samples is sensitive to proton concentration on a different length scale than ESEEM modulation depth. Due to its r^{-6} dependence the ESEEM modulation depth is strongly biased towards distances of closest approach, i.e. to water molecules in van-der-Waals contact with the label. Exactly those protons and the hydrogen-bonded protons are decoupled from the proton spin bath by their hyperfine splitting. Thus there is no proton spin diffusion in direct neighborhood to the electron spin, **[Blumberg 1960]**, **[Zecevic 1998]**, **[Huber 2001]**. Protons are effective in spin diffusion induced relaxation if their

hyperfine coupling is comparable to or smaller than the proton-proton dipole-dipole linewidth of about 100 kHz. The dipolar proton hyperfine coupling falls below 300 kHz at a distance of about 7 Å from the electron spin. Thus the relaxation technique is most sensitive to the concentration of protons in a distance range between approximately 7 and 20 Å, where protons couple both to the electron spin and to the large spin bath of remote protons.

Concentrations of exchangeable protons in this distance range can be calculated according to equation ***49*** and transformation of the obtained concentration to the usual units. They range from 17 M for mutant S123r to 57 M for mutant S59r. These values compare to a proton concentration of 111 M in pure water. The concentration variation between the different mutation positions can be explained by the shielding effect of the spin label environment. This fact allows the use of transversal relaxation for structure determination in LHCIIb. The small concentration values for the reference samples can be explained by a significant relaxation mechanism different from proton spin diffusion, as can be estimated from the shape of the relaxation curve.

For mutant V196r a negative concentration is observed. As it physically does not make sense, this result is explained by the systematic error introduced by the experiment or fitting of the relaxation curves. This mutant is considered to be solvent inaccessible.

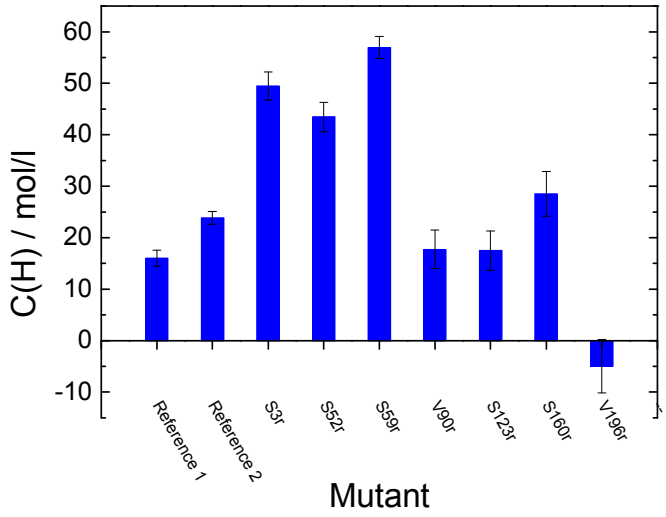


Figure 29: Local concentration of exchangeable protons according to equation ***49*** for different LHCIIb mutants.

In order to increase the amount of information that can be extracted from relaxation experiments, W-band frequency measurements were performed.
Figure 30 shows a characteristic W- band nitroxide ESE spectrum.

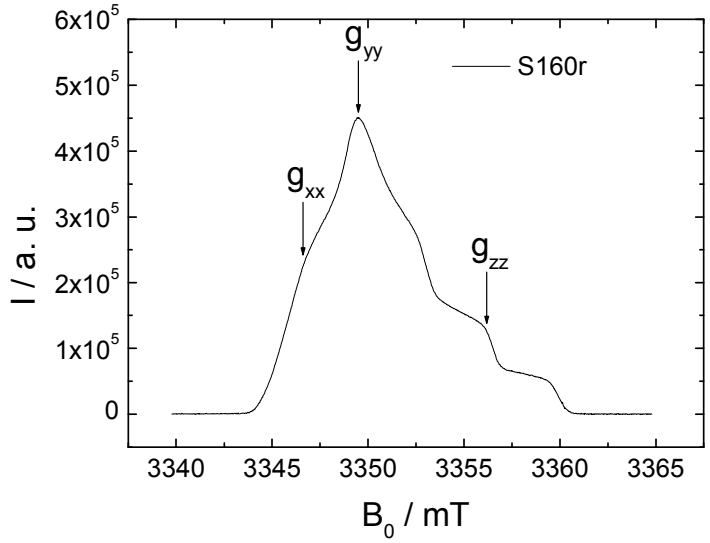


Figure 30: ESE W-band spectrum of the S160r mutant.

The g tensor orientations as adapted from **[Owenius 2001]** are also shown. The spectrum in Figure 30 differs significantly from the one shown in Figure 4 that was measured at X-band frequencies. The ESE spectra in W-band allow the separation of **g** tensor components from the **A** tensor components **[Jeschke 1998]**. This is not possible in the X-band due to the smaller field used.
The shape of the ESE spectrum in the Figure 30 shows features that are characteristic for very slow spin label reorientation **[Leporini 2003]**.
By measuring the ESE spectra with different separation times between the pulses (τ) (see Figure 4 for details), an effective relaxation time (τ_2) is obtained. Such measurements allow for obtaining relaxation information as a function of nitroxide orientations in the external magnetic field that can be further interpreted in terms of spin label libration geometry as described in **[Rohrer 1996]**. Furthermore the measurements in the W-band are more sensitive concerning the g-tensor reorientation due to molecular libration compared to X-band measurements **[Schweiger 2001]**.
In the Figure 31 such ESE measurements are presented.

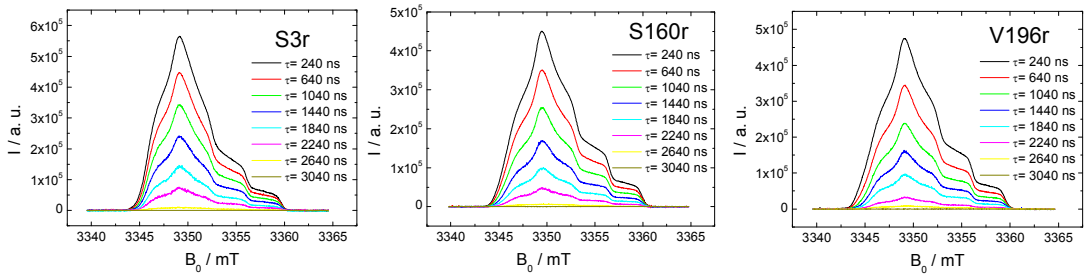


Figure 31: W-band ESE spectra as a function of interpulse position for different LHCIIb mutants.

The intensity of the ESE spectra decreases with increasing interpulse distance due to transversal relaxation. The effects of different relaxation rates at different spectral positions are better seen by normalization of the curve at τ= 2240 ns on the one at τ= 240 ns, as shown in Figure 32.

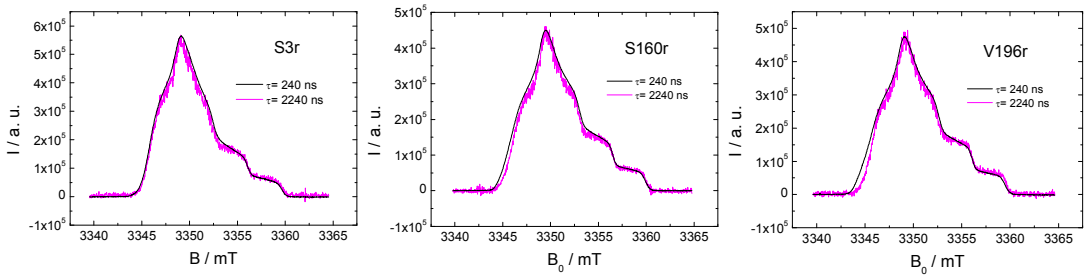


Figure 32: Transversal relaxation time anisotropies of different LHCIIb mutants in W-band.

Figure 32 shows only very slight differences in relaxation behavior at the g_{xx} position for all of the mutants. Otherwise almost no difference in the shape of the spectra measured at different τ values is seen. This result can be interpreted in terms of the isotropic spin label motion geometries at different mutation positions. Furthermore the measurements suggest that the spin label has the same motional geometry in the helix as in the flexible loop region under these conditions, thus indicating that protein secondary structure can not be differentiated from these measurements. This observation supports the idea that the mobility of the spin label under these conditions

is mostly determined by the structure of its solvation shell and not by the secondary structure of the protein (see also next page for details).

Also the effective longitudinal relaxation time τ_1 can provide environmental information. It is known that the Raman process is dominant for the relaxation of nitroxides at X-band in the studied temperature region in water-glycerol mixtures **[Zhou 1999]**. The relevant vibrations were assumed to have intramolecular origin **[Du 1995]**. Furthermore it was shown that at these low temperatures longitudinal relaxation is dominated by modulation of the hyperfine field due to libration motion of nitroxides in a glassy matrix **[Leporini 2003]**.

While effective longitudinal relaxation times (τ_1) do not exhibit significant changes on deuteration of the buffer, they vary significantly among the different LHCIIb mutants (Table 1). In general, relaxation times are longer for sites that are known to be exposed to water (reference samples) or have high water accessibility according to ESEEM. The prolonged τ_1 for both reference samples and for the S59r mutant and the shortened τ_1 for the S160r and V196r mutants can then be understood by considering restriction of the librational motion of the spin label by the neighborhood. In an aqueous environment a stiff cage of ordered water molecules can be formed around the spin label due to the hydrophobic effect. This cage hinders small angle motion of the spin label to a larger extent than an alkyl chain solvation shell or neighboring protein side groups. This stronger hindrance in turn is reflected in a longer τ_1.

There is an alternative explanation for this effect based on the idea that hydrogen bonding of the solvent to the nitroxide changes its dynamics **[Du 1995]**. As there is a linear correlation between τ_1 and ESEEM deuterium accessibility that is known to originate from non-hydrogen bonded deuterium atoms (s. Figure 48), the hypothesis of the hindered libration caused by the water cage *around* the spin label that in term causes the changes in relaxation was favorized. The other effect however can not be excluded.

Further support for the hypothesis of hindered libration is found in lineshape effects in ESE spectra of the different samples (Figure 33) that arise from anisotropy of the transversal relaxation time **[Dzuba 1996]**, **[Kirilina 2001]**, **[Kirilina 2004]**.

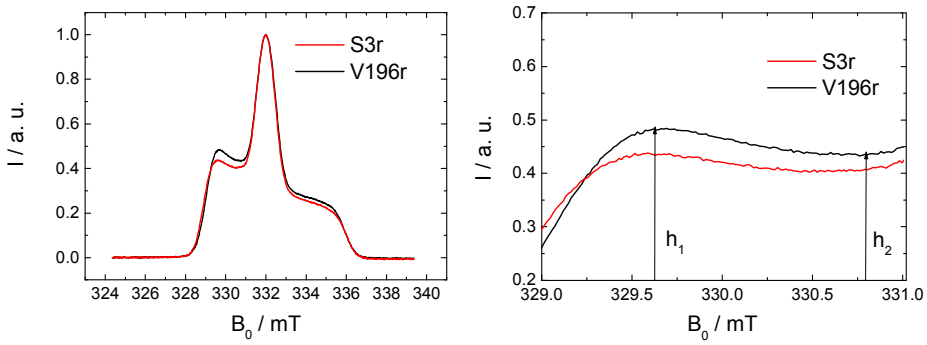


Figure 33: Intensity normalized ESE spectra of different LHCIIb mutants. The ratio of intensities at around 329.6 mT and 330.8 mT was used to determine the product of libration amplitude and libration frequency of the spin label.

The shape of ESE spectra can be explained in terms of spin label libration on a nanosecond time scale **[Dzuba 1996]**. The analysis described in **[Dzuba 1996]** was used in order to get information on the dynamics of the nitroxide label in LHCIIb. Relaxation anisotropy is quantified by the h_1/h_2 ratio of spectral intensities at two positions in the spectrum that roughly correspond to the orientations in the nitroxide molecular frame along the orbital lobes of the nitrogen p_z orbital and perpendicular to it. This ratio is expected to be linearly dependent on the product of the square of the nitroxide libration amplitude and the correlation time of the libration **[Dzuba 1996]**.

The largest error obtained from such analysis is assumed to originate from the selection of h_1 position. The comparison of the h_1/h_2 values at different spectral positions allowed an error estimate of 2%.

In Figure 34 h_1/h_2 anisotropy ratio and longitudinal relaxation rate are correlated with each other for different mutants.

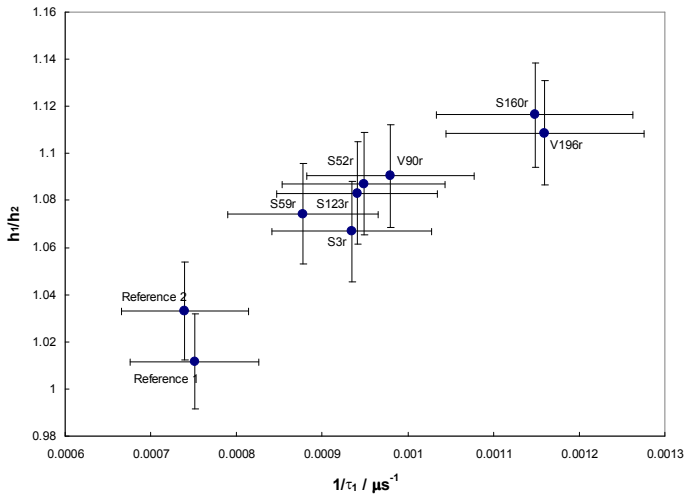


Figure 34: Anisotropy ratio h_1/h_2 for transversal relaxation from ESE spectra plotted against longitudinal relaxation rate $1/\tau_1$ for different LHCIIb mutants and two reference samples.

An increased h_1/h_2 value is observed for the mutants S160r and V196r.

From the previous measurements it is known that these positions are also less solvent accessible. This would mean that the spin label in these positions librates faster or with larger amplitude than in reference samples and some other mutation positions. It should be noted that a similar analysis was not possible with the deuterated samples. The reason for that may be the echo modulation by deuterium nuclei.

The correlation between h_1/h_2 and τ_1 is good, and thus it can be assumed that both parameters are related to stiffness of the solvation cage around the nitroxide spin label.

In order to separate the libration frequency effects from the amplitude effects and to get deeper insight into the local spin label dynamics, W-band longitudinal relaxation measurements were performed. It is known that in the temperature range studied longitudinal relaxation in W-band is determined by a second order Raman process **[Eaton 2001]**, **[Leporini 2003]**. If the relaxation time is determined by a Raman process, which is known to be frequency independent **[Abragam 1961]**, than the trends in the τ_1 relaxation parameter for LHCIIb sample in W-band should be the same as in X-band.

Table 3 shows these τ_1 values for different UZ folded mutants.

Mutant	τ_1 / μs
S3r	993
S160r	800
V196r	716

Table 3: Effective longitudinal relaxation time (τ_1) for different LHCIIb mutants in W-band.

The trends of τ_1 in W-band are similar to the ones obtained in X-band, thus supporting the Raman origin of T_1 relaxation also in the spin-labelled proteins under these conditions.

As the frequency of molecular libration does not influence the relaxation properties due to the Raman origin of the relaxation mechanism, it is possible to assign the changes in τ_1 as a function of position to changes of librational amplitude.

Whether actual motion is the intramolecular libration **[Du 1995]** or the collective motion of the solvent matrix **[Kirilina 2004]**, cannot be clearly distinguished. The experiments to distinguish the two possible relaxation mechanisms, as described in **[Kirilina 2004]**, can be of interest.

In summary it is possible to say that the relaxation measurements on proteins can provide valuable information about the local spin label environment and spin label dynamical properties, and therefore are important sources of information for membrane protein structure studies.

4.1.5. DEER

As with the help of DEER spectroscopy intramolecular distances can be measured, this technique provides a convenient tool for obtaining structural information about LHCIIb on larger length scales than by the methods presented previously.

4.1.5.1. New monomer mutants

In order to get maximum information about LHCIIb, and later to get the most detailed information about its folding mechanism, new double mutants, in addition to the ones described in **[Jeschke 2005]** were produced as described in the experimental part.
The new mutation positions were pre-selected so that conformational freedom of the spin label was restricted as much as possible by neighboring amino acids. This should lead to a narrower distance distribution that allows exacter distance determination. The pre-selection was based on conformational flexibility analysis of the spin labels attached in the computer to the X-ray structure of LHCIIb **[Polyhach, Ye., Volkov, A., Dockter, C., Paulsen, H., Jeschke, G. unpublished results]**.
The new double mutants with corresponding dipolar DEER data, as well as the distance distributions obtained, are shown in the Figure 35.
The background correction parameters were as following: 376 ns for S59r/V90r mutant, 224 ns for V90r/V196r mutant, 280 ns for V90r/S123r mutant, 208 ns for S123r/V196r mutant and 320 ns for S106r/S160r mutant.

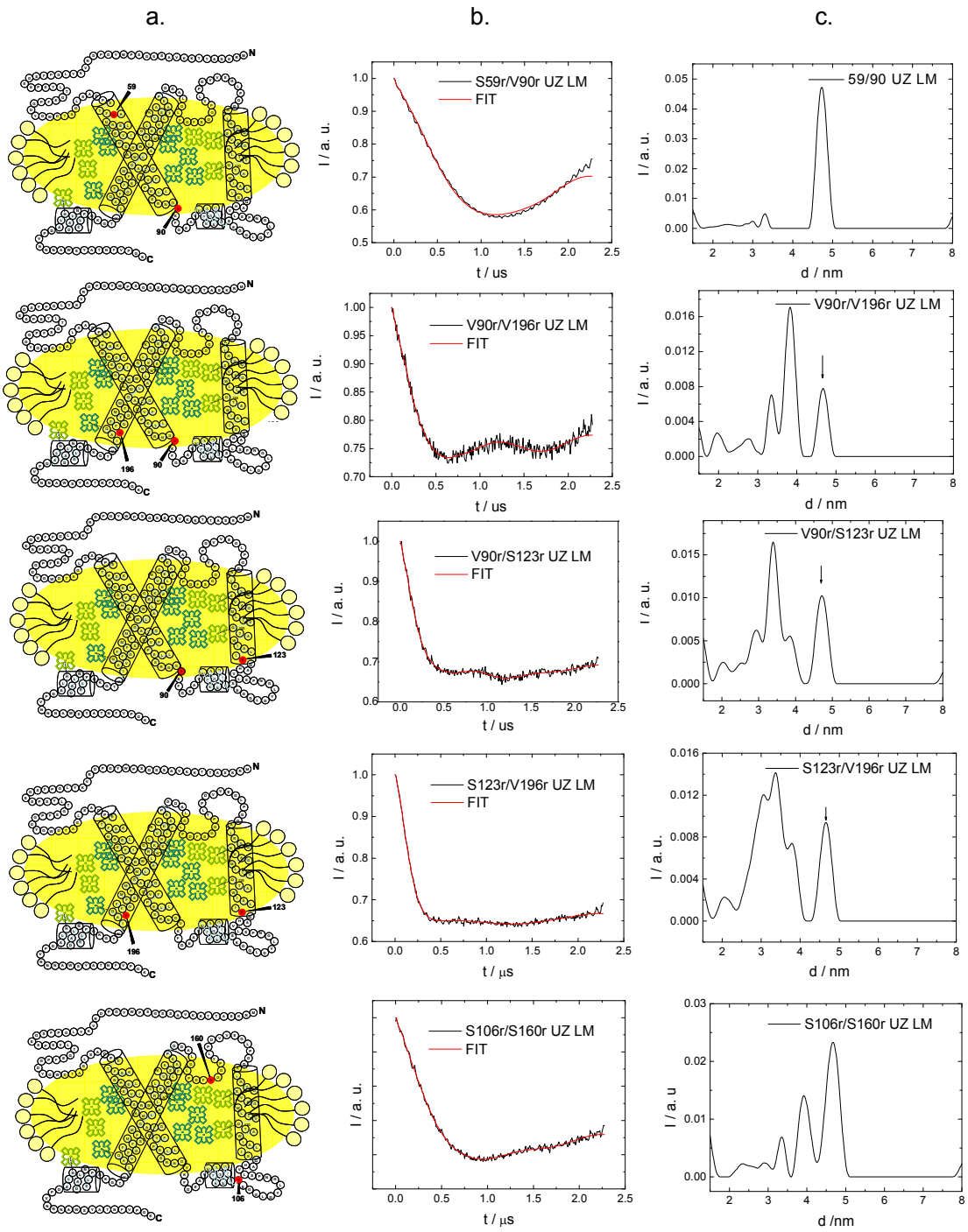


Figure 35: DEER data of new LHCIIb double mutants. a. Studied double mutant, b. Background corrected dipolar data assuming a three-dimensional spin distribution for background correction (the red curve represents the fitting function), c. Distance distribution obtained by Tikhonov regularization.

From Figure 35 b. it is seen that the fits of the dipolar evolution data are reasonably good. Different double mutants show as expected different distance distributions. The mutants S59r/V90r and S106r/S160r thus allow for the determination of a distance along a single helix while V90r/V196r, V90r/S123r and S123r/V196r allow for distance measurements between neighboring helices.

The S106r/S160r mutant shows a similar distance distribution as described before **[Jeschke 2005]**. Differences between the earlier measurements using a TEMPO-based spin label and the new measurements using the slightly smaller and more stable PROXYL spin label are minor.

The simulation of the distance distribution between the spin labels in the particular positions using a rotamer library **[Polyhach 2008]** shows good correlation with experimentally determined data proving the correctness of protein preparation, measurement conditions and data analysis procedure. Such good correlation also legitimizes the use of a rotamer library based approach for the simulation of DEER spectra.

Differences between the measurement and theoretical prediction were however observed for S59r/V90r mutant where the theoretical distance distribution is broader, than the measured one, and S123r/V196r mutant, where the peak between 4 and 5 nanometers that can be an artifact as it gets smaller when the experimental data is analyzed with the other type of background correction function (see later this section) is present in the theoretical distance distribution (Figure 36) **[Polyhach 2008]**.

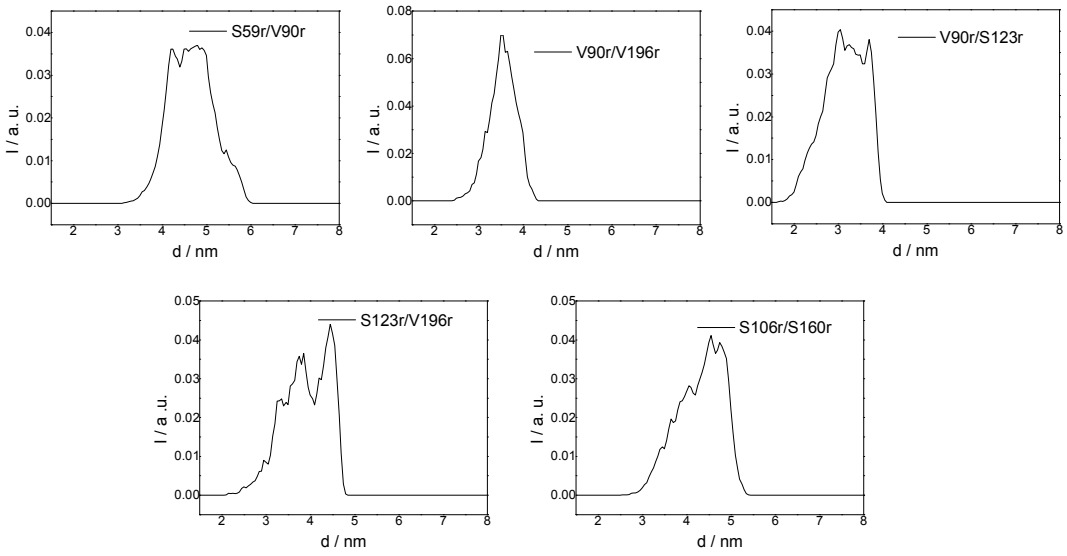


Figure 36: Distance distributions determined from a rotamer library based simulation.

The experimental distance distributions of many double mutants show a peak between 4.5 nm and 5 nm marked with an arrow in Figure 35 c. As the reproducibility of this peak was good, its origin was investigated.

From Figure 35 b. it is seen that the fit on the DEER data at long evolution times is not very good. The reason for that may be the background correction procedure. In order to proof this hypothesis the following experiment was performed:

A DEER experiment on a mixture of V90r and V196r single mutants (60 μM) was performed. In such a mixture no intramolecular distances are expected, and thus this measurement should represent the perfect background signal, as no dipolar modulation should be observed. The obtained DEER data for this sample is shown in Figure 37. The background correction parameter was 224 ns.

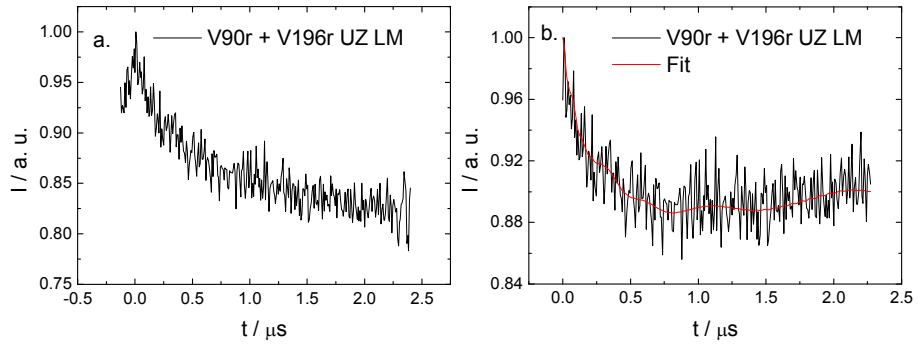


Figure 37: DEER data for the V90r and V196r mixture of single mutants. a. original data, b. background corrected data.

The data of the mixture of single mutants cannot be fitted completely by a background function corresponding to a homogeneous distribution of LHCIIb, showing that this function is not perfectly suitable for background correction.

As there is some dipolar frequency in the mixture of the single mutants, which was not expected, the effect was investigated in more detail.

In many cases during this work different states of sample had to be compared with respect to a change in the distance distribution, while the detailed shape of the distance distributions was not needed, or was not possible to be determined due to the insufficient signal-to-noise ratio. In such cases it proved to be reasonable to compare the dipolar spectra obtained by Fourier transformation of the background corrected time-domain DEER data, as they are less subjected to error than distance distributions.

At the same time differences between different datasets are easier to recognize and quantify in dipolar spectra than in time-domain data.

In the Figure 38 the normalized dipolar spectra of V90r/V196r mutant and mixture of V90r and V196r mutants are shown. Background correction time was 224 ns.

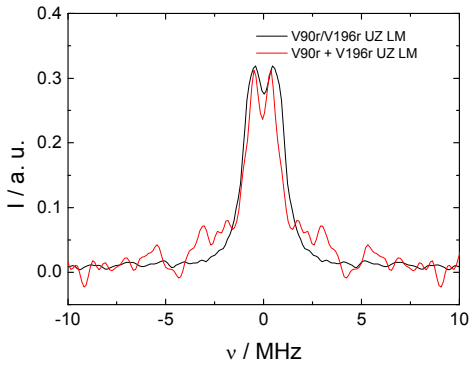


Figure 38: Comparison of normalized DEER dipolar data for the V90r/V196r and mixture of V90r and V196r mutants. Normalized on integral from -10 to +10 MHz. The slight zero frequency shift of the red curve is due to the small error in an older version of the "DeerAnalysis2006" program.

Figure 38 shows that the frequency of the dipolar modulation in the sample consisting of mixture of V90r and V196r mutants does contain similar components as in the case of V90r/V196r mutant, however the distribution of frequencies is much broader and also other dipolar frequencies are present.

Furthermore the background corrected data in Figure 37 b. has much smaller modulation depth than for the V90r/V196r double mutant. The modulation depth depends on the number of coupled spins **[Milov 1984]**, **[Schweiger 2001], [Hilger 2005]**, **[Bode 2007]**. Hence it is possible to conclude that although the number of coupled spins in the mixture of single mutants is much smaller than in the double mutants there is still some aggregation that may lead to spurious features in the dipolar data.

The described observations poof that the effects measured in the double mutants actually originate from intramolecular distances.

For the empirical background correction that takes into account the deviation from a homogeneous distribution of the spins in the sample the data in Figure 37 a. was fitted with a 4th order polynomial (not shown). The fitting was performed from 128 ns to 2400 ns, with background correction starting at 128 ns. This polynomial was saved, and used as a background correction function of the DEER data of double mutants. The V90r/V196r, V90r/S123r and S123r/V196r data analyzed in such a way with the same parameters as in Figure 35, however with the experimentally determined polynomial as a background correction function are shown in Figure 39.

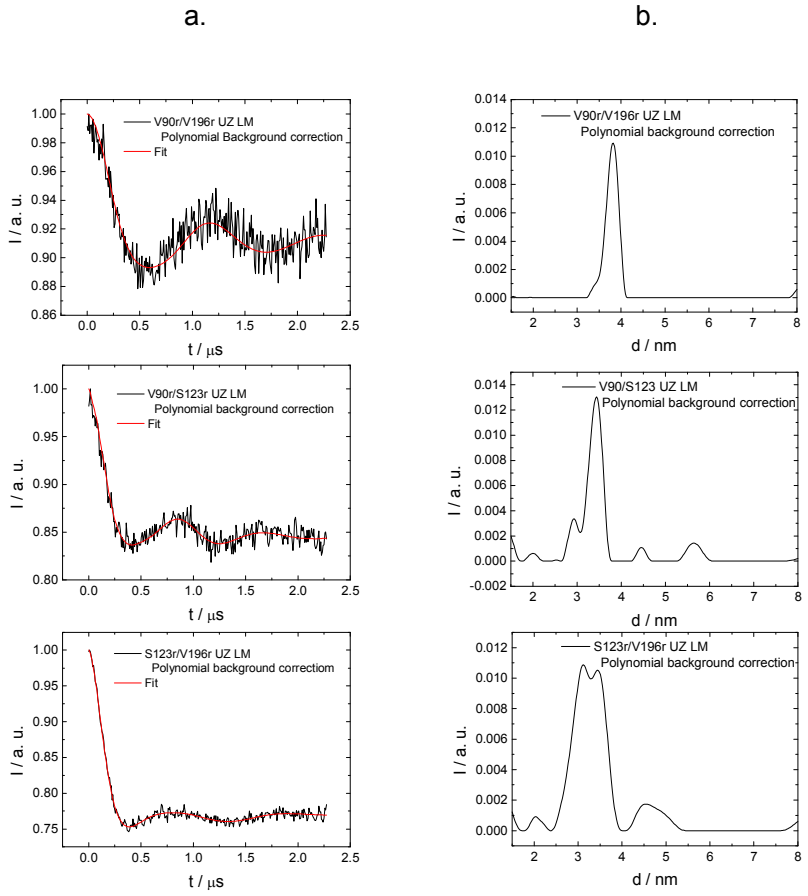


Figure 39: DEER data for V90r/V196r, V90r/S123r and S123r/V196r mutants obtained by background correction with an experimentally determined polynomial from a mixture of the single mutants. a. background corrected DEER data. b. determined distance distribution

From Figure 39 it is seen that the subtraction of an experimentally determined polynomial as a background correction function provides data with very marked dipolar modulation for all of the shown mutants. The distance distribution obtained from this modulation data shows the same features as in Figure 35, where the three dimensional model was used for background fitting, however the peak between 4.5 nm and 5 nm is reduced for all the samples confirming the hypothesis of its artifact origin. Also the small peaks between 1.5 and ca. 3 nm are reduced, which also hints on their artifact origin. The highest peaks in the distance distribution remain independent on the background correction function. They can thus be safely assigned to real intramolecular distances in protein. This legitimizes the use of the theoretical three-dimensional background correction function for qualitative discussions and semi-quantitative analysis.

A similar background correction procedure for the S59r/V90r and S106r/S160r mutants provided datasets that could not be fitted with good accuracy (data not shown). The reason for that can be that the distribution of the spins in the mixture of single mutants is not completely independent on mutation position.

That is why for further distance data analysis still the theoretical background correction function assuming a three-dimensional homogeneous distribution was used, although some artifacts can be introduced by that. A further reason for using the theoretical background correction is that it is not possible to measure the experimental background correction function for RR TX mutants used for folding experiments due to the ill-defined sample composition.

The DEER measurements presented in this work allowed to increase the structural information on LHCIIb in a conformation that is expected to be similar to the one in thylakoid membranes. This structural information proved the suitability of the sample preparation, measurement and data analysis procedures. The measurements also allowed to test the models for the theoretical prediction of the DEER spectra based on a rotamer library.

The obtained information will be further needed for the investigation of LHCIIb folding.

4.1.5.2. Reproducibility of DEER experiments.

In order to further use the DEER experiments to characterize structural changes in the protein the reproducibility of the data has to be controlled.

Figure 40 shows the dipolar spectra of different double mutants that, as described previously, allow for easy comparison of distance distribution data.

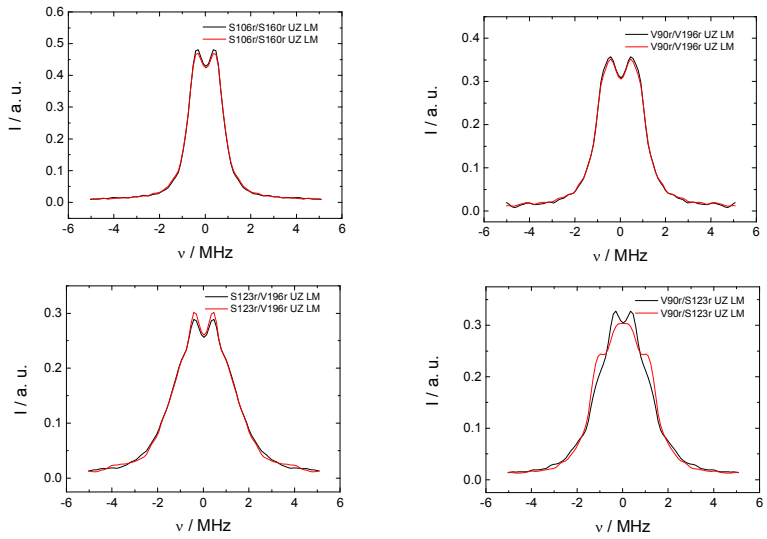


Figure 40: Proof of reproducibility of DEER experiments for different LHCIIb double mutants. Normalized on integral from -5 to 5 MHz.

The reproducibility of DEER experiments for the UZ LM samples seems to be sufficient for the presented mutants, although concentration and repetition rate were not optimal for some of the mutants shown.

The only exception was the V90r/S123r mutant, where significant differences in the dipolar data between the samples were observed. The discrepancies between two different V90r/S123r preparations can be explained in terms of marginal stability of helical structure in this region.

Hints for that are found in the X-ray structural data. The structure presented by **[Liu 2004]** shows no helix in the selected region, whereas the structure presented by **[Standfuss 2005]** does. As the crystallization buffer composition between the two different crystal preparations differed, it is possible that the fragile region between positions V90 and S123 was distorted by preparation of the crystal in one of the published structures.

Hints for the changes in the protein structure in this region were also found in the longitudinal relaxation data (τ_1) of the S106r mutant. Sometimes longer and sometimes shorter longitudinal relaxation times were observed meaning that the accessibility of this mutant position changed from sample to sample. The samples with smaller accessibility also showed a small shoulder in CD spectra at about 475 nm (data not shown).

These observations show the possibility of using EPR to study conformational changes in the protein without having protein crystals. Thus DEER and also the local accessibility measurements can be important in pre-scanning the lipid compositions to see whether there are structural changes before protein crystallization. This can be useful in finding the best suitable lipid composition for protein crystallization.

4.1.5.3. New trimers

During this work also some data on LHCIIb trimers was obtained. This information is important as the trimer represents the native state of the protein **[Liu 2004]** and direct correlation with X-ray data is possible **[Liu 2004]**, **[Standfuss 2005]**, which is not exactly the case for monomer samples. As DEER spectroscopy provides information about the protein in a glassy matrix, it is of interest to compare the DEER data with the X-ray data in order to check whether the crystal packing effects change the protein conformation.

With many of the spin-labelled mutants prepared in this work trimerisation of LHCIIb monomer samples was not possible, which may be explained by the distortions introduced by mutations on the trimerisation motif.

Figure 41 shows the DEER data obtained on a new UZ LM LHCIIb trimer sample. The background correction was 336 ns.

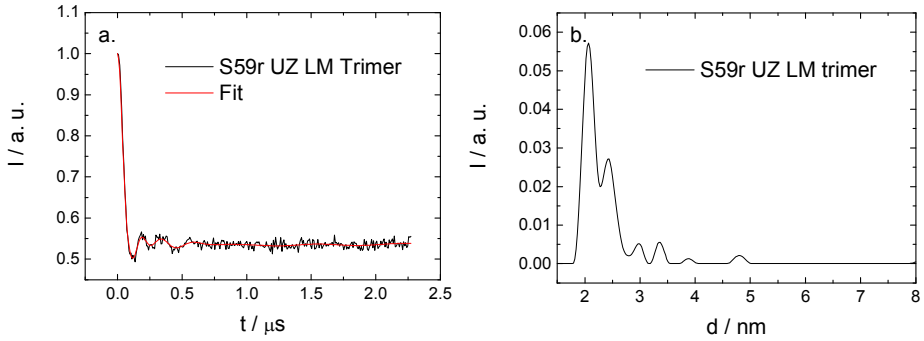


Figure 41: Background corrected DEER data (a.) and distance distribution (b.) obtained on a S59r mutant trimer.

DEER data for this S59r trimer mutant exhibits several oscillations corresponding to a narrow distance distribution that can be directly compared with X-ray crystallography data.

The distance distribution data predicted by the rotamer library approach **[Y. Polyhach, unpublished results]** correlates satisfyingly with the experiment as can be seen by comparing Figure 41 b. with Figure 42.

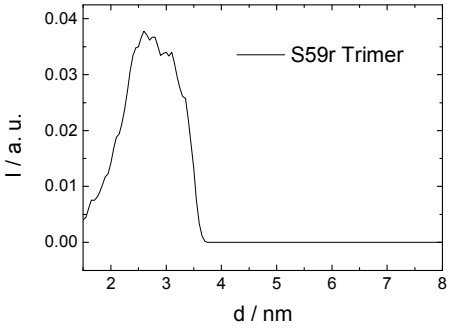


Figure 42: Simulated distance distribution for the S59r mutant trimer

This demonstrates that the rotamer library approach is suitable also for LHCIIb trimers. The experimentally determined distance distribution however seems to be sharper than the one theoretically predicted on the basis of the X-ray structure. This indicates that the rotamer library approach overestimates flexibility of the spin label in this case.

4.1.5.4. Change of LHCIIb structure as a function of micelle composition

Monomers

The adjustment of the membrane proteins to their lipid environment is still a topic of discussion **[Bowie 2005]**. In this section it is shown how with the help of the DEER technique this question can be addressed.

Information on geometry change of LHCIIb monomers as a function of the micelle composition is shown in the Figure 43.

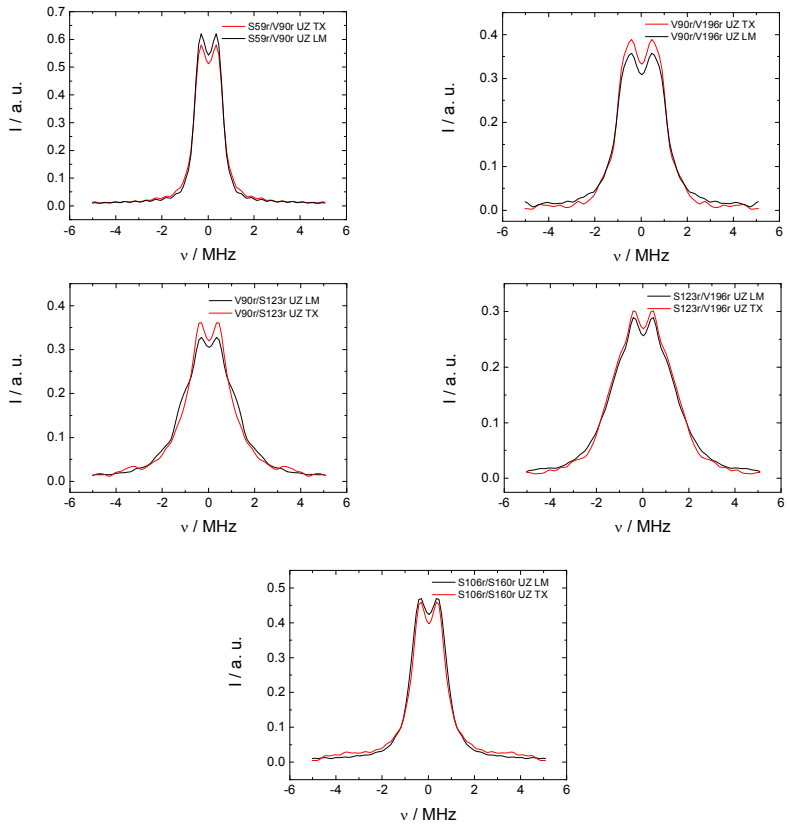


Figure 43: Effects of different micelle composition on the distance distribution of different LHCIIb monomer mutants. Intensity is normalized by the integral from -5 to 5 MHz.

From Figure 43 it is seen that there are small but not negligible structural changes for all the investigated double mutants when different detergents are used. The most significant changes are observed for the V90r/V196r and V90r/S123r mutants. This observation hints to the largest structural change at position V90r. Unfortunately as the reproducibility of V90r/S123r mutant was not sufficient, it is not possible to definitely assign the structural change at different lipid compositions to this position.

Interestingly no significant changes between the fluorescence spectra of LM and TX mutants could be observed.

The CD spectra also do not differ strongly from each other however some minor changes are still recognizable as can be seen from the Figure 44.

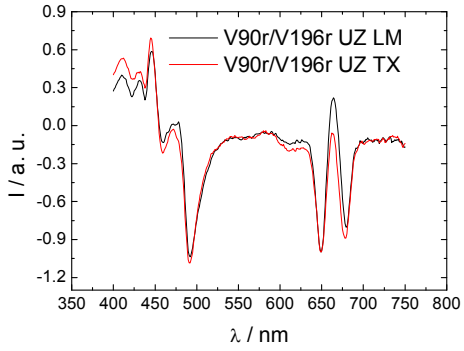


Figure 44: Change in CD spectra as a function of micelle composition. Intensity is normalized by the amplitude of the peak at 649 nm.

The fact that protein structural changes that are not visible in fluorescence spectra and cause only very small changes in CD spectra can be observed and partially explained by pulse EPR, proofs the strength of this spectroscopic technique for this type of protein structural investigations.

Trimers

After distance change in monomers was observed, although it was not expected from CD and fluorescence spectra, change of the distance distribution in trimers as a function of micelle composition was investigated.

The fluorescence spectra of LM and TX trimer samples do not differ from each other in their shape. However the decreased energy transfer in TX micelles is observed, which hints on some structural changes. Unfortunately the same experiment could not be performed on the monomers. A more detailed discussion of this effect can be found in **[Dockter 2008]**.

In the case of LHCIIb trimers also changes in CD spectra are observed as a function of different micelles **[Boggash 2006]** as can be seen from Figure 45

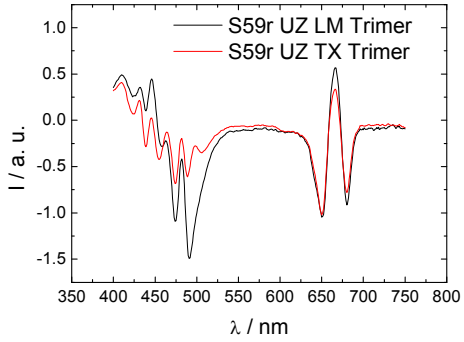


Figure 45: Difference in CD spectra for LHCIIb trimers as a function of micelle composition. Intensity is normalized by the amplitude of the peak at 649 nm.

These differences suggest a change of the protein structure in different micelles.

The DEER experiment was performed on these samples in order to check this assumption. In Figure 46 the dipolar spectra of the trimers analyzed with identical parameters in solvents with different micelle compositions are shown.

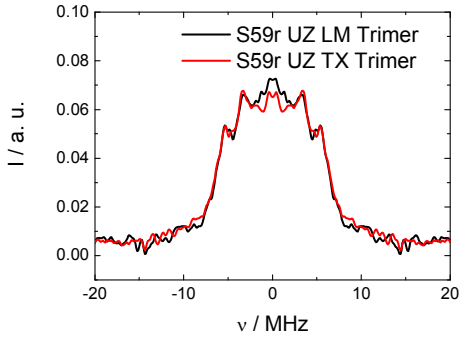


Figure 46: Effects of different micelle composition on the distance distribution in trimers

From Figure 46 it is seen that for the S59r trimer no significant changes in dipolar spectra are observed, meaning that no changes in distance distribution occur in this protein region. These EPR measurements provide a hint that the changes in the structure seen in fluorescence energy transfer and CD measurements happen in the lumenal region of LHCIIb trimer. Note that the changes in monomer structure were also observed in this region. Thus with the help of CD and fluorescence measurements it is possible to say *that* the change of geometry happens, whereas with EPR and site directed spin labeling it is possible to say *where* this change takes place or does not take place.

The information obtained with the presented methods is of high importance as the lipids are assumed to be an important factor in giving the membrane protein its final shape **[Bowie 2005]** and thus a more detailed picture of lipid protein interaction can be obtained.

As already noted in the section 4.1.5.2., EPR can also be a helpful method in the pre-selection of micelles for protein crystallization, as the micelles that lead to narrow reproducible distance distributions should also provide better crystallization results.

4.1.6. Comparison between methods.

The EPR methods used in this work can be separated in two types:

- Methods for local environment (up to 2 nm) and dynamics determination.

These methods include CW EPR, CW progressive power saturation, ESEEM, and relaxation measurements.

- Methods for global structure determination (up to 8 nm).

The method used here is DEER spectroscopy.

As to the methods for characterizing local environment and dynamics, well established CW EPR methods were used together with rather modern pulse EPR methods. Their simultaneous use allows to significantly increase the information content that can be obtained with EPR on membrane proteins, as these methods are based on different physical interactions. The choice of the described methods depends on the complexity of the problem, desired exactness and laboratory restrictions.

CW EPR measurements allow for the determination of solvent accessible amino acid residues in LHCIIb membrane proteins as A_{zz} is sensitive to solvent polarity. The advantage of this method is the low technical complexity and the straightforward data analysis.

A big disadvantage of this method is that large changes of the dielectric constant induce minor effects in the CW spectra, and that A_{zz} determination is rather inexact due to the broad lines. Also other effects like line broadening due to the unresolved hyperfine splitting may cause additional errors. Thus the measurement of accessibilities by CW EPR spectroscopy is rather inexact compared to the other methods presented in this work. Another disadvantage of this method is that the effect of charged amino acids in direct neighborhood to spin label that may also change the polarity of the environment cannot be separated from the overall polarity effects. However when other accessibility data is available this drawback of CW EPR spectroscopy can be used as an advantage to obtain hints on the presence of such charged amino acids.

CW progressive power saturation measurements provide accessibility information of the spin label to paramagnetic molecules. The great advantage of this method is that the protein is studied at physiological temperatures and in an aqueous environment, which is not the case for any other method described here. Thus CW saturation measurements provide useful information about arrangement of LHCIIb at physiological temperatures and can be in principle extended to LHCIIb in liposomes and thylakoid membranes.

Similar information as from CW progressive power saturation measurements can be obtained by ESEEM measurements. ESEEM is assumed to be more direct than CW progressive power saturation measurements in determining the solvent accessible protein parts, as coupling to deuterated solvent molecules is measured, and not the collision rate with a bulky metal complex. When using the negatively charged chromium oxalate complex an attraction by positively charged amino acids or lipids, and repulsion by negatively charged amino acids or lipids may be induced. This can influence the collision frequency with the spin label and lead to a systematic error. This is not the case for deuterated buffer. Also no heavy metal ion complexes have to be introduced during ESEEM measurements, and thus the investigated system is less perturbed.

In a way environmental information obtained by ESEEM is similar to CW measurements, however compared to A_{zz} data, the spread of the ESEEM accessibility parameter is much larger, and thus the exactness of position determination is improved. This improved ESEEM sensitivity compared to CW was also reported in **[Noethig 2004]**. Also as a pulse method ESEEM selects one particular interaction from the Hamiltonian, whereas in the case of CW EPR measurements several interactions influence the spectrum, which can be an additional source of errors.

Additional structural information can be obtained from relaxation measurements. In this work it was possible to see the environmental effect on the relaxation also for the medium size membrane protein. The comparison between transverse relaxation rates for deuterated and non-deuterated spin label allowed the water availability determination on the larger scale compared to ESEEM as explained in detail in section 4.1.4. The simultaneous use of ESEEM and transverse relaxation measurements provides detailed information about accessibility of the spin label in the range from 0.35 to 2 nm. The excellent complementation of ESEEM and transverse relaxation measurements is illustrated in Figure 47.

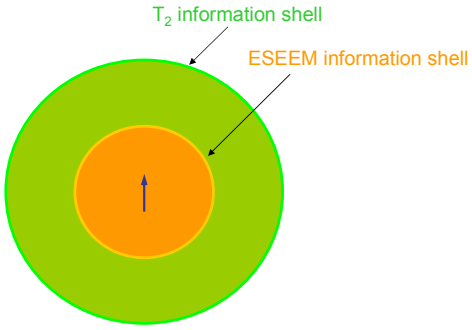


Figure 47: Complementation of ESEEM and transverse relaxation (T_2) measurements of water accessibility. The blue arrow represents an electron spin.

The simultaneous use of these two techniques may be particularly important for environmental localisation of residues that are on the membrane water interface, as these positions should be water inaccessible on a short length scale, and should show water accessibilities on long length scales. Some hints for that could be observed on the example of S160r mutant (Figure 48).

Orientation-selective transversal relaxation measurements at W-band frequencies in principle allow for getting information about the geometry of spin label libration. These measurements provide a hint that the libration of the nitroxide in different protein positions in LHCIIb is isotropic and is not influenced by secondary structure.

The studies of longitudinal relaxation in combination with ESE measurements provide further information about the local environment and dynamics of the spin label. The obtained information can be interpreted in terms of local rigidity of the spin label surrounding. Analysis of ESE spectra shows good correlation with the effective longitudinal relaxation time.

Unfortunately, from these relaxation measurements the helix and loop regions can not be differentiated, probably because the librational amplitude is mostly determined by the stiff cage of ordered solvent molecules around the spin label, and not the amino acid neighbors or secondary structure. Also no correlation between longitudinal relaxation times and the number of possible spin label rotamers at labeled site could be observed probably due to the same reason.

The measurements of longitudinal relaxation at W-band frequencies provided deeper insight into the details of the relaxation mechanism and allowed to interpret the changes of relaxation as a function of mutation position in terms of a change of libration amplitude.

If the explanation of the dominating contribution to longitudinal relaxation is indeed correct, then the experiments shown here may be used to determine the accessibility of the spin label to water. As local density of water molecules in the direct neighborhood to the spin label can be probed, it may also be possible to use this method to determine the availability of protein pockets near the spin label.

From the results presented in this work it is possible to separate effects of local proton concentration, determined by the measurement of the effect of deuteration on transversal relaxation, from local spin label dynamics in the glassy environmental cage determined by measurements of longitudinal relaxation time.

The disadvantage of the relaxation measurements is that the nature of these processes and their theoretical description is complicated. They are thus analysed phenomenologicaly and the obtained data should be interpreted with caution.

In order to better compare the different methods for determination of spin label accessibility, parameters from all the experimental techniques that proved to give this information were plotted against the three-pulse ESEEM accessibility parameter, which is assumed to give the most direct information on water accessibility. Also the correlation between ESEEM accessibility and theoretical accessibility determined with the “Deep View” program based on the **[Standfuss 2005]** crystal structure were compared. Figure 48 shows these correlation plots.

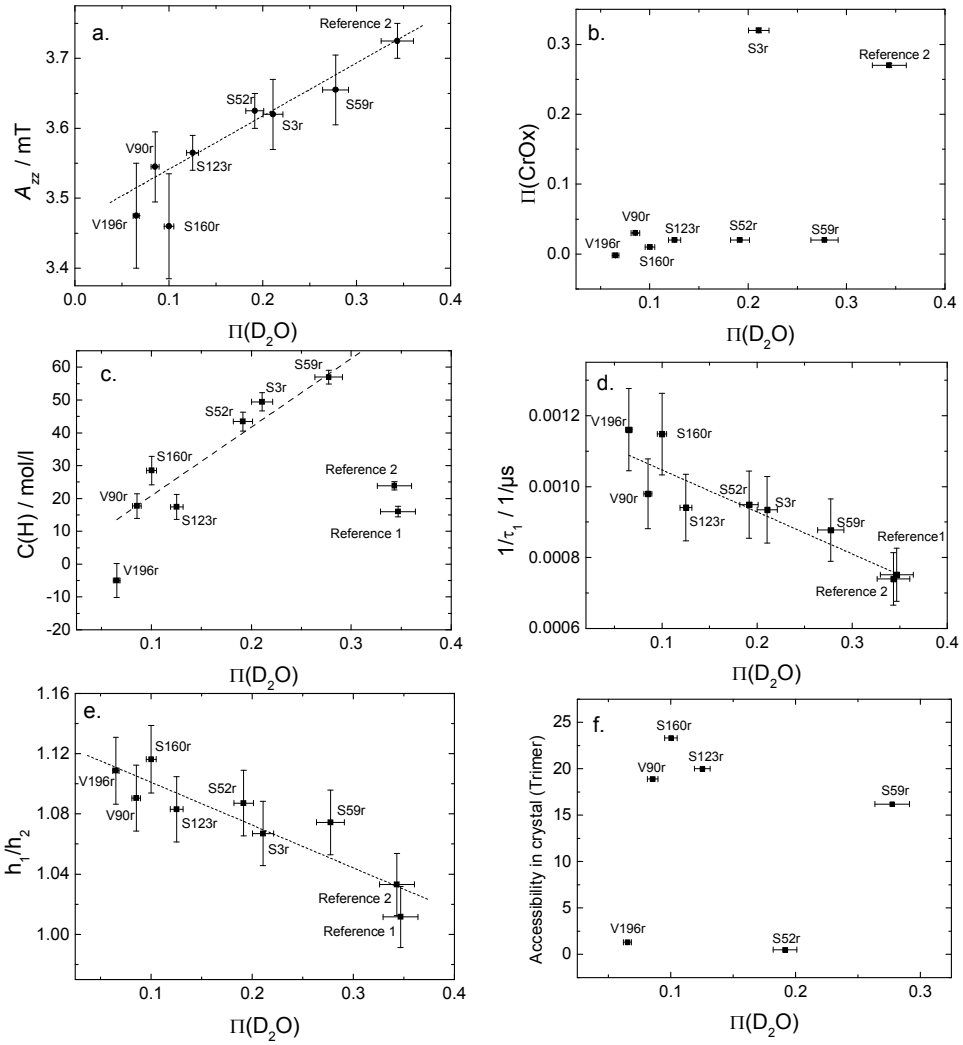


Figure 48: Parameters from different local accessibility determination methods plotted against deuterium accessibility parameter determined by three-pulse ESEEM ($\Pi(D_2O)$). a. A_{zz} from CW EPR vs. $\Pi(D_2O)$ b. Π(CrOx) from CW progressive power saturation measurements vs. $\Pi(D_2O)$ c. Proton concentration at intermediate distance from transversal relaxation measurements vs. $\Pi(D_2O)$ d. longitudinal relaxation rate vs. $\Pi(D_2O)$ e. Relaxation anisotropy parameter from ESE measurements vs. $\Pi(D_2O)$ f. Accessibility determined from the crystal structure of trimer vs. $\Pi(D_2O)$. The dashed line is a linear regression of the data (through zero point by c.). If the error bars are not shown, they are assumed to be smaller than the data points. No error for the theoretical accessibility prediction was assumed.

An approximately linear correlation between ESEEM accessibilities on the one hand and A_{zz}, $1/\tau_1$ and h_1/h_2 on the other hand is observed (Figure 48 a. d. and e.). This legitimizes the use of A_{zz} as an approximate parameter for water accessibility of spin labels in membrane proteins. Also the longitudinal relaxation time and ESE can be used for the determination of water accessibility of a spin label however based on a completely different physical principle.

Also correlation is observed between ESEEM accessibilities and local proton concentration determined from deuteration effects on transverse relaxation, meaning that the short-range and long-range accessibility of the spin labels to water are similar for most investigated mutants (Figure 48 c.). The mutant S160r that is expected to reside near the interface shows some tendency to higher proton concentrations on longer scales, as it is expected due to the considerations above. The mutant V196r shows smaller accessibility to water on a large length scale. That may be due to the occasional deuteration of the exchangeable protons in protein in direct vicinity to this position. Such deuteration may be an additional source of errors by ESEEM measurements. The points that also deviate in their relaxation behavior are the two reference samples. This discrepancy may be attributed to a different relaxation mechanism in these samples or to the systematic error introduced by the relaxation curve fitting as these samples, unlike the LHCIIb samples, were showing strongly non-exponential echo decay. These points were thus not included by linear regression.

Interestingly there is no linear correlation of CrOx accessibility data with ESEEM accessibilities (Figure 48 b.). The S3r mutant and Reference 2 sample do show high CrOx accessibilities, as also expected from ESEEM, however mutants S52r and S59r seem to be hardly accessible to CrOx. This is in contrast to all the other measurements shown here. The change of protein conformation during the freezing process may be a possible reason for that. However the comparison of DEER distance data for this protein determined at the same conditions with simulation data based on X-ray crystallography structure showed good correlation (see DEER section) thus excluding this possibility at least for the S59 position.

The effects of dissolved oxygen can also be excluded, as explicit check of sample degassing was performed for ESEEM and relaxation measurements on some deuterated and non-deuterated samples and no significant changes were observed. This is in contrast to an observation made in **[Noethig 2004]**. However the measurements presented here are assumed to be trustable as the main mechanism for relaxation enhancement by oxygen in solution depends on diffusional collisions with the spin label that are suppressed in the solid state. Thus all measurements were

performed without degassing. This is a significant simplification of the experiments, as the preparations tend to foam strongly during this procedure.

A possible explanation for the discrepancy between CrOx and true water accessibilities is that the local CrOx concentration differs from the one of deuterium nuclei thus also influencing the local accessibility. Steric and charge effects of protein residues surrounding the spin label that change the local CrOx concentration can be a reason for that. Thus it is assumed that although saturation measurements represent the trends in spin label accessibility, they are more subjected to systematic errors than ESEEM.

The comparison of the measured ESEEM data with theoretical prediction based on the X-ray structure of the LHCIIb trimer **[Standfuss 2005]** shows very poor correlation (Figure 48 f.). Although some correlation of mutants S59r and V196r was observed, all the other positions of the protein chain did not correlate. As the theoretical accessibility prediction is developed for the soluble proteins, it is assumed, that the effects of micelles on the accessibility is completely neglected in this case. That in turn is assumed to be the main reason for the bad correlation. What can however be concluded from this data, is that the V196r mutant is shielded from the solvent also by its amino acid neighbors. This should also be the case for the S52r mutant. As all the other accessibility measurements showed a rather high accessibility for this mutant, it may be that the micelles change the conformation of this protein part to such extent that it gets solvent accessible.

The bad correlation of ESEEM with accessibilities determined from X-ray crystallography data may be also caused by the crystal packing effects, differences of chemical composition or local changes of the protein accessibility due to the oligomerization.

In comparison to all of the methods listed above DEER provides global protein structure information. This is of high importance as in addition to environmental information intramolecular distances are obtained without significant changes of sample preparation.

4.1.7. New information about LHCIIb structure

In addition to method development new information about the LHCIIb protein structure was obtained during this work.

Information about the structurally not resolved N-terminus that is known to be important for its regulatory function was obtained. The results presented here suggest that it is not buried in the membrane but is most probably highly solvent accessible up to 2 nm. This observation is supplementary to information about N-terminus previously obtained in our group **[Dockter 2005]**, **[Jeschke 2005]**.

The determined accessibility data can be used to compare the structure of LHCIIb monomers in detergent micelles used here with the X-ray structure of LHCIIb trimers **[Standfuss 2005]**. The position S52 that is not in a helix shows high water accessibility. The position S160 that is known to be in the protein loop unexpectedly shows rather low water accessibility. From the polarity analysis of direct neighbors it is not possible to clearly say whether this part of the protein chain is polar or unpolar and thus to predict its preferential location. From the measurements presented here it seems that the unpolar effects seem to dominate. The positions V90, S123, V196 that are expected to be in the transmembrane helices are also showing the expected low solvent accessibility.

Interestingly the position S59, which is as position V196 in a transmembrane helix, shows very high solvent accessibility. A closer look at amino acid sequence shows that the amino acids in direct neighborhood of position S59 are mostly polar. Thus it is possible that this part of the helix is not micelle buried, due to this increased amino acid side chain polarity. The direct neighbors of position V196 are mostly non-polar, thus suggesting this part of the helix to be buried in the micelle. The larger protrusion length of LHCIIb from the membrane on the stromal side was also reported by **[Standfuss 2005]**. It is nice to see that this small effect could also be reproduced with the presented methods with reconstituted LHCIIb monomer in artificial micelles. This also supports the notion of a similarity of micelles to lipid bilayers which was put forward in **[Roth 1989]**.

The accessibility for the S106 position was not as well reproducible as for the other positions in LHCIIb. It was showing good and poor solvent accessibilities although the preparation techniques were the same. These accessibility changes could be correlated with changes in CD spectra. The reason for that may be the marginal helix stability at position S106. However in order to better understand this effect further investigations are needed.

Distance information about the new mutants was obtained with the DEER experiment. That allowed for extending our knowledge about the structure of LHCIIb monomers in a glassy state.

The measurements on LHCIIb trimers gave new information about its structure in a glassy state and allowed direct comparison of the DEER data with the X-ray crystallography data and rotamer library based simulations.

The reproducibility of the DEER data was rather good, which proved the robustness of the sample preparation, experiment and data analysis. Only the reproducibility of DEER of V90r/S123r mutant was not sufficient. This observation was explained in terms of the conformational lability of this region. Other hints for that were obtained by comparison of two different protein crystal structures from X-ray data and longitudinal relaxation data on position S106 as described previously.

The DEER data with the S106r/S160r mutant were still used without any restriction as the distance distribution for this mutant is broad and thus changes caused by marginal helix stability are expected to be negligible compared to the whole distance distribution. Furthermore the distance changes that can originate from marginal helix stability are expected to be small compared to the distance change during protein folding. Therefore also kinetic studies were performed with these mutants.

DEER also proved to be a powerful technique in the study of protein conformation change caused by its interaction with lipids. Change in the distance distribution as a function of different micelle compositions was observed for LHCIIb monomers. The comparison of DEER data with fluorescence and CD measurements showed that in some cases DEER is more sensitive to these structural changes and that more detailed information about the structural changes can be obtained.

4.2. Investigation of LHCIIb folding with EPR

After the effects determining the EPR signal of spin labelled LHCIIb were mostly understood, the methods described in the previous sections were applied for the study of the LHCIIb folding/self-assembly process.

4.2.1. Characterization of the samples for kinetic measurements

Before the folding process could be investigated significant effort had to be invested in order to get reproducible folding conditions. The samples for EPR kinetic studies (RR TX) differed significantly in their concentration from the ones used in the kinetic study with optical methods **[Horn 2004]**. That is why the sample preparation method presented there could not be adopted directly. The increased concentration for EPR measurements led to requirements for changes in chemical composition (especially of the detergents) in the sample. As described in the experimental part these samples also showed significant differences in their composition compared to the UZ mutants. The exact procedure of optimization of the new protein preparation is described in **[Dockter 2008]**.

The reproducibility of the folding experiments was rather poor. That is why reliable parameters were needed in order to see whether the protein attained its correct three-dimensional structure. As free pigments and unfolded protein are present in the RR TX samples to some extent even in good preparations no simple sample characterization via fluorescence and CD measurements was possible.

In order to perform fluorescence and CD spectroscopy of the RR TX samples a technique was developed in cooperation with C. Dockter (University of Mainz) that allowed for the separation of folded protein in these samples from unfolded protein and free pigments.

The purification is performed as follows. The RR TX samples were let to fold completely. Then they were dissolved 1:1 with lithium borate buffer (pH 8.1) and concentrated in Amicon ultra centrifugal filter units (30 kDa, Millipore) to the initial volume. This procedure was repeated four times. The sample was then loaded on the ultracentrifugation gradient consisting of 0.6 M sucrose, 0.1 % n-dodecyl-β–D-maltoside, Tris- HCl (pH 7.8) and ultracentrifugated at 230000 g for 16 hours at 4 °C.

The representative CD and fluorescence spectra of the RR TX samples purified as described above are shown in the Figure 49.

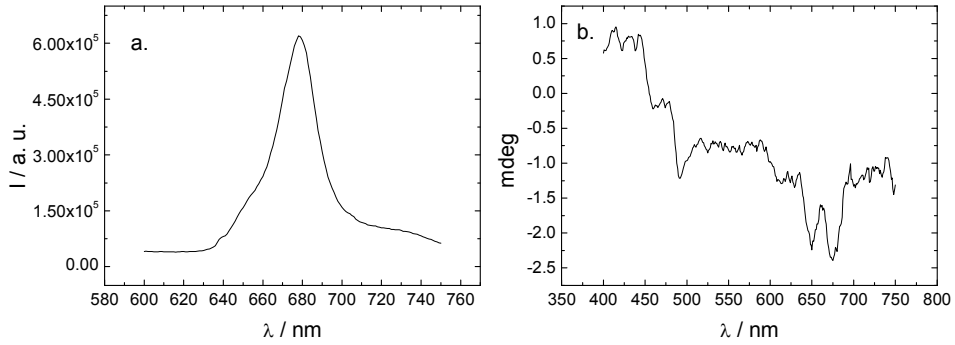


Figure 49: Fluorescence (a.) and CD (b.) spectra of the RR TX mutant after purification

The fluorescence spectrum of the purified RR TX sample shows the characteristic features also observed for UZ samples. However the position of the maximum is shifted by two nanometers. This effect might be caused by changes of sample composition. Also a small shoulder at short wavelengths is observed for the RR TX sample. It is assumed to originate from unfolded protein and shows that the purification procedure is not perfect.

The characteristic peaks in the CD spectrum of the purified RR TX sample are similar to those observed for the UZ sample. The positions of the CD peaks are however not exactly the same. The reason for this shift may be wrongly folded protein as already seen by the fluorescence measurements.

From the measurements shown in Figure 49 it can be seen that in the RR TX samples a significant amount of folded protein is present. From these observations it was concluded that it is possible to use of the new protein preparation method for kinetic measurements. However also unfolded or wrongly folded proteins are probably present in the sample.

Due to the complexity of the purification procedure, a technically less demanding method was needed in order to determine the quality of protein folding. For that the partially denaturating gel electrophoresis as described in **[Paulsen 1990]**, however using a 10% polyacrylamide gel, was used. This method is known to be sensitive to the availability of folded LHCIIb protein **[Paulsen 1990]**.

The characteristic gel of the LHCIIb RR TX samples that were used for the folding studies is shown in the Figure 50.

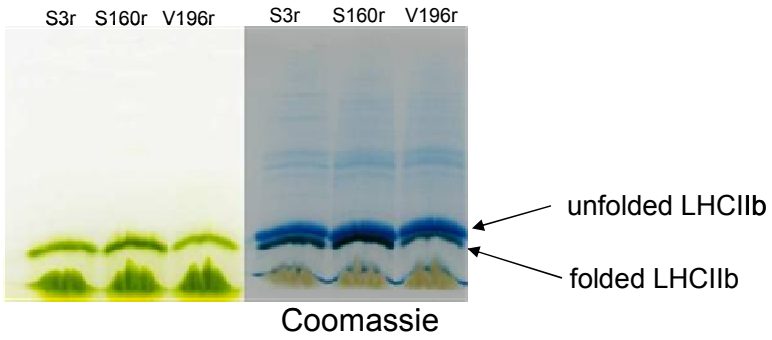


Figure 50: Proof of availability of folded protein in studied samples via gel electrophoresis.

The fluorescence and CD spectra of RR TX samples were correlated with the gel electrophoresis pattern. The samples that showed folded mutants in the fluorescence spectra usually also showed a characteristic band in the gel. The only aberration was the mutant S59r/V90r. That showed a good gel, but very poor CD and fluorescence characterization after the EPR measurement. However as for all the other samples the correlation was good, this result is considered not to be representative, as the S59r/V90r mutant also showed wrong chlorophyll a to chlorophyll b ratio, and changes in the distance distribution data (see section 4.2.1.1.) which hints on structural distortions.

Thus the main biochemical characterization technique for the quality of folding of the RR TX samples was gel electrophoresis. It was used routinely for all the samples for kinetic measurements. Only the samples that showed a band in the gel as shown in Figure 50 were used for the kinetic studies.

The disadvantage of this method as sample test during the protein folding study is that part of the protein molecules are being denaturized on the gel. It was also observed that different mutation positions can destabilize the LHCIIb to different extent. That is why the S160r mutant has a more intensive band than V196r mutant in Figure 50. Thus just a rough estimate of the amount of folded mutant is possible by this method.

It is however necessary to have a technique that would at least allow to determine whether the amount of folded protein changes from sample to sample. The differing amount of unfolded protein during the time resolved measurements can cause an error

in folding time constant determination as extra signal in the unfolded sample region can not be unambiguously assigned. Thus further characterization methods that also allow the determination of the amount of folded mutant were developed.

As EPR showed to be a powerful technique to determine small protein changes (see section 4.1.5.2.) it was utilized to find reliable parameters to characterize the quality of protein folding and its reproducibility.

The parameters that proved to be sensitive to the protein folding quality were the ESE-detected EPR spectrum and transversal relaxation time. The samples, whose ESE spectrum and effective transversal relaxation time were strongly deviating from the reference sample (for the definition of reference sample see next page) were not used. Thus these measurements were performed routinely for each protein sample used in the kinetic measurements. Small differences in effective transversal relaxation time could in principle be tolerated, as with the length of the DEER pulse sequence the fast relaxing components should not be visible. Such samples were still not used in this work as their composition is ill defined.

The explanation for the sensitivity of ESE and effective transversal relaxation time in determination of protein folding quality is that the aggregates that were shown to be characteristic for the RR TX unfolded state (See sections 4.2.3.1., 4.2.4., and 4.2.5.1. for details) have a very characteristic relaxation mechanism (see section 4.2.4.) that leads to decreased relaxation times. The strongly coupled spins of the aggregates also have an influence on the ESE spectra (see section 4.2.4.).

It was sometimes observed that samples that were well folded according to gel electrophoresis did not show the desired ESE and effective transversal relaxation time properties. These samples were not used for the protein folding investigation, as it is assumed that they have a large amount of unfolded protein that cannot be quantified sufficiently well by the gel, but that is visible via relaxation properties. Only the samples with a strong band in the gel for the folded mutant, and ESE and effective transversal relaxation time comparable to the reference sample were used for the kinetic measurements.

The reference samples for the folding measurements were V90r/V196r RR TX after 930 seconds folding and S106r/S160r RR TX after 929 seconds folding. The reason for taking these samples was that the folding is assumed to be finished after this time **[Booth 1996]**, **[Horn 2002]**. Thus direct comparison of the protein geometry by DEER of RR TX samples with the corresponding UZ samples, which are fully structurally characterized, was possible.

Figure 51 shows the comparison of dipolar spectra of the V90r/V196r RR TX sample 930 seconds folding with the V90r/V196r UZ LM sample.

The background correction time was 192 ns for the V90r/V196r sample for an evolution time of 1400 ns and 224 ns for an evolution time of 2500 ns.

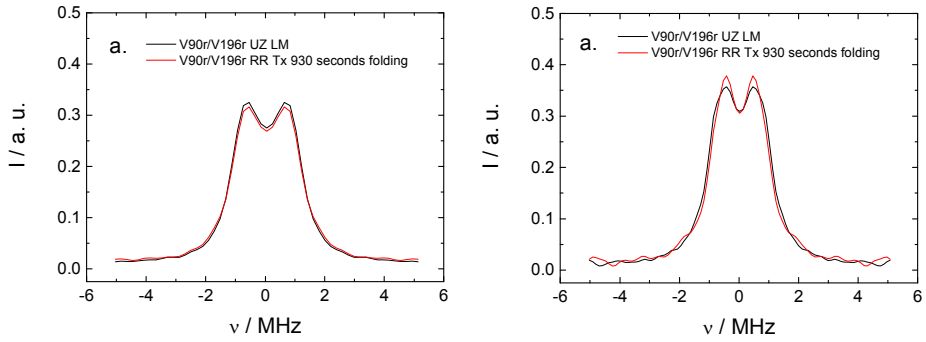


Figure 51: Comparison of DEER dipolar spectra of V90r/V196r UZ LM sample with V90r/V196r RR TX sample after 930 seconds folding. a. 1500 ns evolution time b. 2500 ns evolution time. Normalized on the integral from -5 to 5 MHz.

The spectra at two different evolution times are shown. While for UZ samples the concentration was sufficient to always perform measurements at long evolution times, for RR TX samples, due to a smaller concentration caused by difficulties in sample preparation, a compromise in the length of the dipolar evolution time had to be made. The measurements with longer evolution time allow for a more reliable determination of the distance distribution. Measurements with shorter evolution times are preferable when the sample signal is weak. This comes at the cost that long distances may be partially suppressed in the distance distribution due to imperfect background correction. As longer evolution times are preferable, only such spectra will be compared further, unless noted.

Figure 51 shows that the spectra of RR TX and UZ LM samples are similar. This is further proof for the correct protein structure in the samples prepared for kinetic

measurements. This also legitimizes the use of the ESE and T_2 spectra of this RR TX sample as a reference for the other kinetic samples.

Figure 52 shows the comparison of dipolar spectra of the S106r/S160r RR TX protein sample after 929 seconds folding with the S106r/S160r UZ LM protein.

The background correction time was 320 ns.

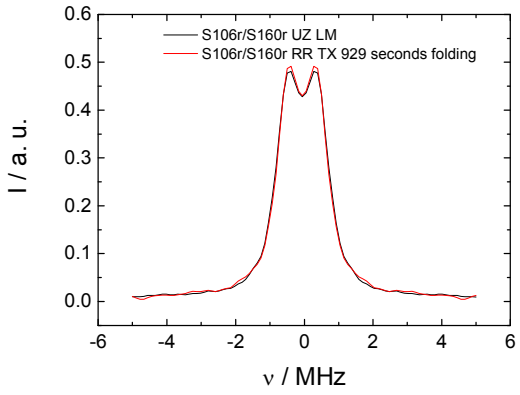


Figure 52: Comparison of the DEER dipolar data of S106r/S160r UZ LM mutant with S106r/S160r RR TX mutant after 929 seconds folding (used as a reference). Normalized on the integral from -5 to 5 MHz.

The similarity of the DEER dipolar data of the RR TX sample after 929 seconds folding time with the UZ LM mutant is obvious. This is further proof of the correct protein structure of the samples prepared for kinetic measurements. This legitimizes the use of ESE spectra and effective transversal relaxation time also of this RR TX sample as a reference.

Although the S106r mutant showed bad reproducibility in its accessibility (see section 4.1.5.2. for details), dipolar data of the S106r/S160r mutant are sufficiently reproducible. That is why the S106r/S160r mutant was used without restrictions for protein folding studies.

In the case of kinetic measurements with single mutants no comparison with the distance distribution between RR TX and UZ LM sample was possible. However as the double mutant measurements proved that the RR TX preparation method provides good distance correlation with UZ samples for mutants V90r, S160r and V196r, these single mutants were assumed to fold correctly and were used for the kinetic study. For the kinetic measurements with single mutants only the samples with good gels and ESE spectra and T_2 relaxation times comparable to the reference sample were used, unless further noted.

4.2.1.1. Mutants not suitable for protein folding investigations.

Unfortunately not all of the studied double mutants showed such a good correlation between distances in the RR TX sample and UZ LM sample as the ones described above.

Figure 53 shows the comparison of dipolar spectra of the S59r/V90r UZ LM sample with the S59r/V90r RR TX sample after 1835 seconds folding (a.), of the V90r/S123r UZ LM sample with the V90r/S123r RR TX sample after 1829 seconds folding (b.) and of the S123r/V196r UZ LM sample with the folded S123r/V196r RR TX sample (c.).

For S59r/V90r mutant the background correction time was 184 ns. For V90r/S123r mutant the background correction time was 160 ns. For S123r/V196r mutant the background correction time was 160 ns.

The dipolar evolution time was 1500 ns for all samples.

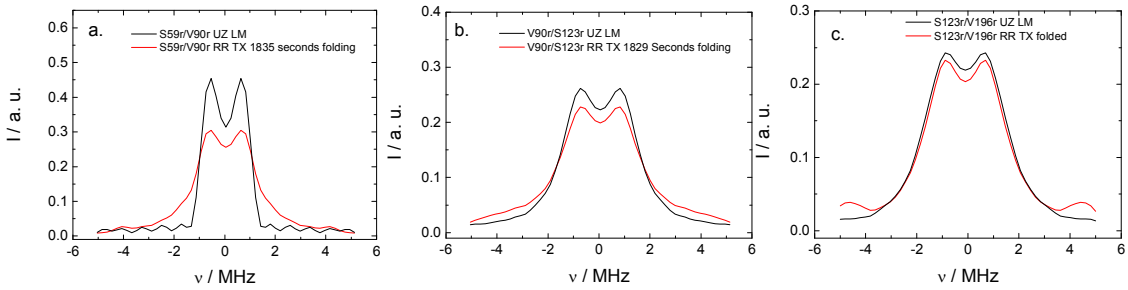


Figure 53: Comparison of the DEER dipolar data of different UZ LM LHCIIb samples with RR TX LHCIIb samples. Normalized on the integral from -5 to 5 MHz.

Figure 53 shows differences between the UZ LM samples and the RR TX samples. The effect is most prominent for the S59r/V90r mutant. This result is rather unexpected, as the gel as well as EPR characterization showed features characteristic for folded protein.

The explanation for the different distances in UZ LM and RR TX samples can be conformational change at these particular positions due to a change in micelle composition, an effect already observed and described in section 4.1.5.4. The reason for such behavior can also be the destabilization or altering of the membrane proteins in some detergent micelles as shown in **[Bowie 2005 and references there]**. The fact that the S106r/S160r and V90r/V196r RR TX mutants showed good agreement with the

UZ LM sample can be explained by the assumption that the protein regions studied with these mutants are less influenced by this specific micelle composition.

The fact that the biochemical characterization methods as well as the comparison of EPR data showed the presence of folded protein once again shows the sensitivity of DEER for detecting small structural changes.

As the folded state could not be well defined for the mutants shown in Figure 53 kinetic DEER measurements of these samples can not be analyzed with confidence.

Indeed, the time-resolved folding data did not show any significant change between different folding states of the S59r/V90r mutant, V90r/S123r and S123r/V196r mutant.

These samples are thus unsuitable for folding experiments.

4.2.1.2. Changes of protein geometry as a function of micelle composition in samples used for kinetic measurements.

As there is a TX detergent in the RR TX sample and it was previously shown that there are conformational changes, when the micelle composition is changed (see section 4.1.5.4.), the RR TX samples were compared with the UZ TX sample in order to learn more about the trends of this conformational change.
Figure 54 shows the comparison of dipolar spectra of the V90r/V196r UZ TX sample with the V90r/V196r RR TX after 930 seconds folding sample.
The background correction time was 224 ns.

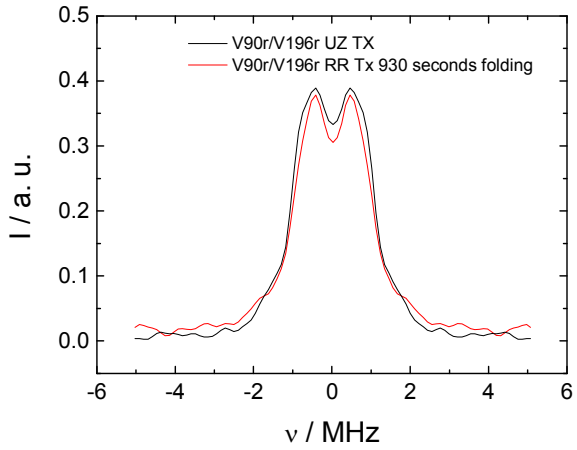


Figure 54: Comparison of DEER dipolar spectra of a V90r/V196r UZ TX sample with a V90r/V196r RR TX sample after 930 seconds folding (reference). Normalized on the integral from -5 to 5 MHz.

From Figure 54 it is seen that the RR TX sample shows minor differences in distance distribution compared to the UZ TX sample. This result is interpreted in terms of a similarity of the structure of the protein in the studied position in RR TX micelles and UZ LM samples as can be seen by comparing Figure 54 with the Figure 51.

Also the S106r/S160r UZ TX sample was compared with the S106r/S160r RR TX sample in order to detect any changes in protein structure as a function of micelle composition. Figure 55 shows the comparison of dipolar spectra of the S106r/S160r UZ TX sample with the S106r/S160r RR TX sample after 929 seconds folding.
The background correction time was 320 ns.

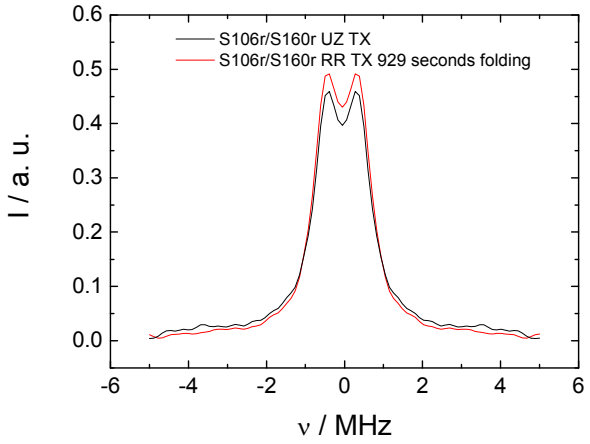


Figure 55: Comparison of DEER dipolar data of a S106r/S160r UZ TX sample with a S106r/S160r RR TX sample after 929 seconds folding. Normalized on the integral from -5 to 5 MHz.

Figure 55 again shows differences between the dipolar spectra of S106r/S160r RR TX and UZ TX samples. This observation again proves that the structure of the RR TX sample is more similar to UZ LM sample than to the UZ TX sample.
The comparison of UZ TX structures with RR TX structures shows that there are no clear trends in protein geometry as a function of micelle composition. It was however shown that with the help of DEER spectroscopy the complicated topic of protein lipid interaction can be addressed with high precision.

4.2.1.3. Changes of protein samples for kinetic measurements in liquid nitrogen.

The samples for the kinetic measurements were shock-frozen in liquid nitrogen in order to stop the folding after a certain time. However in order to be sure that the folding actually did stop after this procedure it was also checked whether the kinetic samples change their conformation in liquid nitrogen (Figure 56). The starting time for background correction was 136 ns and the evolution time was 1500 ns.

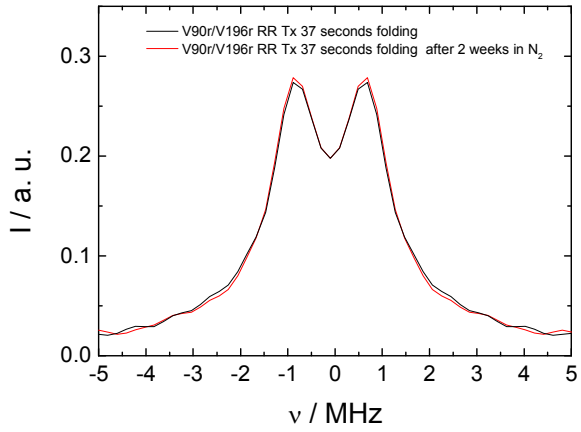


Figure 56: Control experiment, whether the folding takes place in liquid nitrogen.

From the Figure 56 it can be seen, that the shape of the dipolar curve did not change after 2 weeks of storage in liquid nitrogen.

This proves that the protein folding does not continue at 78 K. Furthermore it proves that the sample is not melted to a significant degree during the insertion in the spectrometer. Thus the folding time of the protein is indeed defined by the period between mixing the unfolded protein with the pigments **[Booth 1996]** and inserting it into liquid nitrogen.

There was also no change in the effective transversal relaxation time proving that an altering of the protein does not happen either in liquid nitrogen, or during the insertion in the spectrometer.

4.2.2. CW spectroscopy for LHCIIb folding studies

In section 4.1.1. it was shown that CW spectroscopy can provide information about the local environment of the spin label in LHCIIb by characterizing polarity of the environment by its influence on the ^{14}N hyperfine coupling.

As the composition of UZ samples differs from the one of RR TX samples, only RR TX samples are compared with each other in this chapter in order to exclude solvent effects on the hyperfine coupling.

It is expected that the spin label environment should change during the folding process, and that this should have an effect on the A_{zz} value. There is also evidence for protein aggregation in the unfolded state (see sections 4.2.3.1., 4.2.4., and 4.2.5.1.) and effects of that might be expected to be seen in CW spectra.

From Figure 57 it is clearly seen that CW spectra of the S3r, S160r and V196r mutants in different folding states are practically identical. The fact that the A_{zz} of the folded mutant is the same as for the unfolded mutant means that a polarity change during the folding does not take place or CW EPR is not sensitive enough to detect it under these experimental conditions.

Also the height ratio of the outer peaks to the center peak, which is known to be sensitive to the presence of short distances and to sample concentration [**Kokorin 1972]** does not change as a function of protein folding.

For the S160r mutant also the 1:4 mixture of the unfolded mutant with the unfolded unlabelled mutant showed no difference in CW spectra, proving that there is no effect of aggregation on CW EPR lineshapes.

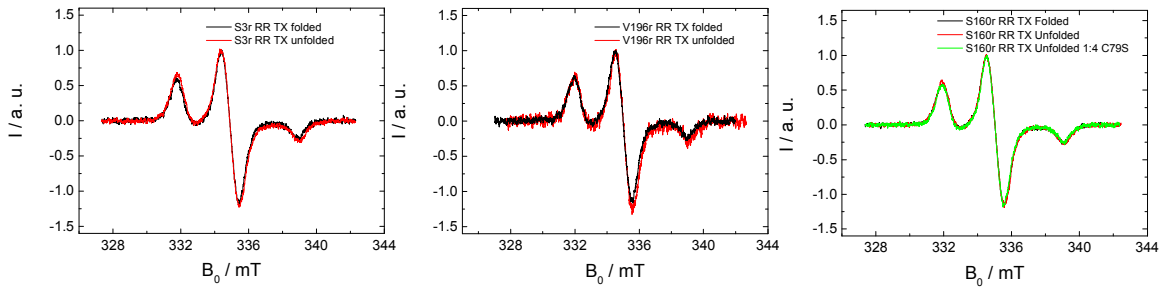


Figure 57: CW EPR spectra of LHCIIb mutants at different folding states. Normalized on intensity of the central line and field shifted to obtain coinciding of maximum field positions.

Having no success with single mutants, CW EPR was applied for the study of folding of double mutants (Figure 58).

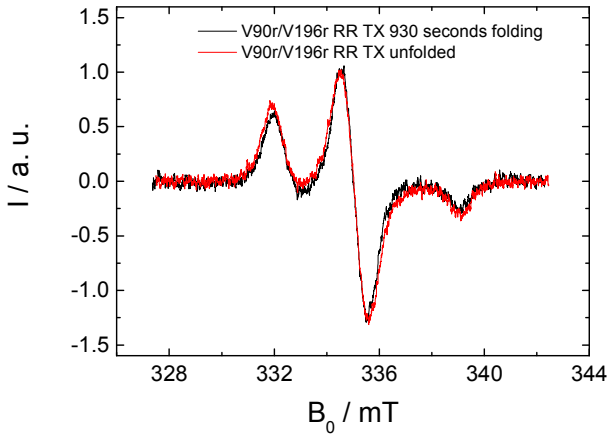


Figure 58: Normalized CW spectra of the folded and unfolded V90r/V196r mutant.

From Figure 58 it can be seen that the CW spectra of the folded and unfolded V90r/V196r mutant differ only insignificantly, and thus can not be used to study protein folding.

The conclusion from these observations is that it is not possible to use CW spectroscopy to study protein folding either for single or for double mutants under these experimental conditions.

The potential of CW progressive power saturation measurements to study protein folding was not investigated, as in section 4.1.6. it was shown that ESEEM is capable of giving similar information and is expected to be more precise.

That is why for further studies of LHCIIb folding kinetics only the pulse EPR methods were used.

4.2.3. ESEEM for LHCIIb folding studies

As explicitly shown in section 4.1.3. ESEEM can be used to study the local water accessibility of the spin label, and is assumed to be the most direct and exact method for this purpose. This property is utilized to study the structure of the LHCIIb in the unfolded state, as well as the local structural changes of the protein during the folding process.

4.2.3.1. ESEEM on unfolded LHCIIb

It is important to have detailed structural information about the unfolded proteins for the easier interpretation of kinetic and thermodynamic data of protein folding, better description of protein folding pathways, verification of the theoretical models for unfolded proteins and eventual understanding of the processes causing protein amyloid formation **[Havlin 2005]** that are suspected to be the reason for a number of human diseases **[Santucci 2008]**.

For that reason three-pulse ESEEM measurements were performed on unfolded deuterated proteins. The noise does not seem to introduce a significant error in the height of deuterium peak in the Fourier transform spectra, which proofs the robustness of the method. The results of these measurements are shown in the Figure 59.

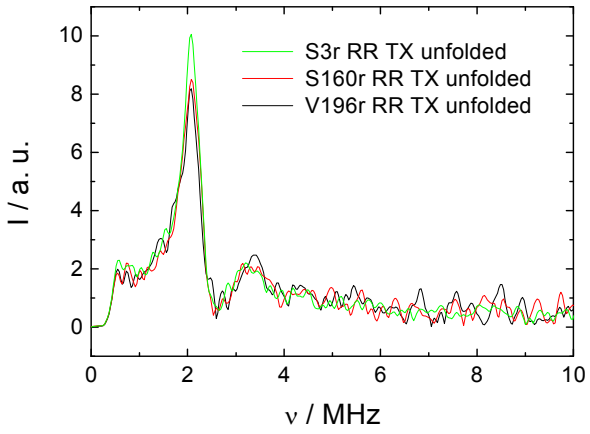


Figure 59: Fourier transform ESEEM spectra of three different unfolded mutants.

From Figure 59 it is clearly seen that S160r and V196r unfolded samples show the same deuterium accessibility. The S3r mutant shows a bit higher accessibility, however the difference to the other mutants is small. Thus it is possible to conclude that the unfolded mutants do not show any significant preferential orientation in the micelle, compared to the folded protein. However some minor differences in the structure of different unfolded protein domains can still be recognized.

In section 3.1.4. it was shown that the unfolded protein can be prepared with different micelle compositions in slightly different solvents. It is assumed that mainly the micelle composition can influence the structure of the unfolded state of the protein. The comparison between mutants in different unfolded states (Figure 60) shows differences in their ESEEM deuterium accessibility.

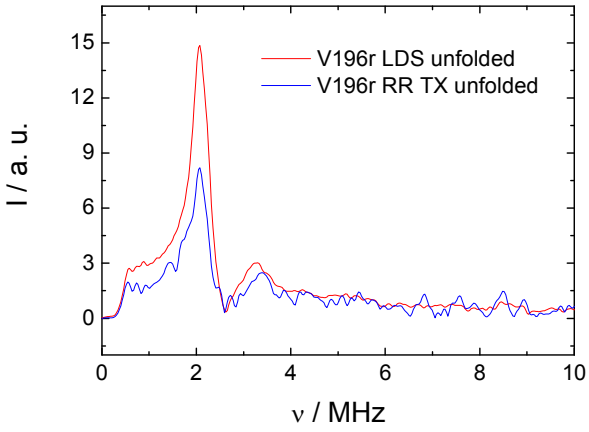


Figure 60: Change of deuterium peak intensity in Fourier transform ESEEM spectra as a function of micelle composition for unfolded mutants.

This observation supports the hypothesis that different unfolded states are structurally not identical. It seems that the spin label is more solvent accessible in V196r LDS unfolded sample than in the V196r RR TX unfolded state, which hints on a more open structure.

The less accessible structure of the RR TX unfolded mutant is consistent with the aggregation hypothesis presented in sections 4.2.4. and 4.2.5.1. It is expected that spin labels in the aggregated structure will show reduced deuterium accessibility compared to the open structures as the spin label will be shielded from the deuterated buffer by other protein molecules of the aggregate.

These observations shed some light on the structure of unfolded protein and show the strength of ESEEM in structural studies.

4.2.3.2. Folding kinetics determination with ESEEM

The next step of the folding investigation is to use the environmental sensitivity of ESEEM to get information about the changes of the local environment of spin labels due to protein folding.

The necessary condition for this type of studies is the difference in the environment in the folded and unfolded state. In the Figure 61 the RR TX folded and unfolded samples are compared with each other.

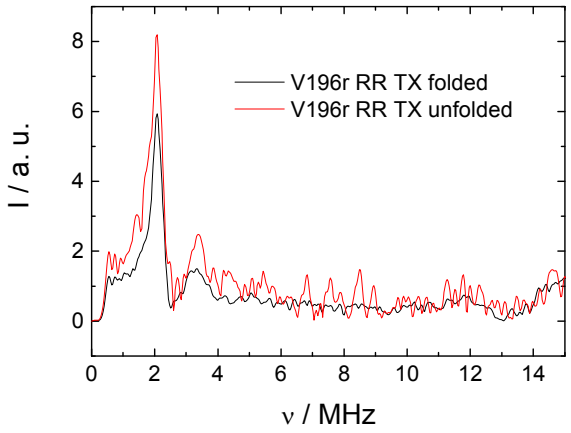


Figure 61: Change of deuterium peak intensity in Fourier transform ESEEM spectra as a function of folding state.

Figure 61 shows that the intensity of the ESEEM peak at about 2 MHz, which was shown to be sensitive to the number of surrounding deuterium nuclei, changes as the protein folds.

The measurement conditions and sample composition were the same. The relaxation behavior for the folded sample was not very good but still showed features that allowed for assuming that there was at least some amount of folded protein in the sample.

ESEEM intensity comparison with the UZ folded samples shows higher deuterium peak intensity for the RR TX samples. Further comparison of the ESEEM intensities of RR TX folded S59r, S160r, and V196r mutants shows the same trends as for UZ mutants, except for V90r, where the deuterium peak intensity in the RR TX sample was higher than expected. This can be explained by the fact that some unfolded protein is also present in RR TX samples. Thus the ESEEM signal in RR TX samples corresponds to a mixture of folded and unfolded species. As the accessibility trends are similar as in

UZ samples it is assumed that the protein in RR TX samples at least partially resembles the correctly folded protein.

The Figure 62 shows the gel that proves the availability of folded protein in the studied sample.

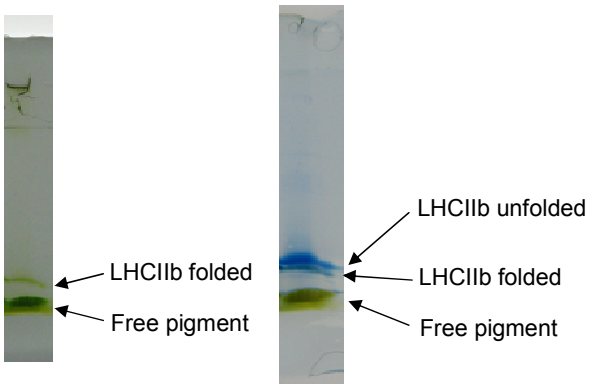


Figure 62: Gelelectrophoretical control of the folding quality

From all the facts listed above Figure 61 can be safely interpreted in terms of local structural changes during the protein folding process. This allows for the first time to use ESEEM as a method for membrane protein folding investigations. The study of local changes of the spin label environment during protein folding is thus possible. This newly developed method opens new perspectives in protein folding investigations.

The kinetics of protein insertion in the membrane that is one of the stages in the two-stage membrane protein folding model **[Bowie 2005]**, can in principle be studied by the method developed in this work. Also the processes of translocation mechanisms of proteins in the membrane can be addressed and possibly understood with the method presented here. This is important as these processes are poorly understood to date **[Buchner 2005]** although of high importance, as for example many protein toxins work via membrane insertion **[Savitsky 2004]**.

Similar information can also be obtained from fluorescence measurements **[Booth 2006b and references there]**.The advantages of using ESEEM as a method for the characterization of the local environment were explicitly discussed in section 4.1.6.

After proof of principle that it is possible to use ESEEM to differentiate between folded and unfolded structure and to study the local accessibility change during the folding process, kinetic information about this structural change was obtained.

Figure 63 shows ESEEM results on protein folding kinetics for the V196r single mutant. The samples were prepared as described in section 3.1.4.

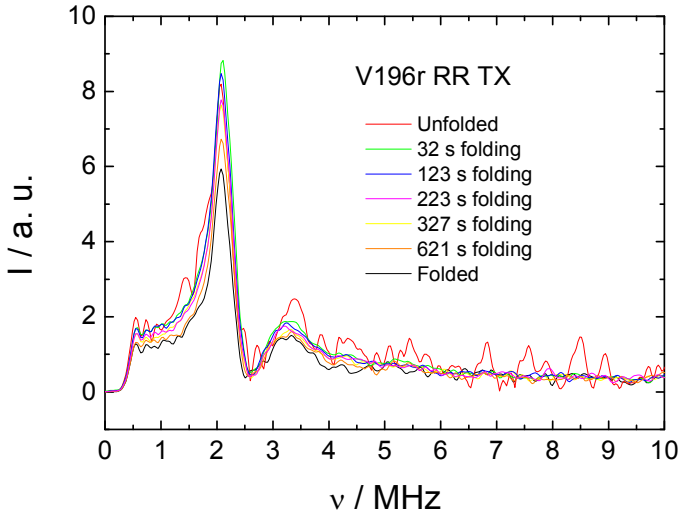


Figure 63: Fourier transform ESEEM spectra as a function of folding time.

The effective transversal relaxation time of the samples showed a continuous decrease as a function of folding. That is assumed to be linked to the environmental change of the spin label as the protein folds. That is why this parameter cannot be used to determine sample quality during these measurements as described in section 4.2.1. As the ESE spectrum of the studied samples was comparable to the reference sample and gel electrophoresis showed the presence of folded protein, the samples were considered to fold properly.

Figure 63 shows that there is a change in ESEEM intensity as the folding proceeds. The intensity of the unfolded mutant is rather small. This is assumed to be an aggregation effect, as discussed in detail in sections 4.2.3.1., 4.2.4., and 4.2.5.1. After 32 s of folding time an increase of ESEEM intensity is observed. This can be explained by the deaggregation of the folded protein. The aggregation may thus be a parallel process that competes with folding or the aggregated proteins can deaggregate when the pigments are added. The high intensity of the peak at small folding times shows that the partially folded protein structure is quite open. This can be explained by the fact that this part of the protein is not buried in the micelle at short folding times or by the expanded structure of the molten globule state that can allow water penetration to some extent **[Damaschun 1986]**, **[Santucci 2008]**. After 32 s of folding a continuous

decrease of ESEEM intensity is observed, meaning that the position V196r inserts more deeply into the micelles or is shielded by neighboring amino acids or that the structure gets tighter. The folded protein shows the smallest accessibility of all mutants. From this figure it could be shown that also the rate of environmental change of a certain site in the protein during folding can be measured with ESEEM. To our knowledge this is the first time that ESEEM was used for this purpose.

By plotting the deuterium peak intensity from Figure 63 as a function of folding time a folding time constant (*k*) can be determined. Usually the folding kinetics function could be fitted by a single exponential function $y= y_0 + A_1 \exp-(x/k)$, with y_0, A_1, and k as variable fitting parameters.

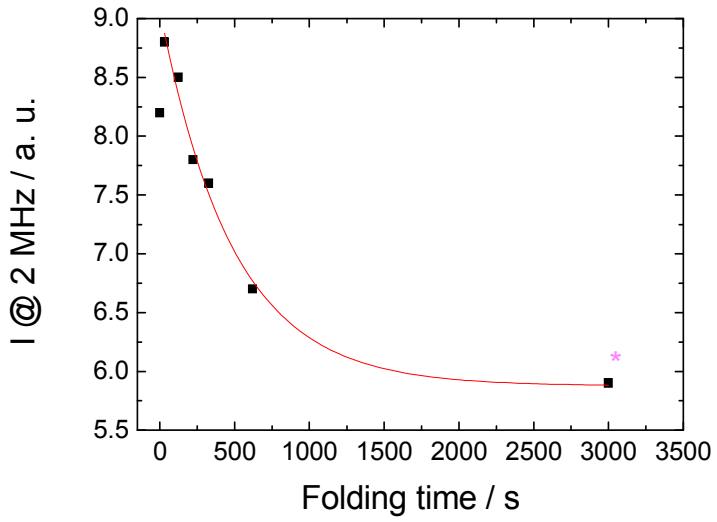


Figure 64: Determination of the time constant of environmental change of V196r mutant during protein folding with ESEEM. Red curve is the single exponential fit function on the data.

The completely folded protein is shown in Figure 64 as a point at a folding time of 3000 s. It was checked that the folding time constant does not change its value when assuming even larger folding time values for this sample.

For the environmental change of the V196r mutant the folding time constant is determined to be about 500 s. The data without the point for the unfolded sample (that is assumed not to be representative due to the aggregation of the protein in the unfolded state) could be fitted by a single exponential function, which suggests first order reaction kinetics. However as the pigments were in excess, also pseudo first order reaction kinetics is consistent with these results.

After determining the rate of environmental change for the V196r mutant, other mutants were investigated.

From Figure 65 it is seen that the V90r mutant is also suitable for kinetic investigation with ESEEM, as the deuterium peak intensities for the folded and unfolded state are different.

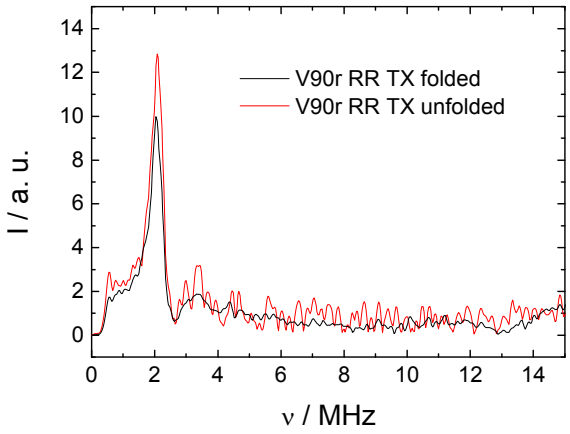


Figure 65: Change of deuterium ESEEM intensity as a function of protein folding for the V90r mutant.

The main restriction for using ESEEM for protein folding kinetics determination is that there must be an accessibility and thus deuterium intensity change between the folded and unfolded protein. This was not always the case, as shown in Figure 66 for the S160r mutant.

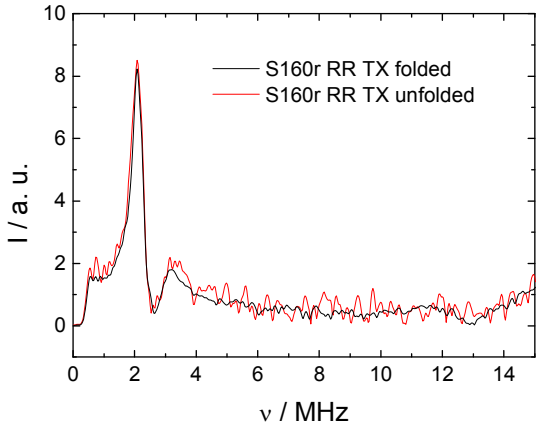


Figure 66: Fourier transform ESEEM spectrum of the S160r mutant in different folding states as an example of a mutant that is not suitable for ESEEM protein folding studies.

Figure 66 shows that the Fourier transformed ESEEM intensity in the folded and unfolded state is the same. Thus this mutant is not suitable for folding investigation by ESEEM.

Other mutation positions were not investigated, as their ESEEM accessibilities in the folded state are intermediate as can be estimated from Figure 27. It was thus assumed that change between the ESEEM spectra of folded and unfolded protein would be too small for quantification.

Unchanged deuterium accessibility is not the only hindrance in the ESEEM studies of protein folding. Although a difference in the deuterium ESEEM intensity between the folded and unfolded state of the S59r mutant is observed, no kinetic measurements with this mutant were performed as there are hints for a possible structural change in this protein region (see section 4.2.1.1.). Thus data for mutant S59r could not be interpreted safely.

Mutant S106r was also not used for characterizing folding kinetics by ESEEM, as it was shown to be in a labile protein region (see section 4.1.5.2.).

In summary it is possible to say that ESEEM is a useful method for structural investigation of folded, unfolded and partially folded proteins. The information could be obtained also for some mutants that showed no differences in the CW EPR spectra. This again shows the strength of pulse EPR spectroscopy for this kind of studies.

4.2.4. Relaxation measurements for LHCIIb folding studies

As shown in section 4.1.4., relaxation times are sensitive to the environment of the spin label. Effective relaxation times were compared between the unfolded and folded state for different mutants in Table 4 in order to find typical trends. The mutants presented here are considered to be representative.

Mutant	Condition	τ_1 / µs	τ_2 / ns
S3r	RR TX folded	941	2334
S3r	RR TX unfolded	442	384
S3r	RR TX unfolded deuterated	493	1548
S160r	RR TX folded	903	2153
S160r	RR TX unfolded	506	463
S160r	RR TX unfolded deuterated	434	719
S160r	RR TX unfolded 1:4 mixture with C79S	633	985
V196r	RR TX folded	627	1517
V196r	RR TX unfolded	409	409
V196r	RR TX unfolded deuterated	361	797
V196r	Unfolded LDS	1231	3024
V196r	Unfolded LDS deuterated	1198	4148

Table 4: Dependence of relaxation times on the protein folding state and solvent composition.

The effective longitudinal relaxation time (τ_1) for all the samples was determined as usual (See section 3.2.2.3. for details).

Fits for the determination of the effective transversal relaxation time (τ_2) of folded mutants were monoexponential. They were fitted regardless of proton modulation, as the error introduced by modulation is estimated to be around 0.5% and can thus be neglected.

Fits for τ_2 determination of unfolded mutants were biexponential. They were also fitted regardless of proton modulation, as the error introduced by such modulation for this type of samples is estimated to be around 6%. The investigated effect is much bigger than this error. Interestingly τ_2 for the RR TX unfolded sample mixed with unlabelled C79S protein showed an almost monoexponential decay function.

For the unfolded sample with LDS micelles τ_2 was defined as the time when the echo amplitude decayed to 1/e of the initial value as neither monoexponential nor a biexponential fit was good.

The fits for τ_2 determination of unfolded deuterated mutants were monoexponential. They were performed over the deuterium modulation as compared to protons the effect of deuterium modulation may introduce significant errors to the measured relaxation time.

The fact, that different samples are fitted with different shapes of relaxation curves can be a hint for different dominating relaxation mechanisms.

From Table 4 it can be seen the RR TX folded mutants show τ_1 and τ_2 values comparable with UZ mutants. This is a further hint for a reasonable quality of RR TX samples. It is also obvious that relaxation times are sensitive to the protein folding state.

The discussion of Table 4 is split in two parts. In the first part transversal relaxation is considered. In the second part longitudinal relaxation is considered.

Table 4 clearly shows that τ_2 in the unfolded samples is always at least by a factor of 3.7 smaller. This is a significant effect. It seems improbable that such a large effect could be caused only by the environmental change during the protein folding, as in section 4.1.4. it was shown that environmentally caused τ_2 changes are much smaller.

Table 4 also shows increased τ_2 for the unfolded S160r mutant mixed with unlabelled C79S mutant and for the V196r unfolded LDS sample. A closer look at RR TX unfolded samples showed that they were more turbid compared to samples in LDS.

These observations led to the idea of possible aggregation of proteins in the unfolded RR TX state. In order to prove this hypothesis dynamic light scattering experiments were performed. However no interpretable results could be obtained with this method, probably due to the scattering of the micelles that are also present in the sample (see section 3.1.4. for details).

A possibility to test the aggregation hypothesis was given by the nature of relaxation itself.

At increased spin concentrations relaxation measured by an echo experiment is apparently enhanced by instantaneous diffusion**.** In this case there is a dependency of the transversal relaxation time on the flip angle β of the pulse as can be seen from equation ***50***

$$\frac{1}{T_{ID}} = \frac{C\pi\mu_0 g_1 g_2 \beta_e^2}{\hbar 9\sqrt{3}} \sin^2\left(\frac{\beta}{2}\right) \qquad \textit{Eq. 50}$$

Here C is the number of spins per cubic meter and g_1 and g_2 are the g values of observed and excited spins. **[Schweiger 2001]**

g_1 and g_2 were taken equal to 2.

According to this equation, by measuring the transversal relaxation time as a function of pulse flip angle β and plotting its rate against $\sin^2(\beta/2)$ positive linear dependence should be observed if instantaneous diffusion is significant. The slope of this linear function should be directly proportional to the local spin concentration C.

The pulse angle was varied by changing the microwave attenuation and leaving the pulse length the same. This is required to keep the excitation bandwidth constant, so that the same spin packets contribute to the signal at all flip angles.

The resulting flip angle β was calculated according to equation ***51*** **[Schweiger 2001]**:

$$\beta(\mathrm{a}) = \beta(\mathrm{r})\sqrt{\left(10^{\frac{\mathrm{dB(r)}}{10}-\frac{\mathrm{dB(a)}}{10}}\right)} \quad \textit{Eq. 51}$$

Here β(a) is the actual flip angle after changing the microwave attenuation, β(r) is the reference flip angle (was always set as near as possible to 180°), dB(r) is the reference microwave attenuation (was adjusted so that no echo was seen when the amplitude of the pulse was set to zero) and dB(a) is the actual microwave attenuation.

The relaxation parameters were determined with the weighted average of a second order exponential decay curve (except for the smallest angle where a single exponential fit was possible) fitted over proton modulation for the V196r RR TX unfolded sample, by the time the curve decays to 1/e for the V196r LDS unfolded sample, and by an exponential decay curve fitted over proton modulation for the V196r RR TX folded sample.

Figure 67 shows the transversal relaxation behavior of different samples as a function of $\sin^2(\beta/2)$.

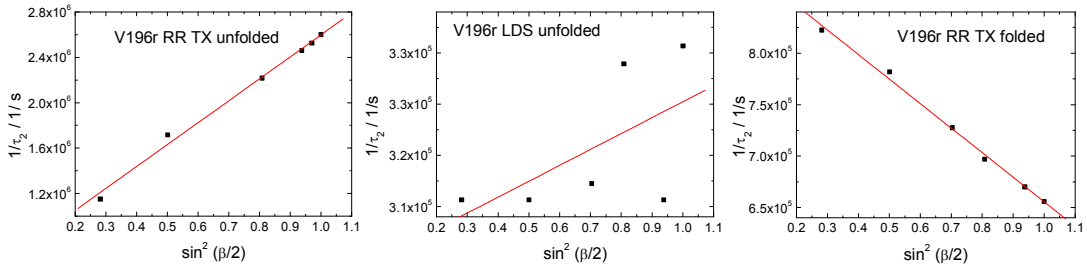


Figure 67: Transversal relaxation rate behavior as a function of pulse flip angles for different samples. The red lines are linear regression fits. The regression fits at x-values larger than one have no physical meaning.

The parameters for linear regression fits on these samples are summarized in Table 5.

Mutant	A	B	R
V196r RR TX unfolded	665295.41	$1.931 \cdot 10^6$	0.997
V196r LDS unfolded	309695.56	$0.016 \cdot 10^6$	0.593
V196r RR TX folded	894052.98	$0.238 \cdot 10^6$	-0.998

Table 5: Parameters for linear regression fits obtained from Figure 67. The linear regression is given by y=A+Bx. R is the regression coefficient.

Figure 67 clearly shows, that the V196r mutant in the RR TX unfolded state shows a linear dependence of $1/\tau_2$ on $\sin^2(\beta/2)$. The quality of the fit is very good, as can be seen from the R value in Table 5. This is considered to be a proof for instantaneous diffusion contributing significantly to transversal relaxation in this sample.

Due to this fact and good quality of the regression the local spin label concentration in the unfolded sample can be determined. According to equation ***50*** it is 3800 μM. This is about 100 times higher than the expected concentration which confirms the idea of protein aggregation in this buffer composition when pigments and lipids are not present. Interestingly the concentration effects begin to influence the line height in CW EPR spectra starting only at concentrations higher than 7500 μM [**Kokorin 1972]**. That can be the reason why no differences in CW spectra between the folded and unfolded LHCIIb molecules are observed in Figures 57 and 58. Thus only pulse EPR can

provide reliable information about the local sample concentration in the observed range.

In comparison to the unfolded RR TX sample the unfolded LDS sample shows very weak dependence of $1/\tau_2$ on $\sin^2(\beta/2)$. The linear regression fit performed on the data is very poor, which is also seen by the R value. The comparison between the slopes of the unfolded RR TX sample and the unfolded LDS sample shows a difference of a factor of 100. It is concluded, that in the unfolded LDS sample no significant instantaneous diffusion takes place and thus it is assumed that no aggregation is present in this sample type.

For the folded sample there is a linear dependence of $1/\tau_2$ on $\sin^2(\beta/2)$, however the slope is negative. According to equation ***50***, such behavior is not expected, as the concentration C can not be negative. Thus it can be concluded that instantaneous diffusion is not present in the folded sample. The mechanism that causes the negative slope in the Figure 67 however can not be easily explained due to the high complexity of the sample. A systematic study with nitroxides in model solvents would be necessary in order to understand the mechanism causing this relaxation behavior.

From the presented measurements it is possible to conclude that the presence of protein aggregates as well as the local concentration of spins in these aggregates can be determined. This increases the method palette for obtaining structural information about the unfolded protein state.

The fact that there is an aggregation in the unfolded state does not necessarily exclude also an environmental change during protein folding. In order to separate the two effects from each other the samples were deuterated as it was shown in section 4.1.4. that by deuteration the effects of neighboring protons can be explicitly studied.

In Table 6 the relaxation rate difference for unfolded mutants that are deuterated and not deuterated are shown together with the local proton concentration according to equation ***49***.

Mutant	$\Delta(1/\tau_2)\cdot10^{-4}$ / 1/ns	C(H) / mol/l
S3r RR TX unfolded	19.6	488
S160r RR TX unfolded	7.7	192
V196r RR TX unfolded	11.9	297
V196r LDS unfolded	0.9	22

Table 6: Effective transversal relaxation rate differences for unfolded mutants in deuterated and non-deuterated solvents, and the local proton concentration determined from them.

From Table 6 it is seen that deuteration causes large effects on relaxation rates for all of the studied RR TX unfolded samples, with the biggest effect observed for the S3r mutant. This hints to the high solvent accessibilities of these protein states. The local proton concentration determined for all the unfolded RR TX samples from the relaxation rate difference has no physical meaning, as it is significantly higher than proton concentration in pure water. This suggests a huge error in these measurements. The error is probably caused by the fact that only a few deuterium modulations are visible, and that effects of instantaneous diffusion relaxation and proton spin diffusion interfere with each other. Thus only the accessibility trends can be discussed.

A rather minor effect on the relaxation rate difference observed for the unfolded LDS sample is interpreted as moderate deuterium accessibility on the scales probed by transversal relaxation. For this sample the determined proton concentration is assumed to be realistic as from the previous measurements it was shown not to be aggregated.

It is of interest to compare the accessibilities determined from the transversal relaxation measurements with the ones determined by ESEEM. The high solvent accessibility for the RR TX unfolded state with increased accessibility for the S3r mutant was also observed by ESEEM measurements as shown in section 4.2.3.1. These observations legitimize the use of the relaxation rate differences, although they are subjected to high error, to determine the local environment of sites in proteins also in the unfolded state. The reason for the increased solvent accessibility of unfolded S3r RR TX mutant may be the hydrophilicity of this protein part, as also for the folded protein it shows high solvent accessibility. Interestingly the deuterium accessibility for the V196r unfolded LDS mutant determined from transversal relaxation measurements is smaller than the one determined from the ESEEM measurement. This corresponds to high deuterium accessibility on short length scales, and moderate accessibility on longer length scales.

Due to the complexity of the sample, and just a few structural constraints this observation cannot currently be explained in terms of a structural model.

The longitudinal relaxation parameter τ_1 in unfolded RR TX samples was always smaller compared to the one in folded RR TX samples as can be seen from Table 4. The effect is, however, not as significant as for τ_2.

Hints of influence of protein aggregation on τ_1 are present in Table 4 as τ_1 increases when the unfolded S160r mutant is mixed with unlabelled C79S mutant. Also the unfolded LDS mutant, which from τ_2 measurements is known to be in the non-aggregated state, shows increased τ_1.

The reason for a τ_1 decrease in the unfolded samples can be spin diffusion, a process that spreads the z magnetization from one spin to the other when they are dipolar coupled. The efficiency of spin diffusion depends also on the average distance between the coupling spins. **[Schmidt-Rohr 1994]**, **[Schweiger 2001]**

As the average interspin distance is assumed to decrease for the aggregated protein samples that would also explain their τ_1 relaxation behavior.

The effects of spectral diffusion on both relaxation times are also possible, they were however neglected, as experiments on other nitroxide samples indicate that they are negligible at 50 K with the measurement protocol used here **[Schweiger 2001], [Jäger 2008]**.

The reason for the decreased τ_1 in the unfolded samples may be also the environmental change of the spin label as the protein folds, as it was shown in section 4.1.4. that τ_1 is sensitive to the local environment. Deuteration does not have any effect on τ_1 of the RR TX unfolded sample, which was also the case for the folded UZ protein. Thus also in the unfolded RR TX samples the librational effects most probably influence the τ_1 relaxation time.

Comparison of the ESE spectra of the folded and unfolded protein shows drastic changes in their shape (Figure 68).

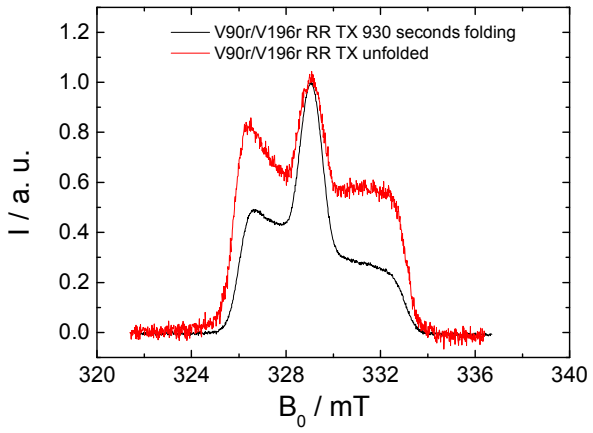


Figure 68: Comparison of the ESE spectra for the folded and unfolded RR TX mutants. Intensity normalized and B_0-field shifted.

The differences in ESE spectra can be explained by the change in the environment of the spin label as the h_1/h_2 values change for some of the mutants, or by the effects caused by the strong coupling between the spins due to aggregation. It is noteworthy that these ESE spectra were also used to test the quality of protein folding (See section 4.2.1.). It is now clear that they are sensitive to protein aggregation or environmental change, which legitimizes their use for quality control of the samples.

The presented results show that it may be in principle possible to use τ_1 for studying protein folding kinetics if the local environment or aggregation state changes as the protein folds.

Summarizing the observations it is possible to say that effective longitudinal relaxation time τ_1 and effective transversal relaxation time τ_2 can be used to study the structure of unfolded proteins as well as protein folding. The effective transversal relaxation time can be used to characterize the aggregation state. By τ_2 comparison in deuterated and non-deuterated unfolded samples it is also possible to get hints about the local environment of the spin label, and thus to obtain more information about the structure of the unfolded protein.

The effective longitudinal relaxation time measurements are also capable of determining the aggregation state of the protein. This effect is however not as strong as in the case of the effective transversal relaxation time. The local environment information on RR TX samples, as also described in section 4.1.4. for UZ LM samples,

might be obtained from τ_1 and ESE measurements. However it is not very clear whether and to which extent the observed aggregation can introduce errors to that.
As τ_2 and probably also line shapes in ESE spectra are sensitive to protein aggregation, and as it was shown that unfolded proteins are aggregated, the methods were used routinely to control the quality of the samples for kinetic measurements. When τ_2 was small or ESE spectra showed features similar to that of unfolded protein the sample was not investigated as already mentioned in section 4.2.1.

4.2.4.1. Folding kinetics determination with relaxation time measurements

In the previous section it was shown that relaxation times are sensitive to the aggregation state of the protein and the local environment of the spin label. The relaxation times were thus used to study protein folding.
Time-resolved measurements of effective transversal relaxation time (τ_2) were performed in combination with the DEER measurements and thus are always listed for the double mutants.
For the V90r/V196r RR TX mutant the τ_2 increased rapidly from 344 ns in the unfolded state, to 2312 ns for the 37 s sample. For all larger folding times τ_2 remained almost constant. The fit for the unfolded sample was biexponential, whereas for all of the later folding times (see DEER kinetics for details) it was monoexponential. The curves were fitted regardless of proton modulation.
This result can be explained by assuming that the sample deaggregates on a time scale faster that 37 s, and after that the local changes of the spin label environment are not sufficient to cause any further effects on τ_2 relaxation. The different fit functions for the folded and unfolded state have already been discussed in section 4.2.4.
There is no significant change in the τ_2 relaxation data after 39 s of folding time for the S106r/S160r RR TX mutant. For the unfolded RR TX sample mixed 1:1 with C79S mutant a smaller relaxation time is expected according to decreased signal-to-noise ratio compared with the mutants after 39 s of folding. The fitting after the 39 s of folding could not be performed by the mono- or biexponential function. The relaxation time was estimated to be above 2000 ns.
From these experiments it is concluded that the change in τ_2 happens on a faster time scale than the changes observed during DEER measurements (see section 4.2.5.2.). This fast kinetic step is referred to the aggregation of the unfolded proteins in the RR TX buffer without the pigments. Whether the aggregation step is a prerequisite before

the folding, or is a dead-end parallel reaction, cannot be clearly identified. It was however observed, that the mutants that did not fold well showed also shorter τ_2.
The change of effective longitudinal relation time (τ_1) as a function of folding time for the deuterated V196r RR TX mutant is shown in Table 7.

Folding time / s	τ_1 / μs
0	361
32	849
123	809
223	717
327	717
621	642
Folded	511

Table 7: Effective longitudinal relaxation time (τ_1) as a function of folding time for deuterated V196r mutant.

From Table 7 it is seen that there is a strong change in τ_1 between 0 and 32 s of folding. This can be explained in terms of aggregation of the unfolded protein samples, as it was also seen by τ_2 measurements. After that a continuous decrease of the relaxation time as a function of folding time is observed. The data without the unfolded mutant could be fitted by a single exponential function. The folding time constant from this type of measurements is determined to be around 550 s. This time constant is very similar to the one determined by the ESEEM kinetic measurements on the same samples. This gives a further hint to the environmental sensitivity of the effective longitudinal relaxation time and proves the possibility of using it to study the local environmental changes as a function of folding time. The exact type of structural change is difficult to pinpoint as the relaxation time can depend on many parameters. However it is most likely that it is the solvent accessibility that changes during folding (See sections 4.1.4. and 4.1.6. for details).

4.2.5. DEER for LHCIIb folding studies

4.2.5.1. DEER on unfolded LHCIIb

In the previous sections of this work it was shown that the unfolded RR TX proteins are most probably aggregated. Here DEER spectroscopy is utilized to prove this statement, and also to get more detailed structural information about these molecular aggregates.
DEER on unfolded RR TX samples cannot be easily performed due to the decreased transversal relaxation time. This means that at the detection time the signal from those spins that are very close to each other has already decayed and thus aggregated structures are underrepresented compared to non-aggregated structures that have a longer relaxation time. In order to be able to measure such samples the transversal relaxation time should be increased. In Table 4 it was shown that it is possible by mixing the unfolded protein sample with the unlabelled C79S mutant. Thus all the unfolded RR TX double mutants for DEER measurements were prepared like that. Additional deuteration of such samples did not lead to further signal improvement and thus was not performed routinely.
In order to get deeper understanding of the aggregation effects on the DEER data measurements were performed with different DEER interpulse delay times, as shown in Figure 69. Raw data is shown that was not subjected to any analysis in order to be sure not to introduce any artifacts.

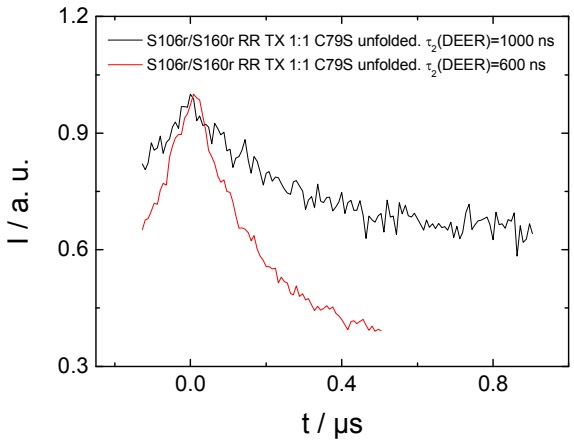


Figure 69: Dependence of the DEER modulation depth on the interpulse delay for a RR TX sample in unfolded state. Raw DEER data.

The τ_1 and τ_2 DEER interpulse delay times (see Figure 9 for their definition) for the black curve were 200 ns and 1000 ns, and for the red curve they were 200 ns and 600 ns respectively. Note that these τ_1 and τ_2 are the interpulse delay times and have nothing in common with effective relaxation times discussed earlier.

Figure 69 shows that the modulation depth changes when the interpulse delay is changed even though the labelled protein is diluted by C79S protein.

A similar comparison of the dependence of the modulation depth on the interpulse delay for a folded mutant (raw data) is given in the Figure 70.

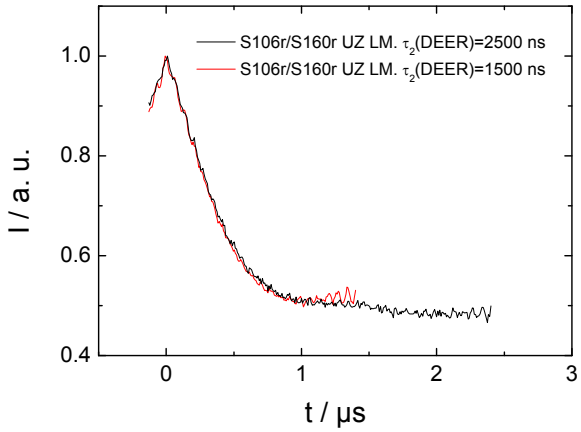


Figure 70: Dependence of the DEER modulation depth on the interpulse delay for folded UZ LM sample. Raw DEER data.

The τ_1 and τ_2 DEER interpulse delay times (see Figure 9 for their definition) for the black curve were 200 ns and 2500 ns, and for the red curve they were 200 ns and 1500 ns respectively. Also here τ_1 and τ_2 are the interpulse delay times and have nothing in common with effective relaxation times discussed earlier.

Figure 70 shows that for the UZ LM sample no change in the modulation depth is observed when the DEER interpulse delay is changed.

It is known that the amount of coupled spins can be determined from the modulation depth **[Milov 1984]**, **[Schweiger 2001]**, **[Hilger 2005]**, **[Bode 2007]**. As there is a change in the modulation depth with changing interpulse delay time only for the unfolded mutant it is assumed to be related with the relaxation effects due to sample aggregation. In a heterogeneous sample, configurations with shorter average distance cause a faster decay of the DEER signal, but are also subjected to faster transversal

relaxation due to instantaneous diffusion. At longer evolution times the echo signal for these configurations is strongly suppressed, they do not contribute to the DEER signal, and the decay of the DEER curve thus becomes slower. This data thus provides evidence for the aggregation hypothesis described in the previous sections.

The next step in the investigation of the unfolded state is to obtain information about its structure. For that DEER experiments on different double mutants in the unfolded state were performed.

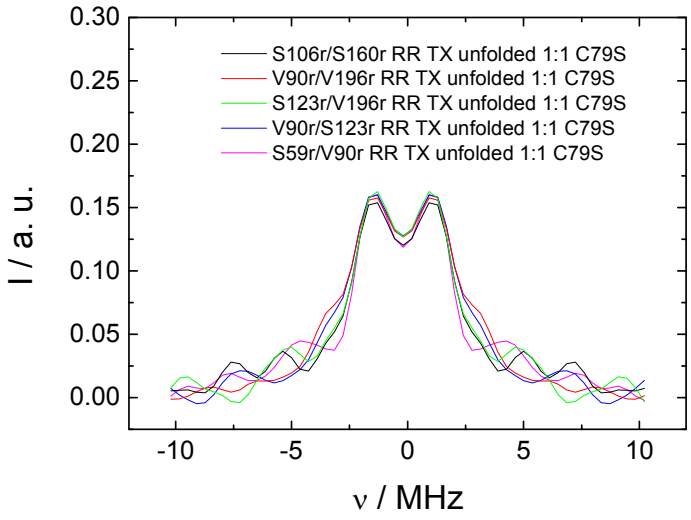


Figure 71: Comparison of the DEER dipolar spectra of different RR TX mutants in the unfolded state. Spectra were normalized with respect to integral intensity in a range from -10 to 10 MHz.

The comparison of dipolar DEER spectra in Figure 71 (Cutoff at 800 ns, background correction at 88 ns) of different unfolded mutants shows that the main frequency component is the same for all the studied double mutants. This means that the distances in the unfolded mutants are rather similar. The similarity of the distances for all the mutation positions can be explained by the aggregation of the mutants in RR TX buffer. It is expected that the intermolecular distances in a large aggregate are being measured and they are almost independent on mutation position. The spectra are rather broad, which means that a broad distribution of distances is present, as it would also be expected for the unfolded state. Care is however needed in interpretation of these DEER measurements, as the signal-to-noise ratio is rather poor, and the evolution time of the DEER experiment was always under 1 μs, which excludes determination of long distances.

For unfolded protein the conformation is expected to behave according to the random coil polymer model **[Fitzkee 2004]**. The radius of gyration (R_g), which is the root mean square of atoms from their common center of gravity **[Flory 1969]**, is expected to follow equation ***52***

$$R_g = R_0 N^{\nu} \qquad \textit{Eq. 52}$$

Here R_0 is a constant related to the persistence length and ν is a scaling factor that depends on the polymer solvent interaction. R_0 was taken as 1.98 Å according to **[Fitzkee 2004]**. The scaling factor ν ranges from 0.33 for a poor solvent through 0.5 for ideal solvent to 0.6 for a good solvent **[Fitzkee 2004], [De Gennes 1979]**. In the case at hand the degree of polymerization N is the number of amino acids between the mutation positions including the second mutated amino acid. The mean square end-to-end distance for such a random coil is six times as large as the squared radius of gyration, determined from the equation above.

The MATLAB simulation procedure of the experimental DEER data based on the described model was kindly provided by Gunnar Jeschke.

The DEER data was not background corrected. The reason for that was that it was not possible to find a satisfactory background correction parameter. Thus the intermolecular effects on the data have to be kept in mind in further comparison. The effects of neighboring molecules can be mostly taken in account by fitting the data with differing modulation depth. The largest error due to the not performed background correction comes from the aggregated molecules.

In Figure 72 the comparison between theoretically determined DEER dipolar curve and experimental data is shown for unfolded RR TX mutants.

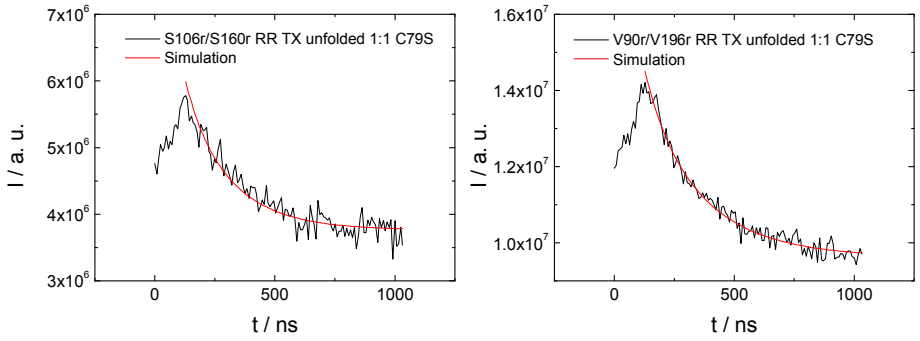


Figure 72: Raw DEER data of RR TX unfolded mutants (black) and simulation of this data based on the random coil model (red) for different LHCIIb unfolded RR TX mutants.

Unfortunately due to the short measurement time and poor signal to noise ratio it was not easy to obtain a good fit for the S106r/S160r mutant as the proton modulation was also fitted. The simulation was thus performed by keeping the modulation depth parameter under 0.6 and varying the scaling factor. A reasonable fit as shown in the Figure 72 could be obtained with a modulation depth of 0.6 and a scaling factor of 0.475. The error for the determined fitting parameters is assumed to be rather large. The simulation of the theoretical distance distribution for V90r/V196r RR TX unfolded mutant was possible without fixing the modulation depth parameter. The modulation depth parameter of the fit shown in the Figure 72 is 0.522 and the scaling factor is 0.425.

The fact that the experimental distance distribution data could be fitted by the simulation based on the theory described above allows the assumption that the random coil theory is also suitable in explaining the structure of the aggregated proteins.

Different scaling factors for two different mutation positions are assumed not to be representative due to the small measuring period of the experimental data and poor signal to noise ratio and thus were not further analyzed.

In order to see whether the theory is generally applicable simulations of the data for the unfolded LDS samples were performed. In the previous sections it was shown that the unfolded protein in this solvent does not aggregate. Thus these samples should be a nice system to test the validity of the random coil theory for unfolded proteins. The experimental data and the simulations of this data are sown in the Figure 73.

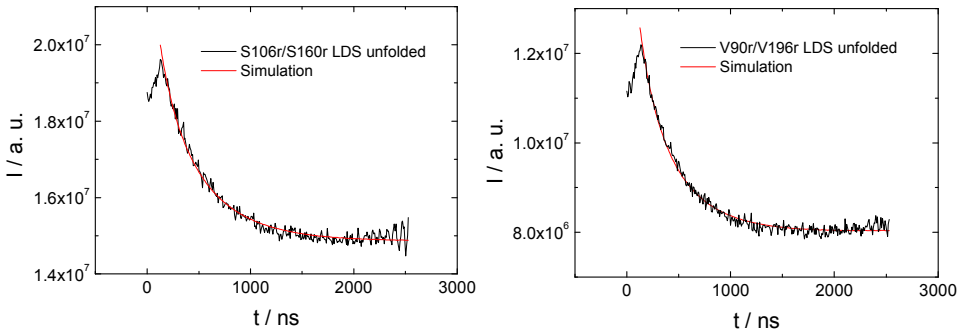


Figure 73: Raw DEER data of unfolded LDS mutants (black) and simulation of this data based on the random coil model (red) for different LHCIIb unfolded LDS mutants.

Figure 73 shows that the S106r/S160r unfolded LDS mutant data can be fitted sufficiently well with the random coil model with a modulation depth of 0.357 and a scaling factor of 0.546. Also the V90r/V196r unfolded LDS mutant can be fitted by the random coil model with the parameters being 0.497 for the modulation depth and 0.455 for the scaling factor.

The fact that the fitting of the experimental data with the theory presented above is possible hints that the domains between the labeled residues in these samples behave like random coils.

The comparison of the scaling factors for two different unfolded LDS mutants shows differences depending on the mutation position. This would suggest different solvent behavior of different protein parts in the unfolded state. Such behavior could be also caused by helix pre-forming, as the pre-formed helix could have different scaling behavior than a random coil.

From the CD measurements it was suggested that there is some helical structure in the unfolded LHCIIb protein in SDS micelles, presumably in the regions that are helical in the folded protein **[Horn 2002]**. The composition of the samples studied here differed from the ones described in **[Horn 2002]**, however it was assumed that the anionic micelles have the main influence on the structure formation. Thus it is assumed considering the different scaling factors for the random coil simulation of unfolded LDS mutants that the helix pre-organization can also take place in the solvent used here.

In order to check this hypothesis and to eventually localize the helix formation region the DEER dipolar evolution curves of UZ LM samples are compared with the ones of unfolded LDS samples in Figure 74.

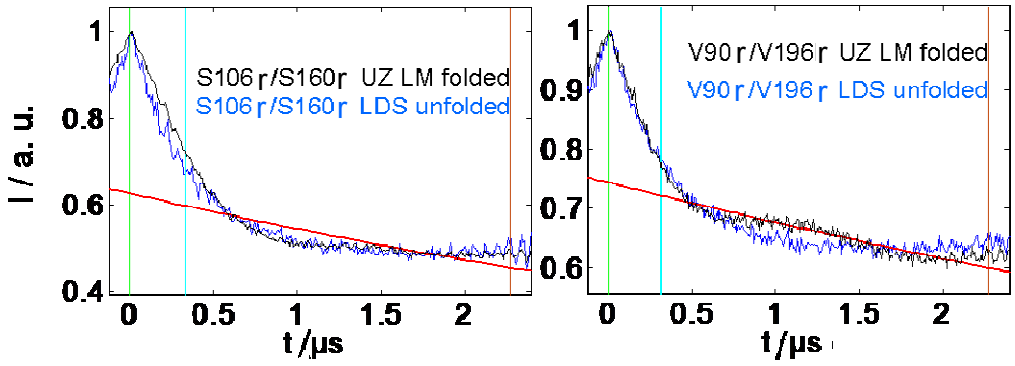


Figure 74: Comparison of DEER dipolar evolution curves for two different double mutants for folded UZ LM sample and unfolded LDS sample. The original data is presented. The dipolar evolution curves were scaled by their modulation depth using "DeerAnalysis2006" program.

Figure 74 shows similarity between the dipolar evolution function of UZ LM folded and LDS unfolded S106r/S160r sample. This is not the case for the V90r/V196r sample. This observation can be a hint for the helix preorganization of LHCIIb in LDS micelles, as it is known that positions S106r and S160r are on the opposite sides of a transmembrane helix. The fact that the unfolded S106r/S160r LDS sample could still be fitted by the random coil model can be explained by the flexible loops around the pre-formed helix.

Summarizing these observations it is possible to say that in addition to information on aggregation, structural information about the unfolded structure was obtained with DEER spectroscopy. This information is of fundamental interest as it allows to explicitly test theoretical models of unfolded protein structures. Some hints on helix pre-forming in a specific protein domain were obtained. The DEER data of unfolded proteins will be further useful in interpretation of folding kinetics with DEER spectroscopy in section 4.2.5.2.

4.2.5.2. Folding kinetics determination with DEER

The next step after studying the unfolded protein of LHCIIb with DEER is to apply this technique to study protein folding kinetics. Compared to previously used optical techniques DEER measurements have the advantage that structurally resolved kinetic information about protein domain organization during the folding can be obtained. Thus the kinetics of specific secondary and tertiary structure assembly can potentially be determined. This in turn can provide direct information about the protein folding mechanism.

The first step in establishing this technique was to compare the DEER spectra of the folded and unfolded mutant with each other, as it is shown in Figure 75.

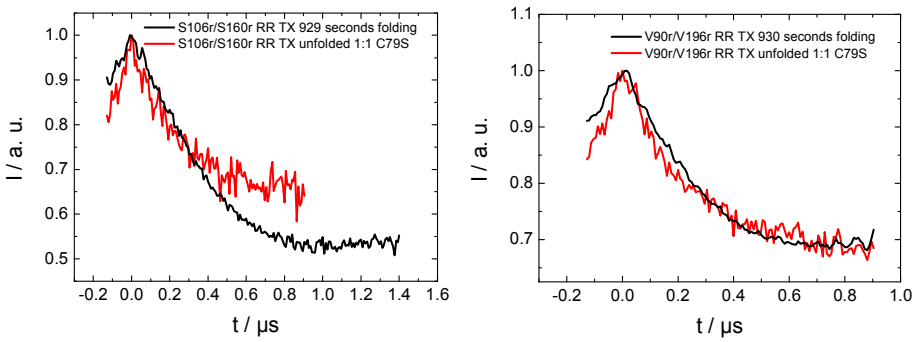


Figure 75: DEER data for folded and unfolded protein samples for different double mutants. Raw data.

The data in Figure 75 was not subjected to any analysis, in order to be sure to exclude any artifacts.

Figure 75 clearly shows significant differences in the DEER data as a function of protein folding time. Differences in the modulation depth for the unfolded S106r/S160r mutant, as well as decreased signal-to-noise ratio for both double mutant spectra are observed. This was already interpreted in terms of aggregation of unfolded protein (see sections 4.2.4. and 4.2.5.1.). The new information in Figure 75 is that the shape of the dipolar curve differs for the folded and unfolded protein. This is especially well seen in the case of the V90r/V196r sample. These differences in the dipolar evolution curves mean that there is a distance change when the protein folding proceeds. Thus the geometry change of protein during the folding process can be studied. The S106r/S160r mutant allows to study the rate of helix/helix organization combined with

helix formation and V90r/V196r mutant allows to study the rate of helix/helix organization, as can be seen from Figure 11 a.

This is, to our knowledge, the first proof of the possibility of using DEER for protein folding studies. The advantages offered by this method were already described in section 2.1.2.4.

Samples of the RR TX series corresponding to different folding times were prepared as described in section 3.1.4. with the only difference that for the DEER measurements only double mutants were used.

For the reasons described in section 4.1.5.1., dipolar spectra were analyzed that were obtained by Fourier transformation of background-corrected DEER data. Figure 76 shows such dipolar spectra for S106r/S160r and V90r/V196r mutants as a function of folding time.

The parameters for the S106r/S160r mutant were an 800 ns cutoff and background correction starting at 300 ns. The parameters for V90r/V196r mutant were a 900 ns cutoff and background correction starting at 136 ns. The discussion concerning the choice and influence of these parameters on the kinetic data is described later.

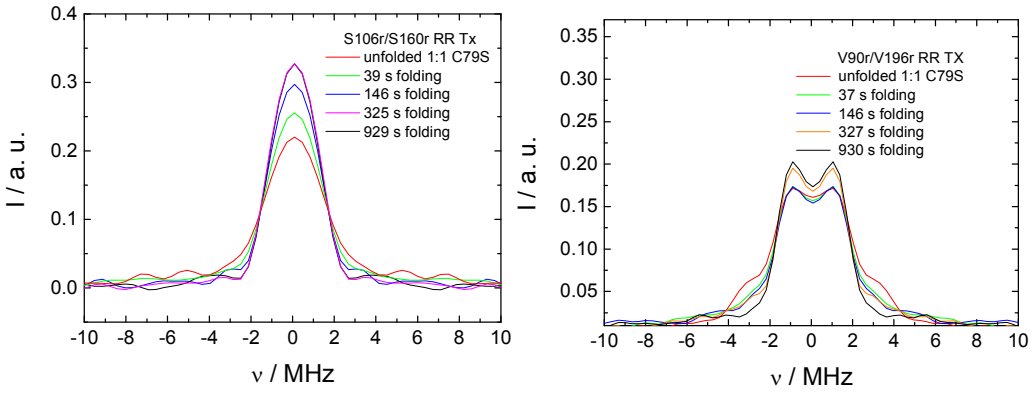


Figure 76: DEER dipolar spectra of the S106r/S160r mutant and the V90r/V196r mutant as a function of folding time. Spectra are normalized on the integral from -10 to 10 MHz.

There are clear and continuous differences between the dipolar spectra at different folding times for both S106r/S160r and V90r/V196r double mutants. The differences can thus be interpreted as time-dependent changes of the spin label distance at the particular mutation position. In analogy to studies of kinetics by UV/vis or vibration spectroscopy the presence of an isosbestic point, where intensity in the spectrum does

not change during folding, can indicate that there are no intermediates during folding **[Baumann 1962]**.

A similar continuous change in the dipolar data was also observed for a S106r/S160r sample with different micelle composition (data not shown), proving the reproducibility of the presented experiments.

As the different frequencies in the dipolar spectrum represent different distances it is possible to observe the decrease of the contribution of short distances by plotting the integral of the dipolar spectra at high frequencies as a function of folding time. Alternatively an increase of the dipolar peak intensity at low frequencies is interpreted as an increase of long distances during folding.

Orientation selection can also contribute to the changes in intensity in the dipolar spectrum in different protein states. A significant contribution from this source is however rather unlikely as the rather broad distance distributions imply broad orientation distributions of the labels.

The folding time constant for S106r/S160r mutant was determined by plotting the integral of the normalized dipolar spectrum at frequencies higher than the one of the isosbestic point (1.6 MHz) as a function of folding time (Figure 77 a.). An upper limit of 10.1 MHz was used for integration to avoid unnecessary accumulation of noise. Alternatively the inverse intensity of the dipolar peak maximum as a measure for the increase of the contribution at long distances was plotted as a function of the folding time (Figure 77 b.)

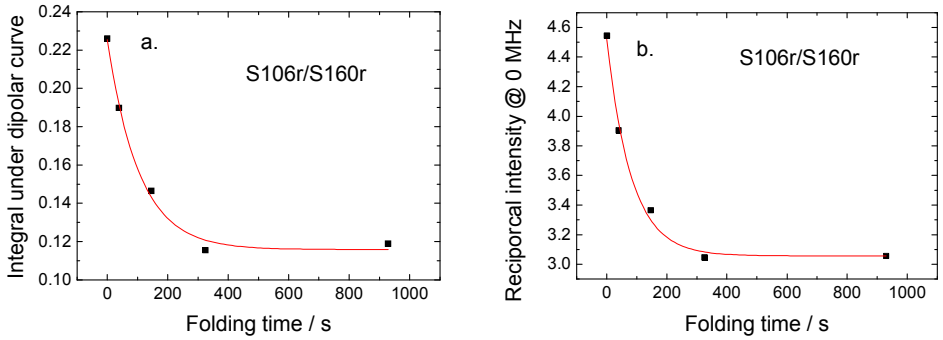


Figure 77: Determination of the folding time constant for the S106r/S160r mutant from DEER dipolar spectra. a. Integral of the dipolar spectrum from 1.57 MHz to 10.1 MHz as a function of folding time. b. Inverse intensity at 0 MHz as a function of folding time. The red curve is an exponential fit on the experimentally determined data points.

Figure 77 a. shows that the data can be fitted quite well by a monoexponential function. The time constant obtained for this process is 105 s. The relative error due to the deviation of the measured points from exponential behavior is 15%.

The data in the Figure 77 b. can also be fitted with a monoexponential function. In this case a time constant of 82 s is obtained and the relative error due to the deviation of the measured points from exponential behavior is 16%. The differences in the time constants determined from the different positions of the dipolar spectrum can originate from the fact that different subsets of the unfolded ensemble that correspond to different initial label-to-label distances fold with different rates. Such heterogeneity of the folding process becomes observable by DEER as different frequencies in the dipolar spectrum represent different distances. However due to the small difference in the time constants, which is still within the error range, such heterogeneity cannot be definitely proved.

Furthermore the determined time constant cannot be attributed solely to helix formation, as the flexible loop near position S160 might also change its conformation as the protein folds. The effect of helix formation is however assumed to dominate.

The same analysis was also performed on the dipolar data of the V90r/V196r mutant.

The folding time constant for the V90r/V196r mutant was determined by plotting the integral of the normalized dipolar spectrum from the isosbestic point at a frequency of 1.7 MHz to a frequency of 10.1 MHz where the intensity has decayed to the noise level, as a function of folding time (Figure 78 a.). Alternatively the inverse intensity of the dipolar peak maximum at 1.05 MHz as a measure for the contribution of long distances was plotted as a function of the folding time (Figure 78 b.)

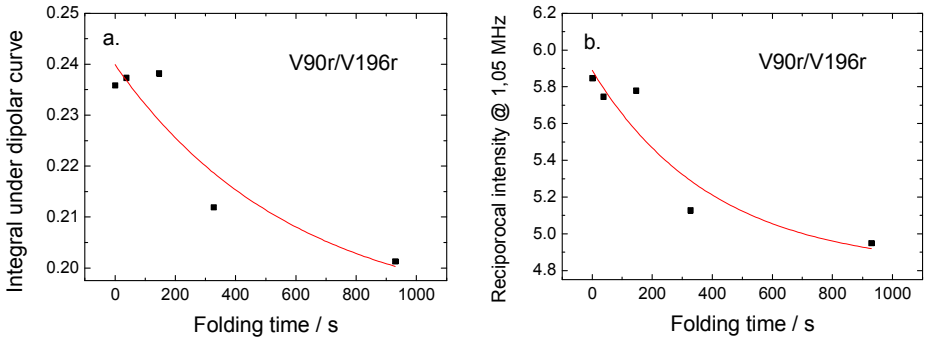


Figure 78: Estimation of a folding time constant for the V90r/V196r mutant from DEER dipolar spectra. a. Integral of the dipolar spectrum from 1.7 MHz to 10.1 MHz as a function of folding time. b. Intensity at 1.05 MHz as a function of folding time. The red curve is an exponential fit on the experimentally determined data points.

Within experimental error a continuous decrease of both parameters is observed.

The exponential fit of the folding data in the Figure 78 a. is poor compared to the one of the S106r/S160r sample. A folding time constant of 593 s with a relative error in the time constant determination due to the deviation of the measured points from exponential behavior of 130% is found.

The data in Figure 78 b. was also fitted by a monoexponential decay. Also in this case the fit is rather poor. The time constant obtained like that is 402 s. The error in the time constant determination due to the deviation of the measured points from exponential behavior is 90%. Whether the two different time constants originate from the real effect of different rates for different conformational subensembles cannot be clearly said as the difference is within experimental error. The data are better fitted by a sigmoidal curve. However the amount of data points does not allow for a firm conclusion whether the dependency is indeed sigmoidal rather than exponential.

An alternative way to process DEER data is to analyze directly the distance distributions. This is an intuitively more direct way as the distance distribution better visualizes the change in the protein geometry as it folds. In order to get reliable kinetic information from the distance distribution data, the parameters for DEER analysis between different mutants should be as similar as possible. This was however not always the case, as the best data fits for different mutation positions required different fitting parameters. The choice of the parameters for the analysis of the DEER data is not a trivial problem. For the V90r/V196r mutant the background correction time point was chosen automatically by the “DeerAnalysis2006” program. As the mutant is expected to have a rather short distance the short background correction time of 136 ns is assumed to be reasonable. For the S106r/S160r mutant the automatic background correction time was suggested to be 152 ns. However the fit for the folded mutant with this background parameter was not good. Reasonable fits were obtained by manually setting the background correction time to 300 ns. This makes sense, as the S106r/S160r mutant is known to have larger distances in the folded state than V90r/V196r mutant. Thus background correction fitting must start at longer dipolar evolution times **[Jeschke 2007]**. The problems concerning the ill-posedness of the conversion of background-corrected DEER data to distance distributions addressed in section 3.2.2.5. also have to be taken in account when using this data analysis procedure.

Figure 79 shows the background corrected DEER data, the fits and the distance distribution obtained from the fit for different folded and unfolded mutants.

The background correction time for S106r/S160r samples was 300 ns and the cutoff was at 800 ns, as the unfolded sample could not be measured much longer.

The background correction time for V90r/V196r samples was at 136 ns and the cutoff was at 900 ns.

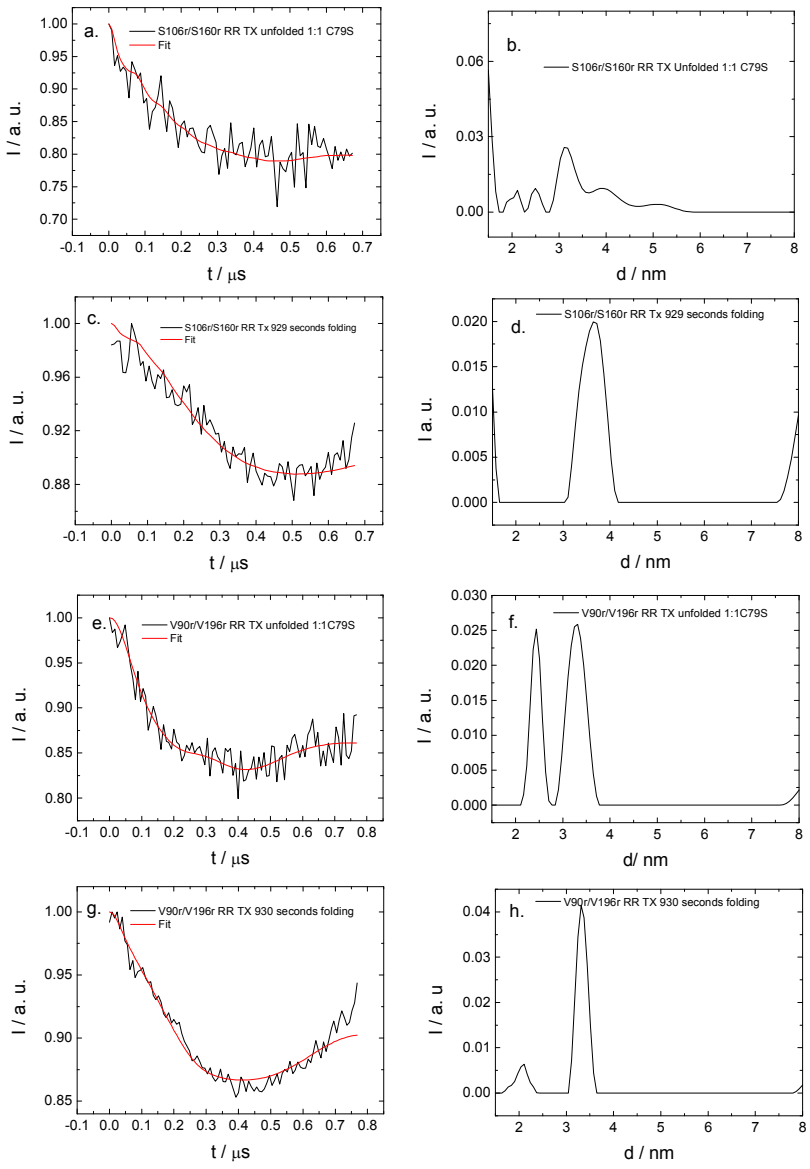


Figure 79: Background-corrected dipolar data with the fit and distance distribution obtained for different double mutants and different folding states.
a. background corrected DEER data with fit for the unfolded S106r/S160r RR TX mutant mixed 1:1 with C79S mutant. b. corresponding distance distribution.
c. background corrected DEER data with fit for the S106r/S160r RR TX mutant after 929 seconds of folding. d. corresponding distance distribution.
e. background corrected DEER data with fit for the unfolded V90r/V196r RR TX mutant mixed 1:1 with C79S mutant. f. corresponding distance distribution.
g. background corrected DEER data with fit for the V90r/V196r RR TX mutant after 930 seconds of folding. d. corresponding distance distribution. The red curves in a., c., e. and g. reperesent the fit.

From Figure 79 it is seen that the fits of the DEER data for the S106r/S160r and V90r/V196r mutant are reasonable for the two extreme cases of unfolded and completely folded protein. This proves that parameters for the analysis of the DEER data were chosen correctly.
The strength of this analysis method is clearly seen from Figure 79 as it is possible to say that the protein folds from a rather compact state into a less compact state. It is also possible to determine the average distance and amount of spins coupled during the protein folding, as shown in Tables 8 and 9.
There is a clear difference between the distance distribution of the folded and unfolded S106r/160r mutant. The main effects are seen in the intensity of the peaks corresponding to different distances. In order to analyze folding kinetics of this protein domain a folding coefficient was defined as

$$\text{Folding coefficient (S106r/S160r)} = A/(A+B)$$

Here A is the integral intensity of the distance distribution between 1.73 nm and 2.88 nm and B is the integral intensity of the distance distribution between 2.88 nm and 4.48 nm. These are the regions where most of the changes are observed during the folding. The error due to the overlap of the distance distributions in the integration region has to be taken into account for some of the samples. It is however assumed that the error introduced by the ill-posedness of the problem, as explicitly shown later, is larger.
By plotting the folding coefficient as a function of protein folding time a kinetic curve is obtained with a time constant of 90 s, as shown in Figure 80. The relative error due to the deviation of the measured points from exponential behavior is 22%.

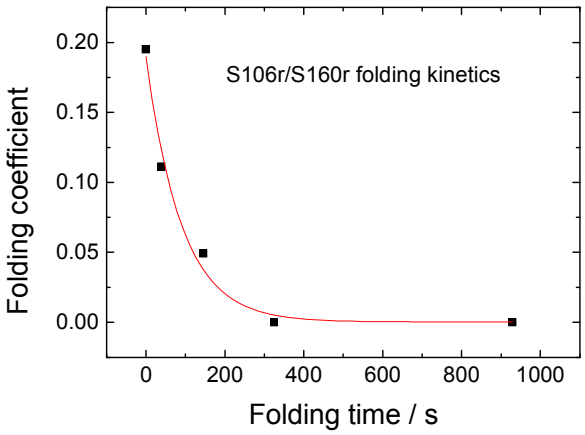


Figure 80: Determination of the folding time constant for the S106r/S160r mutant based on the integration of distance distribution. The red curve is an exponential fit on the experimental data points.

In Table 8 also the change of the average distance and the number of spins that are coupled to each other as a function of folding time is shown.

S106r/S160r		
Folding time / s	**Average distance / nm**	**Number of spins**
0	3.14	1.43
39	3.62	1.26
146	3.78	1.18
325	3.84	1.17
929	3.98	1.21

Table 8: Average distance and number of coupled spins as a function of folding time for S106r/S160r mutant.

From the Table 8 a continuous change of the average distance as a function of the folding time is clearly seen. This proves the idea of folding from the aggregated compact state.

No correlation between the number of coupled spins and folding time is observed. This is expected as the extent of spin labeling should not change during folding. The increased number of spins for the 0 seconds folding time can be due to the protein aggregation.

Also the V90r/V196r mutant shows differences in the distance distribution as a function of folding time.

A similar analysis as described previously was performed with the V90r/V196r mutant. In order to determine the folding kinetics of this protein part a folding coefficient was defined as

$$\text{Folding coefficient (V90r/V196r)} = A/(A+B)$$

A is the integral intensity of the distance distribution between 1.77 nm and 2.84 nm and B is integral intensity of the distance distribution between 2.84 nm and 3.84 nm. These are the regions where most of the changes are observed during folding, and also as few as possible signal coming from the unwanted proton modulation is analyzed.

The Figure 81 shows the plot of the folding coefficient of V90r/V196r mutant as a function of the folding time.

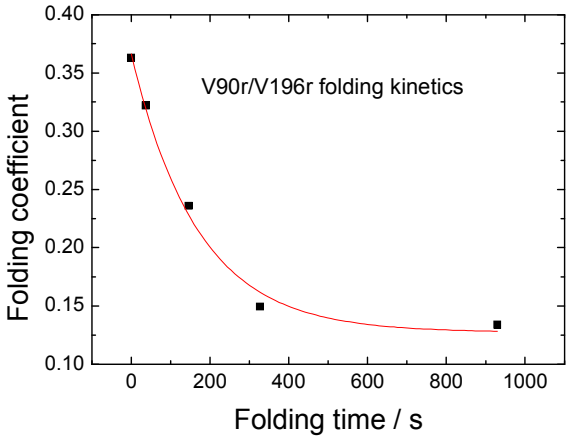


Figure 81: Determination of the folding time constant for the V90r/V196r mutant based on the integration of distance distribution. The red curve is an exponential fit of the experimental data points.

The time constant for V90r/V196r mutant determined by that procedure is 169 s. The relative error due to the deviation of the measured points from exponential behavior is 16%.

In Table 9 also the change of the average distance and the number of spins that are coupled to each other is shown.

V90r/V196r		
Folding time / s	**Average distance / nm**	**Number of spins**
0	3.05	1.3
37	3.04	1.26
146	2.91	1.27
327	3.06	1.27
930	3.2	1.22

Table 9: Average distance and number of coupled spins as a function of folding time for V90r/V196r mutant.

The average distance does not seem to change as a function of folding time. This can be explained by the similarity of average distances in different protein states. However as clear changes were observed in the Figure 78 the kinetic effect is proved to be real. For this double mutant the distance distribution narrows during folding without significantly changing the mean value. This shows a further advantage of DEER for such kind of studies, as it allows to determine in addition to the average distance between the spins also the distance distribution.

The number of coupled spins also does not seem to change. In the unfolded state the number of spin is not increased, as it was observed for the S106r/S160r mutant. The average spin number may be not representative for the unfolded mutants due to the decreased measurement quality.

The next step in the folding kinetics analysis is to compare the folding rates of different protein segments with each other. This is important, as the described folding models can be explicitly tested like that. There are some hints in the DEER data that the S106/S160 protein region shows faster kinetics. Whether this can be used as the proof for the two stage-folding model is not clear as in order to compare the folding time constants of the two protein segments with each other the error of the measurements has to be estimated.

The first error source, as already described earlier, is the one due to the deviation of the measured points from fitted exponential behavior. Because of this error the data analysis via dipolar spectra does not allow to draw a firm conclusion about the difference in the folding time constants of the studied protein regions. However in the

case of kinetic analysis of the distance distribution the difference in the time constants between two double mutants are outside this error range, which gives support to the two-stage folding model.

Another source of error is the one introduced by the background correction of the DEER data. In order to estimate the error introduced by this procedure analysis was performed on the S106r/S160r data with DEER parameters of V90r/V196r data and vice versa. The folding time constants determined from the analysis of dipolar spectra determined like that were quite similar to the ones determined with other background parameters leaving the deviation of the measured points from exponential behavior as a main error source.

The analysis of the distance distribution of S106r/S160r mutant with V90r/V196r parameters showed a poor fit for the folded mutant. The folding time constant determined with these parameters was 20 s, corresponding to a 600% error due to the background correction and Tikhonov regularization procedure. This result shows the main advantage of the dipolar spectra analysis in order to determine small changes in the distance distribution as the Fourier transformation is less sensitive to errors in the separation of the intermolecular background from intramolecular distances compared to Tikhonov regularization.

As there should be an additional error in the folding time constants determined from the analysis of the distance distributions it may thus be that the small difference in the folding time constants between S106r/S160r mutant and V90r/V196r mutant is an artifact. Thus in order to determine whether the formation of a helix in LHCIIb happens faster than mutual helix arrangement only the dipolar data will be considered.

As seen from the Figure 78 the kinetic data cannot be well fitted by the exponential function. This introduces significant errors in the determined folding time constant. In order to see whether one of protein segments folds faster than the other the kinetic data points for two different samples were normalized according to equation ***53***.

$$C=I_i/(I_0-I_\infty) - I_\infty/(I_0-I_\infty) \qquad \textit{Eq. 53}$$

Here C is a normalized folding coefficient that allows to compare the relative amount of distance change, I_i is the integral value from figures 77 a. and 78 a. at the particular time point, I_0 is the integral value from figures 77 a. and 78 a. at time zero and I_∞ is the integral value from figures 77 a. and 78 a. at the last measured time point that should correspond to complete folding.

The folding coefficient determined by the integration of the dipolar curves from isosbestic point was used. This method was chosen as it provides similar information as the analysis of central peak intensities of dipolar curve however the better distance weighting due to the integration is expected.

The introduction of this normalization parameter allows to directly compare the kinetic changes in distances for different samples as a function of the folding time as shown in Figure 82. Also the kinetic changes obtained with different methods can be compared like that as will be shown later.

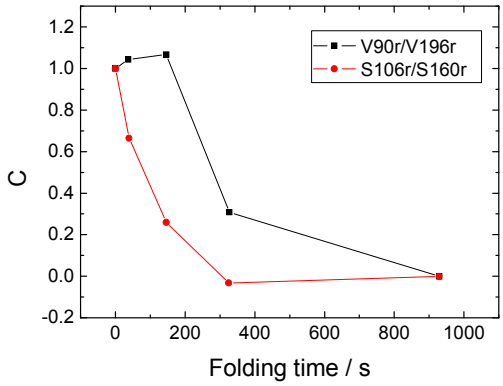


Figure 82: Comparison of the relative changes in the dipolar spectra for different double mutants. The points are connected in order to better see the trends.

The values bigger than one and smaller than zero are considered to be due to the experimental errors.

From Figure 82 it is seen that the changes in the S106r/S160r mutant happen faster than in the V90r/V196r mutant. This can be an evidence for the faster formation rate of the helix, which is monitored by S106r/S160r mutant than the rate of interhelix arrangement that is monitored by V90r/V196r mutant.

In order to further verify this assumption the reason for the strong deviation of V90r/V196r kinetic data from exponential behavior was searched. This reason could be the short measurement time for the unfolded sample and the poor signal-to-noise ratio of these data. As the measurement at zero time is subjected to the biggest error, an analysis of the DEER data without this point was performed. This has a further advantage that the data could be analyzed for a longer evolution time, and thus the dipolar evolution curve contains more information. Background correction was chosen to give the best possible fit of the experimental data. The fitting parameters were 404

ns background correction and 1400 ns cut off time for the S106r/S160r mutant and 184 ns background correction and 1400 ns cut off time for the V90r/V196r mutant.

The folding time constants were determined by integration of the dipolar curves, normalized on the integral from -5 to 5 MHz, from the isosbestic point (0.9 MHz for S106r/S160r mutant and 1.3 MHz for V90r/V196r mutant) to 5 MHz and plotting the value of this integral as a function of folding time as shown in the Figure 83.

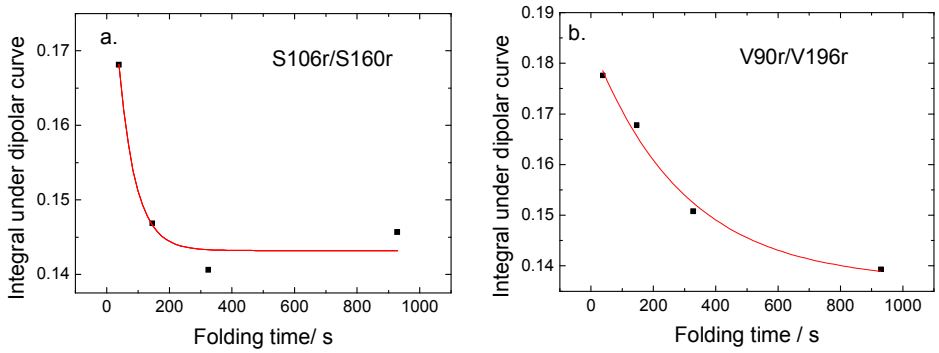


Figure 83: Determination of a folding time constant of S106r/S160r (a.) and V90r/V196r (b.) mutants based on the integration of their DEER dipolar spectrum for different folding times. The data point at zero time was disregarded. The red curve is an exponential fit of the experimental data points.

The time constants determined like that were 54 s for the S106r/S160r mutant (67% error due to the deviation of the measured points from exponential behavior) and 294 s for V90r/V196r mutant (29% error due to the deviation of the measured points from exponential behavior). This is further proof that the measured kinetic effect is real. The kinetic constants in this time range still show differences between the two double mutants. This gives further evidence that the folding of different protein parts happens on different time scales.

The results of this section allow the conclusion that the protein folding kinetics investigation with DEER spectroscopy is possible. The method allows following the structurally resolved protein folding process of certain spin labelled protein segments. The segments that could be investigated were defined by mutation positions.

In the case of the S106r/S160r mutant the time dependent changes in the DEER spectrum are mostly assumed to be determined by helix formation. However also the movement of other flexible structural elements and interhelix arrangement can contribute to the measured distance change. In the case of V90r/V196r mutant kinetic changes are assumed to be determined by interhelix arrangement. The folding time

constant determination based on the experimental DEER data suggests faster kinetics for the S106r/S160r mutant than for the V90r/V196r mutant.
This in turn can be a hint that helix formation happens faster than interhelix arrangement which supports the two-stage folding model for LHCIIb.
Furthermore structural information about folding intermediates can potentially be obtained with the method developed here. This is a great advantage, as the structural features of folding intermediates are assumed to be a central problem of structural biology **[Gianni 2007]**.
The developed method of investigating protein folding with DEER can potentially provide experimental information on the folding mechanism of any protein of interest if the site-directed spin labeling and freeze-quench techniques can be applied.

4.2.6. New information about LHCIIb folding

At this point it is of interest to compare the folding time constants determined by DEER spectroscopy with the ones determined by ESEEM and longitudinal relaxation measurements and also with alternative methods like fluorescence spectroscopy described in **[Dockter 2008]** and to present a detailed LHCIIb folding model based on these measurements.
The folding time constants determined disregarding the data at zero time that are influenced by aggregation (see previous section for details) are 54 s (±67% error) for the S106r/S160r mutant and 294 s (±29% error) for the V90r/V196r mutant. As described previously, the folding time constants obtained by DEER give information about the global changes in the secondary and tertiary structure of the protein.
The folding time constant obtained with ESEEM for the V196r mutant, which gives the information of local environment change, was determined to be 486 s (±30% error). This is clearly longer than the time constant determined with DEER for the S106r/S160r mutant and in the same range as the time constant determined for the V90r/V196r mutant. The differences in folding time constants might originate from the histidin tags on C-terminus of some of the mutants. However due to the fact that both the V196r and the S106r/S160r mutant have a histidine tag, and show different folding time constants, and V90r/V196r has no histidine tag and shows time constant similar to the V196r mutant allows to safely exclude the effect of the histidine tag on folding rate.
Thus the differences in determined folding time constants may be a hint that there is a hierarchy of structure formation when the protein folds. However, as the last data point

in ESEEM and T_1 folding kinetics is measured at a longer folding time than in DEER measurements, care in the comparison of different time constants is needed.

Figure 84 shows the C values, defined in equation ***53***, for different pulse EPR techniques and mutants as a function of folding time. It was shown previously that by defining the C value differences in the rate of folding can be recognized without uncertainties involved in data fitting.

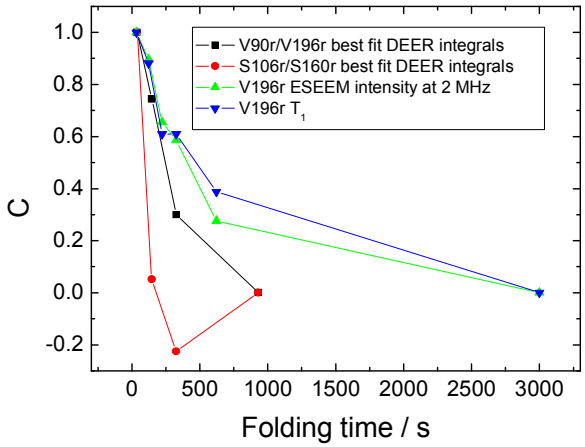


Figure 84: Kinetic master plot for the different mutants and different pulsed EPR techniques used in studying LHCIIb folding. For the DEER data the points from Figure 83 a. and 83 b. were used. The points are connected in order to better see the trends.

Figure 84 again shows that the folding kinetics observed by DEER is most probably faster for the S106r/S160r mutant than for the V90r/V196r mutant.

The dependencies of the ESEEM intensity and of T_1 on folding time are similar, which supports the finding that they both represent the environmental change during folding.

The master plot in Figure 84 also indicates that the time scale of changes is similar for DEER of the V90r/V196r mutant and ESEEM and T_1 of the V196r mutant. This can be a hint that these processes of interhelix arrangement and environmental change may be coupled. Different positions of the last measurement point should be taken into account.

It is known that fluorescence measurements can be used to determine kinetics of LHCIIb assembly with pigments **[Horn 2004]**. As the experimental conditions during this work differed from the ones described in the reference, the fluorescence measurements were repeated with the samples prepared for EPR as described in the experimental section. They were labelled with a fluorescence dye on position S160.

The experiments were performed by C. Dockter and G. Grundmann in the group of Prof. Dr. H. Paulsen.

A time dependent change of the fluorescence transfer is observed when the protein is mixed with the pigments as cofactors. The fact that there is a fluorescence transfer under these experimental conditions again proofs the correctness of the new protein preparation method. The time constants obtained like that are between 13 and 37 seconds for the fast step, and between 85 and 324 seconds for the slow step. This is in the same order of magnitude as in the previously described measurements **[Horn 2007]**. The two time constants observed by fluorescence spectroscopy are similar as observed with the DEER for two different mutation positions. Whether this can be a proof that the chlorophyll a binding as a fast kinetic step **[Horn 2007]** is connected to helix formation, assumed to be determined by DEER on the S106r/S160r mutant, and chlorophyll b binding kinetics as a slower step **[Horn 2007]** is associated with interhelix arrangement, assumed to be determined by DEER on the V90r/V196r mutant, cannot be decided on the basis of the present data.

According to all the described facts a model for protein folding during LHCIIb self assembly was constructed. It is shown in the Figure 85. The fluorescence data provides only indirect information about protein folding as the pigment organization is being observed and not the changes in protein backbone. Therefore only the EPR data was used to construct it.

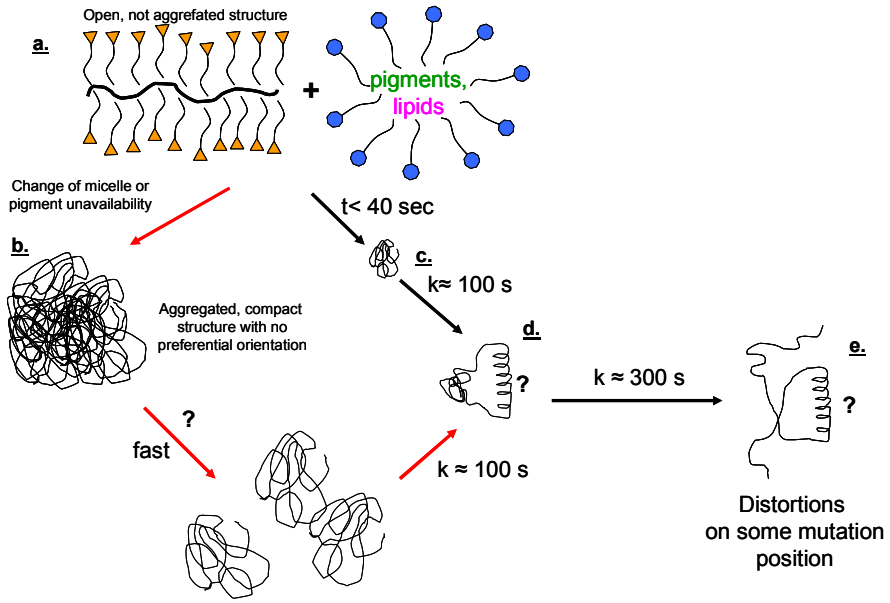


Figure 85: LHCIIb folding model determined by EPR.

The current state of knowledge on the mechanism of protein folding during *in vitro* self assembly of LHCIIb is the following:

The LHCIIb protein in LDS micelles (Figure 85 a.), which shows a rather open structure that is not aggregated, and probably behaves like a random coil, is mixed with pigments and lipids. Then two processes can occur. When the micelle composition is changed and no or not sufficient amount of pigments is present the protein aggregates (Figure 85 b.). The aggregated structure is rather compact, has no preferential orientation in the micelle and also behaves like a random coil. Whether the aggregated structure can transform to the folded state when pigments become available later is not clear.

When the micelle composition is changed and a sufficient amount of pigments is present the protein assumes a more compact state, as explicitly shown by DEER spectroscopy (Figure 85 c.). Such a hydrophobically collapsed state during membrane protein folding was already reported on another protein **[Kleinschmidt 2002]**. It might thus be a rather general effect. Then a first kinetic step happens that is associated with helix formation (Figure 85 d.). The time constant of this step is around 100 s. Whether other structure arrangement kinetics can influence the determined folding is not very clear, that is why a question mark is shown near the helix. Afterwards a slower step that is assumed to be associated with the organization of tertiary structure with a time constant of around 300 s happens (Figure 85 e.). The local environment change at position V196 happens rather slowly, with a time constant of around 500 s. This time constant is in the range of the assumed tertiary structure organization step. This indicates that interhelix organization and local environmental change at position V196 can be coupled.

5. Conclusion and Outlook

Conclusion

In this work different EPR methods were applied to determine details of structure and folding of the major light harvesting complex LHCIIb of green plants. By applying the technique of site-directed spin labeling the protein domains of particular interest could be studied thus significantly reducing complexity of the problem.

The structural investigations revealed the possibility of applying ESEEM for quantifying water accessibility of protein sites in a medium sized membrane protein in a micelle that simulates the membrane. This water accessibility can be interpreted with some caution in terms of positioning of the membrane protein within a detergent micelle. By systematic comparison with other EPR methods for characterizing water accessibility, three-pulse ESEEM on samples prepared with deuterated water was shown to be the method of choice for obtaining this kind of information. Further it was shown that combination of ESEEM with these other methods gives detailed information concerning the local environment of the spin-labelled residue in a state that is similar to the one of the protein under ambient conditions.

Confirmation of high water accessibility of the structurally not resolved LHCIIb N-terminus, which is known to be important for regulatory function, was obtained.

Furthermore investigations of global protein structure were performed with the DEER experiment that can measure distance distributions between two spin labels in the range between about 1.5 and 8 nm. These investigations provided new geometrical constraints on protein structure in a state that is similar to the one in thylakoid membranes. The constraints are in satisfying agreement with simulated data based on the X-ray structure of the LHCIIb and on a rotamer library approach to account for the conformational distribution of the spin label. This indicates, first, that the structure of LHCIIb in detergent micelles closely resembles the crystal structure, and, second, that the width of the distance distributions can be predicted rather well by the rotamer library approach.

Furthermore changes in protein geometry as a function of micelle composition were observed and investigated. In most cases these changes were slight, but for certain domains of the protein the results indicate that tertiary and possibly even secondary structure can be different in different types of micelles. These studies are of relevance, as protein-membrane interactions are important issues in structural biology. The

approaches developed during this work can provide better understanding of this phenomenon. Also some hints were obtained that EPR can be used as a tool for the selection of micelles that are most suitable for protein crystallization, as micelle compositions that lead to narrow reproducible distance distributions and reproducible accessibilities should probably also provide better crystallization results.

After understanding the effects that dominate the EPR data, the developed arsenal of methods was applied to obtain structural information about unfolded and partially folded LHCIIb as well as structurally resolved kinetic data for the LHCIIb folding/self-assembly process.

Special attention was paid to the unfolded state of the apoprotein, as not much experimental data is available on it. Structural information on the LHCIIb apoprotein was obtained with the EPR methods presented here. Also hints were found and investigated that the structure of the unfolded apoprotein changes as a function of solvent and micelle composition.

Relaxation measurements proved to be useful in detecting the presence of aggregated protein, which was not easily possible by light scattering due to the signal from micelles. ESEEM showed that, unlike in the folded state, different residues in the unfolded state have similar water accessibilities, suggesting that there is no preferential orientation of the unfolded protein in the micelle for the RR TX sample composition. DEER experiments in the unfolded state gave very similar results for different mutation positions with unexpectedly strong contributions at short distances, suggesting that the unfolded membrane protein is in a more compact or even collapsed state for the RR TX sample composition. Simulations of the experimental DEER dipolar evolution data of the unfolded protein based on the random coil model showed good correlation. Thus it was concluded that the unfolded proteins probably behave as random coil polymers. Furthermore differences in the scaling exponents in fits of the coil model for different double mutants were obtained that hint on helix preorganization in some regions for unfolded LDS samples.

Thus rather comprehensive structural information on the unfolded protein could be obtained. The developed approach may be of interest for studies on the causes of the protein misfolding.

Finally the kinetics of formation of different structural elements during protein folding was addressed.

Simple and fast EPR techniques, such as field-swept echo-detected EPR and measurement of the transversal relaxation time, were shown to be useful for determining the quality of protein folding, and thus were routinely used for this purpose.

DEER experiments allowed for measuring the kinetics of formation of a single transmembrane helix (secondary structure) and of interhelix arrangement (tertiary structure), whereas ESEEM as well as longitudinal and transversal relaxation measurements allowed for observing and quantifying changes of the environment of particular residues during folding. Thus the developed methods allow for getting local and global structural information on the folding process.

By combining all the experimental EPR information a structurally resolved LHCIIb folding model could be constructed. According to this model helix formation during LHCIIb folding happens somewhat faster than interhelix arrangement, which is in agreement with the two-stage folding model. Furthermore hints were obtained that interhelix arrangement and changes in local water accessibility of residues in the protein happen on a similar time scale.

This work thus showed the advantages and perspectives that EPR offers for the investigation of protein structure and folding. The approaches developed during this work are in principle generally applicable to any protein or other self-assembled structure of interest, provided that site-directed spin labeling is possible and the folding process proceeds on a time scale that is accessible to freeze quench techniques.

Outlook:

Structure

The measurements presented in this work were performed on LHCIIb reconstituted into micelles. Such an environment is similar but not identical to the native lipid bilayer. A comparison of accessibilities in detergent micelles and natural thylakoid membranes would be of great interest as it would allow for a better understanding of the effect of lipids on protein conformation.

In addition to DEER the methods developed in this work could also be used to characterize conformational changes in the N-terminal domain of LHCIIb that are expected to occur during the regulatory processes.

Folding

After establishing an EPR toolbox for protein folding investigations in this work, the following issues can be addressed:

- Investigation of protein folding at different temperatures would allow for the determination of the activation energy of different processes in the protein that happen during folding.

- Investigation of folding kinetics of LHCIIb mutants to determine the influence of particular amino acid residues on stabilization of the structure and folding behavior. This can provide information about the transition state during protein folding (Φ analysis **[Lesk 2004]**)

This work showed that formation of the secondary structure of the transmembrane helix between positions S106 and S160 is faster than formation of the tertiary structure element of crossed helices defined by positions V90 and V196. However the difference between the folding time constants is surprisingly small, possibly because formation of the helix was accessed via the distance between the residues S106 and S160. The final structure between these residues is established only after parts of a lumenal and stromal loop have also assumed their final conformation. Proper folding of these loop parts may in turn depend on formation or at least on preorganization of tertiary structure.

New double mutants with spin labels on the opposite site of the helix close to the helix ends that do not lead to structural distortions or incomplete folding have to be found and studied in order to obtain a better estimate of the folding time constant of helix formation. If singly spin-labelled mutants are found for whom mobility of the spin label at ambient temperature is restricted in the folded but not in the unfolded state it would also be of interest to apply room temperature CW EPR as an alternative method. In

principle, such experiments are feasible as the mobility of spin labels in general depends on the secondary structure around it **[Morin 2006]**.

The methods developed during this work can be applied to other proteins and other self-assembling structures. In particular, it would be of interest to perform similar studies as in this work on other transmembrane and soluble proteins and compare them with the results obtained here. On the one hand this would further increase our knowledge about the principles of protein folding. On the other hand it would also clarify to which extent the current uncertainties of the results are due to complexity of the self-assembly process of LHCIIb and to which extent they are due to limitations of the site-directed spin labeling EPR methodology.

EPR

As the potential of DEER spectroscopy is not yet fully utilized, studies on DEER method development were also performed during this work.

The orientation averaged DEER experiment as introduced in **[Godt 2006]** was shown to give reproducible results for further rigid model biradicals. This method may be of interest for other spin pairs with non-random relative orientations.

TRIER spectroscopy was tested as a potential method for measuring angles between spin-spin vectors, based on the idea initially described in **[Pannier 2000 b]**. This is a two-dimensional experiment that relies on separate excitation of three spins by three microwave frequencies.

Although measurements performed on asymmetric triradicals provided raw data that were modulated along both dimensions, the spectra obtained by two-dimensional Fourier transformation did not allow for unambiguous peak assignment and extraction of the angles in the triangle. This was traced back to ambiguities in the background correction procedure of the raw data in two dimensions, which seems to be a non-trivial procedure. Other theoretical approaches for analyzing such data can be of interest.

Information about relative spin label orientations can be obtained with W-band DEER **[Polyhach 2007]**. In fact it may be of interest to perform distance measurements in W-band on membrane proteins, even when no orientation correlation between the spin labels is observed. The increased absolute sensitivity of W-band EPR spectroscopy allows for measurements with a smaller sample amount. Thus the method could provide the same distance information as obtained in X-band with the sensitivity of W-band measurement. However if the amount of sample is sufficient for an X-band measurement and the sample concentration is limited, as in this work, sensitivity is better at X-band.

It would be of great interest to perform pulse EPR studies on the spin labelled proteins in shock frozen living cells in order to study protein geometry changes in vivo. For that the sensitivity of the methods should be significantly increased. The development of high field techniques with shorter microwave pulse times can be one way for such a sensitivity increase.

6. References

[Abragam 1961] Abragam, A. The Principles of Nuclear Magnetism (1961) 406-411 Clarendon Press, Oxford

[Arora 2001] Arora, A., Tamm, L. K. *Curr. Opin. Struct. Biol.* (2001) **11** 540-547

[Altenbach 1989] Altenbach, C., Flitsch, S. L., Khorana, H. G., Hubbell, W. J. *Biochemistry* (1989) **28** 7806-7812

[Altenbach 1994] Altenbach, C., Greenhalgh, D. A., Khorana, H. G., Hubbel, W. J. *Proc. Natl. Acad. Sci.* (1994) **91** 1667-1671

[Allen 2001] Allen, J. F., Forsberg, J. *Trends Plant Sci.* (2001) **6** 317-326

[Antonov 1984] Antonov, O. F., Sabirov, R. K. *Phys. Stat. Sol. (b)* (1984) **125** 117-120

[Bartucci 2003] Bartucci, R., Guzzi, R., Marsh, D., Sportelli, L. *Biophys. J.* (2003) **84** 1025-1030

[Bartucci 2006] Bartucci, R., Erilov, D. A., Guzzi, R., Sportelli, L., Dzuba, S. A., Marsh, D. *Chem. Phys. Lipids* (2006) **141** 142-157

[Baumann 1962] Baumann R. P. Absorption Spectroscopy (1962) John Wiley & Sons

[Bender 2004] Bender, A. PhD Thesis (2004) Mainz University

[Berliner 1984] Berliner, L. J., Reuben, J. Biological Magnetic Resonance (1984) **6** Plenum Press, New York

[Berliner 2000] Berliner, L. J., Eaton, G. R., Eaton, S. S. Biological Magnetic Resonance (2000) **19** "Distance Measurements in Biological Systems by EPR" Kluwer Academic/Plenum Publishers, New York

[Blumberg 1960] Blumberg, W. E. *Phys. Rev.* (1960) **119** 79-84

[Bode 2007] Bode, B. E., Margraf, D., Plackmeyer, J., Dürner, G., Prisner, T. F., Schiemann, O. *J. Am. Chem. Soc.* (2007) **129** 6736-6745

[Boggash 2006] Boggash, S. PhD Thesis (2006) Mainz University

[Booth 1996] Booth, P. J., Paulsen, H. *Biochemistry* (1996) **35** 5103-5108

[Booth 2006b] Booth, P. J., Curnow, P. *Curr. Opin. Struc. Biol.* (2006) **16** 480-484

[Borbat 1999] Borbat, P. P., Freed, J. H. *Chem. Phys. Lett.* (1999) **313** 145-154

[Bowie 2005] Bowie, J. U. *Nature* (2005) **438** 581-589

[Brown 1979] Brown, I. M. in: Time-domain electron spin resonance Eds. L. Kevan. R. N. Schwartz (1979), Wiley, New York

[Buchner 2005] Buchner, J., Kiefhaber, T. Protein Folding Handbook (2005) **2** WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

[Carmieli 2006] Carmieli, R., Papo, N., Zimmerman, H., Potapov, A., Shai, Y., Goldfarb, D. *Biophys. J.* (2006) **90** 492-505

[Carpentier 2004] Carpentier, R. Photosynthesis Research Protocols (2004) 93-104 Humana Press

[Carrington 1967] Carrington, A., McLachlan, A. D. Introduction to Magnetic Resonance with applications to chemistry and chemical physics (1967) Harper & Row New York; John Weatherhill, INC., Tokyo

[Cashmore 1984] Cashmore, A. R. *Proc. Natl. Acad. Sci.* (1984) **81** 2960-2964

[Chen 2008] Chen, Y., Ding, F., Nie, H., Serohijos, A. W., Sharma, S., Wilcox, K. C., Yin, S., Dokholyan, N. V. *Arch. Biochem. Biophys.* (2008) **469** 4-19

[Damaschun 1986] Damaschun, G., Gernat, C., Damaschun, H., Bychkova, V. E., Ptitsyn, O. B. *Int. J. Biol. Macromol.* (1986) **8** 226-230

[De Gennes 1979] de Gennes, P.-G. Scaling Concepts in Polymer Physics (1979) Cornell University Press, Ithaca, New York

[De Simeone 2007] De Simeone F., Guzzi R., Sportelli L., Marsh D., Bartucci R. *Biochim. Biophys. Acta* (2007) **1768** 1541-1549

[Dockter 2005] Dockter, C. Diploma thesis (2005), Mainz University

[Dockter 2008] Dockter, C. PhD thesis (in preparation), Mainz University

[Du 1995] Du, J.-L., Eaton, G. R., Eaton, S. S. *J. Magn. Reson.*, Series A (1995) **115** 213-221

[Dzuba 1996] Dzuba, S. A. *Phys. Lett. A* (1996) **213** 77-84

[Eaton 2001] Eaton, S. S., Harbridge, J., Rinard, G. A., Eaton, G. R., Weber, R. T. *Appl. Magn. Reson.* (2001) **20** 151-157

[Engelman 2003] Engelman, D. A., Chen, Y., Chin, C.-N., Curran, A. R., Dixon, A. M., Dupuy, A. D., Lee, A. S., Lehnert, U., Matthews, E. E., Reshetnyak, Y. K., Senes, A., Popot, J.-L. *FEBS Lett.* (2003) **555**122-125

[Erilov 2005] Erilov, D. A., Bartucci, R., Guzzi, R., Shubin, A. A., Maryasov, A. G., Marsh, D., Dzuba, S. A., Sportelli, L., *J. Phys. Chem. B* (2005) **109** 12003-12013

[Fanucci 2006] Fanucci, G. E., Cafiso, D. S. *Curr. Opin. Struc. Biol.* (2006) **16** 644-653

[Farahbakhsh 1992] Farahbakhsh, Z. T. , Altenbach, C., Hubbell, W. L. *Photochem. Photobiol.* (1992) **56** 1019-1033

[Fitzkee 2004] Fitzkee, N. C., Rose G. D. *Proc. Natl. Acad. Sci.* (2004) **101** 12497-12502

[Flory 1969] Flory, P. J., Statistical Mechanics of Chain Molecules (1969) John Wiley and Sons, Inc.

[Gianni 2007] Gianni, S., Ivarsson Y, Jemth, P., Brunori, M., Travaglini- Allocatelli, C. *Biophys. Chem.* (2007) **128** 105-113

[Godt 2006] Godt, A., Schulte, M., Zimmermann, H., Jeschke, G. *Angew. Chem. Int. Edit.* (2006) **45** 7560-7564

[Griffith 1974] Griffith, O. H., Dehlinger, P. J., Van, S. P., *J. Membrane Biol.* (1974) **15** 159-192

[Hansen 1992] Hansen, P. C., *SIAM Rev.* (1992) **34** 561- 580

[Havlin 2005] Havlin, R. H., Tycko, R. *Proc. Natl. Acad. Sci.* (2005) **102** 3284-3289

[Henderson 2004] Henderson, R. *Q. Rev. Biophys*. (2004) **37** 3-13

[Heuck 2002] Heuck, A. P., Johnson, A. E. *Cell Biochem. Biophys.* (2002) **36** 89-101

[Hilger 2005] Hilger, D., Jung, H., Padan, E., Wegener, C., Vogel, K.-P., Steinhoff, H. -J., Jeschke, G. *Biophys. J.* (2005) **89** 1328–1338

[Hinderberger 2004] Hinderberger, D., Spiess, H. W., Jeschke, G. *J. Phys. Chem. B* (2004) **108** 3698-3704

[Hobe 1994] Hobe, S., Prytulla, S., Kühlbrandt, W., Paulsen, H. *EMBO J.* (1994) **13** 3423-3429

[Hobe 1995] Hobe S., Förster R., Klingler J., Paulsen H., *Biochemistry* (1995) **34** 10224-10228

[Horn 2002] Horn, R., Paulsen, H. *J. Mol. Biol.* (2002) **318**, 547-556

[Horn 2004] Horn, R., Paulsen, H. *J. Biol. Chem.* (2004) **279** 44400-44406

[Horn 2007] Horn, R., Grundmann, G., Paulsen, H., *J. Mol. Biol.* (2007) **366** 1045-1054

[Hubbell 2000] Hubbell, W. L., Cafiso, D. S., Altenbach, C. *Nat. Struct. Biol.* (2000) **7** 735-739

[Huber 2001] Huber, M., Lindgren, M., Hammarström, P., Mårtensson, L.-G., Carlsson, U., Eaton, G. R., Eaton, S. S. *Biophys. Chem.* (2001) **94** 245-256

[Ichikawa 1979] Ichikava, T., Kevan, L., Bowman, M. K., Dikanov, S. A., Tsvetkov, Y. D., *J. Chem. Phys.* (1979) **71** 1167-1174

[Ionita 2008] Ionita, P., Volkov, A., Jeschke, G., Chechik, V. *Anal. Chem.* (2008) **80** 95-106

[Jäger 2008] Jäger, H., Koch, A., Maus, V., Spiess, H. W., Jeschke, G. *submitted*

[Jeschke 1998] Jeschke, G., Spiess, H. W. *Chemical Physics Letters* (1998) **293** 9-18

[Jeschke 2000] Jeschke, G., Pannier, M., Godt, A., Spiess, H. W. *Chem. Phys. Lett.* (2000) **331** 243-252

[Jeschke 2002] Jeschke, G. *Macromol. Rapid Comm.* (2002) **23** 227-246

[Jeschke 2002b] Jeschke, G. *ChemPhysChem* (2002) **3** 927-932

[Jeschke 2004] Jeschke, G., Wegener, C., Nietschke, M., Jung, H., Steinhoff, H.-J. *Biophys. J.* (2004) **86** 2551-2557

[Jeschke 2005] Jeschke, G., Bender, A., Schweikardt, T., Panek, G., Decker, H., Paulsen, H. *J. Biol. Chem.* (2005) **280** 18623-18630

[Jeschke 2006] Jeschke, G. "DeerAnalysis2006" manual

[Jeschke 2006b] Jeschke, G., Zimmermann, H., Godt, A. *J. Magn. Reson.* (2006) **180** 137-146

[Jeschke 2007] Jeschke, G., Polyhach, Ye. *Phys. Chem. Chem. Phys.* (2007) **9** 1895-1910

[Jeschke 2007b] Jeschke, G. (2007) Instrumentation and Experimental Setup, in: M. A. Hemminga, L. J. Berliner (Eds.) ESR Spectroscopy in Membrane Biophysics, Springer, New York.

[Kirilina 2001] Kirilina E. P., Dzuba S. A., Maryasov A. G., Tsvetkov Yu. D. *Appl. Magn. Reson.* (2001) **21** 203-221

[Kirilina 2004] Kirilina E. P., Grigorjev I. A., Dzuba S. A. *J. Chem. Phys.* (2004) **121** 12465-12471

[Kleinschmidt 2002] Kleinschmidt, J. H., Tamm, L K., *J. Mol. Biol.* (2002) **324** 319-330

[Kokorin 1972] Kokorin, A. I., Zamaraev, K. I., Grigoryan, G. L., Ivanov, V. P., Rozantzev, E. G. *Biofizika* (1972) **17** (1) 34-40 (in russian)

[Kroll 1999] Kroll, C., PhD Thesis (1999) Humboldt- University Berlin

[Kühlbrandt 1994] Kühlbrandt, W., Wang, D. N., Fujiyoshi, Y. *Nature* (1994) **367** 614-621

[Kurad 2003] Kurad, D., Jeschke, G., Marsh, D. *Biophys. J.* (2003) **85** 1025-1033

[Lacapere 2007] Lacapere, J.J., Pebay-Peyroula, E., Neumann, J. M., Etchebest, C. *Trends Biochem. Sci.* (2007) **32** 259-270

[Leporini 2003] Leporini, D., Schädler, V., Wiesner, U., Spiess, H. W., Jeschke, G. *J. Chem. Phys.* (2003) **119** 11829-11846

[Lesk 2004] Lesk, A. M. (2004) Introduction to protein science, Oxford University Press

[Liu 2004] Liu, Z., Yan, H., Wang, K., Kuang, T., Zhang, J., Gui, L., An, X., Chang, W. *Nature* (2004) **428** 287-292

[Lindgren 1997] Lindgren, M., Eaton, G. R., Eaton, S. S., Jonsson, B. -H., Hammarström, P., Svensson, M., Carlsson, U. *J. Chem. Soc. Perk. T. 2* (1997) **12** 2549-2554

[Lindsey 1991] Lindsey, Y. S. *New J. Chem.* (1991) **15** 153-180

[Mao 2008] Mao, Q., Schleidt, S., Zimmermann, H., Jeschke, G. *Phys. Chem. Chem. Phys.* (2008) **10** 1156–1167

[Millhauser 1984] Millhauser, G. L., Freed, J. H. *J. Chem. Phys.* (1984) **81** 37-48

[Milov 1984] Milov, A. D., Ponomarev, A. B., Tsvetkov, Yu. D. *Chem. Phys. Lett.* (1984) **110** 67-42

[Morin 2006] Morin, B., Bourhis, J.-M., Belle, V., Woudstra, M., Carrière, F., Guigliarelli, B., Fournel, A., Longhi, S. *J. Phys. Chem. B* (2006) **110** 20596-20608

[Munishkina 2007] Munishkina, L. A., Fink, A. L. *BBA-Biomembranes* (2007) **1768** 1862-1885

[Nilsson 1997] Nilsson, A., Stys, D., Drakenberg, T., Spangfort, M. D., Forsén, S., Allen, J. F. *J. Biol. Chem.* (1997) **272** 18350-18357

[Noethig 2004] Noethig-Laslo, V., Cevc, P., Arčon, D., Šentjurc, M. *Appl. Magn. Reson.* (2004) **27** 303-309

[Owenius 2001] Owenius, R., Engström, M., Lindgren, M., Huber, M. *J. Phys. Chem. A* (2001) **105** 10967-10977

[Panek 2006] Panek, G., Schleidt, S., Mao, Q., Wolkenhauer, M., Spiess, H. W., Jeschke, G. *Macromolecules* (2006) **39** 2191-2200

[Pannier 2000] Pannier, M., Veit, S., Godt, A., Jeschke, G., Spiess, H. W. *J. Magn. Reson.* (2000) **142** 331-340

[Pannier 2000 b] Pannier, M. PhD Thesis (2000) University of Mainz

[Pannier 2001] Pannier, M., Schöps, M., Schädler, V., Wiesner, U., Jeschke, G., Spiess, H. W. *Macromolecules* (2001) **34** 5555-5560

[Paulsen 1990] Paulsen, H., Rümler, U., Rüdiger, W. (1990) *Planta* **181** 204-211

[Paulsen 1993] Paulsen, H., Finkenzeller, B., Kühlein, N. *Eur. J. Biochem.* (1993) **215** 809-816

[Plumley 1987] Plumley, F. G., Schmidt, G. W. *Proc. Natl. Acad. Sci.* (1987) **84** 146-150

[Polyhach 2007] Polyhach, Ye., Godt, A., Bauer, C., Jeschke, G. *J. Magn. Reson.* (2007) **185** 118-129

[Polyhach 2008] Polyhach, Ye., Volkov, A., Dockter, C., Paulsen, H., Jeschke, G. Unpublished results

[Ponti 1997] Ponti, A. *J. Magn. Reson.* (1997) **127** 87-104

[Popot 1990] Popot, J. –L., Engelman, D. M. *Biochemistry* (1990) **29** 4031-4037

[Reinsberg 2001] Reinsberg, D., Ottmann, K., Booth, P. J., Paulsen, H. *J. Mol. Biol.* (2001) **308** 59-67

[Rohrer 1996] Rohrer, M., Gast, P., Möbius, K., Prisner, T. F. *Chem. Phys. Lett.* (1996) **259** 523-530

[Roth 1989] Roth, M., Lewit-Bentley, A., Michel, H., Deisenhofer, J., Huber, R., Oesterhelt, D. *Nature* (1989) **340** 659-662

[Ruban 2007] Ruban, A. V., Berera, R., Ilioaia, C., van Stokkum, I. H. M., Kennis, J. T. M., Pascal, A. A., van Amerongen, H., Robert, B., Horton, P., van Grondele, R. *Nature* (2007) **450** 575-578

[Salnikov 2006] Salnikov, E.S., Erilov, D. A., Milov, A. D., Tsvetkov, Yu. D., Peggion, C., Formaggio, F., Toniolo, C., Raap, J., Dzuba, S. A. *Biophys. J.* (2006) **91** 1532-1540

[Santucci 2008] Santucci, R., Sinibaldi, F., Fiorucci, L. *Mini-Rev. Med. Chem.* (2008) **8** 57-62

[Savitsky 2004] Savitsky, A., Kühn, M., Duché, D, Möbius, K., Steinhoff, H. -J. *J. Phys. Chem. B* (2004) **108** 9541-9548

[Schiemann 2007] Schiemann, O., Prisner, T. F. *Q. Rev. Biophys.* (2007) **40** 1-53

[Schmidt-Rohr 1994] Schmidt-Rohr, K., Spiess H. W. Multidimensional solid- state NMR and polymers (1994) Academic Press, London

[Schweiger 2001] Schweiger, A., Jeschke, G. Principles of Pulse Electron Paramagnetic Resonance (2001) Oxford University Press

[Standfuss 2005] Standfuss, J., van Scheltinga Terwisscha, A. C., Lamborghini, M., Kühlbrandt, W. *EMBO J.* (2005) **24** 919-928

[Steed 2000] Steed, J. W., Atwood J. L. Supramolecular Chemistry (2000) John Wiley & Sons, Ltd.

[Steinhoff 2000] Steinhoff, H. J., Savitsky, A., Wegener, C., Pfeiffer, M., Plato, M., Möbius, K. *Biochim. Biophys. Acta* (2000) **1457** 253-262

[Szajdzinska 1984] Szajdzinska-Piętek, E., Maldonado, R., Kevan, L., Jones, R. R. M. *J. Am. Chem. Soc.* (1984) **106** 4675-4678

[Tamm 2005] Tamm, L. K., Hong, H. Protein Folding Handbook (2005) **2** 9981-1031 Wiley- VCH Verlag GmbH&Co.KGA, Weinheim

[Tamm 2006] Tamm, L. K., Liang, B. *Prog. Nucl. Mag. Res. Sp.* (2006) **48** 201-210

[Thóren 2004] Thóren P. E. G., Persson D., Ersbjörner E. K., Goksör M., Lincoln P., Nordén B. *Biochemistry* (2004) **43** 3471-3489

[Vamvouka 2008] Vamvouka, M., Cieslak, J., Van Eps, N., Hubbell, W., Gross, A. *Protein Sci.* (2008) **17** 506-517

[Voet 2004] Voet, D., Voet, J. G. Biochemistry (2004) 3rd edition, John Wiley & Sons, Inc.

[Wallin 1998] Wallin, E., Heijne, G. *Protein Sci.* (1998) **7** 1029-1038

[Wilhelm 1973] Wilhelm, E., Battino, R. *Chem. Rev.* (1973) **73** 1-9

[Wilhelm 1977] Wilhelm, E., Battino, R., Wilcock, J. R. *Chem. Rev.* (1977) **77** 219-262

[Windrem 1980] Windrem, D. A., Plachy, W. Z. *Biochim. Biophys. Acta* (1980) **600** 655-665

[Xu 2006] Xu, Q., Ellena, J. F., Kim, M., Cafiso, D. S. *Biochemistry* (2006) **45** 10847-10854

[Young 1996] Young, A. J., Frank, H. A. *Photochem. Photobiol.* (1996) **36** 3-15

[Zecevic 1998] Zecevic A., Eaton G. R., Eaton S. S., Lindgren M. *Mol. Phys.* (1998) **95** 1255-1263

[Zhang 2008] Zhang, X., Settembre, E., Xu, C., Dormitzer, P. R., Bellamy, R., Harrison, S. C., Grigorieff, N. *Proc. Natl. Acad. Sci.* (2008) **105** 1867-1872

[Zhou 1999] Zhou , Y., Bowler, B. E., Eaton G. R., Eaton S. S. *J. Magn. Reson.* (1999) **139** 165-174