Chapter III

Accurate long-range distance measurements in a doubly spin-labeled protein by a four-pulse, double electron-electron resonance method
Abstract

Distance determination in disordered systems by a four-pulse double electron-electron resonance method (DEER or PELDOR) is becoming increasingly popular because long distances (several nm) and their distributions can be measured. From the distance distributions eventual heterogeneities and dynamics can be deduced. To make full use of the method, typical distance distributions for structurally well-defined systems are needed. Here, the structurally well-characterized protein azurin is investigated by attaching two 1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl)methanethiosulfonate spin labels (MTSL) by site-directed mutagenesis. Mutations at the surface sites of the protein Q12, K27, and N42 are combined in the double mutants Q12C/K27C and K27C/N42C. A distance of 4.3 nm is found for Q12C/K27C and 4.6 nm for K27C/N42C. For Q12C/K27C the width of the distribution (0.24 nm) is smaller than for the K27C/N42C mutant (0.36 nm). The shapes of the distributions are close to Gaussian. These distance distributions agree well with those derived from a model to determine the maximally accessible conformational space of the spin-label linker. Additionally, the expected distribution for the shorter distance variant Q12C/N42C was modeled. The width is larger than the calculated one for Q12C/K27C by 21%, revealing the effect of the different orientation and shorter distance. The widths and the shapes of the distributions are suited as a reference for two unperturbed MTSL labels at structurally well-defined sites.

The results in this chapter have been published in:

Introduction

Distance determination by EPR techniques became gradually more important to solve structural problems that are not easily accessible by standard structural techniques. Usually, two spin labels are introduced such that their distance reflects the structural property of interest. Double electron-electron spin resonance (DEER or PELDOR\textsuperscript{1-4}), a pulsed EPR method by which distances in the nm regime can be determined, is more and more used in that context, as reviewed recently\textsuperscript{5,6}. Besides the long distance range accessible by the technique, the possibility to determine distance distributions is increasingly appreciated. Such distributions contain information about the uniformity of the molecular ensemble and can be used as a measure for flexibility, parameters that could cast light on the dynamics of macromolecules. Such dynamical aspects are of particular interest in a biological context. Here, DEER and related methods are spreading rapidly, also because methods to incorporate labels at the desired positions in many biological systems are now available.

Model systems in which the approach has been tested in a biological context are scarce. In particular, data of well characterized systems are needed as a reference for flexible, dynamic systems. We present distance measurements between two spin labels in the small protein azurin. The structure of this protein is well defined and known from X-ray crystallography\textsuperscript{7}. Spin labels were introduced by spin label mutagenesis, as pioneered by Hubbell \textit{et al.}\textsuperscript{8} Two double mutants were investigated. The first one, in which Q12 and K27 were each replaced by a cysteine (Cys) (Q12C/K27C), and the second one, in which K27 and N42 were replaced by Cys (K27C/N42C). The singly labeled mutant protein K27C was used as a reference. All residues concern surface residues, i.e., residues that are expected to interfere least with the structure of the protein.

Two steps have to be accomplished. From experimental data, the distance (distribution) is determined, then, the distances are related to the protein structure. For the latter step, the length of the spin-label linker joining the spin label with the protein backbone is important\textsuperscript{9-11}.

Multiple conformations and dynamics can best be identified through the width of the distance distribution, but the uncertainty of the spin-label linker conformation makes it difficult to discriminate between the effect of the linker and the sought-for width of the distribution of the attachment points. For structurally well characterized proteins the former effect should dominate and therefore can be used to calibrate widths observed in unknown systems.

In the present study we show that distances in the 4 nm region can be measured with high accuracy. We propose a simple model to account for the conformation of the spin label linker (represented in Figure 3.1) and show that it yields a good agreement with the data.
Furthermore, we use the model to calculate the distance distribution of the spin-label combination Q12C/K42C to assess the effect of the location of the spin-labels, which are adjacent, i.e., on the same hemisphere of the protein, (Q12C/K42C) versus opposing, i.e., at opposite ends of the protein (Q12C/K27C and K27C/N42C) (Figure 3.1).

The range of the distance distributions is between 0.25 and 0.45 nm, where the latter value includes the effect of a wider distribution in the case of (Q12C/K27C). The width of these distance distributions is at the lower end of several of the distance distributions determined previously\textsuperscript{12-17}, suggesting that in the latter cases the flexibility of the protein plays an important role.

**Materials and methods**

**Mutants and spin labeling**

The mutation N42C was introduced in the gene encoding K27C azurin (plasmid pChH02, kindly provided by Prof. G. W. Canters, Leiden). First, part of the gene was amplified by PCR using an oligonucleotide containing the N42C mutation (CCTGCGAAGTGCGTCATGGGTCACAACTGGG) and an oligonucleotide binding downstream of the gene (CATGCACGGATCGTCGCGC). The resulting megaprimer was used in a second PCR reaction in combination with an oligonucleotide encoding a region upstream of the SalI restriction site at residue 22 (ACGACCAGATGCTGATCTAC). The product was digested with *Kpn*I and *Sal*I and inserted into the pChH02, digested with the same enzymes, yielding pMGF02.

The Q12C/K27C double mutant was obtained by restriction of a fragment of the gene with only the Q12 mutation (pAZQ12C, kindly provided by Prof. G. W. Canters, Leiden) and insertion into the K27C azurin gene, taking advantage of the *Sal*I site, located between the mutation sites. This resulted in pMGF01. The mutations were confirmed by sequence analysis. Protein expression and purification was carried out under reducing conditions, as described\textsuperscript{18}. The naturally present

Figure 3.1 A surface model of azurin (PDB entry 4AZU) is shown with the MTSL oxygens of all sterically allowed conformations of the spin labels. Red: spin label attached to C12; green: to C27; yellow: to C42. Insert: structure of MTSL, including cysteine sulphur and C\textsubscript{\beta} - atom.
metal ion Cu(II) was replaced by Zn(II) to avoid interference from spin interaction or redox chemistry\textsuperscript{19}. Spin labeling of the mutants with the spin label 1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl)methanethiosulfonate (MTSL) was performed as described in ref. 20.

**Sample preparation**

The protein concentrations for the samples of the double mutants of azurin (0.175 mM, i.e. 0.35 mM in spin label, volume 150 µl, including 30% v/v glycerol) and the single mutant reference (K27C, 0.35 mM, volume 150 µl, including 30% v/v glycerol) were chosen according to the maximum concentrations suggested in reference\textsuperscript{4} for the distance range of interest. Wilmad suprasil quartz tubes (inside diameter 3 mm, outside diameter 4 mm) were used.

**DEER experiments**

The DEER experiments were performed on a Bruker Elexsys 680 spectrometer (Bruker Biospin GmbH Rheinstetten, Germany) with the modifications described in reference\textsuperscript{20}. Measurement conditions were analogous to those in reference\textsuperscript{21}. The DEER experiment was performed at 40 K using a dielectric ring resonator and a helium flow system by Oxford, model CF 935. The four-pulse DEER sequence\textsuperscript{2} \( p_1-t_1-p_2-t_2-p_3 \) with a pump pulse inserted after \( p_2 \) was employed. The pulse power of the observer pulses \( p_1, p_2 \) and \( p_3 \) (lengths: 32 ns) was adjusted to obtain a \( \pi/2 \) pulse for \( p_1 \) and \( \pi \)-pulses for \( p_2 \) and \( p_3 \). The pump-pulse (length: 36 ns) was adjusted for maximum inversion of the echo by varying the pump power. Delay times were \( t_1 = 200 \) ns, \( t_2 = 2000 \) ns and the time \( T \), at which the pump pulse was inserted after \( p_2 \) was varied. The observer field was set to the low-field edge of the spin-label EPR spectrum and the pump frequency adjusted to coincide with the maximum of the electron-spin-echo detected EPR spectrum (frequency separation: \( \Delta \nu = 65 \) MHz). The total measurement time for the DEER curves was 15 h.

**Analysis of DEER results**

For the analysis of the distance distributions, first, Gaussian distance distributions were simulated using the program DEERfit\textsuperscript{2,22,23}. The parameters were adjusted manually; errors for single Gaussian distributions (Q12C/K27C: \( \pm 0.03 \) nm; K27C/N42C: \( \pm 0.02 \) nm) were estimated by determining the magnitude of the changes in the parameters that resulted in simulated curves outside the noise limit of the experiment. For the fitting procedure with arbitrary distance distributions, the methods provided in the program DeerAnalysis 2006\textsuperscript{24} were used. The background was calculated from the DEER curves with different starting times (Q12C/K27C: > 980 ns; K27C/N42C: > 1300 ns). Some of the characteristics in the data of the Q12C/K27C mutant also point to the possibility that the degree of double labeling is smaller in this mutant than in the
K27C/N42C mutant. Therefore, for the data of Q12C/K27C, also the experimental background from the K27C sample was tried but did not result in significant changes of the fits. Tikhonov regularization was employed and the first and second moments of the distance distributions are given.

**Determining the conformational space of the spin label**

Models of azurin with MTSL on Cys27 and either C12 or C42 were constructed using PDB entry 4AZU. Then, using XPLOR-NIH, the five torsion angles between the Cα of the Cys residue and the ring of the spin label were rotated systematically in steps of 30°. All orientations with steric clashes were discarded. Then, the distances from the MTSL nitrogens of all orientations of Cys27 to the nitrogens of all the MTSL orientations of C12 or C42 were determined and binned (binning classes: 0.1 nm).

**Results**

The intensity of the echo as a function of the time T (DEER time trace) for all three samples is shown in Figure 3.2. The modulations observed for the two double mutants reflect the spin-spin interaction. They are absent in the single mutant that contains only one spin label.

The difference in the modulation period of the two mutants is clearly visible, indicating a smaller dipolar interaction and thus a longer distance for the K27C/N42C mutant. For the analysis, the programs of Jeschke et al. were used.

In Figure 3.3, panels a (Q12C/K27C) and b (K27C/N42C) the fits obtained are shown, superimposed on the baseline corrected time traces from Figure 3.2. The corresponding distance distributions are displayed in panels c (Q12C/K27C) and d (K27C/N42C).

For Q12C/K27C, the result of a single Gaussian fit is shown as trace 1 in Figure 3.3a. The deviation from the measured curve for the second period of the modulation around 2.5 µs is significant, suggesting that the distribution deviates from a single Gaussian. The agreement is better when two Gaussian functions are used, an example for which is shown in Figure 3.3a, trace 2 (distance distribution: Figure 3.3c, trace 2). Different combinations
of the two Gaussians give simulations of similar quality. The larger fraction peak is always centered at the value given in Table 3.1.

This peak has a position that is similar to that of the single Gaussian fit in trace 1 but it has a smaller width. Tikhonov regularization yields a fit (Figure 3.3a, trace 3) that is slightly better in quality than trace 2. The distribution (Figure 3.3c, trace 3) has a main peak that is similar to the main peak in trace 2 and an additional peak at 2.4 nm of low intensity. In Table 3.1, the distance parameters corresponding to the main peak of the Tikhonov regularization are also listed.

For K27C/N42C (Figure 3.3b) a single Gaussian distribution (trace 1 in Figure 3.3b and 3.3d) yields the proper period of the modulation albeit with larger damping than the experimental curve, suggesting that the width of the Gaussian distribution is too large. Tikhonov regularization results in the fits (Figure 3.3b, trace 2) and the distance distribution shown (Figure 3.3d, trace 2). The agreement with the experimental data is slightly better than in the case of the single Gaussian (Figure 3.3b, trace 1). As seen from the parameters corresponding to these two distributions (Table 3.1) the peak resulting from Tikhonov regularization is indeed narrower than the single Gaussian.

Figure 3.3 DEER time traces (background corrected) and fits with different methods for Q12C/K27C (a) and K27C/N42C (b) (curves are shifted by constant amounts to avoid overlap); distance distributions for Q12C/K27C (c) and K27C/N42C (d). Models shown: (a, and c, trace 1): single Gaussian fit, parameters: 4.22 nm (centre), 0.35 nm (width), trace 2: two-Gaussian simulation, parameters see Table 3.1, trace 3: Tikhonov regularization, parameters see Table 3.1, trace 4: Distance distribution from model (see text). Panels b and d give the corresponding distance distributions. The distance distribution derived from the structural model is shown for Q12C/K27C in c, trace 4 and for K27C/N42C in d, trace 3.
Overall, a smaller distance and a smaller width of the distribution are found for the Q12C/K27C mutant than for the K27C/N42C mutant. There is some indication for a deviation of the distribution from Gaussian or a bimodal distribution for Q12C/K27C.

Table 3.1 Distances and parameters of distance distributions from the X-ray structure of azurin, DEER experiments, and the model

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Structure</th>
<th>Model</th>
<th>Gaussian</th>
<th>Tikhonov</th>
<th>$\Delta r(C_\beta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C_\alpha$</td>
<td>$C_\beta$</td>
<td>Distance (nm)</td>
<td>Width (nm)</td>
<td>Distance (nm)</td>
</tr>
<tr>
<td>Q12C/K27C</td>
<td>3.38</td>
<td>3.50</td>
<td>4.58</td>
<td>0.42</td>
<td>4.27</td>
</tr>
<tr>
<td>K27C/N42C</td>
<td>3.51</td>
<td>3.69</td>
<td>4.72</td>
<td>0.35</td>
<td>4.62</td>
</tr>
<tr>
<td>Q12C/N42C</td>
<td>1.44</td>
<td>1.70</td>
<td>2.80$^b$</td>
<td>0.53</td>
<td>nd$^i$</td>
</tr>
</tbody>
</table>

$^a$ $C_\alpha$: $C_\alpha(1)$ - $C_\alpha(2)$
$^b$ $C_\beta$: $C_\beta(1)$ - $C_\beta(2)$
$^c$ half width at half-height
$^d$ standard deviation
$^e$ (distance from Tikhonov) – ($C_\beta(1)$-$C_\beta(2)$)
$^f$ additional Gaussian, 31%: 3.8 nm centre, 0.4 nm width
$^g$ parameters of the central peak of the distribution
$^h$ maximum of distribution
$^i$ nd: not determined; na: not applicable

From the crystal structure\(^7\), the distances between the $C_\beta$ atoms of the two residues that are mutated were obtained. The measured distances are larger than the distances between the $C_\beta$ atoms ($C_\beta(1)$-$C_\beta(2)$) listed in Table 3.1.

A simple model was used to determine the distribution of spin-label distances. By systematic rotation around the five torsion angles that connect the nitroxyl radical with the protein backbone, an ensemble of sterically allowed conformations was obtained for each spin label. The model results in the distributions shown in Figure 3.3c, trace 4 and Figure 3.3d, trace 3.

The combination Q12C/K27C is also analyzed, and it is shown along with the other calculated distributions in Figure 3.4. These distributions are slightly skewed with a foot at shorter distances, but the deviation from a symmetrical Gaussian distribution is small. The distributions overlap the experimental ones and differences are discussed in more detail below.

**Figure 3.4** Distance distribution calculated from model. The distributions for Q12C/K27C and K27C/N42C are identical to those shown in Figure 3.3 (c) and (d).
Discussion

The distance distributions for two double mutants of Zn-azurin were obtained. In the following we will first discuss the analysis of the data and compare them to a model describing the spin-label linker conformations. The measured distances can then be related to the protein structure. Finally, we will compare the distance distributions to those obtained for other proteins.

The DEER data were analyzed using the currently available methods. For Q12C/K27C the parameters of the Gaussian distribution and the Tikhonov regularization agree within experimental uncertainties, for K27C/N42C the single Gaussian has a distance comparable to the Tikhonov regularization, but the smaller width of the latter yields a better agreement with the data. Thus, although the distance distributions from these approaches differ slightly, the overall deviation from Gaussians is small. The distributions are narrow and do not indicate significant bimodal character. The distance distribution of Q12C/K27C has a smaller distance and a smaller width than that of K27C/N42C.

As pointed out before\textsuperscript{9-11}, interpretation of the measured distances requires structural information about the spin-label. From the X-ray structure, only the position of the C\textsubscript{α} and C\textsubscript{β} atoms of the aminoacid-residue replaced by the cysteine is known, but the distance obtained from the experiment corresponds to the center of spin density on the two nitroxide groups of the spin label. The distance from the nitroxide group of the spin label to the C\textsubscript{β} atom is between 0.5 and 0.6 nm for extended conformations of the spin-label linker\textsuperscript{10}. Several approaches have been suggested to account for the conformation of the spin-label linker.

A straightforward approach\textsuperscript{10} is to analyze the difference in the distances between the C\textsubscript{α} and the C\textsubscript{β} atoms of the two residues (residues (1) and (2)) that are mutated. If the C\textsubscript{α}(1)-C\textsubscript{α}(2) distance is smaller than the C\textsubscript{β}(1)-C\textsubscript{β}(2) distance, it is assumed that the spin labels are pointing away from each other, suggesting a distance larger than the C\textsubscript{β}(1)-C\textsubscript{β}(2) distance, if the opposite is true the distance should be shorter. For both mutants, the C\textsubscript{α}(1)-C\textsubscript{α}(2) distance is smaller than the C\textsubscript{β}(1)-C\textsubscript{β}(2) distance (see Table 3.1), suggesting that the spin labels are pointing away from each other. In agreement with this, the measured distances are larger than the C\textsubscript{β}(1)-C\textsubscript{β}(2) distances (see values \(\Delta r(C_{\beta})\) in Table 3.1). Assuming a spin-label-linker length of 0.5 nm per spin label, the distances observed experimentally, i.e. \(\Delta r(C_{\beta}) < 1\) nm, are in the range of what the X-ray structure predicts.

To model the conformation of the linker directly, a model is proposed that is described in ‘Experimental section’. It aims to explore the maximally accessible conformational space for the spin label by generating all possible conformations of the linker chain and only exclude those that have sterical clashes. The result may not represent the real distribution of distances, because any
molecular interactions that may result in preferred orientations of the MTSL are neglected (conformations, see Figure 3.1). For Q12C/K27C, the distance distribution derived from the model (Figure 3.3c) overlaps the distribution derived from the DEER experiment, but the model predicts a significant fraction at a larger distance that is not observed in the experiment. The trend towards model distributions that are broader than the experimental is also discussed in a recent publication. For K27C/N42C (Figure 3.3d), the distribution from the model is closer in width and centre position to the experimental distribution than in the case of Q12C/K27C. It could thus be argued that the experimentally observed shift to shorter distances for Q12C/K27C represents conformations of the spin-labeled residue C12 that are determined by favorable molecular interactions. The good overall agreement of the model with the experimental results made it attractive to address the remaining combination of spin label positions Q12C/K42C in silico, to investigate the effect of a different relative orientation and a smaller distance of the spin labels. The comparison of the three distributions from the model (Figure 3.4) shows that the width of the distribution of Q12C/K42C is larger, which we attribute to the location of the two residues, which are adjacent rather than at opposing hemispheres of the protein, and the shorter distance between the nitroxides.

Previously, the model had been used in the context of distance determination in NMR, in which the paramagnetic relaxation caused by an MTSL spin label is used. Because of the $r^{-6}$ weighting of the distances in the NMR data, for the interpretation, particular attention has to be paid to the conformational space occupied by the spin label. The present data show that for the three positions of the spin label in the mutants a good agreement between the model and the experimental results is achieved as shown in Figure 3.3 and detailed in the Results section. This presents independent evidence for the validity of the model.

Other approaches to obtain the spin-label conformation are modelling the conformational degree of freedom of the spin label as free rotation about two bonds, Molecular Dynamics (MD) simulations, Monte Carlo energy minimization, and rotamer libraries. A comparison of different approaches has been discussed previously. Recently, an approach has been presented that uses rotamer libraries and rigid-body refinement to determine the structure of a dimer by direct calculation of the DEER time traces. For de novo-structure determination the method has the advantage that it avoids the step to derive distance distributions from the DEER time traces. For the time being, we find the model employed in the present study a good compromise that allows us to include the local structure of the protein without having to pay the price of a full-fledged MD simulation.

The spin labels in the Zn-azurin mutants investigated here are on the surface of the protein and therefore not conformationally restrained. The protein azurin has a well-defined structure and,
as a consequence, should not contribute to the width of the distance distribution. Therefore, the widths of the distributions observed here, i.e. 0.24 nm and 0.36 nm, reflect exclusively the flexibility of the spin-label linkers. For a given range of conformations of the individual spin labels, the width of the distribution will, however, also depend on the relative location of the spin labels on the protein. Both the absolute distance but also the relative orientation of the protein surfaces can play a role. The distribution calculated for the pair Q12C/N42C gives an indication because, as seen in Figure 3.1, here two spin labels are in the same hemisphere of the protein, rather than at opposing ends of the protein. The calculated width of the distribution is 21 % larger than that of Q12C/K27C, which, in terms of the largest measured width (K27C/N42C, Tikhonov regularization), would correspond to 0.44 nm.

Several systems reported in the literature have comparable widths. A width of 0.3 nm was found for the interaction of two MTSL labels in ubiquitin\textsuperscript{12}, another structurally well characterized protein. In contrast, the buried spin-label sites in T4 lysozyme revealed broader, and in some cases bimodal distance distributions by continuous-wave EPR using a spectral-deconvolution method\textsuperscript{13}. In one protein-oligomer system, a distribution with the parameters (6.15 ± 0.14) nm\textsuperscript{14}, was observed, however, in that case, the width may not be a reliable parameter given that the total length of the trace is shorter than the modulation period. Otherwise distance distributions in proteins tend to be broader. For example, in a membrane protein (LHCII), distance distributions with widths between 0.5-2 nm (width at half height) were found\textsuperscript{15}. For the interaction between protein-monomers in an oligomeric system, depending on the conditions, distributions as narrow as 0.2 ± 0.1 nm, around 0.4 nm, or broad distance distributions ranging from 1.5 nm to 6 nm were found\textsuperscript{17}. Almost structureless distributions with decreasing intensity between 2 and 4 nm were observed between spin labels attached to a single membrane protein\textsuperscript{16}. Similarly, also DNA/RNA systems were investigated by DEER techniques, see for example reference \textsuperscript{33}, but chemically different spin labels and linkers used for these systems make the comparison difficult.

Our study indicates that broad distributions (> 0.45 nm) are not likely to be caused by the spin-label-linker conformations alone. This is not self evident because the length of the spin-label linker of 0.5 nm could, in principle, cause widths as large as 2 nm for the distance between two spin labels. Taking the widths measured for the relatively unrestricted spin labels at the surface of the protein of Q12C/K27C and K27C/N42C as a reference, structural heterogeneities or dynamics in the proteins investigated are the most likely source for distributions with widths larger than 0.45 nm.
Summary and conclusions

We measured the distance between two spin labels located at the surface of a well characterized protein, azurin. The system serves as a model for distance determination in a biological context. We applied different methods of analysis for the data and found relatively narrow distance distributions that are well described by Gaussians.

A model to determine the maximum conformational space of the spin-label linker yielded a good agreement with the experimental distance distribution, suggesting that the model gives a realistic picture of the linker conformation. We suggest that for proteins labeled with two MTSL spin labels that are not conformationally restricted, widths between 0.25 and 0.45 nm should be typical, indicating that larger widths are caused by flexibility or dynamics of the protein investigated.

The problem of absolute structure determination remains difficult as long as the label is attached by linkers of the given length, emphasizing the usefulness of conformationally rigid spin labels, such as used in synthetic peptides, proteins and DNA or RNA systems. The investigation of conformational changes or of interactions of protein subunits under different conditions suffers much less from the absolute distance uncertainty and has therefore been successfully addressed by the MTSL approach, as has been demonstrated in several studies.

Compared to Förster resonance energy transfer (FRET), a method that is well established for distance determination in biological systems, the spin label is smaller than the conventional fluorescent labels, and the spin density is localized on two atoms enabling structural determination to high precision. Also, there is no need for differential labeling, as required to introduce the FRET donor and acceptor pairs, and, in spin-label EPR, the same paramagnetic labels give access to the distance range from several Å to several nm.

The DEER method extends spin-label EPR to longer distances and thus has aided to turn spin-label EPR into a tool that complements the existing structural methods as NMR and FRET.
Reference List


2. Jeschke,G. Distance measurements in the nanometer range by pulse EPR. *Chemphyschem* 3, 927-932 (2002).


