Published in final edited form as:

Methods Enzymol. 2015 ; 563: 531–567. doi:10.1016/bs.mie.2015.07.031.

A Straightforward Approach to the Analysis of Double Electron–Electron Resonance Data

Richard A. Stein, Albert H. Beth, and Eric J. Hustedt

Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville TN 37232

Abstract

Double electron–electron resonance (DEER) is now widely utilized to measure distance distributions in the 20 to 70 Å range. DEER is frequently applied to biological systems that have multiple conformational states leading to complex distance distributions. These complex distributions raise issues regarding the best approach to analyze DEER data. A widely used method utilizes *a priori* background correction followed by Tikhonov regularization. Unfortunately, the underlying assumptions of this approach can impact the analysis. In this chapter, a method of analyzing DEER data is presented that is ideally suited to obtaining these complex distance distributions. The approach allows the fitting of raw experimental data without *a priori* background correction as well as the rigorous determination of uncertainties for all fitting parameters. This same methodological approach can be used for the simultaneous or global analysis of multiple DEER data sets. This global analysis uses varying ratios of a common set of components allowing direct correlation of distance components with functionally relevant conformational and biochemical states. Examples are given throughout to highlight this robust fitting approach.

Introduction

Since its introduction over 25 years ago (Altenbach, Flitsch, Khorana & Hubbell, 1989), site-directed spin labeling (SDSL) has developed into an important method for investigating structure and dynamics with applications to both soluble and membrane bound proteins as well as to nucleic acids. Numerous reviews have discussed the power of SDSL and the diversity of protein structures, protein-protein interactions, and dynamic structural equilibria to which it has been applied (e.g. Hubbell, Mchaourab, Altenbach & Lietzow, 1996; Hubbell, Gross, Langen & Lietzow, 1998; Hubbell, Cafiso & Altenbach, 2000; Columbus & Hubbell, 2002; Cafiso, 2002; Hubbell, Altenbach, Hubbell & Khorana, 2003; Klug & Feix, 2005; Fanucci & Cafiso, 2006; Klug & Feix, 2008; Zhang, Cekan, Sigurdsson & Qin, 2009; Phuong & Qin, 2012). A great deal of site-specific information including spin-label side-chain dynamics and solvent accessibility can be obtained from conventional continuous wave EPR measurements on singly-labeled molecules. In addition, it is also possible to obtain inter-probe distance and distance distribution information in the 5 to 20 Å range from continuous wave EPR measurements (Rabenstein & Shin, 1995; Hustedt & Beth, 1999; Hustedt & Beth, 2000; Persson et al., 2001). With the development of pulsed EPR methods such as double electron–electron resonance (DEER, also known as PELDOR) and double quantum coherence (DQC) the maximum distance measurable has been extended

considerably. The availability of commercially available, turnkey instrumentation capable of pulsed measurements has led to a rapid rise in the use of DEER during the 21st century.

Numerous reviews have focused on the methodology of DEER and its applications in structural biology (e.g. Jeschke, 2002; Fajer, Brown & Song, 2007; Bhatnagar, Freed & Crane, 2007; Borbat & Freed, 2007; Reginsson & Schiemann, 2011b; Reginsson & Schiemann, 2011a; Jeschke, 2012). This chapter will focus on the analysis of DEER data. We acknowledge that our perspective on the fitting of DEER data is somewhat contrary to what has become standard practice. Our goal here is to challenge the conventional wisdom, to offer an alternative approach to the one that is commonly employed, to stimulate further discussion, and to motivate further development of methodology for the analysis of DEER data.

DEER

Electron–electron dipolar coupling will produce significant broadening of continuous wave EPR lineshapes at distances up to 20 Å at X-band (for reviews see Hustedt & Beth, 1999; Hustedt & Beth, 2000; Persson et al., 2001) and even longer distances at lower frequencies (Kittell, Hustedt & Hyde, 2012). Methods have been developed for extracting an interspin distance, or distribution of distances, from these spectra (Rabenstein & Shin, 1995; Hustedt, Smirnov, Laub, Cobb & Beth, 1997; Hustedt, Stein, Sethaphong, Brandon, Zhou & Desensi, 2006).

At distances longer than 20 Å, pulsed techniques can be used to separate the weak electron–electron dipolar coupling from the electron Zeeman and hyperfine interactions. Figure 1A shows a standard DEER 4-pulse sequence. The three pulses at the observer frequency, ν_1 , selectively excite a population of spins, designated A spins, producing an echo at $2(\tau_1 + \tau_2)$ due to the refocusing of the Zeeman and hyperfine inhomogeneities. Inhomogeneities due to electron–electron dipolar coupling will also be refocused except for those interactions with a second population of spins, designated B spins, excited by the pump pulse at ν_2 . The extent to which the dipolar interaction between A and B spins is refocused varies with the time, t , of the pump pulse leading to a modulation of a fraction of the echo amplitude.

DEER can be performed on a variety of different systems including single proteins, dimers and higher homo-oligomers of proteins, and a variety of hetero-oligomeric species. Typically DEER is used to determine the distance distribution between the pair of spins within such an object or the full set of distances for an object with more than two spins. The spins within a given object also experience dipolar couplings with spins on all of the other objects within the sample giving rise to a background signal which must be accounted for in any analysis. This chapter will focus on the analysis of DEER data from doubly-labeled objects (for further discussion of multi-spin interactions see Jeschke, Sajid, Schulte & Godt, 2009; Junk, Spiess & Hinderberger, 2011; Dalmas, Hyde, Hulse & Perozo, 2012; von Hagens, Polyhach, Sajid, Godt & Jeschke, 2013; Giannoulis, Ward, Branigan, Naismith & Bode, 2013; Edwards et al., 2014).

The great utility of DEER spectroscopy comes from the fact that the result is not a single number corresponding to an average interspin distance but is instead the full distance distribution defining the probability of finding a pair of spins at a given interspin distance. DEER is highly sensitive to both the average distance and the width of the distribution for unimodal (single component) distributions (see Figure 2A). Also, DEER data can show strong sensitivity to multimodal distributions depending on the centers and widths of the various components (see Figure 2B). However, it must be remembered that the inverse problem of determining the correct distance distribution from DEER time-domain data is an ill-posed problem (Chiang, Borbat & Freed, 2005; Jeschke, Koch, Jonas & Godt, 2002). Thus, many different distance distributions may give reasonable agreement with a given data set and the solution for any particular data may be highly dependent on the details of the random errors in that data.

During the initial development of DEER methodology, data from experiments on organic biradicals was analyzed by Fourier transformation of the data collected in the time domain. For samples with a fixed interspin separation, the distance could be calculated from the frequency separation from the two singularities in the Pake pattern (Pake, 1948; see also Borbat & Freed, 2007). This approach required that the time-domain data be corrected to remove the large background signal due to inter-object spin interactions. As DEER was applied to samples with broader, more complex distributions, techniques were developed to fit DEER data in the time domain assuming the distance distribution could be modeled as a sum of Gaussian functions (Pannier, Schädler, Schöps, Wiesner, Jeschke & Spiess, 2000). Such an approach can be performed with *a priori* background correction or by including an exponential factor in the fitting function to account for the background. Finally, approaches using Tikhonov regularization have been developed that do not rely on a specific model function for the distance distribution but instead incorporate an additional term in the least-squares minimization that imposes a variable degree of smoothness on the distribution (Bowman, Maryasov, Kim & DeRose, 2004; Chiang et al., 2005; Jeschke et al., 2006; Jeschke, Panek, Godt, Bender & Paulsen, 2004). These approaches require the use of *a priori* background correction.

The preceding paragraph emphasizes the two fundamental choices that must be made by anyone wishing to analyze DEER data, whether to apply an *a priori* background correction and whether to use a model-based (i.e. sum of Gaussians) or model-free (i.e. Tikhonov regularization) approach to define the intra-object distance distribution. Both the model-based and Tikhonov regularization approaches address the ill-posed nature of the problem. With the widespread use of the **DeerAnalysis** software developed by Jeschke and colleagues (Jeschke et al., 2006), *a priori* background correction followed by Tikhonov regularization has become widely used in the field. While the model-free approach offered by Tikhonov regularization is appealing, it is important to realize that the method is not assumption-free and not necessarily distortion-free. The choice of the value of the regularization parameter, α , is subjective requiring a compromise between distortions of the distance distribution by noise artifacts for small α versus over-broadening of the distribution for large α . In particular, Tikhonov regularization methods can produce significant distortions when the true distance distribution is a mix of both narrow and broad components (Jeschke et al.,

2006). Figure 3 shows the analysis of simulated DEER data for such a distance distribution (shaded area in Figure 3C) using Tikhonov regularization in **DeerAnalysis**. Depending on the choice of α , either the data is poorly fit and the representation of the narrow component overly broad ($\alpha = 100$; dashed blue lines) or the data is well fit and the representation of the broad component artificially spiked ($\alpha = 100$; solid red lines). Tikhonov regularization also requires the use of an *a priori* background correction using a particular model for the background signal. Depending on the quality of the data and the nature of the intra-object distance distribution, it can be difficult to properly isolate and fit the background factor from the total signal. Improper background correction can lead to distortions of the fit and the resulting distance distribution as shown below in Figure 11 (see also Brandon, Beth & Hustedt, 2012). **DeerAnalysis** does include a validation tool to examine the effects of the *a priori* selection of a background signal on the resulting distance distribution. Two approaches are available. The first allows variation of the spin concentration, modulation depth, and dimensionality (d) over user-specified ranges but does not fully take into account correlations between these parameters. The second allows variation of the time range of the data used for background fitting. Depending on the time range of the data and the range of distances in the distribution this approach may never access the true background component. See the **DeerAnalysis** User Manual (Jeschke, <http://www.epr.ethz.ch/software/index>) for additional discussion of the validation tool.

Based on our experience with a variety of approaches, we strongly believe that fitting DEER data without *a priori* background correction and using a sum of Gaussians to model the distance distribution offers important advantages that are often overlooked. These advantages include:

1. The information available from modeling the distance distribution as a sum of Gaussians can be directly used to build structural and dynamic models. Specific Gaussians can be assigned to specific conformational states within a dynamic equilibrium or different spin pairs within a multi-labeled complex. In fact, in some cases distance distributions obtained by analyzing DEER data using Tikhonov regularization have been subsequently fit to a sum of Gaussians (e.g. Georgieva, Borbat, Ginter, Freed & Boudker, 2013; Kear, Blackburn, Veloro, Dunn & Fanucci, 2009; Sircar et al., 2015) in order to develop structural models from the DEER data. As shown in Figure 3, distance distributions obtained by Tikhonov regularization may not represent the true number of components in the distance distribution.
2. Using a model-based approach, rigorous error analysis can be performed on all fitting parameters to give their uncertainties. These uncertainties can be used, for example, to determine whether a change in the value of a given parameter between different samples is statistically

significant. The error analysis is performed in a way to take into account all correlations between parameters. See Figures 6, 8, 9, and 12 below.

3. By not using *a priori* background correction, data can be analyzed in situations where the combination of data collected over a short time span and the presence of long distance components in the distribution make *a priori* background correction challenging. Error analysis can be used to directly determine how uncertainties in the background contribution influence the distance distribution taking into account all correlations between the various parameters. See Figures 11 and 12 below.
4. By not using *a priori* background correction, data can be fit in situations where the true background signal is neither an exponential nor a stretched exponential using a model that takes into account excluded volume effects. See Figure 10 below.
5. Using a model-based approach, global (simultaneous) analysis can be performed on multiple data sets collected for different samples or different conditions. Global analysis allows a *direct* test of specific models for the combined data and allows for rigorous error analysis of the model parameters. See Figures 13 through 17 below.

Building on the successful work of Jeschke and colleagues (Jeschke et al., 2006) and Fajer and colleagues (Sen, Logan & Fajer, 2007), we (RAS and EJH) have independently developed software that is optimized to fit DEER data using this alternative approach (Brandon et al., 2012; Masureel et al., 2014; Mishra, Verhalen, Stein, Wen, Tajkhorshid & McHaourab, 2014). While there are slight differences in our implementations of this approach, for clarity the rest of this chapter will focus on the **DD and GLADD** software. Table 1 gives a list of these and additional software for analyzing DEER data that has been made available by various research groups.

Theory

Details of the fitting approach have been previously published (Brandon et al., 2012). Briefly, data are fit to Equation 1,

$$F(t)=O(t) \times B(t) \times E(t) \quad [1]$$

where $O(t)$ is the DEER signal for all of the spins within a given object, $B(t)$ is a background signal calculated for a given arrangement of spins or objects within the sample, and $E(t)$ is

an exponential or stretched exponential. Either $B(t)$ or $E(t)$ is used to fit the background component with the other factor fixed to 1.

For an object with a single pair of spins,

$$O(t) = 1 - \Delta \int_0^\infty P(R) G(\omega_d t) dR \quad [2]$$

where Δ is the modulation depth, $G(\omega_d t)$ has been defined previously (Brandon et al., 2012), $P(R)$ is the distribution of distances between spins, and all orientation effects have been ignored. Modeling $P(R)$ as a sum of n Gaussian components,

$$P(R) = \begin{cases} a_1 p_1 & \text{for } n=1 \\ \left(\prod_{k=1}^n a_k \right) p_n + \sum_{j=1}^{n-1} \left(\prod_{k=1}^j a_k \right) (1 - a_{j+1}) p_j & \text{for } n>1 \end{cases} \quad [3]$$

where $a_1 \equiv 1$ and the Gaussian components are given by

$$p_j(R) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(R-r_{0j})^2}{2\sigma_j^2}} \quad [4]$$

The mean and variance of $P(R)$ are given by $\langle R \rangle$ and σ_R respectively (Brandon et al., 2012). For a given number of components, n , the unknown parameters are the set of r_{0j} , σ_j , and a_j which define $P(R)$.

The exponential function is given by

$$E(t) = e^{(-\alpha|t|)^\beta} \quad [5]$$

where $\alpha = 10^\lambda$ and $\beta = d/3$ with $d = 3$ for an ideal 3-dimensional homogeneous solution. When fitting using an exponential or stretched exponential background model, the unknown parameter is λ for fixed d . $B(t)$ can be calculated by a numerical approach for simulating a homogeneous solution of spheres each with a single spin at their center. This approach has been successfully used to account for excluded volume effects (Brandon et al., 2012). More recently, Hinderberger and colleagues have developed an elegant and computationally efficient analytical method for computing excluded volume effects for a variety of geometries (Kattinig, Reichenwallner & Hinderberger, 2013).

For a given data set and a given model, the best-fit parameters are obtained by finding the set of parameters defined in Equations 1 through 5 that minimizes the reduced chi-squared value given in Equation 6,

$$\chi^2_\nu = \frac{1}{(N-q)} \sum_{i=1}^N \frac{[V(t_i) - F(t_i)]^2}{s_i^2} \quad [6]$$

where $V(t)$ is the experimental data, $F(t)$ is the fit function, N is the total number of data points, q is the number of parameters varied in the fit, and s_i is the estimated noise level of the i^{th} data point, and t_i is the time value of the i^{th} data point. Custom written Levenberg-Marquardt and simulated annealing algorithms have been implemented which perform better than the standard routines available in Matlab (Brandon et al., 2012). The reduced chi-squared takes into account the number of data points (N), the model complexity via the number of fit parameters (q), and the inherent noise in the data as estimated by s_i . As a general rule, $\chi^2_\nu > 1$ indicates a poor fit to the data, $\chi^2_\nu \approx 1$ indicates that the error in the fit is fully explained by the noise in the data and the fit is good, and $\chi^2_\nu < 1$ suggests that either the model is over-fitting the data or that the noise in the data has been overestimated.

The choice of a best model to fit a given data set is made using the Akaike information criterion corrected for finite sample size ($AICc$) given in Equation 7

$$AICc = N \ln \left(\frac{\sum_{i=1}^N [V(t_i) - F(t_i)]^2}{N} \right) + 2K + \frac{2K(K+1)}{N-K-1} \quad [7]$$

where $K = q + 1$. As data are fit using distance distributions with increasing number of Gaussian components the sum of the squared residuals in the first term of Equation 7 decreases as the value of K in the second term increases. The third term is a minor correction for a finite N value. According to information theory, the best model from a given set of models is the one with the lowest $AICc$ (Burnham & Anderson, 2002). Thus, $AICc$ can be used to determine the optimum number of Gaussians needed to fit a given data set without overfitting. $AICc$ values are used to compare the relative merit of different models but individually they do not give any indication of the quality of a particular fit.

Using DD

Originally, the program **GLADD** was developed to provide a flexible framework for the global analysis of data from multiple DEER experiments (Brandon et al., 2012). Global analysis and **GLADD** will be discussed in detail below. More recently, **DD** was developed to provide a convenient tool for analyzing data from single experiments. In this and the two following sections, the use of **DD** will be illustrated using example data sets. A third program, **DEERsim**, has been developed for simulating DEER data. For the first example, data was collected at X-band from T4 lysozyme labeled at residues 82 and 132 (T4L

82-132) (Kazmier, Alexander, Meiler & Mchaourab, 2011). For the second example, data was collected at X-band from the dimer of cytoplasmic domain of the band 3 protein labeled at residue 341 (CDB3 341) (Zhou et al., 2005; Zhou et al., 2007). The programs, sample data sets, and additional instructions for their use are available on the World Wide Web (www.mc.vanderbilt.edu/root/vumc.php?site=hustedtlab&doc=42160). Names, labels, and commands associated with the program will be bolded. Values obtained for the particular data set will be given in parentheses.

- 1 Make sure that the current folder in Matlab is set to the directory containing **DD**. Starting the program **DD** in Matlab will open a graphical user interface (GUI) similar to that shown in Figure 4.

The **DD** GUI is divided from top to bottom into several different sections. The first (**Data File**) controls the loading and pre-processing of data. The second (**Distance Distribution**) defines the model and parameters of the $P(R)$ for the intra-object distances. The number of components is controlled by the 3 checkboxes on the far left. Currently, **DD** is limited to $n = 3$. The three central numerical columns give current values for r_0 , σ , and a . The column on the right gives the calculated values of the relative amplitudes of the components (see Equation 3). For $n = 1$, $f_1 = 100\%$. For $n > 1$

$$f_n = 100 \times \prod_{k=1}^n a_k \text{ and } f_j = 100 \times (1 - a_{j+1}) \prod_{k=1}^j a_k \text{ for } j > n \quad [8]$$

The bottom row allows shapes other than Gaussian to be tried (**Square**, **Triangle**, **Circle**, and **Rice** (Kohler, Spitzbarth, Diederichs, Exner & Drescher, 2011)). The next section (**Background**) controls the parameters for the inter-object contribution to the DEER signal. The checkboxes on the far left control whether an **Exponential** background ($E(t)$) or a **Calculated** background ($B(t)$) that can include excluded volume effects is used. The next section controls a number of different constants that control the calculation of the DEER signal and generally do not need to be changed. The next section controls the modulation **depth** () and optional **scale** and **time shift** parameters. The next section contains pushbuttons for various commands and the final section controls output from **DD**. Checkboxes to the right of any parameter control whether the parameter will be varied (*floated*) during the fit. When checked, the parameter will be floated. When unchecked, the parameter will be held constant (*fixed*) at the value displayed.

- 2 Use the **Find** pushbutton to locate and load a given data set (C:\Example1\t4l82-132). **DD** can read Bruker formatted or ASCII data files. A second window will open which initially displays the unprocessed (raw) data (Figure 5, upper left). If necessary, points can be truncated from the data set under the **MISC** tab in **OPTIONS**.
- 3 Use the **Process** pushbutton to phase adjust (-5.3°), and time shift (9.79 ns), and scale the raw data. The first two operations

can be disabled by unchecking the **Phase Adjust** or **Zero Time Adjust** checkboxes.

The DEER data is phase-corrected by finding the phase that minimizes the sum of squared differences between the imaginary component and an unknown constant. This minimum sum of squared differences is subsequently used to estimate the noise level ($s_f = 2.11 \times 10^{-3}$) in the real component. Experimental DEER data is measured to an arbitrary scale factor and the true maximum value of the real component may be shifted from $t = 0$. Taking advantage of the symmetry of the DEER signal about $t = 0$ (Jeschke et al., 2006), data at small values of $|t|$ is fit directly to a Gaussian function (Figure 5, lower left). Using values from this Gaussian function, the data are scaled and time shifted such that the signal is now has a maximum value of approximately 1.0 at $t \approx 0$ (see Brandon et al., 2012 for additional details). This procedure does not force a data point to be equal to 1.0 at $t = 0$ since doing so can introduce artificial components at short distances into the $P(R)$. Note that obtaining a significant number of points at $t < 0$ is highly useful for accurately scaling and shifting the data. Also, since the data for small $|t|$ is highly sensitive to the distance distribution, all of the data including that for $t < 0$ is included in the subsequent fit to determine $P(R)$.

The procedures described in the preceding paragraph represent a minimal amount of processing prior to fitting the data. The scaling and time shifting that are applied in the **Process** step generally will suffice, however additional **scale** factor and **time shift** parameters can be determined during the least-squares fitting.

After the initial processing of the data, an initial simulation will be overlaid on the data calculated for all of the parameter values currently loaded in the GUI.

- 4 Use the **FIT** pushbutton to attempt to find the set of parameters that minimize χ^2_ν using a Levenberg-Marquardt algorithm. The results from fitting the T4L 82-132 data using a single Gaussian are given in Table 2 and Figure 5.
- 5 The **WRITE** pushbutton will output the results of the fit. By default, results are written to a single Microsoft Excel file with five different worksheets.
 - a. The contents of the **PARAMETERS(1)** worksheet include the initial fit parameters, the final fit parameters, and information on what parameters were varied.
 - b. The **OPTIONS** worksheet contains various options that control how **DD** operates and how the fit was performed.

- c. The **DATA(1)** worksheet contains information on the raw data.
- d. The **FIT(1)** worksheet includes the processed data and the final fit.
- e. The **DISTRIBUTION(1)** worksheet contains $P(R)$.

The relatively large value of $\chi^2_{\nu}=2.067$ and the presence of a significant oscillation in the plot of residuals, $V(t_i) - R(t_i)$, shown in the lower right corner of Figure 5 indicate that the data for T4L 82-132 is not well-fit using a unimodal Gaussian model for $P(R)$.

As will be discussed in detail in the next section, confidence intervals can be used to determine the uncertainties in the various fit parameters. Confidence interval calculations also provide a way to systematically search through parameter space for a better fit than the one found in the original minimization. In cases where the χ^2_{ν} surface contains multiple local minima, the results of any Levenberg-Marquardt minimization will depend on the initial set of parameter values. These calculations provide an efficient means of searching for and locating the global minimum.

6

The **Conf. Int.** pushbutton will bring up another GUI which can be used to control how confidence intervals for the various fit parameters will be calculated. There are four columns and a check box per parameter. The checkbox indicates whether a confidence interval will be carried out on that parameter. The columns contain the best-fit value of the parameter and then, the lower limit, the upper limit, and the increment that the parameter will be stepped over. The starting and ending values are input as the deviation away from the best fit parameter. For the T4L 82-132 data the best fit for the center of the single Gaussian fit is 26.5. To perform a series of confidence interval calculations over the range 25.5 to 27.5, -1.0 and $+1.0$ should be input into the lower and upper limit column, respectively. A value of 0.1 for the increment will calculate 21 points on this confidence interval. As the confidence interval calculations are performed for each parameter the results are plotted in new Matlab figures (Figure 6). As the confidence interval for each parameter is completed, the information is added to a new worksheet in the output Excel file.

Here, all four confidence interval curves pass through the minimum point (large circle) determined in the original fit. In the **DD** GUI, the χ^2_{ν} value from the original fit is displayed in green. The absolute minimum χ^2_{ν} value from all of the confidence interval calculations is shown in red to the right (Figure 4). These confidence interval calculations and other

calculations performed over a broader range of the parameters (not shown), all indicate that parameters from the original fit ($\chi^2_{\nu}=2.067$) are, in fact, the absolute best fit parameters for this data set with the unimodal Gaussian model.

The next step is to fit using a bimodal Gaussian model for $P(R)$.

- 7 Changing the fit model to a 2 component Gaussian is done by checking the **2nd component** checkbox. Doing so automatically places check marks to the right of the r_{0_2} , σ_2 , and a_2 parameters so that they will float during the fit. The results of any Levenberg-Marquardt minimization can depend on the initial values used if the algorithm becomes trapped in a local minimum. One approach to initiate a two component fit is to choose initial values for the r_0 parameters that bracket the value found for the one component fit. For this example, $r_{0_1} = 26.5 \text{ \AA}$ for the one component fit so $r_{0_1} = 24 \text{ \AA}$ and $r_{0_2} = 29 \text{ \AA}$ would be appropriate initial values for the two component fit. After inputting these values, use **Initialize** to perform an initial simulation, **FIT** to find the parameters that minimize χ^2_{ν} , **WRITE** to output the results to a new Excel file, and **Conf. Int.** to perform confidence interval calculations.

The results from fitting the data for T4L 82-132 using a bimodal Gaussian model are shown in Table 2 and Figure 7. The value of $\chi^2_{\nu}=1.0051$ and the lack of a significant oscillation in the plot of residuals (Figure 7B) indicate that the data for T4L 82-132 is well-fit by a bimodal Gaussian model for $P(R)$. The $AICc$ value for the bimodal Gaussian model (-1899.4) is significantly lower than that for the unimodal Gaussian model (-1791.2), indicating that the bimodal Gaussian model is preferred. Confidence interval calculations (not shown) demonstrate that the original fit is the best fit to the bimodal Gaussian model. The next step is to try a trimodal Gaussian model. Given the high quality of the bimodal Gaussian fit, it is not surprising that there is little improvement in the fit by adding a third component and it is difficult to find the best-fit parameters for the trimodal Gaussian model. Specifically, the final values obtained from the initial minimization do depend strongly on the initial values. One approach to initializing the parameters for the trimodal Gaussian model is to split the highest amplitude or the broadest component into two separate Gaussians.

- 8 Check the **3rd component** checkbox and select initial values for the distance distribution parameters. Press the pushbuttons **Initialize**, **FIT**, **WRITE**, and then **Conf. Int.** Regardless of the initial values used for this fit, the confidence interval calculations consistently reveal an improved fit to the data. The **Transfer** button will replace all of the parameters with the best-fit values found in the confidence interval calculations and perform an initial calculation. The process can then be repeated

(**FIT, WRITE, Conf. Int.**, and then **Transfer**) until it appears that the global minimum has been found.

The parameters from fitting the data for T4L 82-132 using a trimodal Gaussian model are given in Table 2. The $AICc$ value for the trimodal Gaussian fit (-1897.4) is higher than that obtained for the bimodal Gaussian model (-1899.4), thus the bimodal Gaussian model is preferred. The lack of improvement in the fit is supported by the fact that confidence intervals for the parameters r_{0_3} , σ_3 , and a_3 (not shown) contain multiple local minima that are statistically indistinguishable at the 66% confidence level.

Error Analysis

There are numerous approaches to estimating uncertainties for the parameters obtained from a non-linear least-squares fit (see discussion in Beechem, Gratton, Ameloot, Knutson & Brand, 1991; Bevington & Robinson, 1992). The simplest approach, estimation from the so-called curvature matrix at the χ^2_ν minimum, assumes that the uncertainty is symmetric about its best-fit value and that there are no correlations between parameters. Often neither of these assumptions is valid. The most computationally intensive approach is to perform an exhaustive search along each parameter axis while allowing all of the other parameters to vary in order to reach the χ^2_ν minimum at each point. Confidence intervals were introduced in the section above as a tool for fully exploring the χ^2_ν surface and will be described in detail here in the context of error analysis.

Figure 8 shows the calculation of a confidence interval for r_{0_1} for the bimodal Gaussian with exponential background fit to the T4L 82-132 data that is discussed above (Table 2 and Figure 7). Additional calculations are performed over a series of r_{0_1} values spanning a range around the value obtained at the χ^2_ν minimum (large circle). At each r_{0_1} value, another Levenberg-Marquardt minimization is performed to find the best-fit values of the other 6 parameters. Figure 9 shows how the other parameters vary with r_{0_1} . For this data set, neither λ nor σ_1 vary dramatically as a function of r_{0_1} while the other parameters (σ_1 , r_{0_2} , σ_2 , and a_2) are all highly correlated to r_{0_1} . The horizontal dashed lines in Figure 8 show where statistically significant changes in χ^2_ν occur. The 66% and 95% levels correspond approximately to 1 and 2 standard deviations, respectively. The vertical lines have been added to show how the uncertainty in r_{0_1} can be read from the graph. The asymmetry in the uncertainty is evident in this case. At 1 standard deviation

$$r_{0_1} = 23.4^{+1/2}_{-0/8} \text{ \AA} \text{ and at 2 standard deviations } r_{0_1} = 23.4^{+2.8}_{-1.7} \text{ \AA}.$$

This approach to determining the parameter uncertainty fully accounts for correlations between parameters and allows for asymmetric uncertainties. Given these advantages, the computational demands are more than tolerable.

The Background Signal

The general method of background correction is to assume that the background can be parameterized by an exponential for a homogeneous 3-dimensional solution or a stretched

exponential for other dimensions. This assumption is strictly valid for point spins and works for the majority of samples. However, for very large objects there can be a significant excluded volume effect (Milov & Tsvetkov, 1997; Junk et al., 2011; Brandon et al., 2012). This effect occurs when the size of the objects limits their distance of closest approach, reducing the contribution of short distances (and high frequencies) to the background, and thus producing an apparent flattening of the background at early times. It would be difficult, at best, to account for an excluded volume effect while applying an *a priori* background correction. **DD** includes an option for including a background signal calculated to include an excluded volume effect in the fit. The following example for CDB3 341 highlights the use of this calculated background option.

1. In the **Background** Section click in the checkbox for **Calculated**. This will fit the background using a concentration (**[]**) describing the spin concentration (in μM) and a radius, **ρ** (in $\text{\AA}$), that characterizes the excluded volume. The distance of closest approach is given by 2ρ . The default option is for both the concentration and **ρ** to float during the fit.
2. Continue the fit as described above starting with Step 4.

Table 3 gives the results for 6 different models used to fit the CDB3 341 data. The *AIC_c* values for these models clearly show that the preferred model is a bimodal Gaussian distribution with a large excluded volume effect ($\rho = 29 \text{ \AA}$). The fits to the data from CDB3 341 using a bimodal Gaussian distance distribution and either an exponential background or a calculated background are shown in Figure 10. There is a modest difference in the distance distributions obtained using the two different background corrections, but the fit to the data is poor for the exponential background. Similarly poor results are seen applying an exponential background *a priori* using **DeerAnalysis**. Larger excluded volume effects have been seen in other CDB3 samples (Brandon et al., 2012) depending on the site labeled. Given the dimensions of the CDB3 dimer observed by X-ray crystallography (Zhang, Kiyatkin, Bolin & Low, 2000), $75 \times 55 \times 45 \text{ \AA}$, and a Stokes radius of $53 \text{ \AA}$ (Appell & Low, 1981), the large excluded volume seen from some CDB3 samples is not surprising. For smaller proteins (e.g. T4L), attempts to fit with an excluded volume background typically give small best-fit values of ρ with no significant improvement in the fit versus an exponential background. Background signals calculated for $\rho = 0 \text{ \AA}$ are exponential decays.

It is an axiom of data analysis that data should not be transformed or manipulated, apart from trivial scaling operations, prior to fitting (Beechem, 1992). Systematic errors introduced into the data by an improper *a priori* background cannot be undone by the subsequent fitting and can lead to significant errors in the distance distribution obtained. While the fitting of DEER data to a model that simultaneously includes the intra-object distance distribution and the inter-object background is moderately challenging, we have found that with scaling of the parameters (e.g. fitting to λ and not α in Equation 5) and improvements to the Levenberg-Marquardt routine (Brandon et al., 2012) the problem is tractable. In the following discussion, we present examples of (simulated) data where the combination of a long interspin distance and a short data time span (as determined by τ_2 in

the 4-pulse DEER sequence of Figure 1) make isolating the background from the total DEER signal somewhat difficult (Figure 11).

The data for the longest τ_2 (~6 μ s) are fit well by both **DD** and using *a priori* background correction as performed in **DeerAnalysis** (Figure 11A). The *a priori* background correction (dashed purple line) misses the true background (solid black line) and there are some minor peaks that appear in the distribution (inset), but the main peak is in reasonable agreement with the true distribution. The fit, background contribution, and distance distribution obtained from **DD** (black lines) all agree with the simulated data.

At an intermediate τ_2 (~4 μ s, Figure 11B), **DD** slightly overestimates the width of the distribution but gives excellent fits to the data (and the background component). The *a priori* background correction misses the true background resulting in a distance distribution that is both too narrow and at the wrong average distance.

At the shortest τ_2 (~3 μ s), **DD** continues its trend of broadening the distance distribution while giving excellent fit to the data and matching the true background. On the other hand, *a priori* background correction again misses the background and now gives a bimodal distribution. Using the Validation Tool of **DeerAnalysis** it is possible to modify the time range fit to determine the *a priori* background. However, in this case no significant change in the distance distribution is obtained.

Fitting DEER data with a model-based distance distribution together with a background component also allows for a more thorough analysis of the certainty in the fitting parameters. Figure 12 shows error analysis and the χ^2 plots for the four variables (r_0 , λ , r_{0-1} , and σ_1) from the fits in Figure 11. For the longest τ_2 , all four parameters have well-defined minima and the curves are symmetric about the minima (red squares). The uncertainty in all of the parameters increases with decreasing τ_2 as expected. However, the only undefined parameter is σ_1 at the shorter data collection times. On the other hand, the mean of the Gaussian distribution is accurately and precisely defined for all three time spans

($r_{0-1}=60.00^{+0.16}_{-0.19}$ for $\tau_2 = 6 \mu$ s, $r_{0-1}=60.1^{+1.1}_{-0.4}$ for $\tau_2 = 4 \mu$ s, and $r_{0-1}=59.9^{+1.9}_{-0.7}$ for $\tau_2 = 3 \mu$ s). Unlike the results shown above for T4L 82-132 (Figures 9A and 9B), in this case for the shorter τ_2 values there is a strong correlation between r_{0-1} and both r_0 and λ (not shown). By including the background in the fit function and performing an exhaustive search during the calculation of the confidence intervals, the calculations take into account all correlations between parameters. This method gives the maximum possible variation in a fit parameter consistent with a given confidence level. Nonetheless, these results demonstrate that even under challenging combinations of τ_2 and $P(R)$, estimates of r_0 values and their uncertainties can be obtained that would allow firm conclusions to be drawn regarding changes seen between different samples (see Figures 10 and 11 in Brandon et al., 2012 for an example of how results from different samples can be compared using confidence intervals).

Global Analysis

It has long been known that multiple spectroscopic data sets from a variety of techniques could be analyzed globally (simultaneously) in order to enhance one's ability to recover

relevant information more accurately and more precisely than from the analysis of a single experiment. Moreover, these methods permitted straightforward rigorous error analysis of the recovered parameters as well as an assessment of their possible covariance. Perhaps nowhere were these global analysis methods more widely used than in the area of fluorescence spectroscopy as demonstrated in a number of seminal papers published in the late 1970's and throughout the 1980's (reviewed in Beechem, 1992; Beechem et al., 1991). We first adopted the methods of global analysis, based upon non-linear least-squares and the Levenberg-Marquardt algorithm to simultaneously analyze Saturation Transfer EPR (ST-EPR) data and transient-phosphorescence anisotropy decay data to determine the rotational correlation time for proteins in solution (Hustedt, Cobb, Beth & Beechem, 1993). These early studies were followed by analyzing ST-EPR data on the spin-labeled anion exchange protein in human erythrocytes, collected at different Zeeman field modulation frequencies, in order to determine the proteins uniaxial rotational diffusion coefficient (Hustedt & Beth, 1995). This work was extended to analysis of ST-EPR data obtained from the EGF receptor in A431 cells a few years later (Stein, Hustedt, Staros & Beth, 2002). This same global analysis approach was also employed to determine the distance between two dipolar coupled spin labels as well as all 5 angles relating the two labels relative to the inter-electron axis from CW-EPR data obtained at different (X-, Q-, and W-band) microwave frequencies (Hustedt et al., 1997). This study was instrumental in demonstrating the capabilities of global analysis for determining the confidence intervals for all six fitting parameters. More recently, we have extended global analysis to the fitting of time domain DEER data (Brandon et al., 2012). In all of these cases, the simultaneous analysis of multiple data sets using the methods of global analysis has significantly increased our ability to carry out rigorous analysis of experimental data from a wide range of systems.

Global analysis is performed by modifying the equation for χ^2_ν to include a sum over all the experimental data sets included in the fit so that

$$\chi^2_\nu = \sum_{m=1}^M \sum_{i=1}^{N_m} \frac{[V_m(t_i) - F_m(t_i)]^2}{s_{mi}^2 [\sum_m N_m - Q - 1]} \quad [9]$$

where M is the total number of data sets being fit, N_m is the number of data points in the m^{th} data set, s_{mi} is the estimated noise level for the i^{th} point in the m^{th} data set, and Q is the total number of parameters being varied in the fit (Beechem et al., 1991).

Global analysis of DEER data will be illustrated here with two different examples. The rationale behind these analyses is the concept that changes in the biochemical state of the protein lead to shifts in equilibrium between two or more structural conformations. These structural conformations manifest as multiple Gaussians components in the distance distribution. The assumption is that the multiple data sets can be fit with the same Gaussian components with varying amplitudes for the different biochemical states.

The first analysis uses DEER data from the multidrug transporter LmrP labeled at residues 160 and 310 (LmrP 160-310) collected at eight different pH values. These data reveal a shift

in the conformational equilibrium of the protein as the protonation state of acidic residues changes. Representative data are shown in Figure 13. Preliminary fits to the individual data sets suggest that the eight data sets can be globally fit using distance distributions expressed as the sum of the same two or three different components with varying amplitudes. The insets in Figure 13 show the distance distributions obtained from these fits to the individual data sets. At pH = 4.5, there are two distinct components at $r_0 \approx 36$ Å and $r_0 \approx 46$ Å. At pH = 6.0, there are two components at roughly the same distances as those obtained at pH = 4.5, $r_0 \approx 37$ Å and $r_0 \approx 47$ Å, with different relative amplitude. At pH = 8.0, the component at the shorter distance is not present and the region of the longer distance peak now contains both a narrow and broad component. These results, together with the analyses of data collected at other pH values (not shown), all indicate that as the pH is raised the equilibrium shifts towards the conformation at the longer distance.

In a previously published report a bimodal global fit of these data was presented (Mishra et al., 2014). Here, trimodal distributions will be used. Global analysis of DEER data is performed using **GLADD**. Details of how to use **GLADD** are available for download www.mc.vanderbilt.edu/root/vumc.php?site=hustedtlab&doc=42160. Briefly, output Excel files obtained from fitting the individual data sets using **DD** can be combined into a single Excel that will be used as input to **GLADD**. Modifications are made to this combined Excel file to determine which parameters from which experiments are *linked*. Parameters for different data sets that are linked are required to have the same value. One of the virtues of **GLADD** is that it allows total flexibility for how different parameters for different data sets are linked as well as what parameters will be floated during the fit and what parameters will be fixed. Here the distance distribution is modeled as

$$P(R) = (1 - a_2)p_1 + a_2[(1 - a_3)p_2 + a_3p_3] \quad [10]$$

where the p_j are the Gaussian components defined in Equation 4. For all eight data sets, the values of the r_{0j} and σ_j were linked. The best-fit values are given in Table 4. Also, a_3 was linked across all data sets (best-fit value $a_3 = 0.36$) so that the model corresponds to a two-state model with one state (outward-closed) corresponding to the short distance component (p_1) and the other (outward-open) corresponding to the two longer distance components ($p_2 + p_3$). The values of a_2 varied across the different data sets. Numerous experimental factors can lead to variations in the background and modulation depth so λ and ν were also allowed to vary (Table 5).

Excellent fits were obtained for all eight experiments. Representative fits are shown overlaid on the data in Figure 13. The distance distributions for all eight global fits are shown in Figure 14. Figure 15 shows the best-fit values of a_2 plotted as a function of pH with error bars obtained from confidence interval calculations. This global analysis allows all eight data sets to be simultaneously fit to a simple model directly related to the pH-dependent shifts in the conformational equilibrium underlying the transport mechanism (Masureel et al., 2014).

The second example of global data analysis is for the heterodimeric, ABC multidrug transporter BmrCD labeled at residues 348 and 532 (BmrCD 348-532). This protein utilizes the energy of ATP hydrolysis to export molecules out of the cell. In this case the changes in the biochemical state have been observed in the absence of nucleotide (apo) or in various nucleotide-bound states. The use of the non-hydrolyzable ATP analogue, AMP-PNP, is a mimic for nucleotide binding; the use of ATP and Vanadate (Vi) traps the ATPase domain in a post hydrolysis state; and ATP is used to explore the predominant states during turnover. Figure 16 shows DEER data sets from BmrCD under these four different conditions (ATP, AMP-PNP, ATP-Vi, and apo). Preliminary analyses of the individual data sets, the results of which are shown as insets in Figure 16. The ATP-Vi sample alone was fit with a bimodal Gaussian $P(R)$ with both components at short distances while the AMP-PNP sample was fit with two components at longer distances. The ATP sample shows a contribution from the two short components (as a single broader peak) and from the longer components as well, while the apo sample clearly is dominated by the longer components. All four data sets were successfully fit in a single global analysis as shown by the fits overlaid on the data in Figure 16 using a total of 4 Gaussians. The four distance distributions from the global fit are shown in Figure 17. While the results from the individual fits to the four experiments suggest a common set of components between all four samples, the different Gaussians from the four fits do not precisely overlap. The quality of the fits from the global analysis supports the hypothesis of a common set of components and conformations across all conditions.

Conclusions

There has been a tremendous growth in the utilization of DEER measurements for characterizing the structures and conformational dynamics of biomolecules in the past decade. As these studies advance into problems with ever increasing complexity, there is a clear need for robust analysis methods that allow the extraction of inter-probe distance and distance distribution data from systems where multiple conformational states exist and thus multiple average inter-probe distances and distance distributions exist. In this chapter, methods have been described which are ideally suited to these types of problems. The **DD** program presented here allows the fitting of raw experimental data without *a priori* background correction as well as the rigorous determination of confidence intervals for all fitting parameters. Including the background contribution in the fit itself arguably reduces the ambiguity of the results. By taking into account correlations between parameters, the confidence intervals, give a realistic representation of the uncertainties in the fit parameters.

This same methodological approach is used in the **GLADD** program allowing for simultaneous fitting of multiple experimental DEER data sets. A central theme of this chapter stems from the longstanding appreciation in the experimental data analysis community, particularly the fluorescence spectroscopy community, that direct fitting of experimental data without *a priori* transformation of the data provides a powerful way for obtaining unbiased results from the analyses. If the goal is to analyze data sets collected under different conditions in terms of varying ratios of a common set of components, then the full set of time-domain data can and should be directly and globally fit to that model.

Improvements to both DD and GLADD are being made continually. We encourage others to download and try these programs. Feedback is welcome.

Acknowledgments

H. Mchaourab, K. Kazmier, and Z. Zhou provided us with experimental data used in this work. The National Institutes of Health provided funding for this work through P01 GM080513 (EJH and AHB), S10 RR019120 (AHB), S10 RR027091 (AHB). RAS received support through NIH grants U54 GM087519 and R01 GM077659. H. Mchaourab provided invaluable insight and encouragement.

References

- Altenbach C, Flitsch SL, Khorana HG, Hubbell WL. Structural studies on transmembrane proteins. 2. Spin labeling of bacteriorhodopsin mutants at unique cysteines. *Biochemistry*. 1989; 28(19):7806–7812. [PubMed: 2558712]
- Appell KC, Low PS. Partial Structural Characterization Of The Cytoplasmic Domain Of The Erythrocyte-Membrane Protein, Band-3. *J Biol Chem*. 1981; 256(21):1104–1111.
- Beechem, JM.; Gratton, JM.; Ameloot, M.; Knutson, JR.; Brand, L. The global analysis of fluorescence intensity and anisotropy decay data: second generation theory and programs. In: Lakowicz, JR., editor. *Top Fluoresc Spectrosc (Vol. 2: Principles)*. New York: Plenum Press; 1991.
- Beechem JM. Global analysis of biochemical and biophysical data. *Methods Enzymol*. 1992; 210:37–54. [PubMed: 1584042]
- Bevington, PR.; Robinson, DK. *Data reduction and error analysis for the physical sciences*. 2. New York: McGraw-Hill; 1992.
- Bhatnagar J, Freed JH, Crane BR. Rigid body refinement of protein complexes with long-range distance restraints from pulsed dipolar ESR. *Method Enzymol*. 2007; 423:117–133.
- Borbat PP, Freed JH. Measuring distances by pulsed dipolar ESR spectroscopy: Spin-labeled histidine kinases. *Method Enzymol*. 2007; 423:52–116.
- Bowman MK, Maryasov AG, Kim N, DeRose VJ. Visualization of distance distribution from pulsed double electron–electron resonance data. *Appl Magn Reson*. 2004; 26(1–2):23–39.
- Brandon S, Beth AH, Hustedt EJ. The global analysis of DEER data. *J Magn Reson*. 2012; 218:93–104. [PubMed: 22578560]
- Burnham, KP.; Anderson, DR. *Model selection and multimodel inference: a practical information-theoretic approach*. 2. New York: Springer-Verlag; 2002.
- Cafiso DS. Peptide-membrane interactions determined using site-directed spin labeling. *Peptide-Lipid Interactions*. 2002; 52:3–29.
- Chiang YW, Borbat PP, Freed JH. The determination of pair distance distributions by pulsed ESR using Tikhonov regularization. *J Magn Reson*. 2005; 172(2):279–295. [PubMed: 15649755]
- Columbus L, Hubbell WL. A new spin on protein dynamics. *Trends Biochem Sci*. 2002; 27(6):288–295. [PubMed: 12069788]
- Dalmas O, Hyde HC, Hulse RE, Perozo E. Symmetry-Constrained Analysis of Pulsed Double Electron–electron Resonance (DEER) Spectroscopy Reveals the Dynamic Nature of the KcsA Activation Gate. *J Am Chem Soc*. 2012; 134(39):16360–16369. [PubMed: 22946877]
- Edwards DT, Huber T, Hussain S, Stone KM, Kinnebrew M, Kaminker I, Matalon E, Sherwin MS, Goldfarb D, Han SI. Determining the Oligomeric Structure of Proteorhodopsin by Gd3+-Based Pulsed Dipolar Spectroscopy of Multiple Distances. *Structure*. 2014; 22(11):1677–1686. [PubMed: 25438671]
- Fajer, PG.; Brown, L.; Song, L. *Practical pulsed dipolar ESR (DEER)*. Vol. 27. Kluwer Academic/Plenum Publishers; New York, NY: 2007.
- Fanucci GE, Cafiso DS. Recent advances and applications of site-directed spin labeling. *Curr Opin Struct Biol*. 2006; 16(5):644–653. [PubMed: 16949813]
- Georgieva ER, Borbat PP, Ginter C, Freed JH, Boudker O. Conformational ensemble of the sodium-coupled aspartate transporter. *Nat Struct Mol Biol*. 2013; 20(2):215–221. [PubMed: 23334289]

- Giannoulis A, Ward R, Branigan E, Naismith JH, Bode BE. PELDOR in rotationally symmetric homooligomers. *Molecular Physics*. 2013; 111(18–19):2845–2854. [PubMed: 24954956]
- Hubbell WL, Mchaourab HS, Altenbach C, Lietzow MA. Watching proteins move using site-directed spin labeling. *Structure*. 1996; 4(7):779–783. [PubMed: 8805569]
- Hubbell WL, Gross A, Langen R, Lietzow MA. Recent advances in site-directed spin labeling of proteins. *Curr Opin Struct Biol*. 1998; 8(5):649–656. [PubMed: 9818271]
- Hubbell WL, Cafiso DS, Altenbach C. Identifying conformational changes with site-directed spin labeling. *Nat Struct Biol*. 2000; 7(9):735–739. [PubMed: 10966640]
- Hubbell WL, Altenbach C, Hubbell CM, Khorana HG. Rhodopsin structure, dynamics, and activation: A perspective from crystallography, site-directed spin labeling, sulfhydryl reactivity, and disulfide cross-linking. *Membrane Proteins*. 2003; 63:243–290.
- Hustedt EJ, Cobb CE, Beth AH, Beechem JM. Measurement of rotational dynamics by the simultaneous nonlinear analysis of optical and EPR data. *Biophys J*. 1993; 64(3):614–621. [PubMed: 7682452]
- Hustedt EJ, Beth AH. Analysis of saturation transfer electron paramagnetic resonance spectra of a spin-labeled integral membrane protein, band 3, in terms of the uniaxial rotational diffusion model. *Biophys J*. 1995; 69(4):1409–1423. [PubMed: 8534811]
- Hustedt EJ, Smirnov AI, Laub CF, Cobb CE, Beth AH. Molecular distances from dipolar coupled spin-labels: the global analysis of multifrequency continuous wave electron paramagnetic resonance data. *Biophys J*. 1997; 72(4):1861–1877. [PubMed: 9083690]
- Hustedt EJ, Beth AH. Nitroxide spin-spin interactions: applications to protein structure and dynamics. *Annu Rev Biophys Biomol Struct*. 1999; 28:129–153. [PubMed: 10410798]
- Hustedt, EJ.; Beth, AH. Structural information from CW-EPR spectra of dipolar coupled nitroxide spin labels. In: Berliner, LJ.; Eaton, SS.; Eaton, GR., editors. *Biological Magnetic Resonance*. New York: Kluwer Academic/Plenum Publishers; 2000. p. 155–184. Vol. Volume 19: Distance Measurements in Biological Systems by EPR
- Hustedt EJ, Stein RA, Sethaphong L, Brandon S, Zhou Z, Desensi SC. Dipolar coupling between nitroxide spin labels: the development and application of a tether-in-a-cone model. *Biophys J*. 2006; 90(1):340–356. [PubMed: 16214868]
- Jeschke G. Distance measurements in the nanometer range by pulse EPR. *Chemphyschem*. 2002; 3(11):927–932. [PubMed: 12503132]
- Jeschke G, Koch A, Jonas U, Godt A. Direct conversion of EPR dipolar time evolution data to distance distributions. *J Magn Reson*. 2002; 155(1):72–82. [PubMed: 11945035]
- Jeschke G, Panek G, Godt A, Bender A, Paulsen H. Data analysis procedures for pulse ELDOR measurements of broad distance distributions. *Appl Magn Reson*. 2004; 26(1–2):223–244.
- Jeschke G, Chechik V, Ionita P, Godt A, Zimmermann H, Banham J, Timmel CR, Hilger D, Jung H. DeerAnalysis2006 - a comprehensive software package for analyzing pulsed ELDOR data. *Appl Magn Reson*. 2006; 30(3–4):473–498.
- Jeschke G, Sajid M, Schulte M, Godt A. Three-spin correlations in double electron–electron resonance. *PCCP*. 2009; 11(31):6580–6591. [PubMed: 19639133]
- Jeschke, G. DEER Distance Measurements on Proteins. In: Johnson, MA.; Martinez, TJ., editors. *Annual Review of Physical Chemistry*. Vol. 63. 2012. p. 419–446.
- Junk MJN, Spiess HW, Hinderberger D. DEER in biological multispin-systems: A case study on the fatty acid binding to human serum albumin. *J Magn Reson*. 2011; 210(2):210–217. [PubMed: 21450500]
- Kattinig DR, Reichenwallner J, Hinderberger D. Modeling excluded volume effects for the faithful description of the background signal in double electron–electron resonance. *J Phys Chem B*. 2013; 117(51):16542–16557. [PubMed: 24245922]
- Kazmier K, Alexander NS, Meiler J, Mchaourab HS. Algorithm for selection of optimized EPR distance restraints for de novo protein structure determination. *J Struct Biol*. 2011; 173(3):549–557. [PubMed: 21074624]
- Kear JL, Blackburn ME, Veloro AM, Dunn BM, Fanucci GE. Subtype Polymorphisms Among HIV-1 Protease Variants Confer Altered Flap Conformations and Flexibility. *J Am Chem Soc*. 2009; 131(41):14650–14651. [PubMed: 19788299]

- Kittell AW, Hustedt EJ, Hyde JS. Inter-spin distance determination using L-band (1–2 GHz) non-adiabatic rapid sweep electron paramagnetic resonance (NARS EPR). *J Magn Reson.* 2012; 221:51–56. [PubMed: 22750251]
- Klug, CS.; Feix, JB. SDSL: A survey of biological applications. In: Eaton, SS.; Eaton, GR.; Berliner, LJ., editors. *Biological Magnetic Resonance*. New York, NY: Kluwer Academic/Plenum Publishers; 2005. p. 269–308. Vol. Volume 24: Biomedical EPR, Part B: Methodology, Instrumentation, and Dynamics
- Klug, CS.; Feix, JB. Methods and applications of site-directed spin Labeling EPR Spectroscopy. In: Correia, JJ.; Detrich, HW., editors. *Biophysical Tools for Biologists: Vol 1 in Vitro Techniques*. Vol. 84. 2008. p. 617–658.
- Kohler SD, Spitzbarth M, Diederichs K, Exner TE, Drescher M. A short note on the analysis of distance measurements by electron paramagnetic resonance. *J Magn Reson.* 2011; 208(1):167–170. [PubMed: 21044853]
- Masureel M, Martens C, Stein RA, Mishra S, Ruyschaert JM, Mchaourab HS, Govaerts C. Protonation drives the conformational switch in the multidrug transporter LmrP. *Nat Chem Biol.* 2014; 10(2):149–155. [PubMed: 24316739]
- Milov AD, Tsvetkov YD. Double electron–electron resonance in electron spin echo: Conformations of spin-labeled poly-4-vinylpyridine in glassy solutions. *Appl Magn Reson.* 1997; 12(4):495–504.
- Mishra S, Verhalen B, Stein RA, Wen PC, Tajkhorshid E, Mchaourab HS. Conformational dynamics of the nucleotide binding domains and the power stroke of a heterodimeric ABC transporter. *eLife.* 2014;3.
- Pake GE. Nuclear Resonance Absorption in Hydrated Crystals: Fine Structure of the Proton Line. *J Chem Phys.* 1948; 16(4):327–336.
- Persson M, Harbridge JR, Hammarstrom P, Mitri R, Martensson LG, Carlsson U, Eaton GR, Eaton SS. Comparison of electron paramagnetic resonance methods to determine distances between spin labels on human carbonic anhydrase II. *Biophys J.* 2001; 80(6):2886–2897. [PubMed: 11371461]
- Phuong N, Qin PZ. RNA dynamics: perspectives from spin labels. *Wiley Interdisciplinary Reviews-RNA.* 2012; 3(1):62–72. [PubMed: 21882345]
- Rabenstein MD, Shin YK. Determination of the distance between two spin labels attached to a macromolecule. *Proc Natl Acad Sci U S A.* 1995; 92:8239–8243. [PubMed: 7667275]
- Reginsson GW, Schiemann O. Studying bimolecular complexes with pulsed electron–electron double resonance spectroscopy. *Biochem Soc Trans.* 2011a; 39:128–139. [PubMed: 21265760]
- Reginsson GW, Schiemann O. Pulsed electron–electron double resonance: beyond nanometre distance measurements on biomacromolecules. *Biochem J.* 2011b; 434:353–363. [PubMed: 21348855]
- Sen KI, Logan TM, Fajer PG. Protein dynamics and monomer-monomer interactions in AntR activation by electron paramagnetic resonance and double electron–electron resonance. *Biochemistry.* 2007; 46(41):11639–11649. [PubMed: 17880108]
- Sircar R, Borbat PP, Lynch MJ, Bhatnagar J, Beyersdorf MS, Halkides CJ, Freed JH, Crane BR. Assembly States of FliM and FliG within the Flagellar Switch Complex. *J Mol Biol.* 2015; 427(4): 867–886. [PubMed: 25536293]
- Stein RA, Hustedt EJ, Staros JV, Beth AH. Rotational dynamics of the epidermal growth factor receptor. *Biochemistry.* 2002; 41(6):1957–1964. [PubMed: 11827542]
- von Hagens T, Polyhach Y, Sajid M, Godt A, Jeschke G. Suppression of ghost distances in multiple-spin double electron–electron resonance. *PCCP.* 2013; 15(16):5854–5866. [PubMed: 23487036]
- Zhang DC, Kiyatkin A, Bolin JT, Low PS. Crystallographic structure and functional interpretation of the cytoplasmic domain of erythrocyte membrane band 3. *Blood.* 2000; 96(9):2925–2933. [PubMed: 11049968]
- Zhang, X.; Cekan, P.; Sigurdsson, ST.; Qin, PZ. Studying RNA using site-directed spin-labeling and continuous-wave electron paramagnetic resonance spectroscopy. In: Herschlag, D., editor. *Methods in Enzymology*, Vol 469: Biophysical, Chemical, and Functional Probes of RNA Structure, Interactions and Folding, Pt B. Vol. 469. 2009. p. 303–328.
- Zhou Z, DeSensi SC, Stein RA, Brandon S, Dixit M, McArdle EJ, Warren EM, Kroh HK, Song L, Cobb CE, Hustedt EJ, Beth AH. Solution structure of the cytoplasmic domain of erythrocyte

membrane band 3 determined by site-directed spin labeling. *Biochemistry*. 2005; 44(46):15115–15128. [PubMed: 16285715]

Zhou Z, DeSensi SC, Stein RA, Brandon S, Song L, Cobb CE, Hustedt EJ, Beth AH. Structure of the cytoplasmic domain of erythrocyte band 3 hereditary spherocytosis variant P327R: band 3 Tuscaloosa. *Biochemistry*. 2007; 46(36):10248–10257. [PubMed: 17696498]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. The 4-pulse DEER sequence. The refocused echo at the fixed time $2(\tau_1 + \tau_2)$ is measured as t is varied.

Figure 2.

Simulated DEER time-domain data for the different distance distributions shown in the insets. A) DEER signals for unimodal Gaussian $P(R)$ with different $\langle R \rangle$ and σ_R . $r_0 = 30$ Å and $\sigma = 4$ Å (solid black line); $r_0 = 30$ Å and $\sigma = 1$ Å (dotted red line); $r_0 = 40$ Å and $\sigma = 4$ Å (dashed green line). B) DEER signals for a unimodal Gaussian $P(R)$ ($r_0 = 30$ Å and $\sigma = 4$ Å, solid black line) and three bimodal Gaussian $P(R)$ ($r_{0,1} = 26.7$ Å, $\sigma_1 = 3$ Å, $r_{0,2} = 33.3$ Å, $\sigma_2 = 1$ Å, and $f_2 = 0.5$, dotted red line; $r_{0,1} = 26.7$ Å, $\sigma_1 = 1$ Å, $r_{0,2} = 33.3$ Å, $\sigma_2 = 3$ Å, and $f_2 = 0.5$, dashed green line; $r_{0,1} = 26.7$ Å, $\sigma_1 = 2.25$ Å, $r_{0,2} = 33.3$ Å, $\sigma_2 = 2.25$ Å, and $f_2 = 0.5$, dash-dot-dashed blue line) all with $\langle R \rangle = 30$ Å and $\sigma_R = 4$ Å. Signals do not include a background component. All calculations were performed with the **DEERsim** program in Matlab R2015a. All parameters are defined in Equations 1 through 5.

Figure 3.

Analysis of simulated DEER data using **DeerAnalysis** (version 2015 using Matlab R2015a) and Tikhonov regularization. A) Background corrected simulated DEER data (black dots) and fits obtained for two different regularization parameters, $\alpha = 1$ (solid red line) and $\alpha = 100$ (dashed blue line). Data was simulated using **DEERsim** for a bimodal Gaussian $P(R)$ ($r_{0_1} = 25 \text{ \AA}$, $\sigma_1 = 1 \text{ \AA}$, $r_{0_2} = 40 \text{ \AA}$, $\sigma_2 = 5 \text{ \AA}$, and $f_2 = 0.5$), an exponential background with $\lambda = 5$ (corresponding to $\sim 300 \text{ \mu M}$ spin concentration), and a modulation depth of 0.3. Random noise was added to give a signal-to-noise ratio of 200. B) The L-curve obtained from **DeerAnalysis**. C) The bimodal Gaussian $P(R)$ used to generate the simulated data (shaded region) and those obtained from fitting using $\alpha = 1$ (solid red line) and $\alpha = 100$ (dashed blue line).

Figure 4. The **DD** GUI after fitting DEER data for T4L 82-132 using a unimodal Gaussian with exponential background model. Data courtesy of Drs. K. Kazmier and H. Mchaourab (Kazmier et al., 2011).

Figure 5.

Graphical output from **DD** after fitting DEER data for T4L 82-132 using a unimodal Gaussian with exponential background model.

Figure 6.

Confidence interval plots generated by **DD** for A) , B) λ , C) r_{0_1} , and D) σ_1 while fitting DEER data for T4L 82-132 using a unimodal Gaussian with exponential background model.

Figure 7.

Results from fitting DEER data for T4L 82-132 with a bimodal Gaussian with exponential background model using **DD**. A) The data (black dots) together with the best fit (solid red line) and the background component (dashed blue line). The parameters are given in Table 2. B) The residuals of the fit. C) The distance distribution obtained from the **DD** fit (solid red line) and, for comparison, that obtained from fitting the same data using Tikhonov regularization in **DeerAnalysis** (dashed green line).

Figure 8.

Confidence interval plots generated by **DD** for r_{0_1} for T4L 82-132 using a bimodal Gaussian with exponential background model (purple line with asterisk symbols). The χ^2_{ν} minimum point is shown as a large open circle. The vertical lines have been added to show how the uncertainty in r_{0_1} can be read from the graph for two different confidence levels.

Figure 9.

Plots showing how the 6 other fit parameters vary during the calculation of the confidence interval for r_{0_1} for T4L 82-132 fit to a bimodal Gaussian with exponential background. The fit is shown in Figure 7 and the best-fit parameters given in Table 2. The confidence interval for r_{0_1} is shown in Figure 8.

Figure 10.

Results from fitting DEER data for CDB3 341 (black dots). A) Data was fit using **DD** using a bimodal Gaussian $P(R)$ and either a calculated background including an excluded volume effect (top panel, solid red line) or an exponential background (bottom panel, solid green line). The parameters are given in Table 3. For comparison, the two different background signals, calculated (dashed red line) and exponential (dotted green line), are both plotted on the upper panel. B) The distance distribution obtained from the fit (solid red line) using a bimodal Gaussian with calculated background model and that obtained using a bimodal Gaussian with exponential background (dashed green line). The data is taken from (Zhou et al., 2005).

Figure 11.

Analysis of simulated DEER data calculated for a unimodal Gaussian $P(R)$ over three different time spans. Data was simulated for $r_0 = 60 \text{ \AA}$ and $\sigma = 2 \text{ \AA}$, an exponential background with $\lambda = 4.75$ (corresponding to $\sim 175 \mu\text{M}$ spin concentration), and a modulation depth of 0.2. Random noise was added to give a signal-to-noise ratio of 500. Data was calculated for A) $t = -0.18 - 5.82 \mu\text{s}$ (red dots), B) $t = -0.12 - 3.88 \mu\text{s}$ (green dots), and C) $t = -0.09 - 2.91 \mu\text{s}$ (blue). The true background signals for each simulation are also shown as colored dots. Excellent fits (overlaid black lines) were obtained using **DD** for all three simulated data sets using a unimodal Gaussian with exponential background model. The background component to each of the fit is also shown as a solid black line. Insets show

the true $P(R)$ (shaded region), the best-fit $P(R)$ obtained using **DD** (solid black line), and the best-fit $P(R)$ obtained using Tikhonov regularization in **DeerAnalysis** (dashed purple line). The *a priori* background corrections from **DeerAnalysis** are also shown as dashed purple lines. All simulated data were generated using **DEERsim**.

Figure 12.

Confidence intervals for the 3 fits in Figure 10. Confidence intervals for the $t = -0.18 - 5.82$ μs data (Figure 10A) are plotted as red squares. Confidence intervals for the $t = -0.12 - 3.88$ μs data (Figure 10B) are plotted as green circles. Confidence intervals for the $t = -0.09 - 2.91$ μs data (Figure 10C) are plotted as blue triangles. All confidence intervals were adjusted slightly to have the same minimum $\chi^2_{\nu} = 1$. The 95% confidence level as determined using the F -test is shown as a horizontal line.

Figure 13.

Three of the eight DEER data sets from LmrP 160-310 that were collected as a function of pH and analyzed globally using **GLADD**. Data are plotted as black dots and fits overlaid as solid red lines. Details of the global analysis are given in the text and Tables 4 and 5. The *insets* show distance distributions obtained from analyzing the data *individually* using a bimodal Gaussian distance distribution. Data courtesy of Dr. H. Mchaourab (Masureel et al., 2014).

Figure 14.

The eight best-fit distance distributions that were obtained from the global analysis of DEER data from LmrP 160-310 collected as a function of pH. Representative data and fits are shown in Figure 14.

Figure 15.

The best-fit a_2 values from the results in Table 5. The a_2 values give the fraction of $P(R)$ in the two longer distance components (see Equation 10). Error bars were obtained from 95% confidence levels.

Figure 16.

Fits to four data sets from BmrCD 348-532 collected under different conditions. All four data sets were fit using the same 4 Gaussians with different relative amplitudes. A) ATP (red dots). B) AMP-PNP (green dots). C) ATP-Vi (blue dots). D) apo (purple dots). All fits are shown as solid black lines. The *insets* show distance distributions obtained from analyzing the data *individually* using a bimodal Gaussian distance distribution. Data courtesy of Dr. H. Mchaourab (Mishra et al., 2014).

Figure 17.

The four distance distributions obtained from the global analysis of the BmrCD 348-532 data in Figure 17 (Tables 6 and 7).

Table 1

Software available for the analysis of DEER data.

Name	Author ¹	<i>a priori</i> backgr. corr. ²	$P(R)$ ³	website ⁴
DeerAnalysis	Jeschke	recommended yes	various ⁵ Tik. reg. ⁶	www.epr.ethz.ch/software
DEFit	Fajer	yes	Gaussians ⁷	www.sb.fsu.edu/~fajer/Programs/DEFit/defit.html
LongDistances	Altenbach	yes yes	various Tik. reg.	sites.google.com/site/altenbach/labview-programs/epr-programs/long-distances
DeerA	Stein	no	Gaussians ⁷	http://www.mathworks.com/matlabcentral/fileexchange/46729-deera2012-zip/content/DeerA.m
DD/GLADD	Hustedt	no	various	www.mc.vanderbilt.edu/root/vumc.php?site=hustedtlab&doc=42160

¹ Primary author of software.² Whether *a priori* background correction is used.³ Method for determining distance distribution.⁴ Website where software can be downloaded.⁵ $P(R)$ can be fit using a variety of functional forms.⁶ $P(R)$ determined using Tikhonov regularization.⁷ $P(R)$ fit as a sum of Gaussians.

Table 2
Parameters obtained from various fits to the data from T4L 82-132. All parameters as defined in text.

Gaussians	λ	r_{ij} (Å)	σ_j (Å)	f	q	χ^2_{ν}	$AICc$
1	0.280	5.130	26.5	3.7	1	2.067	-1791.2
2	0.279	5.133	29.0 23.4	2.1 2.3	0.55 0.45	7	-1899.4
3	0.279	5.134	28.7 23.1 29.7	2.3 2.2 0.1	0.57 0.40 0.03	10	-1897.4

Table 3
Parameters obtained from various fits to the DEER data from the CDB3 341. All parameters as defined in text.

Gaussians	λ	[spin] (μM)	ρ ($\text{\AA}$)	r_{θ} ($\text{\AA}$)	σ ($\text{\AA}$)	f	q	χ^2_{ν}	AICc
1	0.222	5.012	-	34.2	2.6	1	4	1.846	-2236.6
2	0.222	5.013	-	35.4	1.4	0.70	7	1.468	-2288.5
3	0.222	5.010	-	30.4	2.4	0.37	10	1.394	-2297.6
1	0.253	-	416	28.7	3.4	1	5	1.401	-2300.4
2	0.254	-	420	33.1	4.9	0.65	8	1.041	-2370.3
3	0.254	-	423	35.4	1.0	0.35	11	1.033	-2368.5

Table 4

Parameters for the 3 Gaussians obtained from the global analysis of the DEER data from LmrP 160-310 collected as a function of pH.

Mode	r_{0j} (Å)	σ_j (Å)
1	36.7	1.1
2	46.8	2.4
3	45.4	8.0

Table 5

Parameters obtained from the global analysis of the DEER data from LmrP 160-310 collected as a function of pH.

pH		λ	a_2
4.5	0.145	4.39	0.44
5.0	0.138	4.38	0.65
5.5	0.138	4.50	0.91
6.0	0.139	4.39	0.88
6.5	0.158	4.48	0.94
7.0	0.178	4.54	0.97
7.5	0.199	4.56	1.00
8.0	0.130	4.30	0.99

Table 6

Parameters for the 4 Gaussians obtained from the global analysis of the DEER data from BmrCD 348-532.

Mode	r_{0j} (Å)	σ_j (Å)
1	29.1	1.4
2	33.3	2.0
3	47.7	4.3
4	66.0	13.7

Table 7
Parameters obtained from the global analysis of the DEER data from BmrCD 348-532.

Sample	λ	f_1	f_2	f_3	f_4
ATP	0.090	4.39	0.55	0.19	0.10
AMP-PNP	0.071	4.38	0.03	0.00	0.89
ATP-Vi	0.091	4.50	0.80	0.16	0.03
apo	0.139	4.39	0.00	0.02	0.14