

DeerAnalysis 2006 User's Manual

G. Jeschke

Max Planck Institute for Polymer Research
Postfach 3148, 55131 Mainz, Germany
jeschke@mpip-mainz.mpg.de

January 5, 2006

1 Purpose of the program

The program DeerAnalysis 2006 can extract distance distributions from dead-time free pulse ELDOR data (constant-time and variable-time four-pulse DEER) [1, 2]. Furthermore, it can be used for direct comparison of primary data of similar samples [3]. For a series of related samples, an average distance distribution can be computed taking into account the signal-to-noise ratio of the individual data sets. With some caution [4] the program may also be applied to the analysis of dead-time free double-quantum coherence EPR experiments [5]. It should not be used for data from experiments that have a significant dead time,

$$t_d > 2r^3 \text{ ns nm}^{-3}, \quad (1)$$

where r is the shortest distance in the distribution.

The underlying mathematical problem is (moderately) ill-posed, i.e., quality of the analyzed data is very crucial. Pre-processing tools are implemented to correct for experimental imperfections (phase errors, displacements of the time origin of the modulation) and to separate the intramolecular distances of interest from the intermolecular background contribution. Furthermore, the program provides several independent approaches for extracting the distance distribution, which helps to get a feeling for the reliability of the distribution. Characterization of the distance distribution in terms of its mean value $\langle r \rangle$ and width (standard deviation σ_r) is usually reliable [3] and is therefore a standard output. The performance of the different approaches for data analysis depends on the type of distance distribution (narrow or broad peaks or both) and was discussed in some detail in Ref. [3].

DeerAnalysis2006 is based on experience with earlier programs DeerFit, DeerTrafo, and in particular DeerAnalysis2004 as well as with model-specific fitting of data [6, 8]. It supersedes the earlier programs with respect to reliability and functionality. At the present time, DeerAnalysis2006 is released only as

source code that can be run within MATLAB but not as a stand-alone application. The older programs DeerAnalysis2004, DeerFit and DeerTrafo can still be downloaded at <http://www.mpip-mainz.mpg.de/jeschke/distance.html>.

2 Changes with respect to DeerAnalysis2004

The user interface of the new version DeerAnalysis2006 was written from scratch, while many of the computational subroutines are well tested subroutines of DeerAnalysis2004. The choice of analysis techniques was narrowed down to the ones that we and others found most reliable. An optimum regularization parameter for Tikhonov regularization is now predicted from the L curve as suggested by Freed's group [10], while the stabilizing constraint of a purely positive distance distribution is maintained as in our previous approach. Recent results of Tsvetkov's group on the effects of finite microwave pulse amplitude [11, 12] were also taken into account. An optional excitation bandwidth correction is now included that will be described in detail elsewhere. The following list gives an overview of the most important changes

- two data sets can be directly compared on screen (dual display)
- reasonably fast excitation bandwidth correction
- easy work with experimental background functions
- significance check for minor peaks in the distance distribution
- computation and display of L curves in Tikhonov regularization
- the total number n of coupled spins is now displayed, not $n - 1$
- results are no longer saved automatically
- user-defined models with up to six parameters can be fitted

3 Installation

DeerAnalysis2006 requires Matlab 6.5 or higher and was tested only in Windows environments. It should also run with Matlab for Linux or other Unix systems. However, design of the user interface may not be optimum for this platform. The windows package can be installed either from the small package DeerAnalysis-Compact.exe or from the full package DeerAnalysisFull.exe. In the former case, kernels should be computed on your own computer directly after installation (auxiliary program make tables.m), which may take half an hour or so. To do it, start Matlab, change to the installation directory (e.g., by `cd(c:/Program Files/DeerAnalysis2006)`), and type `mktables` at the Matlab prompt. With the full package, the kernels are already contained and installation takes only a few seconds. To install for Linux, you should run the full package on a Windows PC (sorry) and copy the whole installation directory to the Linux computer.

In both cases, you run the downloaded file on the Windows PC, which is a self-extracting ZIP archive that automatically starts the setup program. The setup program guides you through the process. An uninstall program is provided that can cleanly remove DeerAnalysis from your computer. Please note that the

DeerAnalysis directory must not be write protected, as the program uses this directory for file exchange with the external program FTIKREG.

4 The user interface

To run the program, start Matlab, change to the directory where it is installed (e.g., by `cd(c:/Program Files/DeerAnalysis2006)`) and call it by typing `DeerAnalysis` at the Matlab prompt (*Tip:* It may be useful to add the path to DeerAnalysis to your default Matlab path in the Matlab startup script `startup.m` by `addpath(c:/Program Files/DeerAnalysis2006)`).

The graphical user interface shown in Fig. 1 opens, of course first without a loaded data set.

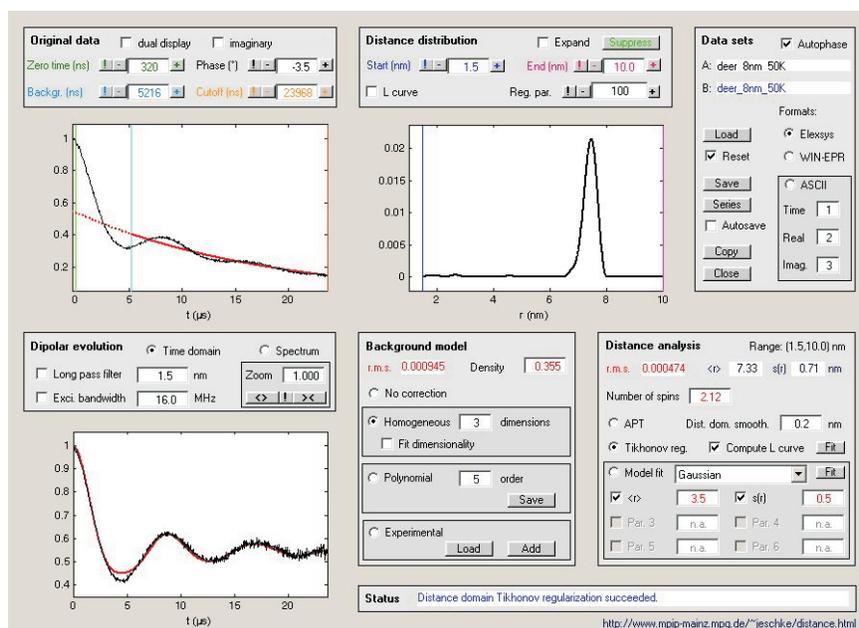


Figure 1: Graphical user interface of DeerAnalysis 2006.

The user interface has been programmed with the following ideas in mind

- no *unnecessary* complexity
- no hidden functionality (no menus)
- default behavior should give reasonable results for most data
- experienced users can easily override default behavior.

Default behavior is to read Elexsys (Xepr) data files, assume that the last three quarters of the data can be used for the background fit, adjust the phase automatically, and correct for exponential background decay (homogeneous spatial distribution of nanoobjects). Initially, no points are cut off at the end of the

data set, the distance distribution is obtained by approximate Pake transformation (APT) [4], and the mean distance $\langle r \rangle$ and standard deviation σ_r by moment analysis within the range from 1.5 to 8 nm using distance domain smmoothing with a filter width of 0.2 nm. A suggestion for cutting off noisy or distorted data points at the end of the data set is made. All this happens automatically once you load a data set via the **Load** button.

Different models for the background can be selected in the **Background models** panel (center of bottom half of Fig. 1) as described in more detail below. Similarly, Tikhonov regularization or fitting of the data by a model distance distribution can be selected in the **Distance analysis** panel. As these approaches are time-consuming, fitting is not started automatically but only after clicking on the corresponding **Fit** button. Adjustable parameters can be edited directly (the most common errors, such as non-digit input or values out of range, are corrected automatically) or incremented or decremented by **+** and **-** buttons, respectively. Several parameters can be adjusted or reset by automatic procedures (described below). This is done with the **!** buttons.

Display in each of the plot windows can be toggled. In the plot below the **Original data** panel display of the imaginary part (magenta trace) can be switched on or off by clicking on the **imaginary** checkbox. If two data sets have been loaded, the real part of the previous set (data set B) can be displayed as a blue trace by clicking on the **dual display** checkbox. This automatically suppresses display of the imaginary part of the active data set (data set A) and changes the **imaginary** checkbox into a **mod. depth scaling** checkbox. Dual display and modulation depth scaling also effect the other two plots, where results corresponding to data set B are also displayed as blue traces.

The plot below the **Dipolar evolution** panel can be toggled between time-domain display and display of the dipolar spectrum by checking the corresponding radiobuttons. Finally, the plot below the **Distance distribution** panel can alternate between display of the distance distribution and the L curve, after a Tikhonov regularization with L curve computation has been performed.

There is no **Help** function, but the controls are provided with short explanations that will show up when you move the cursor above them.

5 Pre-processing

5.1 Loading data

Data input and output is initiated by buttons in the **Data sets** panel. By default the program expects Bruker Elexsys data (binary format). It recognizes automatically if the data are complex (quadrature detection) or real (single-channel detection, discouraged). If the data set is one-dimensional, it is interpreted as output of a (classical) constant-time DEER experiment [1], see also Fig. 2. If the data set is two-dimensional with exactly two traces, it is interpreted as a variable-time DEER experiment [2] with the first trace being the reference trace and the second trace being the recoupled trace. For any other

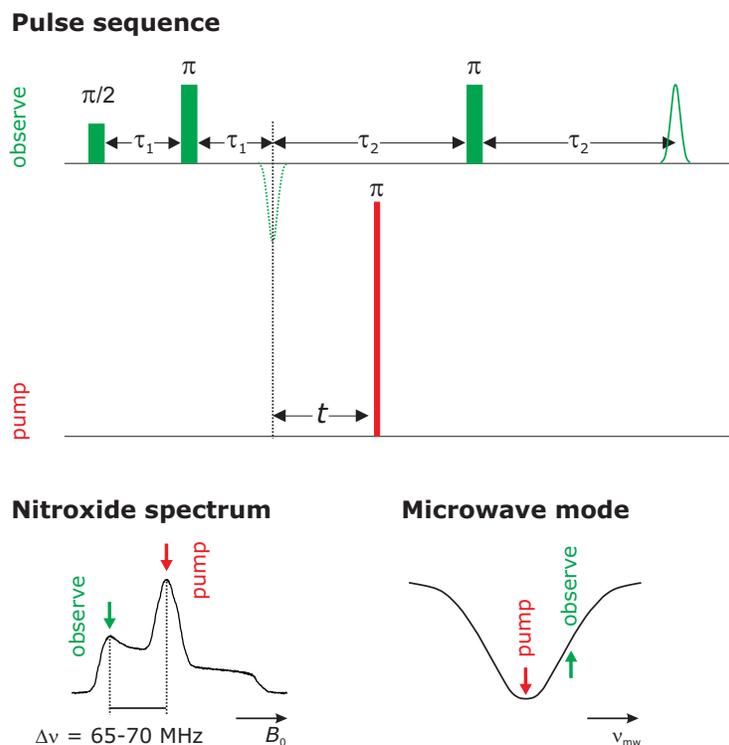


Figure 2: Pulse sequence and positions of the observer and pump frequency with respect to the nitroxide spectrum and to the microwave mode for the four-pulse DEER experiment.

size of experimental data, program response is undefined. If you unintentionally load a data set of some other experiment, it is advisable to close the program and restart it.

Mainly as a support for ESP 380 machines, the program has the capability to read data in WIN-EPR binary format (select by radio button in the **Formats** column of the). As the binary number format of the ESP 380 is somewhat obscure, this mode requires that the data are first read into WIN-EPR on a PC and saved again from WIN-EPR. This mode is less well tested than the Eleksys mode and completely untested for two-dimensional data. Alternatively you can convert ESP 380 data to ASCII data (also possible in WIN-EPR with command sequence `1D processing/Parameters/List data file.../Save`).

From an ASCII file, only one-dimensional data can be read. If there are any header lines before the numerical data, they must start with a percentage character (%). By default, the program expects the time axis (in nanoseconds) in the first column, the real part of the data in the second column, and the imaginary part (if present) in the third column. These assignments can be adapted

in the edit fields below the `ASCII` radio button. For ASCII data exported from WIN-EPR, the proper settings are 2, 3, and 4 instead of 1, 2, and 3. The first six lines (header lines) have to be deleted or commented out by a `%` character. The program automatically recognizes if there is no imaginary part. After successfully loading data, the `Status` panel shows a short characterization of the data set (const-time/variable-time DEER, complex/real, number of data points). The filename is included in the title of the DeerAnalysis main window and is also shown in line `A:` of the `Data sets` panel.

5.2 Determining zero time

The time origin of the dipolar evolution function corresponds to $\tau_1 = \tau_2$ (see Fig. 2). Because pulse lengths are finite, the relation between this equation and actual delays in the pulse sequence may not be trivial. We therefore suggest determination of the time origin (zero time) from experimental data with a good signal-to-noise ratio (SNR) for the pulse lengths and τ_1 delay that you actually use. If you later measure on the same spectrometer with the *same* pulse lengths and τ_1 you can use the same value. Knowing this value is important for data with poor SNR where automatic determination is likely to fail. Automatic determination of zero time t_0 is based on the expectation that the real part of the signal should be symmetric about the time origin. For the proper choice of t_0 , the first moment of the signal in a range symmetric about t_0 should thus be zero. In a first step, the program approximates zero time by the time t_{\max} at which the real part is maximum. Then the first moment is determined in a window $t_x t_{\max}/2$, where t_x is shifted through the whole data set. The optimum value of t_0 is the time t_x where the first moment is minimum. This algorithm should work well for good SNR and distances up to ≈ 5 nm. If it fails under such conditions, τ_1 is too short (expected symmetry of the data is spoiled by interference between adjacent microwave pulses). The algorithm may fail for very long distances where data close to the maximum are pretty flat. For such long distances small mis-settings have only minor influence on the distance distribution.

You may correct the automatically determined zero time by the `+` and `-` buttons right and left from the value or by direct input of a new value in the edit field (fit by the eyes). A wrong choice may be easier to detect when you switch the `Dipolar evolution` plot to frequency domain (`spectrum`).

5.3 Phase correction

In a properly adjusted DEER experiment, the signal should be entirely in the real part of the data set. If receiver offsets are canceled by `[(+x)-(-x)]` phase cycling of the first pulse, as we strongly suggest, the imaginary part is zero. It is therefore tempting to acquire and process only the real part. We discourage this. For very weak signals, as you occasionally encounter with membrane proteins, it is difficult to adjust signal phase exactly during setup. Consequently, part of the signal will be in the imaginary part. Furthermore, depending on stability

of your spectrometer, there may be small phase drifts during the experiment. It is better to correct for these drifts than to ignore them. Finally, unexpected artifact signals are likely to manifest in the imaginary part (see Fig. 3). If the imaginary part *after* phase correction strongly deviates from zero at early times, it is advisable to acquire data with a longer τ_1 value (see Fig. 2).

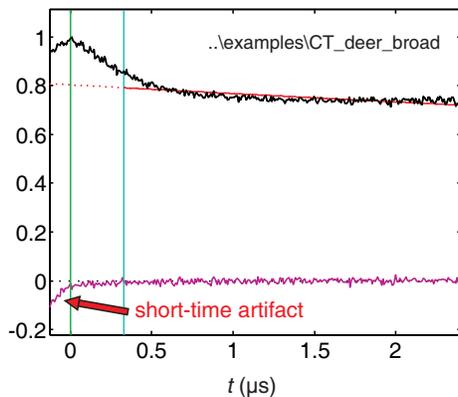


Figure 3: Imaginary-part artifact at early times (see red arrow) due to mw pulse interference. Interpulse delay τ_1 should be long enough for the artifact to have almost completely decayed at $t = 0$ (green vertical line).

Automatic phase correction can be based on the expectation that the imaginary part should be zero at sufficiently long times. By default, the program determines the corresponding phase correction directly after loading complex data by minimizing the root mean square deviation of the imaginary part for the last three quarters of the data (part between the blue and orange cursors). The phase correction in degree is displayed in the **Original data** panel.

You may correct phase manually by using the + and - buttons right and left from the value or by direct input of a new value in the edit field. Phase is automatically restricted to the range $(-180, +180)^\circ$. If you did not phase cycle and do have a receiver offset, you may aim to flatten the imaginary part and put all modulation into the real part. Note however, that in this case you are likely to have a receiver offset in the real part, too. This will be detrimental to data analysis. Automatic phase correction can be reactivated by the ! button left from the value. It will always relate to the part of the data between the blue and orange cursors. If you move any of these cursors, the result may differ from the result that you got directly after loading.

Automatic phase correction after loading can be deactivated by unselecting the check box **Autophase** in the **Data sets** panel.

5.4 Cutting data

For several reasons, you may want to exclude points at the end of your data set from analysis. First, some people prefer to acquire data up to delays t , where the pump pulse starts to interfere with the last observer pulse or even overlaps with it. In this case, the last data points are spoiled. Second, if at maximum t the signal has decayed to a very small value (say 0.1 times maximum intensity), the dipolar evolution function after background correction will be rather noisy, as correction involves division by the background decay. Third, SNR in variable-time DEER data increases with t even before background correction. It may be wise to cut the data at a time where noise is still tolerable.

By default no data points are cut off at the end, but a suggestion for cutoff is displayed as an orange vertical cursor in the `Dipolar evolution` plot (see Fig. 4). This suggestion is derived from the difference D between the experimental dipolar evolution function and its fit by the APT result. The mean square deviation M_k of eleven consecutive points $D_{k-5} \dots D_{k+5}$ around the k th data point is computed for all indices k . The minimum of M is a measure for the noise level. An acceptable noise level of $6\min(M)$ is assumed. The program then searches for a range of consecutive points at the end of the data set that *all* fulfil the condition $M(k) > 6\min(m)$. If such a range of points exists, the program suggests to cut it off. Otherwise the orange cutoff cursor is set to the end (right border) of the trace.

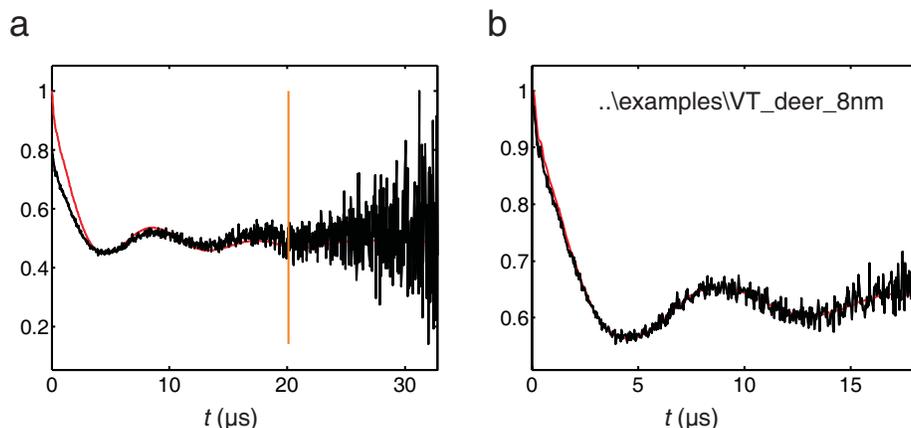


Figure 4: Cutting off the noisy part at the end of variable-time DEER data. (a) `Dipolar evolution` plot for the whole data set. The orange cursor shows the suggested cutoff time. (b) Dipolar evolution function (black) and fit (red) by a distance distribution obtained with APT after cutting the data at the suggested time.

The suggestion can be accepted by clicking on the `!` button of the `Cutoff` controls in the `Original data` panel. Note that this may in turn improve

the fit, thus leading to a smaller value $\min(M)$ and a new cutoff suggestion. Therefore, it is advisable to click on the ! button several time to iteratively approach the optimum cutoff. Furthermore, the cutoff suggestion depends on correct settings of other parameters (zero time, phase). For instance, for the variable-time DEER data set shown in Fig. 4 the zero time must be zero, while the program automatically determines 96 ns. If this is not corrected, not good fit is obtained and unnecessarily many data points are cut off. Note also that this data set with relatively poor SNR was intentionally selected for explanation of data cutoff. For many data sets, no cutoff at all may be required and DeerAnalysis2006 immediately sets the cutoff cursor to the right border.

Generally, cutting off a significant amount of data will suppress noise but will also cause a suppression of long distances by background correction. Proper background correction may become more difficult.

5.5 Background correction

In most cases, EPR distance measurements are performed to elucidate the structure of a nanoscopic object. Only distances within this object are of interest. The contribution of distances to neighboring objects should be suppressed. If you think about a biradical or bilabelled protein molecule, you want to measure the intramolecular distance and suppress contributions from intermolecular distances.

Such a separation of the signal $V(t) = \{1 - [1 - \Delta D(t)]\}B(t)$ into a dipolar evolution function $D(t)$ for the nanoobject itself and a background decay $B(t)$ due to neighboring objects requires a criterion for distinguishing the two contributions. Furthermore, the functional form of the background decay has to be known. This functional form is related to the spatial distribution of the nanoobjects. A separation can only be successful if distances within the object are typically shorter than distances to neighboring objects. The wanted contribution is then confined to the earlier part of the time domain data, while later parts are dominated by the background decay. The decay can only be fitted properly if the maximum time t in the pulse sequence (Fig. 2) is significantly longer than the time at which the dipolar modulation has decayed. A more detailed discussion can be found in Ref. [3].

Separation into the two contributions is simple and reliable if the distance distribution is dominated by distances shorter than 4 nm. In protein samples, it becomes challenging for distances between 4 and 6 nm, and near to impossible for distances longer than 6 nm, unless protons around the spin labels can be strongly diluted by deuteration [2]. Note that one can still get a quite reliable estimate of a distance of closest approach if separation fails. However, the width and shape of the distance distribution should not be discussed in such a situation.

In simple cases (short distances and homogeneous distribution of the nanoobjects in three dimensions), separation depends only weakly on the choice of parameters. Default behavior of the program should then be sufficiently good. By default, an exponential background decay corresponding to a homogeneous

three-dimensional distribution is fit to the last three quarters of the data. The fit parameter is the decay time constant, which is proportional to the concentration of nanoobjects. With proper calibration such fits can be used to determine local concentrations (see Section 5.6).

Generally, the background is shown as a red line in the **Original data** plot. A continuous line is plotted in the range where the background was fitted (between the blue and orange cursors), a dotted line is plotted where the fit was extrapolated. The r.m.s. value of the background fit is displayed in the **Background model** panel.

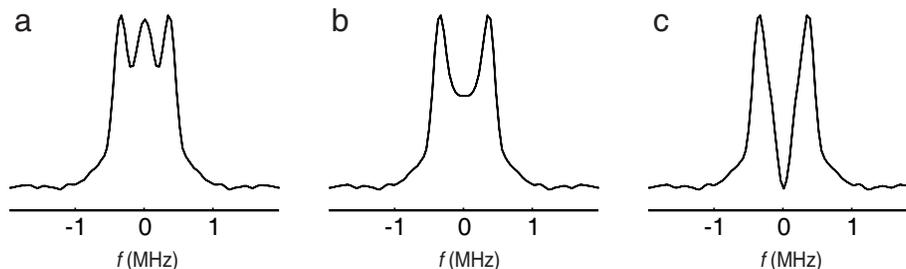


Figure 5: Manifestation of different background fits in the dipolar spectrum (example data set CT_DEER_5nm). (a) Part of the background is attributed to the biradical. (b) Good separation of intra- and intermolecular contributions, as obtained with automatic correction (! button). (c) Part of the biradical contribution is attributed to the background.

For distances of ≈ 4 nm and longer, choice of the time range for background fitting may decide whether you obtain artifacts in the distance distribution at long distances. Unlike the other problems in determining a distance distribution, this problem is most severe for narrow distributions of distances. In this case the modulation decays more slowly and thus interferes more strongly with the background fit. Our automatic determination of the optimum fit range is based on the assumption that the longest detectable distance exceeds the largest distance within the nanoobject. If this condition is met, the distance distribution after correct background correction is zero at the maximum detectable distance. This can be checked by approximate Pake transformation (APT, see below). APT is sufficiently fast to be applied at all possible choices of the starting time for the background fit. For any selected background model, this search for the optimum starting time can be initiated by clicking on the blue ! button in the **Original data** panel. Depending on the length of your data set and the speed of your computer, this optimization can take up to a few minutes.

The starting time for background fitting can also be adjusted manually with the blue + and - buttons or by direct input into the edit field. The consequences can best be judged when switching the bottom left plot below to frequency domain. For a narrow distance distribution, the black trace should look like a Pake pattern. Deviations are best seen at zero frequency. There should be

neither a positive spike nor an obvious hole in the center of the Pake pattern (see Fig. 5).

Background correction can be switched off completely by selecting the **No correction** radiobutton in the **Background model** panel. In this modus the input data are interpreted as a dipolar evolution function which is already separated from background. The modus is intended for compatibility with external pre-processing programs, for polynomial fitting of single-label data to derive an experimental background function (see below), or for fitting by a user model that explicitly contains the background contribution. User models consisting of a single Gaussian peak with 3D homogeneous background (**Gaussian_hom**) or of two Gaussian peaks with 3D homogeneous background (**Two_Gaussians_hom**) are already included in DeerAnalysis2006. However, we strongly discourage fitting background and distance distribution simultaneously, as such fits are very likely to end up in local minima of the error hypersurface. Whenever a separation of the background contribution from the contribution of the nanoobject can be performed with some confidence, it should be done before analysis of the distance distribution.

In the following we shortly discuss the possible choices for the spatial distribution of nanoobjects. They can be selected by checking the corresponding radiobutton in the **Background model** panel.

5.5.1 Homogeneous

This model is strongly suggested for all cases where you do not have experimental background functions from singly labelled molecules. The general background function in this model is

$$B(t) = \exp\left(-kt^{d/3}\right) \quad (2)$$

where k quantifies the density of the spins and d is the dimensionality of the homogeneous distribution. Unless there is a confinement on length scales below 10 nm, the distribution is homogeneous in $d = 3$ dimensions. This case applies to most solutions. Membrane proteins in a liposome may be confined to $d = 2$ dimensions. If possible, such confinement should be established by control measurements on singly labelled proteins, for which $d = 2$ is expected give a better fit than $d = 3$. For labels attached to a stretched polymer chain, $d = 1$ may be appropriate. Note also that a choice of $d = 6$ corresponds to a Gaussian background decay, as it has been observed with the single-frequency SIFTER experiment [7]. The dimension is not necessarily an integer number- if experimental data of a singly labelled sample can be nicely fitted with a fractal dimension, it is advisable to use the same fractal dimension for background correction of the corresponding doubly labelled sample.

When the **Fit dimensionality** checkbox is selected, both k and d are fitted. This mode is suggested only for determining the fractal dimension of purely homogeneous (singly-labelled) samples. In this case the **Bckg.** control in the **Original data** panel should be set to zero (green and blue cursors coincide), as the early decay of the data is most sensitive to the parameter d .

5.5.2 Polynomial

Short distances are underrepresented in the intermolecular distance distribution, if the spin labels are attached to nanoobjects that cannot penetrate each other. As a result, the intermolecular contribution decays more slowly at early times than would be expected for a homogeneous distribution. If singly labelled objects are available, the intermolecular part can be measured separately and an experimental background function can be derived. Directly using the noisy experimental data set of the singly labelled sample would introduce significant statistical errors. It is therefore prudent to use a smooth fit function for that purpose.

Almost any intermolecular decay can be reproduced by fitting a polynomial to the *logarithm* of the original data. DeerAnalysis2006 allows for polynomials with an order of up to 15, but note that the lowest order should be selected that still gives a good fit (flat trace in the **Dipolar evolution** plot. Polynomial fits are mainly implemented for deriving and afterwards saving experimental background functions from singly labelled samples, not for direct background correction.

5.5.3 Experimental

Once experimental background functions have been derived from singly labelled samples, they can be used for correcting the background in corresponding doubly labelled samples. In this mode, the relative magnitudes of the polynomial coefficients are kept fixed. The background model is given by

$$B(t) = \exp\left(-k \sum_{n=0}^o c_n t^n\right) \quad (3)$$

where k is the density (concentration) parameter, o the order of the polynomial, and the c_n are the polynomial coefficients determined previously on the singly labelled samples. The only fit parameter is k .

In principle, background data should be individually measured for both label positions in a doubly labelled sample, as the suppression of short distances depends on how deep the label is buried in the nanoobject. The weighted sum of both background functions is a better approximation for the actual background in the doubly labelled sample than each individual background function. Several background polynomials can be added using the **Add** button in the **Background model** panel. A weighting factor can be specified in a dialog box that opens after clicking on this button. Note that the different labeling efficiencies at the two positions are already accounted for with weighting factor 1.0 if both singly labelled samples were measured with the same protein concentration.

5.6 Determining local concentrations

The parameters of the background fit are related to the number of coupled spins within the nanoobject (modulation depth after background correction) and to

the density of nanoobjects (parameter k). For calculation of the number of spins and of absolute densities, the modulation depth parameter λ has to be known, which depends strongly on the excitation position, length, and flip angle of the pump pulse and weakly on line broadening in the nitroxide spectrum and shape of the resonator mode. Reliable quantification therefore requires a calibration with known samples and proper adjustment of the flip angle of the pump pulse (see Section 10). The calibration should be repeated if the resonator or the length of the pump pulse is changed. Protonated and deuterated nitroxide spin labels also require separate calibrations. Determination of the number of coupled spins is more reliable when based on Tikhonov regularization or a fit of the data by a model distribution and is therefore discussed later on (Section 7.3).

For a 3D homogeneous distribution of objects, the density is proportional to the local concentration. The term local refers to the length scale of the DEER experiment, which extends to approximately 1020 nm for the background. Measurements of local concentrations can be calibrated with a solution of an appropriate spin label (e.g., protonated or deuterated TEMPOL) in toluene. An example data set from our own calibration (CT_DEER_tempol_2500uM) is provided. This data set was acquired with a 2mM TEMPOL solution in toluene, which corresponds to a concentration of 2.5 mM at 80 K, as toluene shrinks to approximately 80% of its room temperature volume when freeze-quenched in liquid nitrogen.

To calibrate 3D background fitting for determination of concentrations, select **Homogeneous** as the background model, set **dimensions** to 3, and load a data set for a sample with known concentration. Adjust zero time and phase, if necessary. Now input the concentration (in the units you prefer) into the edit field **Density**. The color of the density value then changes to green. When you now load other experimental data sets that have been measured with the same resonator and experimental settings and use the same background model, you can directly read off concentrations from the edit field **Density**. Note that the program loses calibration on restart.

5.7 Long-pass filtering

The major artifact contribution to DEER time-domain signals is usually nuclear modulation due to matrix protons. At X-band frequencies, such proton modulation corresponds to a distance of approximately 1.5 nm. By restricting the distance range for analysis to (1.75, 8) nm, contributions by nuclear modulation can be suppressed. However, as computation of distance distribution is an ill-posed problem, an out-of-range artifact may still influence the result within the range of interest. Very strong proton modulations, as they are sometimes encountered for membrane proteins in liposomes or detergent micelles, should thus be eliminated by filtering.

This can be achieved by completely eliminating contributions above a certain maximum frequency, which roughly corresponds to suppressing distances below a certain minimum distance. Such complete suppression was described

in Ref. [2]. For broad distance distributions with contributions both below and above 1.75 nm, complete suppression may introduce an artificial hole at $t = 0$ into the time-domain data and may thus replace the nuclear modulation artifact with a suppression artifact. To avoid this, filtering in DeerAnalysis2006 is performed by fitting a third-order polynomials to the real and imaginary parts of the frequency-domain data between the cut-off frequency and the Nyquist frequency. The frequency-domain data in this range are then replaced by the polynomial. This suppresses the sharp nuclear modulation peak as well as high-frequency noise, while keeping the high frequency contributions of broad distance distributions intact.

Filtering is enabled by selecting the `Long pass filter` checkbox in the `Dipolar evolution` panel. The cut-off distance (lower limit, default 1.6 nm) can be changed in the edit field right from this check box. When working with broad distributions of short distances, the default value is often a good compromise between residual proton modulation and partial suppression of short distances.

6 Extracting distance distributions

6.1 General remarks

The computation of a distance distribution $P(r)$ from a dipolar evolution function $V(t)$ is an ill-posed problem. For such problems, small variations in the input data (e.g., noise) can cause large variations in the output data. In other words, significantly different distance distributions may correspond to very similar dipolar evolution functions. Data analysis therefore depends strongly on striking a good compromise between improving resolution and decreasing the influence of experimental noise. First and foremost, data should be acquired with as good as possible SNR. Reproducing results for a given sample is usually a good idea. Second, ill-posedness must be taken into account in data analysis. There are several ways of doing this, which all have one feature in common: one tries to find a resolution in distance domain at which a good fit of the experimental data is obtained without introducing strong noise artifacts into the distance distribution.

6.2 Approximate Pake Transformation (APT)

A very fast algorithm relies on an approximate integral transformation to dipolar frequency domain, subsequent correction of cross-talk artifacts, and mapping to distance domain (APT) [4]. Ill-posedness is moderated by proper discretization in dipolar frequency domain. If SNR is too small, the distance distribution may still be influence by strong noise artifacts. A better compromise between reliability of the distribution and resolution can then be achieved by distance-domain smoothing, i.e., by giving up resolution in favor for a smoother distribution. As APT is very fast, it can also be used to generate starting values for

fit procedures. The disadvantage of APT with respect to other techniques is that it cannot incorporate the constraint $P(r) > 0$ (for all r). This disadvantage, however, is significant, as the constraint strongly stabilizes the solution. For this reason, two other approaches for data analysis are incorporated into DeerAnalysis2006.

6.3 Tikhonov regularization

Other approaches rely on computation of a simulated time-domain signal $S(t)$ from a given distance distribution $P(r)$ by

$$S(t) = K(t, r) P(r), \quad (4)$$

where K is the kernel function. For the DEER experiment with ideal pulses, the kernel function is known analytically

$$K(t, r) = \int_0^1 \cos [(3x^2 - 1) \omega_{\text{dd}} t] dx, \quad (5)$$

with

$$\omega_{\text{dd}}(r) = \frac{2\pi \cdot 52.04 \text{ MHz nm}^{-3}}{r^3}. \quad (6)$$

The case of non-ideal pulses is discussed in Section 6.5.

The most elegant response to ill-posedness is Tikhonov regularization. In this approach, the compromise between smoothness (artifact suppression) and resolution of the distance distribution is quantified by a regularization parameter α . The optimum distance distribution $P(r)$ is found by minimizing the objective function

$$G_\alpha(P) = \|S(t) - D(t)\|^2 + \alpha \left\| \frac{d^2}{dr^2} P(r) \right\|^2 \quad (7)$$

for a given α . The first term on the right hand side of eqn (7) is the mean square deviation between the simulated and experimental dipolar evolution function while the second term is the regularization-parameter weighted square norm of the second derivative of $P(r)$, which is a measure for the smoothness of $P(r)$. The larger α the less noise artifacts are introduced. However, a larger α also causes a stronger broadening of peaks in the distance distribution. Therefore, small α are required for samples with well defined distances (narrow peaks) and large α for very broad distributions, which otherwise disintegrate into many narrow peaks. Unfortunately, the correct width of the peaks is often not known in advance.

There are different ways for mathematically defining an optimum regularization parameter. The past version DeerAnalysis2004 used the self-consistency criterion [14, 15]. However, determination of an optimum α is itself influenced by noise [3], and the self-consistency criterion appears to be more sensitive to noise distortions than the L curve criterion [10]. The L curve is a plot of $\log \eta(\alpha)$ versus $\log \rho(\alpha)$, where

$$\rho(\alpha) = \|S(t) - D(t)\|_\alpha^2 \quad (8)$$

quantifies the means square deviation and

$$\eta(\alpha) = \left\| \frac{d^2}{dr^2} P(r) \right\|_{\alpha}^2 \quad (9)$$

the smoothness. For well behaved data (good signal-to-noise ratio, relatively narrow peaks in the distribution), this plot is L-shaped as is illustrated in Fig. 6a. In the range of small regularization parameters α (left of the corner, undersmoothing) the slope is steep and negative, as increasing α and thus the smoothing strongly decreases the norm of the second derivative of $P(r)$ without strongly affecting the mean square deviation. In contrast, right of the corner (oversmoothing) the mean square deviation increases strongly with increasing α as the simulation is no longer a good fit of the data. At the same time, η decreases only gradually as noise-related spikes in $P(r)$ are already smoothed out. If the SNR is worse and the peaks in the distance distribution are broader, the corner of the L curve is somewhat less pronounced (Fig. 6b).

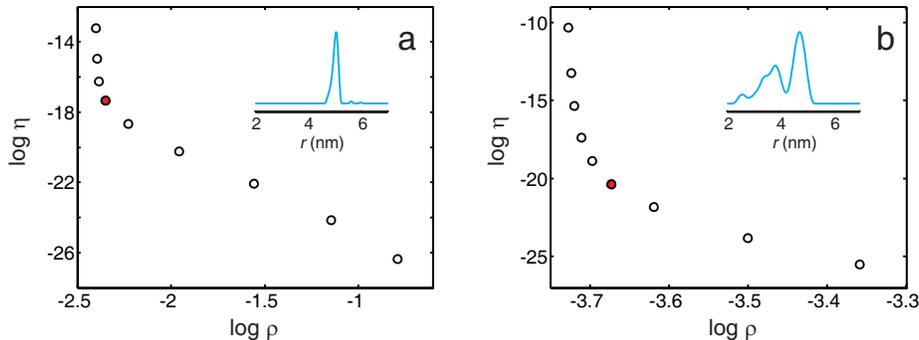


Figure 6: Tikhonov L curves. The red data points correspond to the optimum regularization parameter. The insets show the distance distribution obtained with this parameter. a) Data set dOTP_5nm, $\alpha = 1$. b) Data set CT_DEER_broad, $\alpha = 100$.

The computationally most efficient implementation of the L curve criterion does not allow for additionally introducing the constraint $P(r) > 0$. As this constraint strongly stabilizes the solution, DeerAnalysis2006 relies on the Fortran program FTIKREG, written by J. Weese and distributed by the Materials Research Center Freiburg, which allows for using it. The L curve criterion is then implemented by computing Tikhonov regularization for a pre-defined set of regularization parameters

$$\vec{\alpha} = (0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000, 100000) . \quad (10)$$

Our experience suggests that this set is sufficient for all cases of practical interest. If required, Tikhonov regularization can also be performed for intermediate

values or values that are smaller or larger than the limits of this set. After an L curve has been computed, the distance distribution and simulated dipolar evolution function can be inspected for all values of α , which is helpful in cases where this curve does not exhibit such a clear corner as in Fig. 6. In such cases, automatic recognition of the corner may fail.

Tikhonov regularization is performed by clicking on the corresponding **Fit** button in the **Distance analysis** panel. By default, L curve computation is disabled, as it is time consuming. The regularization parameter (default: 1) can be changed in the corresponding edit field in the **Distance distribution** panel. The distance range for Tikhonov regularization is determined by the blue and magenta start and end values in the **Distance distribution** panel, which can also be edited. Computation of the L curve can be requested by clicking on the **Compute L curve** checkbox in the **Distance analysis** panel and subsequently clicking on the **Fit** button. After such a computation, the L curve is automatically displayed instead of the **Distance distribution** plot with the automatically derived selection of the corner highlighted in red and the corresponding regularization parameter shown in the **Reg. par.** control. The selection of the corner can be shifted with the + and - buttons of the **Reg. par.** controls. Such changes update the fit in the **Dipolar evolution** panel and the r.m.s. value in the **Distance analysis** panel. The distance distributions for different regularization parameters can be inspected in the same way after unselecting the **L curve** checkbox in the **Distance distribution** panel.

6.4 User models

Generally, the solution of an ill-posed problem can be stabilized by introducing additional constraints. A distance distribution $P(r)$ that conforms to a simple model with only a few parameters, for example a distribution consisting of one or two Gaussian peaks, is strongly constrained. Fitting of the data by a model distribution can thus improve reliability of the analysis. Furthermore, by comparing the parameters for a series of related samples trends can be easily recognized. This approach is offered in DeerAnalysis2006 by an interface for fitting pre-processed data by user-defined models for the distance distribution $P(r)$. Model functions with one and two Gaussian peaks are already implemented. The model library can be extended by the user as described below.

In applying this approach one should be aware that a model can impose constraints that do not apply to the true distance distribution and may thus suppress information contained in the original data. For instance, the example data set **dOTP_5nm** can be fitted relatively well by a distance distribution consisting of a single Gaussian peak, but this imposes a symmetry on the peak that is not a feature of the true distribution. The true distribution decays more steeply towards high distances than towards low distances as seen in the inset in Fig. 6a (and the reason for this asymmetry is well understood). It is thus advisable to perform a model-independent analysis by Tikhonov regularization first. From a set of distance distributions for the same class of samples, it is then often possible to derive a model function that does not impose undue constraints but

does make use of additional information on the sample that comes from other characterization techniques.

6.4.1 Fitting with existing models

When DeerAnalysis2006 starts, the program checks the subdirectory `models` for existing Matlab scripts (extension `.m`). The current distribution contains the scripts

- `Gaussian.m`
- `Gaussian_hom.m`
- `Two_Gaussians.m`
- `Two_Gaussians_hom.m`.

These models, and any models implemented by the user, are included in the `model fit` pop-up menu of the `Distance analysis` panel. On selecting an entry of this menu, the parameter definitions, default values and limits of the corresponding model are read and the parameter controls in the `model fit` subpanel are updated. A model can have up to six parameters. If it has less, superfluous parameter controls are disabled.

Before fitting, select the `model fit` radiobutton in the `Data analysis` panel. The `Distance distribution` plot now shows the APT result as a black dotted narrow line and the distance distribution corresponding to the current model and parameter values as a red dotted bold line. The `Dipolar evolution` plot displays the experimental data (black line) and the data simulated with the current model (red dotted line). You may now edit the starting values of the fit parameters in the `model fit` subpanel until you obtain a reasonable agreement between experimental and simulated data. Of course, this step can be skipped and fitting can be started immediately, but by first improving your starting values you decrease the probability to get stuck in a local minimum of the error hypersurface. Before fitting you can also decide whether you want to fit all parameters (default behavior) or whether you want to keep some parameters fixed at their starting values. To fix a parameter, unselect the corresponding checkbox.

Fitting is started by clicking on the `Fit` button in the `model fit` subpanel. During fitting, the `Status` panel displays the current r.m.s. value. Note that fitting can be rather slow if the excitation bandwidth correction (see Section 6.5) is switched on. After the fit is completed, the parameter values are updated, the `Distance distribution` plot shows the fitted distance distribution as a black bold line, and the `Dipolar evolution` plot displays the experimental data (black line) and the fit (red line).

Model fitting considers the distance distribution in the range between 1 and 10 nm. For data sets extending to times longer than $4 \mu\text{s}$, an upper limit of 10 nm may be too short if the homogeneous background is also fitted. As mentioned earlier, we strongly suggest to remove the background contribution before fitting.

6.4.2 Implementing a new model

The interface between DeerAnalysis2006 and the model scripts was designed to allow for writing model scripts without knowledge on the inner working of the main program. A model script has two input variables, a vector of distances `r0` at which values of the distance distribution have to be computed and a vector of parameters `par`. The only output parameter is the distance distribution, which is a vector of the same length as `r0`.

Note that the integral of the distance distribution can be arbitrary, as DeerAnalysis2006 internally renormalizes the distribution to an integral of 0.01 for simulations and later computes the number of coupled spins from the modulation depth of the experimental data. This means that no amplitude parameter is needed. Only if the distribution corresponds of more than one contribution (for instance two Gaussian peaks), a parameter for the relative amplitude of an additional component with respect to the first component has to be defined. Consequently, a Gaussian distribution is defined by only two parameters, the mean distance $\langle r \rangle$ and the width (standard deviation) σ_r . A distribution consisting of two Gaussian peaks thus has the parameters $\langle r(1) \rangle$, $\sigma_{r(1)}$, the relative contribution of the first peak $p(1)$, and $\langle r(2) \rangle$, $\sigma_{r(2)}$. It is convenient to define the relative contributions so that they relate to the integral of the peaks (number of spins) and that $p(1) + p(2) = 1$. The model script `Two_Gaussians.m` is written this way.

A model script needs to declare its parameters to DeerAnalysis2006 and provide default values as well as lower and upper limits for them. This is done in a comment section. As an example consider the script `Gaussian.m`:

```
function distr=Gaussian(r0,par),
%
% Model library of DeerAnalysis2006: Gaussian
%
% single Gaussian peak with mean distance <r> and width (standard
% deviation) s(r)
% (c) G. Jeschke, 2006
%
% PARAMETERS
% name symbol default lower bound upper bound
% par(1) <r> 3.5 1.5 10
% par(2) s(r) 0.5 0.05 5
gauss0=(r0-par(1)*ones(size(r0)))/par(2); distr=exp(-gauss0.^2);
```

The first line is the function declaration, which is the same for all user models except for the function name (here `Gaussian`). The following lines, which start with the `%` character, are all comment lines, as far as Matlab is concerned. However, when the model is selected, DeerAnalysis2006 scans these comment lines in the source file for parameter declarations. A parameter declaration line begins with the `%` character, followed by at least one space and the parameter name. Valid parameter names are `par(1)`, `par(2)`, `par(3)`, `par(4)`, `par(5)`,

and `par(6)`. Only as many parameters have to be declared as are needed for the model (here 2). The parameter name is followed by at least one space and then by the parameter symbol. The symbol consists of at least one non-space character. It is shown as identification of the parameter control in the `model fit` subpanel. A symbol of up to five non-space characters can always be displayed, longer symbols are completely displayed only if some of the characters are narrow. The symbol is followed by at least one space and then the default value of this parameter. The default value is displayed in the edit field of this parameter and is the starting value for the fit if the user does not make any input before clicking on the `Fit` button. A good set of starting values provide for a distribution that is mainly confined between 1.5 and 8 nm and that clearly exhibits all relevant features of the model. The default value is followed by at least one space and the lower limit. No input smaller than this value is accepted by the edit field. Likewise, the value is used as a lower boundary in parameter fitting. The lower limit is followed by at least one space and the upper limit, which is analogous to the lower limit. Note that definition of the default values and limits is *mandatory*. Program response is undefined if the parameter line is incomplete.

6.5 Accounting for limited excitation bandwidth

Analysis of DEER distance measurements is usually based on analytical expressions, such as eqn (5), that assume ideal pulses. Past versions of our analysis programs accounted for this by suggestion a lower limit of 1.75 nm for the reliability of the distribution. Maryasov and Tsvetkov [11] first suggested to use corrected expressions to get more reliable results for short distances. Their approach considered the full Hamiltonian during the pulse, except for the pseudosecular contribution of the dipole-dipole coupling. They still assumed that the observed spins are not excited by the pump pulse and the pumped spins are not excited by the observer pulse. With these remaining assumptions, which are however not very well fulfilled, they could still obtain analytical expressions for the three-pulse DEER experiment. Based on these expressions, the effect of finite pulse lengths on determining distance distributions was assessed in a later contribution by Milov et al. [12].

To relax the remaining assumptions and extend the approach to four-pulse DEER, we examined the dependence of the modulation depth λ on the dipolar frequency ω_{dd} for typical lengths of the observer and pump pulses. Numerical density matrix computations of the full pulse sequence were performed for this purpose. Details will be published elsewhere. The dependence of λ on ω_{dd} can be approximated quite nicely by a Gaussian function

$$\lambda(\omega_{dd}) = \exp\left(-\frac{\omega_{dd}^2}{\Delta\omega^2}\right), \quad (11)$$

where $\Delta\omega$ is an effective excitation bandwidth with respect to dipolar frequencies. For a four-pulse DEER experiments with a pulse length of 24 ns for all

pump and observer pulses and for an experiment with a 12 ns pump pulse and 32 ns observer pulses, we find the same excitation bandwidth of 16 MHz. For a four-pulse DEER experiments with a pulse length of 24 ns for all pump and observer pulses the excitation bandwidth is 12 MHz. The expression in eqn (11) can be used as a correction of the kernel function, eqn (5):

$$K(t, r; \Delta\omega) = \int_0^1 \exp\left(-\frac{\omega_{dd}^2}{\Delta\omega^2}\right) \cos[(3x^2 - 1)\omega_{dd}t] dx, \quad (12)$$

so that effects of finite pulses length can be accounted for without much additional computational effort if the kernel is anyway computed during fitting, as DeerAnalysis2006 does it during Tikhonov regularization.

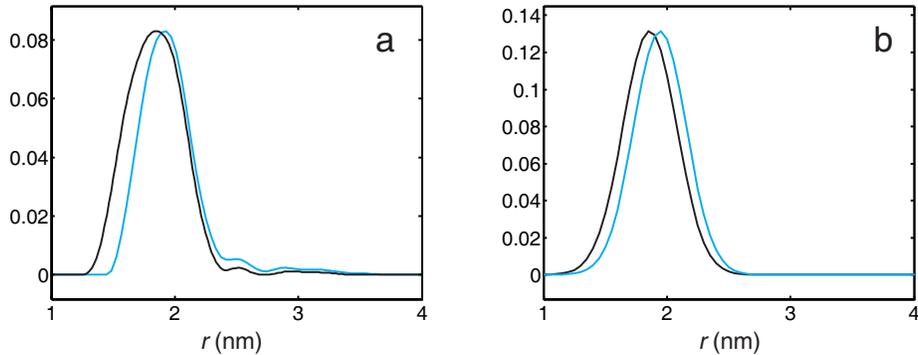


Figure 7: Excitation bandwidth correction. Blue distance distributions were obtained without, black ones with correction. a) Tikhonov regularization with optimum regularization parameter $\alpha = 1$. b) Fit by a single Gaussian peak.

However, simulations of the dipolar evolution function from a distance distribution, as they are required in model fits or at the end of Tikhonov regularization, can be performed with a pre-computed kernel for the expression given by eqn (5), while the kernel must be computed "on-the-fly" for the expression given by eqn (12). This is because the latter expression depends on an additional variable parameter $\Delta\omega$ and, furthermore, does not allow for scaling. In the former expression, a scaling of the t axis by a factor x can be compensated by scaling of the distance axis by a factor $x^{1/3}$. Without bandwidth correction, DeerAnalysis2006 uses fast computations with a pre-computed ideal kernel. Therefore, bandwidth correction considerably slows down simulations and model fits and is thus not selected as default behavior of the program. It can be activated by selecting the `Exci. bandwidth` checkbox in the `Dipolar evolution` panel.

The effect of excitation bandwidth correction is illustrated in Fig. 7 for data set `deer_bi_oligo_n8_50K` from the `calibdepth` subdirectory. Data were cut off at 1504 ns to improve the background fit. Without correction (blue distributions) distances below 1.75 nm are strongly suppressed. With correction they

are recovered. The r.m.s. deviation improves from 0.000320 without correction to 0.000286 with correction in Tikhonov regularization and from 0.000396 without correction to 0.000335 with correction for a Gaussian fit. An improvement in the r.m.s. value may not always be found. The mean distance obtained with Tikhonov regularization changes from 1.97 to 1.85 nm. For a slightly longer flexible biradical (data set `deer_bi_oligo_n10_50K`), the correction is somewhat smaller, as the mean distance changes from 2.07 to 1.98 nm (data not shown). Note also that the Gaussian fits do not account very nicely for the true shape of the distribution in this case.

7 Post-processing

For many cases, one wants to quantify the distance distribution in terms of a few numbers, i.e., mean distance and width of the whole distribution or of individual peaks. For oligomers of membrane proteins and self-assembled supramolecular systems, it may also be of interest to derive the number of spins within an individual nanoobject. All these values can be obtained by post-processing.

7.1 Moment analysis and peak picking

Analysis of a number of simulated and experimental DEER data sets suggested that the first moment (mean distance) and second moment (variance, square of the standard deviation) of the distance distribution are stable parameters. In other words, these values are only very slightly influenced by noise-induced artificial splittings in the distance distribution. This applies in particular to the results of those techniques that incorporate the constraint $P(r) > 0$ (Tikhonov regularization and model fitting). Moment analysis of the distance distribution in the range of interest (default: 1.58 nm) is therefore performed automatically. The mean distance ($\langle r \rangle$) and standard deviation ($s(r)$) are displayed in the **Distance analysis** panel. To exclude obvious artifacts at the short or long end of the distance range (due to nuclear modulations or errors in background correction), you may change the range for analysis using the + and - buttons for the blue and magenta cursor in the **Distance distribution** panel or direct input into the corresponding edit fields. This option can also be used for extending the distance range if very long distances have been measured or for selecting only a single peak in a multimodal distance distribution and determining its mean distance and width. When the **Expand** checkbox is selected, the distance distribution is displayed only between the cursors.

7.2 Checking for the relevance of small peaks

With Tikhonov regularization, one sometimes observes small peaks in the distance distribution that may be related to noise, to errors in background correction, or to genuine small contributions to the distance distribution. It is

instructive to check the contribution of such peaks to the simulated dipolar evolution function or dipolar spectrum. To suppress such peaks, move the blue and magenta cursors so that they include them (see Fig. 8) and click on the green **Suppress** button. The distance distribution without these peaks is shown as a green curve and the corresponding fit of the experimental data is displayed in the **Dipolar evolution** plot also as a green curve. In the case illustrated in Fig. 8, the small peaks are obviously artifacts. The original (red) fit has a slightly better r.m.s. value, but is not perfect (see first minimum of the oscillation). The green fit is better at the first minimum but worse at the second maximum. In this case, the small peaks should thus be disregarded in interpretation.

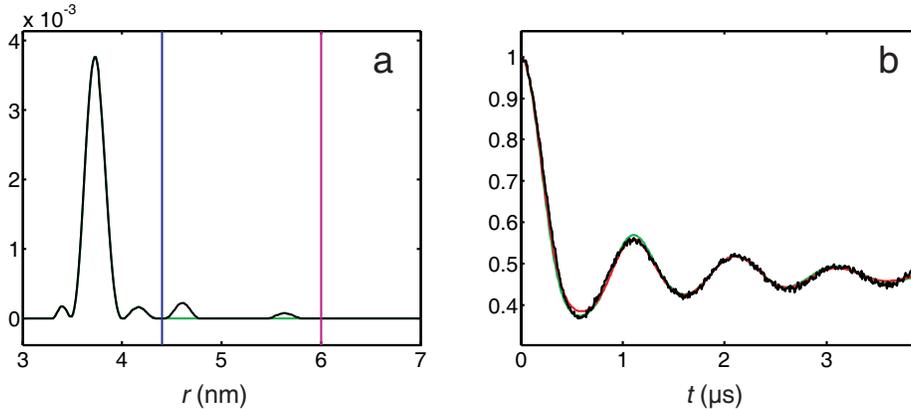


Figure 8: Suppressing small peaks in data set `deer_bi_36_50K` from the `calibdepth` subdirectory. a) Distance distribution obtained by Tikhonov regularization (black) and after suppressing the peaks between the blue and magenta cursor by clicking on the green **Suppress** button (green). b) Experimental dipolar evolution function (black), fit by Tikhonov regularization (red), and fit after suppressing the two small peaks between the blue and magenta cursor.

7.3 Number of coupled spins

The number of spins within a nanoobject can be derived from the (calibrated) modulation depth if decay due to spins in other nanoobjects can be neglected, as was shown early on by the Novosibirsk group [16]. The same applies for the modulation depth in the dipolar evolution function after appropriate correction of the background decay [3]. The total modulation depth is given by

$$\Delta = 1 - \exp[\lambda(\langle n \rangle - 1)] , \quad (13)$$

where $\langle n \rangle$ is the average number of spins in the observed nanoobjects.

To use this information, `DeerAnalysis2006` therefore retains information on the modulation depth in the dipolar evolution function. Quantification requires

knowledge of the modulation depth parameter λ , which depends strongly on the excitation position, length, and flip angle of the pump pulse and weakly on line broadening in the nitroxide spectrum and shape of the resonator mode. Reliable quantification therefore requires a calibration with known samples and proper adjustment of the flip angle of the pump pulse (see Section 10). Spectra from our own series of calibration samples (six biradicals and one triradical) are provided in the folder `calibdepth`. They correspond to 12 ns π pump pulses irradiated at the maximum of the nitroxide spectrum (see Fig. 2) using a Bruker 3mm split-ring resonator. Note that not all example spectra in other folders were measured under the same conditions. To calibrate modulation depths for your own applications, you should measure at least one genuine biradical with close to 100% degree of spin-labeling under your measurement conditions. Analyse the data for this biradical, preferably with Tikhonov regularization and change the number of spins in the corresponding edit field of the `Distance analysis` panel to 2. The number is then displayed in green instead of red color. If another data set, measured under the same conditions, is loaded and processed, the displayed number of spins should correspond to the true average number $\langle n \rangle$ of spins in the nanoobject.

Note that this calibration is lost on restarting DeerAnalysis and that it is unreliable when using excitation bandwidth correction. Also consult Section 5.6.

7.4 Comparing data sets (dual display)

To compare two data sets of the same sample or of similar samples first load one of the data sets and process it as usual. To keep the same processing parameters for the second data set, you may then want to uncheck the `Reset` checkbox below the `Load` button in the `Data sets` panel. After loading the second data set, its file name is shown in line `A:` of the `Data sets` panel. This is the active data set. The file name of the previous data set is shown in line `B:`. The original data and processing results can now be compared by selecting the `Dual display` checkbox in the `Original data` panel. Traces corresponding to the previous data set are now shown in blue in all plots. In the `Dipolar evolution` plot, only experimental data, but no fits are shown for the previous data set.

If the two data sets differ considerably in their modulation depth, but have similar distance distribution, the samples may just differ in the extent of spin labelling or the measurement conditions (flip angles, resonator, pulse lengths) may have been slightly different. To check for this, use modulation depth scaling [3] by selecting the `mod. depth scaling` checkbox in the `Original data` panel. Differences in the distance distribution are noise-related if the original data are not significantly different after such modulation depth scaling.

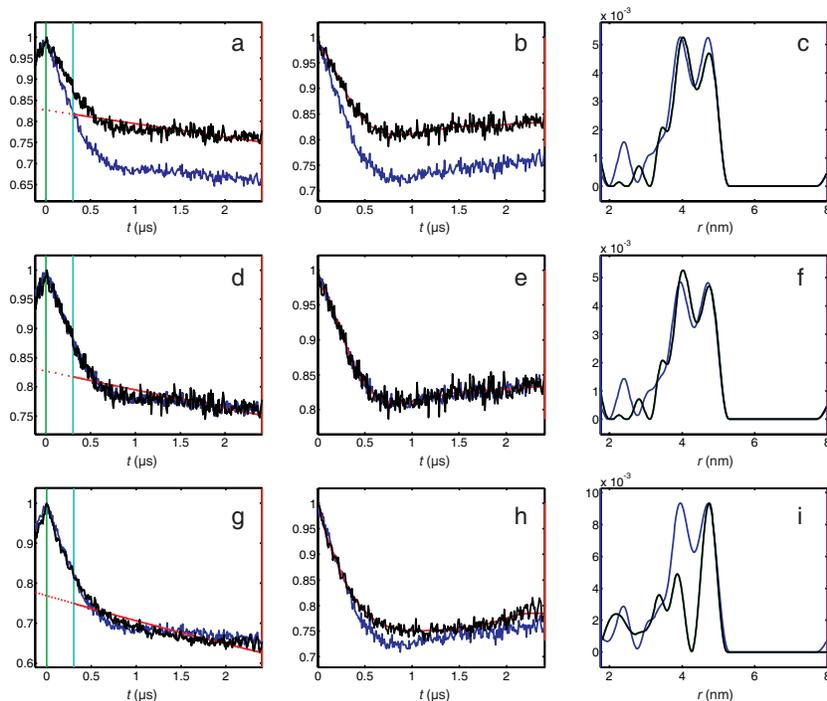


Figure 9: Dual display for comparison of two data sets. All data sets are from the subdirectory `examples\series`. Left column: Original data. Middle column: Dipolar evolution functions after background correction. Right column: Distance distributions. a-c) Comparison of data sets `series2` (active set A, black traces) and `series1` (set B) without modulation depth scaling. d-f) Comparison of data sets `series2` (active set A, black traces) and `series1` (set B) with modulation depth scaling. g-i) Comparison of data sets `series8` (active set A, black traces) and `series1` (set B) with modulation depth scaling.

8 Output

8.1 Saving data

Unlike its predecessor program `DeerAnalysis2004`, the new version `DeerAnalysis2006` does not automatically save results, except as an option during processing series of data sets (see Section 9). On attempt to close the program after time-consuming fits (Tikhonov regularization, model fits) without saving results, the user is reminded. The whole set of data including background correction, experimental and fitted dipolar evolution function and spectrum, distance distribution, processing parameters, results of moment analysis and fitted parameters and L curve (if available) are saved together with the same basis file

name, but into different ASCII files.

After clicking on the **Save** button, the user is asked for the file name. The last extension and, if present, a suffix `_res` are removed to derive the basis name `basname` (this is useful for overwriting old results by selecting their name in the displayed file list). The following files are then saved:

- `basname_res.txt`
a summary of the program settings and the results
- `basname_bckg.dat`
the phase-corrected original data and background fit
1st column: time axis (in μs),
2nd column: real part of original data,
3rd column: background fit
4th column: imaginary part of original data (if present)
- `basname_fit.dat`
the dipolar evolution function and its fit
1st column: time axis (in μs),
2nd column: dipolar evolution function after background correction,
3rd column: fit of the dipolar evolution function
- `basname_spc.dat`
the dipolar spectrum and its fit
1st column: frequency axis (in MHz),
2nd column: experimental dipolar spectrum,
3rd column: fit of the dipolar spectrum
- `basname_distr.dat`
the distance distribution
1st column: distance axis (in nm),
2nd column: distance distribution $P(r)$
- `basname_Lcurve.dat`
the L curve of Tikhonov regularization (only if computed)
1st column: $\log(\rho)$,
2nd column: $\log(\eta)$,
3rd column: corresponding regularization parameters α

The results file `basname_res.txt` protocols all relevant program settings, the mean distance, width of the distance distribution, and third moment, and for Tikhonov regularization, the regularization parameter. For model fits, the values of all fit parameters are also saved here.

8.2 Copying or printing individual plots

The three current plots of DeerAnalysis2006 can be copied into individual Matlab figures by clicking on the **Copy** button in the **Data sets** panel. Using the figure menu, the plots can then be rescaled, edited, annotated, printed, exported as different graphics formats or copied into the Windows clipboard (item **Copy figure** in the **Edit** menu). Matlab has a good help system that explains these possibilities.

9 Processing a series of similar data sets

A global analysis of several data sets is useful when measurements on the same sample have been reproduced or when samples have been prepared under slightly different conditions and one wants to check whether structural changes have occurred (see also Section 7.4). The first case requires computation of an average distance distribution that takes into account the signal-to-noise ratio of the individual data sets. In the second case the comparison should be performed for modulation-depth normalized primary data rather than for distance distributions as it is difficult to estimate what degree of change in the distance distribution is significant [3]. For both tasks a text file *listname.txt* has to be prepared that contains a list of filenames (without extension) of all the data sets that are to be processed together (for an example, see the file *series.txt* in the subdirectory *example\series*).

List processing starts with analysis of a pilot data set, which should ideally be the data set with the best signal-to-noise ratio. This data set with best signal-to-noise ratio should also be the first set in the list, as the first data set is used as a reference for modulation depth scaling. After loading the pilot data set it is processed as usual. Series processing is then initiated by the **Series** button in the **Data sets** panel. Progress is reported in the **Status** panel and line **A:** of the **Data set** panel. Plots are also updated (with a slight delay) during series processing. The program will return after the last data set has been processed. This data set is now the active data set.

The average distance distribution and average dipolar evolution function after series processing as well as average results of moment analysis are not displayed on screen, but are saved automatically. These files have the following formats:

- *listname_res.txt*
a summary of the program settings and the results for the average of all data sets
- *listname_mean.dat*
the mean dipolar evolution function
1st column: time axis (in μs),
2nd column: mean dipolar evolution function after background correction,
- *listname_cmp.dat*
modulation-depth normalized primary data (without background correction)
1st column: time axis (in μs),
 n remaining columns: primary data (real part) for data sets 1 \cdots n ,
- *listname_diff.dat*
 $n \times n$ matrix quantifying the difference between data sets
large values in element (k, j) indicate that data sets k and j differ significantly

Primary data sets and distance distributions are averaged with a weighting factor that is inversely proportional to the mean square deviation of the fit of the

dipolar evolution function. This corresponds to a maximum likelihood estimate of the average.

By default results for the individual data sets are not automatically saved. Automatic saving can be initiated by selecting the **Autosave** checkbox below the **Series** button. Note that even with this checkbox selected, automatic saving takes place only during series processing, not when processing individual data sets via the **Load** button.

10 Hints for Data Acquisition

Conversion of a dipolar evolution function as measured by a magnetic resonance experiment to a distance distribution is an ill-posed mathematical problem [4]. This means that even small deviations from the theoretical function (noise, phase problems, an intensity offset) can cause significant distortions in the distance distribution. Thus, it is of utmost importance to acquire experimental data with the *best* quality possible within a reasonable measurement time.

The choice of a number of experimental parameters has been discussed earlier [9]. From our own experience we suggest to perform measurements in the following way. A temperature of 80 K is a good compromise for most samples, but sensitivity is often somewhat better at 50 K. For critical samples such as membrane proteins, cooling to 50 K is often worth the effort. Unless the sample really has a strong signal, one should plan for measuring two samples in 24 hours, one during the day and one over night. Spectrometers tend to be stable enough over a period of several hours and the quality of the distance distribution tends to be limited by the signal-to-noise ratio except for synthetic model compounds with very narrow distance distributions. The observer and pump frequencies should be stable within 1 MHz during the measurement time, and this should be checked. It is good practice to acquire data with quadrature detection and to adjust the detector phase properly at the beginning. That way instability of the spectrometer can be recognized by the appearance of a significant imaginary part of the signal. Note that a small phase drift (corrections up to 20° for a measurement extending over several hours) is no cause for alarm.

For four-pulse DEER on pairs of nitroxides at X-band frequencies we suggest that the pump pulse has a length of 12 ns. This can be achieved with a Bruker 3-mm-split-ring resonator. We also suggest that all the observer pulses have the same length of 32 ns. These conditions cannot be met at all spectrometers and with all probeheads. Using a length of 32 ns for all pulses, or a length of 16 ns for the $\pi/2$ pulses and a length of 32 ns for the π pulses also provides good results. If your pump π pulse has the same length as the π observer pulses, you may want to set the observer frequency to the center of the resonator mode and the pump frequency into the flank. Note however, that the opposite setting as suggested by Fig. 2 allows for a shorter pump pulse and hence larger modulation depth.

The power of the pump pulse should be adjusted for optimum flip angle (optimum echo inversion) using an inversion recovery sequence $\pi_{\text{pump}} - T -$

$\pi/2_{\text{obs}} - \tau - \pi_{\text{obs}} - \tau - \text{echo}$. This has to be done with coinciding pump and observer frequency at the position in the microwave mode where the pump pulse is applied. After this step the pump frequency should not be changed anymore. If this procedure is not followed, modulation depths are ill-defined and should not be compared between samples. The step is also an absolute requirement if concentrations are to be determined. We suggest that the pump pulse is applied at the maximum of the nitroxide spectrum, which maximizes modulation depth. This minimizes artifacts due to nuclear modulations, phase noise, and spectrometer imperfections.

The observer pulses are then applied at the low-field local maximum which corresponds to increasing the observer frequency (spectrometer frequency) by approximately 65 MHz. You may measure the field difference ΔB_0 between the two maxima and multiply it by 2.8 to obtain the exact frequency difference for your particular nitroxide. A phase cycle $(+x) - (-x)$ should be applied to the first observer pulse to eliminate offsets in the detector channels. If this phase cycling is omitted, any phase correction of the primary data will not be exact and hence background correction by program DeerAnalysis2006 will not be exact. Furthermore, modulation depth information is not reliable. In principle, the problems could be solved by introducing the offset as an additional parameter in background correction, but we strongly discourage such an approach, as it complicates separation of the dipolar evolution function from the background which may be difficult anyway for long distances.

For the interpulse delays in the four-pulse DEER experiment $\pi/2(\nu_{\text{obs}})-\tau_1-\pi(\nu_{\text{obs}})-t'-\pi(\nu_{\text{pump}})-(\tau_1+\tau_2-t')-\pi(\nu_{\text{obs}})-\tau_2-\text{echo}$ we suggest $\tau_1 = 200$ ns for protonated solvents/matrices and $\tau_1 = 400$ ns for deuterated solvents/matrices (at X band). To suppress proton modulations it is advantageous to perform the experiment at eight different values of τ_1 spaced by 8 ns and starting with the values given above. The signals of the eight experiments are added. In variable-time DEER [2] we suggest initial values $\tau_{2,0} = 300$ ns for protonated and $\tau_{2,0} = 500$ ns for deuterated samples. In constant-time DEER, $\tau_2 = 800$ ns is usually convenient for setup (adjustment of the detector phase). For the actual measurement, the choice of τ_2 depends on transverse relaxation, signal strength, and on the longest distances that have to be measured. It is difficult to give general suggestions, but the problem has been discussed in some detail in Ref. [2].

The integration gate should match the width of the observer echo, which is similar to the width of the longest observer pulses. The gate should be centered at the echo maximum. If you can save data in Xepr (Elexsys) format, DeerAnalysis 2004 can directly import binary data. For ESP 380 data we suggest importing them into WIN-EPR and saving them in the *binary* WIN-EPR format (this step converts the coding of binary float numbers to a format legible by MATLAB). If you use another data acquisition system, you should save your data in an ASCII representation.

References

- [1] M. Pannier, S. Veit, A. Godt, G. Jeschke, H. W. Spiess, *J. Magn. Reson.* **142** (2000) 331–340.
- [2] G. Jeschke, A. Bender, H. Paulsen, H. Zimmermann, A. Godt, *J. Magn. Reson.* **169** (2004) 1–12.
- [3] G. Jeschke, G. Panek, A. Godt, A. Bender, H. Paulsen, *Appl. Magn. Reson.* **26** (2004) 223–244.
- [4] G. Jeschke, A. Koch, U. Jonas, A. Godt, *J. Magn. Reson.* **155** (2001) 72–82.
- [5] P.P. Borbat, J.H. Freed, *Chem. Phys. Lett.* **313** (1999) 145–154.
- [6] G. Jeschke, A. Godt, *ChemPhysChem* **4** (2003) 1328–1334.
- [7] G. Jeschke, M. Pannier, A. Godt, H. W. Spiess, *Chem. Phys. Lett.* **331** (2000) 243–252.
- [8] D. Hinderberger, H. W. Spiess, G. Jeschke, *J. Phys. Chem.* **108** (2004) 3698–3704.
- [9] G. Jeschke, *ChemPhysChem* **3** (2002) 927–932.
- [10] Y. W. Chiang, P. P. Borbat, J. H. Freed, *J. Magn. Reson.* **172** (2005) 279–295.
- [11] A. G. Maryasov, Y. D. Tsvetkov, *Appl. Magn. Reson.* **18** (2000) 583–605.
- [12] A. D. Milov, B. D. Naumov, Y. D. Tsvetkov, *Appl. Magn. Reson.* **26** (2004) 587–599.
- [13] A. N. Tikhonov, V. Y. Arsenin, *Solutions of Ill-Posed Problems* (Wiley, New York, 1977).
- [14] J. Honerkamp, J. Weese, *Cont. Mech. Therm.* **2** (1990) 17.
- [15] J. Weese, *Comput. Phys. Commun.* **69** (1992) 99–111 (1992).
- [16] A. D. Milov, A. B. Ponomarev, Y. D. Tsvetkov, *Chem. Phys. Lett.* **110** (1984) 67–72.