

# A short note on the analysis of distance measurements by electron paramagnetic resonance

Silvia Domingo Köhler<sup>a</sup>, Martin Spitzbarth<sup>a</sup>, Kay Diederichs<sup>b</sup>, Thomas E. Exner<sup>a</sup>, Malte Drescher<sup>a,\*</sup>

<sup>a</sup> University of Konstanz, Department of Chemistry, 78457 Konstanz, Germany

<sup>b</sup> University of Konstanz, Department of Biology, 78457 Konstanz, Germany

## A B S T R A C T

In electron paramagnetic resonance (EPR) distance distributions between site-directedly attached spin labels in soft matter are obtained by measuring their dipole-dipole interaction. The analysis of these distance distributions can be misleading particularly for broad distributions, because the most probable distance deviates from the distance between the most probable label positions. The current manuscript studies this effect using numerically generated spin label positions, molecular dynamics simulations, and experimental data of a model system. An approach involving Rice distributions is proposed to overcome this problem.

### Keywords:

Rice distribution  
Spin label EPR  
DEER  
PELDOR  
Distance measurements

## 1. Introduction

EPR spectroscopy has become a powerful tool for distance measurements in disordered systems, giving access to distance *distributions* in the range between 0.5 nm and 8 nm, especially due to the Double Electron Electron Resonance (DEER or PELDOR) technique [1]. The application to biological systems, e.g. membrane proteins, is of particular interest because structure and function in these systems are intimately linked. Many biomacromolecules do not contain native paramagnetic centers but spin labels can be attached site-directedly.

Inter- and intramolecular distances between two spin labels can be determined by measuring their dipole-dipole interaction either by deconvolution of cw-measurements [2] or by pulsed methods [3].

Depending on the flexibility of the investigated system, but often in particular depending on the length and flexibility of the spin label linker, broad distance distributions can be obtained using either model-based data analysis [4,5] or model-free methods, e.g. Tikhonov regularization [6,7]. These broad distributions do by no means reflect the accuracy of the method which can be estimated to 0.05 nm or better [6].

Recent works deal with the analysis of distance distributions, taking rotamer libraries into account, if a structural model, e.g. crystal structure is available [8] but usually the maximum in the

distance distribution or – leading to similar results – the center of a fitted Gaussian is attributed to the distance constraint.

In the current work we demonstrate that the latter procedures are problematic and result in errors in the case of broad distance distributions, e.g. if the center of the experimental distance distribution is of the same order of magnitude as its standard deviation.

To overcome this problem we propose an analysis based on the Rice distribution, either describing the model-free derived distance distribution or as a model for deriving the distance distribution. The Rice distribution is widely used in other disciplines, such as crystallography [9,10], single-molecule fluorescence measurements [11,12], and magnetic resonance imaging (MRI) [13,14].

The Rice distribution is based on the assumption of normally distributed spin label positions. While this is a crude approximation, because the distribution of label positions is dictated by conformation energetics the current manuscript proves the applicability of the Rice distribution in EPR distance measurement by three approaches: (I) numerically simulated spin label positions (normal distributions), (II) molecular dynamics simulations (spin labeled double-stranded DNA), and (III) experimental data (DEER measurement on a polyproline II (PPII) helical peptide).

## 2. Results

In order to illustrate the difficulties in analyzing broad distance distributions one can initially assume a set of two spin labels (point dipole approximation), each positioned normally distributed in space ( $x$ ,  $y$ , and  $z$ ) with a standard deviation  $\sigma$  around two centers with the defined distance  $\mu$  (see Fig. 1).

\* Corresponding author. Fax: +49 7531 883139.

E-mail address: malte.drescher@uni-konstanz.de (M. Drescher).

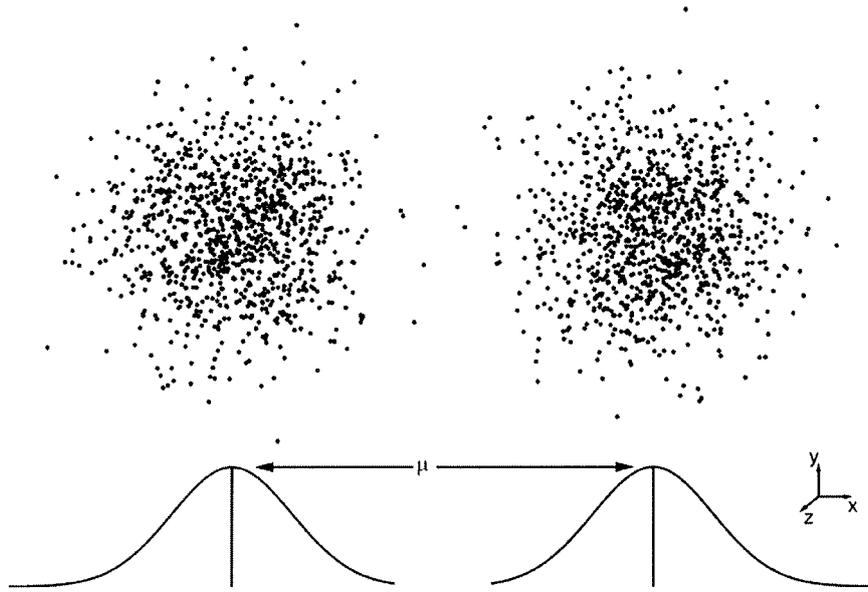


Fig. 1. Schematic illustration of two three-dimensional Gaussian random variables. The centers of the distributions are separated by the distance  $\mu$ , both distributions have the same standard deviation  $\sigma$  in  $x$ ,  $y$  and  $z$  direction.

The resulting distribution between both labels  $p(r)$  with  $r = (\Delta x^2 + \Delta y^2 + \Delta z^2)^{1/2}$  does not correspond to a Gauss but to a Rice distribution [15]. Moreover, the distance distribution exhibits a maximum which is the most probable found distance but deviates from  $\mu$  to larger values. In analogy, fitting this distance distribution by a Gauss distribution the center of the Gaussian is in general larger than  $\mu$ , too.

The distance distribution  $p(r)$  can be described analytically by the Rice distribution in  $n$ -dimensions which is defined by

$$p(r) = \frac{\mu}{\sigma^2} \left(\frac{r}{\mu}\right)^{\frac{n}{2}} \exp\left(-\frac{r^2 + \mu^2}{2\sigma^2}\right) I_{\frac{n}{2} - 1}\left(\frac{r\mu}{\sigma^2}\right) \quad (1)$$

with the modified Bessel function  $I$ . For three dimensions this results in

$$p_{3D}(r) = \sqrt{\frac{2}{\pi}} \frac{r}{\sigma_{\text{rice}} \mu} \exp\left(-\frac{\mu^2 + r^2}{2\sigma_{\text{rice}}^2}\right) \sinh\left(\frac{r\mu}{\sigma_{\text{rice}}^2}\right) \quad (2)$$

with  $\sigma_{\text{rice}} = \sqrt{2}\sigma$  [7].

The effect of varying the widths of the distributions is depicted in Fig. 2, showing data points from numerically simulated distances between normally distributed labels and fits (solid lines) according to equation  $p_{3D}(r)$ . For large  $\mu/\sigma$  values, e.g.  $\mu/\sigma = 10$  (Fig. 2a), the distribution is approximately Gaussian-shaped and the maximum corresponds well to  $\mu$ . As  $\mu/\sigma$  decreases (Fig. 2b–d), i.e. the distribution becomes wider, the curve becomes skewed, broadened, and the position of the maximum differs significantly from  $\mu$ .

This means that for experimentally obtained narrow distance distributions, the maximum of the distribution or the center of a fitted Gaussian curve is in sufficient agreement with the distance between the most probable spin label positions, but deviations occur for broader distributions. This is summarized in Fig. 3, showing the results of fitting either a Gauss or a Rice distribution to numerically simulated ( $\sigma = 0.5$ ) spin label distances in three dimensions. For small  $\mu/\sigma$  values, the biggest discrepancy between Gauss analysis and the distance obtained from the identity function (a straight line through the origin) is obtained. Hence for those values the use of a Gaussian is not justified and results in an error whereas

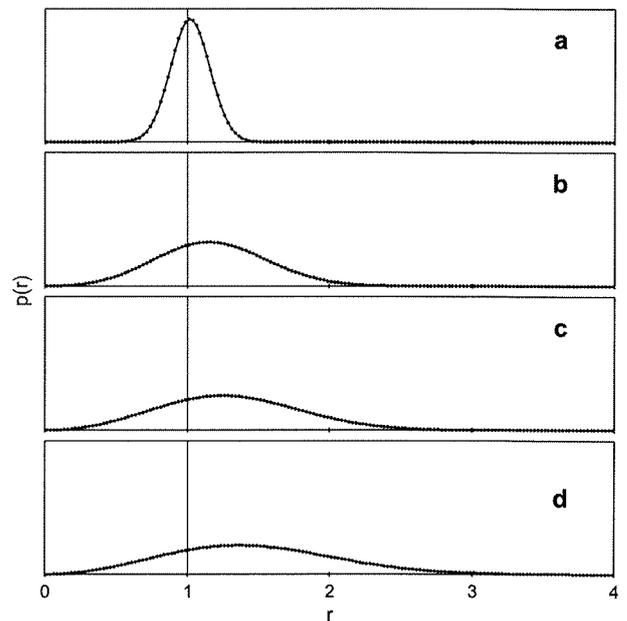
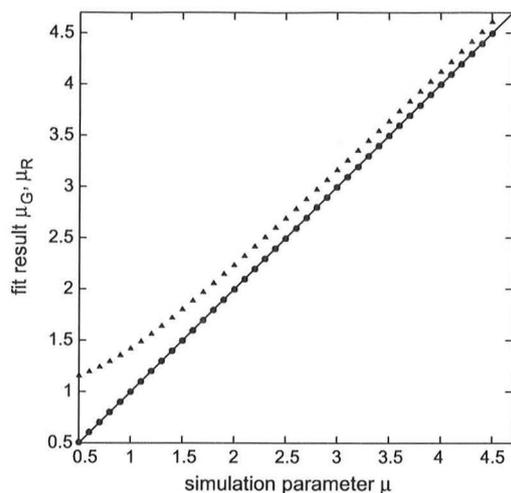


Fig. 2. Numerical simulation of the distance distribution  $p(r)$  between two Gaussian distributed positions in three dimensions (see Fig. 1). The distance between the centers of the two Gauss distributions is  $\mu = 1.0$ , standard deviations are varied: (a)  $\sigma = 0.1$ , (b)  $\sigma = 0.3$ , (c)  $\sigma = 0.4$ , (d)  $\sigma = 0.5$ .

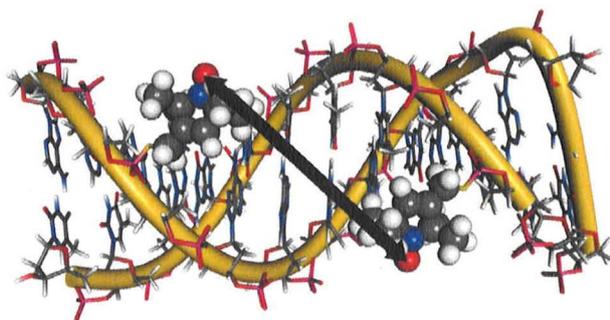
the Rice analysis is in quantitative agreement with the identity function.

In order to study this effect in a more realistic scenario compared to normally distributed spin label positions molecular dynamics (MD) simulations were performed. A double-stranded DNA, doubly labeled with the phosphorothiolate-substituted nitroxide spin label R5 (Fig. 4, see also supporting material) serves as a model system.

For the analysis of the MD data, snapshots were taken every 2 ps and the distance between the oxygen atoms of the spin labels



**Fig. 3.** Comparison of different fit models. For each fixed distance  $\mu$  between the centers of the Gaussian random variables with a fixed standard deviation  $\sigma = 0.5$ , a histogram of the distances between both three-dimensional Gauss distributions was simulated (like in Fig. 2). A Rice distribution and a Gauss distribution were fitted to these histograms. The corresponding fit results  $\mu_i$  are plotted (triangles: Gauss, circles: Rice). The identity function (black line) has been added to the figure for easy comparison of the fit results with the simulation parameter  $\mu$ .

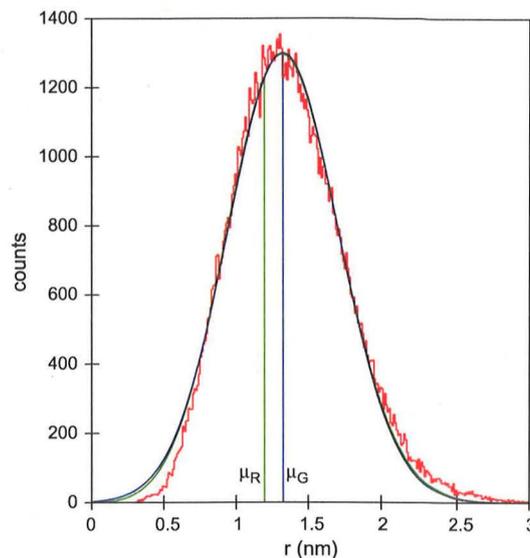


**Fig. 4.** Visualization of the double-stranded DNA dodecamer: Two spin labels (shown as CPK model) are attached to nucleotides four of the first and second strand. The distance is calculated between the oxygen atoms of the two nitroxids (marked by a black arrow).

was measured. Computed distances from 0.5 nm up to almost 3 nm demonstrate the high flexibility of the spin labels. The distances of 25 simulations of 5 ns each were binned into 0.01 nm intervals to derive the distance distribution  $p(r)$  shown in Fig. 5. Additionally, the averaged positions of the oxygen atoms were determined and used to calculate the standard deviations  $\sigma_i$  and the distance  $\mu$  between the most probable spin label positions. We find  $\sigma = 0.27$  nm and 0.26 nm for the oxygen atoms of the first and second spin label, respectively, and  $\mu = 1.24$  nm.

The distance distribution in Fig. 5 was analyzed using a Gaussian ( $\mu_G = 1.33$  nm,  $\sigma_G = 0.38$  nm) and a Ricean ( $\mu_R = 1.20$  nm,  $\sigma_R = 0.40$  nm) fit, respectively, the latter in excellent agreement with the distance between the most probable label positions.

The comparison between the analysis based on the Gauss or Rice distribution, respectively, was also applied to experimental data of an EPR distance measurement (see supporting material). The model system used here consists of a peptide in a polyproline II helix conformation [16,17], which was doubly labeled by site-directed spin labeling using MTSL. While the qualities of both fit curves are indistinguishable on a relevant scale, the derived  $\mu$  values differ significantly ( $\mu_G = 4.20$  nm and  $\mu_R = 4.04$  nm, with  $\sigma_G = 0.80$  nm and  $\sigma_R = 0.82$  nm).



**Fig. 5.** Distribution of the oxygen–oxygen distance of the spin labels taken from MD simulations (red line), including Rice (green line) and Gaussian (blue line) fit curves. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

### 3. Discussion

One of the unique features of distance measurements by spin label EPR is the ability to monitor even broad distance distributions.

The data presented here illustrate that there is a difference between the most probable spin label distance, and the distance between the most probable spin label positions, which is significant at least for broad distance distributions, e.g.  $\mu/\sigma < 4$  (see Fig. 3).

This becomes immediately obvious in the following gedankenexperiment: reduce the distance  $\mu$  between the two clouds of Gaussian distributed spin label positions (with  $\sigma > 0$ ) in Fig. 1 to  $\mu = 0$ . Taking all possible label positions into account, one obtains the distance distribution containing distances equal to or larger than zero. Also the most probable distance is larger than zero and therefore larger than  $\mu$ , too.

The spin label MTSL is widely used in protein studies because it can be covalently linked to any cysteine residue via a sulfur–sulfur bridge. The linker length for this spin label is approximately 0.5 nm. Using this value, the width of the distance distribution caused by the linker flexibility only, can be estimated to be 1 nm. This has also been found in experimentally obtained distance distributions when the MTSL-labels are not significantly sterically hindered. In this case, distance distributions between MTSL labels below 3 nm result in  $\mu/\sigma < 3$ . No matter whether model-based or model-free analysis, e.g. Tikhonov regularization was used, the most probable distance differs from the distance between the most probable spin label positions in this case. Hence, when qualitative analysis using rotamer libraries is impossible, the maximum in the distance distribution  $p(r)$  can be seriously misleading.

This is shown using three different approaches:

- (I) Using numerically generated spin label positions obeying normal distributions the resulting distance distribution can be given analytically by the Rice distribution. Fig. 2 clearly shows the drastic deviations of those distance distributions  $p(r)$  from a Gaussian-like behavior and the shift of the maximum of  $p(r)$  to values higher than  $\mu$ .
- (II) To deal with more realistic spin label positions MD simulations with long simulation times were performed. Again,

even if the spin label positions do not obey normal distributions, the Rice distribution delivers a result which is in good agreement with the distance between the most probable spin label positions. The distance derived by fitting a Gaussian deviates by more than 7%.

- (III) Considering experimental data we presented the analysis of a DEER measurement on a doubly-MTSL-labeled model peptide in a PPII helix conformation by Tikhonov regularization. Even when the  $\mu/\sigma$ -ratio is approximately 4.9, the deviation between  $\mu_G$  as derived by fitting a Gaussian and  $\mu_R$  derived by using the Rice distribution is still significant (0.2 nm).

#### 4. Conclusions

Distance measurements using EPR combined with a spin label approach are a powerful tool to determine long range distance constraints when studying the structure of soft matter. Site-directed spin labeling using nitroxides covalently bound via a linker is commonly applied. The analysis of experimentally obtained distances is not limited by the intrinsic accuracy of the method [1] but rather by the undetermined label positions due to the label linker flexibility. Possible solutions to this problem are rotamer libraries or molecular dynamics simulation, though these are not applicable in many cases.

Often the most probable distance between the spin label positions is used as distance constraint.

The conclusion of this work is that by taking the maximum of the distance distribution, one commits a systematic error, especially for small  $\mu/\sigma$  values.

Therefore an alternative analysis method is proposed here. The Rice distribution describes the distance distribution resulting from two three-dimensional Gaussian distributed spin positions and is applicable to experimentally obtained distance distributions. It allows the determination of the distance between the most probable spin label positions, which is of particular importance in analyzing three-spin correlations [18] in double electron–electron resonance, where the deviation between the most probable distance and the distance between the most probable spin label positions can introduce internal inconsistency.

The Rice distribution as a model-based analysis method for distance measurements in EPR will be implemented in future versions of DEERAnalysis [19].

#### Acknowledgments

This work was financially supported by the Deutsche Forschungsgemeinschaft (DR 743/2-1). The authors thank G. Jeschke for stimulating discussions.

Ian S. Haworth is gratefully acknowledged for the spin label parameters and the bwGRiD in Ulm and Stuttgart for providing the computer resources for the MD simulations.

#### Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmr.2010.10.005.

#### References

- [1] A.D. Milov, A.B. Ponomarev, Y.D. Tsvetkov, *Chem. Phys. Lett.* 110 (1984) 67.
- [2] H.-J. Steinhoff, Inter- and intra-molecular distances determined by EPR spectroscopy and site-directed spin labeling reveal protein–protein and protein–oligonucleotide interaction, *Biol. Chem.* 385 (2004) 913–920.
- [3] A. Schweiger, G. Jeschke, *Principles of Pulse Electron Paramagnetic Resonance*, Oxford University Press, 2001.
- [4] M. Pannier, M. Schöps, V. Schädler, U. Wiesner, G. Jeschke, H.W. Spiess, Characterization of ionic clusters in different ionically functionalized diblock copolymers by CW EPR and four-pulse double electron–electron resonance, *Macromolecules* 34 (2001) 5555–5560.
- [5] M. Drescher, G. Veldhuis, B.D. van Rooijen, S. Milikisyants, V. Subramaniam, M. Huber, *JACS* 130 (2008) 7796–7797.
- [6] G. Jeschke, G. Panek, A. Godt, A. Bender, H. Paulsen, Data analysis procedures for pulse ELDOR measurements of broad distance distributions, *Appl. Magn. Res.* 26 (2004) 223–244.
- [7] Y.-W. Chiang, P.P. Borbat, J.H. Freed, The determination of pair distance distributions by pulsed ESR using Tikhonov regularization, *JMR* 172 (2005) 279–295.
- [8] G. Jeschke, Y. Polyach, Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance, *Phys. Chem. Chem. Phys.* 9 (2007) 1895–1910.
- [9] G.N. Murshudov, A.A. Vagin, E.J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr. D* 53 (1997) 240–255.
- [10] P.H. Zwart, V.S. Lamzin, Distance distributions and electron-density characteristics of protein-models, *Acta Crystallogr. D* 59 (2003) 2104–2113.
- [11] L.S. Churchman, H. Flyvbjerg, J.A. Spudich, A non-Gaussian distribution quantifies distances measured with fluorescence localization techniques, *Biophys. J.* 90 (2006) 668–671.
- [12] J. Antelman, C. Wilking-Chang, S. Weiss, X. Michalet, Nanometer distance measurements between multicolor quantum dots, *Nano Lett.* 9 (2009) 2199–2205.
- [13] H. Gudbartsson, S. Patz, The Rician distribution of noisy MRI data, *Magn. Reson. Med.* 34 (6) (1995) 910–914.
- [14] A.J. den Dekker, J. Sijbers, Implications of the Rician distribution for fMRI generalized likelihood ratio tests, *Magn. Res. Imaging* 23 (2005) 953–959.
- [15] K.S. Miller, *Multidimensional Gaussian Distributions*, Wiley and Sons, New York, 1964.
- [16] S. Domingo Köhler, A. Weber, S.P. Howard, W. Welte, M. Drescher, The proline-rich domain of TonB possesses an extended polyproline II-like conformation of sufficient length to span the periplasm of Gram-negative bacteria, *Protein Sci.* 19 (2010) 625–630.
- [17] M. Kuemin, S. Schweizer, C. Ochsenfeld, H. Wennemers, Effects of terminal functional groups on the stability of the polyproline II structure: a combined experimental and theoretical study, *JACS* 131 (2009) 15474.
- [18] G. Jeschke, M. Sajid, M. Schulte, A. Godt, Three-spin correlations in double electron–electron resonance, *Phys. Chem. Chem. Phys.* 11 (2009) 6580–6591.
- [19] G. Jeschke, V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J. Banham, C.R. Timmel, D. Hilger, H. Jung, DEERAnalysis2006 – a comprehensive software package for analyzing pulsed ELDOR data, *J. Appl. Magn. Res.* 30 (2006).