New Statistical Inference Methods for DEER Spectroscopy on Proteins

Thomas H. Edwards

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Stefan Stoll, Chair

Xiaosong Li

Paul Wiggins

Program Authorized to Offer Degree:
Chemistry

University of Washington

Abstract

New Statistical Inference Methods for DEER Spectroscopy on Proteins

Thomas H. Edwards

Chair of the Supervisory Committee:
Assistant Professor Stefan Stoll
Chemistry

Double Electron-Electron Resonance (DEER) spectroscopy is a pulse-EPR experiment that can provide sub-ångström resolution distance measurements of proteins and other biomacromolecules in the distance range of about 2-16 nm. It is growing in importance as a tool for structural biology and, due to its robustness towards difficult sample conditions, complements existing experimental and theoretical techniques quite well. An outstanding problem in the field is the analysis and interpretation of the data: currently popular workflows invite user bias and there is little to no quantification of uncertainty. This is a significant barrier to more widespread adoption of the technique and may call into question published results.

Transforming time-domain DEER data into distance distributions is an ill-conditioned, inverse problem. This means that there is no analytical solution and some form of regularization is necessary, which precludes straightforward fitting and error analysis. The work in this thesis introduces a Bayesian statistical framework for the processing and analysis of DEER data. The result is a set of numerical tools that allow for automated and statistically robust analysis.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

It takes a village to raise a graduate student, and I am fortunate to be supported by several wonderful communities. This work could not have been accomplished without them. Most significantly, I must thank my PhD advisor, Stefan Stoll, for his seemingly boundless enthusiasm, guidance, energy, support, and patience. I am his second PhD student and it has been a privilege to be a part of the early years of his lab at the University of Washington. This journey has been the most worthwhile of my academic career.

My fellow groupmates (the Stoll Patrol) have been an endless source of encouragement, inspiration, guidance, and fun. It has been an honor to work alongside so many wonderful young scientists. They have all been to me—in their own ways—friends, colleagues, and mentors. Ellen Hayes, Stefan's first student, has been an endless source of guidance, support, and friendship. When Jeremy Lehner switched into our group, Ellen and I were thrilled. He is a joy to be around and it has been fulfilling to watch his project mature into a truly significant contribution to the field. Saralyn Ogden, before her move to greener pastures in industry, was a welcome source of commiseration about the particulars of having moved to Seattle from Utah to pursue a career in science. Donald Mannikko brought with him some much-welcomed hard-won wisdom, humor, and friendship. Lizzy Canarie's arrival catalyzed a marked increase in the intra-group friendships. Sam Jahn's calm kindness, focused intellect, and theoretical strength have pushed the group into new and exciting conceptual landscapes. I could not have asked for a better successor than Sarah Sweger. Her intellectual rigor and work ethic have pushed me to be better and left me hopeful about the future of the ideas explored in this thesis. Hannah DeBerg joined the lab as our first postdoc near the end of my first year and I cannot overstate how much she helped me during my early years in the lab. From providing me with the samples I needed to get started to patiently explaining key concepts and procedures to educating me about how to manage my advisor (and that I had to), her guidance and support was crucial. Claudia Tait elevated the level of theoretical and experimental rigor in the lab to new heights and I am lucky to have overlapped with her time here. Eric Evans is a molecular biology wizard, which has greatly benefited the lab both in terms of sample preparation and homebrew beer supply. I am grateful for his friendship and look forward to crushing many more miles together.

I am incredibly grateful to the Molecular Biophysics Training Grant (MBTG) program for their mentorship and support during my third and fourth years at UW (NIGMS/NIH T32GM008268). Ning Zheng, Rachel Klevit, and Christina Larmore put in a tremendous amount of work to ensure this program's funding, longevity, and success. The training grant enabled me to focus

on my research, provided a very good level of training in biophysics and the major experimental techniques in that field, and provided generous travel funding that helped me to share my work at the Biophysical Society meetings. The program is also supported and made as wonderful as it is by many other faculty members around UW, but most significantly to me, by the other trainees. In particular, to Hannah Baughman, Amanda Clouser, Matthew Cook, Ken Laszlo, Kayla Sapp, Caitlin Cornell, Yang Hsia, James Jaffe, and Julie Cass, I am very grateful for the discussions we held together.

The NSF-funded SharedEPR Network graciously provided me funding to travel to both ACERT at Cornell and the Rocky Mountain Conference of Magnetic Resonance in 2018. At Cornell, Prof. Jack Freed's lab made me feel at home and I was very lucky to receive truly expert tutoring in the ways of the wavelet by Madhur Srivastava. I am grateful to Madhur for his time, patience, and friendship.

Prof. Bill Zagotta's lab, in addition to providing Hannah and Eric, has been a fantastic resource in terms of molecular biological and biophysical knowledge. They have always been very generous and patient with us. Michael Puljung, along with Hannah, helped to provide me with the samples that kick-started this work and ensured that I had enough data to pass my second-year exam.

My committee has provided valuable feedback and support throughout my graduate career, and I am grateful to them. Discussing my work with them has been a true highlight of my time at UW.

My friends, family, and partner have provided constant and unconditional love and support. There are too many of you and too much to say about each one of you to reasonably record in words. I cannot imagine making it this far without you. Thank you.

I moved to Seattle, WA in August of 2012 to begin graduate school. Prior to that, I had lived in Salt Lake City, UT my entire life. It was not easy to leave behind my friends and family to live in a larger city where I knew nobody. I was incredibly lucky to wind up living with Bjorn Hubert-Wallander (Uber) and Ashwin Bhandiwad (NIH) for the first three years of my PhD. Bjorn and Ashwin were PhD students in the Psychology department here at UW, and they truly helped me to feel welcome in my new home. Similarly, the UW Starcraft 2 team provided me with much-needed friendship and recreation outside of school. Although I was only active with the team for one year, I won't soon forget flying to Los Angeles to compete against the 3 other best collegiate teams on the planet. I am grateful for how the team immediately made me feel at home; I continue to value those friendships.

The Stoll Patrol half-jokingly refers to Big Time Brewery as our second office. I offer sincere thanks to Rick McLaughlin and the rest of the Big Time staff for curating a warm and inviting atmosphere fit for both work and play. I am lucky enough to call several of the staff my friends and their support has been invaluable.

Having had no formal training in applied math or statistics, I had a somewhat difficult time getting this project off of the ground. François Monard and Niket Thakkar pointed me in the right directions and saved me perhaps months of effort. Johnathan Bardsley from the University of Montana provided very valuable feedback and guidance towards the efforts that resulted in my first publication.

One example of Stefan's endless patience is his willingness to permit me to attend Burning Man every year in Black Rock City, NV. This yearly ritual has become very important to me and I am humbled to now be a part of the motley community of helpers known as the Black Rock Rangers. A team of mediators, peer counselors, safety volunteers, and first responders, they have encouraged the best in me while helping to diminish the worst. Special thanks to Camp Catalyst and The Contraptionists for providing homes in the Dust.

Chapter 1

INTRODUCTION

1.1 Overview

This thesis describes improvements to data analysis methods for a pulse Electron Paramagnetic Resonance (EPR) spectroscopy experiment called Double Electron-Electron Resonance (DEER). By measuring through-space magnetic dipolar couplings, this experiment can report the distance between two or more unpaired electrons in the 2-16 nm range to sub-ångström resolution [1–3]. DEER is a two-frequency pump-probe experiment that records the amplitude modulation of an electron spin echo as a function of the pump-pulse position. The frequency of this modulation is determined by the magnitude of the magnetic dipolar coupling between the spin angular momentum moments of interest, and is proportional to the inverse cube of the distance. Combined with mutagenic site-directed spin-labeling (SDSL) [4, 5], DEER has become an increasingly important tool for structural biology [6–10], especially for systems that are not amenable to more traditional methods such as solution NMR and X-ray crystallography.

Crucially, DEER provides not just average distances, but full probabilistic distance distributions, $P(r)$, (Fig. 1.1). If multiple molecular conformations are present, this will manifest in the $P(r)$ as broadened and/or multiple peaks, provided those conformations feature distinct spin-spin distances. Thus, even though DEER data can be considered coarse-grained—only one or a few distance restraints per mutant and/or measurement—the explanatory power per measurement can be quite high. This often enables practitioners to answer important biophysical questions with only a few DEER measurements [11].

Although it is now an established tool for biological study, DEER—and related pulse sequences like DQC, SIFTER, and RIDME—is still a relative newcomer and remains under active development. Researchers are continuously making improvements in sample preparation (new spin-labels, spin-labeling chemistry, biomacromolecule reconstitution, microfluidics & RFQ), spectrometer hardware (new frequencies/fields, resonator design, pulse-shaping, more powerful amplifiers), and data analysis (faster algorithms, fitting methods, incorporation of DEER results into molecular modeling).

DEER data analysis can be roughly conceptually divided into processing and interpretation. Both are open research topics. Processing involves the numerical procedures that lead from raw DEER signals to distance distributions. It can be complicated by quadrature phase correction, background signal removal or fitting, orientation selection, multi-spin effects, high-spin effects, pulse overlap, differential phase memory times [12], and fitting parametric or non-parametric models to the data. Interpretation comprises drawing conclusions about a physical system from

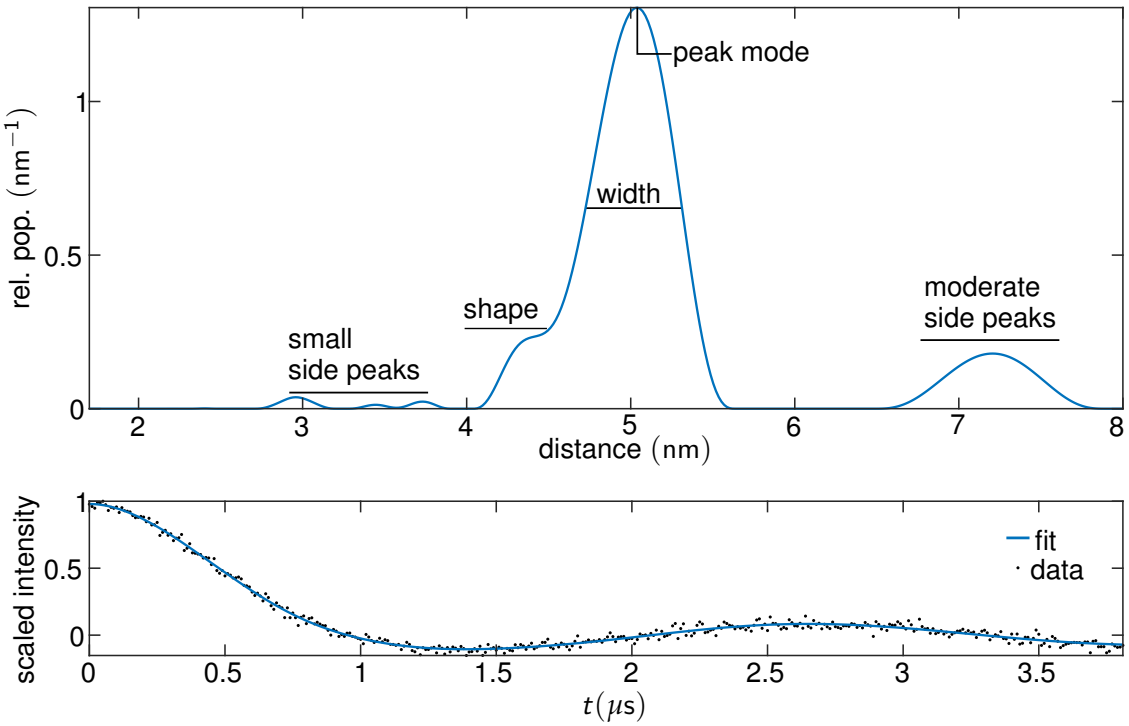

Figure 1.1: Top panel: DEER distance distribution, $P(r)$, inferred from the $S(t)$ in the bottom panel using Tikhonov regularization (Subsection 1.2.2). The figure annotations highlight several characteristics of the $P(r)$ that may contain structural information and are also sensitive to various sources of uncertainty. Bottom panel: background-corrected Q-band 4-pulse DEER data ($S(t)$) displayed as black dots with the corresponding Tikhonov fit in solid blue. The sample was E22C A338C maltose binding protein doubly-labeled with the MTSSL spin label.

the DEER distance distribution or other parameters. Beyond underlying complications from the particular system under study, such as heterogeneous conformations or oligomerization, interpretation of DEER data is hindered by the fundamental disconnect between label-label distances and sample-sample distances due to the spin-label linker.

This dissertation is confined to DEER data processing assuming the most simple (to model) sample reality: spin-$\frac{1}{2}$ nitroxide biradicals uniformly randomly distributed in a perfect frozen glassy solution in the high temperature limit with no orientation selection or other complicating factors. The numerical tasks considered are background correction and inference of distance distributions using a non-parametric model approach. The novel contributions introduced in this body of work are an optimized implementation of Tikhonov regularization [13] and a hierarchical Bayesian model approach to process DEER data using a reproducible and statistically robust methodology [14].

Inferring distance distributions from DEER data is non-trivial due to the inverse and ill-conditioned nature of the problem [15–17]. There exist several distinct currently utilized approaches: Tikhonov regularization [15–18], Gaussian mixture models [19–21], Tikhonov regularization post-processed with Gaussians [22, 23], Tikhonov regularization combined with maximum entropy [24], Bayesian inference (based upon Tikhonov regularization) [14], regularization by limiting the number of points in the distance domain [25], wavelet denoising [26], truncated singular-value decomposition [17, 27], the integral Mellin transform [28], and neural networks [29, 30]. Polynomial fits and the Approximate Pake Transform [15, 31, 32] are no longer in widespread use. Currently, Tikhonov regularization is the most widely employed method, usually through the MATLAB toolbox `DeerAnalysis` [18] or the LabView package `LongDistances` [33].

The Gaussian mixture model approach is distinct from other methods in that it uses a parametric model. Parametric model approaches feature straightforward non-negativity constraints and uncertainty quantification at the cost of more significant assumptions about the shape of the $P(r)$. Determining the optimal model complexity (for example, the number of Gaussians to use to represent $P(r)$) involves solving a statistical model selection problem. One such approach to quantify uncertainty is to assume a particular model for $P(r)$, fit the model to the data, and then statistically evaluate the error in the various fit parameters, usually with some variant of the χ^2 cost function. This method has been employed in `LongDistances` [33], `DeerSim` [34–37], and `DEERconstruct` [38]. Alternatively, `DEFit` fits Gaussian distributions directly to the time-domain data [39]. Hustedt and coworkers have implemented in `GLADD` a model-based approach that simultaneously fits the background function and $P(r)$ to time-domain data, thereby uncovering potential correlations between background parameters and features in $P(r)$ [21, 40].

So far, non-parametric model approaches, such as Tikhonov regularization, have lacked rigorous uncertainty quantification, complicating interpretation and reducing reproducibility [6, 17, 18, 21]. This is a major obstacle to more widespread adoption of the method and decreases the utility of the information gathered. These methods also require background correction as a preprocessing step and it is difficult to estimate the uncertainty introduced. Concerningly, it is

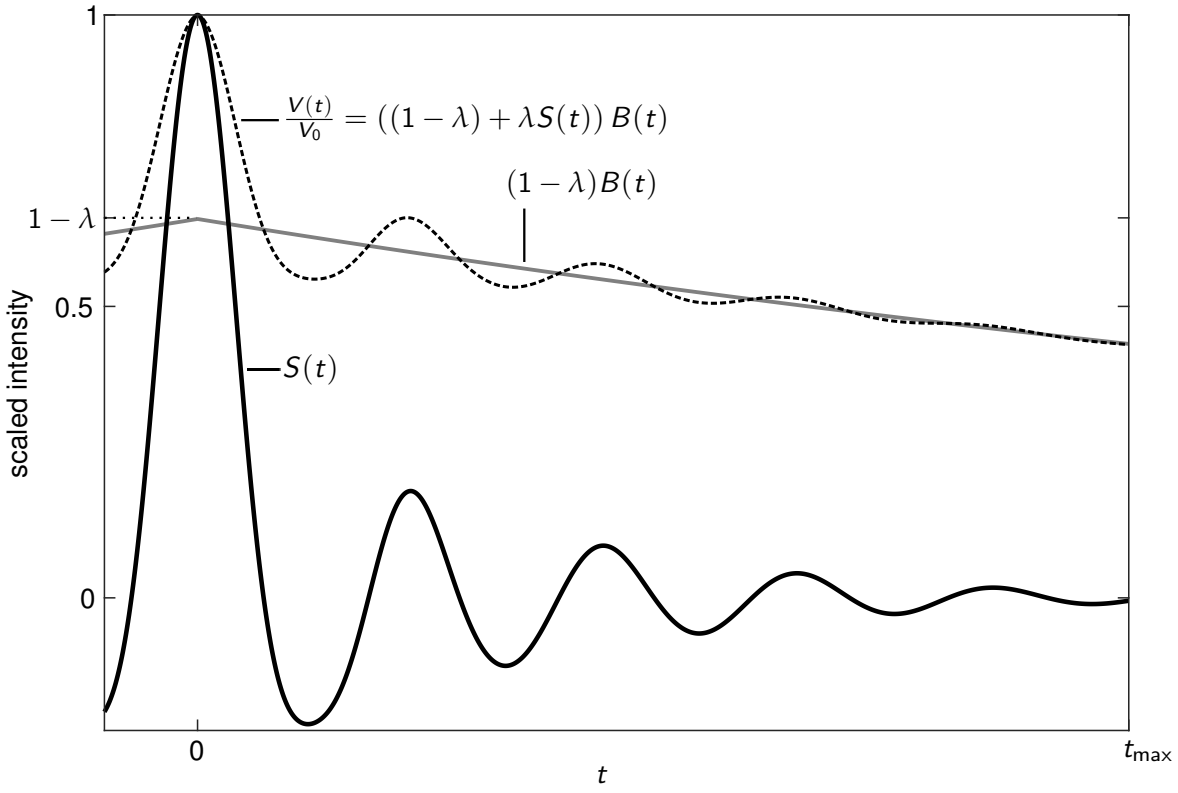

Figure 1.2: Simulated DEER signal assuming randomly-distributed, well-isolated spin-pairs. The corresponding $P(r)$ is a Gaussian distribution with mean 3.5 nm and standard deviation 0.2 nm. The scaled DEER signal, $\frac{V(t)}{V_0}$, is shown in dashed black. Note that it has a maximum value of 1 at $t = 0$ and eventually decays to 0. As one moves further from the noise-free limit, the value of V_0 becomes increasingly difficult to estimate. The intermolecular background contribution, $B(t)$, also has a maximum of 1 at $t = 0$. Shown in grey is $B(t)$ multiplied by the factor $(1 - \lambda)$ where λ is the modulation depth. This is the curve typically fit to the DEER signal for background correction. The intramolecular component, $S(t)$, is shown in solid black. It is a damped oscillation around 0 with a maximum of 1 at $t = 0$.

very uncommon for any error or uncertainty estimates to be displayed in $P(r)$. After surveying over 100 published papers from 2014-2016 containing DEER distance distributions, we have found that nearly 90% lacked error estimates on $P(r)$ (see Appendix C).

The Bayesian method proposed here infers fit parameters and quantifies uncertainty arising from time-domain noise, regularization, and background correction, ameliorating one of the disadvantages associated with Tikhonov regularization compared to parametric model-based approaches. By allowing robust uncertainty quantification and background correction, the work introduced in this dissertation will permit direct comparison between parametric and non-parametric approaches for DEER data processing.

1.2 Theory

This section contains a basic description of the DEER signal and its analysis sufficient to motivate the contents of the subsequent chapters.

1.2.1 DEER Signal

DEER measures the (typically integrated) electron spin echo intensity, V, as a function of the position in time of one or more pump pulses, t (Fig. 1.2). The most common implementation of DEER is a four-pulse sequence [41] with one pump pulse, depicted in Fig. 1.3. For dilute samples of doubly-labelled molecules or complexes, V is the product of an unmodulated spin-echo amplitude V_0, a modulation function due to intra-complex coupling, F, and a background function from the inter-complex interactions, B:

$$V(t) = V_0 \cdot F(t) \cdot B(t) = V_0 \cdot ((1-\lambda) + \lambda S(t)) \cdot B(t) \tag{1.1}$$

where λ is the modulation depth, a parameter related to the excitation efficiency of the DEER experiment. The intermolecular background contribution, $B(t)$, is:

$$B(t) = \exp\left(-|c\rho_{\text{pump}}t|^{D/3}\right) \tag{1.2}$$

where D is a phenomenological stretch constant (usually 3 for soluble proteins). The decay constant $c = \frac{2\pi}{9\sqrt{3}\hbar}\mu_0{\mu_\text{B}}^2 g_\text{A} g_\text{B} c_\text{spins}$ is proportional to the spin concentration c_spins, in units of spins·m^{-3}, μ_B is the Bohr magneton, and g_A and g_B are the isotropic average g values for spins A (probe/observe) and B (pump). The percentage of B-spins excited by the pump pulse, ρ_pump, also affects the decay rate.

It is common to replace the pump-pulse efficiency, ρ_pump, in Eq. 1.2 with the modulation depth, λ. This reflects an outdated understanding of the DEER signal. In 2014, Salikhov pointed out that the modulation depth, λ, is also affected by a few other properties, especially the presence of A-A diradicals (both spins on a protein are excited by the observer sequence)

[42]. The two quantities are deeply related and will be, in most cases, quite similar in numerical value, but they are not the same.

When analyzing DEER data, whether by Tikhonov regularization or a different approach, a background model is typically fitted to $V(t)/V_0$ and divided out, yielding the isolated intra-complex modulation function $S(t)$ after scaling by λ. $S(t)$ is normalized such that $S(0) = 1$ in the noise-free limit. Fig. 1.2 illustrates a simulated DEER signal before and after background correction.

The background contribution, $B(t)$, is not always removed before estimating $P(r)$. Examples include Hustedt and coworkers' global analysis method [40] and the approach explored in Ch. 4.

In the absence of pulse overlap (2+1 modulation, nuclear modulations, pump echoes), orientational correlation between the labels and orientation selection, exchange coupling, differential relaxation, and multi-spin effects, $S(t)$ is related to the distance distribution between the two spins on the complex, $P(r)$, by

$$S(t) = \int_0^\infty \mathrm{d}r\; K(t,r)P(r). \tag{1.3}$$

The dipolar kernel, K, is:

$$K(t,r) = \int_0^{\pi/2} \mathrm{d}\theta\;\; \sin\theta \cos\left[(1-3\cos^2\theta)\omega_\perp t\right] \tag{1.4}$$

$$= \frac{\mathcal{C}(z)}{z}\cdot\cos(\omega_\perp t) + \frac{\mathcal{S}(z)}{z}\cdot\sin(\omega_\perp t) \tag{1.5}$$

where $\omega_\perp = \mu_0\mu_\mathrm{B}^2 g_\mathrm{A} g_\mathrm{B}/(4\pi\hbar r^3)$ is the coupling in angular frequency units between spins A and B when the inter-spin vector is perpendicular to the magnetic field, B_0. The substitution $z = \sqrt{6\omega_\perp t/\pi}$. Because they can be quickly and accurately calculated, we rewrite the kernel using the cosine and sine Fresnel integrals $\mathcal{C}(z)$ and $\mathcal{S}(z)$, respectively [43, 44]. Finally, $P(r)$, being a probability density function (PDF), integrates to one: $\int_0^\infty \mathrm{d}r\; P(r) = 1$.

When S is discretized as an n_t-element vector with elements $S_i = S(t_i)$, and P as an n_r-element vector with elements $P_i = P(r_i)$, the integral transformation from P to S is represented by the matrix-vector multiplication (linear model):

$$S = KP \tag{1.6}$$

where K is the $n_t \times n_r$ kernel matrix. Its elements are $K_{ij} = K(t_i, r_j)\cdot\Delta r$, with the distance increment Δr, assuming a uniform spacing of r_i. Usually, n_r is set equal to n_t.

The matrix K is close to singular, and the calculation of the inverse of $K^\mathsf{T}K$—needed to obtain P from S in an ordinary least-squares fitting procedure—is very inaccurate [31]. To enable the solution, a penalized least-squares approach known as Tikhonov regularization is often used [15, 17].

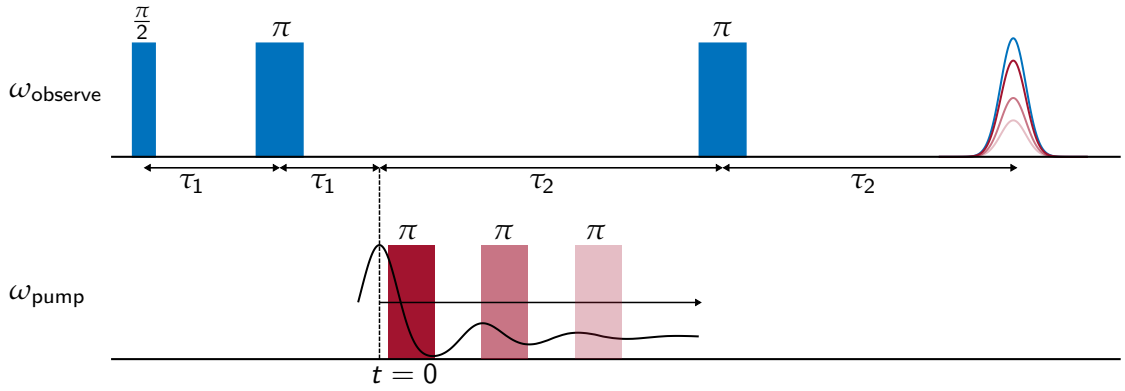

Figure 1.3: Schematic of the DEER pulse sequence. The three-pulse observer sequence (top, blue) forms a refocused echo at time $2\tau_1 + 2\tau_2$. The echo amplitude modulates as a function of the time-position of the pump pulse (bottom, red). The maximum dipolar evolution time is τ_2. The integrated echo amplitude as a function of time is represented by the solid black curve overlaid on the pump pulse channel. The attenuated spin-echoes are plotted in shades of red corresponding to particular pump-pulse times. The schematic is not to scale: the pulses typically range between 5 and 300 ns while τ_2 is typically on the order of 2-10 μs.

1.2.2 Tikhonov Regularization

Tikhonov regularization was introduced to NMR for de-Pake-ing [45], for the extraction of internuclear distances from dipolar time-domain signals such as those from REDOR [46], for the determination of orientational distributions from ^{2}H NMR data [47, 48], and relaxation rate distributions [49, 50]. These approaches used the self-consistent method for selecting α, as introduced and implemented in the program `FTIKREG` [51, 52]. In the context of extracting distance distributions from DEER data, Tikhonov regularization was initally mentioned in 2002 [31, 32], and first applications appeared in 2004 [15, 16]. In these papers, the regularization parameter was selected manually or using `FTIKREG`. A thorough paper examining Tikhonov regularization and introducing the use of the L-curve maximum-curvature criterion appeared in 2005 [17]. A different L-curve method, the minimum-radius criterion, was introduced in 2006 in the program `DeerAnalysis` [18] and is used in its current release (2018).

Tikhonov regularization is expressed by the minimization problem:

$$P_{\alpha L} = \underset{P \geq 0}{\operatorname{argmin}} \left(\|S - KP\|^2 + \alpha^2 \|LP\|^2 \right). \tag{1.7}$$

This is a form of penalized least-squares fitting. The first term is the least-squares term capturing the misfit between the non-parametric model, KP, and the data, S. The second term penalizes for unwanted properties of the solution, P, and depends on a specific form for the regularization operator, L, and a specific value for the regularization parameter, α. The effect of α is summarized in Fig. 1.4. The fit is constrained by the requirement that all elements of P be non-negative ($P \geq 0$). The subscripts in $P_{\alpha L}$ indicate that the solution depends on the particular choices of α and L. An example of a $P(r)$ determined by Tikhonov regularization is provided in Fig. 1.1. The unconstrained solution, which is not applicable for DEER, is given by:

$$P_{\alpha L} = (K^{\mathsf{T}}K + \alpha^2 L^{\mathsf{T}}L)^{-1} K^{\mathsf{T}}S = Q_\alpha K^{\mathsf{T}}S. \tag{1.8}$$

The matrix L defines the criterion by which P should be penalized. In the DEER literature, L has been selected based upon the assumption that physically reasonable distance distributions between spin labels on proteins are smooth on a tenths-of-nanometer scale. There are three common L choices that all encourage smoothness and penalize roughness of P, in one sense or another. The second derivative, represented by the $(n_r - 2) \times n_r$ second-order difference matrix

$$L_2 = \begin{pmatrix} 1 & -2 & 1 & & & 0 \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ 0 & & & 1 & -2 & 1 \end{pmatrix} \tag{1.9}$$

penalizes sharp turns in the distribution, which arise from sharp peaks. The first derivative,

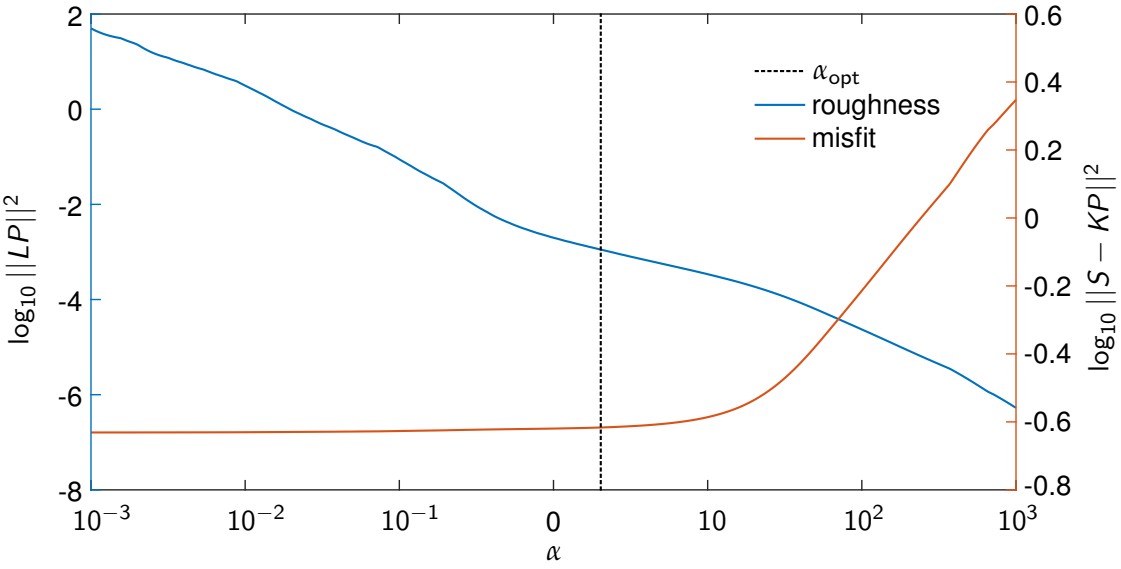

Figure 1.4: The effect of the Tikhonov regularization parameter. The 10-base logarithms of the least-squares error ($\|S - KP\|^2$) and penalty ($\|LP\|^2$) terms in Tikhonov regularization are plotted against α in orange and blue, respectively. The optimal value of α—that is, the one that best recovers the true $P(r)$—is indicated with a black dotted line.

represented by the $(n_r - 1) \times n_r$ first-order difference matrix

$$L_1 = \begin{pmatrix} -1 & 1 & & & 0 \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ 0 & & & -1 & 1 \end{pmatrix} \tag{1.10}$$

penalizes steep slopes, which are also associated with sharp peaks. Note that the matrices defined in Eqs. (1.9) and (1.10) do not contain Δr. Lastly, the $n_r \times n_r$ identity matrix $L_0 = \mathrm{I}$ can be used. It penalizes tall peaks, which tend to be narrow due to the overall normalization of P. Many other penalty matrices exist, and many more are possible. The discussion in this thesis and the associated publications is limited to these three common choices.

The value of α determines the balance between the two terms in Eq. (1.7) (Fig. 1.4). Large α values lead to oversmoothing of P and a poor fit to the time-domain data S, while small α values lead to unrealistically spiky distributions (undersmoothing) and overfitting of the data. Therefore, a proper choice of α is essential. A large number of α selection methods are described in the literature, and we include many of them in our performance analysis. They are described in

Appendix B.

The optimal values of both α and L, α_{opt} and L_{opt}, are the ones which together best recover the underlying true distribution P_0, i.e. the ones for which the model recovery error (mre), defined as

$$\text{mre}(\alpha, L) = \|P_{\alpha L} - P_0\| \tag{1.11}$$

is minimal. Other figures of merit are likely just as valid here, for example, the Kullback–Leibler divergence. For synthetic data, this error and thereby α_{opt} and L_{opt} can be determined, since P_0 is given. However, in experimental practice, P_0 is not known, and one must resort to other methods to choose a form for L and a value for α.

1.2.3 Bayes' Theorem

Uncertainty quantification in ill-conditioned, inverse problems is non-trivial and an active field of research [53–55]. Bayesian statistics provides a framework that permits inference of parameters from complex, hierarchical models. Most numerical implementations of this framework also readily provide the quantification of uncertainty in estimated parameter values with little or no additional theoretical and computational effort. Recently, the utility of Bayesian statistics has been recognized in the estimation and analysis of model parameters related to biophysical measurements [56–58].

For a set of data, X, with associated parameters, ϕ, Bayes' Theorem is an expression of conditional probability:

$$\pi(\phi|X) \propto \pi(X|\phi)\pi(\phi) \tag{1.12}$$

where the notation $\pi(\phi|X)$ means the probability of ϕ given X. The left-hand side is called the posterior probability density function (PDF) and describes the probability of the parameters given the data. The right-hand side is a product of the likelihood function, $\pi(X|\phi)$, and the prior PDF, $\pi(\phi)$. The prior encodes any previous knowledge and assumptions about the parameters.

Conveniently, there exists a well-known and intuitive connection between Bayes' Theorem and penalized least-squares methods like Tikhonov regularization: the Tikhonov functional is the negative logarithm of the posterior PDF $\pi(P|S)$. This idea is explored more fully in Ch. 3.

1.3 Thesis Outline

Ch. 2 addresses the selection of both the regularization parameter, α, and the regularization operator, L, for Tikhonov regularization of background-corrected DEER data [13]. The changes recommended in this chapter will improve the accuracy, reliability, and reproducibility of the most common method of DEER data analysis [14] by pointing to automated methods that give better model-recovery with lower rates of failure. As we found in our first foray into Bayesian hierarchical modeling [14], providing a good estimate of α is necessary to achieve good results in a statistical MCMC approach, as well.

Ch. 3 introduces a Bayesian statistical approach to infer $P(r)$ from background-corrected and modulation-depth-scaled DEER data [14], $S(t)$. The Bayesian approach exploits a well-known connection between penalized least-squares approaches and Bayes' theorem [59, 60]. A numerical Markov-chain Monte-Carlo Gibbs sampler is implemented to estimate the posterior probability distribution that provides both $P(r)$ and its uncertainty, given the data.

Ch. 4 extends the Bayesian hierarchical model from Ch. 3 to infer the intermolecular background function, $B(t)$. This permits the quantification of uncertainty for all fit parameters, including correlations between them. The increased complexity of this extended hierarchical model precludes the use of a simple Gibbs sampler, so Hamiltonian Monte Carlo is employed instead.

Chapter 2

TIKHONOV REGULARIZATION PARAMETER SELECTION

This chapter is based on a 2018 publication in the *Journal of Magnetic Resonance* by Edwards T. H. and Stoll S. titled, “Optimal Tikhonov regularization for DEER spectroscopy," volume 288, pages 58–68, available at https://doi.org/10.1016/j.jmr.2018.01.021 or https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5840305/ [13].

Tikhonov regularization is the most commonly used method for extracting distance distributions from experimental double electron-electron resonance (DEER) spectroscopy data. This method requires the selection of a regularization parameter, α, and a regularization operator, L. We analyze the performance of a large set of α selection methods and several regularization operators, using a test set of over half a million synthetic noisy DEER traces. These are generated from distance distributions obtained from *in silico* double labeling of a protein crystal structure of T4 lysozyme with the spin label MTSSL. We compare the methods and operators based on their ability to recover the model distance distributions from the noisy time traces. The results indicate that several α selection methods perform quite well, among them the Akaike information criterion and the generalized cross validation with either the first- or second-derivative operator. They perform significantly better than currently utilized L-curve methods.

2.1 The synthetic test data set

In order to test α selection methods and regularization operators on DEER data from practically relevant distance distributions, we numerically generated a large set of synthetic noisy DEER time-domain signals based on a crystal structure of T4 lysozyme (PDB ID 2LZM, 1.7 Å resolution, https://www.rcsb.org/structure/2LZM); its structure is shown in Fig. 2.1. We use this protein since it is currently one of the most thoroughly investigated proteins by DEER and other EPR techniques [61–69] and is of a size commensurate with the typical DEER distance range of 1.5 to 6 nm. We *in silico* labeled the 2LZM protein structure with the nitroxide spin label MTSSL which is to date, by a wide margin, the most commonly used label for DEER. In this section, we detail the construction of the test set.

Construction of model distributions. Using scripts based on MMM 2017.1 [70–72] with the default 216-member rotamer library R1A_298K_UFF_216_r1_CASD for the MTSSL ((1-Oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl) Methanethiosulfonate) spin-label at ambient temperature, we calculated rotamer distributions for all 164 sites on the protein. For each site, this

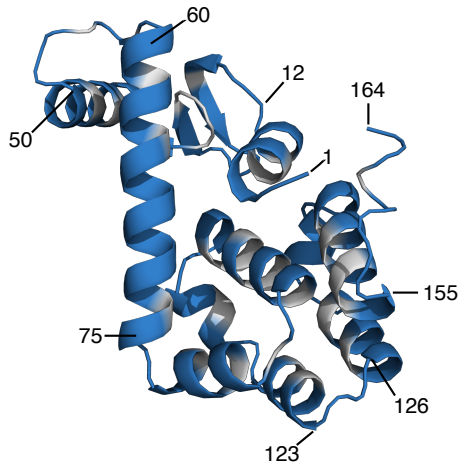

Figure 2.1: Crystal structure of T4 lysozyme (164 residues, PDB 2LZM, 1.7 Å resolution). Residues with low predicted labeling probabilities that were excluded from the test set are shown in grey. Several residues are labeled as guides to the eye.

yielded coordinates of the mid-points of the N-O bonds of all 216 rotamers and their associated populations, as well as the site partition function. In order to keep the test set realistic, we eliminated all sites with low predicted labelling probability (partition function smaller than 0.05), leaving 129 sites. These sites are indicated in Fig. 2.1.

For each of the resulting 8256 site pairs, we calculated the associated model distance distribution, consisting of a sum of Gaussian lineshapes, with centers corresponding to the inter-rotamer distances, intensities corresponding to the products of the respective populations, and with a uniform full width at half maximum (FWHM) of 0.15 nm to account for structural heterogeneity around the energy minimum of each rotamer due to librational motion [70]. Each distribution was discretized with a high resolution of Δr = 0.005 nm, corresponding to 1/30 of the FWHM. To keep the test set experimentally realistic, we discarded all distributions with more than 5% integrated population below 1.7 nm. This guarantees that there is no population below 1.5 nm. Site pairs with population at such short distances are always avoided in experimental studies in order to avoid potential distortions due to exchange coupling and incomplete excitation.

This procedure resulted in a set of 5622 model distance distributions. Their properties are summarized in Fig. 2.2. The distribution modes (location of the distribution maxima) range from 1.7 to 6.1 nm, and the inter-quartile ranges (iqr; range of central 50% integrated population) are between 0.09 and 1.1 nm wide. Long, short, narrow, wide, unimodal, multimodal, symmetric, and skewed distributions are all well-represented in the test set. Skewness is relatively evenly

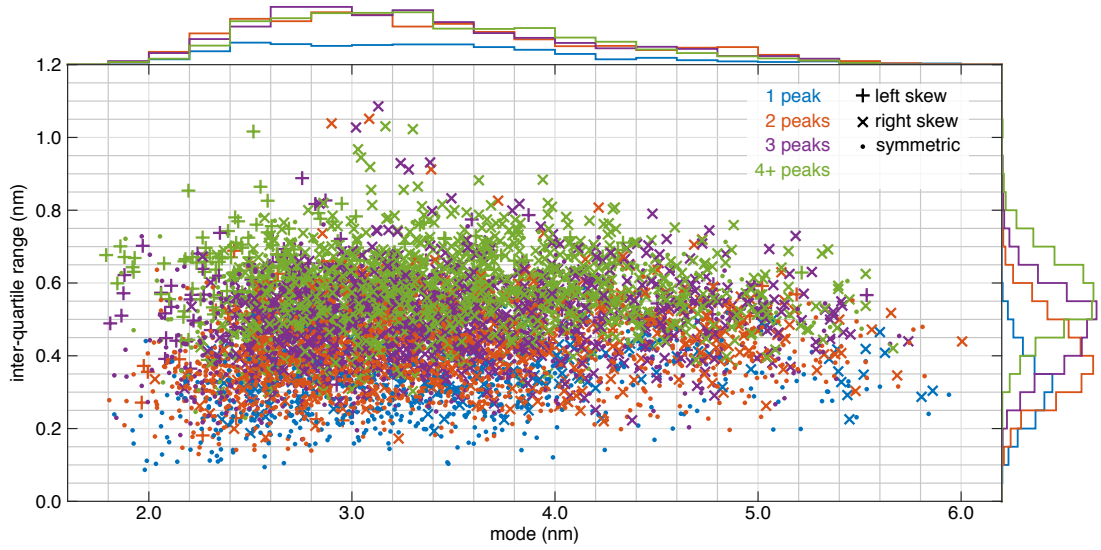

Figure 2.2: Summary statistics of the 5622 model DEER distance distributions derived from the crystal structure of T4 lysozyme (PDB 2LZM). The inter-quartile range (iqr) of r is plotted vs the mode of r for each distribution. Color indicates the number of peaks (significant maxima), and the symbol indicates the type and degree of skew. The histogram colors correspond to the number of peaks.

distributed across both the mode and iqr. The number of significant peaks is evenly distributed across modal distance, but there is a positive correlation between iqr and number of peaks. Fig. 2.3 shows several representative example distributions.

An alternative strategy of constructing a test set would be to generate completely artificial distributions without reference to a protein where the center, width, multi-modality, and skewness is varied either systematically or randomly. However, such a test set will be less experimentally relevant. Figure 2.2 shows that the structure-based test set used here covers a sufficiently diverse range of distribution centers, widths, and shapes.

Generation of time-domain traces. Each model distribution was used to generate a set of time-domain traces S with varying maximum evolution times t_{max}, time increments Δt, noise levels σ, and noise realizations, subject to the constraints that: (1) each S must be long enough to permit accurate recovery of the longest distances in the corresponding P, and (2) that the sampling rate must be high enough to recover the shortest distances. The starting time for

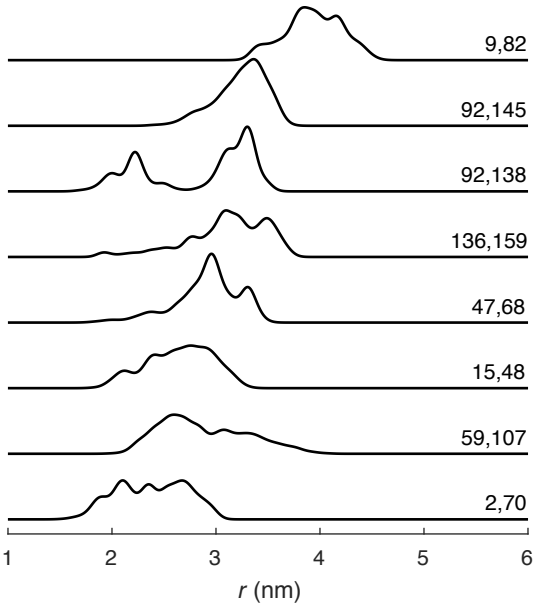

Figure 2.3: Examples of distance distributions calculated via *in silico* double labeling of the 2LZM structure of T4 lysozyme with MTSSL. The site pairs corresponding to each distribution are listed in the figure.

all time traces was set to $t_{\mathrm{min}} = 0$. We used multiple values of t_{max} and Δt for each model distribution in order to be able to assess whether or not any α selection methods display uneven performance with respect to these parameters.

To choose a set of appropriate t_{max} for a given distribution, we take as an upper limit $t_{\mathrm{mm}} = 3T_{\perp,95}$, where $T_{\perp,95}$ is the period corresponding to the perpendicular-orientation dipolar frequency of the 95-percentile distance, i.e., $T_{\perp,95} = (r_{95}/\mathrm{nm})^3/(52\,\mathrm{MHz})$. The factor 3 prolongs the time in order to completely capture the long-distance tail of the distribution, which has the lowest frequencies. With this, we choose the set of t_{max} according to the following procedure: if $t_{\mathrm{mm}} > 6.4\,\mu\mathrm{s}$, then t_{max} is 6.4 and 3.2 µs; if $3.2\,\mu\mathrm{s} < t_{\mathrm{mm}} < 6.4\,\mu\mathrm{s}$, then t_{max} is 3.2 and 1.6 µs; if $1.6\,\mu\mathrm{s} < t_{\mathrm{mm}} < 3.2\,\mu\mathrm{s}$, then t_{max} is 1.6 and 0.8 µs; if $0.8\,\mu\mathrm{s} < t_{\mathrm{mm}} < 1.6\,\mu\mathrm{s}$, then t_{max} is 0.8 µs; and if $t_{\mathrm{mm}} < 0.8\,\mu\mathrm{s}$, then t_{max} is 0.4 µs. Figure 2.4 shows an example distribution for which $t_{\mathrm{mm}} = 2.36\,\mu\mathrm{s}$, hence t_{max} is 1.6 and 0.8 µs. Note that even the shorter t_{max} value of 0.8 µs is long enough to accommodate a full oscillation of period $T_{\perp,95} = 0.786\,\mu\mathrm{s}$.

To find a reasonable set of time increments Δt for each distribution, we first determined the longest allowable increment, Δt_{max}, that correctly samples all frequencies in the signal. This is dictated by the highest dipolar frequency in the time-domain signal, which corresponds to the shortest distance in the distribution. We use the period $T_{\parallel,05}$ of the parallel-orientation dipolar frequency corresponding to the 5-percentile distance. According to the Nyquist theorem, Δt_{max} must be less than $T_{\parallel,05}/2$ to avoid frequency aliasing. To capture the short-distance tail of the distribution and to obtain enough time-domain points to achieve reasonable distance resolution in the Tikhonov fits in all cases, we chose $\Delta t_{\mathrm{max}} = T_{\parallel,05}/6$. This resulted in values from 8 ns to 280 ns. Δt_{max} now guides the choice of Δt: If $\Delta t_{\mathrm{max}} > 200\,\mathrm{ns}$, then Δt is 50, 100, and 200 ns; if $100\,\mathrm{ns} < \Delta t_{\mathrm{max}} < 200\,\mathrm{ns}$, then Δt is 20, 50, and 100 ns; if $50\,\mathrm{ns} < \Delta t_{\mathrm{max}} < 100\,\mathrm{ns}$, then Δt is 10, 20, and 50 ns; if $20\,\mathrm{ns} < \Delta t_{\mathrm{max}} < 50\,\mathrm{ns}$, then Δt is 10 and 20 ns; if $10\,\mathrm{ns} < \Delta t_{\mathrm{max}} < 20\,\mathrm{ns}$, then Δt is 10 ns; and if $\Delta t_{\mathrm{max}} < 10\,\mathrm{ns}$, then Δt is 5 ns. For the example distribution in Fig. 2.4, $\Delta t_{\mathrm{max}} = T_{\parallel,05}/6 = 31.3\,\mathrm{ns}$, so Δt was set to 10 and 20 ns.

To generate the noise-free time trace for a particular combination of t_{min}, t_{max}, and Δt (with n_t = 33, 41, 65, 81, 129, 161, 321, or 641 points), we used in Eq. (1.6) the high-resolution model distribution ($n_0 = 1341$ points) and generated the n_t-element time trace using an $n_t \times n_0$ kernel matrix. Therefore, the model distributions used for this forward modeling have a higher resolution than the ones obtained via Tikhonov regularization (Eq. (1.7)) of the simulated noisy data, with $n_r = n_t$. This avoids circular reasoning and committing the "inverse crime" [59, 73]. Fig. 2.4 shows the synthetic time-domain traces resulting from the above rules for an example distribution.

The combinations of Δt, t_{min}, and t_{max} resulted in a total of 20701 noise-free time-domain traces with between 33 and 641 points. A few traces with fewer than 33 points were thrown out, since such short traces are not acquired in practice.

Finally, we generated 30 noisy traces from each noise-free trace. We utilized uncorrelated Gaussian noise—which is experimentally observed in DEER measurements [14]—with standard deviations $\sigma = 0.02$, 0.05, and 0.1. These noise levels span the range typically obtained

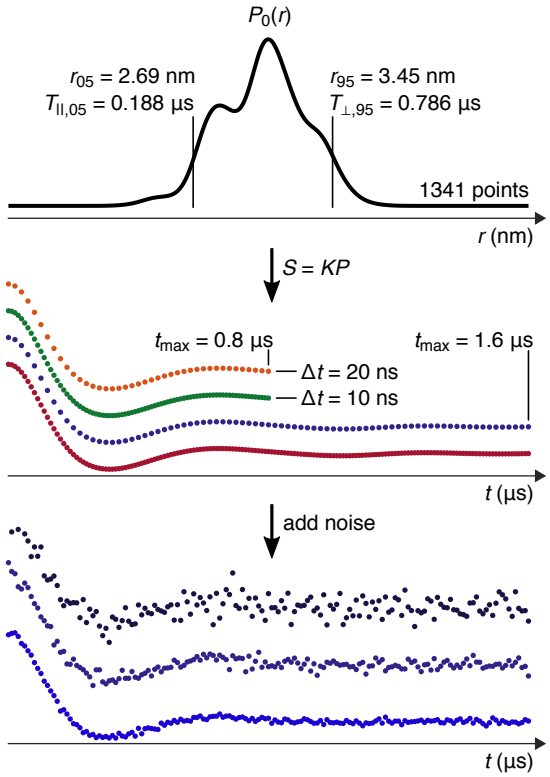

Figure 2.4: Generation of time traces ($S(t)$) from a distance distribution. Top panel: the model distribution (for labeled residues 112 and 143). In this case, $3T_{\perp,95} = 2.36$ µs, so t_{max} was set to 1.6 and 0.8 µs in accordance with the rules given in Section 2.1. $T_{\parallel,05}/6 = 31.3$ ns, so Δt was set to 10 and 20 ns. Middle panel: the four resulting time-domain traces generated using Eq. (1.6). Bottom panel: three noisy realizations with σ = 0.1, 0.05, and 0.02 for the 3rd (dark blue) trace in the middle panel. A total of 120 noisy time-domain traces were generated from this particular distribution.

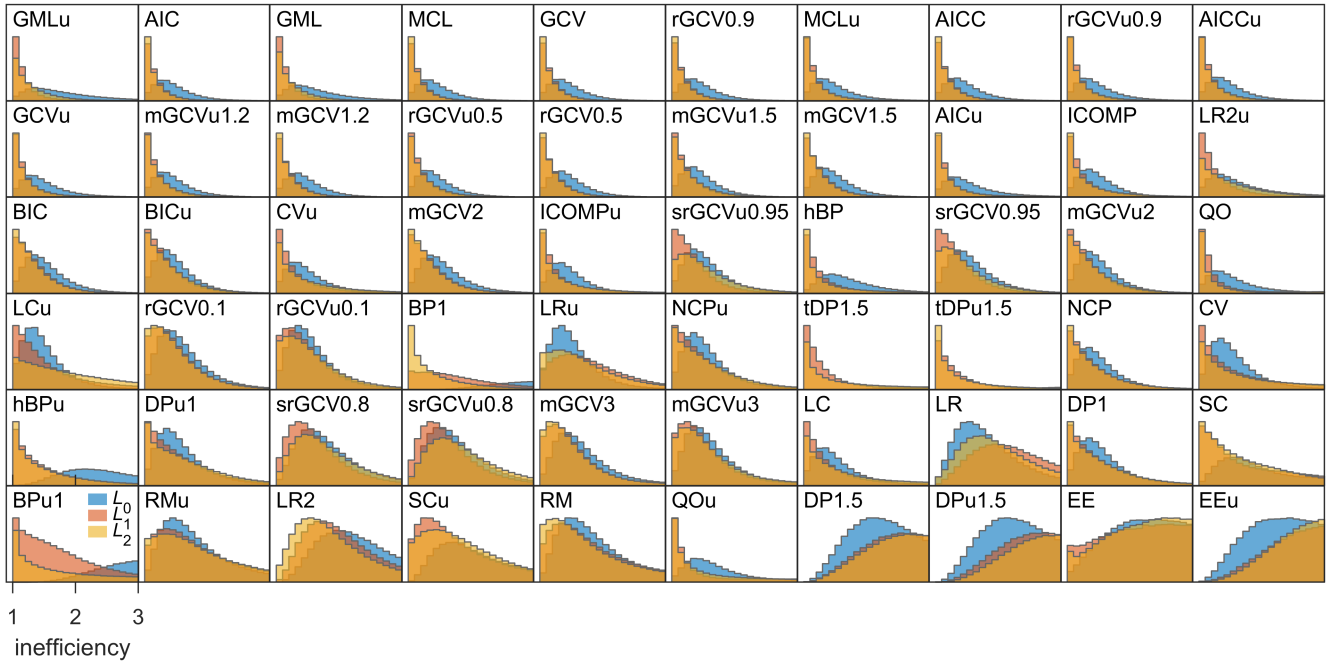

Figure 2.5: Inefficiency distributions for each α selection method (indicated by their acronyms) and regularization operator (color-coded). The panels are arranged from top left to bottom right in increasing order of lowest 75^{th}-percentile inefficiency among L_0, L_1, and L_2. For each histogram, only the part with inefficiency ≤ 3 is shown, and the often prolonged tails are not visible. The histograms are scaled such that the maxima in all panels are identical.

experimentally for proteins. In experimental settings, the background removal step typically increases the magnitude of the noise at the end of the $S(t)$ trace relative to the start. This effect is more pronounced for steeper background functions and can be negligible for shallower ones. We did not attempt to simulate this characteristic of the data. For each noise level, $m = 10$ noise realizations were generated using a random-number generator. For reproducibility, seeds for the random-number generator are stored as part of the test set.

Altogether, with these selections for t_{max}, Δt, σ, and m, the 5622 model distributions P_0 resulted in 621030 noisy time-domain traces S. These constitute the final test data set, which is available for download at https://doi.org/10.6069/H5S75DCG.

2.2 Performance evaluation

For each of the traces in the test set, we use 60 different variants of α selection methods to determine an appropriate regularization parameter value α_{sel} and the associated Tikhonov solution. Mathematical details of the methods are given in the Appendix. Each method is referred to via a short acronym, which is listed in Table 2.1. We evaluate each α selection method for each of the three regularization operators L (identity, first derivative, and second derivative). Thus, we examine Tikhonov regularization along both its degrees of freedom (L and α), with a total of 180 approaches.

Each method determines α by minimizing or thresholding some cost function over α. To search for these points, we use an α range from 10^{-3} to 10^{3} with 61 points on a logarithmic scale, i.e. 10 points per decade (1 dB increments). All methods require the data S as input and need to calculate $P_{\alpha L}$ over the entire α range for each L. For this, $P_{\alpha L}$ can be determined with or without the constraint $P \geq 0$ in Eq. (1.7), and we test both variants in all cases. Several methods require knowledge of the time-domain noise standard deviation, σ. The cost functions of a few methods are equipped with a tuning or scaling parameter that can be adjusted to increase stability.

The analysis procedure also requires the choice of a distance range and resolution for $P_{\alpha L}$. We use a range of 1.0 nm to 7.0 nm for all cases. These limits encompass the full ranges of all model distributions in the test set. The number of points in the distance domain, n_r, is determined separately for each case and is set equal to the number of points in the time-domain trace, n_t.

Since the goal of analyzing the time-domain data is to reveal the underlying model distribution, we base our performance evaluation of the various selection methods not on α (which is a nuisance parameter whose numerical value is physically irrelevant), but on the model recovery error defined in Eq. (1.11). It quantifies how close the calculated Tikhonov solution, $P_{\alpha L}$, for a given α and L is to the model distribution P_0. This error is generally non-zero, since noise in the data and the fundamental ill-posedness of the problem prevent full recovery of the model from the data, no matter which α and L are used. As the actual performance measure, we use the

inefficiency [74]

$$\text{inefficiency} = \frac{\text{rmsd}(\alpha, L)}{\text{rmsd}(\alpha_{\text{opt}}, L_{\text{opt}})} \tag{2.1}$$

with $\text{rmsd}(\alpha, L) = \text{mre}(\alpha, L)/\sqrt{n_r} = \|P_{\alpha,L} - P_0\|/\sqrt{n_r}$ (Eq. 1.11). This measure relates the model recovery error for a given L and the particular α value chosen by a selection method to the smallest possible model recovery error, obtainable with α_{opt} and L_{opt}. This error, and the associated optimal solution P_{opt} are found by minimizing the mre as a function of α and L. The inefficiency equals 1 when $P_{\alpha L} = P_{\text{opt}}$ and is greater otherwise. The closer the inefficiency is to 1, the better the method.

2.3 Results and discussion

Histograms. Figure 2.5 shows an overview of the calculated inefficiencies for all 621030 time traces, 60 α selection methods, and three regularization operators. Each histogram shows the distribution of inefficiencies for a particular combination of method and operator. The methods are sorted by increasing order of lowest 75$^{\text{th}}$-percentile inefficiency from top left to bottom right. There is a wide range of performances. Most methods in the top row achieve low inefficiencies that are sharply peaked at 1. Methods shown lower down have a propensity towards worse inefficiencies. The high-inefficiency tails of the histograms are not shown, but are protracted in many cases, indicating a not insignificant failure rate for those methods. It is apparent from the histograms that the L_0 operator, shown in blue, tends to perform worse than the others for the same method, except for some with overall bad performance. For most methods, L_1 and L_2 appear to give similar results.

Performance comparison. To compare the method/operator combinations more quantitatively, we use the 99$^{\text{th}}$ percentile from each inefficiency histogram as the metric. This stringent choice is motivated by the consideration that a method/operator combination can be deemed good and safe if it delivers overall low inefficiencies with negligible risk of large inefficiencies (i.e. severe under- or oversmoothing). Figure 2.6 shows the results. There is a group of methods that perform similarly well, leading to small 99$^{\text{th}}$-percentile inefficiencies just above 2. Mallows' C_L (MCL), the Akaike information criterion (AIC) and the generalized cross validation (GCV) perform equally well with L_1 and L_2, whereas the generalized maximum likelihood method and its unconstrained variant (GML and GMLu) perform well only with L_1. L_0 is the worst-performing operator for these top methods. The methods ranking below this top group are parameterized modifications of GCV and AIC. Since they are inferior to their parent methods, they can be disregarded. Currently, the most commonly applied α selection methods are based on the L-curve. The results show that the L-curve methods LC, LCu, LR, LRu, LR2, LR2u are not competitive with the top group. Their overall tendency is to oversmooth. Similarly, the SC and SCu methods

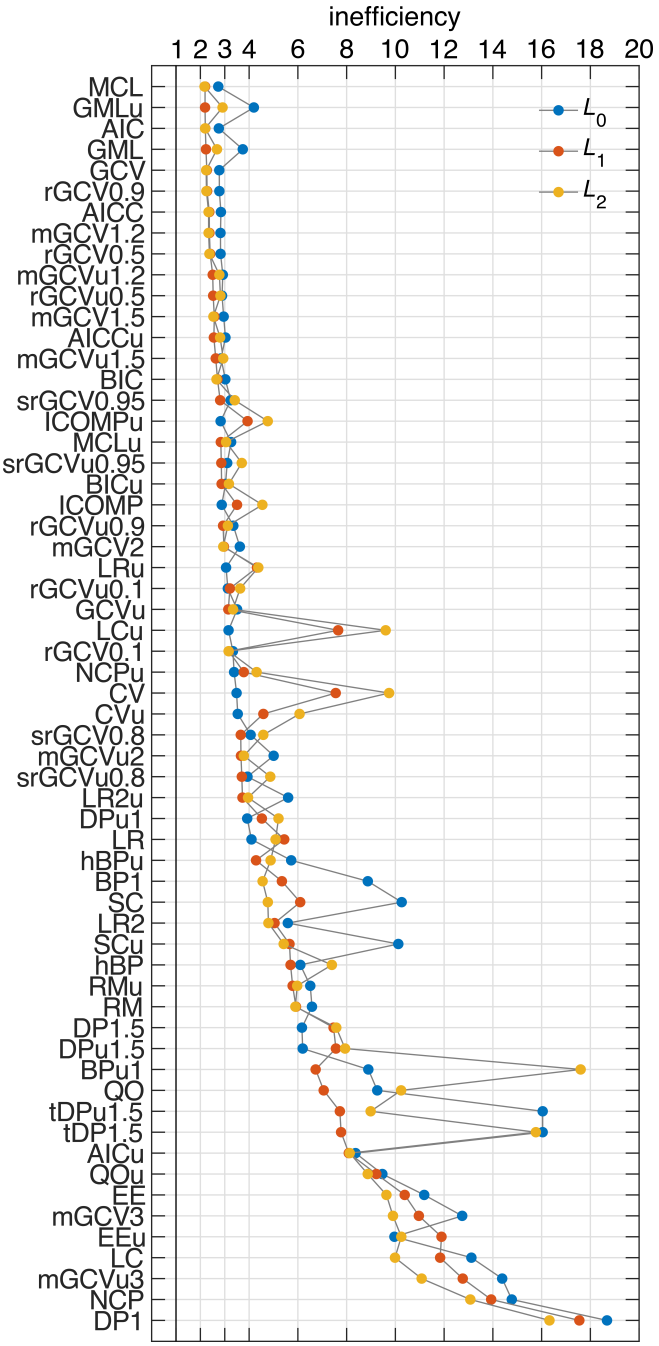

Figure 2.6: Performance comparison of all combinations of α selection methods and regularization operators, based on their 99$^{\text{th}}$-percentile inefficiency. The methods are sorted based on best 99$^{\text{th}}$-percentile inefficiency among L_0, L_1, and L_2. The gray lines serves as guides to the eye.

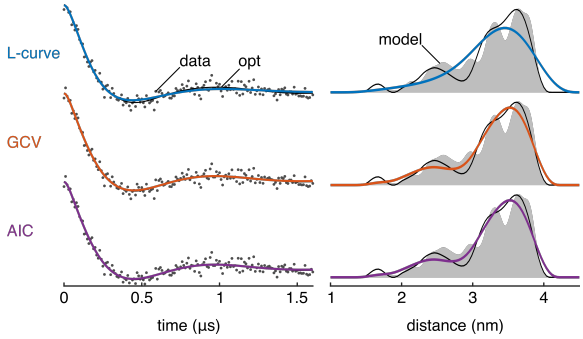

Figure 2.7: Example case comparing the LR2u L-curve, GCV, and AIC methods (dataset 115333). For this particular test case, each method features an inefficiency very close to that method's overall median inefficiency (LR2u 1.38, GCV 1.09, AIC 1.09). Left column: the simulated data are shown as grey dots, the Tikhonov fit with optimum α is shown as a black line, and the Tikhonov fits for LR2u, GCV, and AIC are shown as blue, orange, and purple lines, respectively. Right column: the model P is shown in grey, the Tikhonov solution for P with optimum α is shown as a black line, and the Tikhonov Ps for LR2u, GCV, and AIC are shown as blue, orange, and purple lines, respectively. In all cases, the L_2 operator was used.

(used in `FTIKREG`) display elevated inefficiency for DEER. The computational cost of all methods is essentially identical, since it is dominated by the solution of the Tikhonov minimization problem for each α.

Due to the diversity of distribution shapes and time traces, and due to the wide range of inefficiencies, it is impossible to visualize method performances with a few sample datasets. Nevertheless, for the sake of illustration, Figure 2.7 shows the relative performance of LR2u, GCV, and AIC for a typical test dataset. This example is typical in these sense that the inefficiency of each method is very close to the method's median inefficiency. The figure shows that the L-curve has a tendency to oversmooth, and that AIC and GCV perform similarly well. Fig. 2.8 illustrates the uncertainty in the extracted $P(r)$ based on these α selection methods, quantified using Bayesian inference [14]. The Bayesian statistical methodology used to quantify this uncertainty is introduced in Ch. 3.

Sensitivity to metric. The performance comparison shown in Fig. 2.6 is based on our particular choice of metric, the 99$^{\text{th}}$ percentile of the ratio of rmsds, defined in Eq. (2.1). Varying the percentile to 90, 75, or 50 does not significantly affect the composition of the top group, although it affects the detailed rankings. Altering the definition of inefficiency in Eq. (2.1) by using the difference instead of the ratio, or by using the maximum absolute deviation (mad) instead of the rmsd, yields similar results. The online supplementary material for the paper upon which

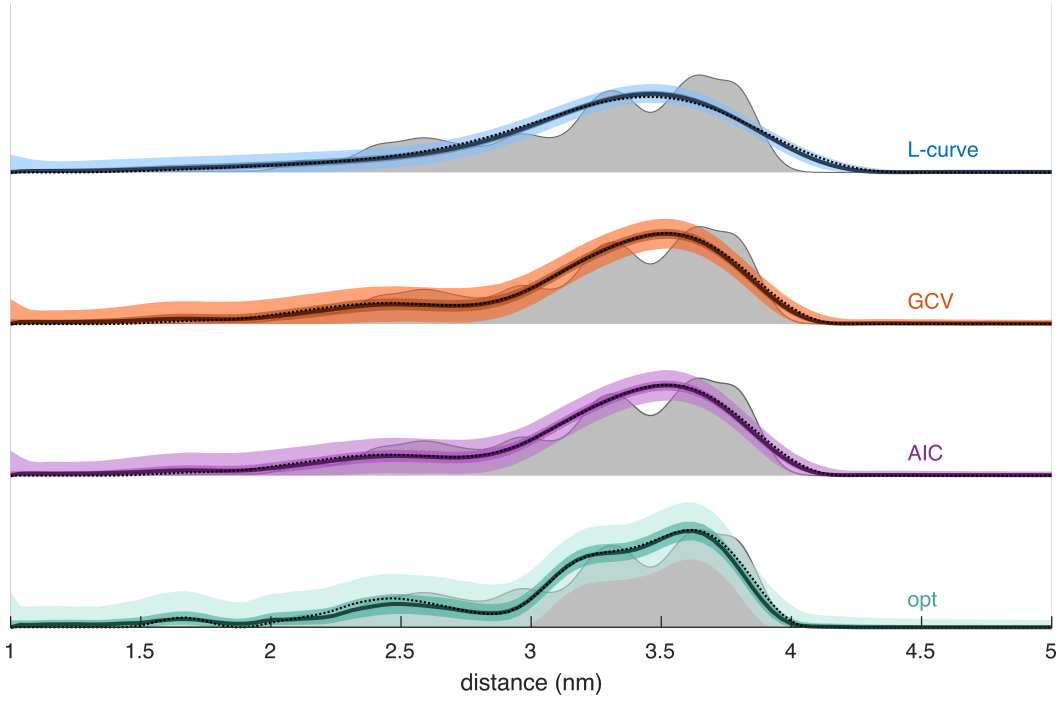

Figure 2.8: Bayesian credible intervals for the case shown in Fig. 2.7. In each plot, the model $P(r)$ is shown in grey and the corresponding Tikhonov solution is in dotted black. The L_2 operator was used in each case.

this chapter is based contains an excel spreadsheet demonstrating these facts [13]. Therefore, we conclude that the identity of the top methods is insensitive towards the particular choice of ranking metric, and that our assessment procedure is overall robust.

Sensitivity to data characteristics. To check for uneven performance for certain subsets of the test set, we examined the performance across noise level, the ratio of iqr to mode (a measure of the damping rate of oscillations in S), number of points, and number of modal periods, using the 99$^{\text{th}}$ percentile rmsd ratio inefficiency metric. The breakdowns are available in the online supplementary material mentioned in the last paragraph [13]. For all values of these variables, MCL, AIC, GCV, GML, and GMLu are usually within 5% of the optimal method/operator combination and never deviate by more than 11%. The specific rankings vary between subsets, but actual performance changes to such a small degree that this is not significant. The conclusions reached above regarding the three operators generalize to the subsets considered in the online supplementary excel spreadsheet mentioned previously [13]. In addition to poor performance relative to L_1 and L_2, L_0 also displays much higher variation in performance across test subsets. Several variations on GCV (mGCV, mGCVu, rGCV, and rGCVu) with varied tuning parameter values, as well as AICC, come within 10% of the best case for many of the test subsets. However, none display sufficient consistency or quality of performance to surpass MCL, AIC, GCV, GML, or GMLu.

Since there is no correlation between the performance of any top method and these subset characteristics, the results are therefore likely to be applicable to situations with different relative representations of distance distribution characteristics, such as other spin labels or proteins with more β sheets than T4 lysozyme.

Method choice. The performance differences among the top method/operator combinations are so small that we cannot identify one as an evident best choice. However, additional considerations can provide some guidance. MCL requires the time-domain noise variance σ^2 as an input. An under- or overestimation of σ^2, which is likely in experimental settings, will affect the performance and likely degrade it. GML and GMLu depend on a thresholding value to remove eigenvalues close to zero, and the choice of this value can affect the performance. In contrast, AIC and GCV do not require *a priori* knowledge of σ^2 nor do they depend on thresholding or tuning parameters. Therefore, the parameter-free AIC and GCV methods with either L_1 or L_2 appear to be the simplest, best, and safest choices for practical applications.

2.4 Conclusions

Our performance analysis of a large number of regularization parameter selection methods over a large set of protein-based synthetic DEER data indicates that there are several method/operator combinations that perform equally well. Among these, the Akaike information criterion (AIC)

and the generalized cross validation (GCV) are preferable, as they are parameter-free and do not require accurate knowledge of the noise level. They work equally well with the first- and second-derivative operator, but not with the identity. L-curve methods, some of which are currently widely employed, perform distinctly worse.

The structure-based test set developed in this work is useful beyond Tikhonov regularization, as it can be used to assess the performance of other existing analysis methods (truncated singular-value decomposition, Gaussian mixture models) and of any new solution approaches. It is available online at https://doi.org/10.6069/H5S75DCG.

Acknowledgements

This work was supported by the National Science Foundation with grant CHE-1452967 (S.S.) and by the National Institutes of Health with grants R01 EY010329 (S.S.) and T32 GM008268 (T.H.E.). This work was further facilitated though the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system funded by the STF at the University of Washington.

Table 2.1: List of abbreviations for regularization parameter selection methods. To indicate the use of unconstrained $P(r)$ in a method, the suffix 'u' is appended to the abbreviation. If a tuning parameter is used, its value is appended as well.

Acronym	Full name
AIC	Akaike information criterion
AICC	corrected Akaike information criterion
BIC	Bayesian information criterion
BP	balancing principle
CV	leave-one-out cross validation
DP	discrepancy principle
EE	extrapolated error
GCV	generalized cross validation
GML	generalized maximum-likelihood
hBP	hardened balancing principle
ICOMP	information complexity criterion
LC	L-curve, maximum curvature
LR	L-curve, minimum radius
LR2	L-curve, minimum radius 2
MCL	Mallows' C_L
mGCV	modified generalized cross validation
NCP	normalized cumulative periodogram
QO	quasi-optimality criterion
rGCV	robust generalized cross validation
RM	residual method
SC	self-consistency method
srGCV	strong robust generalized cross validation
tDP	transformed discrepancy principle

Chapter 3

BAYESIAN MODEL TO QUANTIFY UNCERTAINTY FROM NOISE AND REGULARIZATION

This chapter is based on a 2016 publication in the *Journal of Magnetic Resonance* by Edwards T. H. and Stoll S. titled "A Bayesian approach to quantifying uncertainty from experimental noise in DEER spectroscopy," volume 270, pages 87–97, available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4996738/ and https://doi.org/10.1016/j.jmr.2016.06.021 [14].

3.1 Introduction

The most common method for inferring distance distributions, $P(r)$, from experimental DEER data is Tikhonov regularization. One disadvantage of this method is that it does not directly provide error bars for the resulting $P(r)$, rendering correct interpretation difficult. Here we introduce a Markov-chain Monte-Carlo (MCMC) method based on Bayesian statistics that quantifies uncertainty in $P(r)$ arising from noise and numerical regularization. This method provides credible intervals (error bars) for α and $P(r)$ at each r. This allows practitioners to answer whether or not small features are significant, whether or not apparent shoulders are significant, and whether or not two distance distributions are significantly different from each other. In addition, the method quantifies uncertainty in the regularization parameter. I introduce the basic probabilistic equations, show their connection to Tikhonov regularization, and describe an MCMC algorithm for generating representative numerical samples of P and α. The methodology is demonstrated on a few test cases.

3.2 The posterior distribution

We start by considering what is observable and given, and what is unknown, uncertain, and wanted. Here, we consider the intramolecular dipolar modulation function $S(t)$ observable and given. $S(t)$ is obtained from the experimental data $V(t)$ after background correction. It is an inseparable sum of the true signal and corrupting noise. A second useful quantity that can be obtained from experiment is the noise variance. We obtain it experimentally by acquiring each phase-cycled DEER trace separately and then statistically analyzing the distribution of values for each t_i. As shown in Appendix C, this noise is Gaussian with amplitude roughly independent of t_i. The noise deviations at adjacent time points are not correlated. Therefore,

we can characterize the noise by a single noise variance, σ_S^2. Note that for steep background functions, the noise at the end of the $S(t)$ vector will be artificially inflated relative to the start of the trace. For notational convenience, we will henceforth use its inverse, called the precision

$$\tau = 1/\sigma_S^2. \tag{3.1}$$

Next, we consider the unknown and uncertain quantities. These include the desired underlying distance distribution P. However, since the problem is not solvable without regularization, a regularization parameter is required as well. There is no absolute certainty about its optimal value. Here, we will use the symbol δ and show later that δ is proportional to α^2.

To summarize, S and τ are observable and fixed, and P and δ are unknown and desired. Our goal is to fully characterize the conditional probability distribution for P and δ, given the measured S and τ: $\pi(P, \delta|S, \tau)$. First, we construct an expression for $\pi(P, \delta|S, \tau)$. Like any conditional probability, it is proportional to the full joint probability

$$\pi(P, \delta|S, \tau) = \frac{\pi(P, \delta, S, \tau)}{\pi(S, \tau)} \propto \pi(P, \delta, S, \tau) = \pi(S, P, \delta, \tau). \tag{3.2}$$

In the second step we omitted the marginal probability $\pi(S, \tau)$, since it is a constant independent of P and δ. In the last step, we reordered the variables. This joint probability can be written as a chain product of four probabilities

$$\pi(S, P, \delta, \tau) = \pi(S|P, \delta, \tau) \cdot \pi(P|\delta, \tau) \cdot \pi(\delta|\tau) \cdot \pi(\tau). \tag{3.3}$$

Next, we introduce specific expressions for the factors on the right-hand side. The first term indicates the conditional probability of a certain S given P, δ and τ. For this, we can use a Gaussian with the weighted sum-of-squares-deviation in the exponent

$$\pi(S|P, \delta, \tau) = \pi(S|P, \tau) = \left(\frac{\tau}{2\pi}\right)^{n_t/2} \exp\left(-\frac{\tau}{2}\|S - KP\|^2\right). \tag{3.4}$$

(We omit δ in the notation, since the expression does not directly depend on it). This expression, which constitutes the likelihood model, captures the essence of least-squares fitting: if P is such that KP fits S well, the magnitude of the exponent is small and the probability large. On the other hand, if the fit is poor, the exponent is large in magnitude, and the probability is small (Fig. 3.1).

The second term in Eq. (3.3) is the conditional probability of a certain P given values for δ and τ. For this, we choose a Gaussian smoothness requirement:

$$\pi(P|\delta, \tau) = \pi(P|\delta) = \left(\frac{\delta}{2\pi}\right)^{n_r/2} \exp\left(-\frac{\delta}{2}\|LP\|^2\right) \tag{3.5}$$

where L is the same operator as in Tikhonov regularization (commonly the second derivative) and δ determines the roll-off rate of the probability as a function of the roughness $\|LP\|^2$. This expression assigns higher probabilities to smoother P, as illustrated in Fig. 3.1.

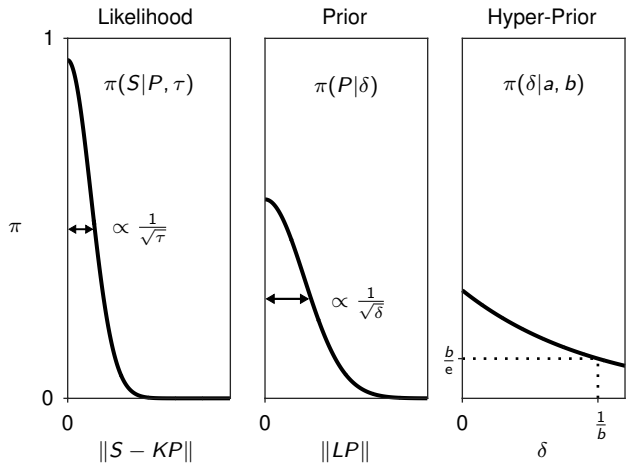

Figure 3.1: Illustration of the factors composing the posterior distribution: the likelihood, prior, and hyperprior. The likelihood function (left) is a Gaussian in $\|S - KP\|$, the fit error. Its width is determined by τ. The prior (middle) is also a Gaussian, this time in $\|LP\|$, the roughness. It biases towards zero roughness with width determined by δ. The hyper-prior (right) is a gamma distribution with the form of a slowly decaying exponential, given $a = 1$ and small b.

The third term in Eq. (3.3) is the probability of a given δ. Since we are uncertain about the value of δ, we use a very broad distribution. For mathematical convenience, we choose a gamma distribution [54]

$$\pi(\delta|\tau) = \pi(\delta|a,b) = \frac{b^a}{\Gamma(a)}\delta^{a-1}\exp(-b\delta) \tag{3.6}$$

where we omitted τ because it is independent of δ. The fixed parameters a and b determine the shape and width of the distribution such that its mean is a/b and its variance is a/b^2. We use $a = 1$, in which case the expression simplifies to a decaying exponential, $b\exp(-b\delta)$ (see Fig. 3.1). We discuss the selection of b in Section 3.4.

Finally, the last term in Eq. (3.3), $\pi(\tau)$, is a constant, since τ is known from experiment. We can therefore drop it and use the experimental τ value.

To summarize, our target expression for the conditional probability for P and δ, given S and τ, is

$$\pi(P,\delta|S,\tau) \propto \pi(S|P,\tau)\cdot\pi(P|\delta)\cdot\pi(\delta|a,b) \tag{3.7}$$

with the right-hand side quantities given by Eq. (3.4), Eq. (3.5) and Eq. (3.6) and illustrated in Fig. 3.1. This expression is a form of Bayes' law and constitutes a Bayesian hierarchical model. $\pi(P,\delta|S,\tau)$ is called the posterior PDF, or simply the posterior. $\pi(S|P,\tau)$ is called the likelihood and describes how likely it is to measure a certain S given a particular P and τ.

$\pi(P|\delta)$ is a called the prior PDF, or simply the prior. It describes our knowledge about P prior to the acquisition of the data (i.e. P is not rough). $\pi(\delta|a,b)$ is a hyper-prior which describes prior knowledge about δ. The reason we use a gamma distribution in δ is that this makes the total expression a gamma function in δ. Eq. (3.6) is therefore called a conjugate hyper-prior [53]. Similarly, the Gaussian prior in P is conjugate with the likelihood, since the total expression is also Gaussian in P.

3.3 Connection to Tikhonov regularization

The particular P that maximizes the posterior probability is the most likely P and is called the maximum-a-posteriori solution: $P_{\text{MAP}} = \max_{P\geq 0} \pi(P,\delta|S,\tau)$. The same P can be obtained by minimizing the negative logarithm of that probability: $\max_{P\geq 0} \pi(P,\delta|S,\tau) = \min_{P\geq 0}(-\log \pi(P,\delta|S,\tau))$. Dropping all P-independent prefactors in $\pi(P,\delta|S,\tau)$, taking the negative logarithm, and dividing by $\tau/2$ gives

$$-\log\pi(P,\delta|S,\tau) \propto \|S-KP\|^2 + \frac{\delta}{\tau}\|LP\|^2 . \tag{3.8}$$

This is identical to the Tikhonov functional in Eq. 1.7 if [53, 75, 76]

$$\alpha^2 = \frac{\delta}{\tau} \tag{3.9}$$

Consequently, the P_α obtained by Tikhonov regularization is equal to P_{MAP}, for a given δ (or α).

3.4 Gibbs sampling

In order to obtain visually interpretable results, we need to compute the average and other measures of P over the distribution given in Eq. (3.7). Due to the multivariate and constrained nature of the expression, this is not possible analytically. Therefore, we need to resort to numerical sampling techniques to generate a set of P and δ such that any expectation (e.g. mean) over this set approximates well the corresponding analytical expectation value over the entire posterior distribution, $\pi(P,\delta|S,\tau)$.

For this, we use the Gibbs sampler, which is a Markov-chain Monte-Carlo (MCMC) algorithm for drawing numerical samples from a multivariate probability distribution [60]. The Gibbs sampler generates samples of P and δ from $\pi(P,\delta|S,\tau)$ by successively sampling from the posterior of each variable (P and δ) in turn. In our case, the necessary posteriors are $\pi(P|\delta,S,\tau)$ and $\pi(\delta|P,S,\tau)$.

To obtain $\pi(P|\delta,S,\tau)$, we take Eq. (3.7) and omit terms that do not contain P. As shown in Appendix C, the resulting expression gives a multivariate normal distribution in P

$$\pi(P|\delta,S,\tau) \propto \exp\left[-\frac{1}{2}(P-P_{\text{MAP}})^{\mathsf{T}}\Sigma^{-1}(P-P_{\text{MAP}})\right] \tag{3.10}$$

with mean $P_{\text{MAP}} = \tau\Sigma K^{\mathsf{T}} S$ and covariance matrix $\Sigma = (\tau K^{\mathsf{T}} K + \delta L^{\mathsf{T}} L)^{-1}$. Drawing a random sample of P from this multivariate normal distribution under the non-negativity constraint $P \geq 0$ can be done by solving the following randomized minimization problem ([76], see Appendix C):

$$\min_{P \geq 0} \left[P^{\mathsf{T}} \Sigma^{-1} P + P^{\mathsf{T}} (\tau K^{\mathsf{T}} S + (C_{\mathsf{L}}^{-1})^{\mathsf{T}} v) \right] \tag{3.11}$$

where v is a random vector sampled from the normal distribution $\text{Normal}(0, I)$ with an identity covariance matrix, and C_{L} is the lower triangular Cholesky factor of Σ such that $C_{\mathsf{L}} C_{\mathsf{L}}^{\mathsf{T}} = \Sigma$ (Appendix C). To solve Eq. (3.11), we use the Fast Non-Negative Least Squares (FNNLS) algorithm [77] (see Section 3.9).

To write down $\pi(\delta | P, S, \tau)$, we again take Eq. (3.7) and omit terms independent of δ. This results in a gamma distribution with parameters $\tilde{a} = a + \frac{n_r}{2}$ and $\tilde{b} = b + \frac{1}{2} \|LP\|^2$:

$$\pi(\delta | P, S, \tau) \propto \text{gamma}(\delta; \tilde{a}, \tilde{b}). \tag{3.12}$$

One common approach with gamma hyper-priors is to use small values for a and b that satisfy $a \ll n_r/2$ and $b \ll \|LP\|^2 /2$. With these values, a and b have little influence on $\tilde{a}$ and $\tilde{b}$ and on the posterior of δ in Eq. (3.12), therefore introducing minimal bias [76]. Initially, to achieve such an an uninformative flat distribution, we chose b values between 10^{-6} and 10^{-3}, which render the decay rate of the exponential very slow [76], as in Fig. 3.1. We found that b had to be 10^{-5} or smaller to avoid influencing the results. This choice resulted in estimates for α that were unreasonably large, so we instead use the GCV estimate of α, explained in Ch. 2, to select b. The AIC estimate would also be suitable. Note that the original publication of this work [14] predated the regularization parameter selection study [13] that formed Ch. 2. In the original paper, we utilized the leave-one-out Cross Validation (abbreviated as CV in Ch. 2) estimate of α. The figures in this chapter feature analysis using the CV estimate, rather than the GCV estimate. Recalling the relationship between δ and α from Eq. (3.9), we set the mean of Eq. (3.12) $(\tilde{a}/\tilde{b})$ equal to $\alpha_{\text{GCV}}^2 \tau$ and solve for b to obtain:

$$b = \frac{\tilde{a}}{\alpha_{\text{GCV}}^2 \tau} - \frac{1}{2} \|LP_{\text{init}}\|^2 \tag{3.13}$$

where P_{init} is calculated via Eq. (3.11) using $\delta = \alpha_{\text{GCV}}^2 \tau$. To draw random samples from the gamma distribution, we use the algorithm of Marsaglia and Tsang [78].

The Gibbs sampler explores the full (P, δ) space by taking random steps (Monte Carlo) along each parameter dimension, using the last value sampled for the other parameter (Markov-chain) [53, 79]:
The symbol $\sim$ denotes drawing a random sample from the distribution to its right, and subscripted t refers to the sample index. Note that δ_t, the sampled value of δ, is used immediately in the generation of a sample of P, P_t.

The Gibbs sampler produces a chain of P and δ values that is guaranteed to converge to the true distribution [53, 54, 60]. Upon convergence, the algorithm yields a large set of

Set $\delta_0 = \tau\alpha_{\mathrm{GCV}}^2$
Set $P_0 = P_{\alpha_{\mathrm{GCV}}}$
Set $t = 1$
while not converged **do**
  $\delta_t \sim \pi(\delta|P_{t-1}, S, \tau)$
  $P_t \sim \pi(P|\delta_t, S, \tau)$
  $t = t + 1$
end while

$(P_t,\ \delta_t)$ samples, distributed over parameter space such that their density is proportional to $\pi(P, \delta|S, \tau)$. The initial part of the chain (called the burn-in) is usually discarded, since initially the distribution is not yet stationary and depends on the particular choice of the starting value.

Assessment of convergence needs to be done carefully. We consider the sampling converged when the 2nd, 25th, 50th, 75th, and 98th percentiles for each point in P change by less than $10^{-3}\,\mathrm{nm}^{-1}$. For more easily computed values such as α and the mean of P, we use a multi-chain convergence test called the Gelman-Rubin statistic [55, 76]. This method requires that multiple chains are run from different starting points. At each convergence check, the intra- and inter-chain variances of some test statistic are compared. When the inter- and intra-chain variances coincide, the sampler is considered converged. The details are discussed in Appendix C.

3.5 Analyzing the Gibbs samples

Next, we calculate the median and spread estimates for the set of P vectors and δ values from the Gibbs sampler to obtain measures of uncertainty for P and δ. For each r_j, the values of $P_t(r_j)$ are sorted into quantiles, using method 8 from [80]. Specifically, we calculate the 2nd, 25th, 50th, 75th, and 98th percentiles. We generate distance distribution plots, where the median P (50th percentile) is shown as a line, and the other quantiles are used to draw error bands. The 2nd and 98th percentiles encapsulate the 96% credible interval. This interval indicates that the parameter in question has a 96% probability of falling within the interval [54]. Likewise, the 25th and 75th percentiles bound the 50% credible interval.

We analyze the distribution of δ, and consequently α, in a similar fashion. From the set of δ samples, we obtain a mean and the root-mean-square deviation. The method therefore automatically determines α, starting from the GCV estimate, without the need of user choices and the danger of user bias and inconsistency. Any systematic error in the GCV estimate will affect the final results. We next show that the Bayesian method described above captures well the effect of noise on P.

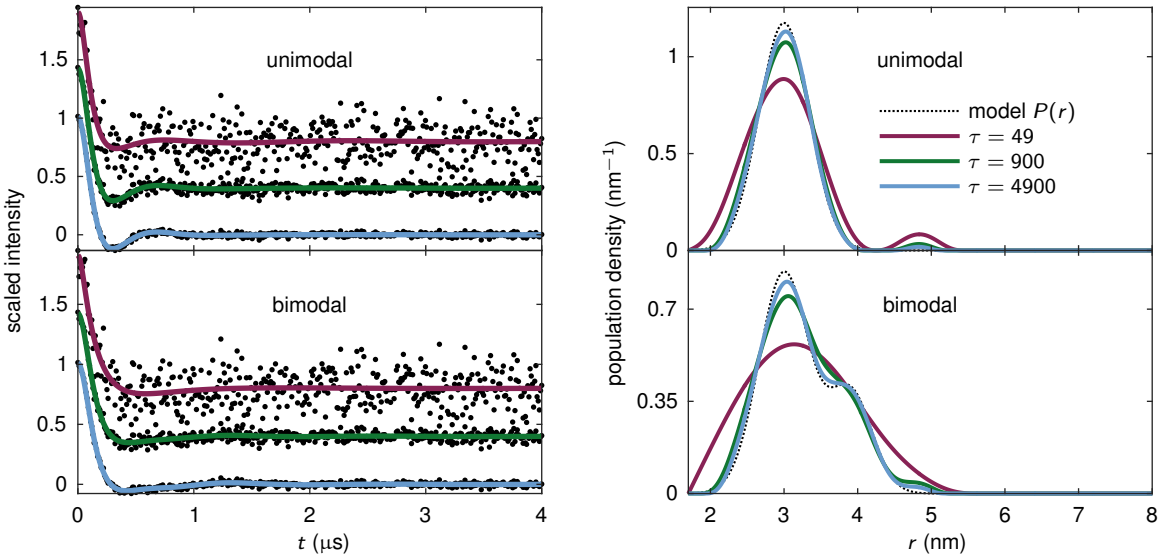

Figure 3.2: Simulated model data. Left panel: DEER time traces with varied SNR generated from unimodal and bimodal distributions are shown as black dots in the top and bottom plots, respectively. The unimodal distribution is a normalized Gaussian centered at 3 nm with a FWHM (full-width at half maximum) of 0.8 nm. The bimodal distribution is a normalized sum of two Gaussians with centers of 3 and 3.9 nm, FWHM of 0.8 and 0.7 nm, and weights of 1 and 0.4, respectively. For all S and P, $n = 400$. Right panel: the corresponding unimodal and bimodal $P(r)$ vectors obtained from Tikhonov regularization are displayed in the top and bottom plots, respectively. The model $P(r)$ is shown as a dotted line and the low-, medium-, and high-SNR solutions are shown as magenta, green, and blue curves, respectively.

3.6 Results

3.6.1 Unimodal and bimodal distributions

Credible intervals Fig. 3.2 shows DEER time-domain traces generated from two model distance distributions (one unimodal and one bimodal) and three different levels of noise. For these data, six independent chains per data set converged within 11,500-20,500 iterations. The sets of P and δ from the resulting chains represent the posterior distribution and are illustrated in Fig. 3.3. The results are plotted in terms of α, rather than δ. This scatter plot of the value of P at $r = 3$ nm vs. α for each iteration of two chains of the Gibbs sampler reveals that most of the samples fall near to the model value for P and the α that best recovered the model through Tikhonov regularization. As the distance from the maximum of the posterior increases, the den-

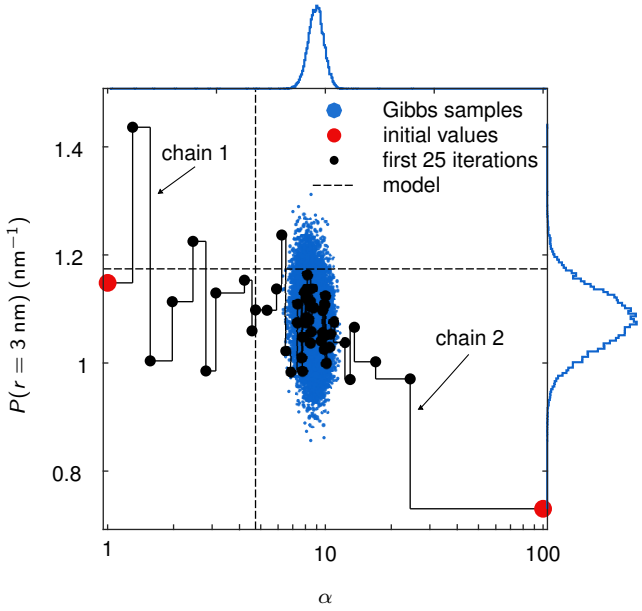

Figure 3.3: Illustrative plot of the result of a Gibbs sampler, showing $P_t(r_j)$ vs. α_t from the unimodal $\tau = 900$ (medium noise) case (Fig. 3.2), including marginal histograms for both variables. The scatter plot shows two chains of Gibbs sampled values for $P(r_j = 3\mathrm{nm})$ vs the sampled δ values (converted to α through Eq. (3.9)). This distance corresponds to the peak maximum of the model P. Chain 1 begins at $\alpha = 1$ and chain 2 begins at $\alpha = 100$. The dashed line in the α dimension corresponds to the α that minimizes the error between the Tikhonov solution and the model P. After a brief burn-in for both chains, the algorithm finds and samples the posterior distribution, visiting regions of high probability more often than those of low probability. The offset from the model values is due to noise and the systematic bias introduced by the CV estimate of α.

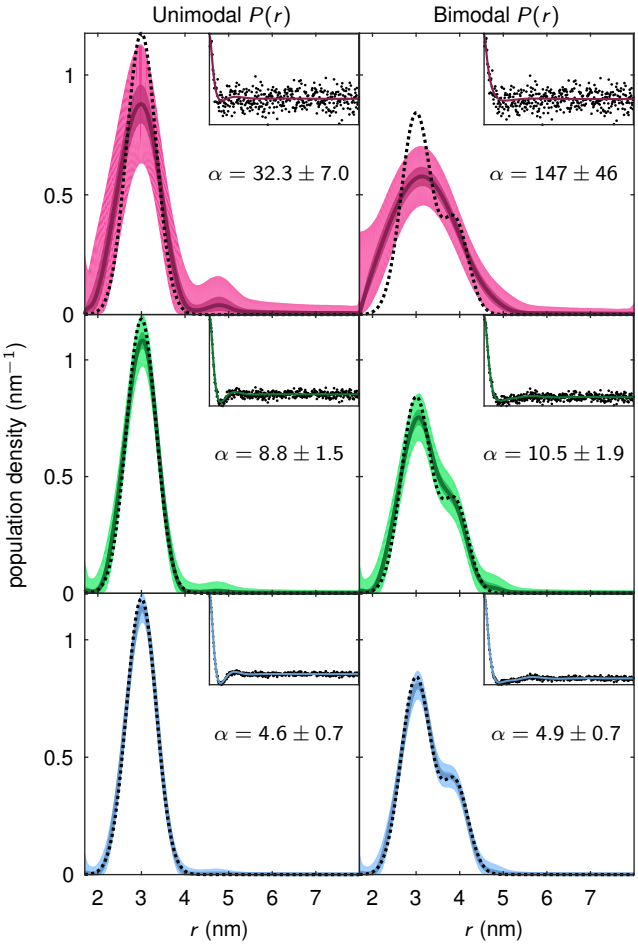

Figure 3.4: Bayesian credible intervals on $P(r)$ from model data (Fig. 3.2). Median P solutions and Bayesian credible intervals for the unimodal and bimodal models are displayed in the left and right columns, respectively. Low-, medium-, and high-SNR cases are in the top, middle, and bottom rows, respectively. The 96% and 50% credible intervals are light and dark bands, respectively. Model S traces with variable SNR (black dots) and time-domain fits from the median Ps are shown as insets. α is reported as the mean plus or minus 2σ.

sity of points decreases. To illustrate the Markov-chains that built this particular posterior, the initial values and the first 25 iterations for two chains are highlighted. The sampler alternates between steps in the δ and P dimensions. The initial value starts the Markov-chain in a region of low probability, and then quickly finds the region of high probability, sampling around the posterior. Though highly informative, this view of the posterior would be difficult to visualize for all $n = 400$ points in each P considered, so we examine the results in terms of quantiles.

The resulting median and the 50% and 96% credible intervals obtained from the Gibbs sampler for simulated data are shown in Fig. 3.4. The width of the credible intervals increases with the noise level and decreases with increasing δ: variance is reduced at the cost of increased bias (Appendix C), but τ dominates. High noise levels may lead to much more drastic deviations in the distance distribution. Bimodal distributions display wider credible intervals than unimodal distributions for a given noise level, again matching our expectations. The average and standard deviations of α, obtained from δ via Eq. (3.9), are given in Fig. 3.4.

Small features adjacent to the primary peaks are likely to be artifactual. For example, the small feature at 4.8 nm in the unimodal $P(r)$s in Fig. 3.2 are shown in Fig. 3.4 to be indistinguishable from zero and therefore artifactual at all noise levels considered.

The credible intervals in Fig. 3.4 show the persistence of a second peak at 3.9 nm for bimodal $P(r)$s in the mid- and high-SNR cases. In the low-SNR case, where Tikhonov regularization is unable to recover a bimodal distribution due to poor data quality, the credible intervals reflect that the second peak cannot be resolved.

Histograms By generating histograms of single-point estimators such as the mean, the mode, or the variance for each P_t, we can gain additional insight. Fig. 3.5 shows an example for the model data from Fig. 3.2. The histograms for the high-noise cases are larger than those for the low and medium noise cases because the Gibbs sampler took many more iterations to converge. In both the unimodal and bimodal cases, the location of the most populated distance does not significantly change with increasing SNR, though it does become more precise. This stability of the mode with respect to noise has been previously observed [15]. In contrast, the peak width is statistically significantly different at different noise levels, and unrelated to the precision of the mode. The mean likewise shifts with increasing noise.

Regularization parameter In Tikhonov regularization, error assessment for α is not performed. One distinct advantage of the method described here is that it generates a distribution of δ and thereby α, allowing statistical analysis of this parameter. Fig. 3.6 shows an example. For the high and medium SNR cases from Fig. 3.2, the Bayesian MCMC approach generates a narrowly distributed set of αs close to the value that minimizes error between P_α and the model P. The discrepancy is greater in the low-SNR case. CV outperforms the L-curve, but both heuristic methods overestimate α. Again, note that this 2016 study that utilized CV predated our 2018 investigation into the regularization parameter selection problem. As discussed

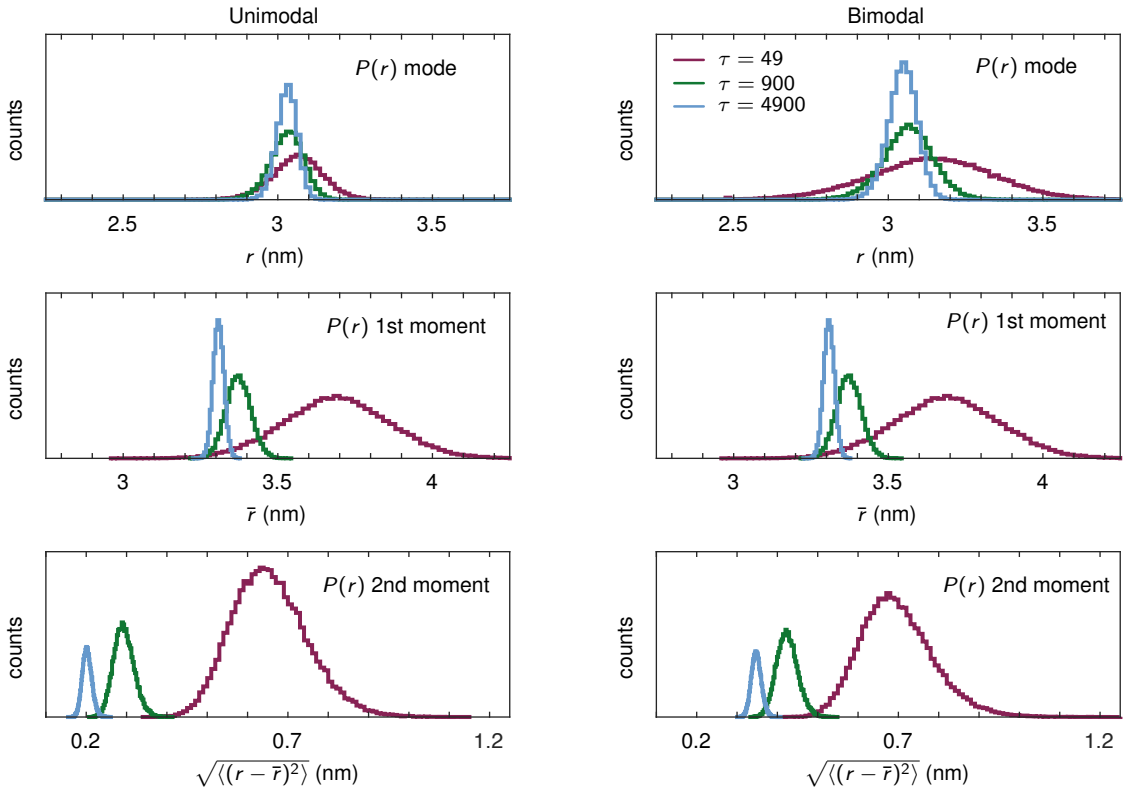

Figure 3.5: Analysis of the Gibbs sampler output. Left column: unimodal case. Right column: bimodal case. Top row: $P(r)$ mode. Middle row: $P(r)$ 1st moment. Bottom row: $P(r)$ 2nd moment. The mode of the $P(r)$ is largely invariant to noise, but its precision does correlate with SNR. Added noise does cause statistically significant shifts in the 1st moment and especially in the 2nd moment. Such analysis can be used to demonstrate the statistical difference or similarity of two separate $P(r)$s.

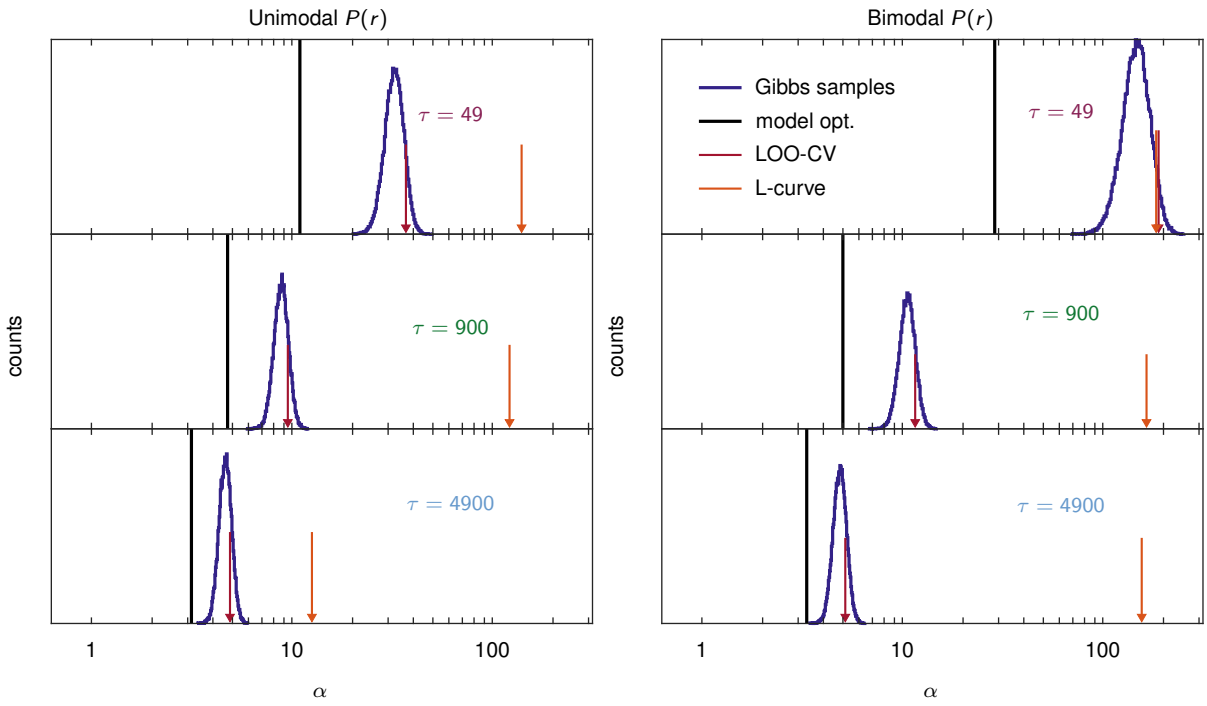

Figure 3.6: Comparison of α values for the model data in Fig. 3.2 determined from the maximum curvature of the L-Curve (orange), CV (dark red), and Gibbs sampling (dark blue), with the value that best recovers the model using Tikhonov regularization (black). Left column: unimodal case. Right column: bimodal case. Top row: low SNR. Middle row: medium SNR. Bottom row: high SNR. The Gibbs sampler out-performs point estimates for α in all cases.

in Ch. 2, GCV and AIC perform better in the setting of DEER data analysis.

3.6.2 Mixed-width distribution

One example that is frequently brought up to describe the limitations of Tikhonov regularization is a bimodal $P(r)$ consisting of a wide and a narrow Gaussian peak of roughly equal area (Fig. 3.7) [15, 21]. From the noisy time-domain signal corresponding to this $P(r)$, Tikhonov regularization predicts the narrow peak well, but represents the wide peak as a combination of features that have widths similar to that of the narrow peak. The smoothness bias in Tikhonov regularization applies approximately equally over the entire range of r and, in this case, this bias prevents faithful recovery of the original mixed-width bimodal $P(r)$. The Bayesian method outlined here provides error bands that visually indicate that the shape of $P(r)$ is quite uncertain

in the area of the wide feature. This shows that the wide peak cannot be recovered as faithfully as the narrow feature.

3.7 Discussion

Although not the only source of uncertainty in DEER distance distributions, noise in the signal contributes spurious features and partially determines the degree of regularization required. Researchers often go to great lengths to minimize the presence of noise. As a first step towards full uncertainty quantification in DEER analysis, the present work successfully quantifies the uncertainty in $P(r)$ stemming from experimental noise. This will make the interpretation of DEER data easier and more robust, as more quantitative arguments about the significance or insignificance of a given feature can be made.

3.7.1 Interpretation guidelines

Bayesian credible intervals are a natural way to visualize the uncertainty in $P(r)$ given a certain noise level in the data. Features lying within a given credible interval are statistically indistinguishable at that level of credibility. The credible intervals can therefore guide interpretation of DEER distance distributions and help guard against overinterpretation. Next, we consider the following typical interpretative questions: (a) Significance of small peaks, (b) presence of shoulders, and (c) relevance of the difference between two distributions.

(a) Small side peaks are not uncommon in experimental DEER distance distributions. With the credible intervals, it is possible to assess whether they are significant or not. If the credible interval of a small peak includes the baseline, then the small peak is statistically indistinguishable from zero at that level of credibility and must be considered insignificant given the data.

(b) Shoulders in distributions can indicate conformational subpopulations. If a straight reference line (or a spline) connecting two points on the median distribution to the left and to the right of the shoulder fully falls within the credible intervals, then the shoulder is not significant. In such a case, a P with a shoulder is statistically indistinguishable from a P without a shoulder.

(c) When comparing two distributions, the credible intervals allows the practitioner to quickly identify whether the distributions are significantly different. If the credible intervals of two P do not overlap, then they are different. If there is overlap in the credible intervals, then significance testing can be employed to answer basic questions such as: are the mean spin-spin distances significantly different between these two samples? For example, one can construct histograms as in Fig. 3.5 and perform a Student's t-test. Fig. 3.5 demonstrates that $P(r)$ peak width can be significantly affected by experimental noise. This implies that quantifying uncertainty in the width of $P(r)$ is crucial in studies that compare the relative widths of distributions.

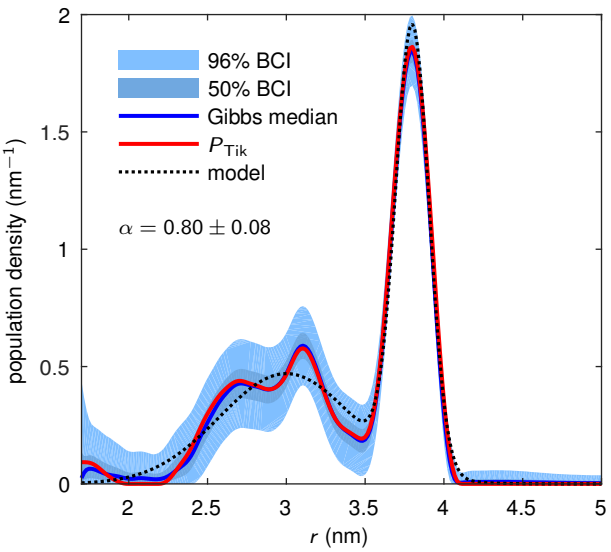

Figure 3.7: Bayesian analysis applied to a bimodal mixed-width distribution. The model was generated with two equally weighted Gaussian peaks centered at 3 and 3.8 nm with FWHM of 1 and 0.25 nm, respectively. Normally distributed noise with $\tau = 2500$ (SNR = 50) was added to the corresponding S. Both the Tikhonov solution and the Gibbs median P fail to recapitulate the model distribution. The Bayesian credible intervals indicate that there is increased uncertainty in the regions where the Tikhonov solution is the most incorrect.

3.7.2 Comparison with other methods

In the literature, the quality of a given $P(r)$ is most often assessed by examining the time-domain data, considering the magnitude of the regularization parameter, and then making a qualitative judgment call (Appendix C). Several attempts have been made to quantify uncertainty in DEER. Those quantitative estimates of error from signal noise that the authors have encountered come from one of three methods: the noise feature of the Validation Tool in the software `DeerAnalysis` [18]; error propagation [18, 52]; and various model-based approaches [21, 33–40].

`DeerAnalysis` `DeerAnalysis` [18] is the most commonly used software for DEER data processing, and it uses Tikhonov regularization to calculate $P(r)$. The method proposed here quantifies uncertainty in the Tikhonov solution for $P(r)$, so `DeerAnalysis` is the most comparable extant tool. `DeerAnalysis` provides the capability to assess error from time-domain noise by repeatedly adding additional normally distributed synthetic noise to the data V/V_0 and then examining the resulting changes in the P vectors. This approach falls under the category of bootstrapping methods. The user sets several parameters, including the magnitude of the added synthetic noise and the number of iterations. The noise is therefore not guaranteed to be equal in magnitude to the experimental noise level, compromising accuracy. Convergence is likewise not guaranteed. The error bars displayed are the $\pm 2\sigma$ bands of the set of calculated P vectors. This visualization is useful if the population is normally, or close to normally, distributed. This is not guaranteed to be the case, especially where P is zero or close to zero. The confidence intervals often include regions of negative probability, which is not physically valid, so negative valued indices of the $\pm 2\sigma$ bands are set to 0. No error analysis regarding the uncertainty in the choice of α can be performed.

Compared to the validation tool in `DeerAnalysis`, the present method is more time consuming because of the large number of required iterations. `DeerAnalysis` also considers additional sources of uncertainty, such as background correction and pulse bandwidth. However, the Bayesian method features convergence criteria, utilizes real noise information, reduces user bias, and provides uncertainty estimates for α.

Propagation The next method is more limited in scope than either `DeerAnalysis` or the proposed Bayesian approach. The variance of the noise in $S(t)$, σ_S^2, can be propagated from the time domain to the distance domain using $\sigma_P^2 = \sigma_S^2 \mathrm{diag}\left(Q'_\alpha Q'^{\mathrm{T}}_\alpha\right)$ where the primes indicate that only rows and columns in Q_α (as defined in Eq. (1.8)) are retained that correspond to the indices where the unconstrained Tikhonov solution is non-negative. Commercial EPR spectrometers store acquired data only after signal averaging, so noise statistics are usually unavailable. In this case, the fit RMSD (root mean square difference) between the Tikhonov fit KP and the data S is often used as a stand-in for σ_S. As a result, any error in the fit will corrupt the calculation. This approach gives no error estimate for regions where the Tikhonov solution is

zero. This method is used in the program `FTIKREG` [52], which was used by `DeerAnalysis` until 2016. The Stoll lab has used this method in the past [81, 82].

Unlike the propagation method, for which true noise statistics are usually unavailable, the proposed method uses observed noise information. The present method also provides estimates for the full r domain, including regions where $P(r)$ has been constrained to non-negative values. Finally, the propagation method considers α to be given and fixed, and so provides no uncertainty estimates for the level of regularization applied.

Model Fitting In addition to the Bayesian approach, it is possible to use classical frequentist methods. One such approach to quantify uncertainty is to assume a particular model for $P(r)$, fit the model to the data, and then statistically evaluate the error in the various fit parameters, usually with some variant of the χ^2 cost function. This method has been employed in `LongDistances` [33], `DeerSim` [34–37], and `DEERconstruct` [38]. Alternatively, `DEFit` fits Gaussian distributions directly to the time-domain data [39]. Hustedt and coworkers have implemented a model-based approach that simultaneously fits the background function and $P(r)$ to time-domain data, thereby uncovering potential correlations between background parameters and features in $P(r)$ [21, 40].

The Bayesian method proposed here quantifies uncertainty arising from time-domain noise and regularization, ameliorating one of the disadvantages associated with Tikhonov regularization compared to model-based approaches. The Bayesian approach so far does not take into account background correction. As with Tikhonov regularization, the proposed method is model-free with respect to $P(r)$. The credible intervals produced by the present method do not include any negative values.

3.8 Conclusions

We introduce a Bayesian statistical method for the analysis of DEER spectroscopic data. The method extends the existing method of Tikhonov regularization: First, it not only determines $P(r)$, but it also fully quantifies the associated uncertainty (error) due to noise in the data. Second, it automatically determines the regularization parameter α and provides an associated error estimate. The uncertainty regarding α is included in the overall error estimates for P. The statistically determined uncertainty estimates are visualized using quantiles. This is an important improvement to the application of DEER spectroscopy as it guides interpretation, reduces practitioner bias, increases reliability, and helps reproducibility.

This treatment is far from complete. There are still many other sources of uncertainty affecting the DEER distance distribution that need to be addressed. The error estimates determined here do not capture uncertainty arising from background correction, time-domain truncation, and discretization. Uncertainty in the background correction, in particular, can give rise to very significant uncertainty in the distance distribution [15, 18]. The proposed method may be com-

bined sequentially with the validation tool of `DeerAnalysis` to assure that a good choice for the background function has been used. The Bayesian statistical approach outlined here is flexible, general and robust enough to be extended to capture the influences of, and correlations between, these other sources of uncertainty.

3.9 Methods

All computations were carried out using MATLAB R2016a [83]. The dipolar kernel matrix K was calculated using Fresnel integrals [84]. Non-negativity constrained minimization for Tikhonov regularization and for Gibbs sampling was done using the Fast Non-Negative Least Squares (FNNLS) algorithm [77]. Each minimization took between 0.01 and 0.1 s on a standard laptop computer. Uniform and normally distributed pseudo-random numbers were generated with the Mersenne Twister pseudo-random number generator (PRNG) as implemented in MATLAB R2016a using a seed value of 19560 to generate model data and 12345 for Gibbs sampler runs. The PRNG was reset before each MCMC run, but not during a run, i.e. not between chains. The gamma-distributed pseudo-random numbers were generated using the algorithm of Marsaglia and Tsang [78, 85].

3.10 Acknowledgements

We thank Jonathan M. Bardsley (University of Montana) [76] for a very helpful discussion over breakfast, François Monard (formerly University of Washington, currently University of California, Santa Cruz) for useful literature recommendations, and Niket Thakkar (University of Washington) for valuable advice. This work was supported by the National Science Foundation (NSF) grant CHE-1452967, the National Institutes of Health (NIH) grant R01 EY010329 (S.S.), and the National Institute of General Medical Sciences (NIGMS) of the NIH under Award Number T32GM008268 (T.H.E.).

Chapter 4

EXTENDED BAYESIAN MODEL INCLUDING BACKGROUND

This chapter summarizes unpublished work that follows up and extends the approach presented in Ch. 3.

Double electron-electron resonance (DEER) signals comprise contributions from both intra- and inter-complex dipolar interactions. The latter is typically fit and divided out before inferring a distance distribution, $P(r)$, from the former. This pre-processing step can introduce significant error into the resulting $P(r)$. Here, we introduce a statistical method, based upon Bayesian inference, to simultaneously fit the intra- and inter-complex components of the DEER signal. The method quantifies uncertainty in the final result, facilitating reasonable interpretation.

4.1 Introduction

In the context of DEER data analysis, approaches that simultaneously infer the intra- and intermolecular signal components are referred to as global. There is not yet a published global analysis approach that utilizes a non-parametric model, such as Tikhonov regularization. Existing implementations use parametric modelling approaches, such as Gaussian mixture models [20, 21]. As the background correction step in a sequential analysis approach can introduce significant error in $P(r)$, finding the best possible estimate of $B(t)$ and quantifying the uncertainty introduced into $P(r)$ is crucial.

The software package, `GLADD`, introduced by Hustedt and coworkers, simultaneously fits $P(r)$ and $B(t)$ to the data, $V(t)$, using Gaussian mixture models [20, 21]. The downside to this approach is that it assumes a particular form for $P(r)$ that may not be physically realistic. A global analysis procedure using a non-parametric model for $P(r)$ would allow a direct comparison between these two approaches.

Here we present a global fully probabilistic model that is able to simultaneously determine a parameter-free distance distribution and a background function. The model is based upon the Bayesian approach introduced in Ch. 3 and provides complete quantification of uncertainty.

4.2 Theory

The DEER signal considered here is identical to that in the preceding chapters and Appendix A. All symbols retain their meanings.

4.2.1 Data Pre-processing

Prior to analysis, DEER data is subjected to a few pre-processing steps. First, quadrature-detected integrated spin-echoes are signal averaged and phase-corrected. Then, estimates for the true t_0 and the non-modulated signal at $V(t_0)$ are obtained by fitting a simple curved function to the early time-points. With the exception of normalization by $V_0 \equiv V(t = 0)$, these pre-processing steps and their effects on the outcome are not treated in the statistical model presented below.

4.2.2 Data model

The general model for the pre-processed DEER time-domain trace is

$$V(t) = [1 - \lambda + \lambda S(t)] \cdot B(t) + N(0, \sigma) \tag{4.1}$$

with the modulation depth parameter λ, the dipolar modulation function $S(t)$, and the background function $B(t)$. N represents stochastic noise and is a random vector drawn from a normal distribution with center 0 and standard deviation σ.

For the components $S(t)$ and $B(t)$, different models can be chosen. We use

$$S(t) = K(t, r)P(r) \tag{4.2}$$

where S, K, and P are the same as in Chs. 1-3. To take the modulation due to interactions between spins on different complexes into account, we employ the widely used exponential-decay model [1, 2, 86, 87] extended with a phenomenological stretching exponent [88]

$$B(t) = \exp\left(-|k\rho_{\text{pump}}ct|^{D/3}\right) \tag{4.3}$$

where ρ_{pump} is the probability of exciting B spins with $k = \frac{8\pi^2}{9\sqrt{3}}\frac{\mu_B^2 g_A g_B}{\hbar}$. It depends on three parameters: the pump-pulse efficiency ρ_{pump} (assumed given), the average spin number (not molar) concentration c, and the dimensionality D (which functions as a stretching exponent). $D = 3$ corresponds to an exponential decay while $D = 6$ results in a Gaussian decay. Note that the $|...|$ in Eq. (4.3) allows it to be applied also for $t < 0$ in the 4-pulse DEER sequence. Alternative background models, such as the excluded-volume model [20, 89] calculated via shell factorization [31] can easily be accommodated.

4.3 Probabilistic model

The probabilistic model used here is an extension of that from Ch. 3 (Eq. 3.7). In Ch. 3, the background-corrected and λ-scaled time-domain vector $S(t)$ was considered the raw data. Perfect background-correction was assumed, or, more accurately, background-correction errors

were neglected. Here, the averaged integrated echo intensity trace $V(t)$ is considered the raw data. The extended Bayesian hierarchical model can be written down in several ways. Here, we will examine the most physically complete model given the same simplifying assumptions about experimental realization made throughout this thesis. The model will have the familiar form:

$$\text{posterior} = \text{likelihood} \cdot \text{priors} \cdot \text{hyperpriors}. \tag{4.4}$$

The likelihood function is a multivariate normal distribution:

$$\pi(V|V_0, \sigma, \lambda, P, c, D) = \text{MvNormal}(V_0[1 - \lambda + \lambda KP] \exp\left(-|\rho_{\text{pump}} ckt|^{D/3}\right), \Sigma_V) \tag{4.5}$$

where the covariance matrix can be written as simply $\Sigma_V = \sigma^2 I$ because the noise in a DEER experiment is independent and normally distributed [14]. Here, as in Ch. 3, σ is assumed given since it can be obtained experimentally. The prior for P is identical to that in Ch. 3:

$$P = \text{constrMvNormal}\left(0, (L^{\mathsf{T}}(\delta I)L)^{-1}\right) \tag{4.6}$$

where δ is the distance-domain precision, defined in Ch. 3, and we explicitly write out the identity matrix. The prior PDF for λ reflects its possible range of 0 to 1 and is:

$$\pi(\lambda) = U(0, 1). \tag{4.7}$$

More informative choices for $\pi(\lambda)$ are possible and could be centered around a value estimated from the pulse excitation bandwidths. The spin-concentration c can be estimated from the sample preparation procedure or, more accurately, from a spin-quantitation CW EPR experiment. Its prior PDF is written as a normal distribution:

$$\pi(c) = N(\bar{c}, \sigma_c) \tag{4.8}$$

where the estimated mean concentration, $\bar{c}$, and associated standard deviation, σ_c, are assumed given. If σ_c is not small compared to $\bar{c}$, then some other distribution with only positive support should be used. The phenomenological stretch constant, D, is estimated from the truncated Normal distribution:

$$\pi(D) = \text{trunc}N(3, 2); 0 < D < \infty \tag{4.9}$$

where the mean has been set to the expected dimensionality of 3 and the standard deviation 2 confines nearly all of the PDF mass to between 0 and 6. The support is limited to only positive values. These values can be adapted based on the nature of the sample: It is expected that $D \approx 3$ for dilute frozen solutions of soluble proteins, but smaller for membrane proteins. The value of the non-modulated spin-echo, V_0, must be estimated from the noisy data V:

$$\pi(V_0|\sigma) = N(\bar{V_0}, \sigma) \tag{4.10}$$

where σ is the time-domain noise standard deviation and $\bar{V}_0$ is estimated by least-squares fitting a simple curved function to the early part of V. Finally, the hyperprior for δ is the same as that in Ch. 3:

$$\pi(\delta|a,b) = \text{Gamma}(a,b) \tag{4.11}$$

with the hyperparameters a and b determined as before. The posterior PDF for P is nearly identical to that in Ch. 3:

$$\pi(P|S,\tau,\delta,\lambda,B(\rho_{\text{pump}},c,D)) = \text{constrMvNormal}(\Sigma_P^{-1}K^{\mathsf{T}}\Sigma_S S_{\text{data}},\Sigma_P) \tag{4.12}$$

where S_{data} is calculated from the raw data V:

$$S_{\text{data},i} = \frac{1}{V_0\lambda}\text{diag}(B)_{ii}\backslash\text{diag}(V)_{ii} - \frac{1}{\lambda} + 1. \tag{4.13}$$

The covariance matrix

$$\Sigma_{S,ii} = \text{diag}(B)_{ii}\backslash\left(\frac{\sigma^2}{V_0}I\right)_{ii} - \frac{1}{\lambda} + 1 \tag{4.14}$$

scales the time-domain variance by B and λ, and

$$\Sigma_P = (K^{\mathsf{T}}\Sigma_S K + \delta L^{\mathsf{T}}L)^{-1}. \tag{4.15}$$

To summarize the full model in terms of numerical sampling:

$$\begin{aligned}
V &\sim \text{MvNormal}(V_0[1-\lambda+\lambda KP]B,\Sigma_V)\\
B &= \exp\left(-|\rho_{\text{pump}}ckt|^{D/3}\right)\\
S_{\text{data},i} &= \frac{1}{V_0\lambda}\text{diag}(B)_{ii}\backslash\text{diag}(V)_{ii} - \frac{1}{\lambda} + 1\\
P &\sim \text{constrMvNormal}(\Sigma_P^{-1}K^{\mathsf{T}}\Sigma_S S_{\text{data}},\Sigma_P)\\
V_0 &\sim \text{N}(\bar{V}_0,\sigma)\\
\lambda &\sim \text{U}(0,1)\\
D &\sim \text{N}(3,2)\\
c &\sim \text{N}(\bar{c},\sigma_c)\\
\delta &\sim \text{gamma}(a,b)
\end{aligned} \tag{4.16}$$

where $\sim$ indicates stochastic sampling and with the parameters τ, ρ_{pump}, K, and L considered given.

4.4 Methods - numerical sampling

Due to the complexity (lack of a standard-form distribution for the posteriors) of the joint expression of the hierarchical model:

$$\begin{aligned}\pi(V, V_0, \lambda, P, \delta, c, D) = \frac{(2\pi^{-n_t/2})}{|\Sigma_V|} \exp\left(-\left(V - V_0[(1-\lambda)+\lambda KP]\exp\left\{-|\rho_B ckt|^{D/3}\right\}\right)^{\mathsf{T}}\right. \\ \left. \ldots\ \Sigma_V^{-1}\left(V - V_0[(1-\lambda)+\lambda KP]\exp\left\{-|\rho_B ckt|^{D/3}\right\}\right)\right) \\ \cdot (2\pi\sigma^2)^{-1/2}\exp\left(-\frac{(V_0-\bar{V}_0)^2}{2\sigma^2}\right) \\ \cdot \frac{(2\pi^{-n_r/2})}{|(L^{\mathsf{T}}(\delta I)L)^{-1}|}\exp\left(-(P-0)^{\mathsf{T}}(L^{\mathsf{T}}(\delta I)L)(P-0)\right) \\ \cdot \frac{b^a}{\Gamma(a)}\delta^{(a-1)}\exp(-b\delta) \\ \cdot \begin{cases}\frac{1}{1-0}, & \text{if } 0 < \lambda \leq 1 \\ 0, & \text{otherwise}\end{cases} \\ \cdot (2\pi\sigma_c^2)^{-1/2}\exp\left(-\frac{(c-\bar{c})^2}{2\sigma_c^2}\right) \\ \cdot (2\pi 2^2)^{-1/2}\exp\left(-\frac{(D-3)^2}{2\cdot 2^2}\right)\end{aligned} \tag{4.17}$$

it is not possible to use direct Gibbs sampling as in Ch. 3, and more general methods have to be used. Hamiltonian Monte Carlo (HMC) is one particular method well-suited to such complex models.

4.4.1 Hamiltonian Monte Carlo

In order to generate a numerical estimate of Eq. 4.17, we use the pseudo-random sampling technique Hamiltonian Monte Carlo [90], which has a statistical accept/reject step (similar to the classic Metropolis algorithm [60]) and therefore does not need to generate perfect samples on every step. For HMC sampling, we use the MATLAB interface [91] of the open-source package Stan [92].

As an initial validation, we compared HMC and the Gibbs sampler on the probabilistic model from Ch. 3. The HMC and Gibbs MCMC results for the case of unconstrained P are identical (Fig. 4.1). Including the non-negativity constraint leads to differences that are due to differences in how the constraint is incorporated in the samplers (Fig. 4.2).

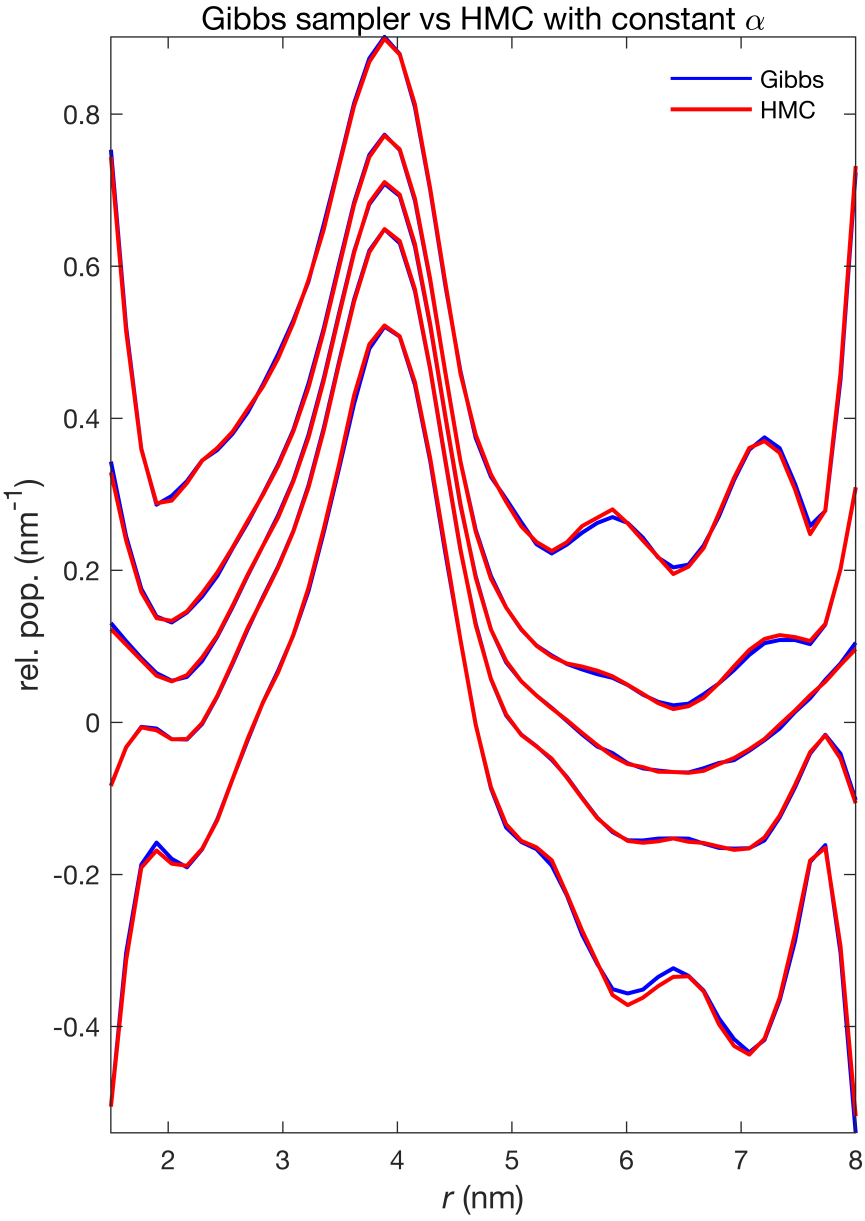

Figure 4.1: Comparison of the Gibbs sampler from 3 with the Stan HMC sampler for unconstrained P samples from a multivariate normal distribution with a constant α ($\alpha = \delta/\tau$) value. The sampled values are nearly identical. The five lines, stacked from bottom to top, are the 2nd, 25th, 50th, 75th, and 98th percentiles of the sampled P values, which correspond to the credible intervals shown in Ch. 3.

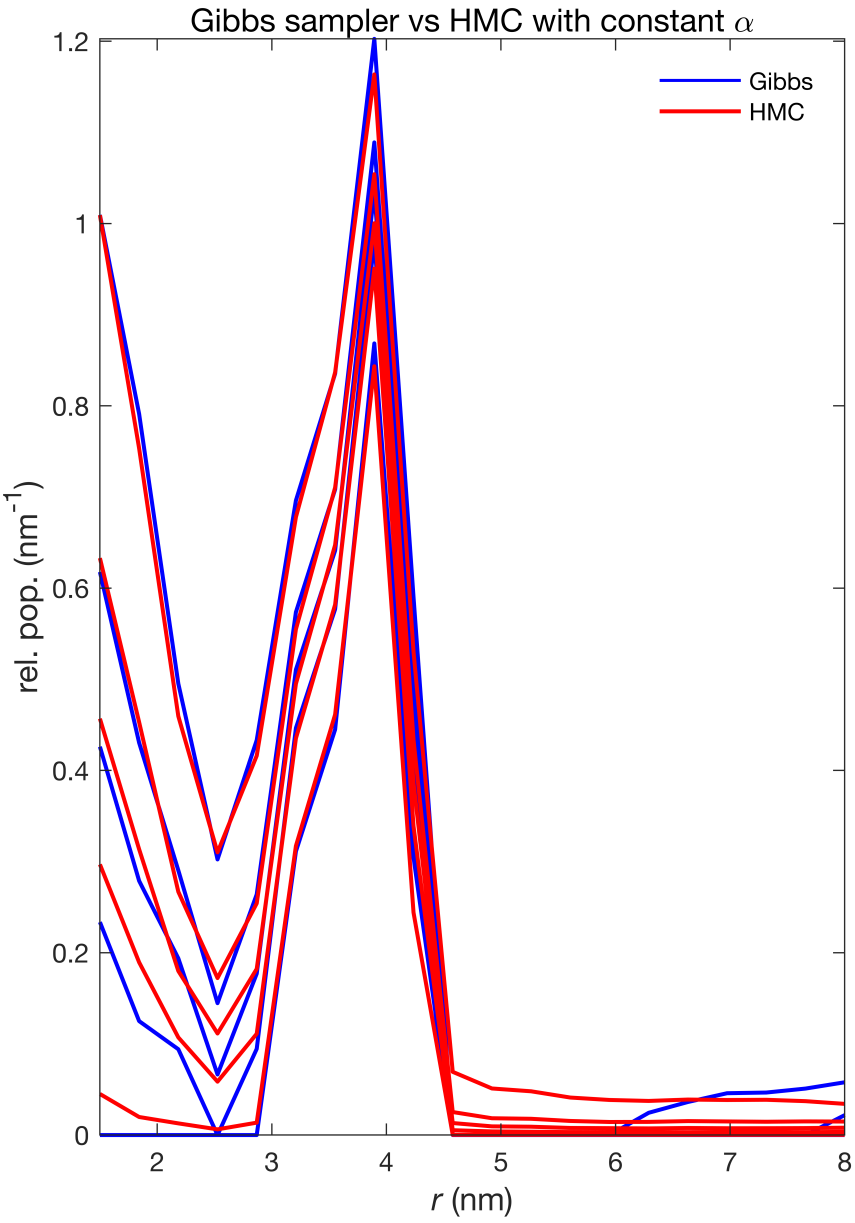

Figure 4.2: Comparison of the Gibbs sampler from 3 with the Stan HMC sampler for non-negativity constrained P samples from a multivariate normal distribution with a constant α ($\alpha = \delta/\tau$) value. The results differ, likely due to differences in how the constraint is implemented. The sampled values are nearly identical. The five lines, stacked from bottom to top, are the 2nd, 25th, 50th, 75th, and 98th percentiles of the sampled P values, which correspond to the credible intervals shown in Ch. 3.

4.5 Numerical Study Design

The numerical investigation will follow similar lines as that in Ch. 2. To build a large, synthetic test data set, the test set introduced in Ch. 2 must be augmented with randomized background functions, $B(t)$, and modulation depths, λ. The results will similarly be analyzed as before: primarily in terms of model parameter recovery.

Chapter 5

SUMMARY AND OUTLOOK

This body of work will enable and simplify statistically defensible interpretations of DEER EPR data. It is now within reach to compare directly parametric and non-parametric model analysis for DEER data, which will allow progression in the field. DEER distance distributions with full uncertainty quantification can be more rigorously incorporated into molecular modeling projects to provide physically-measured restraints.

Ch. 2 describes improvements to Tikhonov regularization for DEER spectroscopy that improve the accuracy and reliability of the method. The results of this work are currently being incorporated into workflows within the community. The design of the study is general and may be applied to other ill-conditioned, inverse problems. We have also made freely and publicly available an extensive test-set of physically- and experimentally-reasonable simulated DEER data that is likely to be useful to other practitioners in the field.

Ch. 3 describes a Bayesian statistical framework for inferring distance distributions with uncertainty quantification from DEER data. This is the contribution that is most important to the field in terms of novelty, as Bayesian statistics had not yet made its way into the field of EPR. It is also significant in that it is the first statistically rigorous method for quantifying uncertainty in a non-parametric DEER $P(r)$.

Ch. 4 outlines an extension to the approach from Ch. 3 that includes background correction. This enables global analysis for a non-parametric model approach for DEER data analysis. It also places non-parametric distance-restraint inference on a similar statistical playing field as parametric approaches, allowing direct comparison.

There is a great deal of work yet to do. The full model presented in Ch. 4 is not yet completely implemented and tweaks may be necessary to ensure computational feasibility. There are also several significant sources of uncertainty in the DEER experiment that could be incorporated into this framework, namely, orientation selection, multi-spin effects, and differential phase memory times. However, it should be possible to account for these eventualities using the Bayesian hierarchical framework employed here.

A still open problem is how to optimally use DEER distance distributions as restraints in molecular modeling studies. It is my hope that upcoming students will uncover a clear and rational way to solve this problem and that quantitative error estimates, such as those made available here, will prove useful to them to those ends. With reliable, quantitative estimates of parameters and their uncertainties, it should be possible to build useful models of reality from which we know exactly what we can and cannot claim. Ultimately, using the numerical results of DEER experiments to create informative structural models of proteins is the only reason to

go through the significant effort to perform such experiments. The work presented here is, in all honesty, simply fulfilling a prerequisite to that more important work.

Appendix A

THEORY OF THE DEER SIGNAL

This appendix provides a derivation of the expressions that describe the time-domain DEER signal for typical—well, idealized—experimental conditions. First, the magnetic dipolar interaction between two electrons located in a strong external magnetic field is explored. Then, the effect of this interaction on the DEER experimental observable—electron spin echoes—is laid out. The microscopic interaction is scaled up to the situation encountered in a macroscopically-sized sample solution through the use of the ṕowder average' integration. Finally, the signal expression is rewritten as a combination of Fresnel integrals, which provides an easy route to fast numerical evaluation.

A.1 The Magnetic Dipolar Interaction

We start by considering the vector potential at a distance r from a classical dipole in units of J/T, given by:

$$A(\vec{r},\vec{\mu}) = \frac{\mu_0}{4\pi}\frac{\vec{\mu}\times\vec{n}}{r^3} \tag{A.1}$$

where μ_0 is the magnetic constant (vacuum permeability), $\vec{\mu}$ is the magnetic dipole moment, and $\vec{n}$ is the spherical unit vector:

$$\vec{n} = \begin{pmatrix} \cos\phi\sin\theta \\ \sin\phi\sin\theta \\ \cos\theta \end{pmatrix}. \tag{A.2}$$

The magnetic field (flux density) from the dipole at some distance r is given in teslas by

$$\vec{B}(\vec{r},\vec{\mu}) = \nabla\times A(\vec{r}) = \frac{\mu_0}{4\pi}\frac{3\vec{n}\cdot\vec{n}^{\mathsf{T}} - \mathsf{I}}{r^3}\vec{\mu} \tag{A.3}$$

where I is the identity and $(3\vec{n}\cdot\vec{n}^{\mathsf{T}} - \mathsf{I})$ is a 3x3 matrix. If we introduce a second dipole and consider the coupling energy between the two dipoles,

$$E(\vec{\mu_1},\vec{\mu_2},\vec{r}) = -\vec{\mu}_2^{\mathsf{T}}\vec{B}(\vec{r},\vec{\mu}_1) = -\vec{\mu}_2^{\mathsf{T}}\frac{\mu_0}{4\pi}\frac{3\vec{n}\cdot\vec{n}^{\mathsf{T}} - \mathsf{I}}{r^3}\vec{\mu}_1 \tag{A.4}$$

where subscripts 1 and 2 refer to two separate dipole moments, then we can write down a Hamiltonian operator for the system. First, let's remember that our dipoles are electrons, and write down their magnetic moments as

$$\vec{\mu} = -\mu_{\mathrm{B}} g\vec{S} \tag{A.5}$$

where μ_B is the Bohr magneton, g is the g-factor of the electron, and $\vec{S}$ is the spin angular momentum vector in atomic units. Assuming negligible exchange coupling, we can then write down the coupling Hamiltonian for the two-spin system in joules:

$$\hat{H}_{dd} = -\frac{\mu_0}{4\pi}{\mu_B}^2 g_1 g_2 \frac{1}{r^3} \hat{\vec{S}}_2^T \cdot [3\vec{n} \cdot \vec{n}^T - I] \cdot \hat{\vec{S}}_1. \tag{A.6}$$

where $\hat{\vec{S}}_i = [\hat{S}_x^i \; \hat{S}_y^i \; \hat{S}_z^i]^T$ is a vector of spin operators. In an EPR experiment, the spins sit in an external magnetic field, B_0, and interact with the field via the Zeeman interaction. This adds terms to the Hamiltonian:

$$\hat{H} = \hat{H}_{Z1} + \hat{H}_{Z2} + \hat{H}_{dd} \tag{A.7}$$

where $\hat{H}_{Zi} = \mu_B B_0 g_i \hat{S}_i$. If we divide both sides of this expression by $\hbar$, we obtain the spin-Hamiltonian in units of angular frequency

$$\begin{aligned} \frac{\hat{H}}{\hbar} &= \hat{\mathscr{H}} \\ &= \hat{\mathscr{H}}_{Z1} + \hat{\mathscr{H}}_{Z2} + \hat{\mathscr{H}}_{dd}. \end{aligned} \tag{A.8}$$

It is now useful to rewrite $\hat{\mathscr{H}}_{dd}$ as

$$\hat{\mathscr{H}}_{dd} = -\omega_\perp \hat{\vec{S}}_2^T d \hat{\vec{S}}_1^T = -\omega_\perp (A + B + C + D + E + F) \tag{A.9}$$

where the matrix $d = [3\vec{n} \cdot \vec{n}^T - I]$ and $\omega_\perp$ is the perpendicular part of the dipolar frequency and is given by:

$$\omega_\perp = \frac{\mu_0}{4\pi} \frac{\mu_B^2 g_1 g_2}{r^3 \hbar}. \tag{A.10}$$

Multiplying $\hat{\vec{S}}_2^T D \hat{\vec{S}}_1^T$ yields the so-called “dipolar alphabet" (Abragam’s Tome, Ch. 4, Sec. II) [86]:

$$\begin{aligned} \hat{A} &= \hat{S}_{1z}\hat{S}_{2z}(3\cos^2\theta - 1) \\ \hat{B} &= \frac{1}{4}(\hat{S}_{1+}\hat{S}_{2-} + \hat{S}_{1-}\hat{S}_{2+})(3\cos^2\theta - 1) \\ \hat{C} &= -\frac{3}{2}(\hat{S}_{1+}\hat{S}_{2z} + \hat{S}_{1z}\hat{S}_{2+})\sin\theta\cos\theta\, e^{-i\phi} \\ \hat{D} &= -\frac{3}{2}(\hat{S}_{1-}\hat{S}_{2z} + \hat{S}_{1z}\hat{S}_{2-})\sin\theta\cos\theta\, e^{+i\phi} \\ \hat{E} &= -\frac{3}{4}\hat{S}_{1+}\hat{S}_{2+}\sin^2\theta\, e^{-2i\phi} \\ \hat{F} &= -\frac{3}{4}\hat{S}_{1-}\hat{S}_{2-}\sin^2\theta\, e^{+2i\phi} \end{aligned} \tag{A.11}$$

where $\hat{S}_+ = \hat{S}_x + i\hat{S}_y$ and $\hat{S}_- = \hat{S}_x - i\hat{S}_y$ are the raising and lowering ladder operators, respectively. These terms are displayed in terms of a two-spin four-level energy diagram in Fig. A.1.

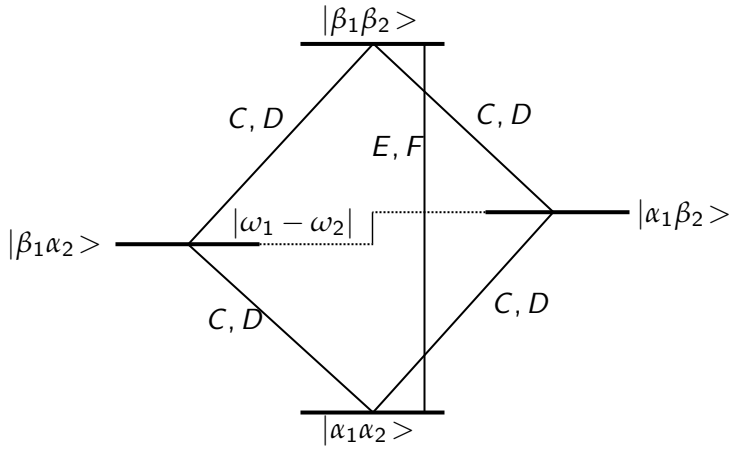

Figure A.1: Energy level diagram for a coupled two-spin system. The single-quantum ($\hat{C}$ and $\hat{D}$) and double-quantum ($\hat{E}$ and $\hat{F}$) EPR transitions are indicated.

This is an intimidating number of terms to worry about in $\hat{\mathscr{H}}_{\text{dd}}$. Fortunately, we can, for most experimental realities, neglect every dipolar alphabet term but $\hat{A}$.

In a Q-band EPR experiment, the sample sits in a strong magnetic field ($\sim$1.2 T), which brings the Zeeman terms, $\hat{\mathscr{H}}_{\text{Zi}}$, up to about 33-34 GHz. In contrast, the magnitude of $\hat{\mathscr{H}}_{\text{dd}}$ is on the order of only a few tens of MHz. This quantizes the electron spin-angular momenta along the B_0 axis, which we define as the z-axis, so the eigenvalues of the overall Hamiltonian are primarily those of the $\hat{S}_z$ operator.

The “alphabet" term $\hat{A}$ is the secular term that describes the coupling between two spins quantized along the z-axis, $\hat{B}$ is the pseudo-secular ‘flip-flop’ term that couples $|\beta\alpha\rangle$ and $|\alpha\beta\rangle$, $\hat{C}$ and $\hat{D}$ are the single-quantum EPR transitions, and $\hat{E}$ and $\hat{F}$ are the forbidden double-quantum transitions. In the high-field limit where the electron Zeeman interaction is much greater than the dipolar coupling, the only non-negligible term is $\hat{A}$. The flip-flop term, $\hat{B}$ can become relevant for high-spin systems [93].

This is our first major assumption, and it results in the simplified spin-Hamiltonian:

$$\hat{\mathscr{H}}_{\text{dd}} = -\omega_{\perp}(3\cos^2\theta - 1)\hat{S}_{1z}\hat{S}_{2z} \tag{A.12}$$

In the DEER experiment, an excess angle, or phase, accumulated by the spin as a function of time is given by

$$\phi_{\text{dd}}(t) = \omega_{\text{dd}} \cdot t. \tag{A.13}$$

It is this phase accumulation in the transverse plane due to the dipolar interaction that forms the basis of the DEER experiment.

A.2 DEER Signal

Here, we follow primarily Marko and Prisner's 2013 paper about the out-of-phase DEER signal because it contains one of the clearest descriptions available in the literature [94]. For a more complete story that begins with the equilibrium density matrix, I recommend that reference. Here, I assume the high-temperature limit where the out-of-phase signal is negligible and have simplified the expressions accordingly. The magnetization along the detection axis owing to a single observer spin coupled to a single pump spin, pumped at time t, is:

$$M(t) = \frac{\mathrm{i} g_i \mu_\mathrm{B}}{4} b_{\mathrm{pol.},1} \cos(\phi_{\mathrm{dd},kl}(t)) \tag{A.14}$$

where $\phi_{dd,kl}(t)$ is the phase offset accumulated at time t due to the dipolar coupling between spins k and l and $b_{\mathrm{pol.},i}$ is a Boltzmann polarization factor

$$b_{\mathrm{pol.},i} = -2 \tanh\left(\frac{\beta \hbar \omega_i}{2}\right) \tag{A.15}$$

where $\beta = (k_\mathrm{B} T)^{-1}$ and ω_i is the resonant frequency of spin i.

For a many-spin system, the magnetization is the sum of the contributions from each observer spin and each observer spin is affected by the product of the contributions from all pump spins:

$$M(t) = \frac{\mathrm{i} g_i \mu_\mathrm{B}}{4} b_{\mathrm{pol.},1} \sum_k^{N_\mathrm{a}} \prod_l^{N_\mathrm{b}} \left((1-\lambda) + \lambda \cos(\phi_{\mathrm{dd},kl}(t))\right). \tag{A.16}$$

where λ is the modulation depth, which characterizes the extent of spin-echo modulation and is a measure of the experiment efficiency. The sum in Eq. A.16 can be rewritten as:

$$M(t) = \frac{\mathrm{i} g_i \mu_\mathrm{B}}{4} b_{\mathrm{pol.},1} N_\mathrm{a} \overline{\prod_l^{N_\mathrm{b}} \left((1-\lambda) + \lambda \cos(\phi_{\mathrm{dd},kl}(t))\right)} \tag{A.17}$$

where $\overline{(\ldots)} = 1/N_\mathrm{a} \sum_k^{N_\mathrm{a}} (\ldots)$. Appealing to properties of statistical expectation values, Marko and coworkers assert that if the observed and pumped spin positions are not correlated, "then the average of the product can be substituted by the product of the averaged multiplicands, which are equal" [94]. This allows Eq. A.16 to be further rewritten as:

$$M(t) = \frac{\mathrm{i} g_i \mu_\mathrm{B}}{4} b_{\mathrm{pol.},1} N_\mathrm{a} \left((1-\lambda) + \lambda \cos(\phi_{\mathrm{dd},kl}(t))\right)^{N_\mathrm{b}}. \tag{A.18}$$

This expression must now be averaged over all possible observer spin positions using a powder integral:

$$I = \frac{1}{V} \int_0^\infty r_{kl}^2 \mathrm{d}r_{kl} \int_0^\pi \sin\theta_{kl} \mathrm{d}\theta_{kl} \int_0^{2\pi} \mathrm{d}\phi \cos(\phi_{\mathrm{dd},kl}(t)) \tag{A.19}$$

which, for a uniformly random homogenous distribution of monoradicals in the high-temperature limit, evaluates to:

$$M(t) = M_0 \exp\left(-\frac{8\pi^2}{9\sqrt{3}}\rho_b c \frac{\mu_0}{4\pi\hbar}\mu_B^2 g_a g_b t\right) = B(t) \tag{A.20}$$

as discussed in Ref. [86], Ch. 4, Sec. 4. This is the intermolecular background signal, $B(t)$, which complicates the analysis of DEER data. However, in a sample containing diradicals or other spin-clusters, the integral cannot be explicitly evaluated without direct knowledge of the intra-cluster spin-spin distance distribution, $P(r)$. Accounting for this additional complication results in the expression:

$$M(t) = B(t)\left((1-\lambda) + \lambda\int_0^\infty \mathrm{d}r\ P(r)\int_0^{\pi/2}\mathrm{d}\theta\ \sin\theta\cos\left[(1-3\cos^2\theta)\omega_\perp t\right]\right) \tag{A.21}$$

where $\omega_\perp = \mu_0\mu_B^2 g_A g_B/(4\pi\hbar r^3)$.

A.3 Expressing the Dipolar Kernel as Fresnel Integrals

If we make the substitution

$$z = \cos\theta$$

in the kernel of the dipolar powder integral, we obtain for the kernel:

$$K(t,r) = \int_0^1 \mathrm{d}z\ \cos\left((1-3z^2)\omega_\perp t\right) \tag{A.22}$$

$$K(t,r) = \int_0^1 \mathrm{d}z\ \cos\left(\omega_\perp t - 3z^2\omega_\perp t\right) \tag{A.23}$$

$$\tag{A.24}$$

Recall that $\cos(a-b) = \cos(a)\cos(b) + \sin(a)\sin(b)$, then:

$$K(t,r) = \int_0^1 \mathrm{d}z\ \left(\cos(\omega_\perp t)\cos\left(3z^2\omega_\perp t\right) + \sin(\omega_\perp t)\sin\left(3z^2\omega_\perp t\right)\right) \tag{A.25}$$

$$K(t,r) = \cos(\omega_\perp t)\int_0^1 \mathrm{d}z\ \cos\left(3z^2\omega_\perp t\right) + \sin(\omega_\perp t)\int_0^1 \mathrm{d}z\ \sin\left(3z^2\omega_\perp t\right) \tag{A.26}$$

With some foresight, we rewrite 3 as $6\pi/2\pi$ and then perform a substitution of variables to obtain

the Fresnel integrals:

$$K(t,r) = \cos(\omega_\perp t)\int_0^1 \mathrm{d}z\ \cos\left(\frac{6\pi}{2\pi}z^2\omega_\perp t\right) + \sin(\omega_\perp t)\int_0^1 \mathrm{d}z\ \sin\left(\frac{6\pi}{2\pi}z^2\omega_\perp t\right) \tag{A.27}$$

$$K(t,r) = \frac{\cos(\omega_\perp t)\int_0^{\sqrt{\frac{6}{\pi}\omega_\perp t}}\mathrm{d}x\ \cos\left(\frac{\pi}{2}x^2\right) + \sin(\omega_\perp t)\int_0^{\sqrt{\frac{6}{\pi}\omega_\perp \mathrm{d}x\ t}}\sin\left(\frac{\pi}{2}x^2\right)}{\sqrt{\frac{6}{\pi}\omega_\perp t}} \tag{A.28}$$

$$K(t,r) = \frac{\cos(\omega_\perp t)\mathcal{C}(\sqrt{\frac{6}{\pi}\omega_\perp t}) + \sin(\omega_\perp t)\mathcal{S}(\sqrt{\frac{6}{\pi}\omega_\perp t})}{\sqrt{\frac{6}{\pi}\omega_\perp t}} \tag{A.29}$$

where $\mathcal{C}(...)$ and $\mathcal{S}(...)$ are the cosine and sine Fresnel integrals, respectively. You might notice that the last two expressions are undefined for $t = 0$. We account for this with an 'if' statement in our Matlab code that sets $K(0, r) = 1$.

Appendix B

REGULARIZATION PARAMETER SELECTION METHODS

Here, we summarize all α selection methods included in Ch. 2 and published in Ref. [13] and give the expressions necessary to implement them. Further details about the methods can be found in the cited statistics literature.

Most methods are derived assuming an unconstrained Tikhonov regularization, i.e. without the $P \geq 0$ constraint in Eq. (1.7). In this case, the solution $P_{\alpha L}$ can be expressed in closed form as

$$P_{\alpha L} = \bar{K}_{\alpha L} S \qquad \text{with} \qquad \bar{K}_{\alpha L} = \left(K^{\mathsf{T}} K + \alpha^2 L^{\mathsf{T}} L \right)^{-1} K^{\mathsf{T}} \tag{B.1}$$

For the physically relevant constrained problem (with $P \geq 0$), a closed-form solution is not possible, and we obtain $P_{\alpha L}$ via an iterative optimization algorithm [77]. In either case, once $P_{\alpha L}$ is obtained, the time-domain fit is

$$S_{\alpha L} = K P_{\alpha L} \tag{B.2}$$

We apply each method described below in two ways, one using the unconstrained solutions and one using the constrained solutions. The methods are referred to by short acronyms (see Table 2.1). A suffix 'u' is appended to the method acronym when unconstrained $P_{\alpha L}$ are used. In addition, if the method contains a tuning or scaling parameter, then its value is appended to the acronym as well.

L-Curve methods (LC, LR, LR2) Several methods select α based on heuristic considerations about a parametric log-log plot of the Tikhonov penalty term, $\eta = \|L P_{\alpha L}\|$, against the residual norm, $\rho = \|S - S_{\alpha L}\|$, as a function of α [95, 96]. Such a plot shows a monotonically decaying $\eta(\rho)$ and is called the L-curve since it tends to form a characteristic "L" shape with a visual corner. The optimal value for α is considered to correspond to this corner, since it intuitively represents a reasonable balance between the fitting error ρ and the regularization error η. The corner is not a mathematically defined quantity, therefore different operational definitions of locating this corner exist.

One of them is the maximum-curvature method (LC) [96]. It selects the α corresponding to the point of maximum (positive) curvature of the L-curve, given by

$$\alpha_{\text{sel}} = \underset{\alpha}{\text{argmax}} \, \frac{\hat{\rho}' \hat{\eta}'' - \hat{\rho}'' \hat{\eta}'}{(\hat{\rho}'^2 + \hat{\eta}'^2)^{3/2}} \tag{B.3}$$

where $\hat{\eta} = \log_{10} \|L P_{\alpha L}\|$, $\hat{\rho} = \log_{10} \|S - S_{\alpha L}\|$, and the primes indicate derivatives with respect

to $\log_{10}\alpha$ or α. This method picks the global maximum of the curvature, even though there can be several local maxima.

Another possible definition of corner is the point closest to the lower left corner of the L-curve plot [95]. `DeerAnalysis` uses one implementation of this idea [18]. The two coordinates $\hat{\rho}$ and $\hat{\eta}$ are evaluated over a range of α values and then rescaled to the interval [0,1]. The corner is determined as the location on the L-curve that is closest to the origin in these rescaled coordinates:

$$\alpha_{\mathrm{sel}} = \underset{\alpha}{\operatorname{argmin}} \left(\left(\frac{\hat{\rho} - \hat{\rho}_{\mathrm{min}}}{\hat{\rho}_{\mathrm{max}} - \hat{\rho}_{\mathrm{min}}} \right)^2 + \left(\frac{\hat{\eta} - \hat{\eta}_{\mathrm{min}}}{\hat{\eta}_{\mathrm{max}} - \hat{\eta}_{\mathrm{min}}} \right)^2 \right) \tag{B.4}$$

The results from this method depend on the α range. We refer to this method as the minimum-radius method and use two variants of it. In the first one (LR), we use the same α range as for all other methods (10^{-3} to 10^{3}), and in the other one (LR2), we employ the same α range as `DeerAnalysis`. The latter extends from the largest generalized singular eigenvalue of K and L, s_{max}, to a value cs_{max} with $c = 16\epsilon \cdot 10^{6} \cdot 2^{\sigma/0.0025}$, where $\epsilon \approx 2.2 \cdot 10^{-16}$ and σ is the noise level. In order to keep the range large enough, we limit c to $\leq 10^{-6}$.

Cross validation (CV, GCV, mGCVc, rGCVγ, srGCVγ) We use several methods based on the idea of cross validation. Leave-one-out cross validation (CV) [97, 98] is conceptually the simplest of them. For a given α, it minimizes the total prediction error, which is obtained as the sum of the prediction errors for each individual data point. For that, a single data point is excluded and a fit to the remaining data is calculated. That fit is judged based upon its ability to reproduce the excluded data point. By repeating the method for each data point, the total prediction error is obtained, and the α value is selected that minimizes it. This procedure can be condensed into the following expression:

$$\alpha_{\mathrm{sel}} = \underset{\alpha}{\operatorname{argmin}} \left(\sum_{i=1}^{n_t} \left| \frac{S(i) - S_{\alpha L}(i)}{1 - H_{\alpha L}(i,i)} \right|^2 \right) \tag{B.5}$$

with the α-dependent influence matrix $H_{\alpha L} = K\bar{K}_{\alpha L}$.

Generalized cross validation (GCV) [99, 100] is very similar to leave-one-out cross validation, but the diagonal matrix element $H_{\alpha L}(i,i)$ in the denominator is replaced by the average of all diagonal elements, rendering the method more stable. The expression simplifies to

$$\alpha_{\mathrm{sel}} = \underset{\alpha}{\operatorname{argmin}} \frac{\|S - S_{\alpha L}\|^2}{\left[1 - \mathrm{tr}(H_{\alpha L})/n_t\right]^2} \tag{B.6}$$

The modified GCV (mGCVc) [101, 102] method is a simple tunable modification to GCV intended to stabilize the method further.

$$\alpha_{\mathrm{sel}} = \underset{\alpha}{\operatorname{argmin}} \frac{\|S - S_{\alpha L}\|^2}{\left[1 - c\,\mathrm{tr}(H_{\alpha L})/n_t\right]^2} \tag{B.7}$$

with the tuning parameter $c \geq 1$. We use this with $c = 1.2$, 1.5, 2, and 3, to test a range of stabilizations.

The robust GCV (rGCVγ) method [103, 104] is also designed to exhibit greater stability than the GCV method. It is given by

$$\alpha_{\text{sel}} = \underset{\alpha}{\text{argmin}} \left(\frac{\|S - S_{\alpha L}\|^2}{\left[1 - \text{tr}(H_{\alpha L})/n_t\right]^2} \left(\gamma + (1-\gamma)\,\text{tr}\left(H_{\alpha L}^2\right)/n_t\right) \right) \tag{B.8}$$

with the tuning parameter $\gamma \leq 1$. With $\gamma = 1$, the method reduces to GCV. As γ gets smaller, the method becomes increasingly stable and less likely to undersmooth. We examine γ values of 0.1, 0.5, and 0.9, with increasing contributions from the second term.

Strong robust GCV (srGCVγ) [105] is another tunable modification to GCV that is based on stronger statistical arguments than rGCV. It selects α via

$$\alpha_{\text{sel}} = \underset{\alpha}{\text{argmin}} \left(\frac{\|S - S_{\alpha L}\|^2}{\left[1 - \text{tr}(H_{\alpha L})/n_t\right]^2} \left(\gamma + (1-\gamma)\,\text{tr}\left(\bar{K}_{\alpha L}^{\mathsf{T}} \bar{K}_{\alpha L}\right)/n_t\right) \right) \tag{B.9}$$

with $\gamma \geq 1$. Like rGCV, with $\gamma = 1$, the method reduces to GCV. We use γ values of 0.8 and 0.95.

Quasi-optimality (QO) This criterion, from Tikhonov and coworkers [106–108], selects α such that a small change in α from that selected value has minimal effects on the resulting $P_{\alpha L}$:

$$\alpha_{\text{sel}} = \underset{\alpha}{\text{argmin}} \left\| \frac{\text{d}P_{\alpha L}}{\text{d}\log_{10}\alpha} \right\| \tag{B.10}$$

The underlying rationale is that the model recovery error is flat at its minimum.

Discrepancy principle (DPτ) This principle [109–111] is predicated upon the idea that the root-mean-square residual, $\|S - S_{\alpha L}\|/\sqrt{n_t}$, should be on the order of the noise standard deviation in the data, σ. It requires *a priori* knowledge of σ. The value for α is selected as the largest value such that

$$\|S - S_{\alpha L}\|/\sqrt{n_t} \leq \tau\sigma \tag{B.11}$$

where τ is a safety factor ≥ 1 to guard against undersmoothing in the case σ is underestimated. We use this principle with $\tau = 1$ and 1.5.

The transformed discrepancy principle (tDPτ) [112–114] performs the comparison in the distance domain, choosing the largest α that satisfies

$$\|\bar{K}_{\alpha L}(S - S_{\alpha L})\|/\sqrt{n_r} \leq \tau\frac{b_0\sigma}{\alpha} \tag{B.12}$$

where $b_0 = 3\sqrt{3}/16 \approx 0.325$. We use $\tau = 1.5$.

Balancing principle (BP, hBP) The balancing principle (BP) [115, 116] balances the propagated noise error with the unknown model recovery error. Using

$$B(\alpha) = \max_{\alpha' < \alpha} \left(\max_{\alpha'' < \alpha'} \frac{\|P_{\alpha'} - P_{\alpha''}\|}{4\sigma\phi(\alpha'')} \right) \tag{B.13}$$

and $\phi(\alpha) = \sqrt{\mathrm{tr}\left(\bar{K}_{\alpha L}\bar{K}_{\alpha L}^{\mathsf{T}}\right)}$, it selects α_{sel} as the largest α value that satisfies $B(\alpha) \leq 1$. The hardened balancing principle (hBP) [116] selects α using

$$\alpha_{\mathrm{sel}} = \underset{\alpha}{\mathrm{argmin}} \left(B(\alpha)\sqrt{\phi(\alpha)} \right) \tag{B.14}$$

Residual method (RM) This method [74, 117] determines α by minimizing a scaled norm of the residual vector

$$\alpha_{\mathrm{sel}} = \underset{\alpha}{\mathrm{argmin}} \frac{\|S - S_{\alpha L}\|^2}{\sqrt{\mathrm{tr}(B^{\mathsf{T}}B)}} \tag{B.15}$$

where $B = K^{\mathsf{T}}\left(\mathrm{I} - H_{\alpha L}\right)$. The scaling penalizes undersmoothing.

Self-consistency method (SC) This method [51, 52], utilized in `FTIKREG`, seeks to minimize the sum of the estimated model recovery error and the propagated data noise variance.

$$\alpha_{\mathrm{sel}} = \underset{\alpha}{\mathrm{argmin}} \left(\|(\mathrm{I} - \bar{K}_{\alpha L}K)P_{\alpha L}\|^2 + \sigma^2 \mathrm{tr}\left(\bar{K}_{\alpha L}\bar{K}_{\alpha L}^{\mathsf{T}}\right) \right) \tag{B.16}$$

This expression is valid for the unconstrained case. In order to include the non-negativity constraint, further steps are necessary. First, α_{sel} for the unconstrained case is determined. Next, the constrained solution is determined using this α value and the indices q of the active non-negativity constraints are stored. Finally, the columns of K and L with indices q are removed and Eq. (B.16) is re-evaluated.

Generalized maximum likelihood (GML) This method [118] selects the α that maximizes the likelihood (or, equivalently, minimizes the negative log-likelihood) and is given by

$$\alpha_{\mathrm{sel}} = \underset{\alpha}{\mathrm{argmin}} \frac{S^{\mathsf{T}}(S - S_{\alpha L})}{\sqrt[m]{\det_{\mathrm{nz}}(\mathrm{I} - H_{\alpha L})}} \tag{B.17}$$

where $\det_{\mathrm{nz}}(\cdot)$ indicates the product of the non-zero eigenvalues, and m is the number of non-zero eigenvalues. To account for numerical errors, we treat all eigenvalues with magnitude below 10^{-8} as zero. If there are no non-zero eigenvalues less than 10^{-8}, then m is equal to the rank of $\mathrm{I} - H_{\alpha L}$).

Extrapolated error (EE) This method [119–121] minimizes an estimate of the regularization error via

$$\alpha_{\text{sel}} = \underset{\alpha}{\text{argmin}} \frac{\|S - S_{\alpha L}\|^2}{\|K^{\mathsf{T}}(S - S_{\alpha L})\|} \tag{B.18}$$

Normalized cumulative periodogram (NCP) This method [122–124] is based on the idea that the power spectrum of the residuals should match the power spectrum of the noise. The unscaled power spectrum (periodogram) for a given α is a vector p with elements

$$p(k) = \left|\text{dft}\left(S - S_{\alpha L}\right)_k\right|^2 \tag{B.19}$$

where $k = 1, \ldots, n_t$, dft refers to the discrete Fourier transform. The normalized cumulative periodogram is an $(n_t - 1)$-element vector c with elements

$$c(i) = \frac{\|p(2\!:\!i+1)\|_1}{\|p(2\!:\!n_t)\|_1} \tag{B.20}$$

where $\|\ldots\|_1$ is the ℓ_1 norm (sum of absolute values). The zero-frequency component, p_1, is omitted. c represents the integrated power spectrum of the residuals. The α value is selected that minimizes the deviation between c and the integrated power spectrum c_{noise} expected for the noise:

$$\alpha_{\text{sel}} = \underset{\alpha}{\text{argmin}} \left\|c(\alpha) - c_{\text{noise}}\right\|_1 \tag{B.21}$$

where c_{noise} is a $(n_t - 1)$-element vector, with elements $c_{\text{noise},i} = i/(n_t - 1)$ for white noise.

Mallows' C_L (MCL) This method [125] minimizes an approximation to the model recovery error, derived under the assumption of unconstrained regularization and uncorrelated Gaussian noise.

$$\alpha_{\text{sel}} = \underset{\alpha}{\text{argmin}} \left(\|S - S_{\alpha L}\|^2 + 2\sigma^2 \text{tr}(H_{\alpha L}) - 2n_t\sigma^2\right) \tag{B.22}$$

This requires the knowledge of the noise level σ.

Information-theoretical criteria (AIC, AICC, BIC, ICOMP) In the context of information theory [126, 127], the set of $P_{\alpha L}$ is regarded as a set of candidate models, and criteria have been developed that select a parsimonious model that balances a small fitting error with a low model complexity. The general expression is

$$\alpha_{\text{sel}} = \underset{\alpha}{\text{argmin}} \left(n_t \ln \frac{\|S - S_{\alpha L}\|^2}{n_t} + c\, \text{tr}(H_{\alpha L})\right) \tag{B.23}$$

where the second term is a measure of model complexity, which decreases with increasing α. The term $\text{tr}(H_{\alpha L})$ can be regarded as the effective number of free parameters in the model. The

constant c depends on the particular criterion: $c = 2$ for the Akaike information criterion (AIC) [128], $c = 2n_t/(n_t - \mathrm{tr}(H_{\alpha L}) - 1)$ for the corrected AIC (AICC) [129, 130], and $c = \ln(n_t)$ for the Bayesian information criterion (BIC) [131]. The information complexity criterion (ICOMP) [132, 133] is yet another information-theoretical procedure and is given by

$$\alpha_{\mathrm{sel}} = \underset{\alpha}{\mathrm{argmin}} \left(\frac{1}{\sigma^2} \left\| S - S_{\alpha L} \right\|^2 + 2\,\mathrm{tr}(H_{\alpha L}) + n_t \ln \frac{\bar{s}_{\mathrm{a}}}{\bar{s}_{\mathrm{g}}} \right) \tag{B.24}$$

where $\bar{s}_{\mathrm{a}}$ and $\bar{s}_{\mathrm{g}}$ are the arithmetic and geometric means of the singular values of $(K^{\mathsf{T}}K + \alpha^2 L^{\mathsf{T}}L)^{-1}$. The last term penalizes for interdependence among model parameters. In contrast to the other information-theoretical criteria, this one requires knowledge of the noise level σ.

Appendix C

SUPPORTING INFORMATION FOR CHAPTER 4

C.1 Literature Search Citations

We examined the DEER distance distributions included in over 100 papers published in 2014, 2015, and early 2016 — including their SIs [12, 21, 37, 69, 81, 134–234] — and found that 90% did not show any kind of quantitative error estimation for $P(r)$. The DEER $P(r)$ was displayed as a bare curve.

Those papers that did display error estimates used the following methods:

- In [137–139, 208], the authors used the DeerAnalysis Validation Tool [18], varying the background dimensionality and fit range.

- In [141], the authors used the DeerAnalysis Validation Tool, varying the background fit range. The authors also used the pruning tool to exclude very bad background choices from the calculation.

- In [151, 193], the authors used the DeerAnalysis Validation Tool, varying the background dimensionality, fit range, and by varying the added noise realizations.

- In [81, 212], the authors linearly propagated the time-domain fit error vector through Tikhonov regularization and plotted the results added to and subtracted from P.

- In [177], the authors used the DeerAnalysis Validation Tool, varying unspecified parameters. In [149], the authors did the same for one data set in the SI.

- In [207, 220], the authors used the DeerAnalysis Validation Tool, varying the background fit range. Error regions were shown only for a few sample data sets in the SI, and none were shown in the main text.

- In [166, 179], the authors indicated that the long-distance regions of the P solutions are less reliable than the short distances, based upon the length of S, by graying/dashing longer regions of P. In [194], the authors grayed out all of the P vectors for $r > 4$ nm, though the data traces S varied in length. This approach does not estimate error, but does indicate the increased uncertainty associated with measuring longer distances.

C.2 Noise Quantification

Fig. C.1 shows background-corrected DEER data collected at Q-band using the 4-pulse sequence [41]. The sample was a frozen glassy solution of 70 μM MTSL-labeled maltose binding protein (MBP) in the presence of substrate (maltose) at 50 K. For both the pump and probe sequences, π-pulses of length 44 ns were used. The initial $\frac{\pi}{2}$-pulse in the probe sequence was 24 ns long. The pump-pulse frequency was set 70MHz above the probe pulse. The pump sequence was carried out on the spectral maximum. The probe frequency was centered at the resonator dip. An 8-step phase cycle was used. Each phase-cycled trace was recorded separately. Each trace was individually phase-corrected before averaging. Because every trace was available, it was possible to sort the data into histograms at each t_i. This revealed that the noise was approximately Gaussian (Fig. C.1 top) and uncorrelated with t (Fig. C.1 bottom). After background correction, a slight correlation with t is introduced, but it remains negligible so long as the background is not too steep. It is therefore justifiable to calculate the standard deviation, σ_S , at each t_i (Fig. C.1), and use the mean value of σ_S to characterize the noise.

C.3 Generating random samples from the posterior PDF

In order to generate random samples for P from Eq. (3.7),

$$\pi(P,\delta|S,\tau) \propto \left(\frac{\tau}{2\pi}\right)^{\frac{n_t}{2}} \left(\frac{\delta}{2\pi}\right)^{\frac{n_r}{2}} \frac{b^a}{\Gamma(a)}\delta^{a-1}\mathrm{e}^{-b\delta} \exp\left\{\left(-\frac{\tau}{2}\left\|S-KP\right\|^2 - \frac{\delta}{2}\left\|LP\right\|^2\right)\right\},$$

we first show that this expression is Gaussian in P. We begin by omitting prefactors independent of P, expanding the norms, and grouping like terms:

$$\begin{aligned}\pi(P|S,\delta,\tau) &\propto \exp\left(-\frac{\tau}{2}\left\|S-KP\right\|^2 - \frac{\delta}{2}\left\|LP\right\|^2\right) \\ &\propto \exp\left(-\frac{1}{2}P^{\mathsf{T}}\left(\tau K^{\mathsf{T}}K + \delta L^{\mathsf{T}}L\right)P + \tau P^{\mathsf{T}}K^{\mathsf{T}}S - \frac{\tau}{2}S^{\mathsf{T}}S\right)\end{aligned} \tag{C.1}$$

where we have used $P^{\mathsf{T}}K^{\mathsf{T}}S = S^{\mathsf{T}}KP$. We then complete the square in P:

$$\pi(P|S,\delta,\tau) \propto \exp\left(-\frac{1}{2}\left(P-\tau\Sigma K^{\mathsf{T}}S\right)^{\mathsf{T}}\Sigma^{-1}\left(P-\tau\Sigma K^{\mathsf{T}}S\right) - \frac{\tau}{2}S^{\mathsf{T}}S - \frac{1}{2}\tau^2 S^{\mathsf{T}}K\Sigma^{\mathsf{T}}K^{\mathsf{T}}S\right) \tag{C.2}$$

with $\Sigma = \left(\tau K^{\mathsf{T}}K + \delta L^{\mathsf{T}}L\right)^{-1}$. The last two terms do not depend on P, so the expression reduces to:

$$\pi(P|S,\delta,\tau) \propto \mathrm{Normal}\left(\tau\Sigma K^{\mathsf{T}}S, \Sigma\right) \tag{C.3}$$

where Σ is the covariance matrix and $\tau\Sigma K^{\mathsf{T}}S$ is equal to the unconstrained Tikhonov solution if $\alpha^2 = \delta/\tau$. Eq. (C.2) shows that Eq. (3.7) is Gaussian in P with well-defined mean, $\tau\Sigma K^{\mathsf{T}}S$, and covariance, Σ, and therefore amenable to sampling by pseudo-random number generation.

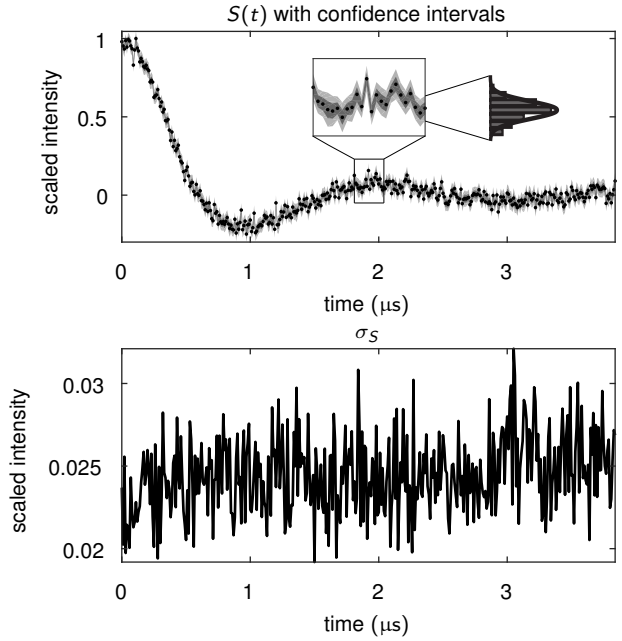

Figure C.1: Noise statistics on DEER time-domain traces. Top: background-corrected time-domain trace. Averages are shown as black points, the 2σ band as light gray, and the 1σ band as dark gray. Inset: blowup of the data and a histogram of a particular point in the time domain, showing that the experimental noise is approximately normally distributed. Bottom: standard deviation of the noise, σ_S, for each t_i in the trace above. The noise level is approximately independent of time, so long as the background is not too steep. A slight correlation with time has been introduced in this case, but it is negligible. The data were collected at Q-band on MTSL-labeled E22C A338C maltose binding protein (MBP) in the presence of maltose.

In order to generate random draws from Eq. (C.3), we will follow a common method for generating random variates x from the multivariate normal where the mean is added to an appropriate random vector [235]:

$$x = \mu + C_{\mathrm{L}}\nu \tag{C.4}$$

where the vector μ is the mean of the distribution, the matrix C_{L} is the lower Cholesky factor of the matrix Σ, and the vector ν is drawn from $\mathrm{Normal}(0, I)$. Following Bardsley, we write a modified version of Eq. (1.7) that will simultaneously generate our mean and random vector [76]:

$$\text{sample } P = \min_{P \geq 0} \left[\frac{1}{2} P^{\mathsf{T}} \Sigma^{-1} P - P^{\mathsf{T}} \left(\tau K^{\mathsf{T}} S + w \right) \right] \tag{C.5}$$

where

$$w \sim \mathrm{Normal}(0, \Sigma^{-1}). \tag{C.6}$$

Because $\Sigma^{-1} = (C_{\mathrm{L}} C_{\mathrm{L}}^{\mathsf{T}})^{-1}$The random vector w can be calculated by:

$$w = (C_{\mathrm{L}}^{-1})^{\mathsf{T}} \nu. \tag{C.7}$$

The argument of Eq. (C.5) is equivalent to the exponent in Eq. (C.1) if the vector w and the constant $\frac{\tau}{2} S^{\mathsf{T}} S$ are set to zero. The minimum found by Eq. (C.5) is therefore the maximum of Eq. (C.1) perturbed by a random vector. As Σ is the precision (inverse-covariance) matrix of the posterior distribution of P, its Cholesky factor C_{L} can be considered the matrix equivalent of a standard deviation. We can see that Eq. (C.5) yields samples in the form of Eq. (C.4) by taking the derivative of Eq. (C.5) with respect to P, setting the result equal to zero, and solving for P to obtain $\Sigma K^{\mathsf{T}} S + C_{\mathrm{L}} \nu$ [236].

The above analysis also holds for the non-negativity constrained case, see the discussion in [76].

C.4 Single Parameter Gibbs Sampling

In order to examine the effect of α upon the resulting Bayesian credible intervals, we ran the Gibbs sampler from Sec. (3.4) with fixed δ (fixed α) for the unimodal mid-SNR case from Fig. (3.2). The results are shown in Fig. C.2. The variance in P varies inversely with the magnitude of α. This is a consequence of the known tradeoff between bias and variance in model fitting [54, 55]. If the regularization procedure is strongly biased towards a particular subset of possible solutions, the potential variance in the observed solutions will be small.

C.5 Convergence Criteria

For testing the convergence in the means of δ and P, we use a multi-chain convergence assessment recommended by Bardsley [76]. This method, developed by Gelman and Rubin [55],

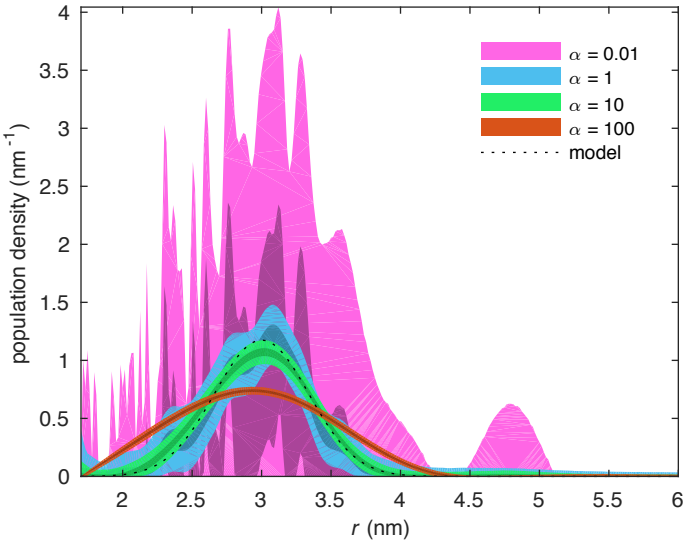

Figure C.2: Bayesian credible intervals resulting from fixed values of α corresponding to the unimodal $\tau = 900$ case from Fig. (3.2). The spread of the credible interval increases with decreasing α. This reflects the regularizing nature of Tikhonov regularization, where larger values of α impose greater degrees of smoothness. This result shows a tradeoff between solution variance and bias: the solutions corresponding to larger αs feature less variance, but they are constrained by greater bias. This is a consequence of the so-called 'bias-variance tradeoff'.

requires that multiple chains be run from different starting points. For a convergence check, the first half of each chain is discarded as burn-in. This is a very conservative choice, as the chains for this application appear to converge to regions of high probability very quickly (Fig. (C.3)). Each truncated chain is split in half, doubling the number of chains. If n total iterations have been run, each truncated chain used in the convergence check will have $k = n/4$ points. Some scalar statistic calculated from the sampled values is statistically analyzed as a function of iteration number and compared both within and between separate chains. Because the first and second halves of each truncated chain from the run are tested against each other, this process also controls for stationarity: if a chain has not yet reached the stationary distribution, then the latter half of the chain will be significantly different from the former.

The particular scalar statistic is denoted ψ_{ij}, where $i = 1, ..., k$ counts the iterations of the truncated chains and $j = 1, ..., m$ counts the chains. We calculate the inter-chain variance:

$$B = \frac{k}{m-1} \sum_{j=1}^{m} (\psi_{.j} - \psi_{..})^2 \tag{C.8}$$

where $\psi_{.j} = \frac{1}{k} \sum_{i=1}^{k} \psi_{ij}$ and $\psi_{..} = \frac{1}{k} \sum_{i=1}^{k} \psi_{.j}$. The intra-chain variance is:

$$W = \frac{1}{m} \sum_{j=1}^{m} s_j^2 \tag{C.9}$$

where $s_j^2 = \frac{1}{k-1} \sum_{i=1}^{k} (\psi_{ij} - \psi_{.j})^2$. Using W and B, we calculate an estimate for $\mathrm{var}(\psi|S)$, the marginal posterior variance of the estimand ψ:

$$\mathrm{var}(\psi|S) = \frac{k-1}{k} W + \frac{1}{k} B. \tag{C.10}$$

For finite k, this quantity over-estimates the marginal posterior variance, and W is an underestimate. Therefore, we can test convergence of the MCMC procedure by comparing $\mathrm{var}(\psi|S)$ and W:

$$\hat{R} = \sqrt{\frac{\mathrm{var}(\psi|S)}{W}} \tag{C.11}$$

which converges to 1 as $k \to \infty$. We set a target value close to 1 for $\hat{R}$ and consider the chains converged when $\hat{R}$ is smaller than that target. For α, we use 1.05 and for the mean of P, we use 1.1. Gelman suggests 1.1 as a reasonable value [55].

Fig. C.3 shows the first 300 iterations of each of the six δ chains, converted to α values, for each of the six cases considered in 3.2. Within a given data set, the six chains all rapidly converge to the same value in less than 50 iterations and explore the same probability space. This demonstrates that our choice of discarding the first half of all iterations for convergence testing is extremely safe. Also plotted is the $\hat{R}$ value, which rapidly drops towards 1 as the chains converge.

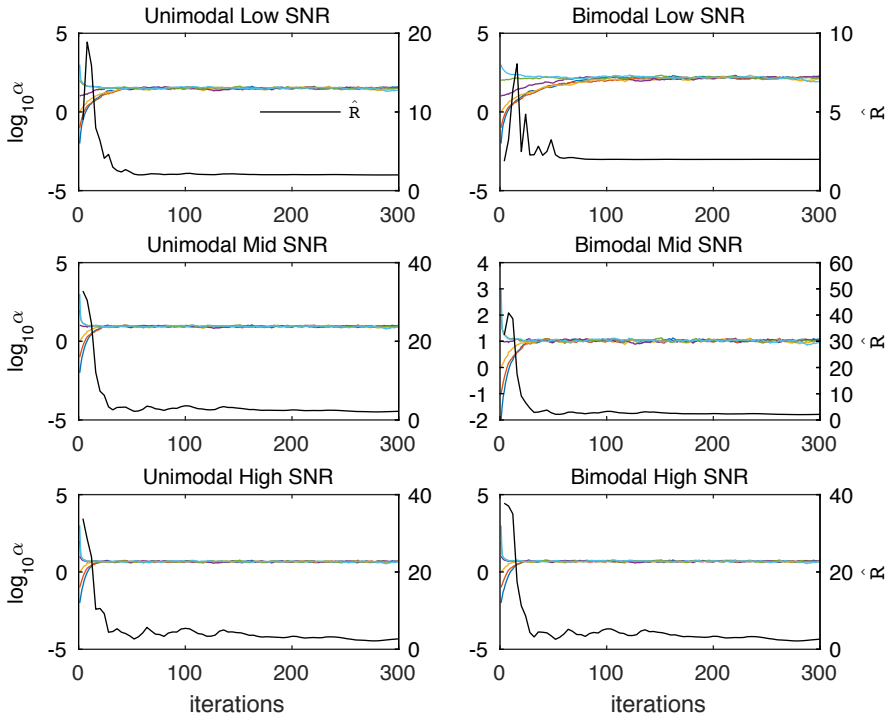

Figure C.3: First 300 iterations of the Gibbs sampler output for α for each of the six datasets from Fig. (3.2) are plotted in color. The convergence assessment parameter $\hat{R}$ is plotted in black. For each MCMC run, six parallel chains were run. The chains quickly find and explore the same high probability region, which is often called good mixing. The chains also remain in the same region in probability space, which indicates that they have achieved stationarity, which is required for convergence.

BIBLIOGRAPHY

[1] Milov A. D., Salikhov K. M., and Shchirov M. D. Application of the double resonance method to electron spin echo in a study of the spatial distribution of paramagnetic centers in solids. *Fiz. Tverd. Tela*, 23:975–982, 1981.

[2] Milov A. D., Ponomarev A. B., and Tsvetkov Yu. D. Electron-Electron Double Resonance in Electron Spin Echo: Model Biradical Systems and the Sensitized Photolysis of Decalin. *Chem. Phys. Lett.*, 110:67–72, 1984.

[3] Schmidt T., Wälti M. A., Baber J. L., Hustedt E. J., and Clore G. M. Long Distance Measurements up to 160 Å in the GroEL Tetradecamer Using Q-Band DEER EPR Spectroscopy. *Angew. Chem. Int. Ed.*, 55:15905–15909, 2016.

[4] Altenbach C., Flitsch S. L., Khorana H. G., and Hubbell W. L. Structural studies on transmembrane proteins. 2. Spin labeling of bacteriorhodopsin mutants at unique cysteines. *Biochemistry*, 28:7806–7812, 1989.

[5] Altenbach C., Marti T., Khorana H. G., and Hubbell W. L. Transmembrane protein structure: spin labeling of bacteriorhodopsin mutants. *Science*, 248:1088–1092, 1990.

[6] Jeschke G. DEER distance measurements on proteins. *Annu. Rev. Phys. Chem.*, 63:419–446, 2012.

[7] Timmel C. R. and Harmer J. R. *Structural Information from Spin-Labels and Intrinsic Paramagnetic Centres in the Biosciences*. Springer, 2014.

[8] Qin P. Z. and Warncke K. *Methods in Enzymology: Electron Paramagnetic Resonance Investigations of Biological Systems by Using Spin Labels, Spin Probes, and Intrinsic Metal Ions, Part A*, volume 563. Elsevier, 2015.

[9] Qin P. Z. and Warncke K. *Methods in Enzymology: Electron Paramagnetic Resonance Investigations of Biological Systems by Using Spin Labels, Spin Probes, and Intrinsic Metal Ions, Part B*, volume 563. Elsevier, 2015.

[10] Jeschke G. The contribution of modern EPR to structural biology. *Emerging Top. Life Sci.*, page ETLS20170143, 2018.

[11] Stoll S., Lee Y., Zhang M., Wilson R. F., Britt R. D., and Goodin D. B. Double electron–electron resonance shows cytochrome P450cam undergoes a conformational change in solution upon binding substrate. *Proc. Natl. Acad. Sci. USA*, 109:12888–12893, 2012.

[12] Baber J. L, Louis J. M., and Clore M. G. Dependence of distance distributions derived from double electron-electron resonance pulsed epr spectroscopy on pulse-sequence time. *Angew. Chem. Int. Ed.*, 54:5336–5339, 2015.

[13] Edwards T. H. and Stoll S. Optimal Tikhonov regularization for DEER spectroscopy. *J. Magn. Reson.*, 288:58–68, 2018.

[14] Edwards T. H. and Stoll S. A Bayesian approach to quantifying uncertainty from experimental noise in DEER spectroscopy. *J. Magn. Reson.*, 270:87–97, 2016.

[15] Jeschke G., Panek G., Godt A., Bender A., and Paulsen H. Data analysis procedures for pulse ELDOR measurements of broad distance distributions. *Appl. Magn. Reson.*, 26:223–244, 2004.

[16] Bowman M. K., Maryasov A. G., Kim N., and DeRose V. J. Visualization of distance distributions from pulse double electron-electron resonance data. *Appl. Magn. Reson.*, 26:23–39, 2004.

[17] Chiang Y., Borbat P. P., and Freed J. H. The determination of pair distance distributions by pulsed ESR using Tikhonov regularization. *J. Magn. Reson.*, 172:279–295, 2005.

[18] Jeschke G., Chechik V., Ionita P., Godt A., Zimmermann H., Banjam J., Timmel C. R., Hilger D., and Jung H. DeerAnalysis2006: a comprehensive software package for analyzing pulsed ELDOR data. *Appl. Magn. Reson.*, 30:473–498, 2006.

[19] Sen K. I., Logan T. M., and Fajer P. G. Protein dyanmics and monomer-monomer interactions in AntR activation by electron paramagnetic resonance and double electron-electron resonance. *Biochemistry*, 46:11639–11649, 2007.

[20] Brandon S., Beth A. H., and Hustedt E. J. The global analysis of DEER data. *J. Magn. Reson.*, 218:93–104, 2012.

[21] Stein R. A., Beth A. H., and Hustedt E. J. A straightforward approach to the analysis of Double Electron-Electron Resonance data. In Qin P. Z. and Warncke K., editors, *Electron Paramagnetic Resonance Investigations of Biological Systems by Using Spin Labels,*

Spin Probes, and Intrinsic Metal Ions, Part A, volume 563, chapter 20, pages 531–567. Elsevier, 2015.

[22] Blackburn M. E., Veloro A. M., and Fanucci G. E. Monitoring inhibitor-induced conformational population shifts in HIV-1 protease by pulsed EPR spectroscopy. *Biochemistry*, 48:8765–8767, 2009.

[23] Casey T. M. and Fanucci G. E. Spin labeling and Double Electron-Electron Resonance (DEER) to Deconstruct Conformational Ensembles of HIV Protease. *Method Enzymol.*, 564:153–187, 2015.

[24] Chiang Y.-W., Borbat P. P., and Freed J. H. Maximum entropy: A complement to Tikhonov regularization for determination of pair distance distributions by pulsed ESR. *J. Magn. Reson.*, 177:184–196, 2005.

[25] Dzuba S. A. The determination of pair-distance distribution by double electron– electron resonance: regularization by the length of distance discretization with Monte Carlo calculations. *J. Magn. Reson.*, 269:113–119, 2016.

[26] Srivastava M., Georgieva E. R., and Freed J. H. A New Wavelet Denoising Method for Experimental Time-Domain Signals: Pulse Dipolar Electron Spin Resonance. *J. Phys. Chem. A*, 121:2452–2465, 2017.

[27] Srivastava M. and Freed J. H. Singular Value Decomposition Method to determine Distance Distributions in Pulsed Dipolar Electron Spin Resonance. *J. Phys. Chem. Lett.*, 8:5648–5655, 2017.

[28] Matveeva A. G., Nekrasov V. M., and Maryasov A. G. Analytical solution of the PELDOR inverse problem using the integral Mellin transform. *Phys. Chem. Chem. Phys.*, 19:32381–32388, 2018.

[29] Kuprov I. Deep neural network processing of DEER data. The 50th Annual International Meeting of the Electron Spin Resonance Group of the Royal Society of Chemistry, 2017.

[30] Worswick S. G., Spencer J. A., Jeschke G., and Kuprov I. Deep neural network processing of DEER data. *Sci. Adv.*, 4, 2018.

[31] Jeschke G., Koch A., Jonas U., and Godt A. Direct Conversion of EPR Dipolar Time Evolution Data to Distance Distributions. *J. Magn. Reson.*, 155:72–82, 2002.

[32] Jeschke G. Distance Measurements in the Nanometer Range by Pulse EPR. *Chem. Phys. Chem.*, 3:927–932, 2002.

[33] Lopez C. J., Yang Z., Altenbach C., and Hubbell W. L. Conformational selection and adaptation to ligand binding in T4 Lysozyme cavity mutants. *Proc. Natl. Acad. Sci. USA*, 110:E4306–E4315, 2013.

[34] Galiano L., Ding F., Veloro A. M., Blackburn M. E., Simmerling C., and Fanucci G. E. Drug pressure selected mutations in HIV-1 protease alter flap conformations. *J. Am. Chem. Soc.*, 131:430–431, 2009.

[35] Blackburn M. E., Veloro A. M., and Fanucci G. E. Monitoring inhibitor-induced conformational population shifts in HIV-1 protease by pulsed EPR spectroscopy. *Biochemistry*, 48:8765–8767, 2009.

[36] de Vera I. M. S., Smith A. N., Dancel M. C. A., Huang X., Dunn B. M., and Fanucci G. E. Elucidating a relationship between conformational sampling and drug resistance in HIV-1 protease. *Biochemistry*, 52:3278–3288, 2013.

[37] Huang X., Britto M. D., Kear-Scott J. L., Boone C. D., Rocca J. R., Simmerling C., Mckenna R., Bieri M., Gooley P. R., Dunn B. M., and Fanucci G. E. The role of select subtype polymorphisms on HIV-1 protease conformational sampling and dynamics. *J. Biol. Chem.*, 289:17203–17214, 2014.

[38] Casey T. DEERconstruct. MATLAB Central File Exchange, 2015. Accessed and downloaded May 19, 2016.

[39] Sen K. I., Logan T. M., and Fajer P. G. Protein dynamics and monomer-monomer interactions in AntR activation by electron paramagnetic resonance and double electron-electron resonance. *Biochemistry*, 46:11639–11649, 2007.

[40] Brandon S., Beth A. H., and Hustedt E. J. The global analysis of DEER data. *J. Magn. Reson.*, 218:93–104, 2012.

[41] Martin R. E., Pannier M., Diederich F., Gramlich V., Hubrich M., and Spiess H. W. Determination of End-to-End Distances in a Series of TEMPO Diradicals of up to 2.8 nm Length with a New Four-Pulse Double Electron Electron Resonance Experiment. *Angew. Chem. Int. Ed.*, 37:2834–2837, 1998.

[42] Salikhhov K. M., Khairuzhdinov I. T., and Zaripov R. B. Three-Pulse ELDOR Theory Revisited. *Appl. Magn. Reson.*, 45:573–619, 2014.

[43] NIST Digital Library of Mathematical Functions. http://dlmf.nist.gov/, Release 1.0.10 of 2015-08-07. Online companion to [44].

[44] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, editors. *NIST Handbook of Mathematical Functions*. Cambridge University Press, New York, NY, 2010. Print companion to [43].

[45] Schäfer H., Mäder B., and Volke F. De-Pake-ing of NMR powder spectra by nonnegative least-squares analysis with Tikhonov regularization. *J. Magn. Reson. A*, 116:145–149, 1995.

[46] Vogt F. G., Aurentz D. J., and Mueller K. T. Determination of internuclear distances from solid-state nuclear magnetic resonance: Dipolar transforms and regularization methods. *Mol. Phys.*, 95:907–919, 1998.

[47] Schäfer H., Mädler B., and Sternin E. Determination of Orientational Order Parameters from ^{2}H NMR Spectra of Magnetically Partially Oriented Lipid Bilayers. *Biophys. J.*, 74:1007–1014, 1998.

[48] Winterhalter J., Maier D., Grabowski D. A., and Honerkamp J. Determination of orientational distributions from ^{2}H NMR data by a regularization method. *J. Chem. Phys.*, 110:4035–4046, 1999.

[49] Schäfer H. and Sternin E. Inverse ill-posed problems in experimental data analysis in physics. *Phys. Can.*, 53:77–85, 1997.

[50] Sternin E. Use of Inverse Theory Algorithms in the Analysis of Biomembrane NMR Data. *Method Mol. Bio.*, 400:103–125, 2007.

[51] Honerkamp J. and Weese J. Tikhonovs regularization method for ill-posed problems. *Continuum Mech. Thermodyn.*, 2:17–30, 1990.

[52] Weese J. A reliable and fast method for the solution of Fredholm integral equations of the first kind based on Tikhonov regularization. *Comput. Phys. Commun.*, 69:99–111, 1992.

[53] Kaipio J. and Somersalo E. *Statistical and Computational Inverse Problems*. Springer, 2005.

[54] Smith R. C. *Uncertainty Quantification: Theory, Implementation, and Applications*. SIAM Computational Science and Engineering, 2014.

[55] Gelman A., Carlin J. B., Stern H. S., Dunson D. B., Vehtari A., and Rubin D. B. *Bayesian Data Analysis*. Taylor & Francis Group LLC, 3rd edition, 2014.

[56] Zhou G. and Voelz V. A. Bayesian inference of conformational state populations from computational models and sparse experimental observables. *J. Comp. Chem.*, 35:2215–2224, 2014.

[57] Hines K. E. A primer on Bayesian inference for biophysical systems. *Biophys. J.*, 108:2103–2113, 2015.

[58] MacCallum J. L., Perez A., and Dill K. A. Determining protein structures by combining semireliable data with atomistic physical models by Bayesian inference. *Proc. Natl. Acad. Sci. USA*, 112:6985–6990, 2015.

[59] Kaipio J. and Somersalo E. *Statistical and Computational Inverse Problems*. Springer, 2005.

[60] Neal R. M. Probabilistic Inference Using Markov Chain Monte Carlo Methods. Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto, 1993.

[61] Mchaourab H. S, Kálai T., Hideg K., and Hubbell W. L. Motion of spin-labeled side chains in T4 lysozyme: Effect of side chain structure. *Biochemistry*, 38:2947–2955, 1999.

[62] Langen R., Oh K. J., Cascio D., and Hubbell W. L. Crystal structures of spin labeled T4 lysozyme mutants: Implications for the interpretation of EPR spectra in terms of structure. *Biochemistry*, 39:8396–8405, 2000.

[63] Columbus L., Kálai T., Jekö J., Hideg K., and Hubbell W. L. Molecular motion of spin labeled side chains in α-helices: Analysis by variation of side chain structure. *Biochemistry*, 40:3828–3846, 2001.

[64] Liang Z., Lou Y., Freed J. H., Columbus L., and Hubbell W. L. A multifrequency electron spin resonance study of T4 lysozyme dynamics using the slowly relaxing local structure model. *J. Phys. Chem. B*, 108:17649–17659, 2004.

[65] Jacobsen K., Oga S., Hubbell W. L., and Risse T. Determination of the orientation of T4 lysozyme vectorially bound to a planar-supported lipid bilayer using site-directed spin labeling. *Biophys. J.*, 88:4351–4365, 2005.

[66] Altenbach C., Froncisz W., Hemker R., Mchaourab H., and Hubbell W. L. Accessibility of nitroxide side chains: Absolute Heisenberg exchange rates from power saturation EPR. *Biophys. J.*, 89:2103–2112, 2005.

[67] Pyka J., Ilnicki J., Altenbach C., Hubbell W. L., and Froncisz W. Accessibility and dynamics of nitroxide side chains in T4 lysozyme measured by saturation recovery EPR. *Biophys. J.*, 89:2059–2068, 2005.

[68] Lopez C. J., Yang Z., Altenbach C., and Hubbell W. L. Conformational selection and adaptation to ligand binding in T4 lysozyme cavity mutants. *Proc. Natl. Acad. Sci. USA*, 110:E4306–E4315, 2013.

[69] Lerch M. T., López C. J., Yang Z., Kreitman M. J., Horwitz J., and Hubbell W. L. Structure-relaxation mechanism for the response of T4 lysozyme cavity mutants to hydrostatic pressure. *Proc. Natl. Acad. Sci. USA*, 112:E2437–E2446, 2015.

[70] Polyhach Y., Bordignon E., and Jeschke G. Rotamer libraries of spin labelled cysteines for protein studies. *Phys. Chem. Chem. Phys.*, 13:2356–2366, 2011.

[71] Polyhach Y. and Jeschke G. Prediction of favourable sites for spin labelling of proteins. *Spectrosc. Int. J.*, 24:651–659, 2010.

[72] Jeschke G. MMM: A toolbox for integrative structure modeling. *Prot. Sci.*, 27:76–85, 2018.

[73] Kaipio J. and Somersalo E. Statistical inverse problems: Discretization, model reduction and inverse crimes. *J. Comput. Appl. Math.*, 198:493–504, 2007.

[74] Bauer F. and Lukas M. A. Comparing parameter choice methods for regularization of ill-posed problems. *Math. Comput. Simulat.*, 81:1795–1841, 2011.

[75] Voelz V. A. and Pande V. S. Calculation of rate spectra from noisy time series data. *Proteins*, 80:342–351, 2012.

[76] Bardsley J. M. and Fox C. An MCMC method for uncertainty quantification in nonnegativity constrained inverse problems. *Inverse Probl. Sci. Eng.*, 20:477–498, 2012.

[77] Bro R. and de Jong S. A fast non-negativity-constrained least squares algorithm. *J. Chemometr.*, 11:393–401, 1997.

[78] Marsaglia G. and Tsang W. W. A simple method for generating gamma variables. *ACM Trans. Math. Softw.*, 26:363–372, 2000.

[79] France J. J., Gur Y., Kirby R. M., and Johnson C. R. A Bayesian approach to quantifying uncertainty in Tikhonov solutions for the inverse problem of electrocardiography. *Comp. Card.*, 41:529–532, 2014.

[80] Hyndman R. J. and Yanan F. Sample quantiles in statistical packages. *Am. Stat.*, 50:361–365, 1996.

[81] Puljung M. C., DeBerg H. A., Zagotta W. N., and Stoll S. Double electron-electron resonance reveals cAMP-induced conformational change in HCN channels. *Proc. Natl. Acad. Sci. USA*, 111:9816–9821, 2014.

[82] DeBerg H. A., Bankston J. R., Rosenbaum J. C., Brzovic P. S., Zagotta W. N., and Stoll S. Structural mechanism for the regulation of HCN ion channels by the accessory protein TRIP8b. *Structure*, 23:734–744, 2015.

[83] The MathWorks Inc. MATLAB Release 2016a, 2016.

[84] D'Errico J. FresnelS and FresnelC. MATLAB Central File Exchange, 2010. Accessed and downloaded November 25, 2013.

[85] Bar-Guy A. RANDRAW. MATLAB Central File Exchange, 2005. Accessed and downloaded May 25, 2015.

[86] Abragam A. *The Principles of Nuclear Magnetism*. International series of monographs on physics. Clarendon Press, 1961.

[87] Klauder J. R. and Anderson P. W. Spectral diffusion decay in spin resonance experiments. *Phys. Rev.*, 125:912–932, 1962.

[88] Milov A. D. and Tsvetkov Yu. D. Double electron-electron resonance in electron spin echo: Conformations of spin-labeled poly-4-vinylpyridine in glassy solutions. *Appl. Magn. Reson.*, 12:495–504, 1997.

[89] Kattnig D. R., Reichenwallner J., and Hinderberger D. Modeling Excluded Volume Effects for the Faithful Description of the Background Signal in Double Electron-Electron Resonance. *J. Phys. Chem. B*, 117:16542–16557, 2013.

[90] M. Betancourt. A Conceptual Introduction to Hamiltonian Monte Carlo. *ArXiv e-prints*, January 2017.

[91] Stan Development Team. MatlabStan: the MATLAB interface to Stan. `http://mc-stan.org`, 2017. Accessed: 2018-05-30.

[92] Carpenter B., Gelman A., Hoffman M. D., Lee D., Goodrich B., Betancourt M., Brubaker M., Guo J., Li P., and Riddell A. Stan: A probabilistic programming language. *J. Stat. Softw.*, 76, 2017.

[93] Cohen M. R., Frydman V., Milko P., Iron M. A., Abdelkader E. H, Lee M. D., Swarbrick J. D., Raitsimring A., Otting G., Graham B., Feintuch A., and Goldfarb D. Overcoming artificial broadening in Gd(3+)-Gd(3+) distance distributions arising from dipolar pseudo-secular terms in DEER experiments. *Phys. Chem. Chem. Phys.*, 18:12847–12859, 2016.

[94] Marko A., Denysenkov V., and Prisner T. F. Out-of-phase PELDOR. *Mol. Phys.*, 111:2834–2844, 2013.

[95] Hansen P. C. and O'Leary D. P. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comp.*, 14:1487–1503, 1993.

[96] Hansen P. C. The L-Curve and its use in the numerical treatment of inverse problems. In Johnston P., editor, *Advances in Computational Bioengineering: Computational Inverse Problems in Electrocardiology*, pages 119–142. WIT Press, 2000.

[97] Wahba G. *Spline Models for Observational Data*. SIAM, 2007.

[98] Hansen P. C. *Discrete Inverse Problems: Insight and Algorithms*. SIAM, 2010.

[99] Wahba G. Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.*, 14:651–667, 1977.

[100] Golub G. H., Heath M., and Wahba G. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21:215–223, 1979.

[101] Cummins D. J., Filloon T. G., and Nychka D. Confidence intervals for nonparametric curve estimates: toward more uniform pointwise coverage. *J. Am. Stat. Assoc.*, 96:233–246, 2001.

[102] Vio R., Ma P., Zhong W., Nagy J., Tenorio L., and Wamsteker W. Estimation of regularization parameters in multiple-image deblurring. *Astron. Astrophys.*, 423:1179–1186, 2004.

[103] Lukas M. A. Robust generalized cross-validation for choosing the regularization parameter. *Inverse Probl.*, 22:1883–1902, 2006.

[104] Robinson T. and Moyeed R. Making robust the cross-validatory choice of smoothing parameter in spline smoothing regression. *Stat. Theory Methods*, 18:523–539, 1989.

[105] Lukas M. A. Strong robust generalized cross-validation for choosing the regularization parameter. *Inverse Probl.*, 24, 2008.

[106] Tikhonov A. and Glasko V. Use of the regularization method in non-linear problems. *U.S.S.R. Comput. Math. Math. Phys.*, 5:93–107, 1965.

[107] Tikhonov A. and Arsenin V. *Solutions of Ill-Posed Problems*. Wiley, New York, 1977.

[108] Hofmann B. *Regularization of Applied Inverse and Ill-Posed Problems*. Teubner, Leipzig, 1986.

[109] Phillips D. L. A technique for the numerical solution of certain integral equations of the first kind. *J. Assoc. Comput. Mach.*, 9:84–97, 1962.

[110] Morozov V. A. On the solution of functional equations by the method of regularization. *Soviet Math. Dokl.*, 7:414–417, 1966.

[111] Morozov V. A. *Methods for Solving Incorrectly Posed Problems*. Springer-Verlag, New York, 1984.

[112] Raus T. An a posteriori choice of the regularization parameter in case of approximately given error bound of data. *Acta et Comment. Univ. Tartuensis*, 913:73–87, 1990.

[113] Raus T. About regularization parameter choice in case of approximately given error bounds of data. *Acta et Comment. Univ. Tartuensis*, 937:77–89, 1992.

[114] Hämarik U. and Raus T. On the choice of the regularization parameter in ill-posed problems with approximately given noise level of data. *J. Inverse Ill-Posed Probl.*, 14:251–266, 2006.

[115] Lepskij O. On a problem of adaptive estimation in gaussian white noise. *Theor. Probab. Appl.*, 35:454–466, 1990.

[116] Bauer F. Some considerations concerning regularization and parameter choice algorithms. *Inverse Probl.*, 23:837–858, 2007.

[117] Bauer F. and Mathé P. Parameter choice methods using minimization schemes. *J. Complexity*, 27:68–85, 2011.

[118] Wahba G. A comparison of gcv and gml for choosing the smoothing parameter in the generalized spline smoothing problem. *Ann. Stat.*, 13:1378–1402, 1985.

[119] Auchmuty G. A posteriori error estimates for linear equations. *Numer. Math.*, 61:1–6, 1992.

[120] Brezinski C., Rodriguez G., and Seatzu S. Error estimates for linear systems with applications to regularization. *Numer. Algorithms*, 49:85–104, 2008.

[121] Brezinski C., Rodriguez G., and Seatzu S. Error estimates for the regularization of least squares problems. *Numer. Algorithms*, 51:61–76, 2009.

[122] Hansen P. C., Kilmer M. E., and Kjeldsen R. H. Exploiting residual information in the parameter choice for discrete ill-posed problems. *BIT*, 46:41–59, 2006.

[123] Rust B. W. Parameter selection for constrained solutions to ill-posed problems. *Comput. Sci. Stat.*, 32:333–347, 2000.

[124] Rust B. W. and O'Leary D. P. Residual periodograms for choosing regularization parameters for ill-posed problems. *Inverse Probl.*, 24:034005, 2008.

[125] Mallows C. L. Some comments on C_P. *Technometrics*, 15:1–6, 1973.

[126] Burnham K. P. and Anderson D. R. *Model Selection and Multimodel Inference*. Springer, 2002.

[127] Zhang Y., Li R., and Tsai C.-L. Regularization parameter selections via generalized information criterion. *J. Am. Stat. Assoc.*, 105:312–323, 2010.

[128] Akaike H. A new look at the statistical model identification. *IEEE T. Automat. Contr.*, 19:716–723, 1974.

[129] Sugiura N. Further analysis of the data by Akaike's information criterion and the finite corrections. *Commun. Stat. Theor. M.*, A7:13–26, 1978.

[130] Hurvich C. M. and Tsai C.-L. Regression and time series model selection in small samples. *Biometrika*, 76:297–307, 1989.

[131] Schwarz G. Estimating the dimension of a model. *Ann. Stat.*, 6:461–464, 1978.

[132] Bozgodan H. Akaike's information criterion and recent developments in information complexity. *J. Math. Psychol.*, 44:62–91, 2000.

[133] Urmanov A. M., Gribok A. V., Bozdogan H., Hines J. W., and Uhrig R. E. Information complexity-based regularization parameter selection for solution of ill conditioned inverse problems. *Inverse Probl.*, 18:L1–L9, 2002.

[134] Lewis A. K., James Z. M., McCaffrey J. E., Braun A. R., Karim C. B., Thomas D. D., and Sachs J. N. Open and closed conformations of the isolated transmembrane domain of death receptor 5 support a new model of activation. *Biophys. J.*, 106:L21–L24, 2014.

[135] Moen R. J., Klein J. C., and Thomas D. D. Electron paramagnetic resonance resolves effects of oxidative stress on muscle proteins. *Exerc. Sci. Sport Rev.*, 42:30–36, 2014.

[136] McCarthy M. R., Thompson A. R., Nitu F., Moen R. J., Olenek M. J., Klein J. C., and Thomas D. D. Impact of methionine oxidation on calmodulin structural dynamics. *Biochem. Bioph. Res. Co.*, 456:567–572, 2015.

[137] Iyalomhe O., Herrick D. Z., Cafiso D. S., and Maloney P. C. Closure of the cytoplasmic gate formed by TM5 and TM11 during transport in the oxalate/formate exchanger from Oxalobacter formigenes. *Biochemistry*, 53:7735–7744, 2014.

[138] Cafiso D. S. Identifying and quantitating conformational exchange in membrane proteins using site-directed spin labeling. *Acc. Chem. Res.*, 47:3102–3109, 2014.

[139] Lu B., Kiessling V., Tamm L. K., and Cafiso D. S. The juxtamembrane linker of full-length Synaptotagmin 1 controls oligomerization and calcium-dependent membrane binding. *J. Biol. Chem.*, 289:22161–22171, 2014.

[140] Carter J. D., Gonzales E. G., Huang X., Smith A. N., de Vera I. M. S., D'Amore P. W., Rocca J. R., Goodenow M. M., Dunn B. M., and Fanucci G. E. Effects of PRE and POST therapy drug-pressure selected mutations on HIV-1 protease conformational sampling. *FEBS Lett.*, 588:22161–22171, 2014.

[141] Edwards D. T., Huber T., Hussain S., Stone K. M., Kinnebrew M., Kaminker I., Matalon E., Sherwin M. S., Goldfarb D., and Han S. Determining the oligomeric structure of Proteorhodopsin by Gd^{3+}-based pulsed dipolar spectroscopy of multiple distances. *Structure*, 22:1677–1686, 2014.

[142] Martorana A., Bellapadrona G., Feintuch A., Di Gregorio E., Aime S., and Goldfarb D. Probing protein conformation in cells by EPR distance measurements using Gd^{3+} spin labeling. *J. Am. Chem. Soc.*, 136:13458–13465, 2014.

[143] Goldfarb D. Gd^{3+} spin labeling for distance measurements by pulse EPR spectroscopy. *Phys. Chem. Chem. Phys.*, 16:9685–9699, 2014.

[144] Gophane D. B., Endeward B., Prisner T. F., and Sigurdsson S. Th. Conformationally restricted isoindoline-derived spin labels in duplex DNA: Distances and rotational flexibility by pulsed electron-electron double resonance spectroscopy. *Chem. Eur. J.*, 20:15913–15919, 2014.

[145] Lumme C., Altan-Martin H., Dastvan R., Sommer M. S., Oreb M., Schuetz D., Hellenkamp B., Mirus O., Kretschmer J., Lyubenova S., Kügel W., Medelnik J. P., Dehmer M., Michaelis J., Prisner T. F., Hugel T., and Schleiff E. Nucleotides and substrates trigger the dynamics of the Toc34 GTPase homodimer involved in chloroplast preprotein translocation. *Structure*, 22:526–538, 2014.

[146] Prisner T. F., Marko A., and Sigurdsson S. Th. Conformational dynamics of nucleic acid molecules studied by PELDOR spectroscopy with rigid spin labels. *J. Magn. Reson.*, 252:187–198, 2015.

[147] Akhmetzyanov D., Plackmeyer J., Endeward B., Denysenkova V., and Prisner T. F. Pulsed electron-electron double resonance spectroscopy between a high-spin Mn^{2+} ion and a nitroxide spin label. *Phys. Chem. Chem. Phys.*, 17:6760–6766, 2015.

[148] Bleicken S., Jeschke G., Stegmueller C., Salvador-Gallego R., Garcia-Saez A. J., and Bordignon E. Structural model of Active Bax at the membrane. *Mol. Cell*, 56:496–505, 2014.

[149] Joseph B., Korkhov V. M., Yulikov M., and Jeschke G. Bordignon E. Conformational cycle of the vitamin B_{12} ABC importer in liposomes detected by double electron-electron resonance (DEER). *J. Biol. Chem.*, 289:3176–3185, 2014.

[150] Duss O., Yulikov M., Jeschke G., and Allain F. H. T. EPR-aided approach for solution structure determination of large RNAs or protein-RNA complexes. *Nat. Commun.*, 5:3669, 2014.

[151] Qi M., Groß A., Jeschke G., Godt A., and Drescher M. Gd(III)-PyMTA label is suitable for in-cell EPR. *J. Am. Chem. Soc.*, 136:15366–15378, 2014.

[152] Duss O., Michel E., Yulikov M., Schubert M., Jeschke G., and Allain F. H. T. Structural basis of the non-coding RNA RsmZ acting as a protein sponge. *Nature*, 509:588–592, 2014.

[153] Raba M., Dunkel S., Hilger D., Lipiszko K., Polyhach Y., Jeschke G., Bracher S., Klare J. P., Quick M., Jung H., and Steinhoff H. J. Extracellular loop 4 of the proline transporter PutP controls the periplasmic entrance to ligand binding sites. *Structure*, 22:769–780, 2014.

[154] Razzaghi S., Qi M., Nalepa A. I., Godt A., Jeschke G., Savitsky A., and Yulikov M. RIDME spectroscopy with Gd(III) centers. *J. Phys. Chem. Lett.*, 5:3970–3975, 2014.

[155] Sahu I. D., Kroncke B. M., Zhang R., Dunagan M. M., Smith H. J., Craig A., McCarrick R. M., Sanders C. R., and Lorigan G. A. Structural investigation of the transmembrane domain of KCNE1 in proteoliposomes. *Biochemistry*, 53:6392–6401, 2014.

[156] Akyuz N., Georgieva E. R., Zhou Z., Stolzenberg S., Cuendet M. A., Khelashvili G., Altman R. B., Terry D. S., Freed J. H., Weinstein H., Boudker O., and Blanchard S. C. Transport domain unlocking sets the uptake rate of an aspartate transporter. *Nature*, 518:68–73, 2015.

[157] Georgieva E. R., Xiao S., Borbat P. P., Freed J. H., and Eliezer D. Tau binds to lipid membrane surfaces via short amphipathic helices located in its microtubule-binding repeats. *Biophys. J.*, 107:1441–1452, 2014.

[158] Sun Y., Borbat P. P., Grigoryants V. M., Myers W. K., Freed J. H., and Scholes C. P. Pulse dipolar ESR of doubly labeled mini TAR DNA and its annealing to mini TAR RNA. *Biophys. J.*, 108:893–902, 2015.

[159] Aluvila S., Mandal T., Hustedt E., Fajer P., Choe J. Y., and Oh K. J. Organization of the mitochondrial apoptotic BAK pore. *J. Biol. Chem.*, 289:2537–2551, 2014.

[160] Kazmier K., Sharma S., Islam S. M., Roux B., and Mchaourab H. S. Conformational cycle and ion-coupling mechanism of the Na+/hydantoin transporter Mhp1. *Proc. Natl. Acad. Sci. USA*, 111:14752–14757, 2014.

[161] Durr K. L., Chen L., Stein R. A., De Zorzi R., Folea I. M., Walz T., Mchaourab H. S., and Gouaux E. Structure and dynamics of AMPA receptor GluA2 in resting, pre-open, and desensitized states. *Cell*, 158:778–792, 2014.

[162] Mishra S., Verhalen B., Stein R. A., Wen P., Tajkhorshid E., and Mchaourab H. S. Conformational dynamics of the nucleotide binding domains and the power stroke of a heterodimeric ABC transporter. *eLife*, 3:e02740, 2014.

[163] Masureel M., Martens C., Stein R., Mishra S., Ruysschaert J., Mchaourab H., and Govaerts C. Protonation drives the conformational switch in the multidrug transporter LmrP. *Nat. Chem. Biol.*, 10:149–155, 2014.

[164] Meyer V., Dinkel P. D., Luo Y., Yu X., Wei G., Zheng J., Eaton G. R., Ma B., Nussinov R., Eaton S. S., and Margittai M. Single mutations in Tau modulate the populations of fibril conformers through seed selection. *Angew. Chem. Int. Ed.*, 53:1590–1593, 2014.

[165] Eps N. V., Thomas C. J., Hubbell W. L., and Sprang S. R. The guanine nucleotide exchange factor Ric-8A induces domain separation and Ras domain plasticity in Gai1. *Proc. Natl. Acad. Sci. USA*, 112:1404–1409, 2015.

[166] Vendome J., Felsovalyi K., Song H., Yang Z., Jin X., Brasch J., Harrison O. J., Ahlsen G., Bahna F., Kaczynska A., Katsamba P. S., Edmondf D., Hubbell W. L., Shapiro L., and Honig B. Structural and energetic determinants of adhesive binding specificity in type I cadherins. *Proc. Natl. Acad. Sci. USA*, pages E4175–E4184, 2014.

[167] Lerch M. T., Yang Z., Brooks E. K., and Hubbell W. L. Mapping protein conformational heterogeneity under pressure with site-directed spin labeling and double electron-electron resonance. *Proc. Natl. Acad. Sci. USA*, 111:E1201–E1210, 2014.

[168] Yang Z., Jimenez-Oses G., Lopez C. J., Bridges M. D., Houk K. N., and Hubbell W. L. Long-range distance measurements in proteins at physiological temperatures using saturation recovery EPR spectroscopy. *J. Am. Chem. Soc.*, 136:15356–15365, 2014.

[169] Lopez C. J., Fleissner M. R., Brooks E. K., and Hubbell W. L. Stationary-phase EPR for exploring protein structure, conformation, and dynamics in spin-labeled proteins. *Biochemistry*, 53:7067–7075, 2014.

[170] Schredelseker J., Paz A., Lopez C. J., Altenbach C., Leung C. S., Drexler M. K., Chen J., Hubbell W. L., and Abramson J. High resolution structure and double electron-electron resonance of the zebrafish voltage-dependent anion channel 2 reveal an oligomeric population. *J. Biol. Chem.*, 289:12566–12577, 2014.

[171] Merz G. E., Borbat P. P., Pratt A. J., Getzoff E. D., Freed J. H., and Crane B. R. Copper-based pulsed dipolar ESR spectroscopy as a probe of protein conformation linked to disease states. *Biophys. J.*, 107:1669–1674, 2014.

[172] Pratta A. J., Shina D. S., Merz G. E., Rambod R. P., Lancaster W. A., Dyer K. N., Borbat P. P., Poole F. L., Adams M. W. W., Freed J. H., Crane B. R., Tainer J. A., and Getzoff E. D. Aggregation propensities of superoxide dismutase G93 hotspot mutants mirror ALS clinical phenotypes. *Proc. Natl. Acad. Sci. USA*, 111:E4568–E4576, 2014.

[173] Cunningham T. F., Shannon M. D., Putterman M. R., Arachchige R. J., Sengupta I., Gao M., Jaroniec C. P., and Saxena S. Cysteine-specific Cu^{2+} chelating tags used as paramagnetic probes in double electron electron resonance. *J. Phys. Chem. B.*, 119:2839–2843, 2015.

[174] Tavenor N. A., Silva K. I., Saxena S., and Horne W. S. Origins of structural flexibility in protein-based supramolecular polymers revealed by DEER spectroscopy. *J. Phys. Chem. B.*, 118:9881–9889, 2014.

[175] Cunningham T. F., Putterman M. R., Desai A., Horne W. S., and Saxena S. The double-histidine Cu^{2+}-binding motif: A highly rigid, site-specific spin probe for electron spin resonance distance measurements. *Angew. Chem. Int. Ed.*, 54, 2015.

[176] Ji M., Ruthstein S., and Saxena S. Paramagnetic metal ions in pulsed ESR distance distribution measurements. *Acc. Chem. Res.*, 47:688–695, 2014.

[177] Aitha M., Richmond T. K., Hu Z., Hetrick A., Reese R., Gunther A., McCarrick R., Bennett B., and Crowder M. W. Dilution of dipolar interactions in a spin-labeled, multimeric metalloenzyme for DEER studies. *J. Inorg. Biochem.*, 136:40–46, 2014.

[178] Dunkel S., Pulagam L. P., Steinhoff H. J., and Klare J. P. In vivo EPR on spin labeled colicin a reveals an oligomeric assembly of the pore-forming domain in E. coli membranes. *Phys. Chem. Chem. Phys.*, 17:4875–4878, 2015.

[179] Gruian C., Boehme S., Simon S., Steinhoff H. J., and Klare J. P. Assembly and function of the tRNA-modifying GTPase MnmE adsorbed to surface functionalized bioactive glass. *ACS Appl. Mater. Inter.*, 6:7615–7625, 2014.

[180] Vöpel T., Hengstenberg C. S., Peulen T. O., Ajaj Y., Seidel C. A. M., Herrmann C., and Klare J. P. Triphosphate induced dimerization of human guanylate binding protein 1 involves association of the C-terminal helices: A joint double electron-electron resonance and FRET study. *Biochemistry*, 53:4590–4600, 2014.

[181] Zhuo Y., Vishnivetskiy S. A., Zhan X., Gurevich V. V., and Klug C. S. Identification of receptor binding-induced conformational changes in non-visual arrestins. *J. Biol. Chem.*, 289:2091–2102, 2014.

[182] Shevelev G. Y., Krumkacheva O. A., Lomzov A. A., Kuzhelev A. A., Rogozhnikova O. Y., Trukhin D. V., Troitskaya T. I., Tormyshev V. M., Fedin M. V., Pyshnyi D. V., and Bagryanskaya E. G. Physiological-temperature distance measurement in nucleic acid using triarylmethyl-based spin labels and pulsed dipolar EPR spectroscopy. *J. Am. Chem. Soc.*, 136:9874–9877, 2014.

[183] Valentin M. D., Albertini M., Zurlo E., Gobbo M., and Carbonera D. Porphyrin triplet state as a potential spin label for nanometer distance measurements by PELDOR spectroscopy. *J. Am. Chem. Soc.*, 136:6582–6585, 2014.

[184] Kerry P. S., Turkington H. L., Ackermann K., Jameison S. A., and Bode B. E. Analysis of Influenza A virus NS1 dimer interfaces in solution by pulse EPR distance measurements. *J. Phys. Chem. B.*, 118:10882–10888, 2014.

[185] Milov A. D., Samoilova R. I., Tsvetkov Y. D., Peggion C., Formaggio F., and Toniolo C. Peptides on the surface. PELDOR data for spin-labeled alamethicin F50/5 analogues on organic sorbent. *J. Phys. Chem. B.*, 118:7085–7090, 2014.

[186] Bowman A., Hammond C. M., Stirling A., Ward R., Shang W., El-Mkami H., Robinson D. A., Svergun D. I., Norman D. G., and Owen-Hughes T. The histone chaperones Vps75 and Nap1 form ring-like, tetrameric structures in solution. *Nuc. Ac. Res.*, 42:6038–6051, 2014.

[187] Ward R., Pliotas C., Branigan E., Hacker C., Rasmussen A., Hagelueken G., Booth I. R., Miller S., Lucocq J., Naismith J. H., and Schiemann O. Probing the structure of the mechanosensitive channel of small conductance in lipid bilayers with pulsed electron-electron double resonance. *Biophys. J.*, 106:834–842, 2014.

[188] Salvadori E., Fung M. W., Hoffmann M., Anderson H. L., and Kay C. W. M. Exploiting the symmetry of the resonator mode to enhance PELDOR sensitivity. *Appl. Magn. Reson.*, 46:359–368, 2015.

[189] Shabestari M. H., Kumar P., Segers-Nolten I. M. J., Claessens M. M. A. E., van Rooijen B. D., Subramaniam V., and Huber M. Three long-range distance constraints and an approach towards a model for the α-Synuclein-fibril fold. *Appl. Magn. Reson.*, 46:369–388, 2015.

[190] Islam S. M. and Roux B. Simulating the distance distribution between spin-labels attached to proteins. *J. Phys. Chem. B.*, 119:3901–3911, 2015.

[191] Meyer V., Swanson M. A., Clouston L. J., Boratynski P. J., Stein R. A., Mchaourab H. S., Rajca A., Eaton S. S., and Eaton G. R. Room-temperature distance measurements of immobilized spin-labeled protein by DEER/PELDOR. *Biophys. J.*, 108:1213–1219, 2015.

[192] Lemmin T., Dimitrov M., Fraering P. C., and Peraro M. D. Perturbations of the straight transmembrane α-helical structure of the amyloid precursor protein affect its processing by γ-secretase. *J. Biol. Chem.*, 289:6763–6774, 2014.

[193] Shenberger Y., Shimshi A., and Ruthstein S. EPR spectroscopy shows that the blood carrier protein, human serum albumin, closely interacts with the N-terminal domain of the copper transporter, Ctr1. *J. Phys. Chem. B.*, 119:4824–4830, 2015.

[194] Sippach M., Weidlich D., Klose D., Abé C., Klare J., Schneider E., and Steinhoff H. J. Conformational changes of the histidine ATP-binding cassette transporter studied by double electron-electron resonance spectroscopy. *Biochim. Biophys. Acta.*, 1838:1760–1768, 2014.

[195] Abdullin D., Florin N., Hagelueken G., and Schiemann O. EPR-based approach for the localization of paramagnetic metal ions in biomolecules. *Angew. Chem. Int. Ed.*, 54:1827–1831, 2015.

[196] Alonso-Garcia N., Garcia-Rubio I., Manso J. A., Buey R. M., Urien H., Sonnenberg A., Jeschke G., and de Pereda J. M. Combination of X-ray crystallography, SAXS, and DEER to obtain the structure of the FnIII-3,4 domains of integrin a6b4. *Acta Cryst.*, D71:969–985, 2015.

[197] Fontes A., Karimi S., Helm L., Yulikov M., Ferreira P. M., and André J. P. Dinuclear DOTA-based GdIII chelates: Revisiting a straightforward strategy for relaxivity improvement. *Eur. J. Inorg. Chem.*, 9:1579–1591, 2015.

[198] Fowler P. W., Orwick-Rydmark M., Radestock S., Solcan N., Dijkman P. M., Lyons J. A., Kwok J., Caffrey M., Watts A., Forrest L. R., and Newstead S. Gating topology of the proton-coupled oligopeptide symporters. *Structure*, 23:290–301, 2015.

[199] Hohl M., Hürlimann L. M., Böhm S., Schöppe J., Grütter M. G., Bordignon E., and Seeger M. A. Structural basis for allosteric cross-talk between the asymmetric nucleotide binding sites of a heterodimeric ABC exporter. *Proc. Natl. Acad. Sci. USA*, 111:11025–11030, 2014.

[200] Babaylova E. S., Ivanov A. V., Malygin A. A., Vorobjeva M. A., Venyaminova A. G., Polienko Y. F., Kirilyuk I. A., Krumkacheva O. A., Fedin M. V., Karpova G. G., and Bagryanskaya E. G. A versatile approach for site-directed spin labeling and structural EPR studies of RNAs. *Org. Biomol. Chem.*, 12:3129–3136, 2014.

[201] Zhang X., Machado A. C. D., Ding Y., Chen Y., Lu Y., Duan Y., Tham K. W., Chen L., Rohs R., and Qin P. Z. Conformations of p53 response elements in solution deduced using site-directed spin labeling and Monte Carlo sampling. *Nuc. Ac. Res.*, 42:2789–2797, 2014.

[202] Schmidt M. J., Borbas J., Drescher M., and Summerer D. A genetically encoded spin label for electron paramagnetic resonance distance measurements. *J. Am. Chem. Soc.*, 136:1238–1241, 2014.

[203] Ambroso M. R., Hegde B. G., and Langen R. Endophilin A1 induces different membrane shapes using a conformational switch that is regulated by phosphorylation. *Proc. Natl. Acad. Sci. USA*, 111:6982–6987, 2014.

[204] Marrale M., Longo A., Barbon A., Brustolon M., and Brai M. Radical distributions in ammonium tartrate single crystals exposed to photon and neutron beams. *Radiat. Prot. Dosim.*, 161:398–402, 2014.

[205] Robotta M., Gerding H. R., Vogel A., Hauser K., Schildknecht S., Karreman C., Leist M., Subramaniam V., and Drescher M. Alpha-Synuclein binds to the inner membrane of mitochondria in an α-helical conformation. *Chem. Bio. Chem.*, 15:2499–2502, 2014.

[206] Shenberger Y., Gottlieb H. E., and Ruthstein S. EPR and NMR spectroscopies provide input on the coordination of Cu(I) and Ag(I) to a disordered methionine segment. *J. Biol. Inorg. Chem.*, 20:719–727, 2015.

[207] Aitha M., Moller A. J., Sahu I. D., Horitani M., Tierney D. L., and Crowder M. W. Investigating the position of the hairpin loop in New Delhi metallo-β-lactamase, NDM-1, during catalysis and inhibitor binding. *J. Inorg. Biochem.*, 156:35–39, 2016.

[208] Joseph B., Sikora A., and Cafiso D. S. Ligand induced conformational changes of a membrane transporter in E. coli cells observed with DEER/PELDOR. *J. Am. Chem. Soc.*, 138:1844–1847, 2016.

[209] Glöz J. P., Bockelmann S., Mayer K., Steinhoff H., Wieczorek H., Huss M., Klare J. P., and Menche D. EPR studies of V-ATPase with spin-labeled inhibitors DCC and Archazolid: Interaction dynamics with proton translocating subunit c. *ChemMedChem*, 11:420–428, 2016.

[210] Dastvan R., Fischer A. W., Mishra S., Meiler J., and Mchaourab H. S. Protonation-dependent conformational dynamics of the multidrug transporter EmrE. *Proc. Natl. Acad. Sci. USA*, 113:1120–1225, 2016.

[211] Grytz C. M., Marko A., Cekan P., Sigurdsson S. Th., and Prisner T. F. Flexibility and conformation of the cocaine aptamer studied by PELDOR. *Phys. Chem. Chem. Phys.*, 18:2993–3002, 2016.

[212] DeBerg H. A., Brzovic P. S., Flynn G. E., Zagotta W. N., and Stoll S. Structure and energetics of allosteric regulation of HCN2 ion channels by cyclic nucleotides. *J. Biol. Chem.*, 291:371–381, 2016.

[213] Abdelkader E. H., Lee H. D., Feintuch A., Cohen M. R., Swarbrick J. D., Otting G., Graham B., and Goldfarb D. A new Gd^{3+} spin label for Gd^{3+}-Gd^{3+} distance measurements in proteins produces narrow distance distributions. *J. Phys. Chem. Lett.*, 6:5016–5021, 2015.

[214] Kumar P., Segers-Nolten I. M. J., Schilderink N., Subramaniam V., and Huber M. Parkinson's protein α-Synuclein binds efficiently and with a novel conformation to two natural membrane mimics. *PLOS ONE*, 10:1–11, 2015.

[215] van Son M., Lindhoud S., van der Wild M., van Mierlo C. P. M., and Huber M. Double electron-electron spin resonance tracks flavodoxin folding. *J. Phys. Chem. B*, 119:13507–13514, 2015.

[216] Abdullin D., Duthie F., Meyer A., Müller E. S., Hagelueken G., and Schiemann O. Comparison of PELDOR and RIDME for distance measurements between nitroxides and low-spin Fe(III) ions. *J. Phys. Chem. B*, 119:13534–13542, 2015.

[217] Shevelev G. Yu., Krumkacheva O. A., Lomzov A. A., Kuzhelev A. A., Trukhin D. V., Rogozhnikova O. Yu., Tormyshev V. M., Pyshnyi D. V., Fedin M. V., and Bagryanskaya E. G. Triarylmethyl Labels: Toward Improving the Accuracy of EPR Nanoscale Distance Measurements in DNAs. *J. Phys. Chem. B*, 119:13641–13648, 2015.

[218] Manukovsky N., Frydman V., and Goldfarb D. Gd^{3+} spin labels report the conformation and solvent accessibility of solution and vesicle-bound Melittin. *J. Phys. Chem. B*, 119:13732–13741, 2015.

[219] Fehr N., Dietz C., Polyhach Y., von Hagens T., Jeschke G., and Paulsen H. Modeling of the N-terminal section and the lumenal loop of trimeric light harvesting complex II (LHCII) by using EPR. *J. Biol. Chem.*, 290:26007–26020, 2015.

[220] Doll A., Qi M., Wili N., Pribitzer S., Godt A., and Jeschke G. Gd(III)-Gd(III) distance measurements with chirp pump pulses. *J. Magn. Reson.*, 259:153–162, 2015.

[221] Vicente E. F., Sahu I. D., Costa-Filho A. J., Cilli E. M., and Lorigan G. A. Conformational changes of the *Hs*DHODH N-terminal microdomain via DEER spectroscopy. *J. Phys. Chem. B*, 119:8693–8697, 2015.

[222] Shenberger Y., Gottlieb H. E., and Ruthstein S. EPR and NMR spectroscopies provide input on the coordination of Cu(I) and Ag(I) to a disordered methionine segment. *J. Biol. Chem.*, 20:719–727, 2015.

[223] Abdullin D., Hagelueken G., Hunter R. I., Smith G. M., and Schiemann O. Geometric model-based fitting algorithm for orientation-selective PELDOR data. *Mol. Phys.*, 113:544–560, 2014.

[224] Rudi K., Ho F. Y., Gilsbach B. K., Pots H., Wittinghofer A., Kortholt A., and Klare J. P. Conformational heterogeneity of the Roc domains in C. *tepidum* Roc-COR and implications for human LRRK2 parkinson mutations. *Biosci. Rep.*, 35:e00254, 2015.

[225] Martorana A., Yang Y., Zhao Y., Li Q.-F., Su X.-C., and Goldfarb D. Mn(II) tags for DEER distance measurements in proteins via C-S attachment. *Dalton Trans.*, 44:20812–20816, 2015.

[226] Hintze C., Schütze F., Drescher M., and Mecking S. Probing of chain conformations in conjugated polymer nanoparticles by electron spin resonance spectroscopy. *Phys. Chem. Chem. Phys.*, 17:32289–32296, 2015.

[227] Abdelkader E. H., Feintuch A., Yao X., Adams L. A., Aurelio L., Graham B., Goldfarb D., and Otting G. Protein conformation by EPR spectroscopy using gadolinium tags clicked to genetically encoded p-azido-L-phenylalanine. *Chem. Commun.*, 51:15898–15901, 2015.

[228] Akhmetzyanov D., Schöps P., Marko A., Kunjir N. C., Sigurdsson S. Th., and Prisner T. F. Pulsed EPR dipolar spectroscopy at Q- and G-band on a trityl biradical. *Phys. Chem. Chem. Phys.*, 17:24446–24451, 2015.

[229] Ching H. Y. V., Demay-Drouhard P., Bertrand H. C., Policar C., Tabares L. C., and Un S. Nanometric distance measurements between Mn(II) DOTA centers. *Phys. Chem. Chem. Phys.*, 17:23368–23377, 2015.

[230] Dalaloyan A., Qi M., Ruthstein S., Vega S., Godt A., Feintuch A., and Goldfarb D. Gd(III)-Gd(III) EPR distance measurements - the range of accessible distances and the impact of zero field splitting. *Phys. Chem. Chem. Phys.*, 17:18464–18476, 2015.

[231] Kaminker I., Bye M., Mendelman N., Gislason K., Sigurdsson S. Th., and Goldfarb D. Distance measurements between manganese(II) and nitroxide spin-labels by DEER determine a binding site of Mn^{2+} in the HP92 loop of ribosomal RNA. *Phys. Chem. Chem. Phys.*, 17:15098–15102, 2015.

[232] Ackermann K., Giannoulis A., Cordes D. B., Slawin A. M. Z., and Bode B. E. Assessing dimerisation degree and cooperativity in a biomimetic small-molecule model by pulsed EPR. *Chem. Commun.*, 51:5257–5260, 2015.

[233] Guo R., Gaffney K., Yang Z., Kim M., Sungsuwan S., Huang X., Hubbell W. L., and Hong H. Steric trapping reveals a cooperativity network in the intramembrane protease GlpG. *Nat. Chem. Biol.*, 12:353–360, 2016.

[234] Syryamina V. N., Samoilova R. I., Tsvetkov Y. D., Ischenko A. V., De Zotti M., Gobbo M., Toniolo C., Formaggio F., and Dzuba S. A. Peptides on the surface: Spin-label EPR and PELDOR study of adsorption of the antimicrobial peptides trichogin GA IV and Ampullosporin A on the silica nanoparticles. *Appl. Magn. Reson.*, 47:309–320, 2016.

[235] Press W. H., Teukolsky S. A., Vetterling W. T., and Flannery B. P. *Numerical Recipes: The Art of Scientific Computing.* Cambridge, 2007.

[236] Bardsley J. M. An Efficient MCMC Method for Uncertainty Quantification in Inverse Problems. 2010.