Determination of Interspin Distance Distributions by cw-ESR Is a Single Linear Inverse Problem

Yun-Wei Chiang,* Tong-Yuan Zheng, Chiao-Jung Kao, and Jia-Cherng Horng
Department of Chemistry, National Tsing Hua University, Hsinchu, Taiwan

ABSTRACT Cw-ESR distance measurement method is extremely valuable for studying the dynamics-function relationship of biomolecules. However, extracting distance distributions from experiments has been a highly technique-demanding procedure. It has never been conclusively identified, to our knowledge, that the problems involved in the analysis are ill posed and are best solved using Tikhonov regularization. We treat the problems from a novel point of view. First of all, we identify the equations involved and uncover that they are actually two linear first-kind Fredholm integral equations. They can be combined into one single linear inverse problem and solved in a Tikhonov regularization procedure. The improvement with our new treatment is significant. Our approach is a direct and reliable mathematical method capable of providing an unambiguous solution to the ill-posed problem. It need not perform nonlinear least-squares fitting to infer a solution from noise-contaminated data and, accordingly, substantially reduces the computation time and the difficulty of analysis. Numerical tests and experimental data of polyproline II peptides with variant spin-labeled sites are provided to demonstrate our approach. The high resolution of the distance distributions obtainable with our new approach enables a detailed insight into the flexibility of dynamic structure and the identification of conformational species in solution state.

INTRODUCTION

In recent decades, electron spin resonance (ESR) spectroscopy with the site-directed spin-labeling (SDSL) technique has been established as a powerful tool that provides new and valuable insight into the dynamics, function, and structure of biomolecules (1–3). The key features of SDSL-ESR for structural biophysics include the following: 1), the nitroxide spin probe is very sensitive to the local environment where it is attached and, therefore, is a useful reporter to probe the heterogeneity in biological system on a molecular level; 2), the timescale of spin-label ESR ranges from micro- to nanoseconds, covering a variety of the important biological molecular mechanisms such as the dynamics of protein in solution; 3), the spectral lineshape displays great sensitivity to the local environment of the spin label, allowing the extraction of a lot of dynamic information from experimental spectra through rigorous theoretical analysis based on the stochastic Liouville equation (4); and 4), SDSL-ESR provides long distance measurement in paramagnetic spin pairs ranging from 0.8 to ~7 nm (2), allowing the studies of structure and dynamics in soluble and membrane-bound proteins of arbitrary molecular weight.

Based on these features, SDSL-ESR is generally considered as a great complement to the classical techniques, e.g., nuclear magnetic resonance and x-ray crystallography, in structural biophysics. Among the variety of ESR techniques, distance measurement experiments that provide a spectroscopic ruler in nanometer range have brought the most substantial impact to structural biology. There are two types of ESR experiments for measuring distances in spin label pairs: continuous-wave (cw) ESR and pulse ESR. The measurable interspin distance ranges are approximately from 0.8 to 2 nm in cw-ESR experiments and 2–7 nm in pulse ESR experiments, making ESR a valuable tool for studying the structural conformations of a protein and the arrangement of protein complexes. It has been proved an essential technique that provides unique insight into dynamics and biological functions of a large protein complex (5–8).

Compared to pulse ESR, which enables long-distance constraints on protein complex, cw-ESR has several advantages (9–12):

1. The cw-ESR spectrometer is more generally accessible and less expensive.
2. The cw experimental measurement is much less technically demanding.
3. The measurement can be performed in solution state at room or physiological temperature, avoiding undesirable effects due to variations in physical state at low temperature, such as from changes in ligand and metal coordination or in the temperature-dependent conformations (13).

In this study, we report the recent theoretical development in cw-ESR distance measurement. The general approach for cw distance measurements is briefly described below. Two spin labels are introduced either on the same monomer proteins or on different subunits in a complex assembly of proteins. Cw-ESR experiments are performed to obtain two spectra—i.e., spectra with, versus that without, the dipolar coupling (which could typically, although not necessarily, be the spectra of doubly- versus singly-labeled molecules). Hereafter, they are denoted as interacting $D(\mathbf{B})$ versus noninteracting $S(\mathbf{B})$ spectrum. It has been generally demonstrated

Submitted March 16, 2009, and accepted for publication May 28, 2009.

*Correspondence: ywchiang@mx.nthu.edu.tw

Editor: Eduardo Perozo.

0006-3495/09/08/0930/7 $2.00

doi: 10.1016/j.bpj.2009.05.030

that the two spectra can be approximately connected by a convolution equation, $D(\mathbf{B}) = S(\mathbf{B}) \otimes M(\mathbf{B})$, where $M(\mathbf{B})$ represents the broadening function due to the dipolar coupling between spins (14). Based on that principle, at least three methods have been developed to obtain the distance information in spin-label pairs. In the method of Rabenstein and Shin (14), the deconvolution problem is solved by Fourier-transforming the equation into the frequency domain. This method, therefore, requires extreme care and experience in determining the cutoff point to avoid the unwanted substantial noise due to division by zeroes or values close to zero. This often suffers from the drawback that one would have to sacrifice and discard the data in the high frequency domain, which corresponds to the information in long distance, to avoid deconvolution instability. In the convolution method of Steinhoff et al. (15), the interspin distance distributions are presumed to be a Gaussian, and are determined by fitting simulated cw-ESR powder spectra to experimental data. In the method of Altenbach and coworkers (13), they avoid deconvolution instability by using a combination of convolution and deconvolution procedures fitting a normalized sum of up to 50 weighted Pake functions, which represent different interspin distances. The highest resolution in the distribution function of interspins is, therefore, limited to the maximum number of the weighted Pake functions. The greater the number of the weighted Pakes used, the more ill-posed the problem becomes.

Unfortunately, the realization that the whole procedure involving the deconvolution of spectra and the determination of distance distributions in cw-ESR distance measurement is, fundamentally, an ill-posed problem has never been conclusively identified and well addressed. Without clear identification and appropriate treatment for these problems, the application of cw-ESR to biology would continue to suffer from time-consuming and complicated technique-demanding data analysis. An easier, unbiased, and more convenient treatment for extracting high-resolution probability distance distributions from cw-ESR experiments has been desired for a long time.

In this report, we treat the problems, which include the deconvolution of cw spectra and the determination of distance distribution of spin pairs, from what is, to our knowledge, a novel point of view. We first point out that they can be, in a more rigorous manner, rearranged to become two Fredholm integral equations of the first kind. The problems can be transformed into one single linear inverse problem requiring only one step of inversion of the problem to obtain the solution. A Fredholm integral equation of the first kind (the left-hand side of which, in discrete form, typically consists of experimental measured quantities; see next section) is a classical example of a linear ill-posed problem requiring a treatment of regularization methods. Note that an ill-posed problem is defined that an arbitrary small perturbation of the data (i.e., the noise in the measured quantities) can give rise to arbitrarily large variations in the solution; in other words, the solution to an ill-posed problem is not unique, since noise always exists in the experimental data. The condition of the illness of the problem depends on the kernel property of the Fredholm equation. It has been recently demonstrated that the equation connecting the distance distribution of spin pairs and the pulse ESR data, in particular the data obtained in double quantum coherence and double electron-electron resonance experiments, is ill-posed and is best solved using the Tikhonov regularization (TIKR) method (16–18). Note that the kernel for the pulse ESR resembles that for cw-ESR, and the condition of illness of the kernel is preserved as well. We apply the TIKR method to solving the problem of extracting interspin distance information in spin pairs from cw-ESR experiments. It is to be demonstrated that the analysis of cw distance measurement, as treated from a novel point of view in this report, cannot get any easier or more straightforward.

We provide some cases of the theoretical and experimental studies for the demonstration of our new approach. A series of model studies are provided to show that the improvements due to our new approach are significant. The deconvolution problem involved in the determination of dipolar broadening function can be transformed into a linear inverse problem to avoid the deconvolution instability. The probability distance distributions of spin pairs that possess high resolution can be unambiguously obtained from the dipolar broadening functions by TIKR method. In the experimental study, our approach is applied to study the dynamic secondary structure of polyproline II (PPII) peptides, which are doubly labeled at variant sites. PPII is a left-handed helix with all-*trans* peptide bonds and has backbone dihedral angles of $(\phi, \psi, \omega) = (-75^\circ, 145^\circ, 180^\circ)$. Formed at the absence of intrachain hydrogen bonds, PPII is quite extended with a helical pitch of 0.93 nm/turn and 3.0 residues/turn, i.e., an axial translation of 0.31 nm per amino acid. PPII helices are common in both folded proteins (19,20) and unfolded polypeptides (21). In the recent years, PPII helices have been demonstrated to be essential to a variety of biological activities such as signal transduction and immune response (22). It has been also found that the flexibility of PPII structure in many proteins and protein complexes plays an important role in the molecular recognition processes (23). The cw-ESR experiments demonstrated in this report would suggest a useful approach to investigate the flexibility and elasticity of PPII conformation in proteins.

MATERIALS AND EXPERIMENTAL METHODS

Peptide synthesis, labeling, and purification

All peptides were synthesized on a 50-μmol or 100-μmol scale by solid-phase methods using FMOC-protected amino acids, HBTU-mediated coupling, and standard reaction cycles on a PS3 automated peptide synthesizer (Protein Technologies, Tucson, AZ). Use of a Rink amide MBHA resin generated an amidated C-terminus after cleavage from the resin with 92.5% trifluoroacetic acid/2.5% triisopropylsilane/2.5% H_2O/2.5% ethyldithiol. Each peptide has an acetylated N-terminus. Peptides were purified by reverse phase

HPLC with a Thermo Biobasic C_{18} semipreparative column (Thermo Electron, Waltham, MA). H_2O/acetonitrile gradients with 0.1% (v/v) trifluoroacetic acid as the counterion were used for the purification of all peptides. Purified peptides were labeled with a 10-fold excess of (1-Oxy-2,2,5,5-tetramethyl-3-pyrroline-3-methyl) methanethiosulfonate spin label (Alexis Biochemicals, San Diego, CA) per Cys residue for overnight in 20 mM MOPS (pH 6.9) and in the dark at 4°C, and then repurified as described above. All peptides and labeled peptides are >90% pure according to HPLC analysis. The identities of all peptides were confirmed by using ESI-mass spectrometry. The concentrations of labeled peptides used for ESR experiments were ~0.4–0.6 mM in a 50% glycerol solution and 0.08–0.12 mM in a 40% sucrose solution. Two experimental temperatures, 250 K and 275 K, were used. For the spectra collected at higher temperatures (e.g., at room or physiological temperature), they display typical fast-motional lineshapes as the 11-amino-acid-long PPII is a rather small molecule. The linewidth due to the dipolar interaction is only slightly revealed in the fast-motional lineshape.

CW-ESR measurement

Samples were loaded in quartz capillaries. ESR spectra were recorded on a model No. EMX-10 spectrometer (Bruker, Billerica, MA) with an operating frequency of 9.4 GHz and the incident microwave power of 2 mW. The swept magnetic range was 200 Gauss.

Theory and mathematical method

In a discrete form, the problems that include the deconvolution of two cw-ESR spectra and the determination of the distance distribution function of spin pairs can be written in one single system of linear equations, $D(\mathbf{B}) = A(\mathbf{B},\mathbf{r})P(\mathbf{r})$, whose left-hand side is a column vector representing an interacting spectrum obtained from experiments, whose right-hand side is composed of a transformation matrix $A(\mathbf{B},\mathbf{r})$ (which is simulated and constructed by Pake doublets and a noninteracting spectrum), and a column vector $P(\mathbf{r})$, representing the spin-pair distribution function to be determined. We will demonstrate later in this article that the above system of linear equations is an ill-posed problem requiring a treatment of regularization methods.

With an appropriate treatment to the equation, the solution $P(\mathbf{r})$ can be obtained in a single inverse step providing fast throughput for extracting $P(\mathbf{r})$ from the cw-ESR experimental spectra. The linear equation is actually composed of two ill-posed inverse problems, which are $D(\mathbf{B}) = T(\mathbf{B},\mathbf{B}')M(\mathbf{B}')$ and $M(\mathbf{B}') = K(\mathbf{B}',\mathbf{r})P(\mathbf{r})$, so that $A(\mathbf{B},\mathbf{r}) = T(\mathbf{B},\mathbf{B}') \times K(\mathbf{B}',\mathbf{r})$. The details for the equations are given later. For the simplicity of the presentation, we present the two ill-posed inverse problems separately before making a combination of the two problems.

The first inverse problem, $D(\mathrm{B}) = T(\mathrm{B},\mathrm{B}')M(\mathrm{B}')$: determination of the function for the sum of Pakes

As shown in Eq. 1, it is known that the interacting $D(\mathbf{B})$ versus noninteracting $S(\mathbf{B})$ cw-ESR spectra can be connected by a dipolar broadening function $M(\mathbf{B})$ in the magnetic domain ($\mathbf{B}$). Under the rigid lattice conditions (e.g., spin-labeled proteins in frozen solutions or solutions of high viscosity), the spectrum $D(\mathbf{B})$ can be well approximated by the convolution of $S(\mathbf{B})$ and $M(\mathbf{B})$. Based on this underlying assumption, the broadening function $M(\mathbf{B})$ can be obtained, therefore, by the deconvolution of Eq. 1:

$$D(\mathbf{B}) = S(\mathbf{B}) \otimes M(\mathbf{B}) = \int_0^{\mathbf{B}} S(\mathbf{B} - B')M(B')dB'. \quad (1)$$

The spectra of $D(\mathbf{B})$, $S(\mathbf{B})$, and $M(\mathbf{B})$ are three column vectors with a dimension of N by 1 and are assumed to have the same range of magnetic field, $[-B_{\max}, B_{\max}]$, centered at zero. The deconvolution problem shown in Eq. 1 is called a Volterra integral equation of the first kind. It can be considered as a special case of linear first-kind Fredholm integral equations with a convolution-type kernel as

$$D(\mathbf{B}) = \int_{-B_{\max}}^{B_{\max}} C_{\text{kernel}}(\mathbf{B}, B')M(B')dB'. \quad (2)$$

The kernel C_{kernel} is zero for $B_i < B'$, where B_i represents the i^{th} element in the vector $\mathbf{B}$. The deconvolution problem in Eq. 1 is thus transformed into a linear inverse problem, which is equivalent to the first-kind Fredholm integral equation with a convolution-type kernel denoted by C_{kernel} in Eq. 2. The advantage of working with Eq. 2 is that the underlying theory of the first-kind Fredholm equation is so well developed. Such a realization of the connection between deconvolution versus the first-kind Fredholm integral equation has never been raised and investigated in ESR spectroscopy. By deconvolution, one means that the solution is inferred from the measured data, which is contaminated with noise and inaccessible to direct measurement. Under this condition, the first-kind Fredholm equation is an ill-posed equation whose treatment requires the use of regularization methods.

To solve an ill-posed equation numerically using TIKR method, we present Eq. 2 in a discrete format:

$$D(\mathbf{B}) = T(\mathbf{B},\mathbf{B}')M(\mathbf{B}'). \quad (3)$$

As a result, the spectral column vector $D(\mathbf{B})$ is approximately equivalent to the product of a matrix $T(\mathbf{B},\mathbf{B}')$ and a column vector $M(\mathbf{B}')$, as shown in Eq. 3. The matrix $T(\mathbf{B},\mathbf{B}')$ represents the matrix form of the convolution-type kernel C_{kernel} and is explicitly shown as

$$T = \begin{bmatrix} t_0 & t_{-1} & t_{-2} & \cdots & t_{1-N} \\ t_1 & t_0 & t_{-1} & \cdots & t_{2-N} \\ t_2 & t_1 & t_0 & \cdots & t_{3-N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{N-1} & t_{N-2} & t_{N-3} & \cdots & t_0 \end{bmatrix}, \quad (4)$$

where $t_i = \Delta B \cdot S(\Delta B \cdot i)$ for $|i| \le (N-1)/2$ and $t_i = 0$ otherwise, and $\Delta B = 2B_{\max}/N-1$. It is a Toeplitz matrix whose element depends only on the difference between the indices and is a persymmetric matrix, i.e., it is symmetric across the anti-diagonal. Note that the matrix $T(\mathbf{B},\mathbf{B}')$ is persymmetric for odd N and it becomes unsymmetric for even N. Both types of a Toeplitz matrix can be easily constructed from an noninteracting spectrum $S(\mathbf{B})$ using a MATLAB (The MathWorks, Natick, MA) function, *toeplitz*. (See the Supporting Material for details.) It is clearly shown that the solution of a linear first-kind Fredholm equation (see Eq. 2) is equivalent to the solution of the deconvolution (Eq. 1). The kernel function that corresponds to the convolution equation takes the form of structured matrices, i.e., a Toeplitz coefficient matrix (i.e., T in Eq. 4). Thus, it is possible to transform the deconvolution problem into a linear inverse problem (see Eq. 3). Discretization of the deconvolution problem thereby leads to a linear inverse problem with a Toeplitz-structured coefficient matrix. Equation 3 defines the first linear inverse problem in this report.

The second inverse problem, $M(\mathrm{B}') = K(\mathrm{B}',\mathrm{r})P(\mathrm{r})$: determination of distance distributions in spin pairs

The broadening function, i.e., $M(\mathbf{B})$ in Eq. 1, represents the weighted sum of the Pake doublets, $\kappa(\mathbf{B},r)$, over the distribution of the interspin distances, $P(r)$, as

$$\begin{aligned} M(\mathbf{B}) &= \int_{R_{\min}}^{R_{\max}} \kappa(\mathbf{B},r)P(r)r^2dr, \\ \text{subject to } &\int_{R_{\min}}^{R_{\max}} P(r)dr = 1. \end{aligned} \quad (5)$$

The Pake doublet $\kappa(\mathbf{B},r)$ is an absorption spectrum as a function of magnetic field for a fixed r (24). It is due to the dipolar interactions for randomly oriented spin pairs at the fixed interspin distance. The splitting in the Pake doublet represents the separation between the two resonances in the three triplet states and possesses the splitting magnitude of $3\gamma\hbar/2r^3$, where γ is

the gyromagnetic ratio of an electron and $\hbar$ is Planck's constant. It is clearly shown that the integral equation in Eq. 5 fits the form of the Fredholm integral equation of the first kind, which is known to be an ill-posed inverse problem. To solve the integral equation by TIKR, Eq. 5 is discretized and transformed into

$$M(\mathbf{B}) \cong \sum_{j=1}^{N} \omega_j r_j^2 \kappa(\mathbf{B}, r_j) P(r_j) = K(\mathbf{B}, \mathbf{r}) P(\mathbf{r}), \quad (6)$$

where ω_j is the corresponding weights of quadrature rule to approximate an integral to a discrete summation. The spin-pair distance distribution, $P(\mathbf{r})$, can then be obtained by solving the second inverse problem that connects $M(\mathbf{B})$ with $P(\mathbf{r})$. Hence, Eq. 6 defines the second linear inverse problem in this report.

Tikhonov regularization

The TIKR method has proved to be a mathematically reliable method for an ill-posed problem such as the Fredholm equation of the first kind. It has been previously demonstrated that TIKR can be applied to an ill-posed problem in the analysis of pulsed dipolar ESR experimental data and obtain a reliable solution to the problem. Equations 3 and 6 are the two systems for the linear inverse problems in which we are interested. They both result from the equations (Eqs. 2 and 5) that belong to the form of a Fredholm integral equation of the first kind, which is ill-posed and whose treatment requires a regularization method. The discretized forms in Eqs. 3 and 6 are ready to be solved using TIKR.

There are two ways to deal with the problems of our interest using TIKR. One is to solve the two linear equation systems, separately, in Eqs. 3 and 6. The solution $M(\mathbf{B})$ to Eq. 3 provides a quick access for estimating the average distances in spin pairs. This distance information is often found adequate for the study of structural biology. If one is also interested in the probability function of interspin distance distributions, the $M(\mathbf{B})$ can be further substituted into Eq. 6 to solve for $P(\mathbf{r})$. Alternatively, one can combine Eqs. 3 and 6 to obtain a single linear equation system, i.e., $D(\mathbf{B}) = A(\mathbf{B},\mathbf{r})P(\mathbf{r})$, where the transformation matrix $A(\mathbf{B},\mathbf{r})$ is the matrix-matrix product of $T(\mathbf{B},\mathbf{B}') \times K(\mathbf{B}',\mathbf{r})$, so that the solution vector $P(\mathbf{r})$ can be directly obtained in a single TIKR regularization step.

TIKR method is based on a modification of the first-kind Fredholm integral equation, i.e., a discrete transformation from an integral equation (for example, Eq. 2 or Eq. 5) to a system of linear equations (for example, Eq. 3 or Eq. 6). This discrete form, in the present of finite noise, is ill-posed, so it cannot be solved in a straightforward manner by directly inverting the matrix. Equation 7 shows the TIKR functional to be minimized to obtain the desired solution $P(\mathbf{r})$,

$$\Phi_{\mathrm{TIKR}}[P(\mathbf{r})] \equiv \|A(\mathbf{B},\mathbf{r})P(\mathbf{r}) - D(\mathbf{B})\|^2 + \lambda^2 \|LP(\mathbf{r})\|^2, \quad (7)$$

where λ is a regularization parameter to be determined and L is a differential operator whose order can be arbitrarily chosen to vary the effect of suppression over the unwanted noiselike oscillations in the regularized solution. The great advantage of employing the TIKR method is that the minimization of the TIKR functional yields a unique and simple solution column vector for a given λ,

$$P(\mathbf{r})_\lambda = \left(A^{\mathrm{T}}(\mathbf{B},\mathbf{r})A(\mathbf{B},\mathbf{r}) + \lambda^2\right)^{-1} A^{\mathrm{T}}(\mathbf{B},\mathbf{r})D(\mathbf{B}), \quad (8)$$

where the kernel matrix $\boldsymbol{A}$ is not necessarily a square matrix and can be decomposed by SVD, as detailed elsewhere (16). The TIKR solution is obtained, therefore, in a direct manner without performing numerical iterations and optimizations. Likewise, the equations shown in Eqs. 3 and 6 can also be solved separately using the TIKR method, following Eqs. 7 and 8.

The regularization parameter of the TIKR functional is determined using the L-curve criterion. The L-curve is a parametric plot whose x and y axes are the norms of the regularized solution $\|LP(\mathbf{r})\|$ and the residual $\|A(\mathbf{B},\mathbf{r})P(\mathbf{r}) - D(\mathbf{B})\|$, respectively. The optimal λ is determined by the point on the L-curve that is at the corner of the curve corresponding to maximum curvature since the minimization of the Tikhonov functional of Eq. 7 is a compromise between minimizing the residual norm, and keeping the solution norm small. Hence, the corner of the L-curve corresponds to a desirable balance between minimized norms, and the corresponding λ is a good one.

There are virtues from the TIKR method. TIKR is a fast and direct regularization algorithm, which is easy to compute numerically. For regularization in obtaining a solution vector of several hundreds of elements, it takes only fractions of seconds, requiring no time-consuming iterative procedure or fitting, and providing, thereby, much higher resolution in the regularized solution. Furthermore, it guarantees the solution uniqueness of an ill-posed problem for a given regularization parameter. The technique-demanding steps involved in extracting distances in spin pairs from cw-ESR experiments are thus greatly eased with our new approach. The procedure is summarized below. The matrix $A(\mathbf{B},\mathbf{r})$ is the result of the matrix-matrix product of $T(\mathbf{B},\mathbf{B}') \times K(\mathbf{B}',\mathbf{r})$, where the former and the latter are constructed from $S(\mathbf{B})$ and the Pake functions, respectively. (See the Supporting Material for the details of practical application.) Once the matrix $A(\mathbf{B},\mathbf{r})$ is constructed, the solution $P(\mathbf{r})$ is ready to be obtained by Eq. 8 with the optimal regularization parameter λ determined in the L-curve-based TIKR minimization (16,25).

It is known that the TIKR method, as employed in a direct manner as described above, cannot incorporate nonnegative constraint for the solution. Thus, the TIKR result is typically characterized by the feature that some noiselike small perturbations always exist as the solution approaches to zero. The noiselike perturbations can be simply ignored. Alternatively, the maximum entropy principle (26) can be incorporated into the TIKR functional to secure a positive solution, which is more physically meaningful. However, it must be performed iteratively, requiring a large amount of computation time. We here suggest a simple refinement to the TIKR result. We propose to refine the TIKR result by fitting a set of Gaussians to the major distribution in $P(\mathbf{r})$. By making this refinement to the TIKR result, not only does one guarantee a physically meaningful probability distribution for $P(\mathbf{r})$, but one also preserves the TIKR method's great advantages. We will employ the Gaussian-based refinement to the TIKR results of the experimental data later in Results and Discussion.

RESULTS AND DISCUSSION

Application to simulated data

We first describe the production of the simulated data that was used to demonstrate the improvements to the data analysis by our novel approach. Five distance distributions, whose average distances range from 0.8 to over 2 nm, were simulated and plotted by solid line in Fig. 1 *a*. Each of the distributions is composed of two Gaussians with various linewidths. The average distances of the two Gaussians in the simulated distributions are, respectively (from *top* to *bottom traces*), (0.8, 0.82), (1, 1.1), (1.05, 1.1), (1.5, 1.6), and (1.2, 2.0) nanometers. The dipolar broadening functions that represent the sums of Pakes due to the simulated distributions were simulated using Eq. 6 and plotted by solid lines in Fig. 1 *b*. All of the simulated functions $M(\mathbf{B})$ show clear and distinct patterns for the sums of Pake functions and are ready to be convoluted with a noninteracting spectrum by Eq. 3 to generate the interacting spectra. In the model studies shown here, we used the experimental spectrum of a singly-labeled PPII peptide (described later; see Fig. S2) as the noninteracting spectrum. The overall magnetic field scanned was $\pm$100 Gauss relative to the resonance center. Thus, the resulting lower limit for the measurable distance range is 0.6 nm. The upper limit is set

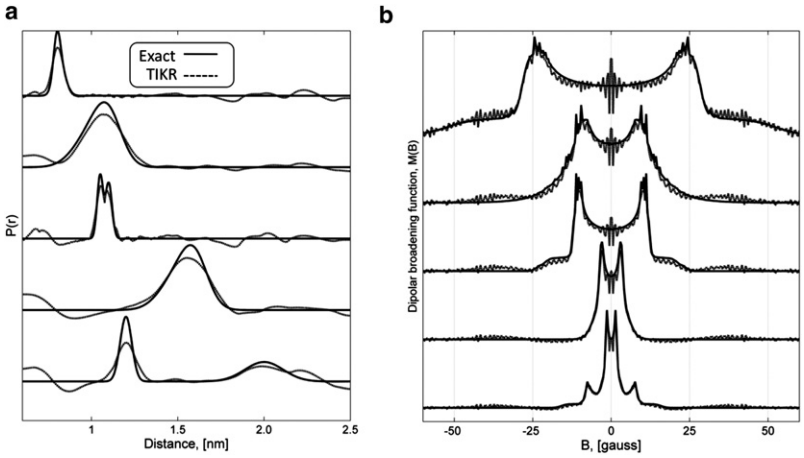

FIGURE 1 Model studies used for the demonstration of the new approach. We first generate the simulated data, which includes the distributions $P(\mathbf{r})$, plotted by solid lines in panel *a*, and the $M(\mathbf{B})$ functions, plotted by solid lines in panel *b*, that represent the sums of Pakes. The simulated $M(\mathbf{B})$ are used to generate the simulated interacting spectra $D(\mathbf{B})$ (not shown) using Eq. 3. To demonstrate our method, we solve Eq. 3 using TIKR to obtain the $M(\mathbf{B})$ functions plotted by dashed lines in panel *b*. The recovered $M(\mathbf{B})$ functions are substituted into Eq. 6. We then apply TIKR to solve Eq. 6 to obtain the $P(\mathbf{r})$ functions shown by dashed lines in panel *a*.

2.5 nm. For a clear presentation, we only plot the broadening function in the range of ± 60 Gauss in Fig. 1 *b*. To reveal a high-resolution distribution function that would be potentially informative for studying the dynamic structure of spin-labeled molecules, the length of the solution vectors, e.g., $P(\mathbf{r})$ and $M(\mathbf{B})$, is >512 in all of our simulations as well as in all of the experimental data shown later.

Fig. 1 *b* shows the regularized solution $M(\mathbf{B})$ (*dashed lines*) obtained by solving the deconvolution problem, i.e., the inverse problem shown in Eq. 3, using TIKR method. The broadening functions $M(\mathbf{B})$, recovered by TIKR method from the simulated interacting spectra, show remarkably good agreements with the simulated broadening functions (i.e., *solid lines* in Fig. 1 *b*). As a result, the average distance (i.e., $\langle r \rangle$) for each simulation case is ready to be accurately determined by the doublet splitting (which is equal to $3\gamma\hbar/2\langle r\rangle^3$) in the recovered $M(\mathbf{B})$ function. The computational time required for the deconvolution took only fractions of a second. It is clearly demonstrated in the model study that our direct treatment of TIKR method in the deconvolution of noninteracting versus interacting spectra is a fast and reliable mathematical approach. It does not require experience-biased judgment to process the experimental data, and hence provides great improvement in extracting distance information from cw-ESR experiments. For biologically relevant applications mainly focused on obtaining the average value of distances, $\langle r\rangle$, our approach is here demonstrated to be an easy, fast, and direct treatment to the analysis of cw-ESR spectra.

For the demonstration of our novel treatment to the second inverse problem of this report, the obtained broadening functions $M(\mathbf{B})$, shown in Fig. 1 *b* by dashed lines, were used and plugged into Eq. 6 to solve for the distance distributions $P(\mathbf{r})$. The solutions for $P(\mathbf{r})$, obtained by the TIKR method, are shown in Fig. 1 *a* by dashed lines. First of all, the average distances of the TIKR results (*dashed lines*) are well consistent with the average distances of the exact solutions (*solid lines*). Although there are some noiselike perturbations in the $M(\mathbf{B})$ functions, they are uniformly distributed and, therefore, do not seem to have any effect on the regularized solutions $P(\mathbf{r})$. The recovered distance distributions are very well consistent with the simulated distributions (*solid lines*). In particular, the result displayed in the third trace from the top in Fig. 1 *a* shows that the two narrow Gaussians separated by only 0.05 nm in distance are unambiguously recovered in our TIKR analysis. The result displayed in the bottom trace demonstrates a great success in the recovery of a distribution consisting of two extreme cases, i.e., narrow versus broad, separating in long distance.

It should be noted that there are always some oscillations of insignificant intensity found in the wings of the recovered $P(\mathbf{r})$ functions, whereas, in the wings of the $M(\mathbf{B})$ functions obtained by TIKR, there are no such similar oscillations. We think that the appearance of the peripheral oscillations in the recovered $P(\mathbf{r})$ somewhat reflects a higher extent of the kernel illness. The kernel matrix $\boldsymbol{T}$ for the first inverse problem is much less ill-posed than the kernel matrix $\boldsymbol{K}$ for the second inverse problem, since the former is transformed from a well-posed deconvolution problem. Fortunately, we found in all of our model studies that the peripheral oscillations in the recovered $P(\mathbf{r})$ were always of small intensity as compared to the major distributions. Thus, the small oscillations in the wings of the $P(\mathbf{r})$ function, recovered by TIKR, can simply be ignored. Alternatively, one may fit a set of Gaussians to the major distributions in the TIKR result so as to obtain an all-positive probability distance distribution.

We emphasize one of the great improvements to the determination of interspin distance distributions due to our approach. It is that our approach enables high-resolution $P(\mathbf{r})$ distance distribution that is essentially important to the study of dynamics. The problem addressed in the linear system, $D(\mathbf{B}) = A(\mathbf{B},\mathbf{r})P(\mathbf{r})$, is ill-posed. The instability of the problem can be easily tested by solving the equation using a linear least-squares method. Here, we employed a MATLAB function, *lsqnonneg*, which is a powerful subroutine for linear least-squares fitting with nonnegative constraints. Trace *a* in Fig. 2 is the TIKR result (i.e., the *third trace* from the *top* in Fig. 1 *a*) displayed in the distance range of 0.9 and 1.3 nm. The distribution displays a smoothing curve of high resolution

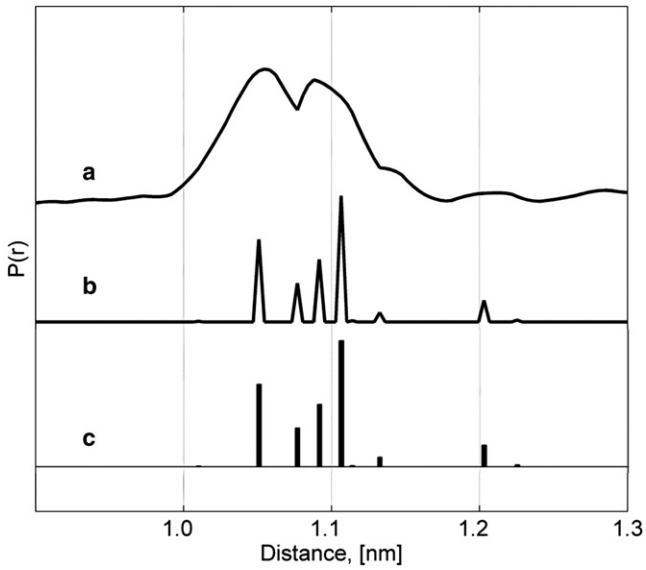

FIGURE 2 Trace *a* is the third trace from top in Fig. 1 *a* within the range of 0.9 and 1.3 nm. It shows that the bimodal distribution is unambiguously recovered by TIKR. With the same given condition of the numerical experiment as trace *a*, we solve the problem by performing linear least-squares fitting with MATLAB and obtain the result shown as trace *b*. Trace *c* shows the histogram plot of trace *b*. TIKR provides the highest resolution for $P(\mathbf{r})$.

that unambiguously reveals a bimodal distribution in the result. Given the same simulation condition as that for trace *a*, our linear least-squares fitting with MATLAB returned a noiselike solution. The result was plotted as trace *b* in Fig. 2. The resolution is greatly reduced though lengths of the solution vectors for traces *a* and *b* are the same. Compared with traces *a* and *b* in Fig. 2, it clearly demonstrates that treating an ill-posed problem with an appropriate regularization method is essentially important. The computation time required for trace *b* is ~500 times greater for trace *a*. TIKR is a fast and efficient method to obtain a reliable solution for an ill-posed problem. Trace *c* in Fig. 2 shows the histogram plot of trace *b*. The information about the width of the distance distribution, which is of importance to the study of dynamics, is completely missing from the result of the least-squares fits. The poor resolution in the distance distribution by lest-squared fits makes it impossible to extract useful information about dynamic structure from $P(\mathbf{r})$.

The study of secondary structure of polyproline peptides

We performed cw-ESR experiments at temperatures 250 K and 275 K to obtain the spectra of the spin-labeled peptides. They are singly-labeled peptides (Ac-P_4CP_6-NH_2) and doubly-labeled peptides (Ac-$P_4CP_nCP_{5-n}$-NH_2 for $n = 0, 1, 2, 3, 4$), where *Ac* indicates an acetylated N-terminus and NH_2 indicates an amidated C terminus. (See Fig. S1, Fig. S2, and Fig. S3 for the experimental spectra.) Hereafter, the above spin-labeled peptides are denoted by C5R1 and C(5,6+*n*)R1. The spectrum of C5R1 was assigned to the column vector $S(\mathbf{B})$ and used to construct the transformation matrix $A(\mathbf{B},\mathbf{r})$, as described in Materials and Experimental Methods. The doubly-labeled spectra were directly used to be the column vector $D(\mathbf{B})$ in the Tikhonov functional (see Eq. 7) to solve for the probability distance distribution of spin pairs $P(\mathbf{r})$. The recovered $P(\mathbf{r})$ by TIKR is plotted by solid (for 250 K) and dashed (275 K) lines in Fig. 3, panels *a* and *b*, respectively, for a 40% sucrose and 50% glycerol solution. The separations between intraspins for each of the peptides and the respective average distances, calculated from the recovered $P(\mathbf{r})$, are marked next to the distributions in Fig. 3, *a* and *b*. The uncertainty value of the average distance indicates the sum of the standard deviations of the Gaussians that best fit to the recovered $P(\mathbf{r})$. The Gaussian-based refinement to the solution is described in Materials and Experimental Methods. The results are well consistent with our general expectation. The average distance in spin label pairs increases with the increasing separation between the spin-labeled sites on a PPII peptide. The width of the obtained distance distribution shows a distinct increasing trend with respect to the increasing distance between the spin-labeled sites. In Fig. 3 *b*, the result also shows the

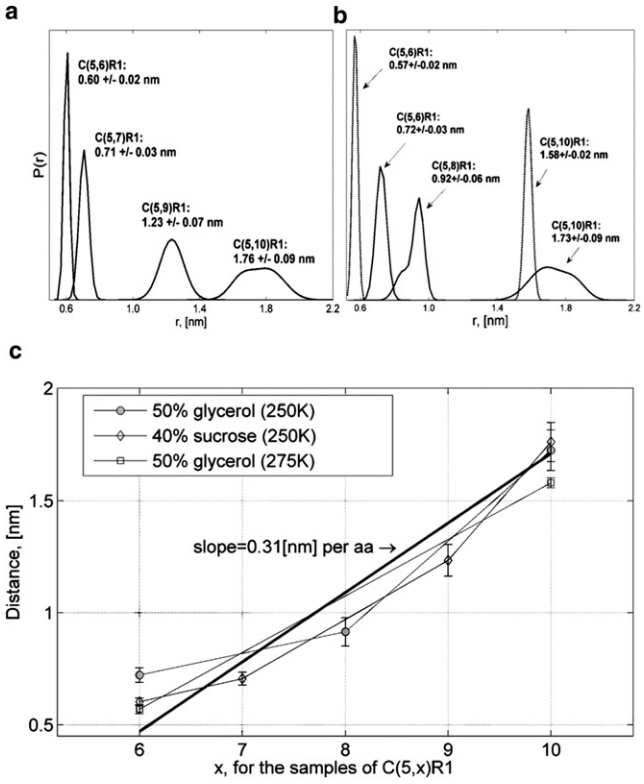

FIGURE 3 Recovered probability distance distributions by TIKR for the spin-labeled PPII peptides in (*a*) 40% sucrose and (*b*) 50% glycerol solutions at 250 K are plotted by solid lines. The dashed lines in panel *b* represent the results at 275 K. (*c*) The plot of the average distances of the TIKR results with variant spin-label separations from C5R1. The bold line shows the second-order polynomial curve with a fixed slope value 0.31 derived from crystallographic models.

distances obtained at 275 K (*dashed lines*) are consistently shorter than the distances at 250 K.

Fig. 3 *c* shows the plot of the average distances of the TIKR results with respect to the position of spin-labeled sites other than C5. The obtained distances increase with the increasing interspin label separation. The bold solid line represents a second-order polynomial curve with a slope value of 0.31, which is the vertical rise (in nanometers) per amino acid along a PPII helix derived in crystallographic molecular models. Remarkably, our TIKR analyses show a consistent trend over the experimental series studied, although the respective slopes differ slightly from the experimental conditions. The results suggest that the secondary structure of PPII is well maintained and found to show insignificant variations with the studied conditions. Further study on the PPII dynamic structure in solution by ESR is warranted.

In summary, we have provided a novel mathematical treatment to the analysis of cw-ESR distance measurements. The equations involved in the complicated spectral analysis have been transformed into one single linear inverse problem, which is ready to be solved with Tikhonov regularization. It has been demonstrated that the approach enables accurate measurements in the distance range of ~0.5–2.5 nm and provides high-resolution interspin distribution. The enhanced resolution enables a more detailed picture of the structure of spin-labeled biomolecules in solution state. Additionally, it is suggested that the novelty of our approach concept would greatly benefit other fields, for example, that of fluorescence spectroscopy, for which the equations of interest also belong to the first-kind Fredholm integral type.

SUPPORTING MATERIAL

Three figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(09)01036-4.

Electron spin resonance measurements were conducted at the National Science Council Research Instrument Center of Taiwan located at National Tsing Hua University.

This work was supported by Taiwan NSC grants No. NSC96-2113-M-007-033-MY2 (to Y.-W.C.) and No. NSC97-2113-M-007-008-MY2 (to J.-C.H.).

REFERENCES

1. Altenbach, C., A. K. Kusnetzow, O. P. Ernst, K. P. Hofmann, and W. L. Hubbell. 2008. High-resolution distance mapping in rhodopsin reveals the pattern of helix movement due to activation. *Proc. Natl. Acad. Sci. USA.* 105:7439–7444.
2. Schiemann, O., and T. F. Prisner. 2007. Long-range distance determinations in biomacromolecules by EPR spectroscopy. *Q. Rev. Biophys.* 40:1–53.
3. Fanucci, G. E., and D. S. Cafiso. 2006. Recent advances and applications of site-directed spin labeling. *Curr. Opin. Struct. Biol.* 16:644–653.
4. Freed, J. H. 2000. New technologies in electron spin resonance. *Annu. Rev. Phys. Chem.* 51:655–689.
5. Georgieva, E. R., T. F. Ramlall, P. P. Borbat, J. H. Freed, and D. Eliezer. 2008. Membrane-bound α-synuclein forms an extended helix: long-distance pulsed ESR measurements using vesicles, bicelles, and rodlike micelles. *J. Am. Chem. Soc.* 130:12856–12857.
6. Park, S. Y., P. P. Borbat, G. Gonzalez-Bonet, J. Bhatnagar, A. M. Pollard, et al. 2006. Reconstruction of the chemotaxis receptor-kinase assembly. *Nat. Struct. Mol. Biol.* 13:400–407.
7. Hanson, S. M., E. S. Dawson, D. J. Francis, N. Van Eps, C. S. Klug, et al. 2008. A model for the solution structure of the rod arrestin tetramer. *Structure.* 16:924–934.
8. Milov, A. D., R. I. Samoilova, Y. D. Tsvetkov, F. Formaggio, C. Toniolo, et al. 2007. Self-aggregation of spin-labeled alamethicin in EPC vesicles studied by pulsed electron-electron double resonance. *J. Am. Chem. Soc.* 129:9260–9261.
9. Cuello, L. G., D. M. Cortes, and E. Perozo. 2004. Molecular architecture of the KvAP voltage-dependent K^+ channel in a lipid bilayer. *Science.* 306:491–495.
10. Crane, J. M., C. Mao, A. A. Lilly, V. F. Smith, Y. Suo, et al. 2005. Mapping of the docking of SecA onto the chaperone SecB by site-directed spin labeling: insight into the mechanism of ligand transfer during protein export. *J. Mol. Biol.* 353:295–307.
11. Dong, J., G. Yang, and H. S. Mchaourab. 2005. Structural basis of energy transduction in the transport cycle of MsbA. *Science.* 308:1023–1028.
12. Vasquez, V., M. Sotomayor, J. Cordero-Morales, K. Schulten, and E. Perozo. 2008. A structural mechanism for MscS gating in lipid bilayers. *Science.* 321:1210–1214.
13. Altenbach, C., K. J. Oh, R. J. Trabanino, K. Hideg, and W. L. Hubbell. 2001. Estimation of inter-residue distances in spin labeled proteins at physiological temperatures: experimental strategies and practical limitations. *Biochemistry.* 40:15471–15482.
14. Rabenstein, M. D., and Y. K. Shin. 1995. Determination of the distance between two spin labels attached to a macromolecule. *Proc. Natl. Acad. Sci. USA.* 92:8239–8243.
15. Steinhoff, H. J., N. Radzwill, W. Thevis, V. Lenz, D. Brandenburg, et al. 1997. Determination of interspin distances between spin labels attached to insulin: comparison of electron paramagnetic resonance data with the x-ray structure. *Biophys. J.* 73:3287–3298.
16. Chiang, Y.-W., P. P. Borbat, and J. H. Freed. 2005. The determination of pair distance distributions by pulsed ESR using Tikhonov regularization. *J. Magn. Reson.* 172:279–295.
17. Jeschke, G., G. Panek, A. Godt, A. Bender, and H. Paulsen. 2004. Data analysis procedures for pulse ELDOR measurements of broad distance distributions. *Appl. Magn. Reson.* 26:223–244.
18. Jeschke, G., and Y. Polyhach. 2007. Distance measurements on spin-labeled biomacromolecules by pulsed electron paramagnetic resonance. *Phys. Chem. Chem. Phys.* 9:1895–1910.
19. Adzhubei, A. A., and M. J. E. Sternberg. 1993. Left-handed polyproline-II helices commonly occur in globular proteins. *J. Mol. Biol.* 229:472–493.
20. Kleywegt, G. J., and T. A. Jones. 1996. Phi/psi-chology: Ramachandran revisited. *Structure.* 4:1395–1400.
21. Shi, Z. S., R. W. Woody, and N. R. Kallenbach. 2002. Is polyproline II a major backbone conformation in unfolded proteins? *Unfolded Proteins.* 62:163–240.
22. Rath, A., A. R. Davidson, and C. M. Deber. 2005. The structure of ''unstructured'' regions in peptides and proteins: role of the polyproline II helix in protein folding and recognition. *Biopolymers.* 80:179–185.
23. Bochicchio, B., and A. M. Tamburro. 2002. Polyproline II structure in proteins: identification by chiroptical spectroscopies, stability, and functions. *Chirality.* 14:782–792.
24. Pake, G. E. 1948. Nuclear resonance absorption in hydrated crystals—fine structure of the proton line. *J. Chem. Phys.* 16:327–336.
25. Hansen, P. C., and D. P. O'Leary. 1993. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput.* 14:1487–1503.
26. Chiang, Y.-W., P. P. Borbat, and J. H. Freed. 2005. Maximum entropy: a complement to Tikhonov regularization for determination of pair distance distributions by pulsed ESR. *J. Magn. Reson.* 177:184–196.