Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Magnetic Resonance 162 (2003) 141–157
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

Inverse methods in two-dimensional NMR spectral analysis

Jacco D. van Beek,[a] Beat H. Meier,[a,*] and Hartmut Schäfer[b,1]

[a] *ETH Zurich, Physical Chemistry, ETH-Hönggerberg, Zurich CH-8093, Switzerland*
[b] *Institute of Physical Chemistry, University of Nijmegen, Toernooiveld 1, Nijmegen, The Netherlands*

Received 25 July 2002; revised 4 December 2002

Abstract

Solid-state NMR is a valuable technique for the study of disordered materials. Analysis of such spectra usually involves solution of so-called *ill-posed inverse problems*. Here we present a strategy for the analysis of two-parameter two-dimensional NMR problems and test it on 2D DECODER and DOQSY experiments. Using Monte Carlo tests, constraints are determined for the resolution and accuracy of the analysis for both experiments. The methods are finally applied to spectra of spider dragline silk, a heterogeneous solid fibrous protein.

Keywords: Tikhonov regularization; 2D NMR; DECODER; DOQSY; (torsion) Angle distribution; Ill-posed problems; Fredholm integral of the first kind

1. Introduction

Solid-state nuclear magnetic resonance (NMR) is able to study disordered and partially ordered materials like glasses, zeolites, polymers, liquid crystals, and biological materials. In such materials the structural and dynamical parameters are not sharply defined but characterised by probability distribution functions (PDFs) instead. Such distributions are mapped into distributions of NMR parameters, such as the chemical shift, the dipolar and the quadrupolar interaction, relaxation, spin-diffusion and exchange rates, and others.

The extraction of the structural information from NMR spectra falls in the category of *ill-posed inverse problems*. The solution of the *direct problem* (e.g., calculation of the spectra from a parameter distribution) is of course a prerequisite for solving the inverse problem but is not sufficient for a unique solution. Strategies to solve such problems are well established for several fields [1–4] and have been applied to solid-state NMR [5–14].

The selected NMR examples mentioned above are concerned with one-dimensional inverse problems, where a distribution in one parameter is extracted from NMR data, which may be multi-dimensional. Two-dimensional inverse problems have been addressed to determine distributions in chemical-shift and quadrupolar parameters from MAS and MQMAS spectra in inorganic disordered materials [15,16], determine distributions in torsion angles and orientational angles in polymers [17], and determine T_1–T_2 relaxation-time distributions in porous materials [18]. In this paper, we give a full description of our attempts to extract distributions in dihedral angles from DOQSY spectra of solid proteins [19,20] and two-angle orientation distributions from DECODER spectra [20].

From these comparatively few examples it is clear that the potential of methods to address two-dimensional inverse problems in solid-state NMR is significant. The first part of this paper introduces the concepts and notations used in this paper and the strategy for the analysis of the two-dimensional (in parameter space) problems. In the second part of this paper the application of this approach will be demonstrated with two practical examples, the 2D DOQSY and 2D DECODER experiments, which were used to extract structural information from natural silks. As silks are

* Corresponding author. Fax: +41-1-632-1621.
E-mail address: beme@nmr.phys.chem.ethz.ch (B.H. Meier).
[1] Present address: BRUKER BioSpin GmbH, Analysis Group, Silberstreifen, Rheinstetten D-76287, Germany.

1090-7807/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S1090-7807(02)00193-3

less than 30% crystalline a non-trivial distribution of structural parameters is expected. The experimental data of silk fibroins, presented in the final part of the paper, indeed could not be explained by using a single narrow distribution of the appropriate parameters.

2. Theory

2.1. Defining inverse problems and ill-posedness

To introduce and illustrate the definitions and the notation important in the context of our approach to inverse problems, we will use a simple one-dimensional example illustrated in Fig. 1. We describe an NMR experiment that yields a single Gaussian line, the position of which is defined by a parameter θ. Coexistence of n sites with different θ_i generates a spectrum (see Fig. 1c) that is a superposition of n lines, weighted by their relative abundance $g(\theta)$. The spectrum $s(\omega)$ is measured at m discrete frequency points ω_j. The lineshape function $K(\omega,\theta)$ for this experiment is calculated for the different values of θ_i and ω_j, and gives a set of basis spectra (Fig. 1a). If the abundances g are known, the spectrum s is given by

$$s = Kg, \tag{1}$$

where s is the data vector $[s_1, s_2, \ldots, s_m]^{\mathrm{T}}$, g is the vector $[g_1, g_2, \ldots, g_n]^{\mathrm{T}}$ of the relative abundances, and K is an

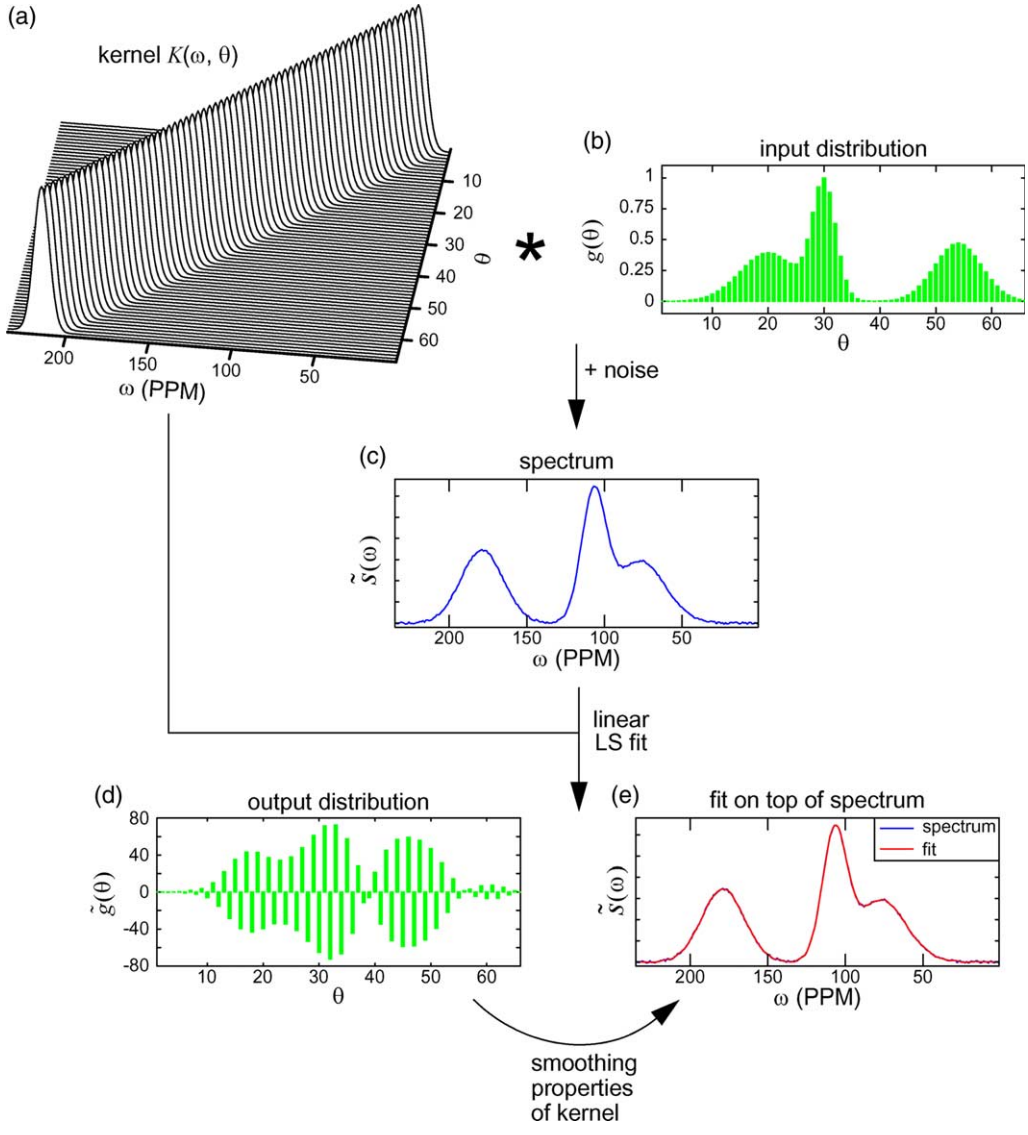

Fig. 1. An example NMR experiment was simulated with a spectrum dependent on one parameter θ. The kernel (a) is the set of basis spectra covering the parameter domain θ. By defining an input distribution function (b), and by adding some Gaussian noise to the frequency domain, the spectrum (c) was obtained. Note that θ_i is a discrete variable that covers the range of θ, i.e., $\theta_i = \theta_{\min} + (\theta_{\max} - \theta_{\min}) * (i-1)/(n-1)$ with n the total number of basis spectra ($n = 66$ in this example). Reversing the procedure, i.e., determining a distribution function from a spectrum, using the known kernel, was done by applying an unconstrained linear LS fit. (d) A strongly oscillatory solution was obtained then even though the fit (e) was excellent. This demonstrates the smoothing properties of the kernel. Numerical details are given in Section 3.1.

$m \times n$ sized matrix whose columns correspond to the basis spectra at $[\theta_1, \theta_2, \ldots, \theta_n]$. In practice one usually wants to determine the abundances $\tilde{g}(\theta)$ from the knowledge of the basis spectra in $K(\omega, \theta)$ and a measured spectrum $\tilde{s}(\omega)$. A least-squares (LS) approach minimizes the discrepancy between the measured and a simulated spectrum, according to

$$\|K\tilde{g} - \tilde{s}\|^2 \rightarrow \min! \tag{2}$$

Using the arbitrarily chosen distribution function $g(\theta)$ from Fig. 1b, the spectrum $\tilde{s}(\omega)$ of Fig. 1c was obtained, by adding some Gaussian noise to the model spectrum $s(\omega)$. Fig. 1e shows the result of an unconstrained linear LS fit of spectrum $\tilde{s}(\omega)$. The fit is excellent indeed, but the obtained output distribution function $\tilde{g}(\theta)$ (Fig. 1d) is different from the input distribution $g(\theta)$.

The continuous version of Eq. (1), $\hat{K}(\omega, \theta)g(\theta) = s(\omega)$, can be stated as a Fredholm integral equation of the first kind:

$$s(\omega) = \hat{K}(\omega, \theta)g(\theta) = \int_\theta g(\theta)K(\omega, \theta)\,\mathrm{d}\theta, \tag{3}$$

i.e., the distribution function $g(\theta)$ is mapped linearly onto the spectrum $s(\omega)$. ω and θ may assign one- or multi-dimensional spectral and parameter domains, i.e., $\theta = \{\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(x)}\}$ and $\omega = \{\omega^{(1)}, \omega^{(2)}, \ldots, \omega^{(y)}\}$, and Eq. (3) covers one- and multi-dimensional cases. The integral kernel $K(\omega, \theta)$ coincides with the spectrum $s(\omega)$ resulting from a sharply defined value θ. $g(\theta)$ is normalised according to $\int_\theta g(\theta)\,\mathrm{d}\theta = 1$ and represents the probability of finding values of the parameter(s) θ within the interval $[[\theta^{(1)}, \theta^{(1)} + \mathrm{d}\theta^{(1)}], [\theta^{(2)}, \theta^{(2)} + \mathrm{d}\theta^{(2)}], [\ldots], [\theta^{(x)}, \theta^{(x)} + \mathrm{d}\theta^{(x)}]]$.

It is well known though that Fredholm integral equations of the first kind are *ill-posed* in the sense of Hadamard [21]. In particular, for experimental data $\tilde{s}(\omega)$ the formal solution of [3], $g(\theta) = \hat{K}(\omega, \theta)^{-1}s(\omega)$, never exists because the operator $\hat{K}(\omega, \theta)$ will not describe the noise or systematic measurement errors. We, therefore, write

$$\tilde{s}(\omega) = \hat{K}(\omega, \theta)g(\theta) + \sigma(\omega), \tag{4}$$

where ideally $\sigma(\omega)$ describes white noise, but may also include deterministic errors. In this case, the *pseudo-inverse* operator $\hat{K}^\dagger(\omega, \theta)$ provides a generalised solution $\tilde{g}(\theta) = \hat{K}^\dagger(\omega, \theta)\tilde{s}(\omega)$, which in the discrete case corresponds to the LS solution [22]

$$\|\hat{K}(\omega, \theta)\tilde{g}(\theta) - \tilde{s}(\omega)\|^2 \rightarrow \min! \tag{5}$$

Lack of uniqueness results in the coexistence of several solutions $\tilde{g}_i(\theta)$ ($\tilde{g}_i(\theta) \neq \tilde{g}_j(\theta)$ if $i \neq j$)), with equivalent misfits for a given spectrum $\tilde{s}(\omega)$. This is clearly seen in Fig. 1 where the distributions in (b) and (d) both lead to virtually the same $\tilde{s}(\omega)$.

Besides non-uniqueness of the solution, lack of stability is often encountered in ill-posed problems. This is also seen in the example of Fig. 1: from the close proximity, in the spectral data space one cannot conclude on the close proximity of the corresponding generating functions $\tilde{g}_1(\theta)$ and $\tilde{g}_2(\theta)$: the fact that

$$\|\hat{K}(\omega, \theta)\tilde{g}_1(\theta) - \hat{K}(\omega, \theta)\tilde{g}_2(\theta)\| = \|\tilde{s}_1(\omega) - \tilde{s}_2(\omega)\| \tag{6}$$

is a small number does not guarantee that the norm

$$\|\tilde{g}_1(\theta) - \tilde{g}_2(\theta)\| \tag{7}$$

is small as well.

Lack of uniqueness and stability causes that ill-posed inverse problems, in general, have many solutions with similar misfit, within the signal-to-noise, which may look completely different. This fundamental problem can only be solved by adding additional information, as a side constraint, to obtain a unique and stable solution.

2.2. Analysis and evaluation of the ill-posed problem

2.2.1. Singular values

The ill-posedness of a problem can be characterised numerically by applying a singular value expansion (SVE) [1] to the kernel K. This infinite series expansion into orthogonal functions $u(\omega)$ and $v(\theta)$ is defined as

$$K(\omega, \theta) = \sum_{l=1}^{\infty} \eta_l u_l(\omega) v_l(\theta). \tag{8}$$

The singular values η_l and the singular functions $u_l(\omega)$ and $v_l(\theta)$ are related via $\int_\theta K(\omega, \theta)v_l(\theta)\,\mathrm{d}\theta = \eta_l u_l(\omega)$. In the discrete case they are a generalization of the eigenvalues of a square matrix. By projecting the spectrum onto the singular functions the solution to the LS problem of Eq. (5) can be rewritten as

$$\tilde{g}(\theta) = \sum_{l=1}^{\infty} \frac{1}{\eta_l}\Big(u_l(\omega), \tilde{s}(\omega)\Big)v_l(\theta). \tag{9}$$

Here $(\cdot, \cdot)$ denotes the scalar product. Assuming $\tilde{g}_1(\theta)$ and $\tilde{g}_2(\theta)$ to be the distributions producing the spectra $\tilde{s}_1(\omega)$ and $\tilde{s}_2(\omega)$, the LS discrepancy in the parameter domain is

$$\|\tilde{g}_1(\theta) - \tilde{g}_2(\theta)\|^2 = \sum_{l=1}^{\infty} \frac{1}{\eta_l^2}(u_l(\omega), \tilde{s}_1(\omega) - \tilde{s}_2(\omega))^2. \tag{10}$$

Small singular values η_l thus lead to a dramatic error amplification by even the smallest differences in the spectra, as was observed in the example of Fig. 1. The decay of the singular values thus provides a measure for ill-posedness. A problem can be roughly classified to be modestly ill-posed if the η_l decay like l^{-y} (with y a positive real number) and severely ill-posed if an e^{-l}-behaviour is observed. The condition number, defined as the ratio between the highest and lowest singular value, may be used as an indication of the severity of ill-posedness.

2.2.2. Correlations

The decay of the singular values is determined by the degree of correlation between the subfunctions $K(\omega, \theta_1)$

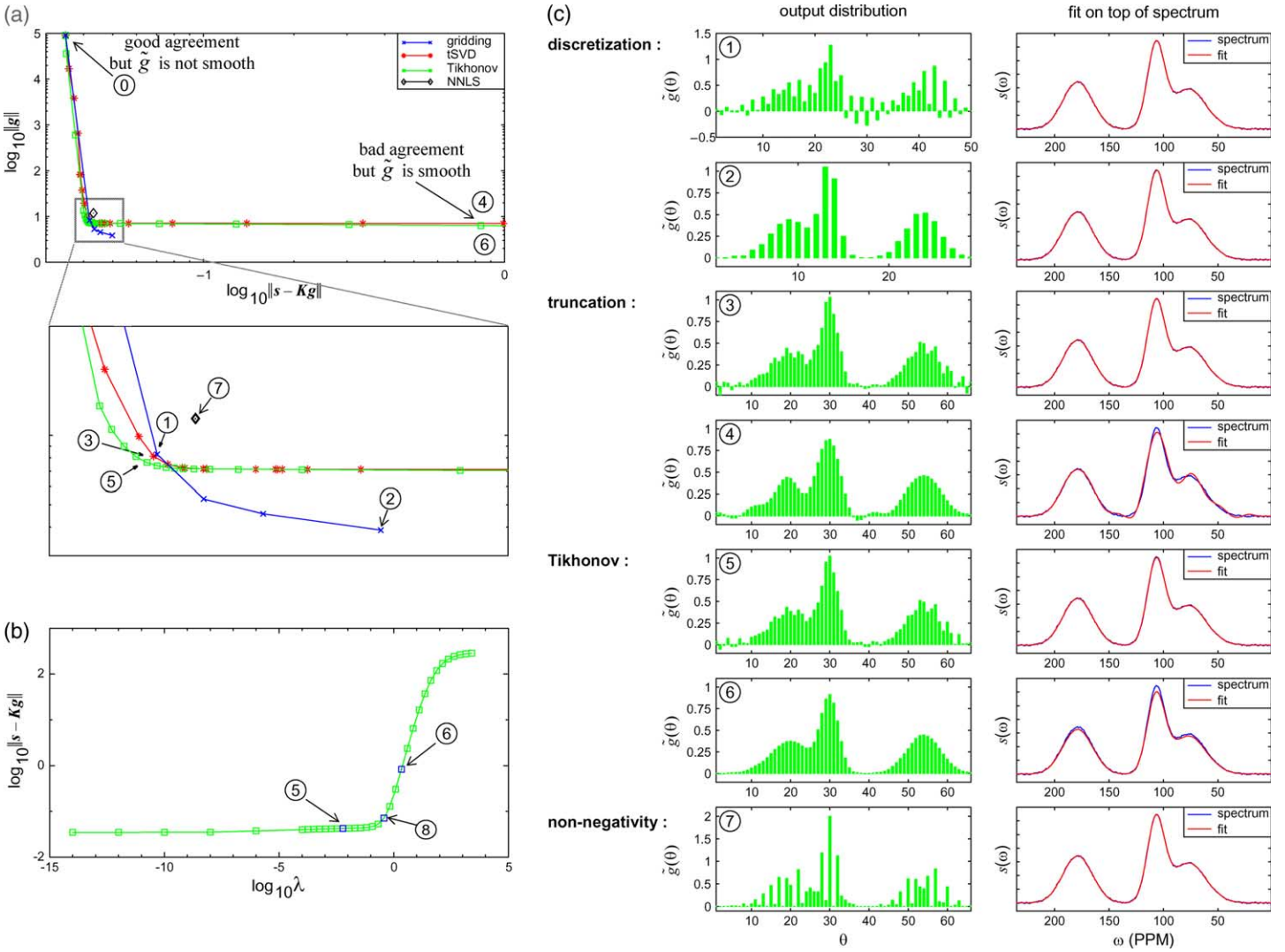

Fig. 2. Regularization increases the stability of ill-posed problems, independently of the approach. (a) The L-plot [1] shows the effects of four different regularization approaches: discretization, truncation, Tikhonov regularization (demanding a smooth solution), and non-negativity (NNLS algorithm). With increasing regularization the norm of the solution $\|\tilde{g}(\theta)\|$ decreases while the discrepancy between spectrum and fit, $\|\tilde{s}(\omega) - K(\omega,\theta)\tilde{g}(\theta)\|$, increases. The enlarged detail shows the sharp bend in all curves. (b) For Tikhonov regularization the application of the discrepancy principle [48] for the determination of the regularization parameter is shown. (c) Eight examples of various degrees of regularization: ⓪ without regularization the result of Fig. 1 is obtained using all four approaches (66 basis spectra). Changed for each particular example was: ① grid density lowered to 50 basis spectra, ② grid density lowered to 29 basis spectra, ③ series truncated for $l > 44$, ④ series truncated for $l > 16$, ⑤ Tikhonov with $\lambda = 6.1 * 10^{-3}$, ⑥ Tikhonov with $\lambda = 2.19$, and ⑦ imposing non-negativity. Situations ④ and ⑥ demonstrate that too much regularization yields an oversmoothed solution. Situation ⑧ shows the result of applying the SC method to determine the regularization parameter.

and $K(\omega,\theta_2)$, which coincide with the basis spectra at θ_1 and θ_2. We characterize the correlation between two functions by the cosine of the angle between the respective functions, i.e.,

$$C_{pq} = \frac{(K_p, K_q)}{\sqrt{(K_p, K_p)(K_q, K_q)}} = \cos(\text{angle}_{pq}), \tag{11}$$

where $K_u = K(\omega,\theta_u)$. It should be noted that the correlation in the example of Fig. 1 is low, and the example is not very strongly ill-posed.

2.3. Regularization

A meaningful solution of the inverse problem requires that we enforce that $\|\tilde{g}_1(\theta) - \tilde{g}_2(\theta)\|$ becomes small, if $\|\tilde{s}_1(\omega) - \tilde{s}_2(\omega)\|$ is small. This is done here by a *regularization* approach. Regularization can realised by extending the minimization problem, defined in Eq. (5), with extra constraint(s), $\hat{\mathscr{L}}(\tilde{g}(\theta))$, i.e.,

$$\|\hat{K}(\omega,\theta)\tilde{g}(\theta) - \tilde{s}(\omega)\|^2 + \lambda\hat{\mathscr{L}}(\tilde{g}(\theta)) \rightarrow \min! \tag{12}$$

Here the operator $\hat{\mathscr{L}}(\tilde{g}(\theta))$ represents the extra constraints to be minimised, e.g., the norm or the second derivative of the solution $\tilde{g}(\theta)$, or an entropy function as is used in maximum-entropy methods [1]. Choosing a proper value for λ is important as it determines the balance of the discrepancy between fit and data (quality of the fit) and the importance of the additional constraint(s).

Fig. 2 shows the effect of four different regularization approaches applied to the example of Fig. 1:

2.3.1. Discretization

Discretization (gridding) of the parameter domain already provides regularization. It means that the continuous functional of Eq. (3) is approximated by

$$s(\omega_j) = \sum_{i=1}^{n} K(\omega_j,\theta_i)g(\theta_i)\Delta(\theta_i) \tag{13}$$

using a discrete grid θ_i and ω_j. $\Delta(\theta_i)$ denotes the (variable) interval size at θ_i.

The singular value decomposition (SVD) of the matrix K is given by

$$K = UWV^{\mathrm{T}}. \tag{14}$$

Here U is an $m \times n$ sized column-orthogonal matrix, W an $n \times n$ sized diagonal matrix with the singular values w_l as its elements, and V is an $n \times n$ sized orthonormal matrix. This modifies the linear LS fit of Eq. (9) to

$$\tilde{g}(\theta) = \sum_{i=1}^{n} \frac{1}{w_l} v_l \cdot u_l^{\mathrm{T}} \cdot \tilde{s} \tag{15}$$

with n is the number of basis spectra in the kernel and u_l, v_l columns of U and V. By decreasing the grid density, i.e., keep n small, oscillations of $\tilde{g}(\theta)$ can be blocked. The regulatory effect of decreasing the grid resolution can be seen from situations ① and ② in Fig. 2. Note that not only the grid density in the parameter domain is of importance. Inappropriate selection of the discrete points in the parameter domain may change the condition number by many orders of magnitude.

2.3.2. Truncation

Alternatively to using a lower grid density one may use the same grid but truncate the series of Eq. (15) at a certain threshold $n_{\text{threshold}} < n$, because this blocks high-frequency components in the spectral domain. This enforces a smoother solution, which is illustrated by situations ③ and ④. This approach still has a discrete step size but this is usually smaller than for discretization. Note the appearance of significant deviations between spectrum and fit for strong regularization in situation ④.

2.3.3. Smoothness

A more intuitive method than truncation is the enforcement of smoothness of $\tilde{g}(\theta)$, as this blocks the oscillations in the parameter domain directly. There are many ways of modelling smoothness, but here we discuss Tikhonov regularization [23] where smoothness is realised by asking for a small norm of $\tilde{g}(\theta)$ leading to the problem

$$\|\hat{K}(\omega,\theta)\tilde{g}(\theta) - \tilde{s}(\omega)\|^2 + \lambda\|\hat{L}\tilde{g}(\theta)\|^2 \rightarrow \min! \tag{16}$$

where $\hat{L}$ is the unity operator. Tikhonov–Phillips regularization is a closely related approach that enforces smoothness, with $\hat{L}$ then being the operator of the second derivative [24].

The constraint of a small norm modifies Eq. (15) to

$$\tilde{g}_\lambda(\theta) = \sum_{i=1}^{n} \frac{w_l}{(w_l^2 + \lambda)} v_l \cdot u_l^{\mathrm{T}} \cdot \tilde{s}, \tag{17}$$

i.e., the $1/w_l$-dependence has become $w_l/(w_l^2 + \lambda)$, which means a reduction of the weight of the high-order singular values (situations ⑤ and ⑥). Note that the choice of λ is not bound to discrete steps.

2.3.4. Non-negativity and other

Enforcing a positive result for $\tilde{g}(\theta)$ also improves stability as is shown by situation ⑦. Finally, in addition to the regularization schemes here, many others have been discussed. and for brevity we refer to [1,24] and references therein.

2.4. Determining the optimum degree of regularization

The best choice of the regularization parameter is such, that it is as large as possible but that it still leads only to a minimal increase in the discrepancy $\|\tilde{s}(\omega) - \hat{K}(\omega,\theta)\tilde{g}(\theta)\|$. The discrepancy, as a function of the regularization parameter can be visualised either in the L-plot or the discrepancy plot [1]. The L-plot shows,

in a double-logarithmic fashion, the discrepancy as a function of the norm of the regularization term, $\|\hat{L}\tilde{g}(\theta)\|$ (Fig. 2a), while the discrepancy-plot shows the discrepancy directly as a function of the regularization parameter λ (for Tikhonov regularization). The best balance between the constraints and the discrepancy is found around the sharp edge for both the L-curve and the discrepancy method.

Alternatively to using a discrepancy in the spectral domain, it is possible to define a discrepancy in the parameter domain. In the context of this work we apply the self-consistency method (SC-method) [25,26] which can be used to estimate both the noise level σ and the optimum regularization parameter λ. The result of applying the SC method is shown by situation ⑧ in Fig. 2.

2.5. Two-dimensional Fredholm equations of the first kind

In the following we want to specifically address the extraction of a distribution of two parameters, e.g., α and β, from a two-dimensional (2D) NMR spectrum. Specifically, we discuss the analysis of two static 2D CSA–

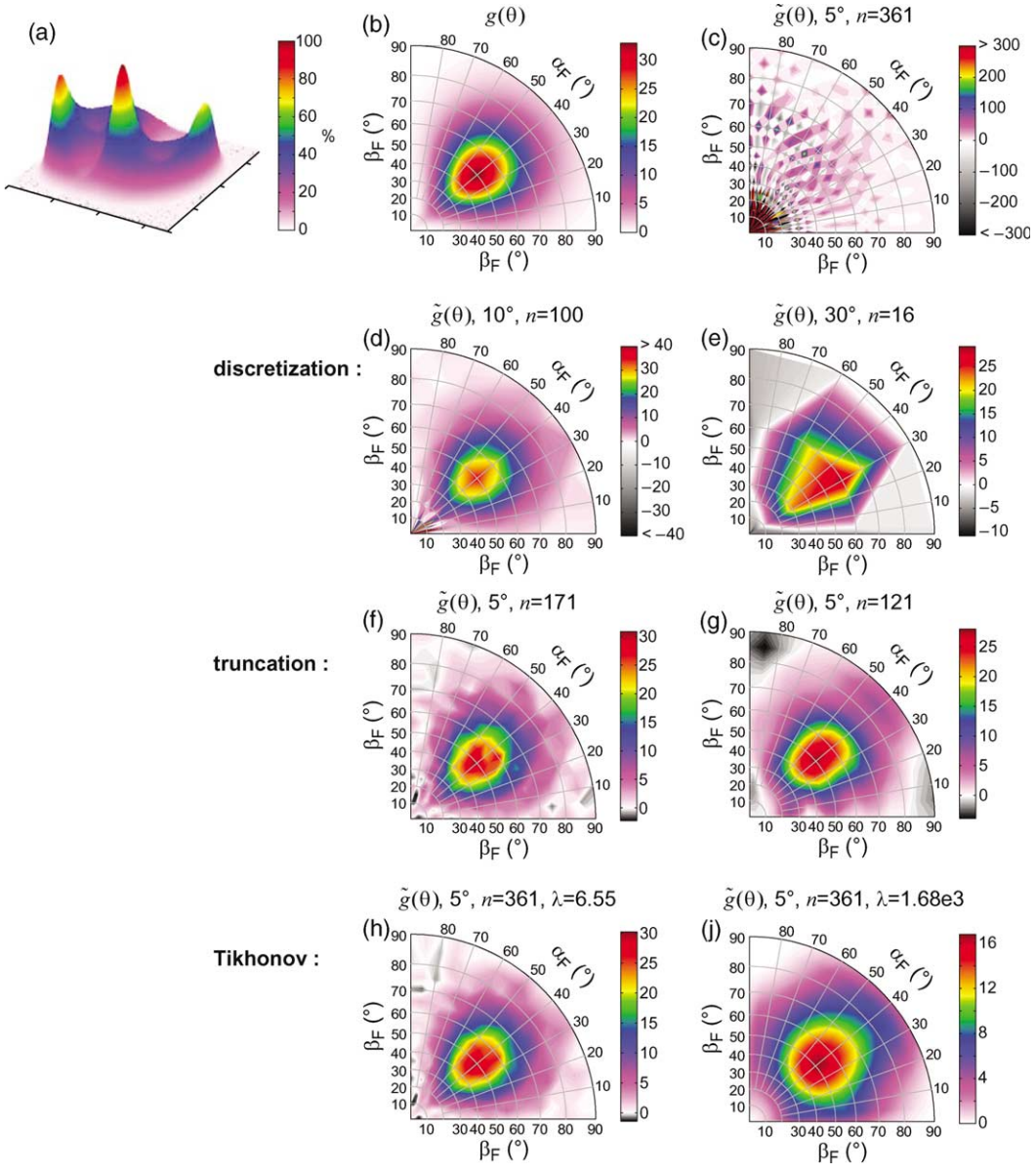

Fig. 3. (a) A 2D DECODER spectrum was simulated using (b) the input distribution with a 5° grid resolution. Gaussian noise was added to the frequency domain. Panel c shows the PDF as obtained from an unconstrained linear LS fit. A highly similar result is obtained when applying virtually no truncation or Tikhonov regularization. Panels d and e show the effect of decreasing the grid resolution to 10° and 30°, respectively. Panel f and g show the effect of truncating Eq. (15) after $l > 171$ and $l > 121$ (of 361), respectively. Finally panels h and j show the effect of Tikhonov regularization with $\lambda = 6.55$ and $\lambda = 1.68\text{e}3$, respectively. The SC method used in this work estimates $\lambda_{\text{opt}} = 12.74$. All three regularization approaches must be applied with care to avoid oversmoothing of the result as demonstrated by e, g, and j. Note the PDF is plotted using a polar projection of α_F and β_F. A uniform distribution (no orientation) would show a uniformly coloured plot. Numerical details are given in the methods section (Example 2).

CSA correlation experiments, DECODER [27–30], and DOQSY [31,32]. The extension from the one-parameter 1D problem to a two-parameter 2D problem can increase the complexity of the problem considerably. The number of parameters and spectral data points are in general approximately squared, correlations may be unevenly spread across the two-dimensional parameter domain and, depending on the algorithm, the size of the problem may induce numerical errors, for example when many orthogonal transformations are calculated. The principles of regularization still hold, however, and the same approaches can be used as for one-parameter problems. This is demonstrated in the example of Fig. 3. An orientation distribution function $g(\alpha,\beta)$ was created (Fig. 3b) and the corresponding 2D DECODER experiment, which is described in more detail below, calculated, and some noise added. Fig. 3c shows the result of an unconstrained linear LS fit. As predicted by theory, the solution is dominated by strong oscillations. The result of the application of either reducing the discretization resolution, truncation, or Tikhonov regularization to the problem (Fig. 3d–j) leads to a more stable solution. This shows that the two-parameter 2D problem basically responds identical to the one-parameter 1D NMR problem of Fig. 2.

2.6. Solution concept and discrete implementation

The previous sections have shown various approaches that can help to solve ill-posed problems, all with similar effects but of different strength and based on different assumptions. Some of these approaches can also be combined in order to achieve better results. In the following, we want to demonstrate the effect of applying the combination of (1) a well-chosen discretization, (2) Tikhonov regularization, and (3) non-negativity, combined into a single linear LS algorithm, to two-parameter 2D NMR problems. The combination performs better, and is more flexible, than using a single scheme, in particular for severely ill-posed problems. The earlier work of Utz [17], on a similar regularization problem, used a truncation approach for regularization.

First, we define the problem by rewriting Eq. (3) into

$$s\big(\omega^{(1)},\omega^{(2)}\big)=\int_{\alpha,\beta} g(\alpha,\beta)K(\omega^{(1)},\omega^{(2)},\alpha,\beta)\,\mathrm{d}\alpha\,\mathrm{d}\beta. \tag{18}$$

The spectrum is discretised into a finite number of m noisy data points, $\tilde{s}_k$, $k=1,\ldots,m$, using a Fourier transform, with

$$\tilde{s}_k = s(\omega_k^{(1)},\omega_k^{(2)})+\sigma_k \tag{19}$$

and where σ_k is given by

$$\sigma_k=\sigma\varepsilon_k, \tag{20}$$

ε_k is a vector containing noise of a given type, e.g., Gaussian, and σ is a scaling factor. For simplicity we assume no systematic errors. Including the Euclidean norm of g as an additional constraint into the least-squares problem yields, according to Eq. (16),

$$\sum_{k=1}^{m}\frac{1}{\sigma_k^2}\left(\tilde{s}_k-\int_{\alpha,\beta}K(\omega_k^{(1)},\tilde{\omega_k^{(2)}},\alpha,\beta)\tilde{g}(\alpha,\beta)\,\mathrm{d}\alpha\,\mathrm{d}\beta\right)^2 +\lambda\|\tilde{g}\|^2\rightarrow\min! \tag{21}$$

In the next step we discretize the integral equation and the regularization term. A natural discretization was defined already by the grid points in the spectral domain due to the discrete sampling and subsequent Fourier transform. Hence, a further discretization scheme had to be introduced for the parameter domain only. This was done using linear splines such that, within the intervals $[[\alpha_i,\alpha_{i+1}],[\beta_j,\beta_{j+1}]]$, $g(\alpha,\beta)$ is approximated by functions that are piecewise linear in α and β, i.e., $g(\alpha,\beta)\cong\bar{g}_{i,j}(\alpha,\beta)$, according to

$$\bar{g}_{i,j}(\alpha,\beta)=c_{ij}^{(1)}(\alpha,\beta)\tilde{g}_{i,j}+c_{ij}^{(2)}(\alpha,\beta)\tilde{g}_{i+1,j} +c_{ij}^{(3)}(\alpha,\beta)\tilde{g}_{i,j+1}+c_{ij}^{(4)}(\alpha,\beta)\tilde{g}_{i+1,j+1}, \tag{22}$$

where $i=1,\ldots,(n_\alpha-1), j=1,\ldots,(n_\beta-1)$, $\tilde{g}$ is the discrete solution vector and

$$\begin{aligned}
c_{ij}^{(1)}(\alpha,\beta)&=\left[\frac{\alpha_{i+1}-\alpha}{\alpha_{i+1}-\alpha_i}\right]\cdot\left[\frac{\beta_{j+1}-\beta}{\beta_{j+1}-\beta_i}\right],\\
c_{ij}^{(2)}(\alpha,\beta)&=\left[\frac{\alpha-\alpha_i}{\alpha_{i+1}-\alpha_i}\right]\cdot\left[\frac{\beta_{j+1}-\beta}{\beta_{j+1}-\beta_j}\right],\\
c_{ij}^{(3)}(\alpha,\beta)&=\left[\frac{\alpha_{i+1}-\alpha}{\alpha_{i+1}-\alpha_i}\right]\cdot\left[\frac{\beta-\beta_j}{\beta_{j+1}-\beta_j}\right],\\
c_{ij}^{(4)}(\alpha,\beta)&=\left[\frac{\alpha-\alpha_i}{\alpha_{i+1}-\alpha_i}\right]\cdot\left[\frac{\beta-\beta_j}{\beta_{j+1}-\beta_j}\right].
\end{aligned} \tag{23}$$

The spectrum $\tilde{s}_k$ was thus approximated by a sum of basis spectra over the intervals $[[\alpha_i,\alpha_{i+1}],[\beta_j,\beta_{j+1}]]$, using the functions $\tilde{g}_{i,j}(\alpha,\beta)$, given by

$$\tilde{s}_k=\sum_{j=1}^{n_\beta}\sum_{i=1}^{n_\alpha}K_{ij}(\omega_k^{(1)},\omega_k^{(2)})\tilde{g}_{ij}, \tag{24}$$

where $K_{ij}(\omega_k^{(1)},\omega_k^{(2)})$ is the basis spectrum belonging to the ijth element of the solution vector at frequency $(\omega_k^{(1)},\omega_k^{(2)})$. The full derivation of Eq. (24) and the exact definitions of the spectra $K_{ij}(\omega_k^{(1)},\omega_k^{(2)})$ are given in Appendix A.

Finally, we can rewrite the LS problem into the matrix equation

$$\chi(\lambda)=\|K\tilde{g}-\tilde{s}\|^2+\lambda^2\|\tilde{g}\|^2. \tag{25}$$

Here $\tilde{s}$ is the data vector $[\tilde{s}_1,\tilde{s}_2,\ldots\tilde{s}_m]^{\mathrm{T}}$, $\tilde{g}$ is the wanted distribution vector $[\tilde{g}_1,\tilde{g}_2,\ldots\tilde{g}_n]^{\mathrm{T}}$, $n=n_\alpha n_\beta$, and the $m\times n$ sized matrix K is the discretised version of the integral kernel. Note that the double index of $\tilde{g}_{ij}$ was transferred into a single index such that $\tilde{g}_l=\tilde{g}_{ij}$ with

$l = n_z(j-1)+i$. The numerical minimization of Eq. (25) follows the strategy outlined by Weese [26]. Both the regularization parameter λ and the noise level σ were estimated using the SC method. Non-negativity was implemented using the NNLS algorithm [22] and SVD as suggested in [33]. All matrix transformations were based on stable Householder transformations and Givens rotations according to [34].

2.7. Experiment 1: 2D DECODER

In the 2D DECODER experiment the principal axis of anisotropic interactions (such as chemical-shift anisotropy, dipolar, or quadrupolar interactions) are determined with respect to the static magnetic-field direction. For this purpose, the orientation of the sample in the field is mechanically altered during the mixing time of the 2D experiment [27–30]. Typically one is interested in the relative orientation of an interaction tensor towards a sample-fixed coordinate system, e.g., the fibre direction, a crystal axis or towards the axis of deformation in a deformed block of polymeric material, and not towards the magnetic field. It is therefore useful to define one or more intermediate coordinate systems between principal axis system (PAS) and the lab frame (LAB), of which only one is of unknown orientation.

In the present case we wanted to investigate the orientation of the chemical-shift anisotropy tensor towards the fibre axis in a uni-axially oriented fibrous sample. By defining a fibre-axis system (FAS), with its z-axis along the fibre direction, the chemical shift tensor in the LAB frame is given by

$$\sigma_{\mathrm{LAB}} = R_L(\alpha_L,\beta_L,\gamma_L)\cdot R_F(\alpha_F,\beta_F,\gamma_F)\cdot \sigma_{\mathrm{PAS}} \cdot R_F^{\mathrm{T}}(\alpha_F,\beta_F,\gamma_F)\cdot R_L^{\mathrm{T}}(\alpha_L,\beta_L,\gamma_L), \tag{26}$$

where R_L and R_F are the Euler rotation matrices describing the transformations from FAS to the LAB and from the PAS to the FAS, respectively. For the Euler rotations the convention of [35] was used. Note that the angle α_F determines the angle of the ^{13}C=O CSA tensor towards the fibre direction, i.e., if $\alpha_F = 0$ then the C=O vector is approximately perpendicular to the fibre direction. β_F specifies the angle of the peptide plane normal towards the fibre direction, i.e., if $\beta_F = 0$ then the normal is parallel to the fibre. In the absence of other interactions, the 2D DECODER spectrum, for a single fixed choice of R_L and R_F, is determined by a two-dimensional delta function

$$s(\omega^{(1)},\omega^{(2)}) = \delta(\omega(\Omega^{(1)})-\omega^{(1)},\omega(\Omega^{(2)})-\omega^{(2)}), \tag{27}$$

with $\Omega^{(1)} = (\alpha_L^{(1)},\beta_L^{(1)},\gamma_L^{(1)},\alpha_F,\beta_F,\gamma_F)$ and $\Omega^{(2)} = (\alpha_L^{(2)},\beta_L^{(2)},\gamma_L^{(2)},\alpha_F,\beta_F,\gamma_F)$. Often, though not necessarily so [30], the change of sample orientation is achieved by a mechanical rotation β_L around an axis perpendicular to the magnetic field. Defining this axis to be the y-axis of the LAB yields $\alpha_L = 0$ and $\gamma_L = 0$. In case of fibre symmetry the angle γ_F is randomly distributed and must be integrated over. The angles α_F and β_F are distributed according to the distribution function $g(\alpha_F,\beta_F)$, which give the abundances of the corresponding tensor orientations towards the fibre, and the spectrum is thus described by

$$s(\omega^{(1)},\omega^{(2)}) = \int_{\alpha_F}\int_{\beta_F} g(\alpha_F,\beta_F) \times \int_{\gamma_F} K(\omega^{(1)},\omega^{(2)},\beta_L^{(1)},\beta_L^{(2)},\alpha_F,\beta_F,\gamma_F)\,\mathrm{d}\gamma_F \times \sin\beta_F\,\mathrm{d}\beta_F\,\mathrm{d}\alpha_F. \tag{28}$$

The fibre symmetry makes that the NMR spectrum is fully described by the parameter ranges $\alpha_F = [0°, 90°]$ and $\beta_F = [0°, 90°]$. All other orientations are related by symmetry. The exact coefficients for the discretised equation, corresponding to Eq. (24), are given in Appendix A.

2.8. Experiment 2: DOQSY

In the 2D DOQSY experiment the double-quantum spectrum of two dipolar-coupled spins is correlated to the respective single-quantum spectra. This experiment has initially been used to determine a molecular torsion angle around two covalently bonded atoms [31,32]. Recently we have applied the experiment to solid silk proteins, labelled at the carbonyl functionality, to simultaneously extract the distribution of the backbone torsion angles (ϕ,ψ) [19,20]. The relative orientation of the two CSA tensors can be in principle uniquely described, bar symmetry-related solutions in the NMR experiment, by a single Euler rotation $R(\alpha,\beta,\gamma)$. Assuming that one knows the orientation of the CSA tensor in the molecular frame, and that only spin pairs of C=O groups neighbouring in the chain contribute to the spectrum, the backbone torsion angles (ϕ,ψ,ω) may be obtained, if the spectrum is sensitive to changes in these angles [19]. Thus

$$R_{1\to 2} = R_{\mathrm{MOL},2}^{\mathrm{T}}(\alpha_2,\beta_2,\gamma_2)\cdot \Gamma(\phi,\psi,\omega) \cdot R_{\mathrm{MOL},1}(\alpha_1,\beta_1,\gamma_1), \tag{29}$$

where $R_{\mathrm{MOL},1}$ and $R_{\mathrm{MOL},2}$ are the Euler rotation matrices to go from the PAS to the respective molecular frame for both spins. From literature the orientation of the PAS in the peptide is approximately known [36–39] and an average value can be used, if the effects on the analysis are small. In the following we assumed that the σ_{33} component is perpendicular to the peptide plane and the σ_{22} component is almost along the C=O bond, with a 5° deviation away from the C–N bond. The implication of these assumptions will be described elsewhere [40].

Note that the Euler rotation $\Gamma(\phi,\psi,\omega)$ depends not only on the torsion angles (ϕ,ψ,ω) but also on the

standard peptide geometry. The geometry of a dipeptide unit is generally considered to be well known. The torsion angle ω is the rotation around the C–N bond in the amide bond and is close to 180°. From a search through 11,876 protein and enzyme crystal structures, determined by X-ray [41], a Lorentzian distribution around 180.1° with a full width at half height (FWHH) of 2.8° was found [40]. The assumption to keep ω fixed at 180° will therefore not perturb the analysis. This means that the spectrum is described by a two-parameter distribution function $g(\phi,\psi)$ according to

$$s(\omega^{(1)},\omega^{(2)}) = \int_{\psi}\int_{\phi} g(\phi,\psi)K(\omega^{(1)},\omega^{(2)},\phi,\psi,\alpha_1,\beta_1,\gamma_1,\alpha_2,\beta_2,\gamma_2)\,\mathrm{d}\phi\,\mathrm{d}\psi. \tag{30}$$

The exact coefficients for the discretised equation, corresponding to Eq. (24), are given in Appendix A.

3. Materials and methods

3.1. Example 1

The convolution kernel K was created by calculating 66 spectra of 235 points in length with a Gaussian line (FWHH = 11 points) centred at point 20 in the first column, and subsequently shifted by 3 points every next column. The input distribution g was generated with three Gaussians of 66 points in length, centred at points 20, 30, and 54, and with FWHH of 12, 5, and 10 points, respectively. The maximum of g was set to 1. The spectrum s was obtained from the matrix multiplication of K and g. Gaussian noise with a maximum amplitude of ± 0.025 was added. By increasing the increment in the shift of the centre of the Gaussian to 4, 5, 6, and 7 points, the size of the parameter domain was decreased to 50, 40, 34, and 29 points, respectively (see also Fig. 2).

3.2. Example 2

A DECODER spectrum was calculated for a CSA tensor with values $(\delta_{11},\delta_{22},\delta_{33}) = (241,178,96)$ PPM, using the same parameters for the integration scheme as given below in Section 3.3. The probability for the orientation (α_F,β_F) was defined to be a two-dimensional Gaussian distribution function centred at 40° and 50° in the α_F and β_F dimensions, respectively, with FWHH = 40° in both dimensions. Line broadening was applied in the time domain by multiplying with a 2D Gaussian (FWHH = 20 points). Gaussian noise was added to the spectrum resulting in a signal-to-noise ratio of 113.6 (signal maximum divided by the standard deviation of the noise). For the analysis the spectrum was reduced to 183×183 points, enclosing the range $[28,210]$ PPM.

3.3. Analysis

DECODER and DOQSY spectra were calculated in frequency space using the GAMMA C++ simulation environment [42]. DECODER spectra $K_{ij}(\omega_k^{(1)},\omega_k^{(2)})$ (see also Appendix A) were calculated on a frequency grid of 1 PPM/point. Each interval $[[\alpha_i,\alpha_{i+1}],[\beta_j,\beta_{j+1}]]$ was integrated using a two-point Gaussian quadrature scheme and a 1° resolution, leading to an approximate 0.5° resolution, effectively. To integrate over γ_F the composite trapezium rule was used, with 1024 steps per 360°. To improve the numerical stability all spectra were minimised in size by taking away points outside the spectral range of the DECODER experiment. Furthermore, all basis spectra $K_{i1}(\omega_k^{(1)},\omega_k^{(2)})$, i.e., all which include $\beta_F = 0$, were added and put into the kernel as a single basis spectrum, because close to $\beta_F = 0$ the spectra are virtually independent of α_F.

DOQSY spectra $K_{ij}(\omega_k^{(1)},\omega_k^{(2)})$ were calculated on a frequency grid of 1 PPM/point with 17710 powder orientations per conformation (ϕ,ψ), using the method by Cheng [43]. Each interval $[[\psi_i,\psi_{i+1}],[\phi_j,\phi_{j+1}]]$ was integrated using a two-point Gaussian quadrature scheme and 5° resolution, leading to an approximate 2.5° resolution, effectively. The frequency-domain simulations were corrected a posteriori for finite RF pulse strengths, which was not included in the frequency-space simulations. It was shown using full quantum dynamical simulations that this provides, in approximation, a good description of the excitation profile of the pulse sequence [40]. After adding line broadening, every second point was taken from the spectra to obtain a 2 PPM/point resolution. To increase the stability of the analysis, it was necessary to take away all points around the spectral range of the DOQSY experiment, thus resulting in a non-rectangular frequency grid.

All spectra have been fitted using a regularization algorithm, written in C. Its reliability has been extensively tested using Monte Carlo studies and it has been successfully applied previously to various other Fredholm integral equations of the first kind [7,8,10,13, 44,45]. For DECODER spectra with a 10° grid resolution (91 basis spectra, 33489 spectral points) typical fitting times were about 25 s. DOQSY spectra, with a 15° resolution (290 basis spectra, 9183 spectral points), needed approximately 1 min on a SUN workstation with a single Ultrasparc 2 processor. Note that the fitting time is determined by the character of the distribution due to the iterative nature of the NNLS algorithm. Different distributions may therefore take longer.

3.4. NMR spectra

NMR spectra were obtained at room temperature on a Bruker DMX 400 spectrometer. The DOQSY spectra

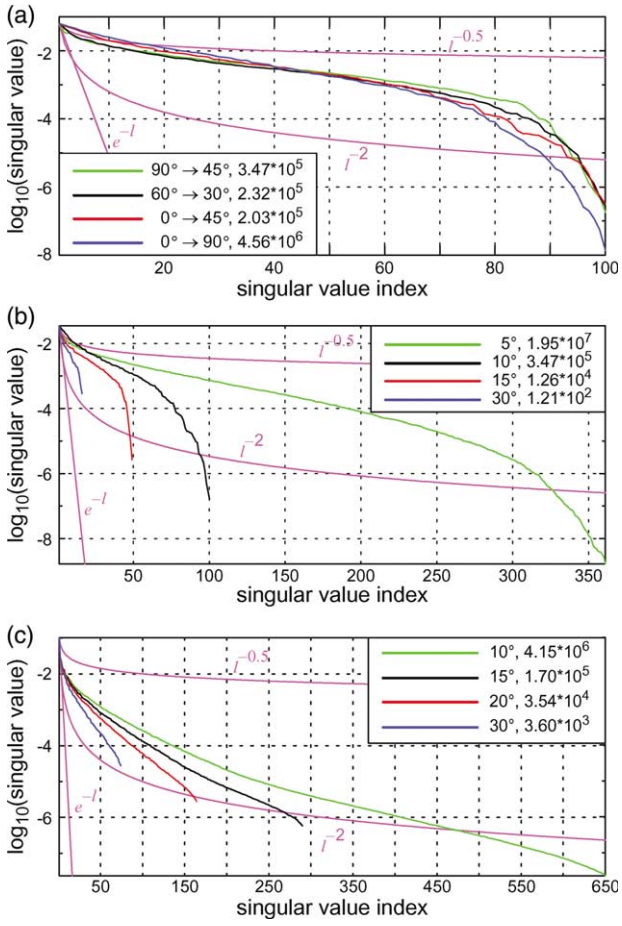

Fig. 4. Singular value plots for (a) four different DECODER experiments and a grid resolution of 10°, (b) a DECODER 90° → 45° experiment and four grid resolutions, and (c) the DOQSY experiment on proteins for four grid resolutions. CSA values as typically found for $^{13}C{=}O$ tensors in proteins, $(\delta_{11}, \delta_{22}, \delta_{33}) = (241, 178, 96)$ PPM, were used.

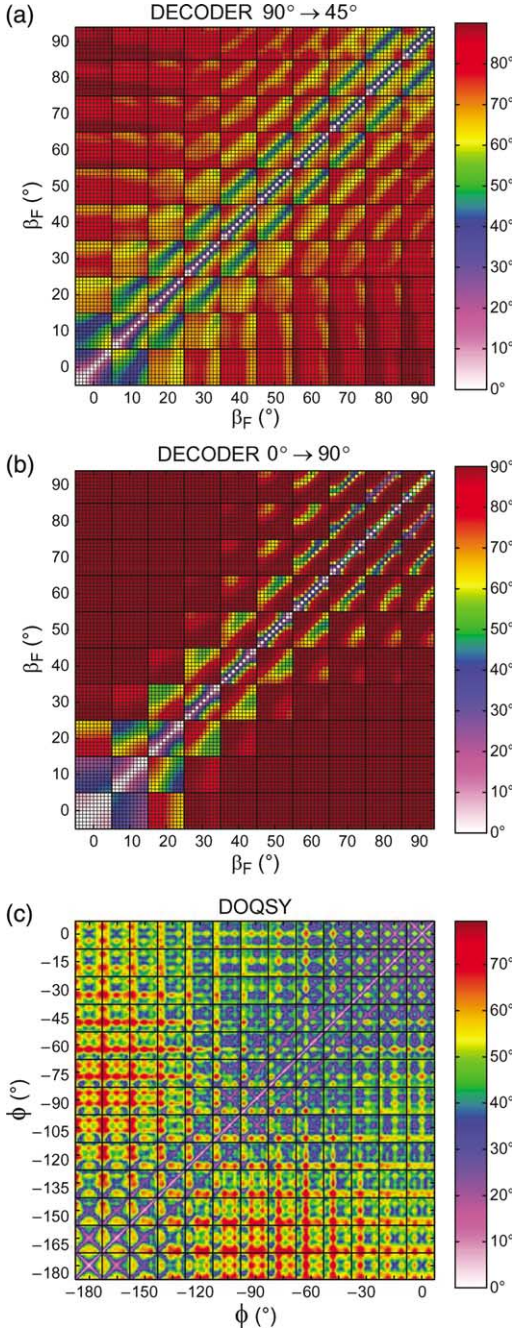

Fig. 5. Angle plots for (a) a DECODER 90° → 45° experiment and (b) a DECODER 0° → 90° experiment both for a grid resolution of 10°. An angle denotes the degree of correlation between two basis spectra as given by Eq. (11). Each large block corresponds to a value for β_F and is subdivided in 10 values for α_F. Panel c shows the angle plot for a DOQSY experiment with a grid resolution of 15°. Due to the inherent symmetry $s(\phi, \psi) = s(-\phi, -\psi)$ the range for ϕ was limited to $[-180°, \ldots, 0°]$. The remaining symmetry-related conformations, visible in the upper right and bottom left squares (angle is 0°), have been left in the angle plots for plotting convenience. Each large block thus corresponds to a value for ϕ and is subdivided in 24 values for ψ.

presented are the same as discussed in [20]. 1D and 2D DECODER spectra of oriented dragline silk were obtained using a homebuilt switched-angle probe setup. Typical RF fields were 50 kHz on both channels. Sample reorientation was done with 2500 °/s and a minimum of 60 ms, including flip, was needed for the stator to stabilize. The total measuring time per spectrum was up to 7 days for samples of approximately 30 mg, using block acquisition to average the effects of long-term instabilities. Spectra were calibrated using a 1D MAS spectrum of adamantane, with peaks centred at 38.6 and 29.5 PPM, respectively [46], and processed using the matNMR [47] processing package.

4. Analysing the experiments

4.1. Kernels

The discretization scheme for both kernels can now be investigated and compared. The DECODER experiment, Eq. (28), represents a set of integral equations due to its flexibility in the choice of the reorientation of the sample, defined by the angles $\beta_L^{(1)}$ and $\beta_L^{(2)}$. Minimum

ill-posedness can, in principle, be achieved by the optimum choice of the reorientation angles, although only small differences are observed in practice (Fig. 4a). Roughly qualified, the experiment is modestly ill-posed, except for the high-order singular values, which show strong ill-posedness. This can be explained, in part, by the discretization scheme. For reasons related to the linear interpolation, the discretization was linear despite the fact that both α_F and β_F are polar angles. For small values of β_F the effect of α_F on the spectrum is very small, creating an artificially high resolution in α_F in the kernel and thus a strong decay in the singular values. The analysis was improved, however, by adding all spectra with $\beta_F = 0$ to a single basis spectrum. This decreased the condition number by almost two orders of magnitude (see also Fig. 10). Better sampling schemes, adapted to the geometry of a sphere, would probably yield a more homogeneous mapping of the parameter domain. For a given choice of the reorientation angles the resolution of the discretization scheme must be optimised. The best value is determined by the Monte Carlo tests described below. Fig. 4b shows the singular values for a DECODER $90° \rightarrow 45°$ experiment for four discretization grid resolutions. A higher resolution dramatically increases the number of basis spectra and the condition number.

Fig. 4c shows the singular value plot for a DOQSY kernel for four grid resolutions. All symmetry-related conformations were removed from the kernel. The decay is generally faster than for DECODER and the condition number is more than an order of magnitude larger, for a comparable grid resolution. The differences between the DECODER and DOQSY experiments can be observed even more clearly from the corresponding angle plots in Fig. 5. For DECODER the correlation decreases with relative distance of the columns, although typically for values of $\beta_F < 20°$ this does not hold. Clear differences between DECODER experiments with different reorientation angles were also observed and confirm the singular value plots. For example, the $0° \rightarrow 90°$ experiment (Fig. 5a) shows lower correlations for $\beta_F > 20°$ but much higher for $\beta_F < 20°$, compared to the $90° \rightarrow 45°$ experiment (Fig. 5b), hence the higher condition number. The angle plot for DOQSY shows a more pathological behaviour with many minima and maxima, indicating a high degree of correlation between various conformations. From Figs. 4 and 5 it is expected that the analysis of DOQSY data will be considerably more difficult than DECODER data, especially considering that the DECODER experiment was markedly improved by adding all spectra with $\beta_F = 0$ to a single basis spectrum. The analysis of DOQSY on the other hand cannot be improved so easily. By discretizing on the (ϕ, ψ) grid the number of basis spectra, and thus some of the numerical problems, was already been reduced considerably compared to a (α, β, γ) grid. The (ϕ, ψ) space spans only a small subspace of the full (α, β, γ) space. The resulting correlations are thus inherent to the molecular fragment of study, i.e., a dipeptide unit.

4.2. Monte Carlo tests

Before applying the algorithm to experimental data, we investigated its behaviour in order to (1) test the reliability, (2) determine the resolution power, and (3) demonstrate the conclusions from the singular value and angle plots. Very often the resolution of an NMR experiment is judged by the (optical) extent of the difference between spectra simulated with different parameters. Although sufficient for delta function distributions, maybe, this certainly is not a good criterion when fitting a non-trivial distribution function, because the noise level and correlations strongly determine the results. The response of the algorithm to both experiments was tested by fitting simulated spectra and evaluating the output distribution function both numerically and graphically. Input spectra were created by simulating a doubly peaked input distribution and calculating the corresponding spectra. After adding line broadening, noise was added to either frequency or time domain to generate an ideal Gaussian or realistic noise representation. Twenty different noise representations were fitted for each noise level.The mean and standard deviation of the normalised squared difference χ between input and output distribution, defined as

$$\chi = \frac{\sum_i (\text{input}_i - \text{output}_i)^2}{\sum_i \text{input}_i^2}, \tag{31}$$

were used as numerical criteria.

Fig. 6 shows a typical example of this procedure for the DECODER experiment. The input distribution consisted of two Gaussian peaks, with FWHH = 15 points, centred at (30°, 30°) and (60°, 60°), respectively. Gaussian noise was added to the frequency domain of the corresponding spectrum. The spectrum was generated using the kernel such that, without noise, the fitting routine was able to recreate the spectrum perfectly. By decreasing the signal-to-noise ratio (S/N) a steady increase in the squared difference between input and output distribution is expected. This is indeed observed. The algorithm was able to reconstruct the main features of the input distribution even at rather high noise levels (Fig. 6h), albeit with increasing number and intensity of small spurious peaks.

In total 80 different input distributions were fitted for $S/N = 3$ to 300. From this limited number of tests it is concluded that for the CSA values and line broadening used here, values as typically found experimentally, the optimum grid density was 10°. Using a resolution of 5° clearly decreased the reproducibility (using identical

input distributions, data not shown here). With the 10° resolution the reliability was rather homogeneous across the parameter domain. Only in regions with small values β_F an increase was observed, with a maximum at $\beta_F = 10°$ (Fig. 6b). This shows that combining all basis spectra with $\beta_F = 0°$ into a single one has helped to stabilize the analysis, but not solved the problem entirely. If accuracy is needed in this region then a better discretization scheme is necessary. The small spurious peaks observed in the PDF usually do not pose problems for a quantitative interpretation of the results, if the distribution function contains peaks that are wider than the grid resolution. For most practical applications this is probably not unrealistic. Extremely sharp-peaked distributions cannot be analysed accurately using a resolution of only 10° and may lead to strong spurious peaks. In these tests the two peaks from the input distribution were always resolved if the distance between

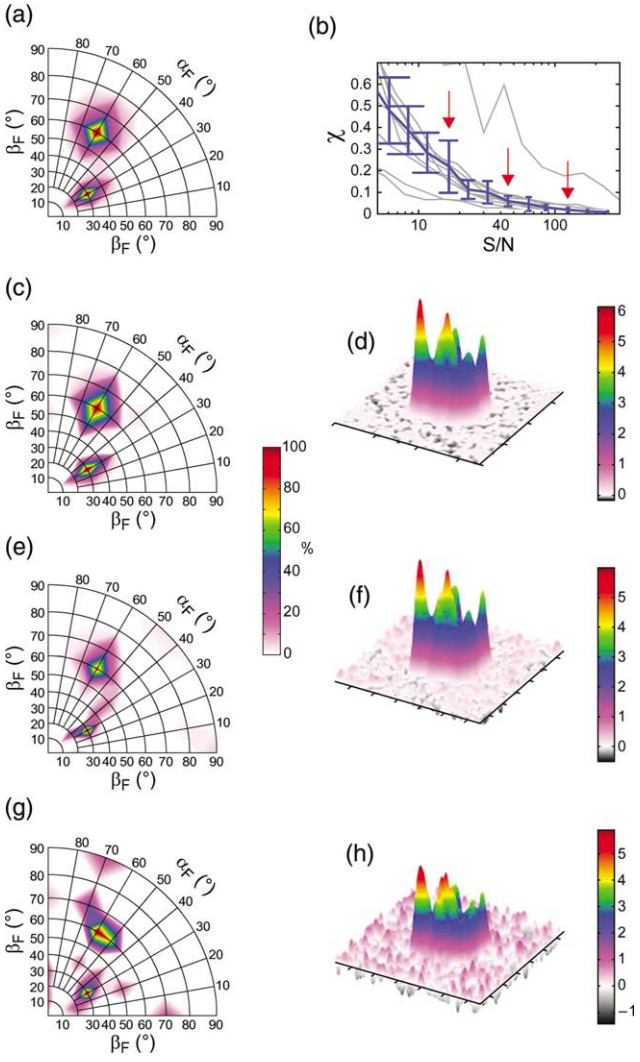

Fig. 6. Typical example of the Monte Carlo tests performed to evaluate the stability of the algorithm towards noise and various input distributions for the DECODER experiment. Panel a shows the input distribution on the same grid density as the kernel, panel b shows the dependency of the squared difference χ, in blue, as a function of the signal-to-noise ratio. The grey lines underneath are the results of nine other input distributions and show the range of results observed in these tests. Note that the grey line with highest squared deviation is a typical example when β_F is centred at 10°, but the two lowest grey lines are obtained by going from 10° to 0°. Panels (c, d), (e, f), and (g, h) show the output distributions and spectra for signal-to-noise ratios of 125, 45, and 17, respectively (marked by red arrows).

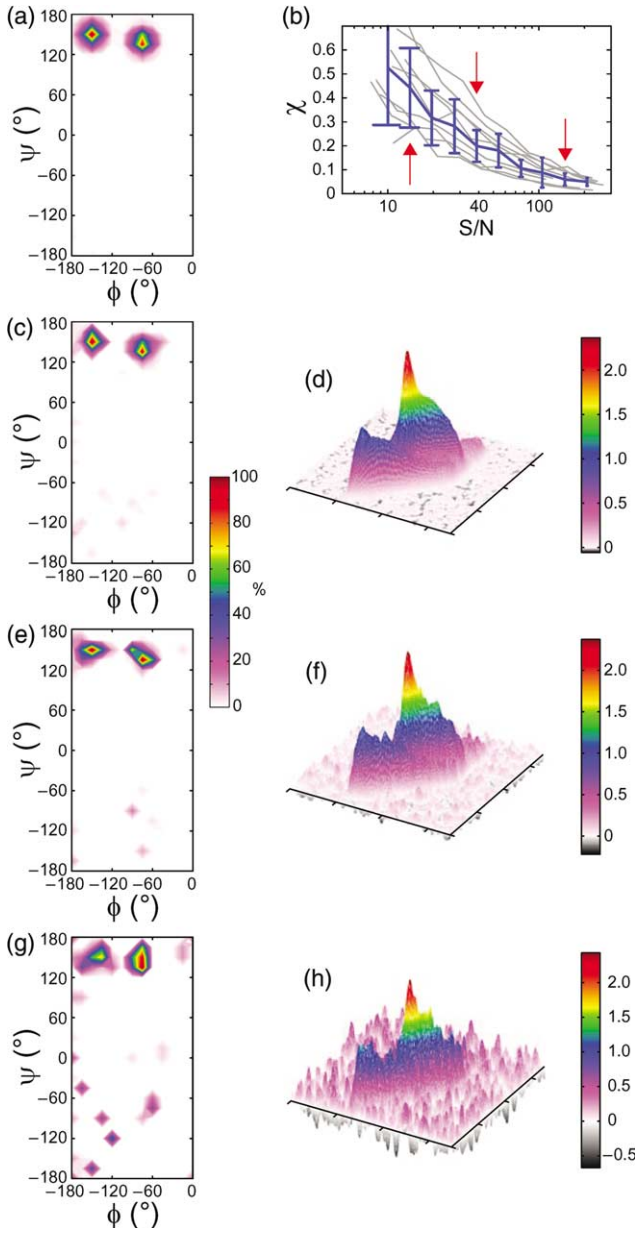

Fig. 7. Typical example of the Monte Carlo tests performed to evaluate the stability of the algorithm towards noise and various input distributions for the DOQSY experiment. Panel a shows the input distribution on the same grid density as the kernel, panel b shows the dependency of the squared difference χ, in blue, as a function of the signal-to-noise ratio. The grey lines underneath are the results of nine other input distributions and show the range of results observed in these tests. Panels (c, d), (e, f), and (g, h) show the output distributions and spectra for signal-to-noise ratios of 148, 39, and 14, respectively (marked by red arrows).

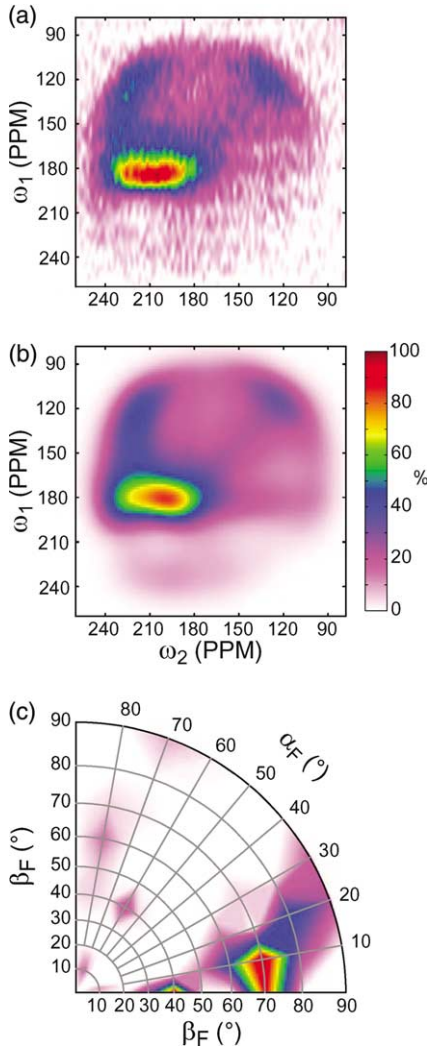

Fig. 8. (a) Experimental 2D DECODER 90° → 45° experiment of oriented spider dragline silk from *Nephila edulis* enriched at [1-^{13}C]alanine, (b) corresponding fit which also included a spectrum from a DECODER 60° → 30° experiment and (c) the corresponding PDF. The signal-to-noise ratio was approximately 30.

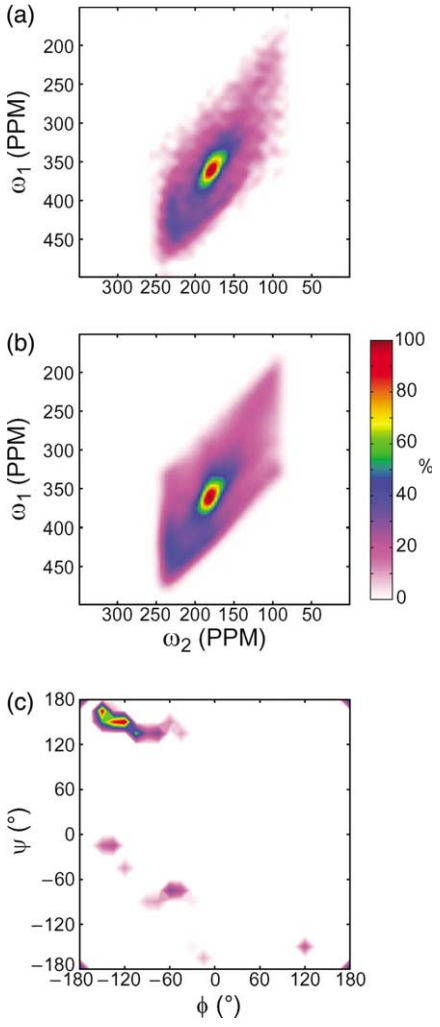

Fig. 9. (a) Experimental 2D DOQSY experiment of unoriented spider dragline silk from *Nephila edulis* enriched at [1-^{13}C]alanine, (b) corresponding fit, and (c) the corresponding PDF. The signal-to-noise ratio was approximately 33. The inherent symmetry of the DOQSY experiment has been broken by including the isotropic $^{13}C_{\alpha}$ chemical shift as measured by a 1D MAS spectrum into the analysis as described in [19].

the centres was at least two grid points, for values of $\beta_F > 20°$. The squared difference increased more than linearly with decreasing S/N.

For the DOQSY experiment, in total 40 different input distributions were tested for $S/N = 8$ to 200. The input distributions consisted of two Gaussian peaks, with FWHH = 25 points. A typical example is shown in Fig. 7. Comparison of Figs. 6a and 7b shows that the normalised squared deviation χ is considerably higher for DOQSY than for DECODER, for similar S/N, which means that the distributions generally showed more spurious peaks and of higher intensity. This is in line with the respective singular-value and angle plots of both experiments. The increased numerical problems allowed a grid resolution of 15° only, which is relatively low. For practical purposes in proteins, where the possible combinations (ϕ, ψ) are limited by the steric hin-

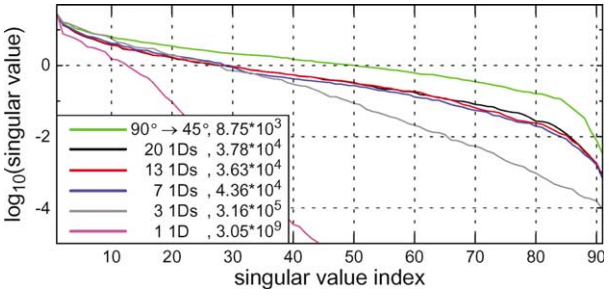

Fig. 10. Direct comparison, using singular-value decay plots, of a 2D DECODER 90° → 45° experiment and various series of 1D spectra taken at different sample orientations. For the 1D series spectra of several orientations between 0° and 90° were used, employing the same CSA values and discretization grid density as for the 2D experiment.

drance in the resulting structures, it is, probably, not realistic to expect peaks in the PDF that are much wider than this. This means that especially for practically

attainable S/N ratios of 20–60 the spurious peaks cannot be avoided, using the current approach. This complicates a quantitative interpretation of the PDF. The algorithm was always able to extract the two peaks at the right position, though. The regions around two conformations show highly specific spectra and can be extracted with higher reliability: the extended conformation $(\pm 180°, \pm 180°)$ and the helical $(\mp 60°, \pm 60°)$. Other regions, e.g., around $(\mp 120°, \pm 120°)$, show extensive correlations with other conformations and may under circumstances be much more difficult to extract faithfully [40].

For samples with comparable isotopic labelling the S/N will in general be lower for DOQSY than for DECODER, due to the reduced efficiency in creating double-quantum coherences. Furthermore, steric energy as a function of the protein geometry leads to relatively sharp distribution functions. The analysis of DOQSY spectra is thus a much bigger challenge, even without systematic errors caused by assumptions made in an experiment (e.g., the orientation of the CSA tensor in the molecular frame, the assumed peptide geometry, the stability of the sample position in a long DECODER experiment, distribution of CSA principal values, etc.).

5. Experimental results

For a fibrous protein like silk, the DECODER and DOQSY experiments allow the correlation of the orientation of secondary-structure elements to the fibre direction. This permits the study of structure–function relationships in materials that are otherwise difficult to study. Fig. 8 shows an experimental DECODER $90° \rightarrow 45°$ experiment of uni-axially oriented silk fibres and the corresponding fitting results. The PDF shows that the alanine C=O groups are strongly oriented towards the fibre direction. From the small value for α_F that was found it follows that the C=O bond is, on average, almost perpendicular towards the fibre. In fact, the single distribution around $(\alpha_F, \beta_F) = (10°, 70°)$, indicates a regular β-sheet conformation oriented with the chain direction along the fibre [20]. The PDF from the corresponding DOQSY spectrum of unoriented silk, shown in Fig. 9 by a single pair of torsion angles, corresponding to a delta-peak in the PDF, is taken to be a confirmation that most Ala–Ala pairs are in a β-sheet conformation. For this we use not only the fact that we observe typical β-sheet torsion angles around $(-150°, 150°)$, but we interpret the intensity at $(\pm 180°, \pm 180°)$ as signal originating from an inter-sheet dipolar coupling. Although we try to exclude these from occurring, using short DQ excitation times to stay in the initial rate regime, such contacts cannot be avoided completely (see also [20,40]). By combining the information from both spectra we can conclude that the β-sheets must be highly oriented towards the silk fibre with the chain direction along the fibre. More details on these results can be found in [20].

Despite the fact, that the resulting PDF from the DOQSY experiment is rather sharpely peaked, an analysis of the DOQSY spectra of Fig. 9 leads to unstatisfactory results (see supporting information to [20] at www.pnas.org).

In the analyses of DECODER and DOQSY spectra from silk we have assumed a single set of CSA tensor values and thus that the CSA tensor is independent of the secondary structure. This assumption was necessary because the exact dependence on secondary structure is not known yet. Fitting with tensor values that deviate from the 'true' values results in spurious peaks even at low noise levels, typically leading to a higher standard deviation and a more irregular mean than observed in Figs. 6b and 7b. The implications of this assumption will be discussed, for DOQSY, in a forthcoming paper [40].

6. 2D DECODER or a series of 1D spectra?

The DOQSY and 2D DECODER experiments were used in this work to provide complementary information on the same silk fibroins. We may, however, think of alternative experiments which may provide similar information. In the following we discuss a simple alternative to the 2D DECODER experiment, namely a series of 1D spectra taken at different orientations of the sample, with respect to the magnetic field.

Fig. 10 shows the singular value decays for kernels consisting of various numbers of 1D spectra and also that of a 2D DECODER $90° \rightarrow 45°$ experiment. The parameter domain was the same for the 1D and 2D spectra and we may therefore compare these results quantitatively. The information content of a 2D DECODER experiment is higher than any number of combined 1D spectra can provide. A minimum of at least seven orientations is required in a series of 1D spectra to obtain a sufficient information content level, although some numerical instabilities were still observed. Using less orientations results in severely increased instability. As shown for both DECODER and DOQSY, the noise level plays a very important role in the reliability of the analysis. Assuming that the total acquisition time for the 1D spectra is equal to that of the 2D DECODER experiment, for our particular experiment the signal to noise is considerably higher in the 1D spectra and the analysis should be more reliable. Numerical tests indeed showed lower and more consistent squared deviations between input and output PDF for a series of 1D spectra (data not shown). Measuring a set of approximately 10 1D spectra thus is a good alternative to the mechanically more-challenging 2D DE-

CODER spectrum, if the measuring time is taken equally long. Note though that for a broad spectrum the phase and baseline corrections can, in our experience, be better separated in 2D spectra than in 1D spectra, which reduces systematic errors.

7. Conclusions

Extracting distribution functions for structural parameters from NMR spectra is challenging and care must be taken in order not to over- or under-interpret the experimental results. Using two practically relevant examples, we have shown that well-known analysis strategies, for solving two-dimensional Fredholm integral equations, can be successfully applied in the context of multi-dimensional NMR. By combining three regulatory tools, optimised discretization, non-negativity, and Tikhonov regularization, we are able to stabilise the inherent numerical problems sufficiently to extract structural information from 2D NMR spectra. By using singular values and correlation between basis spectra, an experiment may be assessed, and possibly improved, such that its information content is optimum. Through numerical testing we can obtain a clear view of the attainable resolution in the analysis, something which is typically disguised, in algorithm-given error bars, by the ill-posed nature of a problem. The algorithm used, shown to be stable and fast in previous applications, works well with the chosen regulatory approach in the context of 2D integral equations and 2D NMR.

The concepts presented here may also, in principle, be applied to any higher-dimensional integral equation and/or nD NMR spectrum, for example by combining DECODER and DOQSY into a single 3D experiment that correlates dipeptide units towards a macroscopic director. The numerical problems rapidly increase with higher dimensionality in parameter space.

The experiments studied in this paper concern static lineshapes of C=O tensors in a semi-crystalline fibrous silk, i.e., the spectra had wide lines, low resolution, and low signal-to-noise ratios. For $S/N = 8$ to 200, with DOQSY the algorithm is in all cases able to reconstruct a doubly peaked input distribution, to a level where the main peaks can clearly be distinguished at the proper positions, and with similar separation level, as the input distribution. Small spurious peaks generally occur in the output distribution, even at low noise levels, which makes the relative intensities of peaks difficult to interpret quantitatively. The DECODER experiment will in practise be less sensitive to such spurious peaks, and thus more reliable, as in most applications the underlying distribution function will be broader than for the DOQSY experiment. Indeed, for similar S/N significantly higher reproducibility was found for DECODER than for DOQSY. Note, however, that the two experiments provide complementary information. Grid resolutions of 10° and 15° were attainable for DECODER and DOQSY, respectively.

We believe that the data evaluation presented here can also be applied to other NMR experiments, in particular also to MAS experiments.

Acknowledgments

This research has been supported by the Swiss National Science Foundation and by the European Science Foundation through the Network: Silk, Properties and Production.

Appendix A

We define a discrete two-dimensional grid $[\alpha_i, \beta_j]$ across the parameter domain with $i = 1, \ldots, (n_\alpha - 1)$ and $j = 1, \ldots, (n_\beta - 1)$. From Eqs. (18) and (19) we then obtain

$$\tilde{s}_k = \sum_{j=1}^{n_\beta - 1} \sum_{i=1}^{n_\alpha - 1} s_{ij}(\omega_k^{(1)}, \omega_k^{(2)}), \tag{A.1}$$

where $s_{ij}(\omega_k^{(1)}, \omega_k^{(2)})$ is the part of the experimental spectrum originating from interval $[[\alpha_i, \alpha_{i+1}], [\beta_j, \beta_{j+1}]]$ and is given by

$$s_{ij}(\omega_k^{(1)}, \omega_k^{(2)}) = \int_{\beta=\beta_j}^{\beta_{j+1}} \int_{\alpha=\alpha_i}^{\alpha_{i+1}} g(\alpha, \beta) K(\omega_k^{(1)}, \omega_k^{(2)}, \alpha, \beta)\, \mathrm{d}\alpha\, \mathrm{d}\beta. \tag{A.2}$$

Using the piecewise approximation of g according to Eqs. (22) and (23) we may rewrite this as

$$s_{ij}(\omega_k^{(1)}, \omega_k^{(2)}) = \tilde{g}_{ij} s_{ij}^{(1)} + \tilde{g}_{i+1,j} s_{ij}^{(2)} + \tilde{g}_{i,j+1} s_{ij}^{(3)} + \tilde{g}_{i+1,j+1} s_{ij}^{(4)} \tag{A.3}$$

with

$$s_{ij}^{(r)} = \int_{\beta=\beta_j}^{\beta_{j+1}} \int_{\alpha=\alpha_i}^{\alpha_{i+1}} c_{ij}^{(r)}(\alpha, \beta) K(\omega_k^{(1)}, \omega_k^{(2)}, \alpha, \beta)\, \mathrm{d}\alpha\, \mathrm{d}\beta, \tag{A.4}$$

$r = 1, \ldots, 4$, $i = 1, \ldots, n_\alpha$, and $j = 1, \ldots, n_\beta$. Note that this step involves going from $(n_\alpha - 1) * (n_\beta - 1)$ coefficients for $\bar{g}(\alpha, \beta)$ to $n_\alpha * n_\beta$ coefficients $\tilde{g}$, without changing the number of independent basis spectra! We can then rewrite Eq. (A.1) as

$$\tilde{s}_k = \sum_{j=1}^{n_\beta - 1} \sum_{i=1}^{n_\alpha - 1} \left[\tilde{g}_{ij} s_{ij}^{(1)} + \tilde{g}_{i+1,j} s_{ij}^{(2)} + \tilde{g}_{i,j+1} s_{ij}^{(3)} + \tilde{g}_{i+1,j+1} s_{ij}^{(4)} \right], \tag{A.5}$$

which can be rewritten as

$$\tilde{s}_k = \sum_{j=1}^{n_\beta-1}\sum_{i=1}^{n_\alpha-1}\tilde{g}_{ij}s_{ij}^{(1)} + \sum_{j=1}^{n_\beta-1}\sum_{i=2}^{n_\alpha}\tilde{g}_{ij}s_{i-1,j}^{(2)} + \sum_{j=2}^{n_\beta} \times \sum_{i=1}^{n_\alpha-1}\tilde{g}_{ij}s_{i,j-1}^{(3)} + \sum_{j=2}^{n_\beta}\sum_{i=2}^{n_\alpha}\tilde{g}_{ij}s_{i-1,j-1}^{(4)}. \quad \text{(A.6)}$$

Eq. (24) is then obtained by collecting the appropriate terms of $s_{ij}^{(r)}$ into $K_{ij}(\omega_k^{(1)},\omega_k^{(2)})$.

The 2D DECODER experiment, is fully described by the parameter ranges $\alpha_F = [0°, 90°]$ and $\beta_F = [0°, 90°]$, due to the fibre symmetry of the silk samples. For this or any other experiment with non-cyclic parameter domains the $K_{ij}(\omega_k^{(1)},\omega_k^{(2)})$ are defined as

$$\begin{aligned}
&K_{1,1}(\omega_k^{(1)},\omega_k^{(2)}) = s_{1,1}^{(1)},\\
&K_{1,n_\beta}(\omega_k^{(1)},\omega_k^{(2)}) = s_{1,n_\beta-1}^{(3)},\\
&K_{n_\alpha,1}(\omega_k^{(1)},\omega_k^{(2)}) = s_{n_\alpha-1,1}^{(2)},\\
&K_{n_\alpha,n_\beta}(\omega_k^{(1)},\omega_k^{(2)}) = s_{n_\alpha-1,n_\beta-1}^{(4)},\\
&K_{i,1}(\omega_k^{(1)},\omega_k^{(2)}) = s_{i,1}^{(1)} + s_{i-1,1}^{(2)}, \quad i = 2,\ldots,(n_\alpha-1),\\
&K_{1,j}(\omega_k^{(1)},\omega_k^{(2)}) = s_{1,j}^{(1)} + s_{1,j-1}^{(3)}, \quad j = 2,\ldots,(n_\beta-1),\\
&K_{n_\alpha,j}(\omega_k^{(1)},\omega_k^{(2)}) = s_{n_\alpha-1,j}^{(2)} + s_{n_\alpha-1,j-1}^{(4)}, \quad j = 2,\ldots,(n_\beta-1),\\
&K_{i,n_\beta}(\omega_k^{(1)},\omega_k^{(2)}) = s_{i,n_\beta-1}^{(3)} + s_{i-1,n_\beta-1}^{(4)}, \quad i = 2,\ldots,(n_\alpha-1),\\
&K_{i,j}(\omega_k^{(1)},\omega_k^{(2)}) = s_{i,j}^{(1)} + s_{i-1,j}^{(2)} + s_{i,j-1}^{(3)} + s_{i-1,j-1}^{(4)},\\
&\quad i = 2,\ldots,(n_\alpha-1),\ \ j = 2,\ldots,(n_\beta-1). \qquad \text{(A.7)}
\end{aligned}$$

The total number of independent coefficients is then $n_\alpha * n_\beta$ with $n_\alpha = 90°/\text{resolution} + 1$ and $n_\beta = 90°/\text{resolution} + 1$, in case of 2D DECODER.

For an experiment with a cyclic parameter domain (e.g., with DOQSY where $\phi = [-180°, 180°]$ and $\psi = [-180°, 180°]$) the $K_{i,j}(\omega_k^{(1)},\omega_k^{(2)})$ are defined as

$$\begin{aligned}
&K_{1,1}\left(\omega_k^{(1)},\omega_k^{(2)}\right) = s_{1,1}^{(1)} + s_{n_\phi-1,1}^{(2)} + s_{1,n_\psi-1}^{(3)} + s_{n_\phi-1,n_\psi-1}^{(4)},\\
&K_{i,1}(\omega_k^{(1)},\omega_k^{(2)}) = s_{i,1}^{(1)} + s_{i-1,1}^{(2)} + s_{i,n_\psi-1}^{(3)} + s_{i-1,n_\psi-1}^{(4)},\\
&\quad i = 2,\ldots,(n_\phi-1),\\
&K_{1,j}(\omega_k^{(1)},\omega_k^{(2)}) = s_{1,j}^{(1)} + s_{n_\alpha-1,j}^{(2)} + s_{1,j-1}^{(3)} + s_{n_\alpha-1,j-1}^{(4)},\\
&\quad j = 2,\ldots,(n_\psi-1),\\
&K_{i,j}(\omega_k^{(1)},\omega_k^{(2)}) = s_{i,j}^{(1)} + s_{1-1,j}^{(2)} + s_{i,j-1}^{(3)} + s_{i-1,j-1}^{(4)},\\
&\quad i = 2,\ldots,(n_\phi-1),\ \ j = 2,\ldots,(n_\psi-1). \qquad \text{(A.8)}
\end{aligned}$$

The total number of independent coefficients is then $(n_\phi-1) * (n_\psi-1)$ with $n_\phi = 360°/\text{resolution} + 1$ and $n_\psi = 360°/\text{resolution} + 1$, in case of DOQSY.

These derivations do not include the removal of symmetry-related basis spectra within the given parameter domains, which may lead to a further reduction of coefficients. Note, also for the numerical integration of $s_{ij}^{(r)}$ within each interval $[[\alpha_i,\alpha_{i+1}],[\beta_j,\beta_{j+1}]]$ an appropriate number of steps must be selected. Besides calculation time considerations regulatory effects may become relevant in some cases. Unreasonable resolution within the intervals will invariably destabilize the problem more and even the choice of the integration scheme can influence the results.

References

[1] P.C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems, SIAM, Philadelphia, 1998.
[2] W. Menke, Geophysical Data Analysis: Discrete Inverse Theory, Academic Press, San Diego, 1989.
[3] I.J.D. Craig, J.C. Brown, Inverse Problems in Astronomy, Adam Hilger, Bristol, UK, 1986.
[4] R. Barakat, G. Newsam, Radio Sci. 19 (1984) 1041–1056.
[5] E. Sternin, M. Bloom, A.L. Mackay, J. Magn. Res. 55 (1983) 274–282.
[6] K.P. Whittall, E. Sternin, M. Bloom, A.L. Mackay, J. Magn. Res. 84 (1989) 64–71.
[7] H. Schäfer, B. Maedler, F. Volke, J. Magn. Res. A 116 (1995) 145–149.
[8] H. Schäfer, B. Madler, E. Sternin, Biophys. J. 74 (1998) 1007–1014.
[9] E. Sternin, H. Schäfer, I.V. Polozov, K. Hawrisch, J. Magn. Res. 149 (2001) 110–113.
[10] H. Schäfer, R. Stannarius, J. Magn. Res. B 106 (1995) 14–23.
[11] J. Winterhalter, D. Maier, D.A. Grabowski, J. Honerkamp, S. Müller, C. Schmidt, J. Chem. Phys. 110 (1999) 4035–4046.
[12] F.G. Vogt, D.J. Aurentz, K.T. Müller, Mol. Phys. 95 (1998) 907–919.
[13] H. Schäfer, H. Bauch, Phys. Lett. A 199 (1995) 93–98.
[14] D.A. Grabowski, J. Honerkamp, J. Chem. Phys. 96 (1992) 2629–2632.
[15] J.W. Zwanziger, Solid State NMR 3 (1994) 219–229.
[16] F. Angeli, T. Charpentier, P. Faucon, J.-C. Petit, J. Phys. Chem. B 103 (1999) 10356–10364.
[17] M. Utz, J. Chem. Phys. 109 (1998) 6110–6124.
[18] Y.-Q. Song, L. Venkataramanan, M.D. Hürlimann, M. Flaum, P. Frulla, C. Straley, J. Magn. Res. 154 (2002) 261–268.
[19] J.D. van Beek, L. Beaulieu, H. Schäfer, M. Demura, T. Asakura, B.H. Meier, Nature 405 (2000) 1077–1079.
[20] J.D. van Beek, S. Hess, F. Vollrath, B.H. Meier, Proc. Natl. Acad. Soc. USA 99 (2002) 10266–10271.
[21] J. Hadamard, Lectures of the Cauchy Problem in Linear Partial Differential Equations, Yale University Press, New Haven, 1923.
[22] C.L. Lawson, R.J. Hanson, Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs, NJ, 1974.
[23] A.N. Tikhonov, V.Y. Arsenin, Solutions of Ill-Posed Problems, Wiley, New York, 1977.
[24] H.W. Engl, M. Hanke, A. Neubacher, Regularization of Inverse Problems, Kluwer Academic Publishers, Dordrecht, 1996.
[25] J. Honerkamp, J. Weese, Cont. Mech. Thermodyn. 2 (1990) 17–30.
[26] J. Weese, Comput. Phys. Commun. 69 (1992) 99–111.
[27] P.M. Henrichs, Macromolecules 20 (1987) 2099–2112.
[28] K. Schmidt-Rohr, M. Hehn, D. Schaefer, H.W. Spiess, J. Chem. Phys. 97 (1992) 2247–2262.
[29] K. Schmidt-Rohr, B.F. Chmelka, H.W. Spiess, Macromolecules 26 (1993) 2282–2296.
[30] R.H. Lewis, H.W. Long, K. Schmidtrohr, H.W. Spiess, J. Magn. Reson. A 115 (1995) 26–34.
[31] K. Schmidt-Rohr, Macromolecules 29 (1996) 3975–3981.
[32] K. Schmidt-Rohr, W. Hu, N. Zumbulyadis, Science 280 (1998) 714–717.

[33] G.H. Golub, C. Rheinsch, Numer. Math. 14 (1970) 403.
[34] A. Kielbasinski, H. Schwetlick, Numerische Lineare Algebra, VEB Deutscher Verlag der Wissenschaften, Berlin, 1988.
[35] D.M. Brink, G.R. Satchler, Angular Momentum, third ed., Clarendon Press, Oxford, 1993.
[36] R.E. Stark, L.W. Jelinski, D.J. Ruben, D.A. Torchia, R.G. Griffin, J. Magn. Reson. 55 (1983) 266–273.
[37] T.G. Oas, C.J. Hartzell, F.W. Dahlquist, G.P. Drobny, J. Am. Chem. Soc. 109 (1987) 5962–5966.
[38] T.G. Oas, C.J. Hartzell, T.J. McMahon, G.P. Drobny, F.W. Dahlquist, J. Am. Chem. Soc. 109 (1987) 5956–5962.
[39] Q. Teng, M. Iqbal, T.A. Cross, J. Am. Chem. Soc. 114 (1992) 5312–5321.
[40] J.D. van Beek, B.H. Meier, manuscript in preparation.
[41] Protein Data Bank, www.rcsb.org.
[42] S. Smith, T. Levante, B.H. Meier, R.R. Ernst, J. Magn. Reson. A 106 (1994) 75–105.
[43] V.B. Cheng, H.H. Suzakawa Jr., M. Wolfsberg, J. Chem. Phys. 59 (1973) 3992–3999.
[44] H. Schäfer, E. Sternin, R. Stannarius, M. Arndt, F. Kremer, Phys. Rev. Lett. 76 (1996) 2177–2180.
[45] P. Bloss, A.S. DeReggi, H. Schäfer, Phys. Rev. B 62 (2000) 8517–8530.
[46] W.L. Earl, D.L. VanDerHart, J. Magn. Res. 48 (1982) 35–54.
[47] MatNMR is a toolbox for processing NMR/EPR data under MATLAB and can be downloaded freely at http://www.nmr.ethz.ch/matnmr.
[48] V.A. Mozorov, Methods for Solving Incorrectly Posed Problems, Springer, New York, 1984.