Open access • Journal Article • DOI:10.1002/ANIE.201707316

Because the Light is Better Here: Correlation-Time Analysis by NMR Spectroscopy

Albert A. Smith, Matthias Ernst, Beat H. Meier

Institutions: ETH Zurich

Published on: 23 Oct 2017 - Angewandte Chemie (Wiley)

Topics: Correlation function (statistical mechanics), Relaxation (physics) and Exponential function

Because the Light is Better Here: Correlation-Time Analysis by NMR Spectroscopy

Originally published in:
Angewandte Chemie. International Edition 56(44), https://doi.org/10.1002/anie.201707316

Funding acknowledgement:
159707 - NMR studies in the Solid State (SNF)
146757 - NMR studies in the Solid State (SNF)

Because the light is better here: correlation-time analysis by NMR

Albert A. Smith,[a] Matthias Ernst,*[a] Beat H. Meier*[a]

Physical Chemistry, ETH Zurich, Vladimir-Prelog-Weg 2, 8093 Zurich, Switzerland
E-mail: maer@ethz.ch, beme@ethz.ch

Abstract: Relaxation data in NMR is often used for dynamics analysis, by modeling motion in the sample with a correlation function consisting of one or more decaying exponential terms, each described by an order parameter, and correlation time. This method has its origins in the Lipari-Szabo model-free approach, which originally considered overall tumbling plus one internal motion and was later expanded to several internal motions. We consider several of these cases in the solid state, and find that if the real motion is more complex than the assumed model, then model fitting is biased towards correlation times where the relaxation data is most sensitive. This leads to unexpected distortions in the resulting dynamics description. We propose using dynamics detectors, which each characterize a range of correlation times, and can be used to give an analysis of protein motion without assuming a specific model of the correlation function.

The story starts: Mullah Nasruddin has lost his ring, and is searching under a street lamp. A passer-by stops to help him find it. After half an hour, he asks “Are you sure you lost your ring here?” “Not here! I lost the ring in the basement of my house.” The passer-by is perplexed. When he asks why they are looking for the ring outside, Nasruddin replies it is because the light is better here!

—Adapted from *The Funniest Tales of Mullah Nasruddin*[1]

Dynamics plays a critical role in understanding the stability and function of proteins.[2] Nuclear magnetic resonance (NMR) can provide such information, since relaxation in NMR is the result of incoherent modulation of NMR interactions via stochastic motion in the sample. Due to the large number of internal degrees of freedom (3N-6 where N is the number of atoms in the protein, typically more than 1000), the interpretation of the data will always have to involve severe approximations.

Fitting NMR relaxation-rate constants allows modeling of the underlying motion. Wangsness-Bloch-Redfield theory details how to calculate relaxation-rate constants from specific motions.[3] However, a complex dependence of each relaxation-rate constant on multiple, orientation dependent correlation functions makes extraction of the original motion nearly impossible. In solid-state NMR, a correlation function with multiple correlation times τ_i, and order parameters S_i^2 is often used to model the dynamic processes especially for proteins:

$$C(t)=\frac{1}{5}\Big[(1-S_1^2)\exp(-t/\tau_1)+S_1^2(1-S_2^2)\exp(-t/\tau_2)+ S_1^2S_2^2(1-S_3^2)\exp(-t/\tau_3)+\ldots+S_1^2S_2^2S_3^2\ldots\Big] \quad (1)$$

which is very similar to and evolved from the Lipari-Szabo model-free approach in solution-state NMR.[4] Note that the coefficients and final term sum to 1 so we can rewrite this more simply as

$$C(t)=\frac{1}{5}\left[S^2+\left(1-S^2\right)\sum_{i=1}A_i\exp\left(-t/\tau_i\right)\right] \quad \text{where } \sum_{i=1}A_i=1,\quad S^2=\prod_{n=1}S_n^2 \quad (2)$$

$1-S^2$ determines the amplitude of the total motion, the A_i specify the contributions from each individual motion with correlation time τ_i. Although the real correlation function may have motion on many time scales, limitations in the amount and precision of experimental data have typically restricted the model of the correlation function to 1-3 exponential terms.

While the original model-free approach was very well justified, the modified solid-state correlation function was proposed more *ad hoc*. This motivated us to reinvestigate its ability to accurately represent protein motion in solids. In particular, we investigate how well a model of the correlation function describes the motion when the model is characterized with fewer correlation times (τ_i) than the real correlation function, which is most probably the case in a typical dynamics analysis of a protein. In a first example, we calculate relaxation-rate constants from a simulated input correlation function having 2-4 correlation times, and then fit those generated rate constants with a model having one fewer τ_i. In this example, both longitudinal, R_1, and transverse, $R_{1\rho}$ rate constants are fitted, and all fits are performed both without and including the total order parameter S^2. The results are shown in Figure 1.

What we find is that, with the exception of Figure 1**B**, the data is well fit despite the model correlation function having fewer correlation times than the input correlation function. If the data is well fit, we expect that the model should be a good representation of the input correlation function. Therefore, each fitted τ_i should be approximately the weighted average of the nearest τ_i from the original correlation, and the amplitude ($(1-S^2)A_i$) of each term in the model should be approximately the sum of the nearest input amplitudes.[4a] What we find, however, is a model having both correlation times far removed from the average of the input correlation times, and fitted amplitudes that are quite different than the sum of the nearest input amplitudes.

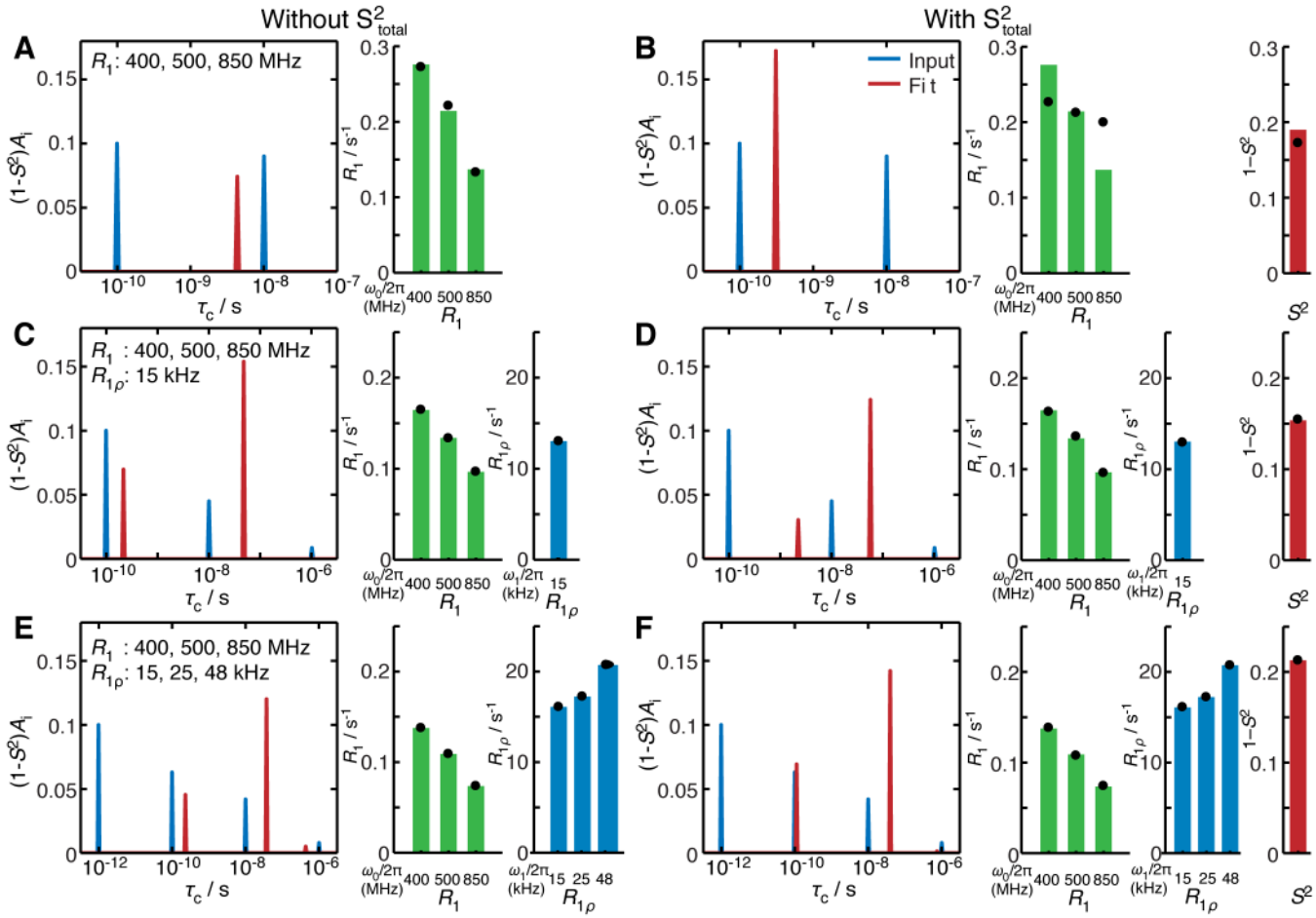

Figure 1. Fits of simulated relaxation data for ^{15}N relaxation. In each plot, several relaxation-rate constants are simulated with 2, 3, or 4 exponentials (τ_i) in the correlation function (each described by τ_i and $(1–S^2)A_i$). They are then fitted with a model having one motion τ_i. less than the input data. Fits are performed without fitting $1–S^2$ in **A**, **C**, and **E** and with fitting $1–S^2$ in **B**, **D**, and **F**. In **A** and **B** correlation functions with 2 τ_i (left plot, blue line) are input, and then fitted with a single τ_i (left plot, red line), where 3 R_1 measurements have been fitted. In **C** and **D**, 3 τ_i are input, and modeled with 2 τ_i, where 3 R_1 and 1 $R_{1\rho}$ measurements are fitted. In **E** and **F,** relaxation resulting from 4 τ_i is simulated and fitted with 3 τ_i, where 3 R_1 and 3 $R_{1\rho}$ are used in the fit. Bar plots for each fit show the calculated rate constants for the input motion, and black dots show the fit. External field for R_1 measurements is indicated below bar plots. Spin-lock strength for $R_{1\rho}$ measurements is also indicated below the bar plots, where the external field is fixed at 850 MHz, and the magic angle spinning (MAS) frequency is fixed at 60 kHz.

To understand where the problems in modeling originate, we consider a simple set of experiments that we can fit using a correlation function having only a single correlation time. We assume a log-uniform distribution of motion (we use 200 A_i in eq. (2) that are all equal with τ_i logarithmically spaced from 10^{-14} to 10^{-3} s), and calculate three $R_{1\rho}$ rate constants. The spin-lock field strength is fixed at 5 kHz, and the spinning frequencies used for probing the relaxation are $0.5\omega_r^0$, ω_r^0, and $1.5\omega_r^0$ where we vary ω_r^0. The values of the three calculated $R_{1\rho}$ rate constants are plotted against ω_r^0 in Figure 2**A** (lines), and are fitted (circles) with a motional model which has only a single correlation time τ_c. The results of the obtained correlation time and order parameter are plotted in Fig. 2**B** and **C** respectively. Although the input motion is always the same, the extracted motion with a single correlation time varies, and is a near linear function of ω_r^0, where $\tau_c \approx 1/\omega_r^0$. If we had generated the relaxation-rate constants using a correlation function with only a single correlation time, we would have always correctly extracted that value. However, since the fit model contains fewer correlation times than the input correlation function, the modeling fails. Then, if a constant distribution function is input, we find that we always extract τ_c approximately where our $R_{1\rho}$ experiments are most sensitive. *One finds motion where the "light" is better*, that is, where the experiments are most sensitive. If the motion is not uniform, but nonetheless given by a correlation function more complicated than the model, then the fitted correlation times may deviate from where experiments are most sensitive but will still be biased. Note the behavior demonstrated here for $R_{1\rho}$ can also be demonstrated for R_1 (see SI, Fig. S2). Since R_1 and $R_{1\rho}$ are largely independent and are mainly sensitive to correlation times at different time scales, a simultaneous fit of such R_1 and $R_{1\rho}$ data with a two time scale model will also result in correlation times where the two experiments are most sensitive (see SI, Fig. S3). Again, *one finds motion where the "light" is better*. Adding an order parameter to the relaxation data makes the fit more complex as will be discussed below.

Based only on the experimental measurement conditions (see SI for details), it is possible to determine towards which τ_c an experimental $R_{1\rho}$ data set is biased. We do so for several dynamics studies, and compare the most sensitive correlation times to the reported τ_c. The biasing/reported τ_c for a number of studies are: HET-s(218-289) ^{15}N 17 μs/19 μs, HET-s(218-289) ^{13}Cα 11 μs/7 μs,[5] Ubiquitin (MPD cryst.) 2.1 μs/1.5 μs (median value),[6] Ubiquitin (PEG cryst.) 26 μs/20 μs (SI Fig. 1).[7] The good agreement between these numbers suggests that the real motion is too complex to be described with a mono-exponential correlation function, so that the results bias towards where the experiments are sensitive.

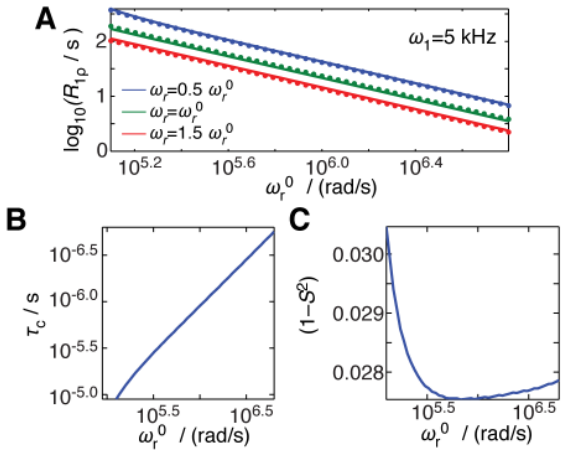

Figure 2. Fitting behavior as a function of experimental settings. A uniform distribution of motion is used to calculate 3 $R_{1\rho}$ rate constants at different spinning frequencies ($0.5\omega_r^0$, ω_r^0, and $1.5\omega_r^0$), with the spin-lock field strength fixed at $\omega_1/2\pi$=5 kHz. $\omega_r^0/2\pi$ is swept from 1 kHz to 1 MHz, where the calculated rate constants are shown in **A** (lines). The rate constants are well-fitted with a mono-exponential correlation function (**A**, circles), with the fitted correlation time (τ_c) shown in **B**, and amplitude, (1–S^2), shown in **C**.

Fitting of only one type of data to a mono-exponential correlation function represents a simple dynamics analysis of solid-state NMR data. Many dynamics studies involve fitting multiple longitudinal (R_1), transverse relaxation-rate constants ($R_{1\rho}$, R_2, etc.), and REDOR-derived order parameters, S^2. This allows a correlation function with two correlation times to be fitted (three, if multiple $R_{1\rho}$ are acquired[8]). Such studies report similar sets of correlation times, typically near $10^{-10.5}$ s and $10^{-7.5}$ s (see Table 1). Figure 3**A** plots R_1 at several typical fields used for such dynamics analyses (lines), where we see that R_1 rate constants are most sensitive for $\tau_c \sim 10^{-8.5}$ s. The same behavior is also seen in generated data sets as shown in the SI (see SI Figs. 4 and 5). Then, the reported correlation times are clearly not near where "the light is better", although the similarity of the correlation times demands further investigation.

Table 1. Median fast (τ_f) and slow (τ_s) correlation times for various proteins studied with solid-state NMR

	SH3[9]	Ubiquitin[10]	GB1[11]	HET-s[5]
τ_s	$10^{-7.6}$ s	$10^{-7.3}$ s	$10^{-6.4}$ s	$10^{-7.4}$ s
τ_f	$10^{-10.7}$ s	$10^{-10.4}$ s	$10^{-10.5}$ s	$10^{-10.7}$ s

In fact, we find an interesting property of these correlation times. If we take a correlation function with the two correlation times fixed such that

$$C(t)=\frac{2}{5}\left((1-S^2)\left(A_1\exp(-t/10^{-10.5}\,\mathrm{s})+A_2\exp(-t/10^{-7.5}\,\mathrm{s})\right)+S^2\right), \quad (3)$$

we find that we obtain a good fit of the three R_1 rate constants at *all* correlation times, by varying $(1–S^2)A_1$ and $(1–S^2)A_2$ (Figure 3**A).** This is reminiscent of an analysis of solution-state NMR relaxation data using fixed correlation times proposed by LeMaster.[12] The possibility to fit three R_1 rate constants with two fixed correlation times also implies that one can fit any real motion with multiple correlation times using a model that fixes τ_i = $10^{-7.5}$ s and $10^{-10.5}$ s. This holds relatively well for any pair of correlation times with one shorter than 10^{-10} s and one longer than 10^{-8} s (see SI for further discussion). We also note that the values of the $(1–S^2)A_i$ are actually highest near $10^{-8.5}$ s, where the R_1 are also most sensitive. So, although the fixed correlation times are not where "the light is better", we actually find the greatest contributions to the amplitudes of the correlation function from the most sensitive correlation times, with the unfortunate side effect that, unlike in previous examples, our fitted τ_i are no longer indicative of where we are sensitive.

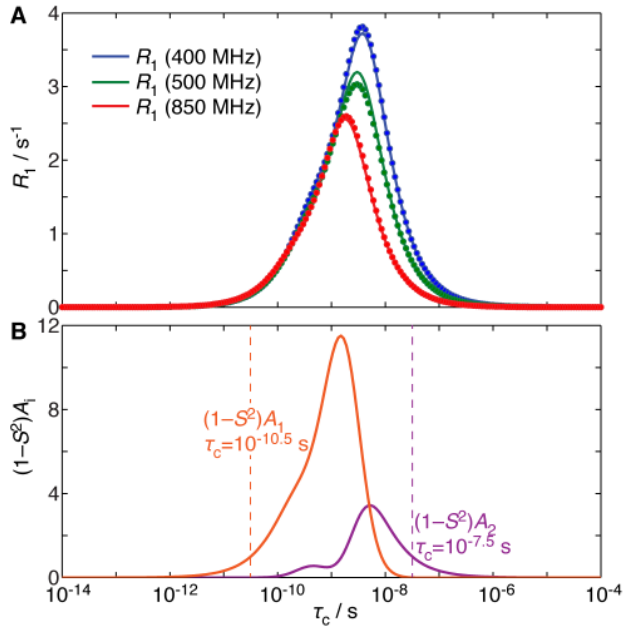

Figure 3. R_1 rate constants fitted with the correlation function given in eq. (3). Lines in **A** show R_1 rate constants for monoexponential correlation functions, as a function of correlation time, τ_c, where $(1–S^2)$=1 (at external fields of 400, 500, and 850 MHz). Circles show fits at each τ_c to eq. (3). **B** shows the weighting of the two terms in eq. (3), with the positions of the τ_i marked as dotted lines.

Although these correlation times can fit any set of high-field R_1 data, the order parameter, S^2, and transverse relaxation-rate constants must also be fitted. Since these analyses do not actually fix the τ_i, one may simultaneously decrease the short correlation and increase $(1-S^2)A_1$ (or vice-versa), to fit S^2 while leaving the fit of R_1 nearly unchanged. Similarly, one may increase the long correlation time and increase $(1–S^2)A_2$ to fit the transverse relaxation, also leaving R_1 unchanged (details in SI). Then, one can fit motion resulting from many correlation functions to a two-timescale correlation function, when combining several R_1 rate constants with S^2 and a single transverse relaxation-rate constant. If the dynamics is actually well described by a model with two τ_i, then the correct τ_i can be extracted, but if the internal motion is more complex, then the fit may bias towards these two "universal" correlation times. The high frequency of fitted correlation times around τ_c=$10^{-7.5}$ and $10^{-10.5}$ s suggests either they have some physical significance or, maybe more likely, these values are an artifact of fitting a complex dynamic process with an oversimplified model.

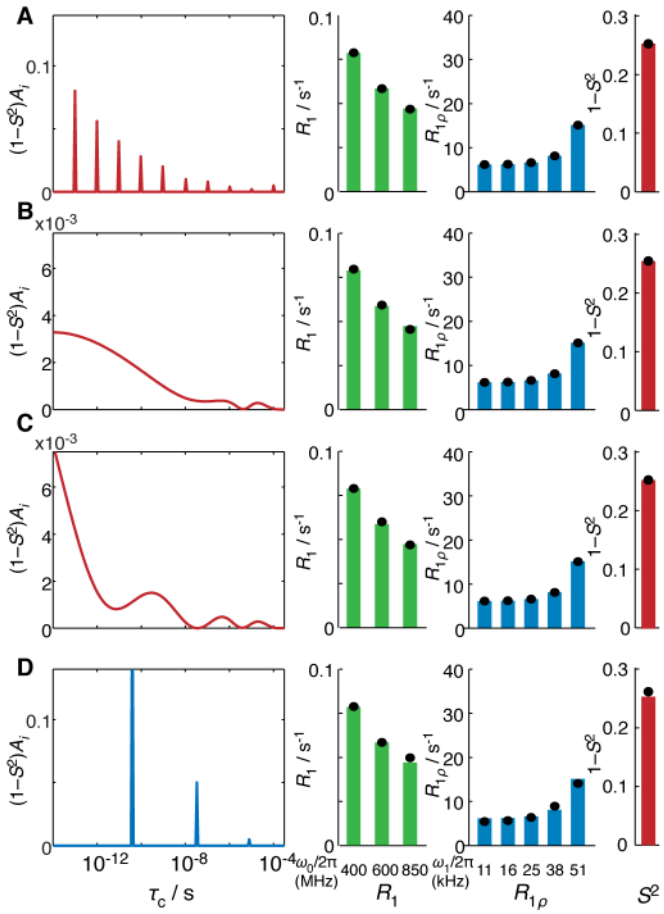

Figure 4. Different distributions of motion with the same relaxation data. Four distributions of motion are plotted as $(1–S^2)A_i$ vs τ_c. Distributions have been chosen in **A**-**C** so that resulting dynamics data (circles, right) yields exactly the same values in each case (bars, right). **D** is restricted to having three non-zero amplitudes, so does not exactly reproduce the data set.

We suspect that real protein motion is not well described by a correlation function having only two to three correlation times. However, if the real motion can have an arbitrary number of correlation times, then different distributions of correlation times will lead to identical relaxation data. This is shown in Figure 4**A**-**C**, where three different distributions of motion produce almost exactly the same set of relaxation data (5 $R_{1\rho}$, 3 R_1, 1 S^2). Figure 4**D** fits a motional model having three correlation times to the relaxation data, yielding a good, although not perfect fit.

When the real motional process is more complex than the model, the model biases towards correlation times that are more sensitive - where "the light is better". Furthermore, when the motion is complex, there is necessarily ambiguity in the data from which we wish to extract dynamical information. Therefore, we introduce the concept of a dynamics *detector,* which contains information only where the light is. A detector would report on the average or total amplitude of motion for a range of correlation times, as opposed to returning an exact amplitude at a specific correlation time – therefore not requiring a specific model of the correlation function. Then, one could describe protein dynamics using multiple detectors that would each characterize protein motion on a different timescale. One would furthermore only have detectors where the dynamics experiments are sensitive.

We then define a detector response as follows, where the response can be calculated from dynamics data:

$$\rho^{(\theta,S)} = (1-S^2)\int_0^\infty \theta(\tau_c)\rho(\tau_c)d\tau_c \qquad (4)$$

$(1–S^2)\theta(\tau_c)$ describes the distribution of motion in the protein- so that $(1–S^2)\theta(\tau_c)$ gives the contribution of each τ_c to the total motion. The detector takes the sum of motion at each τ_c, multiplied by the sensitivity to that correlation time, defined by $\rho(\tau_c)$ to yield the detector response ($\rho(\tau_c)$ is known). A relaxation-rate constant is by this definition a detector, where the sensitivity of the rate constant to different correlation times determines how the rate constant responds to a distribution of motion (the traces in Figure 3**A** are sensitivities, $\rho(\tau_c)$, of the R_1).

Ideally one would like to have a narrower sensitivity than provided by the R_1 rate constants, to say more precisely what range of correlation times is contained in the motion. Then, a detector can be constructed from a linear combination of relaxation data, for example multiple R_1 rate constants may be combined to obtain multiple detectors with narrower ranges (see SI Fig. 6). Previous work along these lines exist: spectral density mapping is performed by taking linear combinations of relaxation data, and yields sensitivities defined by the spectral densities at specific frequencies, $J(\omega,\tau_c)$.[13] Also, LeMaster created detectors by fixing multiple τ_i to carefully chosen values[12] (also see [14]). However, these cases are used for specific applications, requiring

specific sets of measurements. A generalized method for designing detectors, and investigation of properties of those detectors for a variety of data sets, will be presented in a separate report.

In conclusion, NMR dynamics data should be analyzed with great care to yield a picture of protein dynamics that is not biased by selection of a particular correlation function. Based on initial detector analysis, and resulting trends throughout a protein, it may make sense to then introduce a more specific model. However, as we have seen here, we should be aware that modeling can bias our results towards where “the light is better”, and should be suspicious of models describing motion where there is little light.

Acknowledgements

This work was supported by the ETH Zurich and the Swiss National Science Foundation (Grants 200020_159707 and 200020_146757.

Keywords: NMR • dynamics • relaxation • model-free • correlation times

[1] C. Sawhney, *The Funniest Tales of Mullah Nasruddin*, Unicorn Books, New Delhi, **2010**.
[2] K. Henzler-Wildman, D. Kern, *Nature* **2007**, *450*, 964-972.
[3] a) R. K. Wangsness, F. Bloch, *Phys Rev* **1953**, *89*, 728-739; b) A. G. Redfield, *Ibm J Res Dev* **1957**, *1*, 19-31.
[4] a) G. Lipari, A. Szabo, *J. Am. Chem. Soc.* **1982**, *104*, 4546-4559; b) G. M. Clore, A. Szabo, A. Bax, L. E. Kay, P. C. Driscoll, A. M. Gronenborn, *J. Am. Chem. Soc.* **1990**, *112*, 4989-4991.
[5] A. A. Smith, E. Testori, R. Cadalbert, B. H. Meier, M. Ernst, *J. Biomol. NMR* **2016**, *65*, 171-191.
[6] N. A. Lakomek, S. Penzel, A. Lends, R. Cadalbert, M. Ernst, B. H. Meier, *Chemistry* **2017**.
[7] V. Kurauskas, S. Izmailov, O. N. Rogacheva, A. Hessel, I. Ayala, J. Woodhouse, A. Shilova, Y. Xue, T. Yuwen, N. Coquelle, J.-P. Colletier, N. R. Skrynnikov, P. Schanda, *bioRxiv* **2017**, 1-34.
[8] T. Zinkevich, V. Chevelkov, B. Reif, K. Saalwächter, A. Krushelnitsky, *J. Biomol. NMR* **2013**, *57*, 219-235.
[9] V. Chevelkov, U. Fink, B. Reif, *J. Biomol. NMR* **2009**, *45*, 197-206.
[10] a) J. D. Haller, P. Schanda, *J. Biomol. NMR* **2013**, *57*, 263-280; b) P. Schanda, B. H. Meier, M. Ernst, *J. Am. Chem. Soc.* **2010**, *132*, 15957-15967.
[11] J. M. Lamley, M. J. Lougher, H. J. Sass, M. Rogowski, S. Grzesiek, J. R. Lewandowski, *Phys. Chem. Chem. Phys.* **2015**, *17*, 21997-22008.
[12] D. M. LeMaster, *J. Biomol. NMR* **1995**, *6*, 366-374.
[13] J. Peng, G. Wagner, *J. Magn. Res.* **1992**, *98*, 308-332.
[14] S. N. Khan, C. Charlier, R. Augustyniak, N. Salvi, V. Dejean, G. Bodenhausen, O. Lequin, P. Pelupessy, F. Ferrage, *Biophys J* **2015**, *109*, 988-999.