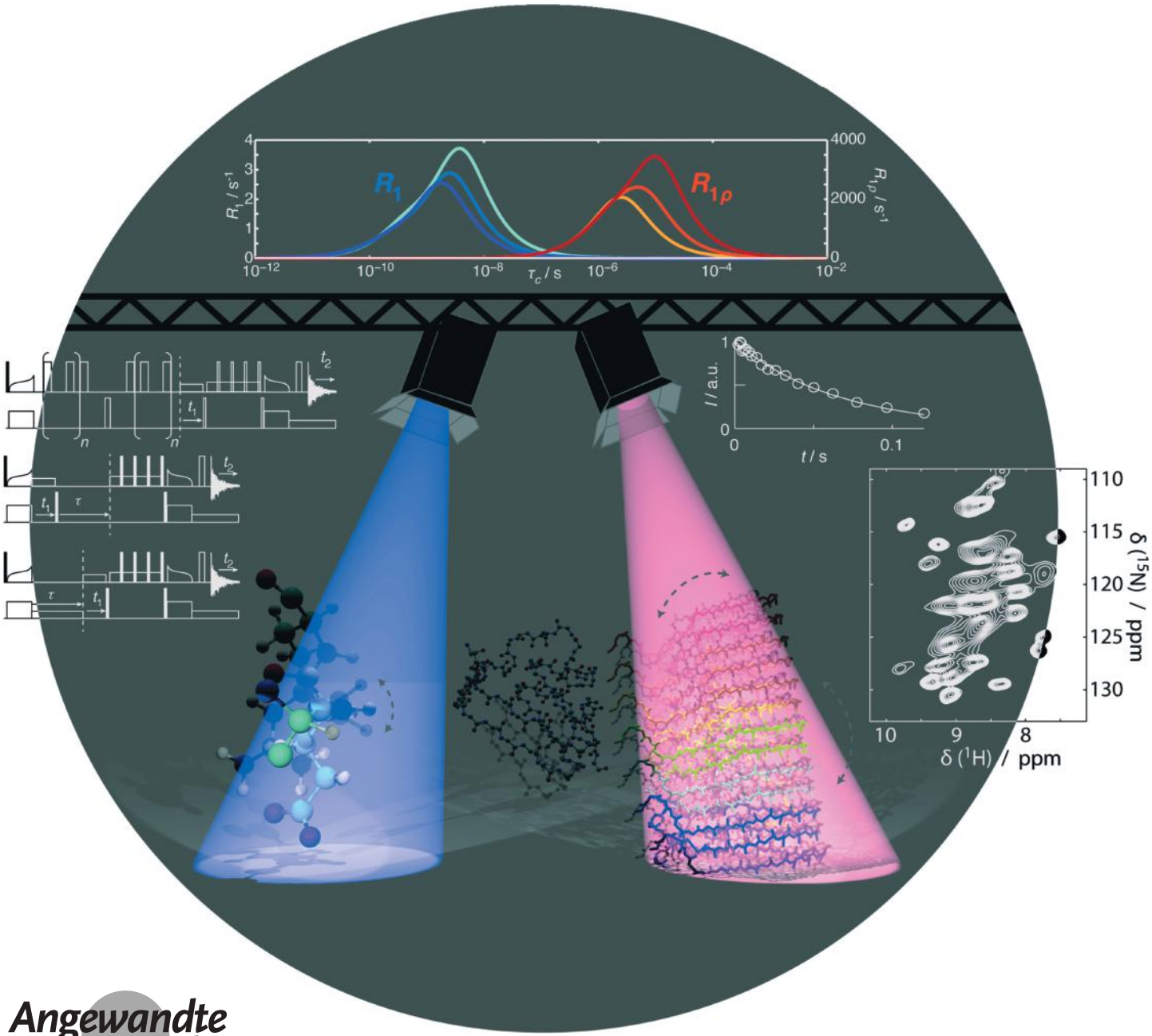
VIP Molecular Motion Very Important Paper

Deutsche Ausgabe: DOI: 10.1002/ange.201707316
Internationale Ausgabe: DOI: 10.1002/anie.201707316

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

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



Abstract: *Relaxation data in NMR spectra are 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 a 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. Considering several of these cases in the solid state it is found that if the real motion is more complex than the assumed model, model fitting is biased towards correlation times where the relaxation data are most sensitive. This leads to unexpected distortions in the resulting dynamics description. Therefore dynamics detectors should be used, which characterize different ranges of correlation times and can help in the 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 play a critical role in understanding the stability and function of proteins.[2] NMR studies can provide such information, since relaxation here is the result of incoherent modulation of NMR interactions through 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 >1000), interpretation of the data must involve severe approximations.

NMR relaxation-rate constants allow modeling of the underlying motion. The Wangsness–Bloch–Redfield theory details how to calculate relaxation-rate constants from specific motions,[3] but dependence of each relaxation-rate constant on multiple, orientation-dependent correlation functions makes extraction of the original motion nearly impossible. In solid-state NMR spectroscopy, a correlation function with multiple correlation times τ_i and order parameters S_i^2 [Eq. (1)] is often used to model dynamic processes

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

especially for proteins, which is similar to and evolved from the Lipari–Szabo model-free approach in solution-state NMR spectroscopy.[4] As the coefficients and final term sum to one, Equation (1) can be rewritten more simply as Equation 2.

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

$1-S^2$ determines the total motion, and A_i stands for the contribution of an individual motion with correlation time τ_i. Although the real correlation function may contain motion on many time scales, limitations in the amount and precision of experimental data typically restrict the model of the correlation function to 1–3 exponential terms.

While the original model-free approach was 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. First, we investigated how well a model of the correlation function describes the motion when the model is characterized by fewer correlation times τ_i than the real correlation function, which is probably the case in a typical analysis of protein dynamics. We calculated relaxation-rate constants from a simulated correlation function having 2–4 τ_i and fitted those rate constants with a model having one τ_i less. In this example, both longitudinal (R_1) and transverse ($R_{1\rho}$) rate constants were fitted, and all fits were performed both with and without the total order parameter S^2; the results are summarized in Figure 1.

With the exception of Figure 1B, the data are well fitted despite the model correlation function having fewer correlation times than the input correlation function. If the data are well fitted, 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] However, we found a model having correlation times far from the average of the input correlation times and fitted amplitudes that are quite different from the sum of the nearest input amplitudes. Two-dimensional plots of the error surface near the fit optima are found in Figure 1 of the Supporting Information (SI), where one sees that many of the fits have well-defined minima despite failing to give a good representation of the input correlation function. Adding an additional correlation time to the model did not improve the situation, resulting in either more fit parameters than observables or poorly defined fit minima (SI, Figure 2).

To understand where the problems in modeling originate, we took a simple set of experiments that we could fit using a correlation function having only a single correlation time. We assumed a log-uniform distribution of motion (we used 200 A_i values in Equation (2) that were all equal with τ_i logarithmically spaced from 10^{-14} to 10^{-3} s) and calculated three $R_{1\rho}$ rate constants. The spin-lock field strength was 5 kHz, and the spinning frequencies used for probing the relaxation were $0.5\omega_r{}^0$, $\omega_r{}^0$, and $1.5\omega_r{}^0$ where we varied $\omega_r{}^0$. The values of the three calculated $R_{1\rho}$ rate constants are plotted against $\omega_r{}^0$ in Figure 2A (lines), and were fitted (circles) with a motion model that had only a single correlation time τ_c. The correlation times and order parameters

[*] Dr. A. A. Smith, Prof. Dr. M. Ernst, Prof. Dr. B. H. Meier
Physical Chemistry, ETH Zurich
Vladimir-Prelog-Weg 2, 8093 Zurich (Switzerland)
E-mail: maer@ethz.ch
beme@ethz.ch
Supporting information and the ORCID identification numbers for authors of this article can be found under: https://doi.org/10.1002/anie.201707316.

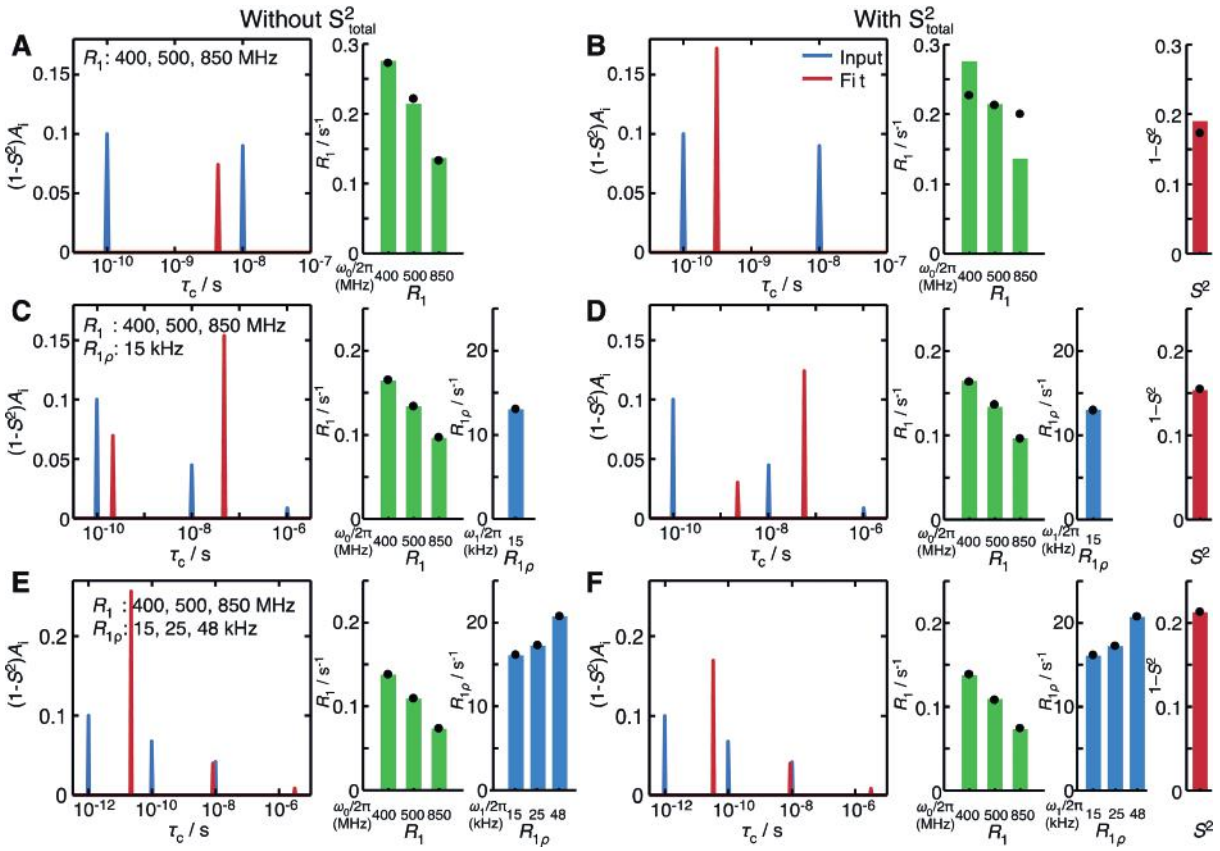


Figure 1. Fits of simulated relaxation data for ^{15}N relaxation. In each plot, several relaxation-rate constants are simulated with 2, 3, or 4 exponentials (τ_i) in the correlation function (each described by τ_i and $(1-S^2)A_i$). They are fitted with a model having one τ_i less than the input data. 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 $2\tau_i$ (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), $3\tau_i$ are input and modeled with $2\tau_i$, where 3 R_1 and 1 $R_{1\rho}$ measurements are fitted. In (E) and (F), $4\tau_i$ are simulated and fitted with $3\tau_i$, where 3 R_1 and 3 $R_{1\rho}$ are used in the fit. Bar plots show the calculated rate constants for the input motion, and black dots show the fit. The external field for R_1 measurements is indicated below the bar plots. The spin-lock strength for $R_{1\rho}$ measurements is also indicated below the bar plots (external field: 850 MHz, magic angle spinning (MAS): 60 kHz).

obtained are plotted in Figure 2B and C, respectively. Although the input motion was always the same, the extracted motion 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 one correlation time, we would have extracted that value (SI, Figure 3). As the fit model contained fewer correlation times than the input correlation function, the modeling failed. With a constant distribution function as input τ_c was always extracted approximately where the $R_{1\rho}$ experiments were most sensitive. *One finds motion where the "light" is better*, that is, where the experiment is most sensitive. If the motion is not uniform, but still described by a correlation function more complex than the model function, the fitted correlation times deviate from where the experiment is most sensitive but are still biased. This behavior can also be demonstrated for R_1 (SI, Figure 4). Since R_1 and $R_{1\rho}$ are largely independent and sensitive to correlation times at different time scales, a simultaneous fit of R_1 and $R_{1\rho}$ data with a two-timescale model will also result in correlation times where the two experiments are most sensitive (SI, Figure 6). Again, *one finds motion where*

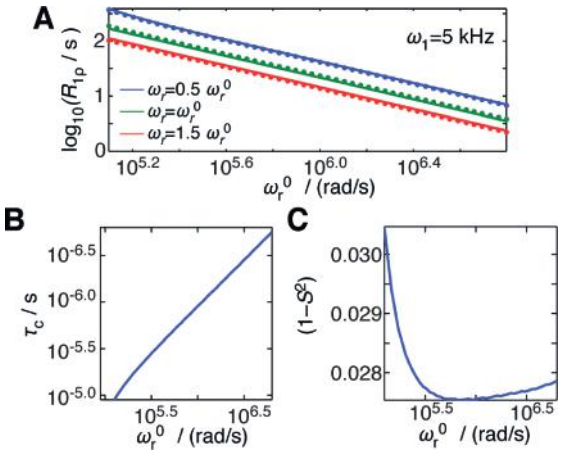


Figure 2. Fitting behavior as a function of experimental settings. A uniform distribution of motion was used to calculate three $R_{1\rho}$ rate constants at different spinning frequencies ($0.5\omega_r^0$, ω_r^0, and $1.5\omega_r^0$), with the spin-lock field strength fixed at $\omega_1/2\pi = 5$ kHz.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

the "light" is better. Adding S^2 to the relaxation data makes the fit more complex as discussed below.

Based only on the experimental measurement conditions, it is possible to determine to which τ_c an experimental $R_{1\rho}$ data set is biased (see the SI). We did so for several dynamics studies and compared the result to the reported τ_c values. 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) $^{13}C\alpha$ 11 μs/7 μs,[5] ubiquitin (2-methyl-2,4-pentanediol (MPD) crystallization) 2.1 μs/1.5 μs (median value),[6] ubiquitin (PEG crystallization) 26 μs/20 μs (SI, Figure 5).[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 results bias to where the experiments are sensitive.

Fitting one data type to a mono-exponential correlation function represents a simple dynamics analysis of solid-state NMR data. Many studies involve fitting multiple longitudinal (R_1) and transverse ($R_{1\rho}$, R_2, etc.) relaxation-rate constants, as well as REDOR-derived order parameters (S^2). This allows a correlation function with two correlation times to be fitted (three, with multiple $R_{1\rho}$[8]). Such studies report similar sets of correlation times, typically near $10^{-10.5}$ s and $10^{-7.5}$ s (see Table 1). This behavior also occurs in generated data sets (SI, Figure 8). From the plot of R_1 at several fields (lines in Figure 3A) it is obvious that R_1 rate constants are most sensitive for $\tau_c \approx 10^{-8.5}$ s. This means that correlation times are 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 spectroscopy.

	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

An interesting property of these correlation times is revealed when a correlation function with the two τ_i values fixed such that Equation (3) results is used. In this case, a good fit of the three R_1 rate constants at *all* correlation times is obtained by varying $(1-S^2)A_1$ and $(1-S^2)A_2$ (Figure 3A).

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

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 $\tau_i = 10^{-7.5}$ s and $10^{-10.5}$ s. This holds relatively well for any pair of correlation times with one time shorter than 10^{-10} s and one time longer than 10^{-8} s (see the SI for details). We note that the values of $(1-S^2)A_i$ are highest near $10^{-8.5}$ s, where R_1 is most sensitive. So, although the fitted correlation times are not where "the light is better", the greatest contributions to the amplitudes in the correlation function come from the most sensitive correlation times, with the unfortunate side effect that, unlike in previous examples, our fitted τ_i values no longer indicate where the experiment is sensitive.

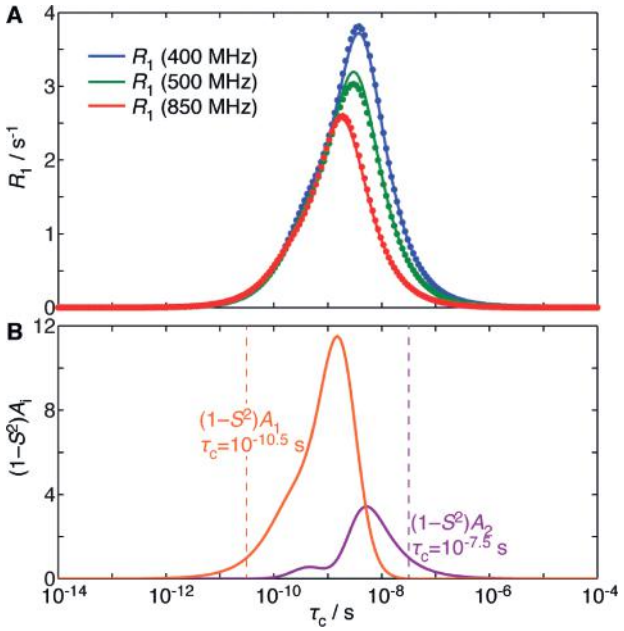


Figure 3. R_1 rate constants fitted with the correlation function given in Equation (3). A) Plots of R_1 rate constants (at fields of 400, 500, and 850 MHz) for monoexponential correlation functions versus correlation time, τ_c, where $(1-S^2)=1$, and circles indicating fits at each τ_c using Equation (3). B) Weighting of the two terms in Equation (3), with the positions of the τ_i values 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 values, one may simultaneously decrease the short τ_i 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 τ_i and increase $(1-S^2)A_2$ to fit the transverse relaxation, also leaving R_1 unchanged (details in the 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 are actually well described by a model with two τ_i, the correct τ_i can be extracted, but if the dynamics are more complex, the fit may bias towards these two "universal" correlation times. The frequent occurence of fitted correlation times around $\tau_c = 10^{-7.5}$ and $10^{-10.5}$ s in the literature suggests either they have some physical significance or, maybe more likely, are an artifact of fitting a complex dynamic process with a simpler model.

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 4A–C, where three distributions of motion produce almost exactly the same set of relaxation data ($5\,R_{1\rho}$, $3\,R_1$, 1 S^2). In Figure 4D a model having three correlation

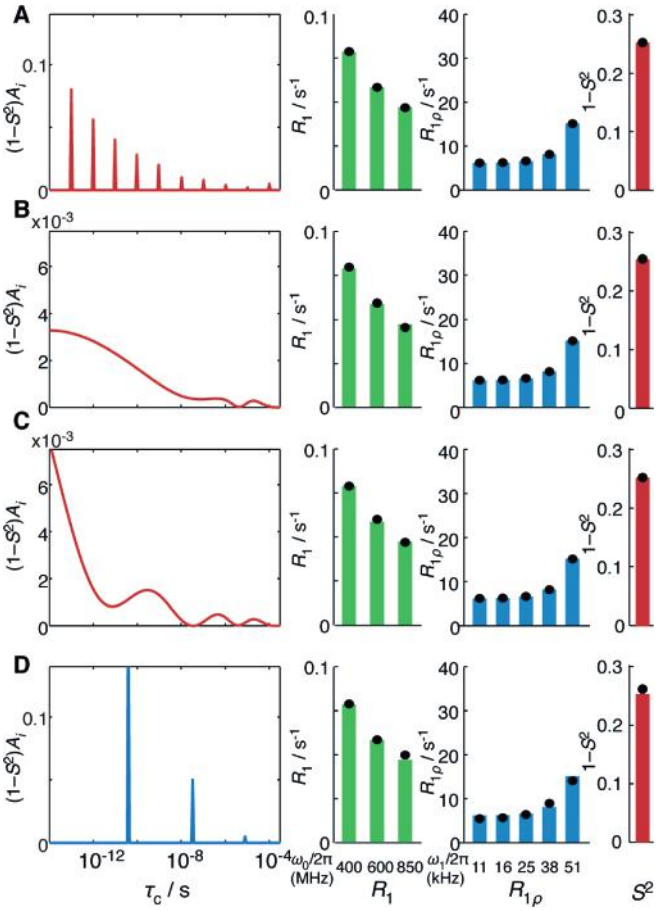


Figure 4. Different distributions of motion with the same relaxation data. Four distributions of motion are plotted as $(1-S^2)\,A_i$ vs. τ_c. Distributions have been chosen in (A)–(C) so that the resulting dynamics data (circles, right) yielded exactly the same values in each case (bars, right). D)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

times is fitted to the relaxation data, yielding a good, but not perfect fit.

When the real motion is more complex than that of 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 the dynamical information is extracted. Therefore, we introduce the concept of a dynamics *detector,* which contains information only where the light is. A detector reports 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. Using multiple such detectors each of which characterizes protein motion on a different timescale, protein dynamics can be described without bias. In addition, there are only detectors where the dynamics experiments are sensitive.

Equation (4) gives a definition of a detector response calculated from dynamics data. $(1-S^2)\theta(\tau_c)$ describes the distribution of motion in the protein, that is the contribution of each τ_c to the total motion. The detector takes the sum of

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

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 R_1).

Ideally detectors would have a narrower sensitivity than the relaxation rate constants and so give more precisely the range of correlation times contained in the motion. 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 (SI, Figure 9). Previous work along these lines exist: spectral-density mapping was performed by taking linear combinations of relaxation data and yielded sensitivities defined by the spectral densities at specific frequencies, $J(\omega,\tau_c)$.[12] LeMaster created detectors by fixing multiple τ_i to carefully chosen values[13] (see also Ref. [14]). However, these cases require 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 selecting 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 shown 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).

Conflict of interest

The authors declare no conflict of interest.

Keywords: correlation times · molecular dynamics · NMR spectroscopy · relaxation

How to cite: *Angew. Chem. Int. Ed.* **2017**, *56*, 13590–13595
Angew. Chem. **2017**, *129*, 13778–13783

[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, *Chem. Eur. J.* **2017**, *23*, 9425 – 9433.
[7] V. Kurauskas, S. A. 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, *Nat. Commun.* **2017**, *8*, 145.
[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] J. Peng, G. Wagner, *J. Magn. Reson.* **1992**, *98*, 308 – 332.
[13] D. M. LeMaster, *J. Biomol. NMR* **1995**, *6*, 366 – 374.
[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.

Manuscript received: July 18, 2017
Accepted manuscript online: August 30, 2017
Version of record online: September 14, 2017

