Journal of Biomolecular NMR, **12:** 471–492, 1998.
KLUWER/ESCOM

Propagation of experimental uncertainties using the Lipari-Szabo model-free analysis of protein dynamics

Danqing Jin[a], Michael Andrec[a,b], Gaetano T. Montelione[b,c,*] and Ronald M. Levy[a,*]
[a]*Department of Chemistry, Wright-Rieman Laboratories;* [b]*Center for Advanced Biotechnology and Medicine and* [c]*Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey, Piscataway, NJ 08855-0939, U.S.A.*

Received 20 April 1998; Accepted 19 June 1998

Key words: model-free formalism, NMR relaxation, order parameter, protein dynamics

Abstract

In this paper we make use of the graphical procedure previously described [Jin, D. et al. (1997) *J. Am. Chem. Soc.*, **119**, 6923–6924] to analyze NMR relaxation data using the Lipari-Szabo model-free formalism. The graphical approach is advantageous in that it allows the direct visualization of the experimental uncertainties in the motional parameter space. Some general 'rules' describing the relationship between the precision of the relaxation measurements and the precision of the model-free parameters and how this relationship changes with the overall tumbling time (τ_m) are summarized. The effect of the precision in the relaxation measurements on the detection of internal motions not close to the extreme narrowing limit is analyzed. We also show that multiple timescale internal motions may be obscured by experimental uncertainty, and that the collection of relaxation data at very high field strength can improve the ability to detect such deviations from the simple Lipari-Szabo model.

Introduction

NMR relaxation (T_1, T_2, NOE) experiments are a very important tool for studying the internal dynamics of proteins (Abragam, 1961; London, 1980). Dynamical information can be extracted from the relaxation data using various analytical models for the dynamics whose parameters may be fit to the relaxation data (Woessner, 1962; Wallach, 1967; Kinosita et al., 1977; Wittebort and Szabo, 1978; Brainard and Szabo, 1981). Another approach involves the direct mapping of the spectral density function from experimental data (Peng and Wagner, 1992). The approach most commonly used is based on the so-called 'model-free' formalism (Levy et al., 1981; Lipari and Szabo, 1982; Clore et al., 1990b). The information contained in the relaxation data is assumed to be completely specified by two quantities: a generalized order parameter, S^2, which is a measure of the spatial restriction of the internal motion, and an effective internal correlation time, τ_e, which is a measure of the rate of the internal motion. This framework, developed by Lipari and Szabo (Lipari and Szabo, 1982), has been applied by many groups to interpret NMR relaxation experiments on proteins (see for example Kay et al., 1989; Clore et al., 1990a; Palmer III et al., 1991; Kördel et al., 1992; Schneider et al., 1992; Stone et al., 1992; Orekhov et al., 1994; Mandel et al., 1995; Li and Montelione, 1995).

In most experimental studies to date, the model-free parameters have been estimated by minimizing the residual sum-of-square error function between the calculated and experimental relaxation parameters (Palmer III et al., 1991). Although this fitting procedure is statistically legitimate, the effects of the uncertainties in the measured relaxation rates on the precision of the model-free parameters is far from transparent.

In earlier work (Jin et al., 1997) we demonstrated how the uncertainties in the relaxation rates may be mapped onto the (S^2, τ_e) plane, allowing the information content of the experimental data to be visualized

*To whom correspondence should be addressed.

graphically. In this paper, we use this method to obtain a better understanding of how these uncertainties propagate, and how this propagation is influenced by the correlation time for the overall macromolecular tumbling (τ_m) and the dynamical regime of the motion. We find that the precision in the estimated motional parameters decreases rapidly as τ_e approaches the tumbling time τ_m. Although the range of validity of the Lipari-Szabo model-free formalism was clearly delineated in the original papers (Lipari and Szabo, 1982), there is a widespread misconception that the model-free approach is applicable only when τ_e is much faster than τ_m. In fact, there is no such *a priori* limit on the range of τ_e. We reexamine this problem using numerical examples to illustrate the precision that is required in order to extract useful information about internal motions on time scales which are not in the extreme narrowing limit. We note that the graphical method has been applied previously in a more limited way to extract protein order parameters and effective correlation times from NMR experiments (Henry et al., 1986; Weaver et al., 1988; Fushman et al., 1994).

Background

NMR relaxation of nuclei in a macromolecule tumbling in solution is determined by the motions through the time correlation function, $C(t)$. Assuming that the overall motion of the macromolecule is isotropic, and that the overall and internal motions are uncoupled, the total time correlation function $C(t)$ can be factored into a product of correlation functions

$$C(t) = C_O(t)C_I(t). \tag{1}$$

The correlation function for overall isotropic rotational motion $C_O(t)$ is given by

$$C_O(t) = \frac{1}{5}e^{-6D_m t} = \frac{1}{5}e^{-t/\tau_m}, \tag{2}$$

where D_m and τ_m are the rotational diffusion constant and correlation time of the macromolecule, respectively. For dipolar NMR relaxation the correlation function for the internal motions $C_I(t)$ is given by

$$C_I(t) = \langle P_2(\hat{\mu}(0) \times \hat{\mu}(t))\rangle, \tag{3}$$

where the unit vector $\hat{\mu}$ points along the internuclear axis between the dipolar coupled nuclei in the macromolecule-fixed frame and $P_2(x)$ is the second Legendre polynomial

$$P_2(x) = \frac{1}{2}(3x^2 - 1). \tag{4}$$

It should be noted that the factorization of the total time correlation function into a product of correlation functions corresponding to overall tumbling and internal motions does not depend on a separation of time scale between these motions. For most native proteins which are sufficiently well structured that their rotational tumbling can be charaterized by a single (time independent) diffusion tensor, we expect that the factorization of Equation 1 is a very good approximation.

The relaxation of protonated ^{15}N nuclei is mediated primarily by dipole-dipole interactions with the attached protons and secondarily by chemical shift anisotropy. If cross-correlation effects are suppressed (Boyd et al., 1990), the longitudinal relaxation rate (R_1), transverse relaxation rate (R_2), and the nuclear Overhauser enhancement (NOE) are given by

$$\begin{aligned} R_1 &= \frac{d^2}{4}[J(\omega_H - \omega_N) + 3J(\omega_N) \\ &\quad +6J(\omega_H + \omega_N)] + \frac{\Delta^2\omega_N^2}{3}J(\omega_N) \end{aligned} \tag{5}$$

$$\begin{aligned} R_2 &= \frac{d^2}{8}[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) \\ &\quad +6J(\omega_H) + 6J(\omega_H + \omega_N)] \\ &\quad +\frac{\Delta^2\omega_N^2}{18}[4J(0) + 3J(\omega_N)] \end{aligned} \tag{6}$$

$$\mathrm{NOE} = 1 + \frac{\gamma_H\,[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)]}{\gamma_N\,[J(\omega_H - \omega_N) + \{3 + 4\Delta^2\omega_N^2/(3d^2)\}J(\omega_N) + 6J(\omega_H + \omega_N)]} \tag{7}$$

where the spectral density function, $J(\omega)$, is the Fourier cosine transform of the total correlation function,

$$J(\omega) = 2\int_0^\infty \cos(\omega t)\, C(t)\, dt, \tag{8}$$

and γ_H and γ_N are the gyromagnetic ratios for the proton and nitrogen respectively; ω_H and ω_N are the Larmor frequencies of ^{1}H and ^{15}N, and the chemical shift anisotropy (CSA) value $\Delta = \delta_\parallel - \delta_\perp$ (the difference between the parallel and perpendicular components of the chemical shift tensor). The constant d is given by

$$d = \frac{\gamma_H\gamma_N\hbar}{r_{NH}^3}\frac{\mu_0}{4\pi}, \tag{9}$$

where $\hbar$ is Planck's constant divided by 2π, r_{NH} is the amide ^{15}N-^{1}H bond length, and μ_0 is the permeability of free space.

If we assume that $C_I(t)$ decays to a plateau value for $t \rightarrow \infty$ and that the decay proceeds in a single-exponential manner characterized by an effective correlation time τ_e, then $C_I(t)$ can be written as

$$C_I(t) = S^2 + (1 - S^2)e^{-t/\tau_e}. \quad (10)$$

As stated in the original paper (Lipari and Szabo, 1982), the plateau value S^2 is a model-independent measure of the degree of spatial restriction of the internal motion, whose value can range from 0 (when the vector samples all possible orientations) to 1 (when the vector is completely restricted); the second parameter, τ_e, is related to the rate of the internal motion. It is important to note that although S^2 has a model-independent significance, τ_e depends on the spatial nature of the motion and can be related to microscopic rate constants only within the framework of a particular dynamic model.

Having obtained an expression for the total time correlation function, one can readily compute the spectral density function in the model-free formalism:

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_m}{1 + (\omega\tau_m)^2} + \frac{(1 - S^2)\tau}{1 + (\omega\tau)^2}\right] \quad (11)$$

where

$$\frac{1}{\tau} = \frac{1}{\tau_e} + \frac{1}{\tau_m}. \quad (12)$$

In the absence of internal motions (e.g., when $S^2 = 1$), the spectral density function is reduced to its isotropic 'rigid body' limit

$$J_0(\omega) = \frac{2}{5}\left[\frac{\tau_m}{1 + (\omega\tau_m)^2}\right] \quad (13)$$

which depends only on the overall tumbling time τ_m of the protein.

If τ_m is known, then the relaxation rates R_i (i = 1, 2, 3; R_3 = NOE) can be directly calculated if S^2 and τ_e are given, whereas it is not possible to evaluate S^2 or τ_e analytically when the relaxation rates R_1, R_2, and NOE are known. The precision of the relaxation data may be represented by an interval of the form $R_i \pm \Delta R_i$, where ΔR_i is the uncertainty in the i-th measurement. In other words, we assume that due to experimental uncertainties, the relaxation measurements are uniformly distributed in the interval ($R_i - \Delta R_i$, $R_i + \Delta R_i$). Our graphical procedure involves the construction of contour lines of constant R_i as a function of S^2 and τ_e. For a given relaxation measurement $R_i \pm \Delta R_i$, the area in the S^2-τ_e plane between the contour lines corresponding to $R_i - \Delta R_i$ and $R_i + \Delta R_i$ contains all (S^2, τ_e) points consistent with that experimental measurement. The intersection of these areas for all three relaxation values defines the complete solution space for the model-free parameters, and contains all (S^2, τ_e) points which are consistent with all three NMR relaxation measurements. Our graphical method can be viewed as a simple example of 'set-theoretic estimation', a general data analysis strategy that has been applied to a wide variety of problems in the electrical engineering literature (Combettes, 1993). Alternatively, our solution space can be interpreted as being proportional to the Bayesian posterior probability density $P(S^2, \tau_e | D)$ under a uniform prior and box-shaped likelihood function (see Bretthorst (1990a) for an introduction to Bayesian statistics in an NMR context). This treatment is somewhat simplistic, in that each (S^2, τ_e) point inside the solution space has an equal probability. As an alternative, we can replace the uniform error distributions with Gaussian (normal) probability distributions. The results presented in this paper are based on uniform error distributions. When Gaussian error distributions are used instead, the conclusions are qualitatively the same as obtained using a uniform error distribution.

In the following section the graphical procedure is used to analyze how experimental uncertainties in the NMR relaxation measurements affect the precision in the estimates of the Lipari-Szabo parameters which describe the protein motions. The results depend upon the assumed precision in the relaxation measurements. The precision reported in publications varies widely; the methods used to estimate the uncertainties vary in statistical rigor, and many publications do not report any estimate of the precision. The synthetic data discussed in this paper assumes uncertainties in the NMR relaxation data in the range of 2% – 5%. The highest precision reported in the current literature for these kinds of measurements is on the order of 1%. Although not considered further here, we note that the approximations upon which Equations 5–7 are based, especially the two-spin approximation and that the CSA Δ is independent of conformation (Tjandra et al., 1996a), need to be examined more carefully when considering the theoretical limits of the precision that it is presently possible to obtain in the relaxation measurements and the meaning of this estimate in the context of these approximations.

Results and discussion

Generic features of the contour maps

In order to provide a general overview of the behavior of the relaxation rates as a function of S^2 and τ_e, contour maps of R_1, R_2, and NOE were constructed for a representative tumbling time ($\tau_m = 4$ ns) and field strength ($\omega_H = 500$ MHz), and are shown in Figure 1. In each map, the increment between two adjacent contour lines was kept constant. There are several features common to all three maps that are worth noticing. First, smaller relaxation rates and NOE values (in the lower left corner of each map) generally correspond to smaller values of S^2 and τ_e. Contour lines corresponding to small R_1, R_2, or NOE values cover a small range of (S^2, τ_e) values, while contour lines corresponding to larger relaxation rates in general cover a broader range. Large S^2 values and/or large τ_e values can only be found along these latter contour lines. This implies that, for a given value of τ_m, larger relaxation rates usually correspond to internal motions that occur with a smaller magnitude or on a longer time scale and is consistent with the general observation that fast internal motions tend to decrease the relaxation rates (Levy et al., 1981). Furthermore, as the relaxation rates increase and approach their 'rigid body' values (denoted by R_1^0, R_2^0, and NOE^0), the spacing between contour lines becomes less dense. This is particularly apparent in the R_1 and NOE contour maps (Figures 1a and 1c), and implies that the relaxation rates are least sensitive to changes in S^2 and τ_e when they are close to their corresponding 'rigid body' values. This graphical analysis demonstrates that when measured relaxation parameters are close to R_1^0 and NOE^0 it is necessary to have much higher precision in the corresponding measurements in order to sufficiently restrict the solution space of motional parameters.

Behavior of the contour maps as a function of overall tumbling time (τ_m)

In order to provide more insight into the impact of changes in τ_m on the observed relaxation rates, contour maps were constructed for several values of the overall tumbling time ($\tau_m = 2$, 4, 8, 16 ns) assuming a field strength of 500 MHz (Figures 2-5). The behavior of the contour map of each type of relaxation parameters (R_1, R_2, and NOE) is described below.

R_1 Contour maps

Figure 2 shows contour diagrams for R_1 relaxation for macromolecular tumbling times of 2 ns, 4 ns, 8 ns, and 16 ns. For $\tau_m = 2$ ns (Figure 2a), the relationship between S^2 and τ_e for a given R_1 value is nearly linear. However, for $\tau_m = 4$ ns (Figure 2b), the contour map is qualitatively different. The special feature of this map is related to the 'rigid body' value, R_1^0. When R_1 is small (i.e. $< 2.0\ \mathrm{s}^{-1}$), there is little difference from the behavior at 2 ns. However, as R_1 approaches R_1^0 ($3.01\ \mathrm{s}^{-1}$ for $\tau_m = 4$ ns), the contour lines become nearly parallel to the S^2 axis. When $\mathrm{R_1}$=$\mathrm{R_1^0}$, the contour is in fact perpendicular to the τ_e axis. Futhermore, R_1^0 is no longer the maximal value of R_1, and contour lines corresponding to $\mathrm{R_1} > \mathrm{R_1^0}$ can be found to the right side of this 'rigid body' contour. In this region of the contour map, the effect of internal motions is to increase R_1 above the corresponding value for the rigidly tumbling protein. This effect can be seen most clearly in Figures 2c and 2d. The origin of these features can be understood by examining the conditions under which the Lipari-Szabo spectral density (Equation 11) reduces to the 'rigid body' limit (Equation 13). They can be found simply by setting Equations 11 and 13 equal to each other. We find that $J(\omega)$=$J_0(\omega)$ only if

$$(\mathrm{S}^2 - 1)(\tau - \tau_m)(1 - \omega^2\tau_m\tau) = 0, \qquad (14)$$

from which we can see that the following are the sufficient conditions:

1. $S^2 = 1$
2. $\tau_e = \infty$
3. $\tau_e = \tau_m/(\omega^2\tau_m^2 - 1)$ (for $\omega\tau_m > 1$).

It follows, therefore, that if $\tau_e = \tau_m\ /(\omega^2\tau_m^2 - 1)$, $J(\omega)$ is equal to $J_0(\omega)$, and is independent of S^2. Since R_1 is dominated by $J(\omega_N)$, R_1 will also be independent of S^2 if the effective internal correlation time of the internal motions equals the 'critical τ_e'. The value of this 'critical τ_e' is approximately 6.9 ns for $\tau_m = 4$ ns and $\omega_N = 50$ MHz; which is somewhat different from the actual value of the 'critical τ_e' (5.03 ns for the same values of τ_m and ω_N) because terms other than $J(\omega_N)$ also contribute to R_1. The true 'critical τ_e' can be found by equating the full expressions for R_1 and R_1^0 obtained using Equation 5. It is clear that the information content of R_1 concerning the internal motional averaging, S^2, is at a minimum when the measured R_1 value approaches R_1^0. There is a simple explanation for why, given a sufficiently

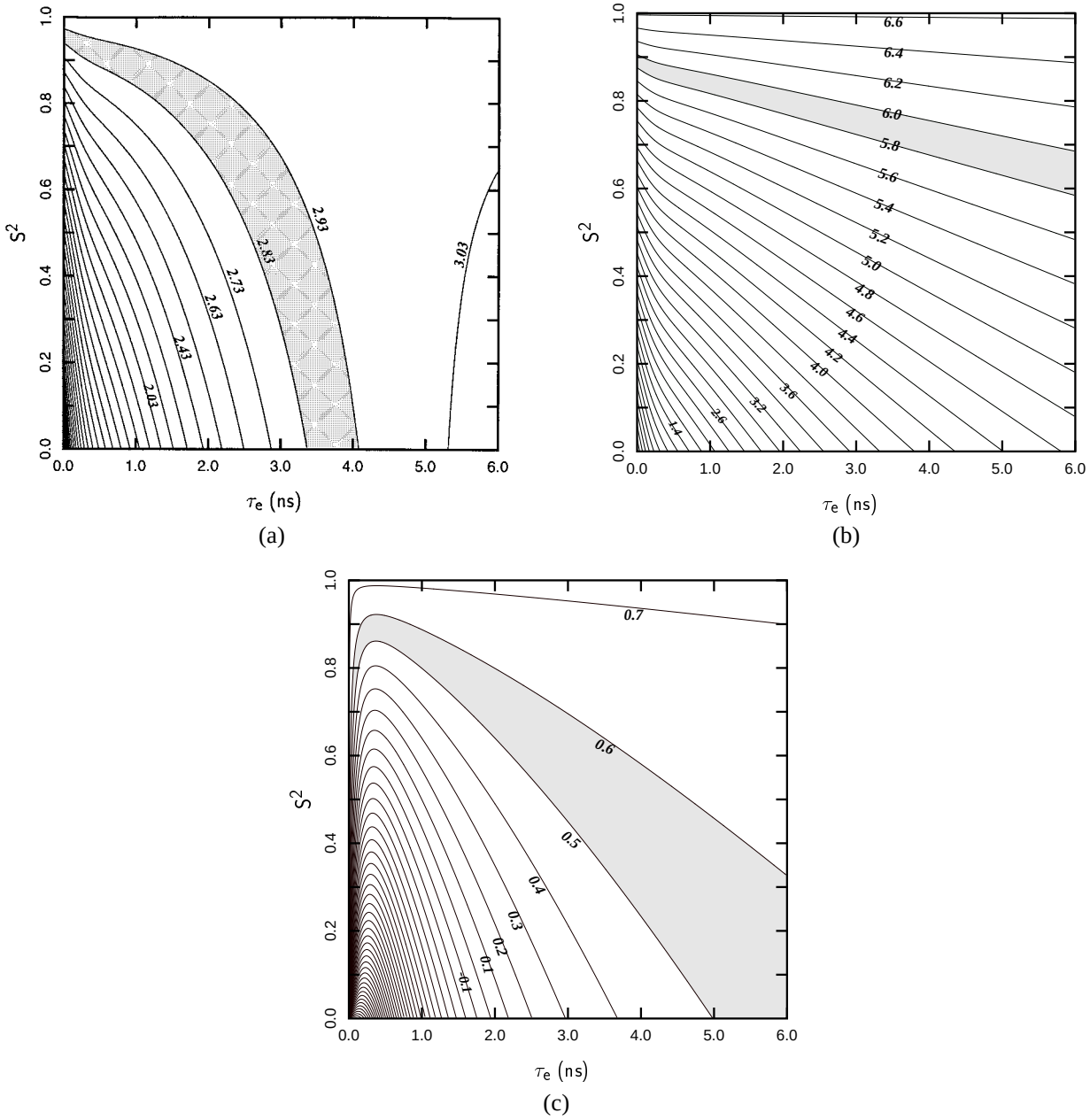

Figure 1. Contour maps of S^2 and τ_e for different types of ^{15}N NMR relaxation measurements: (a) longitudinal R_1 contours, ranging from 0.03 s^{-1} to 3.03 s^{-1} with 0.1 s^{-1} step, $R_1^0 = 3.01$ s^{-1}, (b) transverse R_2 contours, ranging from 0.2 s^{-1} to 6.6 s^{-1} with 0.2 s^{-1}step, $R_2^0 = 6.62$ s^{-1}, (c) heteronuclear NOE contours, ranging from −3.6 to 0.7 with 0.1 step, $NOE^0 = 0.718$. For these contour maps the spectrometer frequencies ω_H and ω_N were chosen to be 500 MHz and 50.7 MHz respectively, and the overall correlation time $\tau_m = 4$ ns. To illustrate the mapping procedure, a region is shaded in each map corresponding to the following sample NMR data: $R_1 = 2.88 \pm 0.05$ s^{-1}, $R_2 = 5.90 \pm 0.10$ s^{-1}, NOE = 0.55 ± 0.05.

long tumbling time such that $\tau_m > 1/\omega$, there exists a value of τ_e for which $R_1 = R_1^0$. In Figure 3, the 'rigid body' longitudinal relaxation rate R_1^0 is plotted as a function of tumbling time τ_m. The maximum relaxation rate R_1^{MAX} occurs when $\tau_m = 1/\omega$. For an ω_N of 50 MHz, R_1^{MAX} (3.13 s^{-1}) occurs at a tumbling time $\tau_m^{MAX} \approx 3.0$ ns. The macromolecular tumbling changes from the faster motion limit to the slow mo-

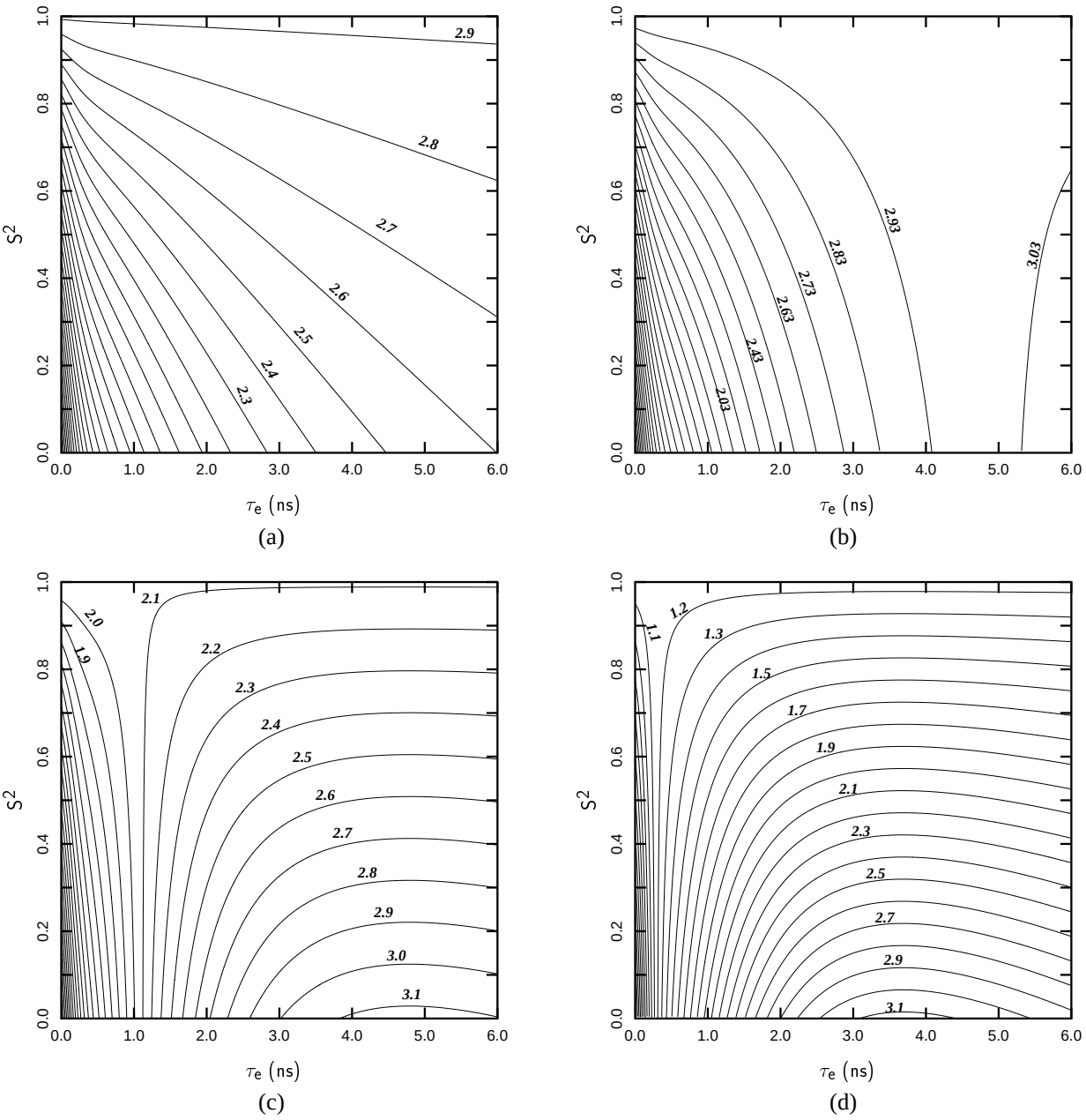

Figure 2. The longitudinal relaxation R_1 contour maps at different values of overall correlation time τ_m: (a) $\tau_m = 2$ ns, R_1 ranging from 0.0 s^{-1} to 2.9 s^{-1} with 0.1 s^{-1} step; (b) $\tau_m = 4$ ns, R_1 ranging from 0.03 s^{-1} to 3.03 s^{-1} with 0.1 s^{-1} step; (c) $\tau_m = 8$ ns, R_1 ranging from 0.0 s^{-1} to 3.1 s^{-1} with 0.1 s^{-1} step; (d) $\tau_m = 16$ ns, R_1 ranging from 0.0 s^{-1} to 3.1 s^{-1} with 0.1 s^{-1} step.

tion limit at τ_m^{MAX}. For any value of $R_1^0 < R_1^{MAX}$, there are two tumbling times corresponding to this relaxation rate; one longer than τ_m^{MAX} and a second one shorter than τ_m^{MAX}. Thus, for ω_N at 50 MHz and tumbling times greater than 3.0 ns, internal motions of a certain frequency (τ'_e) can combine with the overall tumbling ($\tau'^{-1} = \tau'^{-1}_e + \tau_m^{-1}$) to produce the same longitudinal relaxation rate as the overall tumbling alone, $R_1(\tau') = R_1^0(\tau_m)$. When this condition is fulfilled, the longitudinal relaxation rate is independent

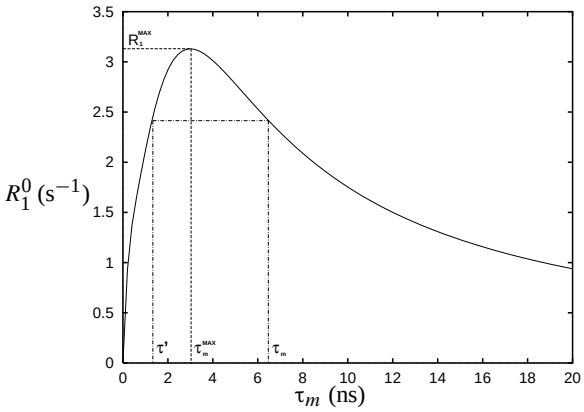

Figure 3. The 'rigid body' longitudinal relaxation rate R_1^0 as a function of the overall tumbling time τ_m. The function is double-valued. When an overall tumbling time τ_m is longer than the τ_m^{MAX}, the longitudinal relaxation rate R_1 for internal motions on the time scale of τ_e' (where $1/\tau' = 1/\tau_e' + 1/\tau_m$) is independent of S^2. A magnetic field strength of $\omega_H = 500$ MHz is used here.

Table 1. Field strength dependence of R_1^{MAX} and τ_m^{MAX}

ω_N (MHz)	R_1^{MAX} (s^{-1})	τ_m^{MAX} (ns)
50.7	3.13	3.0
60.8	2.81	2.5
76.0	2.54	2.0
101.3	2.38	1.5

of S^2 because the two Lorentzians which appear in Equation 11 are equal:

$$\frac{\tau_m}{1+(\omega\tau_m)^2} = \frac{\tau}{1+(\omega\tau)^2}. \qquad (15)$$

Therefore, any combination of coefficients S^2 and $(1 - S^2)$ leaves $J(\omega)$ of Equation 11 unchanged. This explains the transition which appears in the (S^2, τ_e) contour maps (Figure 2) when $\tau_m > 3.0$ ns. The field strength dependence of R_1^{MAX} and τ_m^{MAX} is illustrated in Table 1, where τ_m^{MAX} can be well approximated by $1/\omega_N$.

For $\tau_m = 8$ ns, the R_1 contour map (Figure 2c) differs from the 4 ns contour map only in its quantitative details: the 'rigid body' contour line is shifted towards the left side of the map ($R_1^0 = 2.09$ s^{-1}), with the critical τ_e being only about 1.1 ns, and the R_1^0 value is significantly smaller than the maximum R_1 value (3.13 s^{-1}). The contour map for $\tau_m = 16$ ns (Figure 2d) continues this trend: the 'rigid body' contour is further shifted to the left ($R_1^0 = 1.13$ s^{-1}), with the critical τ_e being about 0.3 ns. The overall qualitative patterns of the contour maps for R_1 at ω_N = 75 MHz are very similar, except that the critical τ_e becomes smaller for each τ_m.

Table 2. Values of R_1/R_1^0 as a function of τ_m for different values of τ_e/τ_m [a]

τ_m (ns)	τ_e/τ_m		
	0.02	0.2	1.0
2	0.81	0.89	0.94
4	0.83	0.91	0.99
8	0.87	1.04	1.09
16	1.01	1.34	1.16

[a] S^2 is assumed to be equal to 0.8 and the magnetic field strength ω_H = 500 MHz.

R_2 contour maps

Compared to the R_1 contour maps, the qualitative patterns of R_2 contour maps (Figure 4) do not change as much as a function of τ_m. This is due to the fact that R_2 is dominated by the $J(0)$ term, which increases monotonically as a function of τ_e. Furthermore, unlike the situation for R_1, $J(0) \leq J_0(0)$ for all values of τ_e, which indicates that the 'rigid body' transverse relaxation rate, R_2^0, is also the maximum R_2 value. In fact, the R_2 contour maps are approximately unchanged with respect to τ_m when the τ_e is plotted in reduced units (i.e., as a function of τ_e/τ_m), since

$$J(0) = \frac{2}{5}\left[\tau_m S^2 + \tau(1 - S^2)\right] \qquad (16)$$

and the 'rigid body' $J_0(0)$ is simply equal to τ_m. If we express τ_e in units of $\tau_m$and assume R_2 is dominated by the $J(0)$ term, then Equation 16 can be written in reduced units:

$$J(0)/J_0(0) = S^2 + \lambda(1 - S^2) = R_2/R_2^0, \qquad (17)$$

where $\lambda = \frac{m}{1+m}$ and $m = \tau_e/\tau_m$. From Equation 17 it is apparent that R_2/R_2^0 as a function of τ_e/τ_m (i.e., λ) is independent of τ_m.

Another feature of R_2 contour maps is that the contours are nearly linear. Again assuming that R_2 is dominated by $J(0)$, Equation 16 can be rearranged to

$$S^2 = C_2\tau_e + C_1 \qquad (18)$$

where $C_1 = J(0)/J_0(0)$ and $C_2 = (C_1 - 1)/J_0(0)$. For any given values of τ_m and R_2, C_1 and C_2 are constants, and S^2 and τ_e are linearly related. Furthermore, since C_2 approaches zero as $J(0)$ approaches

$J_0(0)$, the contour lines tend to become parallel to the τ_e axis as R_2 increases. This implies that R_2 is much less sensitive to changes in τ_e than it is to S^2 when R_2 approches R_2^0. Since R_2 is well approximated by $J(0)$, the qualitative patterns of R_2 contour maps do not change with the field strength.

NOE Contour maps

In contrast to the previous two cases, the NOE cannot be approximated by a single $J(\omega)$ term. Although we could adopt the simplification proposed by Farrow et al. (Farrow et al., 1995), and approximate the NOE by $J(\omega_H)/J(\omega_N)$, the resulting expression is still too cumbersome to predict the behavior of NOE contour maps based on an analysis of the spectral density. Although the more complicated behavior of the NOE contours could be avoided by examining the dependence of the cross relaxation rate

$$R_x = \frac{d^2}{4}[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] \quad (19)$$

on S^2 and τ_e, we do not believe that this is a practical alternative. Even though the cross-relaxation rate could be obtained using the relation

$$R_x = \frac{\gamma_N}{\gamma_H}(\mathrm{NOE} - 1)R_1, \quad (20)$$

the uncertainty in R_x would then be a function of the uncertainties in both the NOE and the R_1 measurements, making our graphical analysis less transparent. Furthermore, our graphical approach assumes that the R_1, R_2, and NOE measurements are independent. If the NOE is converted into cross-relaxation rate this will no longer be the case. Also, we feel that it is best to perform analyses on data in a form which is as close as practical to the raw experimental measurements. The reason for this is particularly apparent here, as any systematic error in R_1 will obviously corrupt the NOE data if the latter is recast as a cross-relaxation rate. Instead we briefly summarize the general features of the NOE contour maps shown in Figure 5. One difference between NOE contour maps and the contour maps for R_1 and R_2 is that the NOE contours are not single-valued functions of S^2. Contour lines with small (or negative) NOE values are usually confined within the lower left region of the map indicating that small NOE values are generally consistent with high internal mobility on short timescales ($\tau_e < 1$ ns). As τ_m increases, the spacing between contour lines with small NOE values decreases while the spacing between those with large NOE values increases.

The behavior of the allowed model-free parameter space (S^2, τ_e) for various τ_e and τ_m regimes

A single relaxation measurement is, of course, not sufficient to allow interpretation in terms of internal motions. By combining the contour maps corresponding to the different relaxation data, we can achieve a better understanding of the shape of the final (S^2, τ_e) solution space and how it is affected by changes in τ_m. Since the results are very different depending on the time scale of τ_e relative to τ_m, we have chosen three τ_e/τ_m ratios to represent three qualitatively different time scales of the internal motions: (a) $\tau_e/\tau_m = 0.02$, for very fast internal motions; (b) $\tau_e/\tau_m = 0.2$, for intermediate time scale internal motions; and (c) $\tau_e/\tau_m = 1.0$, for internal motions that are on the time scale of the overall tumbling time. In addition to the ratio of the internal to the overall correlation time τ_e/τ_m, the ratio of the longitudinal relaxation to its 'rigid body' value R_1/R_1^0 is an important parameter because the structure of the R_1 contours is very different on either side of the critical τ_e. In this section, we use both parameters τ_e/τ_m and R_1/R_1^0 to categorize the interpretation of the data in different regimes. The relationship between the two parameters τ_e/τ_m and R_1/R_1^0 as a function of tumbling time is presented in Table 2.

To investigate the effect of changing τ_m on the (S^2, τ_e) solution space, we synthesized a series of N–H relaxation data for each of the macromolecular tumbling times $\tau_m = 2$, 4, 8, and 16 ns, while setting S^2 to 0.8 and τ_e/τ_m to 0.02, 0.2, and 1.0. A uniform experimental uncertainty of $\pm 5\%$ was added to all the simulated relaxation data. The graphical procedure for determining the model-free parameters (Jin et al., 1997) was then applied to analyze the simulated data. The allowed ranges for S^2 and τ_e are summarized in Table 3.

For very fast internal motions (i.e., $\tau_e/\tau_m = 0.02$) (Figure 6), the (S^2, τ_e) solution space is usually well-determined i.e., with a small uncertainty in both S^2 and τ_e. We found that when $\tau_m = 2$ ns (Figure 6a) the precision of the NOE measurement is particularly crucial. Without any NOE data, the (S^2, τ_e) solution space is very poorly determined, with S^2 ranging from 0.15 to 0.83 and τ_e ranging from 0.01 to 3.6 ns. On the other hand, either R_1 or R_2 (but not both), could be omitted. While R_1 and R_2 data are equally valuable in restricting the solution space in Figure 6a, in simulations with larger values of τ_m, the R_1 data are essential for distinguishing disjoint solution regions

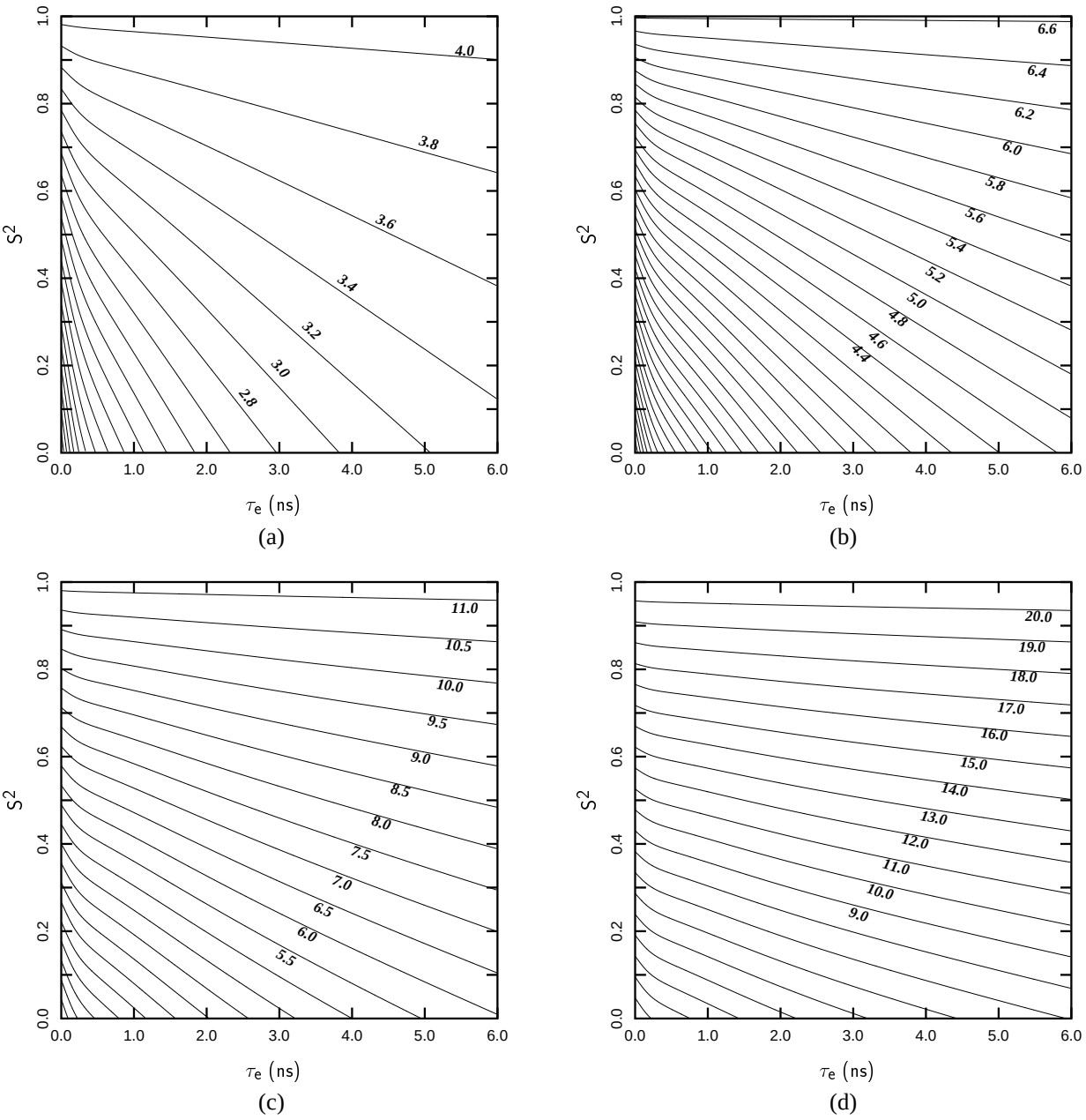

Figure 4. The transverse relaxation R_2 contour maps at different values of overall correlation time τ_m: (a) $\tau_m = 2$ ns, R_1 ranging from 0.0 s^{-1} to 4.0 s^{-1} with 0.2 s^{-1} step; (b) $\tau_m = 4$ ns, R_1 ranging from 0.0 s^{-1} to 6.6 s^{-1} with 0.2 s^{-1} step; (c) $\tau_m = 8$ ns, R_1 ranging from 0.0 s^{-1} to 11.0 s^{-1} with 0.5 s^{-1} step; (d) $\tau_m = 16$ ns, R_1 ranging from 0.0 s^{-1} to 20.0 s^{-1} with 1.0 s^{-1} step. A magnetic field strength of ω_H = 500 MHz is used here.

which result from the characteristic behavior of NOE contours. For example, when $\tau_m = 4$ ns (Figure 6b), the R_1 data are crucial, as they distinguish between two disjoint solution regions — a small region centered around $S^2 = 0.8$ and $\tau_e = 0.02\tau_m$ (i.e., $\tau_e = 80$ ps), and a poorly determined region at larger values of τ_e, that covers a broad range of S^2 and τ_e values. Given that R_1 data are available that allow this distinction, increased precision of either R_1 or R_2 measurements significantly reduces the area of

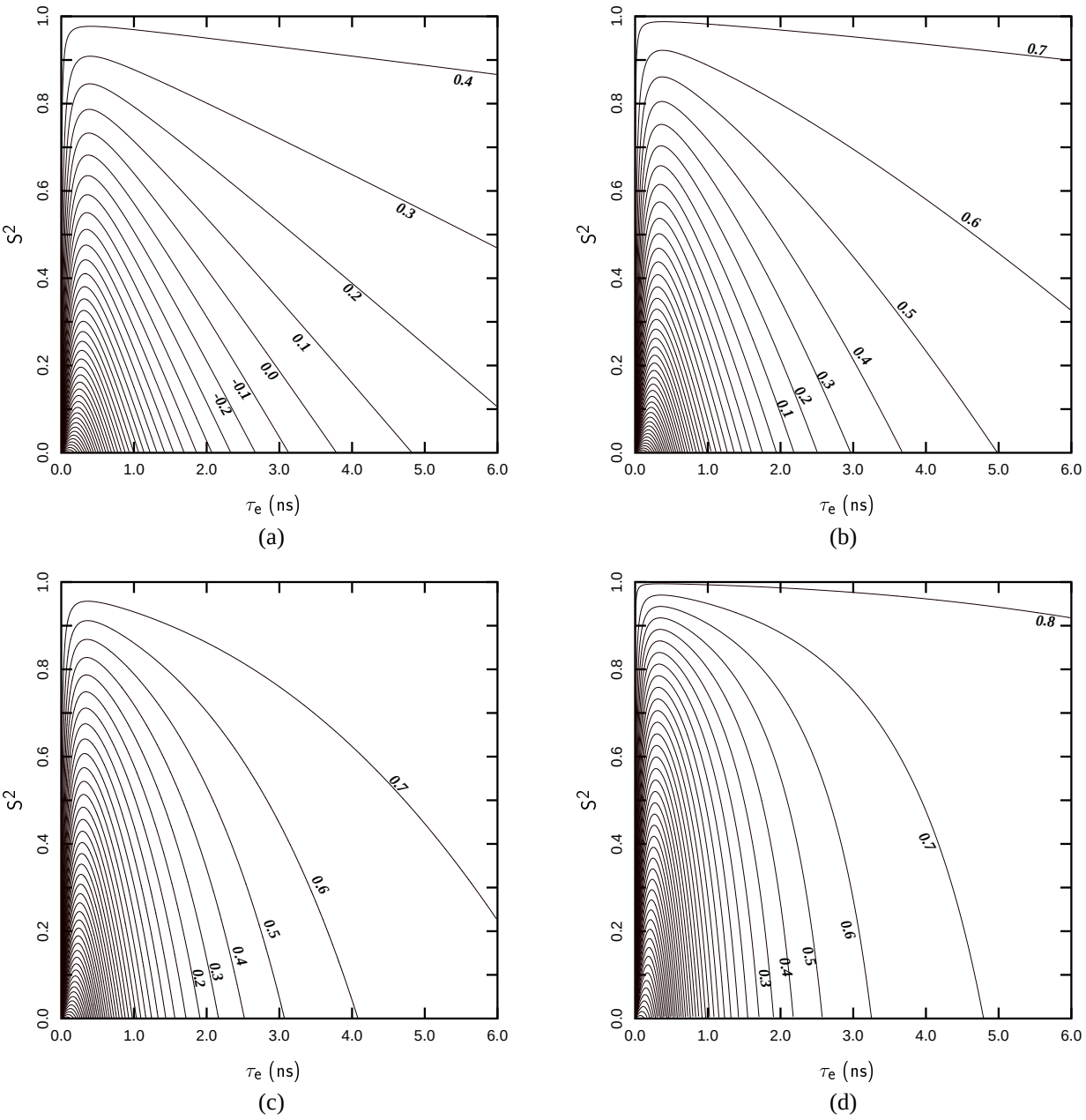

Figure 5. The heteronuclear NOE contour maps at different values of overall correlation time τ_m: (a) $\tau_m = 2$ ns, R_1 ranging from -3.6 to 0.4 with 0.1 step; (b) $\tau_m = 4$ ns, R_1 ranging from -3.6 to 0.7 with 0.1 step; (c) $\tau_m = 8$ ns, R_1 ranging from -3.6 to 0.7 with 0.1 step; (d) $\tau_m = 16$ ns, R_1 ranging from -3.6 to 0.8 with 0.1 step. A magnetic field strength of $\omega_H = 500$ MHz is used here.

the final solution space. Similar conclusions can be drawn from the simulations carried out with larger τ_m values. When $\tau_m = 8$ ns (Figure 6c) or $\tau_m = 16$ ns (Figure 6d), the R_2 data do not contribute to further restricting the allowed (S^2, τ_e) region determined by the combination of R_1 and NOE data with the 5% uncertainty used here. It is interesting that the τ_e range is almost entirely defined by the R_1 measurements, while the S^2 range is almost entirely defined by the R_2 measurements (in the absence of NOE data).

For intermediate time scale motions (i.e., $\tau_e/\tau_m = 0.2$) (Figure 7), the area covered by the solution space

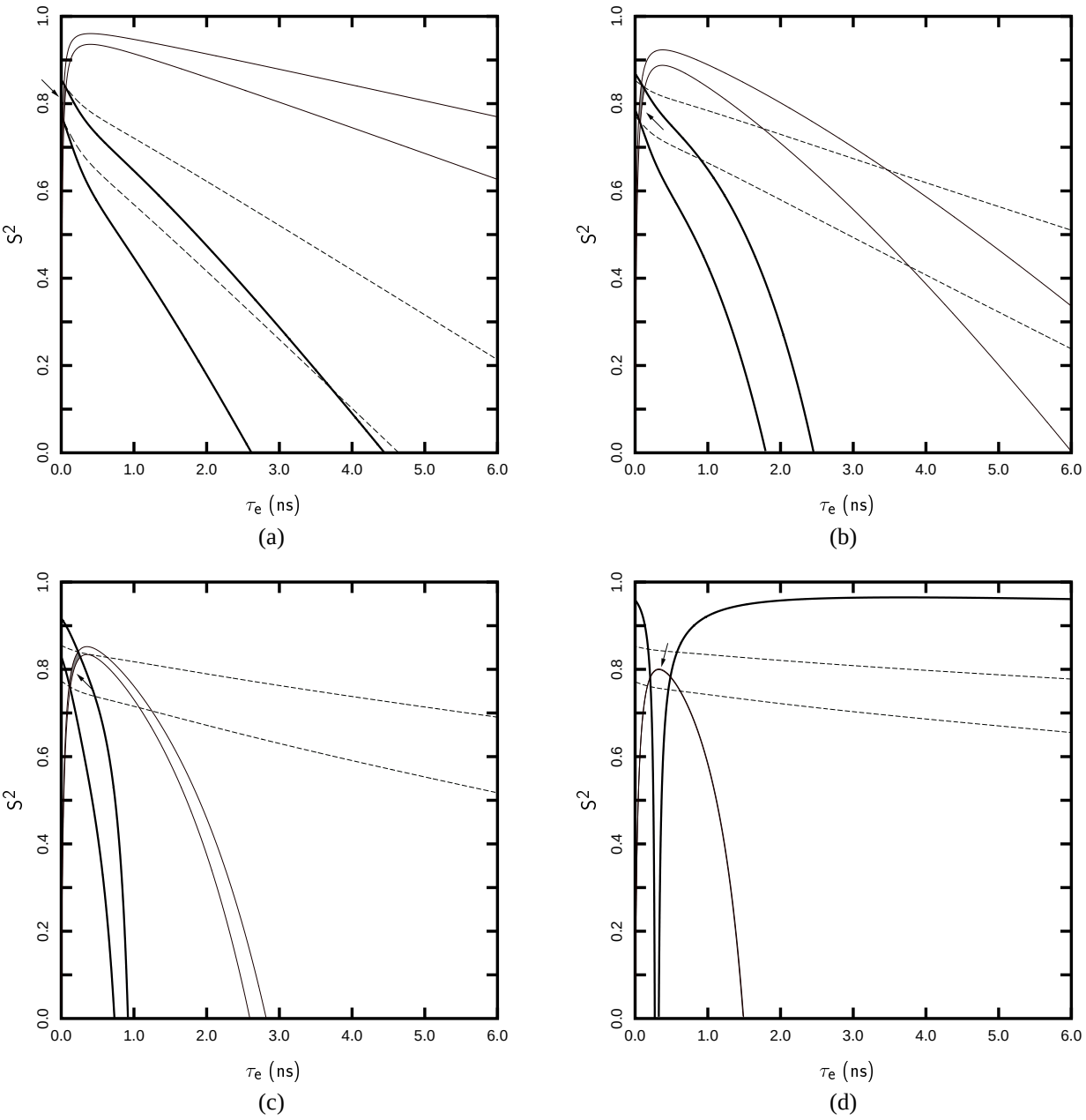

Figure 6. The allowed (S^2, τ_e) solution space for fast time scale internal motions obtained from simulated relaxation data ($\tau_e/\tau_m = 0.02$, see Table 3): contours of R_1, R_2, and NOE are plotted in thick, dashed, and thin lines, respectively. The areas corresponding to the allowed (S^2, τ_e) solution space are shaded in gray and indicated by an arrow when they are very small. (a) $\tau_m = 2$ ns: $R_1 = 2.379 \pm 0.119$ s^{-1}, $R_2 = 3.305 \pm 0.165$ s^{-1}, NOE $= 0.358 \pm 0.018$; (b) $\tau_m = 4$ ns: $R_1 = 2.494 \pm 0.125$ s^{-1}, $R_2 = 5.384 \pm 0.269$ s^{-1}, NOE $= 0.573 \pm 0.029$; (c) $\tau_m = 8$ ns: $R_1 = 1.824 \pm 0.091$ s^{-1}, $R_2 = 9.13 \pm 0.457$ s^{-1}, NOE $= 0.439 \pm 0.022$; (d) $\tau_m = 16$ ns: $R_1 = 1.169 \pm 0.0584$ s^{-1}, $R_2 = 16.97 \pm 0.849$ s^{-1}, NOE $= 0.0546 \pm 0.0027$. A magnetic field strength of $\omega_H = 500$ MHz is used here.

increases considerably. The solution space is largest at τ_m = 2 ns (Figure 7a), where the relative error in S^2 is approximately 40% and the relative error in τ_e is 550%. Higher precision of the R_2 measurements would be required to further restrict the (S^2, τ_e) solution space. The situation for $\tau_m = 4$ ns is somewhat less severe (Figure 7b) – the S^2 range is obtained with an error of $\approx 20\%$ while τ_e is still quite ill-determined.

Table 3. Graphical analysis of the simulated relaxation data with 5% error for different values of τ_m [a]

τ_m (ns)	S^2			τ_e (ns)		
	$\tau_e/\tau_m = 0.02$	$\tau_e/\tau_m = 0.2$	$\tau_e/\tau_m = 1.0$	$\tau_e/\tau_m = 0.02$	$\tau_e/\tau_m = 0.2$	$\tau_e/\tau_m = 1.0$
2	0.75–0.84	0.48–0.81	0–0.92	0.01–0.06 (0.04)	0.20–2.6 (0.4)	≥ 0.12[b] (2.0)
4	0.76–0.85	0.68–0.86	0–0.97	0.05–0.14 (0.08)	0.17–1.7 (0.8)	≥ 0.03[b] (4.0)
8	0.77–0.84	0.72–0.88	0–0.93	0.11–0.25 (0.16)	0.80–2.3 (1.6)	≥ 2.3[b] (8.0)
16	0.78–0.80	0.75–0.85	0–0.95	0.21–0.49 (0.32)	2.4–4.8 (3.2)	≥ 2.8[b] (16)

[a]The true value for S^2 is 0.8, the true values for τ_e are in parentheses, and the magnetic field strength ω_H = 500 MHz.
[b]Defined within $5\tau_m$.

Similar to the results for the tumbling time of 2 ns, improving the precision in R_2 would be most effective in reducing the size of the (S^2, τ_e) solution space. As τ_m increases further (Figure 7c, d), the R_1 value approaches R_1^0 (τ_m = 8 ns) or in fact exceeds R_1^0 (τ_m = 16 ns). As discussed above, R_1 provides very little information about S^2 or τ_e when $R_1 \approx R_1^0$, thus it is not surprising that the allowed region is not significantly restricted by the R_1 measurement. While in this motional regime ($\tau_e/\tau_m = 0.2$) it is observed that the R_1 data is the least constraining of the three measurements, the R_2 and NOE data alone produce two disjoint allowed regions, and R_1 data is required to eliminate this ambiguity. It should be noted that while R_1 and R_2 alone are generally sufficient to obtain the maximal precision in S^2, the NOE data is crucial in reducing the size of the allowed τ_e region.

For very slow internal motions ($\tau_e/\tau_m = 1.0$), the (S^2, τ_e) solution space is quite large with the simulated 5% uncertainties in the relaxation data used here (Figure 8 and Table 3). This is a general consequence of the fact that the relaxation contours become sparse as τ_e approches or exceeds τ_m (Figure 1). However, with sufficient improvement in the precision of the relaxation measurements, it is possible to obtain an acceptable level of precision in S^2 and τ_e even for motions in $\tau_e \approx \tau_m$ regime. In particular, improved precision of the R_2 data is most beneficial, especially for a more precise determination of S^2.

We can summarize the above results for analyzing protein motions which are *not* in the motional narrowing limit in terms of a few empirical 'rules'. In general, precise measurements of R_2 and NOE are most useful for the widest range of internal protein motions (S^2, τ_e). The uncertainty in S^2 is mostly related to the precision in the R_2 measurement. The uncertainty in τ_e is mostly related to the precision in the NOE measurement. These 'rules' follow directly from the shape of the R_1, R_2, and NOE contours seen in Figures 2, 4, 5. For cases where the combination of the R_2 and NOE data results in two disjoint (S^2, τ_e) regions, R_1 measurement (usually even with a crude precision) are also required in order to eliminate one of the disjoint solution regions. In general, these 'rules' are useful throughout the range of the tumbling times we have considered ($2 \leq \tau_m \leq 16$ ns). We repeated the same analysis at higher field ($\omega_H = 750$ MHz) and we found that these general 'rules' still apply.

Application of the graphical method to experimental data

We can take advantages of the above 'rules' in experimental work to increase the efficiency of data collection. In using these methods it is crucial to make reasonable estimates in the uncertainties of the relaxation measurements. For R_1 and R_2 measurements this is best done using Monte Carlo estimates of the uncertainties in decay-curve fitting, based on duplicate measurements of some decay points (Palmer III et al., 1991; Mandel et al., 1995), or rms noise measured in baseline regions of the spectra (Palmer III et al., 1991; Mandel et al., 1995). For NOE measurements, baseline rms noise can be used to estimate uncertainties in peak intensities, and these uncertainties can be propagated through the calculations of NOEs in the standard manner (Li and Montelione, 1995).

After completing a set of relaxation measurements and computing the values and uncertainties of the relaxation parameters, one can apply this graphical analysis procedure using a cutoff of one standard deviation in estimated experimental uncertainties to estimate the allowed (S^2, τ_e) solution space. If the range of S^2 and τ_e values is reasonably small the results can be regarded as final. If the range of S^2 and/or τ_e values is large, and if better definition of the (S^2, τ_e) solution space is desired, one or more measurements need to be repeated with higher precision. By examining each type of relaxation data using the con-

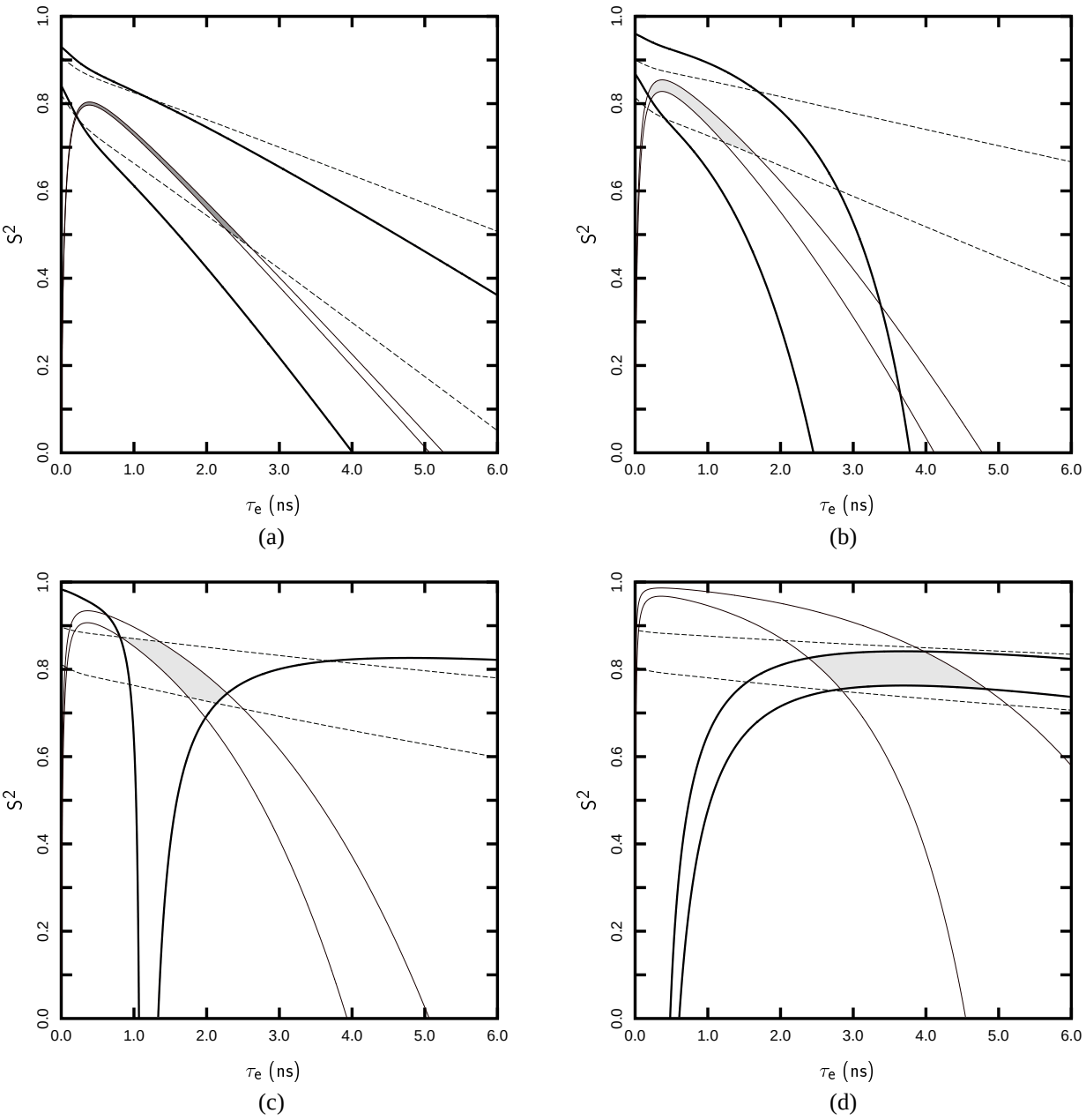

Figure 7. The allowed (S^2, τ_e) solution space for intermediate time scale internal motions obtained from simulated relaxation data ($\tau_e/\tau_m = 0.2$, see Table 3): contours of R_1, R_2, and NOE are plotted in thick, dashed, and thin lines, respectively. The areas corresponding to the allowed (S^2, τ_e) solution space are shaded in gray and indicated by an arrow when they are very small. (a) $\tau_m = 2$ ns: $R_1 = 2.587 \pm 0.129$ s^{-1}, $R_2 = 3.521 \pm 0.176$ s^{-1}, NOE $= 0.123 \pm 0.0062$; (b) $\tau_m = 4$ ns: $R_1 = 2.755 \pm 0.138$ s^{-1}, $R_2 = 5.680 \pm 0.284$ s^{-1}, NOE $= 0.465 \pm 0.023$; (c) $\tau_m = 8$ ns: $R_1 = 2.162 \pm 0.108$ s^{-1}, $R_2 = 9.58 \pm 0.491$ s^{-1}, NOE $= 0.620 \pm 0.031$; (d) $\tau_m = 16$ ns: $R_1 = 1.548 \pm 0.0774$ s^{-1}, $R_2 = 17.72 \pm 0.886$ s^{-1}, NOE $= 0.726 \pm 0.036$. A magnetic field strength of $\omega_H = 500$ MHz is used here.

tour plots and applying the 'rules' mentioned above, it is possible to determine which measurements need to be made more precisely in order to most efficiently improve the precision in the estimate of the motional parameters for a particular residue. Such an analysis would be helpful in planning further relaxation experiments when the precision in the model-free parameters could be improved by a judicious choice of additional

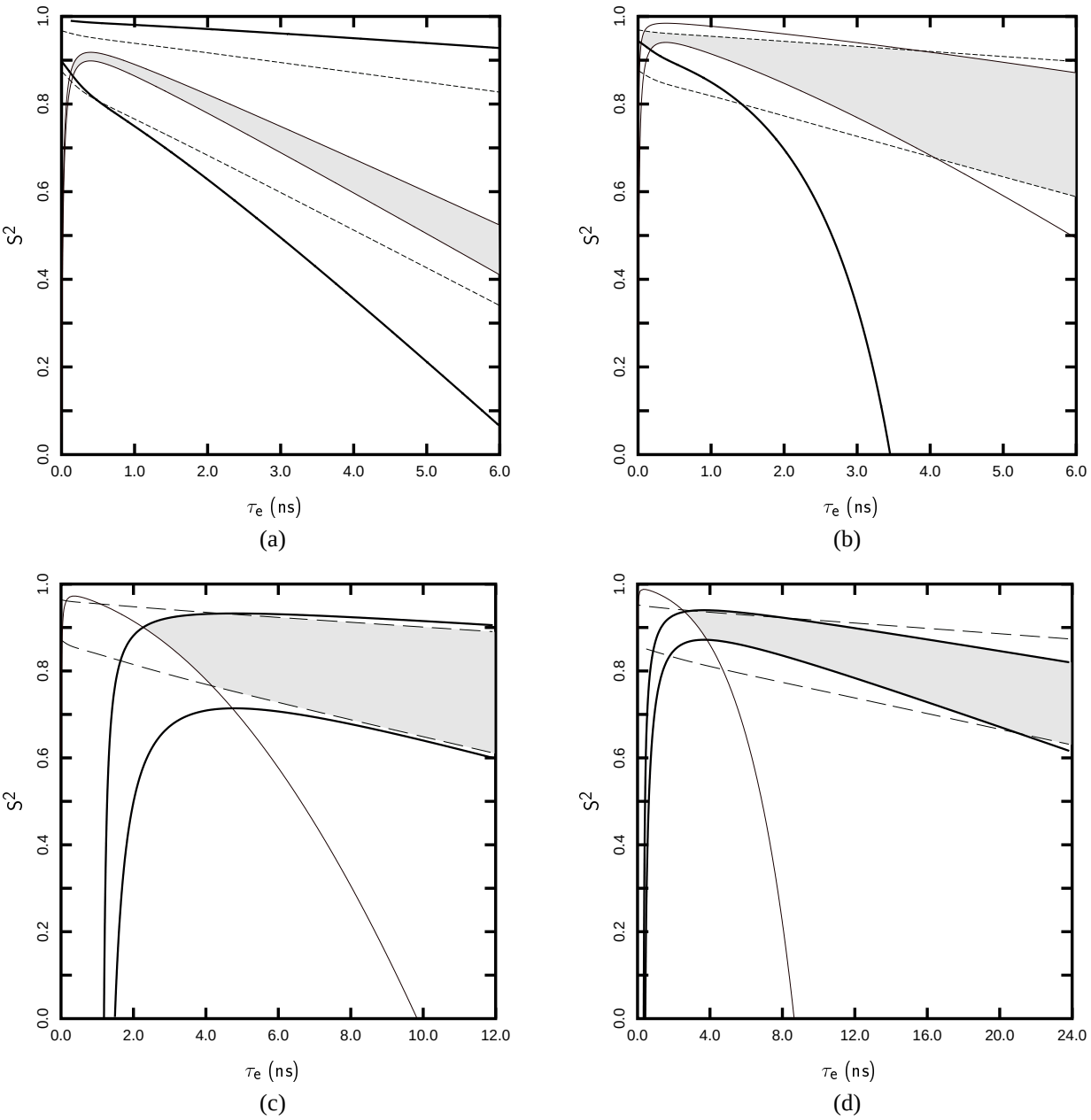

Figure 8. The allowed (S^2, τ_e) solution space for slow time scale internal motions obtained from simulated relaxation data ($\tau_e/\tau_m = 1.0$, see Table III). Contours of R_1, R_2, and NOE are plotted in thick, dashed, and thin lines, respectively. The areas corresponding to the allowed (S^2, τ_e) solution space are shaded in gray and indicated by an arrow when they are very small. (a) $\tau_m = 2$ ns: $R_1 = 2.760 \pm 0.138$ s^{-1}, $R_2 = 3.755 \pm 0.188$ s^{-1}, NOE $= 0.299 \pm 0.015$; (b) $\tau_m = 4$ ns: $R_1 = 2.994 \pm 0.150$ s^{-1}, $R_2 = 6.114 \pm 0.306$ s^{-1}, NOE $= 0.662 \pm 0.033$; (c) $\tau_m = 8$ ns: $R_1 = 2.273 \pm 0.114$ s^{-1}, $R_2 = 10.30 \pm 0.515$ s^{-1}, NOE $= 0.775 \pm 0.039$; (d) $\tau_m = 16$ ns: $R_1 = 1.344 \pm 0.0672$ s^{-1}, $R_2 = 18.97 \pm 0.95$ s^{-1}, NOE $= 0.809 \pm 0.040$. A magnetic field strength of $\omega_H = 500$ MHz is used here.

relaxation measurements; this analysis will be especially useful when the number of residues with large (S^2, τ_e) solution spaces is small, or when there are certain key residues whose dynamics are of particular interest.

We use the ^{15}N relaxation data of Asn^6 and Tyr^{38} of hTGFα discussed previously (Li and Montelione,

1995; Jin et al., 1997) as examples. The (S^2, τ_e) solution space of Asn^6 is quite large – S^2 varies between 0.56 and 0.85 while τ_e varies between 0.1 and 2.3 ns (Figure 9a). The original R_1 relaxation measurement for Asn^6 (2.61 ± 0.71 s^{-1}) has a relative error of more than 25%, which is much greater than that of the R_2 relaxation measurement (5.53 ± 0.58 s^{-1}) (about 10%). Naively, one might think that in this situation an improvement in the precision of R_1 relaxation data would be most useful. However, the simulated results presented in Figure 9b show that even a fourfold increase in the precision of the R_1 measurement offers very little improvement in the determination of the allowed (S^2, τ_e) space. If instead we improve the precision of the R_2 measurement by a factor of 2, we obtain a significant improvement of the range of S^2 (Figure 9c): S^2 then varies between 0.71 and 0.85.

The allowed (S^2, τ_e) solution space of Tyr^{38} is so large that little useful information about the motion can be extracted (Figure 10a). This portion of the hTGFα molecule is within the receptor-binding epitope (Li and Montelione, 1995), and is critical to its biological function. According to the 'rules' described above, since the NOE measurement is already very precise (less than 2% error) and is most restricting (Figure 10a), additional measurements should focus on improving the R_1 and/or R_2 measurement(s). A moderate increase in the precision of the R_2 measurement results in little improvement in the solution space (Figure 10b), while a larger increase in precision of the R_2 measurement results in two disjoint allowed (S^2, τ_e) regions (Figure 10c). In this situation, a moderate increase in the precision of the R_1 measurement eliminates one of the two disjoint (S^2, τ_e) regions and provides a significant improvement in the precision of the allowed (S^2, τ_e) space (Figure 10d). It is interesting to note that in this case improving the R_1 measurement would be sufficient to greatly improve the precision of the estimates of motional parameters.

Based on our experience of applying the graphical analysis to many relaxation data sets available in the literature (Kördel et al., 1992; Li and Montelione, 1995; Yamasaki et al., 1995; Mandel et al., 1996; Stivers et al., 1996; Epstein et al., 1995; Williams et al., 1996; Liu et al., 1996; Cai et al., 1996; Zhou et al., 1996; Hodsdon and Cistola, 1997; Farrow et al., 1997; Papavoine et al., 1997), we suggest that given the restriction of working at a single magnetic field strength, all three (R_1, R_2, and NOE) relaxation parameters should be measured for each site, and that the R_2 and NOE measurements must have high precision ($\leq$ 5%) in order to obtain a good estimate of the model-free parameters while a less precise R_1 measurement is generally sufficient. This proposal is based on the assumption that the chemical exchange contribution to R_2 can be determined precisely.

Internal motions on timescales close to or slower than the overall tumbling time

In their original paper (Lipari and Szabo, 1982), Lipari and Szabo derived the model-free expression for the spectral density function (Equation 11). It should be recognized that Equation 11 makes no assumptions about the relative magnitudes of τ_e and τ_m; the only assumptions are that the overall and internal motions are uncoupled and the correlation functions corresponding to these motions decay as single exponentials. Despite popular misconceptions, under these assumptions, Equation 11 is also valid when $\tau_e \geq \tau_m$. However, as τ_e gets much larger than τ_m, the spectral density functions (Equation 11) go to their 'rigid body' limits, and the relaxation data are insentitive to internal motions much slower than the overall tumbling. They then introduced additional simplifications: (a) if the overall motion is considerably slower than the internal motions ($\tau_e \ll \tau_m$), Equation 11 becomes

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_m}{1+(\omega\tau_m)^2} + \frac{(1-S^2)\tau_e}{1+(\omega\tau_e)^2}\right]. \qquad (21)$$

(b) if τ_e is in the extreme narrowing limit ($(\omega\tau_e)^2 \ll 1$, where ω is the largest frequency at which the spectral density must be evaluated) Equation 21 can be further simplified to

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_m}{1+(\omega\tau_m)^2} + (1-S^2)\tau_e\right]. \qquad (22)$$

Furthermore Lipari and Szabo showed that Equation 22 is exact for the more general situation when $C_I(t)$ is expressed as a linear combination of an arbitrary number of exponential decays

$$C_I(t) = S^2 + \sum_{i=1} a_i e^{-t/\tau_i}, \qquad (23)$$

if (1) the overall motion is isotropic, (2) internal motions are much faster than the overall tumbling (i.e. the internal motions are in the extreme narrowing limit), and (3) τ_e is defined to be the area under the correlation function, i.e.

$$\tau_e(1-S^2) = \int_0^\infty (C_I(t) - S^2)dt. \qquad (24)$$

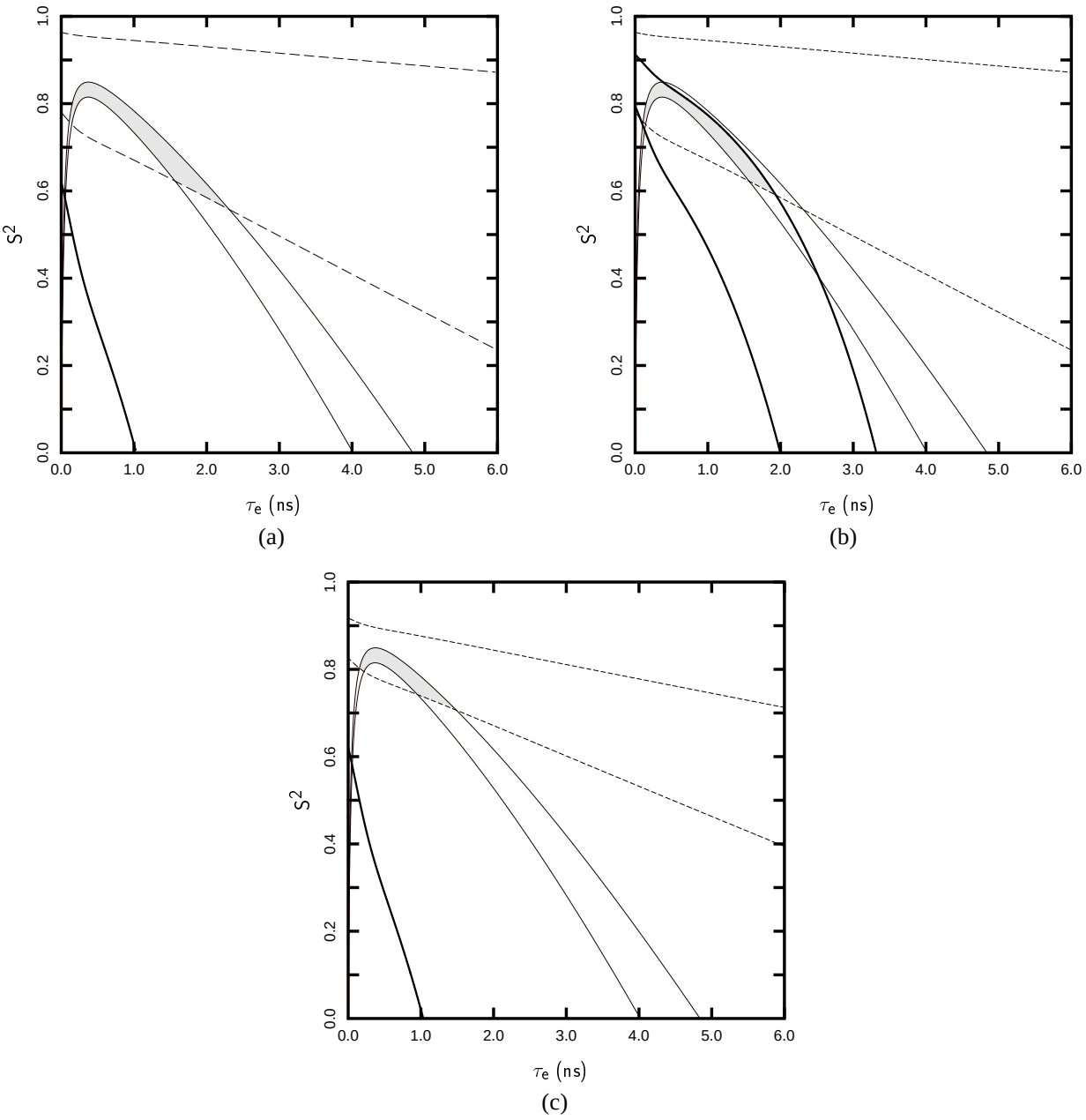

Figure 9. The allowed (S^2, τ_e) space obtained from the experimental relaxation data of residue Asn^6 of hTGFα with improved relaxation measurement ($\tau_m = 3.76$ ns). Contours of R_1, R_2, and NOE are plotted in thick, dashed, and thin lines, respectively. The areas corresponding to the allowed (S^2, τ_e) solution space are shaded in gray and indicated by an arrow when they are very small. (a) Original relaxation data (Li and Montelione, 1995), $R_1 = 2.61 \pm 0.71$ s^{-1}, $R_2 = 5.53 \pm 0.58$ s^{-1}, NOE $= 0.44 \pm 0.03$; (b) fourfold increase in the R_1 precision; (c) twofold increase in the R_2 precision.

Equation 23 also applies in situations in which some $\tau_i > \tau_m$ provided that the order parameter is redefined to exclude the averaging due to those motions.

Lipari and Szabo showed that although Equation 11 was constructed assuming single exponential decay for the internal correlation function, it could still be applied without error to interpret multi-exponential decay provided that the internal motions are either much slower or much faster than the overall tumbling. When the internal motions are close to the tumbling

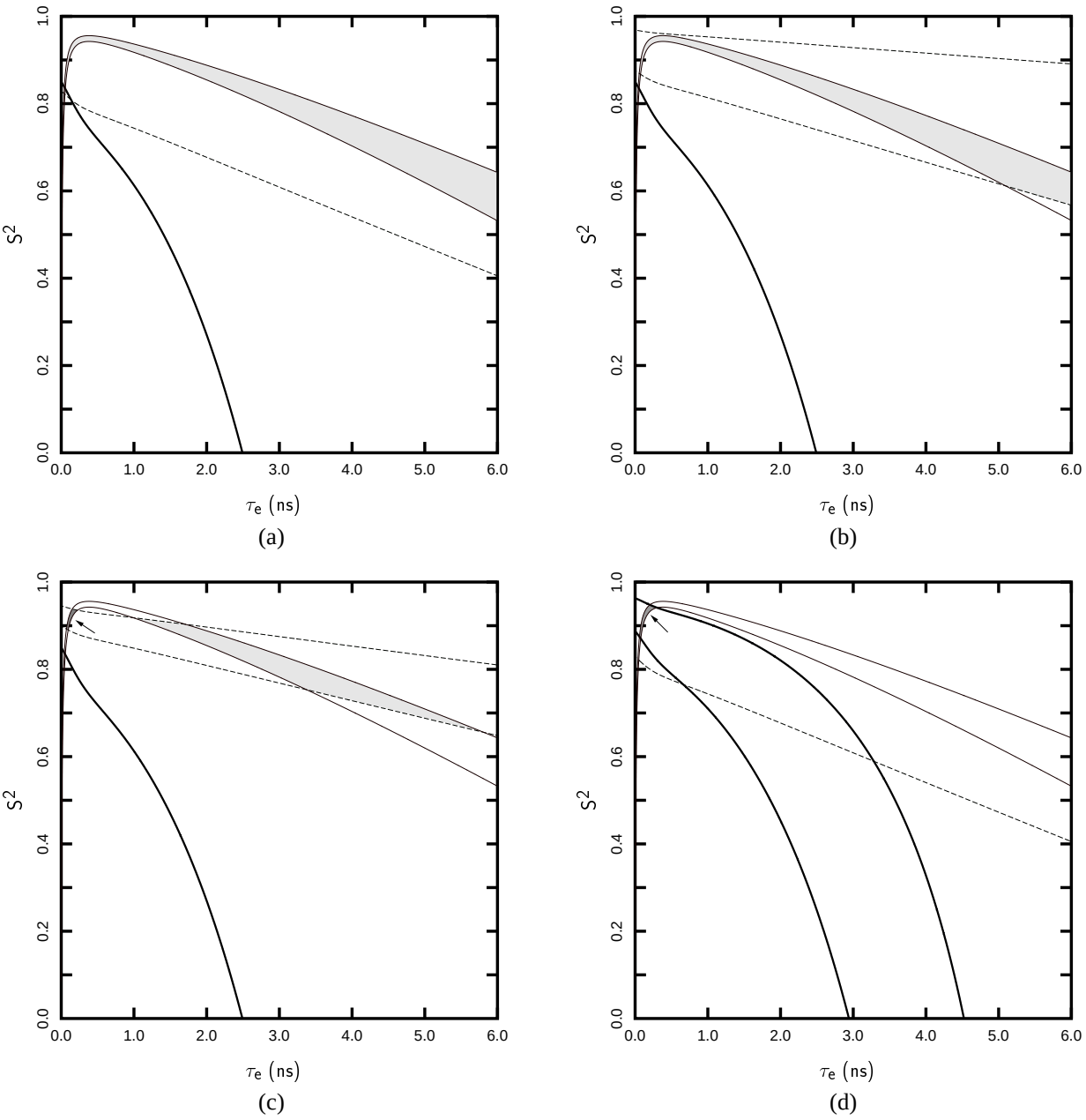

Figure 10. Allowed (S^2, τ_e) space obtained from the experimental relaxation data of residue Tyr[38] of hTGFα with improved relaxation measurement ($\tau_m = 3.76$ ns). Contours of R_1, R_2, and NOE are plotted in thick, dashed, and thin lines, respectively. The areas corresponding to the allowed (S^2, τ_e) solution space are shaded in gray and indicated by an arrow when they are very small. (a) Original relaxation data (Li and Montelione, 1995), $R_1 = 2.83 \pm 0.23$ s^{-1}, $R_2 = 5.85 \pm 0.59$ s^{-1}, NOE $= 0.63 \pm 0.01$; (b) twofold increase in the R_2 precision; (c) fourfold increase in the R_2 precision; (d) twofold increase in the R_1 precision.

time the situation is more complex. This conclusion appears to have led to the assumption by some workers that the model-free formalism is not applicable at all when the correlation function(s) describing the internal motions decay on a time scale approaching that of the overall motions. However, this assumption is incorrect. As long as the overall and internal motions are separable and the correlation function describing

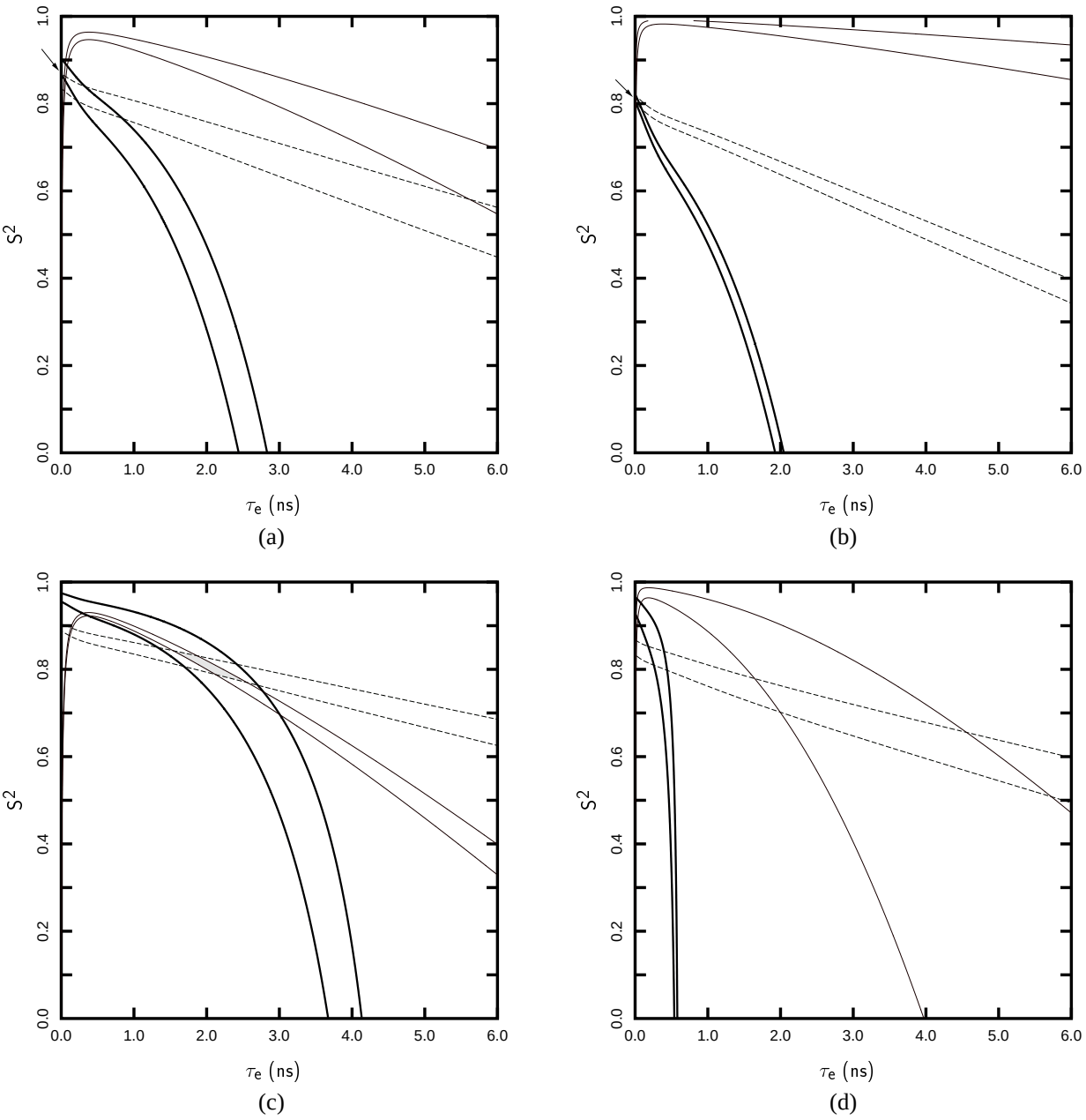

Figure 11. Allowed (S^2, τ_e) solution space obtained from relaxation data with multiple time scale internal motions ($\tau_m = 4$ ns and the relative error = 2%). Relaxation data are simulated using the following inputs: the areas corresponding to the allowed (S^2, τ_e) solution space are shaded in gray and indicated by an arrow when they are very small. Contours of R_1, R_2, and NOE are plotted in thick, dashed, and thin lines, respectively. (a) The extended Lipari-Szabo model (multiple time scale motion), $\tau_f = 10$ ps, $\tau_s = 2$ ns, $S_f^2 = 0.9$, $S_s^2 = 0.9$, $\omega_H = 500$ MHz, (b) the simple Lipari-Szabo model (single time scale motion), $\tau_e = 10$ ps, $S^2 = 0.81$, $\omega_H = 500$ MHz, (c) the simple Lipari-Szabo model (single time scale motion), $\tau_e = 2$ ns, $S^2 = 0.81$, $\omega_H = 500$ MHz, (d) the extended Lipari-Szabo model (multiple time scale motion), $\tau_f = 10$ ps, $\tau_s = 2$ ns, $S_f^2 = 0.9$, $S_s^2 = 0.9$, $\omega_H = 1$ GHz.

the internal motion decays as a single exponential, Equation 11 is exact and may be applied to extract the motional parameters even though the internal motion and overall tumbling time scales are commensurate.

Even when the internal motions cannot strictly be described by a single exponential, the effective order parameter extracted using Equation 11 can sometimes provide information about internal motions on the same time scale as the overall motion. One example corresponds to the use of the extended Lipari-Szabo equation (Clore et al., 1990b) to describe two internal motions, one in the extreme narrowing limit and the other close to the overall motion. For this situation the internal correlation function effectively decays as a single exponential so Equation 11 is again valid. Lipari and Szabo actually considered the problem of detecting internal motions which are slow on the NMR time scale ($(\omega\tau_e)^2 > 1$) as well. They estimated in a qualitative way that order parameters larger than ~ 0.3 could be detected with a precision of $\sim 30\%$. A quantitative examination of this problem requires the kind of analysis reported in this paper. Apart from an examination of the validity of the approximations inherent in the model-free formalism, we can see from the results of the previous sections (Figures 6–8), that when the internal motion is no longer in the extreme narrowing limit, however, very high precision relaxation measurements are required in order to extract any useful information about either the magnitude or the time scale of these slower motions.

Multiple timescale internal motions uncovered by ultra-high field NMR relaxation experiments

This paper has been concerned with the effect of finite precision in the experimental NMR relaxation data on the estimation of the magnitude and time scale of the protein internal motion in cases where the internal motion is predominantly occurring on one time scale. The effects of finite precision in the measurements on the detection of multiple time scale internal motions, an issue related to the simultaneous analysis of relaxation data recorded at multiple field strengths, will be the subject of a separate communication. However, we briefly consider here the advantage of relaxation experiments performed at a single ultra-high field for the detection of two internal motions, one of which is on the picosecond time scale and the other on the nanosecond time scale.

In the original Lipari-Szabo model, only one effective correlation time for internal motion is used. Clore et al. (Clore et al., 1990b) proposed an extension of the Lipari-Szabo model which includes a second correlation time in order to deal with situations where internal motions occur on two distinct timescales which are at least one order of magnitude apart. The spectral density function $J(\omega)$ in the extended model-free formalism is given by

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_m}{1+(\omega\tau_m)^2} + \frac{(1-S_f^2)\tau_f}{1+(\omega\tau_f)^2} + \frac{(S_f^2-S^2)\tau_s}{1+(\omega\tau_s)^2}\right]. \quad (25)$$

What happens when simulated NMR relaxation data is generated using the extended model-free formalism (Equation 25) and interpreted using the single exponential approximation (Equation 11)? To explore this question, we simulated relaxation data that corresponds to motion on two time scales (a fast motion ($\tau_f = 10$ ps) with $S_f^2 = 0.9$ and a much slower motion ($\tau_s = 2$ ns) with $S_s^2 = 0.9$) using Equation 25 ($\tau_m =$ 4 ns and $\omega_H = 500$ MHz). A very small experimental uncertainty (2%) was assumed. We then analyzed the data using the original Lipari-Szabo formalism. Surprisingly, there is a well defined (S^2, τ_e) solution space centered at $S^2 = 0.85$ and $\tau_e = 50$ ps (Figure 11a). These solutions are similar to what would have been obtained from relaxation data generated using the simple Lipari-Szabo model with a single fast internal motion ($S^2 = 0.81$, $\tau_e = 10$ ps; Figure 11b), but very different from what would have been obtained from relaxation data generated with only slow internal motion ($S^2 = 0.81$, $\tau_e = 2$ ns; Figure 11c). Thus, even very high precision relaxation measurements at 500 MHz are sometimes insufficient to detect motions on two time scales. We then repeated this procedure using the same motional parameters with relaxation data simulated at ultra-high field (1 GHz). In these simulations, the results were more encouraging (Figure 11d) – no (S^2, τ_e) solutions exist that are consistent with the single exponential approximation (Equation 11). This implies that higher field spectrometers can be very useful in helping to detect motions on multiple time scales which appear as deviations from the simple Lipari-Szabo model. A systematic investigation of the use of multiple field experiments to probe multiple time scale motions will be the subject of a future communication. It will be desirable to use higher field spectrometers in the future to reduce the ambiguity of the information about internal dynamics of biomolecules that could be extracted from the relaxation measurements.

The fact that a well-defined solution space exists assuming that the motion is occurring on only one time scale when the data was generated from a model

with motion on two distinct time scales in Figure 11a suggests that the model-selection problem may be far more insidious than has been previously appreciated. As is well-known, the 'model-free' formalism as applied in the NMR literature actually consists of a set of nested models which make use of the parameters S^2, τ_e, R_{ex} (a phenomenological exchange contribution to R_2), and S_f^2. Various statistical model-selection procedures, such as significance testing (e.g. Nicholson et al., 1995) and a combination of significance and hypothesis testing (e.g. Mandel et al., 1995; Yamasaki et al., 1995), have been used to determine which subset of the above parameters is needed to adequately describe the relaxation data for a given residue.

These statistical procedures in general do not provide a measure of the amount of evidence in favor of a given model, but rather can only 'accept' or 'reject' a given model. This is problematic, as Figure 11 suggests that in fact there may not be enough information in the three noise-corrupted data points to decisively accept one model in favor of another. The best procedure in such a situation would be to resign oneself to the fact that the choice of model is itself uncertain, and that inferences concerning the parameter values should take into account that uncertainty. Of course, one could invoke some form of Occam's Razor, and choose to consider only the model with the smallest number of parameters which adequately fits the data. However, this provides no guidance in cases where the number of parameters is equal (e.g., (S^2, τ_e, R_{ex}) vs. (S^2, τ_e, S_f^2)). More seriously, the arbitrary selection of one model based upon ambiguous data could result in misleading estimates of the parameter values if the 'best-fit' parameters under the competing models are significantly different.

The situation is further complicated by the fact that the model-selection methods used by Mandel et al. (Mandel et al., 1995) and Yamasaki et al. (Yamasaki et al., 1995) implicitly assume that the overall correlation time τ_m is known precisely. In fact, Korzhnev et al. (Korzhnev et al., 1997) have recently shown that if the bulk of the residues in a protein exhibit slow time scale motion, then the traditional method will mis-estimate τ_m and will result in an incorrect interpretation of the relaxation data. Furthermore, it is well known that even small changes in the assumed value of τ_m can change which model is selected for a given residue (Li and Montelione, 1995). Thus, uncertainty in the estimate of τ_m leads directly to an uncertainty in the model selection.

Unfortunately, the classical statistical methodology used by Mandel et al. (Mandel et al., 1995) and Yamasaki et al. (Yamasaki et al., 1995) does not provide the theoretical machinery needed to adequately deal with model uncertainty. Bayesian statistical methods, on the other hand, are far better equipped to assess the weight of evidence in favor of a model given a set of data, and allow for model uncertainty (Bretthorst, 1990b; Kass and Raftery, 1995). In particular, one may calculate the marginal likelihood of the data given model M_i, which is defined to be

$$P(D|M_i) = \int P(D|\Theta_i)P(\Theta_i)d\Theta_i, \tag{26}$$

where Θ_i is the vector of parameters under the i-th model, $P(D|\Theta_i)$ is the likelihood of the data, and $P(\Theta_i)$ is the prior probability of Θ_i. $P(D|M_i)$ can be interpreted as the weight of evidence in favor of model M_i, and the ratio

$$B_{ij} = \frac{P(D|M_i)}{P(D|M_j)}, \tag{27}$$

known as the 'Bayes factor', is an estimate of the relative odds in favor of M_i vs. M_j (Bretthorst, 1990b; Kass and Raftery, 1995). Development of a practical Bayesian model selection strategy for the the analysis of relaxation data which incorporates uncertainty in τ_m is currently underway in our laboratories.

Conclusions and future directions

The graphical procedure used for the analysis of relaxation data allows one to directly visualize the impact of experimental uncertainties on the precision of estimated model-free parameters. Although the relationship between the precision of relaxation measurements and the precision of (S^2, τ_e) parameters is far from being simple due to its dependence on τ_m, τ_e, and R_1/R_1^0, nevertheless we find that, in general, in order to obtain a good estimate of the motional parameters, one may want to strive for the highest precision in the R_2 and NOE measurements at the expense of a fair precision in the R_1 measurement. The widely employed assumption that the effective correlation time τ_e is required to be significantly faster than the overall tumbling time τ_m in the context of the model-free formalism is not a prior condition for the applicability of the model-free formalism, but a consequence of the limited precision in the NMR relaxation data.

Analysis of relaxation data simulated using multiple time scale internal motions indicates that at typical

field strengths (i.e., 500 MHz) even very high precision relaxation measurements are sometimes not sufficient to detect motions on two time scales, whereas the detection can be achieved at very high field strengths (e.g., 1 GHz). This suggests that the development of ultra-high field (> 800 MHz) spectrometers will improve our ability to characterize protein dynamics on multiple time scales. Finally, we have briefly discussed the advantages of the Bayesian statistical methods over the classical statistical methods presently used for the model-selection problem.

In the analysis presented in this paper, we have assumed that the overall tumbling time τ_m is known with high precision and negligible systematic error. Korzhnev et al. (Korzhnev et al., 1997) have recently shown that the τ_m (often obtained from the ratio R_2/R_1) will be underestimated if the majority of the molecule undergoes internal motions that are not in the motional narrowing limit. They also noted that there are significant discrepancies among the reported rotational correlation times for the same protein obtained from different experiments. For example, values of τ_m varying between 7 ns and 10 ns have been reported for lysozyme using different experimental techniques. The analysis of internal motions in proteins which are not in the motional narrowing limit depends upon an analysis of the overall macromolecular tumbling in ways which are now becoming better understood. In this context we note that several recent studies have demonstrated that even a small amount of rotational anisotropy can result in a misinterpretation of the relaxation data (Schurr et al., 1994; Tjandra et al., 1996b; Luginbühl et al., 1997). A more general analysis of protein dynamics based on the Lipari-Szabo model-free approach must consider both the precision and accuracy in the overall tumbling parameters and its effect on model selection procedure. This will result in an increase in the number of fitting parameters which will require additional relaxation measurements if the problem is not to be underdetermined. We are currently developing general strategies for collecting such data and an appropriate framework for their analysis based on Bayesian statistical methods.

Acknowledgements

We thank Francisco Figueirido for insightful discussions and providing his contour plotting software. This research was supported by grants from the National Institutes of Health GM-30580 to R.M.L. and GM-50733 to G.T.M.

References

Abragam, A. (1961) *The Principles of Nuclear Magnetism*, The International Series of Monographs on Physics, Oxford, Oxford University Press.

Boyd, J., Hommel, U. and Campbell, I. (1990) *Chem. Phys. Lett.*, **175**, 477–482.

Brainard, J.R. and Szabo, A. (1981) *Biochemistry*, **20**, 4618–4628.

Bretthorst, G.L. (1990a) *J. Magn. Reson.*, **88**, 533–551.

Bretthorst, G.L. (1990b) *J. Magn. Reson.*, **88**, 552–570.

Cai, M., Huang, Y., Prakash, O., Wen, L., Dunkelbarger, S.P., Huang, J.-K., Liu, J. and Krishnamoorthi, R. (1996) *Biochemistry*, **35**, 4784–4794.

Clore, G.M., Driscoll, P.C., Wingfield, P.T. and Gronenborn, A.M. (1990a) *Biochemistry*, **29**, 7387–7401.

Clore, G.M., Szabo, A., Bax, A., Kay, L.E., Driscoll, P.C. and Gronenborn, A.M. (1990b) *J. Am. Chem. Soc.*, **112**, 4989–4991.

Combettes, P.L. (1993) *Proc. IEEE*, **81**, 182–208.

Epstein, D.M., Benkovic, S.J. and Wright, P.E. (1995) *Biochemistry*, **34**, 11037–11048.

Farrow, N.A., Zhang, O., Forman-Kay, J.D. and Kay, L.E. (1995) *Biochemistry*, **34**, 868–878.

Farrow, N.A., Zhang, O., Forman-Kay, J.D. and Kay, L.E. (1997) *Biochemistry*, **36**, 2390–2402.

Fushman, D., Weisemann, R., Thüring, H. and Rüterjans, H. (1994) *J. Biomol. NMR*, **4**, 61–78.

Henry, G.D., Weiner, J.H. and Sykes, B.D. (1986) *Biochemistry*, **25**, 590–598.

Hodsdon, M.E. and Cistola, D.P. (1997) *Biochemistry*, **36**, 2278–2290.

Jin, D., Figueirido, F., Montelione, G.T. and Levy, R.M. (1997) *J. Am. Chem. Soc.*, **119**, 6923–6924.

Kass, R.E. and Raftery, A.E. (1995) *J. Am. Stat. Assoc.*, **90**, 773–795.

Kay, L.E., Torchia, D.A. and Bax, A. (1989) *Biochemistry*, **28**, 8972–8979.

Kinosita, K., Kawato, S. and Ikegami, A. (1977) *Biophys. J.*, **20**, 289.

Kördel, J., Skelton, N.J., Akke, M., Palmer III, A.G. and Chazin, W.J. (1992) *Biochemistry*, **31**, 4856–4866.

Korzhnev, D.M., Orekhov, V.Y. and Arseniev, A.S. (1997) *J. Magn. Reson.*, **127**, 184–191.

Levy, R.M., Karplus, M. and McCammon, J.A. (1981) *J. Am. Chem. Soc.*, **103**, 994–996.

Li, Y. and Montelione, G.T. (1995) *Biochemistry*, **34**, 2408–2423.

Lipari, G. and Szabo, A. (1982) *J. Am. Chem. Soc.*, **104**, 4546–4559.

Liu, J., Prakash, O., Cai, M., Gong, Y., Wen, L., Wen, J.J., Huang, J.-K. and Krishnamoorthi, R. (1996) *Biochemistry*, **35**, 1516–1524.

London, R.E. (1980) *Magnetic Resonance in Biology*, Vol. I, John Wiley & Sons, New York, NY.

Luginbühl, P., Pervushin, K.V., Iwai, H. and Wüthrich, K. (1997) *Biochemistry*, **36**, 7305–7312.

Mandel, A.M., Akke, M. and Palmer III, A.G. (1995) *J. Mol. Biol.*, **246**, 144–163.

Mandel, A.M., Akke, M. and Palmer III, A.G. (1996) *Biochemistry*, **35**, 16009–16023.

Nicholson, L.K., Yamazaki, T., Torchia, D.A., Grzesiek, S., Bax, A., Stahl, S.J., Kaufman, J.D., Wingfield, P.T., Lam, P.Y.S., Jadhav,

P.K., Hodge, C.N., Domaille, P.J. and Chang, C.-H. (1995) *Nat. Struct. Biol.*, **2**, 274–280.
Orekhov, V.Y., Pervushin, K.V. and Arseniev, A.S. (1994) *Eur. J. Biochem.*, **219**, 887–896.
Palmer III, A.G., Rance, M. and Wright, P.E. (1991) *J. Am. Chem. Soc.*, **113**, 4371–4380.
Papavoine, C.H.M., Remerowski, M.L., Horstink, L.M., Konings, R.N.H., Hilbers, C.W. and van de Ven, F.J.M. (1997) *Biochemistry*, **36**, 4015–4026.
Peng, J.-W. and Wagner, G. (1992) *Biochemistry*, **31**, 8571–8586.
Schneider, D.M., Dellwo, M.J. and Wand, A.J. (1992) *Biochemistry*, **31**, 3645–3652.
Schurr, J.M., Babcock, H.P. and Fujimoto, B.S. (1994) *J. Magn. Reson. B.*, **105**, 211–224.
Stivers, J.T., Abeygunawardana, C. and Mildvan, A.S. (1996) *Biochemistry*, **35**, 16036–16047.
Stone, M.J., Fairbother, W.J., Palmer III, A.G., Reizer, J., Saier, M.H. and Wright, P.E. (1992) *Biochemistry*, **31**, 4394–4406.
Tjandra, N., Szabo, A. and Bax, A. (1996a) *J. Am. Chem. Soc.*, **118**, 6986–6991.
Tjandra, N., Wingfield, P., Stahl, S. and Bax, A. (1996b) *J. Biomol. NMR*, **8**, 273–284.
Wallach, D.J. (1967) *J. Chem. Phys.*, **47**, 3258.
Weaver, A.J., Kemple, M.D. and Prendergast, F.G. (1988) *Biophys. J.*, **54**, 1–15.
Williams, K.A., Farrow, N.A., Deber, C.M. and Kay, L.E. (1996) *Biochemistry*, **35**, 5145–5157.
Wittebort, R.J. and Szabo, A. (1978) *J. Chem. Phys.*, **69**, 1722.
Woessner, D.E. (1962) *J. Chem. Phys.*, **37**, 647–654.
Yamasaki, K., Saito, M., Oobatake, M. and Kanaya, S. (1995) *Biochemistry*, **34**, 6587–6601.
Zhou, H., McEvoy, M.M., Lowry, D.F., Swanson, R.V., Simon, M.I. and Dahlquist, F.W. (1996) *Biochemistry*, **35**, 433–443.