JOURNAL OF MAGNETIC RESONANCE **125,** 358–363 (1997)
ARTICLE NO. MN971129

COMMUNICATIONS

Measurement of Relaxation Rates from 2D Spectra with Partial Overlap

PRASANNA MISHRA, ZSOLT ZOLNAI,* NENAD JURANIĆ, AND SLOBODAN MACURA†

Department of Biochemistry and Molecular Biology, Mayo Graduate School, Mayo Clinic and Foundation, Rochester, Minnesota 55905

Received August 21, 1996; revised January 13, 1997

Heteronuclear longitudinal and transversal relaxation rates are principal sources of information on molecular dynamics in macromolecular solutions (*1–6*). They became usable after the methods for proton-detected heteronuclear spectroscopy were developed (*7–9*). Large random errors due to poor sensitivity used to be a major obstacle that prevented a reliable analysis of collected data. With proton detection and isotope labeling, the sensitivity has improved several orders of magnitude, and random errors in relaxation rates are reduced sometimes even below 1% (*10*). Despite many achievements in eliminating the other sources of errors [cross-correlation effect (*11*), antiphase contribution (*12*), etc.], the reported data may still contain *systematic* errors that are difficult to identify and correct. This is partly due to the relatively complex data processing and dependence of the results on the chosen motional model. The large number of parameters needed to fit the experimental data to the selected motional model can easily absorb systematic errors which grossly distorts the calculated dynamic parameters.

Spectral overlap is a notorious problem in spectroscopy, and much effort has been devoted to its reduction and elimination. In the present context, the spectral overlap is eliminated using multidimensional techniques. However, the overlap might be a problem even for visually well-resolved peaks when quantitative interpretation is attempted. The tails of spectral lines may extend several linewidths away from the center of the line contributing to the intensity of neighboring lines. This effect might be enhanced by small phase errors which introduce far-reaching dispersive tails as well. When a quantification of intensities with accuracy of a fraction of percent is expected (*10*), spectral overlap may become a significant source of error. Due to spectral overlap, the dependence of peak intensity I on the delay time t becomes multiexponential:

$$I_j^{\exp}(t) = I_j^0[\exp(-R_j^{\mathrm{ac}}t) + \sum_{i\neq j} \varepsilon_i \exp(-R_i^{\mathrm{ac}}t)]. \quad [1]$$

Here R_j^{ac} is the actual value of relaxation rate at spin site j, and ε_i is the fraction of intensity from the neighboring peak i that contributes to the intensity of peak j. For significant overlap and large dispersion in relaxation rates ($R_i^{\mathrm{ac}} \ll R_j^{\mathrm{ac}}$ or $R_i^{\mathrm{ac}} \gg R_j^{\mathrm{ac}}$), the deviation from monoexponential behavior can be easily detected and possibly taken into account. However, for small overlap or for close relaxation rates ($R_i^{\mathrm{ac}} \approx R_j^{\mathrm{ac}}$), the deviation from monoexponential decay may not be noticed, and the calculated relaxation rate R_j^{fit} may contain a systematic error.

In practice, dependence of peak intensity on delay time is fitted to a monoexponential function

$$I_j^{\exp}(t) \approx I_j^0(1 + \sum_{i\neq j} \varepsilon_i)\exp(-R_j^{\mathrm{ac}}t). \quad [2]$$

For consistency, Eq. [2] is derived from Eq. [1] by setting $R_i^{\mathrm{ac}} = R_j^{\mathrm{ac}}$. In real systems, the relaxation rates R_i^{ac} are spread over a range of values, and Eq. [2] may become a poor approximation of Eq. [1] even for small ε_i. Hence, the experimental multiexponential decay fitted to a monoexponential function yields a relaxation-rate constant R_j^{fit} that is different from the actual relaxation-rate constant R_j^{ac}. This implies that vector $\mathbf{R}^{\mathrm{fit}}$ (that comprises all the R values from a given sample) obtained by monoexponential fitting will be, in general, different from the actual vector of relaxation rates $\mathbf{R}^{\mathrm{ac}}$. In the first approximation, difference between the fitted relaxation rate R_j^{fit} and the actual relaxation rate R_j^{ac} is a linear function of spectral overlap ε_i. One can easily deduce from Eqs. [1] and [2] that

$$R_j^{\mathrm{fit}} - R_j^{\mathrm{ac}} = \sum_{i\neq j} \varepsilon_i (R_i^{\mathrm{ac}} - R_j^{\mathrm{fit}}). \quad [3]$$

If the overlap factors on the right-hand side of Eq. [3] are known, then the fitted relaxation-rate constants can be cor-

* On leave from Mathematical Institute, 11000 Beograd, Knez Mihailova 35, Yugoslavia.

† To whom correspondence should be addressed.

1090-7807/97 $25.00

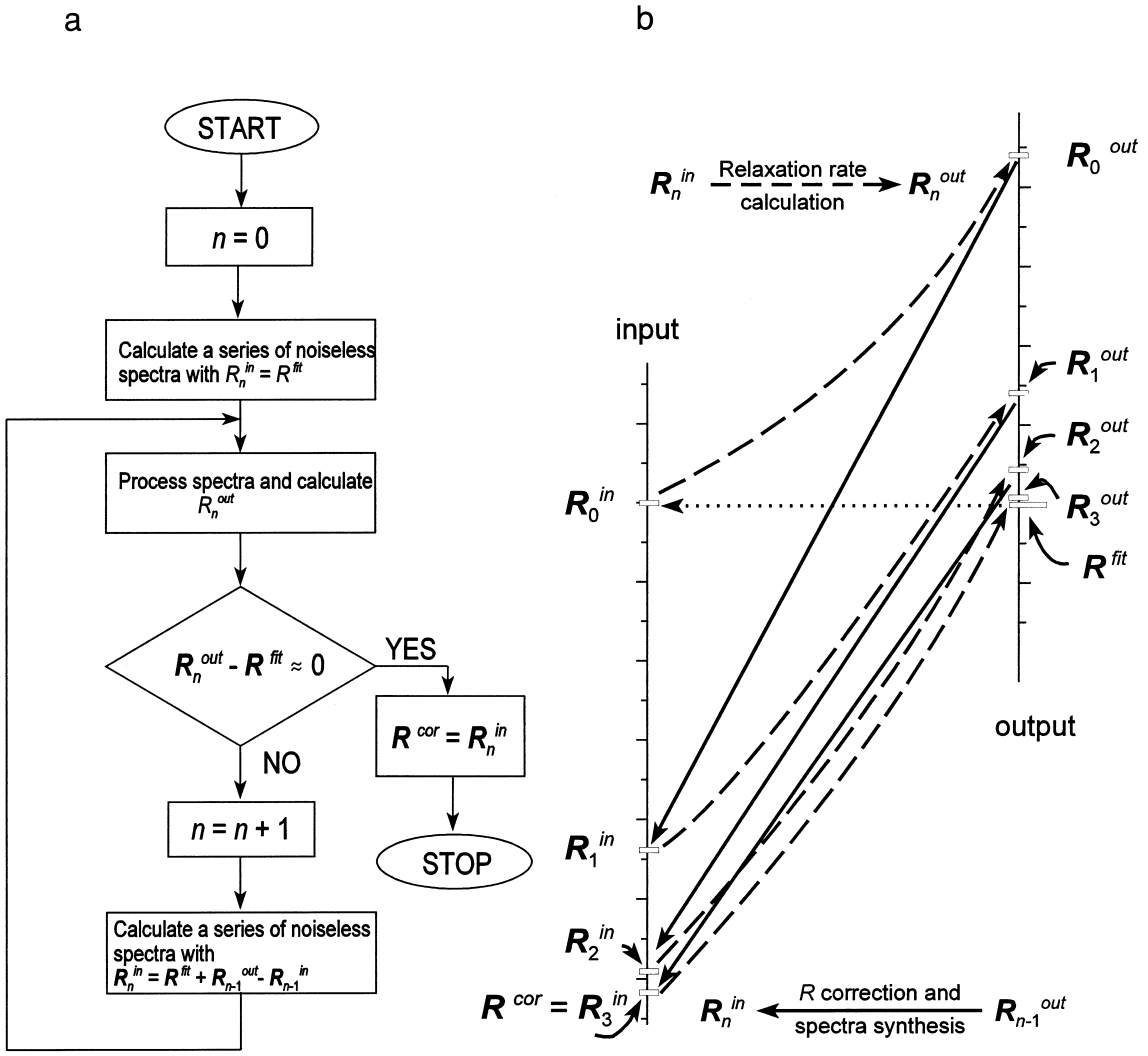

FIG. 1. Schematic representation of the procedure for identifying and correcting the systematic errors due to spectral overlap: (a) Flowchart. (b) Relationship between the input and the output relaxation values during iterations. First, the best experimental values $\mathbf{R}^{\text{fit}}$ are used to generate a series of noiseless spectra. Hence, the input relaxation rates are the same as the experimental values, $\mathbf{R}_0^{\text{in}} = \mathbf{R}^{\text{fit}}$ [dotted arrow (in b)]. The relaxation-rate calculation from vector $\mathbf{R}_0^{\text{in}}$ generates a relaxation-rate vector $\mathbf{R}_0^{\text{out}}$. Due to the spectral overlap, $\mathbf{R}_0^{\text{out}} \neq \mathbf{R}^{\text{fit}}$ and thus the experimental data vector $\mathbf{R}^{\text{fit}}$ is different from the actual relaxation-rate vector. The best approximation of the actual relaxation-rate vector is that which produces an output as close as possible to the fitted experimental data vector $\mathbf{R}^{\text{fit}}$. Hence, the new input $\mathbf{R}_1^{\text{in}}$ is obtained by correcting the previous input $\mathbf{R}_0^{\text{in}}$ according to Eq. [7]. Repeating the same procedure, after n iterations, a relaxation-rate vector $\mathbf{R}_n^{\text{in}}$ is found with corresponding output vector $\mathbf{R}_n^{\text{out}} \simeq \mathbf{R}^{\text{fit}}$. Then, $\mathbf{R}_n^{\text{in}}$ represents the best approximation of the actual relaxation-rate vector.

rected in a single step. If they are not known, the correction can be performed iteratively.

Figure 1a shows the flowchart of an iterative procedure. First, a noiseless set of spectra is generated using experimentally obtained relaxation rates $\mathbf{R}^{\text{fit}}$, spectral linewidths, intensities, and positions. Repeating the same processing on the artificial spectra as on the experimental spectra, a new vector of relaxation rates $\mathbf{R}^{\text{out}}$ is generated. When the effect of overlap is negligibly small, the difference between the input ($\mathbf{R}^{\text{in}} = \mathbf{R}^{\text{fit}}$) and the output relaxation-rate vectors is also negligibly small, $\mathbf{R}^{\text{in}} - \mathbf{R}^{\text{out}} \approx 0$, and no correction is needed. However, in the presence of overlap, the two vectors become different, $\mathbf{R}^{\text{in}} - \mathbf{R}^{\text{out}} \neq 0$, and the input vector must be corrected. The main goal of the correction procedure is to find an input vector $\mathbf{R}_s^{\text{in}}$ that produces an output vector $\mathbf{R}_s^{\text{out}}$ as close as possible to the target vector $\mathbf{R}_s^{\text{fit}}$ obtained by fitting the original experimental data. Thus, $\mathbf{R}_s^{\text{in}} = \mathbf{R}^{\text{ac}}$ implies that $\mathbf{R}_s^{\text{out}} = \mathbf{R}^{\text{fit}}$. Taking a difference of the two equations, we obtain

$$\mathbf{R}_s^{\text{out}} - \mathbf{R}_s^{\text{in}} = \mathbf{R}^{\text{fit}} - \mathbf{R}_s^{\text{ac}} \qquad [4]$$

and

$$\mathbf{R}^{\text{ac}} = \mathbf{R}^{\text{fit}} + \mathbf{R}_s^{\text{in}} - \mathbf{R}_s^{\text{out}}. \qquad [5]$$

Equation [5] cannot be directly solved since $\mathbf{R}_s^{\text{in}}$ is not known. For an arbitrary vector $\mathbf{R}_0^{\text{in}}$, Eq. [5] generates a corrected vector $\mathbf{R}^{\text{cor}}$,

$$\mathbf{R}^{\text{cor}} = \mathbf{R}^{\text{fit}} + \mathbf{R}_0^{\text{in}} - \mathbf{R}_0^{\text{out}}, \qquad [6]$$

that is slightly closer to $\mathbf{R}^{\text{ac}}$ than the input vector $\mathbf{R}_0^{\text{in}}$. This new vector is used as an input for the next iteration $\mathbf{R}_1^{\text{in}} = \mathbf{R}^{\text{cor}}$. In general, the input for nth iteration is

$$\mathbf{R}_n^{\text{in}} = \mathbf{R}^{\text{fit}} + \mathbf{R}_{n-1}^{\text{in}} - \mathbf{R}_{n-1}^{\text{out}}. \qquad [7]$$

The new input vector, $\mathbf{R}_n^{\text{in}}$, is then used to generate a new set of noiseless spectra. The repeated processing of synthetic spectra yields a relaxation-rate vector $\mathbf{R}_n^{\text{out}}$. If the output obtained $\mathbf{R}_n^{\text{out}}$ differs appreciably from the target $\mathbf{R}^{\text{fit}}$, the iteration continues. Since the output vector of relaxation rates converges toward the target vector $\mathbf{R}^{\text{fit}}$, Fig. 1b, the input vector converges toward an actual vector of relaxation rates. In practice, a satisfactory convergence can be obtained in 3–5 cycles.

To show to what extent the spectral overlap interferes with the relaxation-rate measurements and to demonstrate the method to eliminate systematic errors, we conducted a series of simulations on ^{15}N T_1 measurements in a small globular protein (human ubiquitin, $M_w = 8600$). We have used the relaxation-rate values from literature, obtained at $B_0 = 14.1$ T and reported with a precision of 0.2% (*10*), and the chemical shifts (peak positions) from our own spectra recorded at 11 T, 297 K, and pH 4. In case of overlap, the peak positions are obtained, like in spectral assignment, from the position of isolated cross peaks from HSQC/NOESY or HSQC/TOCSY experiments.

To demonstrate how the procedure works, we have simulated the influence of spectral overlap with two sets of filtering functions. The first set is chosen to deliberately broaden spectral lines, to simulate lower fields or larger (or partially denatured) protein [Lorentz-to-Gauss transformation: $\exp(\text{LB} \times t)\exp(-\text{GB} \times t^2)$ with LB = −11, GB = 0.06 in proton domain and LB = −11, GB = 0.05 in nitrogen domain]. The second set consists of cosine squared in both domains and represents the best functions for our own relaxation data. We will refer to those two sets as a ''low field'' and ''high field,'' respectively. For the sake of simplicity, the homonuclear proton couplings were neglected. Due to many differences (digital resolution, filtering functions, chemical shift, number, and spacing of relaxation delays) between simulated and original experimental data, the analysis that follows does not reflect the quality of the actual experiment from Ref. (*10*). However, it may well represent the influence of overlap at higher fields in (partially) denatured or larger proteins.

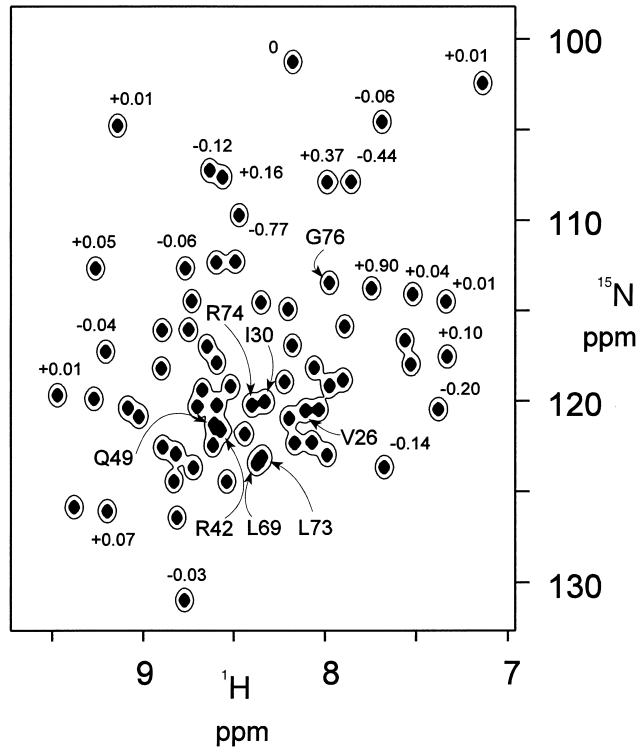

FIG. 2. [^{1}H–^{15}N] HSQC noiseless spectrum of human ubiquitin obtained by Fourier transformation of the time-domain data synthesized with 1024 × 1024 data points using known chemical shifts and linewidths at 14.1 T. Both contours are drawn at 1/5th of the maximal peak intensity. The open contour (''low field'') represents the spectrum with Lorentz-to-Gauss transformation, $\exp(\text{LB}\times t)\exp(-\text{GB}\times t^2)$, with LB = −11 Hz, GB = 0.06 in D1 and LB = −11 Hz, GB = 0.05 in D2, and the filled-in (''high-field'') contour has been obtained using a squared-cosine window over 512 data points in both dimensions. The numerical values indicate the relative errors due to spectral overlap in low field. The relevant numerical values for labeled amino acids are given in Tables 1 and 2.

Figure 2 shows a contour plot of the synthesized [^{1}H–^{15}N] HSQC spectrum of human ubiquitin with the above introduced two sets of filtering functions used. The spectrum has been obtained by 2D Fourier transformation of the time-domain data synthesized from experimental values of chemical shifts, linewidths, and relaxation rates $\mathbf{R}^{\text{fit}}$. It was processed in the same manner as the experimental spectra (see figure legend for details). Although the majority of lines are well resolved, there are several instances of moderate and strong spectral overlap. Numbers in Fig. 2 indicate the percentile deviations between the ^{15}N literature $\mathbf{R}^{\text{fit}}$ values and the values obtained by fitting the series of low-field synthetic spectra. Peaks with deviation larger than 1% are identified by their amino acid residues. Their values are presented in Table 1. As expected, the majority of well-separated peaks have systematic errors less than 0.1%. However, peaks from a more crowded region have errors as high as 5%; see Table 1.

TABLE 1
Systematic Errors of ^{15}N Longitudinal Relaxation Rates in Human Ubiquitin at the "Low Field"

Res. no.	$(\mathbf{R}^{fit} - \mathbf{R}_n^{out})/R^{fit\,a}$				$(\mathbf{R}^{fit} - \mathbf{R}_n^{in})/R^{fit\,b}$			$\mathbf{R}_1$ $(s^{-1})^c$	T_1 $(s)^c$
	$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n = 1$	$n = 2$	$n = 3$		
I30	4.37%	1.42%	0.40%	0.11%	−4.37%	−5.80%	−6.20%	2.232	0.448
L69	3.41%	1.74%	0.82%	0.37%	−3.41%	−5.15%	−5.97%	2.165	0.462
R42	1.79%	0.97%	0.49%	0.24%	−1.79%	−2.76%	−3.25%	2.137	0.468
V26	1.24%	0.32%	0.08%	0.02%	−1.24%	−1.56%	−1.64%	2.222	0.450
Q49	−1.77%	−0.73%	−0.33%	−0.16%	1.77%	2.50%	2.83%	1.984	0.504
G76	−2.07%	−0.03%	0.00%	0.00%	2.07%	2.10%	2.10%	0.791	1.265
R74	−4.92%	−0.89%	−0.21%	−0.05%	4.92%	5.81%	6.01%	1.555	0.643
L73	−5.10%	−2.05%	−0.87%	−0.38%	5.10%	7.15%	8.02%	1.802	0.555
Mean[d]	3.10%	1.00%	0.40%	0.17%	3.10%	4.10%	4.50%	—	—

Note. By strong filtering, lines are deliberately broadened to 44.1 Hz in the proton domain and 29.7 Hz in the nitrogen domain. Shown are only residues with relative errors greater than 1%.

[a] $\mathbf{R}_0^{out}$ is obtained from the spectra synthesized with $\mathbf{R}_0^{in} = \mathbf{R}^{fit}$; thus, the column with $n = 0$ represents a systematic error due to the spectral overlap.

[b] The column with $n = 0$ (not shown) is always zero because $\mathbf{R}_0^{in} = \mathbf{R}^{fit}$.

[c] Average values of the data from supporting information of Ref. (*10*); $B_0 = 14.1$ T at 27°C.

[d] Mean values for reported eight amino acid residues.

Table 1 also shows the progress of iterative error corrections, columns with $n = 0$ to $n = 3$. The column for $n = 0$ shows the percentile deviation between the reported relaxation-rate values $\mathbf{R}^{fit}$ and the values $\mathbf{R}_0^{out}$ obtained from the series of synthesized relaxation spectra using $\mathbf{R}^{fit}$ as an input. The average error for the selected set of residues is 3.1%. After the first correction, the average error falls from 3.1 to 1%, and in subsequent iterations, it falls to 0.4% and finally to 0.17%. The errors of the individual residues have their own paths that do not depend directly on the initial error value. For example, after the second iteration, the error for G76 falls from 2% almost to zero, whereas the error for R42 after the third iteration is still 0.24% even if the initial value is only 1.8%. This nonlinearity demonstrates the complex nature of error propagation caused by spectral overlap. It is even more pronounced in the relative differences between the original vector $\mathbf{R}^{fit}$ and the input vector $\mathbf{R}_n^{in}$. In the course of error correction, the input vector approaches the sought original vector of relaxation rates, and vector $\mathbf{R}_n^{in}$ represents the best (corrected) relaxation-rate values. As is evident from Table 1, the output values that need more than one iteration (except for G76) deviate from the reported values more than the input values. The average deviation of the output values rises from initial 3.1 to 4.5% after the third iteration. Again, for individual residues, these differences vary in a broad range from 1.6% for V26 to 8% for L73. Therefore, for L73, the true relaxation-rate value is 8% higher than the uncorrected one.

Figure 3 provides a better insight into the relationship between the input and the output relaxation-rate values. It shows the correlation between the relative difference of input and output relaxation-rate vectors for low-field data. The horizontal axis represents the relative differences between the reported relaxation-rate vector $\mathbf{R}^{fit}$ and the output vector $\mathbf{R}_0^{out}$ that is obtained from a series of synthetic spectra. The vertical axis represents the relative difference between the vector of reported relaxation rates $\mathbf{R}^{fit}$ and the corrected relaxation-input-rates vector *after n* iterations $\mathbf{R}_n^{in}$. For $n = 0$, the difference is zero because the original data set was used initially, $\mathbf{R}_0^{in} = \mathbf{R}^{fit}$. In the first iteration ($n = 1$), all values were corrected linearly, and according to Eq. [6], all points lie on a straight line with unit slope. For residues not influenced by overlap, the subsequent iterations do not change the differences and every point remains on a straight line. This happens to all residues with initial differences smaller than 1% and exceptionally to G76 for which the initial difference is somewhat larger (2%). The residues with initial differences larger than 1% are driven into a nonlinear regime by subsequent iterations. Because the nonlinearity cannot be identified in the first iteration, more than one iteration is always needed.

The correction of relaxation rates for distortions due to spectral overlap is not unique. Even for a given series of relaxation spectra, it depends on the filtering functions used. Also, slightly different input sets in repeated iterations may converge toward the same target set but yielding somewhat different output values. It is important to note, however, that in most cases the linear correction of the relaxation-rate data set may be sufficient.

For assessment of real high-field experimental data, more relevant is the analysis with optimal filtering functions. With optimal filtering, the overlap is less pronounced but still

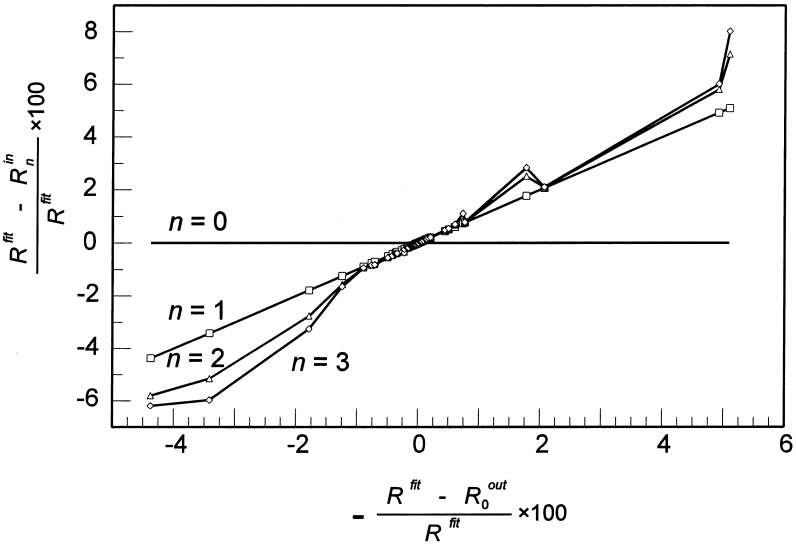

FIG. 3. Correlation among the relative differences between the input and the output relaxation-rate sets in low field. The horizontal axis shows the difference between the fitted experimental data $\mathbf{R}^{fit}$ and the data obtained from synthetic noiseless spectra. The vertical axis shows the difference between the original and corrected data sets: for $n = 0$, there is no difference since $\mathbf{R}^{fit} = \mathbf{R}_0^{in}$; for $n = 1$, all points are on a straight line with slope -1 since a linear correction was applied. The subsequent linear corrections ($n = 2, 3$) do not affect points with small deviations, but for the larger deviations, introduce a nonlinearity in the correction.

present. Table 2 shows the analysis of the same eight residues from human ubiquitin with overlap simulated at high field. Due to better peak separation, the maximal systematic error caused by overlap is only 0.9%. However, this small systematic error must be taken into account if an accuracy on the order of 1% is required. Since the actual overlap error critically depends on the choice of filtering function and the number and spread of relaxation delay increments, results from Table 2 do not apply to the experimental relaxation rates in (*10*) as the processing parameters of the experimental data are not reported. Table 2, however, provides good insight into the magnitude of overlap errors. Even at high fields and optimal filtering, systematic errors for some lines may be on the order of 1% and can easily be corrected in a single iteration.

TABLE 2
Systematic Errors of ^{15}N Longitudinal Relaxation Rates in Human Ubiquitin at the "High Field"

Res. no.	$(\mathbf{R}^{fit} - \mathbf{R}_n^{out})/\mathbf{R}^{fit}$ [a]			$(\mathbf{R}^{fit} - \mathbf{R}_n^{in})/\mathbf{R}^{fit}$ [b]	
	$n = 0$	$n = 1$	$n = 2$	$n = 1$	$n = 2$
I30	0.86%	0.05%	0	−0.86%	−0.91%
L69	0.44%	0.03%	0	−0.44%	−0.47%
R42	0.37%	0.04%	0	−0.37%	−0.41%
V26	0.24%	0.02%	0	−0.24%	−0.25%
Q49	−0.40%	−0.03%	0	0.40%	0.43%
G76	−0.35%	−0.01%	0	0.35%	0.36%
R74	−0.88%	−0.03%	0	0.88%	0.91%
L73	−0.88%	−0.05%	0	0.88%	0.93%
Mean[c]	0.55%	0.03%	0	0.55%	0.58%

Note. Linewidths are 21.8 and 15.7 Hz in the proton and nitrogen domain, respectively. Shown are the same residues as in Table 1.

[a] $\mathbf{R}_0^{out}$ is obtained from the spectra synthesized with $\mathbf{R}_0^{in} = \mathbf{R}^{fit}$; thus, the column with $n = 0$ represents a systematic error due to the spectral overlap.

[b] The column with $n = 0$ (not shown) is always zero because $\mathbf{R}_0^{in} = \mathbf{R}^{fit}$.

[c] Mean values for reported eight amino acid residues.

We believe that the proposed procedure is important for assessing the quality of relaxation-rate data because it can identify and correct systematic errors due to weak spectral overlap. As demonstrated here, when using low fields or in large macromolecules, these errors can far exceed random errors. Sometimes they may be easy to correct by careful optimization of filtering functions. In that case, the proposed procedure (with no iterations) may be used to assess the selected filters. A good filter minimizes the difference $\mathbf{R}^{in} - \mathbf{R}^{out}$.

The proposed method to check and correct the systematic errors is not limited to relaxation-rate measurements but can be applied to any measurement where the information is extracted from peak intensities or peak volumes. For example, it can be equally well applied when measuring coupling constants (*13*) or chemical-shift anisotropy (*14*) as both are based on the analysis of peak intensities from corresponding two-dimensional spectra. The method might be especially important in the study of large or partially denatured proteins where the spectral overlap is particularly strong.

REFERENCES

1. G. M. Clore, A. Szabo, A. Bax, L. E. Kay, P. C. Driscoll, and A. M. Gronenborn, *J. Am. Chem. Soc.* **112,** 4989 (1990).
2. W. E. Hull and B. D. Sykes, *J. Chem. Phys.* **63,** 867 (1975).
3. J. Peng and G. Wagner, *Biochemistry* **31,** 8571 (1992).
4. P. Borer, S. LaPlante, A. Kumar, N. Zanatta, A. Martin, A. Hakkinen, and G. Levy, *Biochemistry* **33,** 2441 (1994).
5. M. Akke and I. Palmer, *J. Am. Chem. Soc.* **118,** 911 (1996).
6. G. Barbato, M. Ikura, L. E. Kay, R. W. Pastor, and A. Bax, *Biochemistry* **31,** 5269 (1992).
7. V. Sklenář, D. Torchia, and A. Bax, *J. Magn. Reson.* **73,** 375 (1987).
8. N. R. Nirmala and G. Wagner, *J. Am. Chem. Soc.* **110,** 7557 (1988).
9. N. R. Nirmala and G. Wagner, *J. Magn. Reson.* **82,** 659 (1989).
10. N. Tjandra, S. Feller, R. W. Pastor, and A. Bax, *J. Am. Chem. Soc.* **117,** 12,562 (1995).
11. L. E. Kay, L. K. Nicholson, F. Delaglio, A. Bax, and D. A. Torchia, *J. Magn. Reson.* **97,** 359 (1992).
12. J. W. Peng, V. Thanabal, and G. Wagner, *J. Magn. Reson.* **95,** 421 (1991).
13. P. R. Blake, B. Lee, M. F. Summers, M. W. W. Adams, J.-B. Park, Z. H. Zhou, and A. Bax, *J. Biomol. NMR* **2,** 527 (1992).
14. N. Tjandra, A. Szabo, and A. Bax, *J. Am. Chem. Soc.* **118,** 6986 (1996).