JOURNAL OF MAGNETIC RESONANCE **84**, 398–409 (1989)

COMMUNICATIONS

How Accurately Can Interproton Distances in Macromolecules Really Be Determined by Full Relaxation Matrix Analysis of Nuclear Overhauser Enhancement Data?

G. MARIUS CLORE AND ANGELA M. GRONENBORN

Laboratory of Chemical Physics, Building 2, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892

Received May 8, 1989

In recent years considerable success has been achieved in determining three-dimensional structures of macromolecules in solution on the basis of approximate interproton distance restraints derived from nuclear Overhauser enhancement measurements (*1–4*). Improvements in the precision of such structure determinations can potentially be obtained by increasing either the number or the accuracy of the experimental restraints. To this end, several groups have been seeking to obtain more accurate interproton distances using full relaxation matrix analysis of the NOE data in order to account fully for multispin effects and overcome the errors arising from the application of a simple two-spin approximation (*5–14*). The strategies generally proposed involve an iterative approach. A trial structure is first computed on the basis of a set of initial approximate interproton distance restraints using one of the available structure determination methods such as metric matrix distance geometry (*15–17*), minimization in torsion angle space (*18, 19*), restrained molecular dynamics (*20–22*), or dynamical simulated annealing (*23–25*). Based on this trial structure, a theoretical two-dimensional NOE (NOESY) spectrum is computed and compared to the experimental one. Adjustment of the interproton distance restraints and further structure refinement yield a new trial structure. The entire process is repeated until the experimental and theoretical NOESY spectra match up (*26*). Alternatively the matrix of theoretical NOESY cross-peak and diagonal-peak intensities derived from the trial structure can be merged with the experimental one to calculate a new set of interproton distances with which to refine the structure and produce a new trial structure; the process is again repeated in an iterative manner until no change in the interproton distances is observed from one cycle to the next (*9–11*). Because structure refinement is computationally intensive, it has generally been considered that one only needs to carry this process out for a single starting structure. The interproton distances derived in this manner are then assumed to accurately represent the true interproton distances and to provide a reliable data set with which to compute an ensemble of structures in order to obtain a measure of the conformational space consistent with the experimental NOE data (*8*). The underlying assumption in such an approach is that the interproton distances, or more specifically the cross-relaxation

0022-2364/89 $3.00

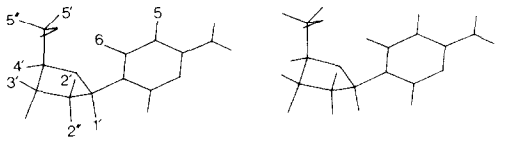

FIG. 1. Stereoview of deoxycytidine in a B-DNA conformation taken from the coordinates of classical B-DNA (*35*). This structure is used to derive the interproton distances used in the present calculations.

rates, are well determined by the NOESY cross-peak build-up curves. To our knowledge, however, this assumption has not been tested.

In this paper we analyze how accurately interproton distances in macromolecules can really be determined from NOE build-up curves. In order to treat multispin effects appropriately and at the same time preserve sufficient simplicity so that the geometric relationship between the interproton vectors can still be readily appreciated and visualized, we have chosen to use deoxycytidine as our model structure. The distances derived from this structure are then used to simulate the time dependence of NOESY cross-peak intensities in the $\omega\tau_c \gg 1$ spin diffusion limit. A stereoview of deoxycytidine is shown in Fig. 1. It has 9 protons and 35 interproton vectors, all of which are less than 6 Å, and only 7 have values greater than 5 Å. The distances for these vectors are given in the first column of Table 1. Although a nucleoside was chosen to extract a set of interproton distances, the general conclusions from the calculations apply equally well to any other multispin system in the $\omega\tau_c \gg 1$ regime.

The behavior of the system in a NOE experiment can be described by a series of coupled linear first-order differential equations (*27, 28*)

$$\frac{da_j}{d\tau_m} = -(R_{1j} + \sum_{i \neq j} \sigma_{ij})a_j + \sum_{i \neq j} a_i \sigma_{ij}, \qquad [1]$$

where a_j is the magnetization of proton j, σ_{ij} the cross-relaxation rate between protons i and j, and R_{1j} the leakage rate to the lattice for proton j. For a 2D NOESY experiment, the intensity of a given diagonal peak and its associated cross peaks, as a function of the mixing time τ_m, is obtained by solving Eq. [1] with the element of the magnetization vector **a** corresponding to the diagonal peak set to 1 and all other elements of **a** set to 0 at $\tau_m = 0$. The interproton distances r_{ij} (in units of Å) are related to the cross-relaxation rates σ_{ij} by the well-known equation (*27*)

$$\sigma_{ij} = \frac{5.7 \times 10^7}{r_{ij}^6}\left(\tau_c - \frac{6\tau_c}{1 + 4\omega^2\tau_c^2}\right), \qquad [2]$$

where τ_c is the effective correlation time of the i–j vector.

Our approach to the problem involves first calculating, under conditions where $\omega\tau_c \gg 1$, the exact NOE build-up curves for the cross-relaxation network of deoxycytidine by numerical integration and adding an appropriate amount of random noise to simulate real experimental data. (For simplicity all leakage rates to the lattice are set to 0.5 s^{-1}.) Only the time dependences of the cross-peak intensities are included in the model data as, in general, the intensities of diagonal peaks cannot be measured accurately for either protein or oligonucleotide spectra owing to extensive spectral overlap.

TABLE 1

Comparison of Target and Optimized Values of the Interproton Distances Obtained from Various Calculations by Best-Fitting Model Data Derived for a Single 5 ns Overall Correlation Time[a]

Distance (Å)	Target value	Full 17 mixing NOE data set calculation			50, 100, 200 ms NOE data	100 ms NOE data	From initial rate
		1	2	3			
H5–H6	2.47	2.47*f*	2.47*f*	2.47*f*	2.47*f*	2.47*f*	2.47*f*
H5–H1′	5.44	4.86(+0.98/−0.82)	10*f*	10*f*	10*f*	10*f*	
H5–H2′	4.17	3.64(+0.35/−0.31)	4.92(+2.82/−1.79)	10*f*	7.22*nd*	12.5*nd*	
H5–H3′	6.29	5.16(+1.94/−1.40)	10*f*	10*f*	10*f*	10*f*	
H5–H4′	6.76	4.65(+0.84/−0.72)	10*f*	10*f*	10*f*	10*f*	
H5–H5′	5.06	5.15(+2.68/−1.76)	10*f*	10*f*	10*f*	10*f*	
H5–H5″	6.17	6.17(+18.2/4.61)	10*f*	10*f*	10*f*	10*f*	
H6–H1′	3.69	3.79(+0.44/−0.40)	3.67(+0.31/−0.30)	3.49(+0.22/−0.21)	4.81(+6.14/−2.70)	12.7*nd*	
H6–H2′	1.87	1.87(+0.02/−0.02)	1.92(+0.02/−0.02)	1.91(+0.02/−0.02)	1.88(+0.03/−0.03)	1.93(+0.06/−0.06)	1.9
H6–H2″	3.43	2.98(+0.21/−0.20)	2.92(+0.17/−0.16)	10*f*	3.88(+3.47/−1.84)	2.68(+2.36/−0.32)	
H6–H3′	3.84	3.68(+0.41/−0.37)	3.44(+0.22/−0.21)	3.46(+0.23/−0.21)	4.22(+2.32/−1.49)	5.45(+0.86/−0.73)	3.0
H6–H4′	4.52	4.32(+1.29/−1.00)	4.03(+0.68/−0.58)	10*f*	4.08(+1.55/−1.12)	7.89*nd*	
H6–H5′	3.31	3.24(+0.21/−0.19)	3.57(+0.35/−0.32)	3.46(+0.29/−0.27)	3.73(+1.02/−0.80)	3.41(+0.61/−0.52)	3.5
H6–H5″	4.07	4.16(+1.43/−1.06)	3.91(+0.65/−0.55)	3.90(+0.72/−0.61)	3.89(+1.30/−0.97)	5.55(+5.25/−3.71)	3.8
H1′–H2′	2.99	3.15(+0.22/−0.21)	3.24(+0.24/−0.22)	3.08(+0.17/0.16)	2.94(+0.28/−0.25)	2.85(+0.24/−0.23)	2.7
H1′–H2″	2.29	2.26(+0.03/−0.03)	2.34(+0.03/−0.03)	2.36(+0.03/−0.03)	2.33(+0.04/−0.05)	2.33(+0.07/−0.06)	2.3
H1′–H3′	3.89	3.83(+0.28/−0.26)	3.89(+0.27/−0.25)	10*f*	4.28(+1.09/−0.87)	3.79(+0.50/−0.46)	
H1′–H4′	3.61	3.52(+0.15/−0.14)	3.79(+0.23/−0.20)	3.61(+0.19/−0.19)	3.59(+0.29/−0.26)	3.71(+0.43/−0.38)	3.7
H1′–H5′	4.48	5.42(+5.38/−2.71)	5.16(+2.16/−1.53)	10*f*	5.08(+4.05/−2.26)	3.93(+0.66/−0.56)	
H1′–H5″	5.10	4.10(+0.62/−0.54)	10*f*	10*f*	10*f*	10*f*	

H2′–H2″	1.78	1.76(+0.02/−0.02)	1.82(+0.02/−0.02)	1.83(+0.02/−0.02)	1.78*f*	1.78*f*	1.8
H2′–H3′	2.39	2.39(+0.04/−0.04)	2.50(+0.05/−0.05)	2.48(+0.04/−0.04)	2.42(+0.08/−0.07)	2.33(+0.09/−0.08)	2.5
H2′–H4′	3.86	4.20(+1.91/−1.39)	3.95(+1.01/−0.81)	10*f*	3.95(+2.48/−1.52)	3.49(+0.92/−0.72)	
H2′–H5′	3.80	5.19(+17.0/−5.71)	3.72(+0.67/−0.57)	3.97(+0.91/−0.75)	3.24(+0.41/−0.36)	4.49(+7.11/−2.75)	
H2′–H5″	3.89	3.61(+0.58/−0.50)	6.44*nd*	3.73(+0.63/−0.54)	6.36*nd*	7.44*nd*	
H2″–H3′	2.70	2.72(+0.07/−0.07)	2.81(+0.08/−0.07)	2.88(+0.08/−0.08)	2.73(+0.13/−0.13)	2.99(+0.34/−0.31)	2.6
H2″–H4′	4.08	3.75(+0.58/−0.51)	4.15(+1.17/−0.91)	3.55(+0.19/−0.17)	4.10(+2.02/−1.34)	3.72(+1.42/−1.03)	
H2″–H5′	4.96	3.91(+0.82/−0.69)	5.28(12.32/−3.70)	10*f*	6.03*nd*	13.2*nd*	
H2″–H5″	4.94	6.48*nd*	3.70(+0.46/−0.41)	10*f*	3.41(+0.48/−0.41)	3.28(+0.47/−0.41)	
H3′–H4′	2.69	2.68(+0.04/−0.04)	2.81(+0.04/−0.04)	2.82(+0.03/−0.03)	2.82(+0.08/−0.08)	2.85(+0.09/−0.08)	2.7
H3′–H5′	3.74	3.57(+0.24/−0.23)	3.42(+0.17/−0.16)	3.32(+0.14/−0.12)	3.75(+0.67/−0.57)	3.76(+0.76/−0.63)	
H3′–H5″	2.83	2.82(+0.06/−0.06)	3.08(+0.09/−0.09)	3.21(+0.11/−0.11)	2.86(+0.11/−0.11)	2.80(+0.10/−0.10)	2.9
H4′–H5′	2.61	2.64(+0.04/−0.04)	2.72(+0.04/−0.04)	2.78(+0.05/−0.05)	2.81(+0.10/−0.10)	2.78(+0.10/−0.10)	2.6
H4′–H5″	2.34	2.33(+0.03/−0.03)	2.38(+0.03/−0.03)	2.41(+0.02/−0.02)	2.34(+0.04/−0.04)	2.29(+0.04/−0.04)	2.3
H5′–H5″	1.78	1.78(+0.02/−0.02)	1.82(+0.02/−0.02)	1.84(+0.02/−0.02)	1.78*f*	1.78*f*	2.0
τ_c (ns)	5.0	4.92(+0.25/−0.24)	5.86(+0.30/−0.28)	6.13(+0.21/−0.21)	5.56(+0.26/−0.23)	5.37(+0.34/−0.32)	
Overall SD of fit		±0.014	±0.0078	±0.0076	±0.006	±0.0047	

[a] The model data have a random noise level of ±0.009 and include the time courses of all cross-peak intensities. The time dependence of the diagonal-peak intensities are not included in the model data. All leakage rates to the lattice are kept fixed at 0.5 s^{-1} both in the generation of the model data and in the optimization calculations, and the spectrometer frequency is set to 500 MHz. The 17 mixing times used are 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 200, 250, 300, and 400 ms. The numbers in parentheses are the 5 and 95% confidence limits in the estimation of the values of the optimized parameters calculated from the variance–covariance marix obtained at the end of the optimization run. *f* indicates that a distance is held fixed during optimization, and *nd* that the value of the distance is not determined by the data to within a factor of 100 (i.e., the corresponding cross-relaxation rate is near zero). See text for further details.

The *complete set* of model NOESY cross-peak build-up curves is then best fitted *simultaneously* by carrying out successive numerical integration runs under control of a nonlinear least-squares optimization routine, varying the values of the interproton distances and overall correlation time. All calculations were carried out using the program FACSIMILE (*29, 30*) which employs Curtis' modified version (*31*) of Gear's backward difference method (*32*) for numerical integration, and Powell's method of optimization (*33*) which does not require the computation of partial derivatives. The level of noise chosen to mimick actual experimental data was ±0.009 (i.e., $\pm \sim 1\%$ relative to the intensity of a diagonal peak at $\tau_m = 0$), and in all cases the values of the distances and overall correlation time to be optimized were set to initial values of 4 Å and 10 ns, respectively. Note that these calculations do not involve any reference to Cartesian coordinates.

To obtain a quantitative measure of how well each varied parameter P_l is determined by the data, we proceeded as follows. When the minimum residual sum of squares is reached, a sensitivity matrix, expressing the dependence of each residual x_{ij} on the natural logarithm $\ln(P_l)$ of each parameter at its optimum value, is calculated by adding 0.2 to each $\ln(P_l)$ in turn and examining the effects on each x_{ij}. (Note the reason that this is carried out on $\ln(P_l)$ rather than on P_l is that it is computationally both more efficient and more reliable to vary $\ln(P_l)$ rather than P_l.) The normal matrix is then calculated from the sensitivity matrix and inverted to obtain the variance–covariance matrix which refers to $\ln(P_l)$. The variance of each $\ln(P_l)$ is given by the diagonal elements of this matrix, and from this the 5 and 95% confidence limits of P_l are obtained (*29, 30*).

The model data at 17 mixing times ranging from 10 to 400 ms calculated for an overall correlation time of 5 ns at a spectrometer frequency of 500 MHz are shown in Fig. 2. This correlation time is not meant in any way to represent the actual correlation time of deoxycytidine but rather to reflect the correlation time of a macromolecule with molecular weight around 10,000 for which $\omega\tau_c \gg 1$. As the cross-relaxation rates are proportional to both the correlation time τ_c and r_{ij}^{-6}, either one distance or the correlation time must be fixed to fit the data. In real systems there are always vectors whose distances are fixed by covalent geometry so that in practice this does not pose a problem. Thus, in the case of deoxycytidine there are three vectors whose distances are conformation independent: $r_{H5-H6} = 2.47$ Å and $r_{H2'-H2''} = r_{H5'-H5''} = 1.79$ Å. Three optimization runs were carried out, fitting all the data simultaneously. In all three cases the overall correlation time was varied and the H5–H6 distance held fixed at 2.47 Å, leaving a total of 34 unknown distances. In the first case, all 34 distances were varied. In the second, the 28 distances less than 5 Å in deoxycytidine were varied and the remaining 6 were set to a value of 10 Å (which is equivalent to giving them near-zero cross-relaxation rates). Finally, in the third run, 14 distances corresponding to those vectors which clearly could not contribute significantly to the relaxation pathway between the corresponding two protons on the basis of geometrical considerations were set to 10 Å, and the remaining 20 were varied. Thus, for example, it is clear from inspection of the structure (Fig. 1) that the direct relaxation pathway between the H2′ proton and the H4′ proton will not contribute significantly to the time development of the H2′–H4′ NOESY cross peak as

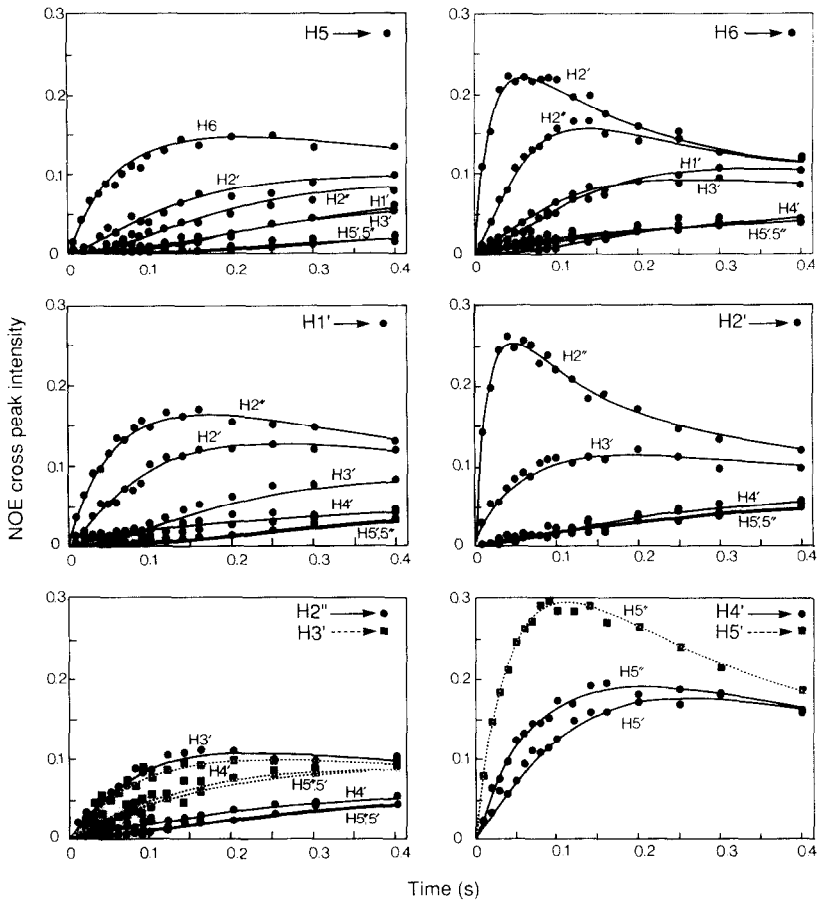

FIG. 2. Comparison of the 5 ns correlation time model data with the computed best-fit time courses of the cross-peak intensities for deoxycytidine. The best-fit parameters for this set of calculated curves are given in Table 1 under calculation number 3 (column 5). The calculated curves are shown as either continuous or interrupted lines, and the cross-peak intensities for the model data by solid circles or squares. The proton at the diagonal of the NOESY spectrum for each set of curves is indicated at the top right hand corner of each panel. Note that the time dependence of the diagonal peaks is not included in the model data as their intensities cannot in general be measured accurately for macromolecules owing to severe spectral overlap. (See text for further details.)

the predominant relaxation pathway will involve the almost linear H2′ → H3′ → H4′ indirect route.

A summary of the results is given in Table 1 and a comparison of the best fit between the data and the third simulation run is shown in Fig. 2. It is clear that agreement between the calculated curves and the model data is within the errors of the data for all three optimization runs. Note that in the third run there are some minor systematic errors in the fits of a few curves at longer mixing times (cf. the H1′–H3′ curve); these errors, however, lie well within the overall standard deviation of the data. The determination of the optimized parameters varies widely. In general, a distance r_{ij} will be well determined within an error of less than ±0.1 Å if the direct $i \rightarrow j$

pathway is the principal source of cross-relaxation (<95%) between protons i and j. This, for example, is true for $r_{H6-H2'}$, $r_{H1'-H2''}$, $r_{H2'-H2''/H3'}$, $r_{H2''-H3'}$, $r_{H3'-H4'/H5'/H5''}$, $r_{H4'-H5'/H5''}$, and $r_{H5'-H5''}$. As the contribution to cross-relaxation between protons i and j from indirect pathways increases, so the accuracy with which the distance r_{ij} can be determined decreases. Providing the contribution from direct cross-relaxation is still significant (>20–30%) the $i-j$ distance can be determined within an accuracy of better than ±0.5 Å, as is true for $r_{H6-H1'/H3'/H5'}$ and $r_{H1'-H2'/H4'}$. Finally, the $i-j$ distance cannot be determined at all by the data if the contribution from the direct $i \rightarrow j$ cross-relaxation pathway is negligible. This will be manifested by a significant lag phase in the time development of the a_{ij} cross peak, as is the case, for example, for all cross peaks arising from NOEs between the H5 base proton and all sugar protons (Fig. 2). Thus, for example, setting the values of these distances to an arbitrarily large value of say 10 Å has little or no effect either on the goodness of fit between the calculated and the model "experimental" data or on the optimized values of the other distances. This applies not only to distances larger than 5 Å (cf. the results of the second optimization run given in column 4 of Table 1) but also to shorter distances in cases where the direct contribution to the corresponding NOESY cross peak is small (cf. the results of the third optimization run given in column 5 of Table 1). It should also be noted that the 5–95% confidence limits in the values of the optimized parameters may underestimate the true errors as there are cases where the target values lie outside the limits (e.g., see the optimized values for $r_{H3'-H5''}$).

At this stage it is also interesting to compare these results with those obtained using a simple two-spin approximation in which the initial slopes of the curves are simply measured manually and the distances calculated using the initial slope of the H5–H6 vector as an internal reference. With this approach, the cross-relaxation rate for a vector $i-j$ can clearly never be determined if the time dependence of the intensity of the a_{ij} cross peak exhibits a lag phase. Where no lag phase is apparent, however, it can be seen that the results of the two-spin approximation compare reasonably well to those of the complete analysis as can be seen from a comparison of the first and last columns of Table 1. This applies not only to very short distances such as $r_{H6-H2'}$ but also to longer ones such as $r_{H1'-H4'}$ for which hardly any indirect contributions occur. When indirect cross-relaxation is sizable, the distance is clearly underestimated as in the case of the distance $r_{H6-H3'}$ where the H6 → H2′ → H3′ pathway contributes significantly to the intensity of the H6–H3′ cross peak. There are also instances where the lag phase is so small that it could easily be overlooked (Fig. 2), in which case a serious underestimation of the actual distance would be obtained. An example of this type of behavior is provided by the H6–H2″ NOESY cross-peak build-up curve where magnetization transfer via the indirect H6 → H2′ → H2″ route is not only the predominant relaxation pathway but is also exceedingly fast owing to the very short H6–H2′ and H2′–H2″ distances. Even the complete analysis, however, does not yield reliable values in such cases, as evidenced by the results of the third optimization run where the value of $r_{H6-H2''}$ is fixed at 10 Å with no untoward effect on the values of the other optimized distances. (Note, of course, that in molecular terms the maximum value of $r_{H6-H2''}$ is given by the sum of $r_{H6-H2'}$ and $r_{H2'-H2''}$.)

In a real experimental case involving a protein or nucleic acid fragment, the accurate measurement of NOESY cross-peak intensities at 17 mixing times between 10

and 400 ms would be nothing less than a tour de force, and in most cases quite impractical. It is therefore useful to investigate how well the distances can be determined from only a few mixing times. To test this we have carried out two further calculations: one with 50, 100, and 200 ms mixing times, the other with only a single 100 ms mixing time. The known values of r_{H5-H6}, $r_{H2'-H2''}$, and $r_{H5'-H5''}$ were held constant, those distances greater than 5 Å were fixed at 10 Å, and the remaining 26 distances were varied together with the overall correlation time. The results of these calculations (columns 6 and 7, Table 1) show that the accuracy with which the distances are determined is considerably decreased for both reduced data sets compared to the full 17-time-point data set. This is most significant for cross peaks where the contributions from both direct and indirect cross-relaxation during the initial growth phase are significant. For example, the distance between the H6 and the H1′ protons which could be determined from the full data to an accuracy of better than ±0.5 Å is ill-determined by both reduced data sets.

The second issue concerning the application of full relaxation matrix analysis to real systems is the effect of potential variations in effective correlation times arising from internal motion. To test what effect this would have on the estimation of interproton distances, we calculated a set of model data using a correlation time of 5 ns for the base–base, base–H1′, and H1′–H4′ vectors and a correlation time of 2.5 ns for the other vectors. Again we varied all distances less than 5 Å, excluding those fixed by covalent geometry, together with an overall correlation time. Three optimization calculations were performed using the full 17-time-point data. In the first only r_{H5-H6} was fixed, in the second $r_{H2'-H2''}$ was fixed, and in the third r_{H5-H6}, $r_{H2'-H2''}$, and $r_{H5'-H5''}$ were fixed. The results are summarized in columns 3, 4, and 5, respectively, of Table 2. When only one known reference distance is fixed, the model data set can be fitted within its errors. The extent to which the distances are determined by the data is the same as that in the model calculations with the single 5 ns correlation time. However, this time, there are obvious systematic errors present. Thus, when r_{H5-H6} is fixed, the distances for those vectors with the smaller correlation time are systematically overestimated and the value for the optimized correlation time is close to 5 ns. When $r_{H2'-H2''}$ is fixed, those distances with the longer correlation time are systematically underestimated and the optimized value of the overall correlation time is 2.8 ns. When the reference distances for the different correlation times are held fixed, the agreement between calculated and model data is still approximately equal to the errors in the data, systematic errors in the distance estimations are present throughout, and the optimized value of the overall correlation time is intermediate between 2.5 and 5 ns. In all cases, however, the magnitude of these systematic errors is rather small and of the order of 0.2–0.3 Å. Thus, significant errors are only introduced into those distances that are determined to an accuracy of better than ±0.2 Å by the data. For example, with the H5–H6 distance fixed, the H2′–H2″ distance is overestimated by 0.21 Å, although its optimized value is determined by the data to a precision of ±0.02 Å. For a distance such as $r_{H6-H5'}$ on the other hand, the difference between the optimized values is within the confidence limits of their estimation. These conclusions also hold for both the 3-time-point and the 1-time-point reduced data sets (columns 6 and 7, respectively, of Table 2). However, as the distances are in any case less well determined by the reduced data sets, the errors arising from the variations

TABLE 2

Comparison of Target and Optimized Values of the Interproton Distances Obtained from Various Calculations by Best-Fitting Model Data Derived for a System with Two Sets of Effective Correlation Times, One of 5 ns for the Base–Base, Base–H1′, and H1′–H4′ Vectors and 2.5 ns for All the Other Vectors[a]

Distance (Å)	Target value	Full 17 mixing time NOE data set calculation			50, 100, 200 ms NOE data	100 ms NOE data
		1	2	3		
H5–H6	2.47	2.47*f*	2.26(+0.02/−0.02)	2.47*f*	2.47*f*	2.47*f*
H5–H1′	5.44	10*f*	10*f*	10*f*	10*f*	10*f*
H5–H2′	4.17	4.31(+0.67/−0.57)	3.51(+0.27/−0.26)	3.11(+0.12/−0.12)	3.13(+0.31/−0.27)	3.09(+0.46/−0.41)
H5–H3′	6.29	10*f*	10*f*	10*f*	10*f*	10*f*
H5–H4′	6.76	10*f*	10*f*	10*f*	10*f*	10*f*
H5–H5′	5.06	10*f*	10*f*	10*f*	10*f*	10*f*
H5–H5″	6.17	10*f*	10*f*	10*f*	10*f*	10*f*
H6–H1′	3.69	3.85(+0.25/−0.24)	3.44(+0.22/−0.20)	3.90(+0.64/−0.55)	4.03(+1.98/−1.32)	4.91(+33.5/−4.28)
H6–H2′	1.87	2.13(+0.02/−0.02)	1.89(+0.02/−0.02)	1.97(+0.02/−0.02)	1.97(+0.03/−0.03)	1.92(+0.05/−0.05)
H6–H2″	3.43	3.29(+0.16/−0.16)	3.61(+0.70/−0.58)	3.43(+0.56/−0.48)	3.03(+0.53/−0.45)	3.78(+5.79/−2.29)
H6–H3′	3.84	4.14(+0.46/−0.41)	3.81(+0.49/−0.43)	4.35(+1.55/−1.14)	4.01(+2.07/−1.36)	4.36(+8.34/−2.86)
H6–H4′	4.52	4.78(+1.22/−0.98)	4.33(+1.20/−0.94)	4.92(+4.36/−2.31)	4.78(+10.5/−3.28)	16.6*nd*
H6–H5′	3.31	3.96(+0.39/−0.35)	3.58(+0.35/−0.32)	3.81(+0.73/−0.62)	3.62(+1.19/−0.89)	3.58(+1.25/−0.93)
H6–H5″	4.07	4.84(+1.89/−1.36)	4.27(+1.43/−1.07)	4.32(+2.13/−1.43)	4.33(+6.47/−2.57)	5.55*nd*
H1′–H2′	2.99	3.37(+0.16/−0.16)	3.42(+0.35/−0.32)	3.15(+0.26/−0.24)	3.05(+0.46/−0.40)	2.81(+0.41/−0.35)
H1′–H2″	2.29	2.54(+0.03/−0.03)	2.30(+0.02/−0.02)	2.33(+0.03/−0.03)	2.34(+0.06/−0.06)	2.36(+0.10/−0.10)
H1′–H3′	3.89	4.32(+0.39/−0.35)	3.95(+0.36/−0.34)	4.01(+0.54/−0.46)	4.14(+1.21/−0.71)	3.64(+1.09/−0.84)
H1′–H4′	3.61	3.69(+0.12/−0.12)	3.36(+0.12/−0.11)	3.42(+0.17/−0.16)	3.25(+0.28/−0.26)	3.31(+0.53/−0.46)
H1′–H5′	4.48	6.14(+6.16/−3.08)	4.74(+1.38/−1.06)	4.50(+1.31/−1.02)	4.22(+2.11/−1.40)	3.57(+0.98/−0.78)
H1′–H5″	5.10	10*f*	10*f*	10*f*	10*f*	10*f*

H2′–H2″	1.78	1.99(+0.02/−0.02)	1.78*f*	1.78*f*	1.78*f*	1.78*f*
H2′–H3′	2.39	2.67(+0.04/−0.04)	2.39(+0.03/−0.03)	2.43(+0.04/−0.04)	2.40(+0.09/−0.09)	2.37(+0.13/−0.12)
H2′–H4′	3.86	4.91(+2.70/−1.74)	3.96(+1.02/−0.82)	3.57(+0.65/−0.55)	3.89(+3.29/−1.78)	3.95(+5.52/−2.30)
H2′–H5′	3.80	3.91(+0.54/−0.49)	3.71(+0.73/−0.62)	3.97(+1.93/−1.30)	3.69(+2.78/−1.58)	16.0*nd*
H2′–H5″	3.89	3.98(+0.69/−0.58)	4.11(+1.81/−1.26)	3.53(+0.83/−0.67)	3.86(+4.78/−2.13)	3.54(+2.58/−1.50)
H2″–H3′	2.70	3.12(+0.09/−0.08)	2.87(+0.09/−0.09)	2.96(+0.14/−0.14)	2.86(+0.25/−0.24)	2.99(+0.55/−0.44)
H2″–H4′	4.08	4.05(+0.48/−0.43)	3.82(+0.61/−0.52)	4.29(+2.23/−1.46)	3.73(+1.77/−1.21)	3.33(+1.09/−0.82)
H2″–H5′	4.96	5.63(+9.07/−3.47)	4.34(+1.91/−1.32)	3.63(+0.76/−0.63)	4.11(+5.75/−2.40)	4.30(+9.21/−2.94)
H2″–H5″	4.94	5.68(+10.62/−3.7)	3.99(+1.04/−0.83)	4.13(+2.23/−1.44)	3.72(+2.53/−1.51)	3.33(+1.09/−0.82)
H3′–H4′	2.69	3.05(+0.04/−0.04)	2.77(+0.03/−0.03)	2.85(+0.05/−0.05)	2.88(+0.14/−0.13)	2.92(+0.23/−0.22)
H3′–H5′	3.74	4.00(+0.29/−0.27)	3.69(+0.30/−0.27)	3.47(+0.27/−0.25)	3.51(+0.69/−0.57)	3.62(+1.33/−0.96)
H3′–H5″	2.83	3.26(+0.08/−0.08)	2.87(+0.05/−0.05)	2.99(+0.10/−0.10)	2.89(+0.19/−0.17)	2.80(+0.22/−0.21)
H4′–H5′	2.61	2.98(+0.04/−0.04)	2.77(+0.03/−0.03)	2.75(+0.05/−0.05)	2.78(+0.14/−0.13)	2.78(+0.21/−0.19)
H4′–H5″	2.34	2.67(+0.03/−0.03)	2.39(+0.02/−0.02)	2.39(+0.02/−0.02)	2.37(+0.06/−0.05)	2.33(+0.07/−0.06)
H5′–H5″	1.78	1.99(+0.02/−0.02)	1.82(+0.02/−0.02)	1.78*f*	1.78*f*	1.78*f*
τ_c (ns)	5.0/2.5	5.04(+0.18/−0.18)	2.87(+0.10/−0.10)	3.12(+0.08/−0.08)	3.04(+0.19/−0.18)	3.00(+0.31/−0.29)
Overall SD of fit		±0.0076	±0.0098	±0.010	±0.010	±0.009

[a] The model data have a random noise level of ±0.009 and include the time courses of all cross-peak intensities. The time dependence of the diagonal-peak intensities are not included in the model data. All leakage rates to the lattice are kept fixed at 0.5 s^{-1} both in the generation of the model data and in the optimization calculations, and the spectrometer frequency is set to 500 MHz. The 17 mixing times used are 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 200, 250, 300, and 400 ms. The numbers in parentheses are the 5 and 95% confidence limits in the estimation of the values of the optimized parameters calculated from the variance–covariance marix obtained at the end of the optimization run. *f* indicates that a distance is held fixed during optimization, and *nd* that the value of the distance is not determined by the data to within a factor of 100 (i.e., the corresponding cross-relaxation rate is near zero). See text for further details.

in effective correlation time are small in relation to the uncertainty in the values of the optimized parameters. Clearly, however, differences in effective correlation times much larger than a factor of 2 may have a sizable effect on the errors.

In conclusion, we have investigated the extent to which interproton distances can be determined from time-dependent NOE data in the presence of multispin effects for a system in the $\omega\tau_c \gg 1$ regime. It is clear that a full relaxation matrix analysis of the data can yield more accurate information than the two-spin approximation for interproton distances where both direct and indirect cross-relaxation pathways contribute significantly to cross-relaxation between the corresponding proton pairs. It does not, however, provide any useful information for interproton distances in cases where indirect effects dominate the cross-relaxation pathways. Moreover, it is important to bear in mind that, with the exception of situations where cross-relaxation between a given proton pair is dominated almost entirely by direct effects, the data can be fitted allowing for quite large variations in the values of the interproton distances. Consequently, any structure determination strategy that relies exclusively on an iterative refinement procedure to obtain a set of more accurate distances starting from a *single* trial structure is likely to introduce significant bias in the final outcome. In particular, the atomic rms distribution of the resulting ensemble of calculated structures will probably be lower than is justified by the data and the conformational space sampled will constitute only a subset of the conformational space consistent with the NOE data. This suggests a more fruitful approach in which a whole series of trial structures generated from the initial approximate interproton distance data are refined against a target function which seeks to minimize, among other terms relating to covalent and nonbonded interactions, the difference between calculated and observed NOESY cross-peak intensities directly, rather than the difference between calculated and estimated interproton distances. Such an approach has been suggested (*34*) but not yet subjected to test. The result would be the generation of an ensemble of refined structures which satisfied the NOE data but were not necessarily identical in terms of their interproton distances.

ACKNOWLEDGMENT

This work was supported by the Intramural Aids Targeted Anti-Viral Program of the Office of the Director of the National Institutes of Health.

REFERENCES

1. K. WÜTHRICH, "NMR of Proteins and Nucleic Acids," Wiley, New York, 1986.
2. G. M. CLORE AND A. M. GRONENBORN, *Protein Eng.* **1** (1987).
3. K. WÜTHRICH, *Acct. Chem. Res.* **22,** 36 (1989).
4. G. M. CLORE AND A. M. GRONENBORN, *CRC Crit. Rev. Biochem.*, in press (1989).
5. J. W. KEEPERS AND T. L. JAMES, *J. Magn. Reson.* **57,** 404 (1984).
6. E. T. OLEJNICZAK, R. T. GAMPE, and S. W. FESIK, *J. Magn. Reson.* **65,** 526 (1986).
7. P. A. MIRAU, "30th Experimental NMR Conference, Asilomar, California, W40, 1989."
8. B. A. BORGIAS AND T. L. JAMES, *J. Magn. Reson.* **79,** 493 (1988).
9. B. A. BORGIAS AND T. L. JAMES, "30th Experimental NMR Conference, Asilomar, California, W54, 1989."
10. R. BOELENS, T. M. G. KONING, AND R. KAPTEIN, *J. Mol. Struct.* **173,** 299 (1988).

11. R. BOELENS, T. M. G. KONING, G. A. VAN DER MAREL, J. H. VAN BOOM, AND R. KAPTEIN, *J. Magn. Reson.* **82,** 290 (1989).
12. D. MARION, M. GENEST, AND M. PTAK, *Biophys. Chem.* **28,** 535 (1987).
13. J. F. LEFEVRE, A. N. LANE, AND O. JARDETZKY, *Biochemistry* **26,** 5076 (1987).
14. C. B. POST, R. P. MEADOWS, J. T. METZ, AND D. G. GORENSTEIN, "30th Experimental NMR Conference, Asilomar, California, W41, 1989."
15. G. M. CRIPPEN AND T. F. HAVEL, *Acta Crystallogr. A* **34,** 282 (1978).
16. T. F. HAVEL, I. D. KUNTZ, AND G. M. CRIPPEN, *Bull. Math. Biol.* **45,** 665 (1983).
17. T. F. HAVEL AND K. WÜTHRICH, *J. Mol. Biol.* **182,** 281 (1985).
18. W. BRAUN AND N. GO, *J. Mol. Biol.* **186,** 611 (1985).
19. M. BILLETER, T. F. HAVEL, AND K. WÜTHRICH, *J. Comput. Chem.* **8,** 132 (1987).
20. G. M. CLORE, A. M. GRONENBORN, A. T. BRÜNGER, AND M. KARPLUS, *J. Mol. Biol.* **186,** 435 (1985).
21. G. M. CLORE, A. T. BRÜNGER, M. KARPLUS, AND A. M. GRONENBORN, *J. Mol. Biol.* **292,** 523 (1986).
22. R. KAPTEIN, E. R. P. ZUIDERWEG, R. M. SCHEEK, R. BOELENS, AND W. F. VAN GUNSTEREN, *J. Mol. Biol.* **182,** 179 (1985).
23. M. NILGES, A. M. GRONENBORN, A. T. BRÜNGER, AND G. M. CLORE, *Protein Eng.* **2,** 27 (1988).
24. M. NILGES, G. M. CLORE, AND A. M. GRONENBORN, *FEBS Lett.* **229,** 317 (1988).
25. M. NILGES, G. M. CLORE, AND A. M. GRONENBORN, *FEBS Lett.* **239,** 129 (1988).
26. D. R. HARE AND R. D. MORRISON, "XIII International Conference on Magnetic Resonance in Biological Systems, Wisconsin, Madison, S22-2, 1988."
27. I. SOLOMON, *Phys. Rev.* **99,** 559 (1955).
28. S. MACURA AND R. R. ERNST, *Mol. Phys.* **41,** 95 (1980).
29. E. M. CHANCE, A. R. CURTIS, I. P. JONES, AND C. R. KIRBY, "A.E.R.E. Report No. R-8775," Harwell, UK, 1977.
30. G. M. CLORE, *in* "Computing in Biological Science," (M. J. Geisow and A. N. Barrett, Eds.), pp. 313–348, Elsevier, Amsterdam, 1983.
31. A. R. CURTIS, "A.E.R.E. Report No. R9352," Harwell, UK, 1979.
32. C. W. GEAR, *Math. Comput.* **21,** 146 (1967).
33. M. J. D. POWELL, *Comput. J.* **7,** 303 (1965).
34. P. YIP AND D. A. CASE, *J. Magn. Reson.*, in press (1989).
35. S. ARNOTT AND D. W. L. HUKINS, *Biochem. Biophys. Res. Commun.* **47,** 1504 (1972).