JOURNAL OF MAGNETIC RESONANCE **126,** 283–286 (1997)
ARTICLE NO. MN971167

Optimal Sampling Strategies for the Measurement of Relaxation Times in Proteins

J. A. Jones

Oxford Centre for Molecular Sciences, New Chemistry Laboratory, Oxford University, South Parks Road, Oxford OX1 3QT, and Merton College, Oxford University, Oxford OX1 4JD, United Kingdom

Received December 23, 1996; revised March 11, 1997

Measurements of relaxation times in proteins are a valuable source of information about molecular motions in biological macromolecules such as proteins (*1–3*). Relaxation times are determined by sampling a relaxing magnetization after various delay times, and it is important to choose these delay times carefully in order to obtain accurate values. Here I describe a strategy for determining the optimal pattern of sampling times for the system under study.

The optimal sampling patterns for measuring a single relaxation time can be determined using Cramér–Rao lower bounds (*4–6*), as described in (*7*), and the essential results are summarized below. The time constant, T, of an exponentially decaying function

$$s(\tau) = Ae^{-\tau/T} \qquad [1]$$

can be determined by sampling the intensity at two time points, τ_1 and τ_2. The optimal sampling pattern for T is that which allows T to be determined with the greatest reliability (defined as the inverse of the fractional error in T), and the optimal locations for the two sampling points are to place one at the beginning ($\tau_1 = 0$), and the other at $\tau_2 = 1.11\ T$. With larger numbers of sample points, the results are similar: it is always best to place some of these sample points at zero time, and the rest at some optimal time, proportional to T. If the number of sample points is very large, the optimal pattern is to place 22% of the points at $\tau = 0$ and 78% at $\tau = 1.28\ T$, or more simply and almost as effectively, one point at $\tau = 0$ and four at $\tau = 1.30\ T$.

The choice of optimal sampling times for measuring relaxation times in proteins is more complicated, as different residues exhibit different relaxation behavior and so have their own optimal sampling patterns. It is not possible to choose an sampling pattern which is optimal for all the different values of T, and it is necessary to choose a compromise pattern which is reasonably good for all residues. In this case, the optimal sampling pattern will depend on how this compromise is made. A particularly simple approach is to measure the efficiency of a sampling pattern applied to several different values of T by the reliability achieved in the worst case. Furthermore, it is easier not to consider individual values of T directly, but simply optimize the sampling pattern for all T values in the range between the largest ($T^{\max}$) and smallest ($T^{\min}$) values considered.

When only two sample points are used, the optimal sampling patterns for a range of T values are easily determined (*7*). For narrow ranges ($T^{\max} \approx T^{\min}$), it is best to place one point at zero time, and the second point at $1.11\ T^{\mathrm{GM}}$, where $T^{\mathrm{GM}} = \sqrt{T^{\max}T^{\min}}$ is the geometric mean of the extreme values of T for the range under consideration. For wider ranges, this second point should be placed at a slightly earlier sampling time, $1.11\ T^{\mathrm{opt}}$. When $T^{\max} = 5\ T^{\min}$, this optimum time is $T^{\mathrm{opt}} = 0.9\ T^{\mathrm{GM}}$, and the worst-case reliability (which occurs at $T = T^{\max}$ and $T = T^{\min}$) is 69% of the best-case reliability (which occurs at $T = T^{\mathrm{opt}}$).

With larger numbers of sampling points, the optimal sampling patterns are more complicated. For narrow ranges of T, the optimal pattern remains equivalent to the optimal sampling pattern for some value, T^{opt}, which is approximately equal to, but slightly smaller than, T^{GM}. For wider ranges, it may be more efficient to place the sample points at several different times, and this can be explored numerically. As the optimization is over two or more variables, it is necessary to use multidimensional search routines. Several such routines are available (*8*), but many ''fast'' search routines are unstable when attempting to maximise worst-case reliabilities, and the best results were achieved with the slow but reliable AMOEBA algorithm (*8*). Furthermore, it is necessary to perform all numerical calculations using at least double-precision arithmetic, and in some cases, it is necessary to use quadruple precision.

The results for three sample points are shown in Fig. 1. For narrow ranges, it is best to place one sample point at zero and the other two points at the same time, $1.19\ T^{\mathrm{opt}}$. For very narrow ranges, $T^{\mathrm{opt}} \approx T^{\mathrm{GM}}$, but for wider ranges T^{opt} is slightly smaller as before. When $T^{\max} \geqslant$

1090-7807/97 $25.00

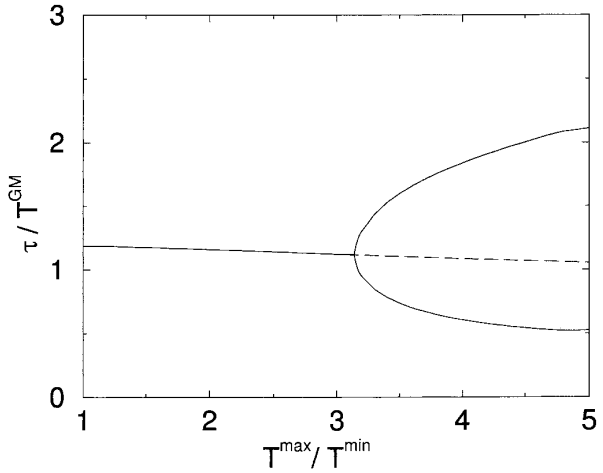

FIG. 1. Optimal sampling pattern for a range of values of T between $T^{\min}$ and $T^{\max}$ with three sample points. One point should be placed at zero, and the other two should be placed at multiples of $T^{GM} = \sqrt{T^{\min}T^{\max}}$, as indicated by the solid lines. For narrow ranges, the two sample points should be placed together, while for larger ranges they should be placed at two different times. The dashed line indicates the optimum location for these two points if they are constrained to lie together.

3.15 $T^{\min}$, the plot bifurcates, and it is no longer optimal to keep the two sample points together. As the range is increased further, the two sample times move further apart so that when $T^{\max} = 5\ T^{\min}$ they lie at 0.53 T^{GM} and 2.11 T^{GM}. Interestingly, the geometric mean of these two sample times lies close to the optimal sampling time if the two samples are constrained to occur at the same time (1.06 T^{GM}).

In any multidimensional search, there is a danger that the algorithm will not locate the true global maximum but rather some lesser local maximum. This can be guarded against by running the search algorithm with several different start points and also by performing crude searches involving the calculation of the function at many different points on an evenly spaced grid. Checks of this kind did not locate any better maxima, and it is probable that the maxima which have been located are the true global maxima, and so do indeed correspond to optimal sampling patterns.

With four sampling points the results are more complicated, as shown in Fig. 2. For narrow ranges, the optimal pattern comprises one point at zero and three at 1.25 T^{opt} as expected (*7*). When $T^{\max} \geqslant 2.92\ T^{\min}$, the plot bifurcates, and two sample points should be placed at an earlier time, while the last point should be placed at some later time. Finally, when $T^{\max} \geqslant 3.51\ T^{\min}$, the plot bifurcates again, and all three points should be placed at different times. For $T^{\max} = 5\ T^{\min}$, these times are 0.52 T^{GM}, 1.09 T^{GM}, and 2.69 T^{GM}.

As before, the geometric mean of these three times is close to the optimal time if the three samples are constrained to lie together. Furthermore, the quality of the results is only weakly dependent on the detailed values of the sampling times as long as their geometric mean lies close to the appropriate value. For this reason, it is difficult to locate the optimal sampling times exactly, and it is necessary to use quadruple-precision arithmetic to obtain accurate results, especially close to the two bifurcations.

With five sample points, it is not possible to locate the optimal sample times exactly, even using quadruple-precision arithmetic. The approximate results which can be obtained suggest that a bifurcation occurs when $T^{\max} \geqslant 2.75\ T^{\min}$, much as expected. A second bifurcation, however, is not observed, at least for the ranges considered ($T^{\max} \leqslant 5\ T^{\min}$). In this case, the optimal sampling pattern involves one point at zero and two points at each of two different sampling times, and this pattern (which has only two variable parameters) can be optimized with comparative ease. The results of this approach are shown in Fig. 3. When $T^{\max} = 5\ T^{\min}$, the optimal sample times are 0.60 T^{GM} and 2.31 T^{GM}. This result, which was previously discovered by an *ad hoc* search process (*7*), is now seen to be a member of a more general family.

Next it is necessary to consider the overall quality of these optimized patterns. Following earlier work (*7*), the efficiency of a sampling pattern is defined as the ratio of the worst-case reliability to the maximum reliability which can be achieved in the same time using a sampling pattern optimized for a single value of T. The results for patterns with two, three, and five sample points are shown in Fig. 4. For a two-point sample pattern, the efficiency is never higher than 88% and falls off rapidly as the width of the distribution of T values is increased. A three-point pattern is more efficient for narrow ranges (up to

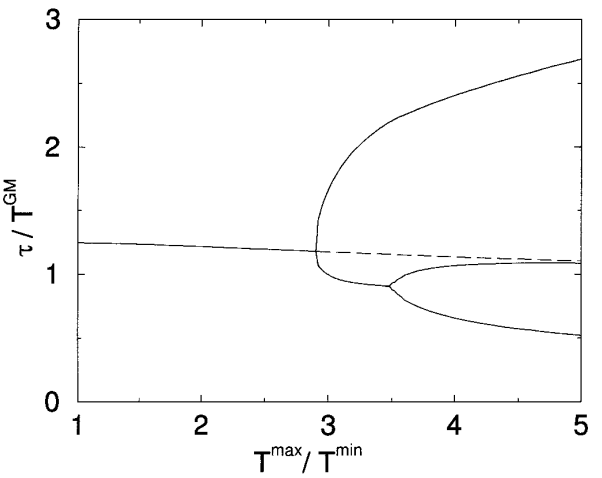

FIG. 2. Optimal sampling pattern for a range of values of T with four sample points. For details, see the legend to Fig. 1.

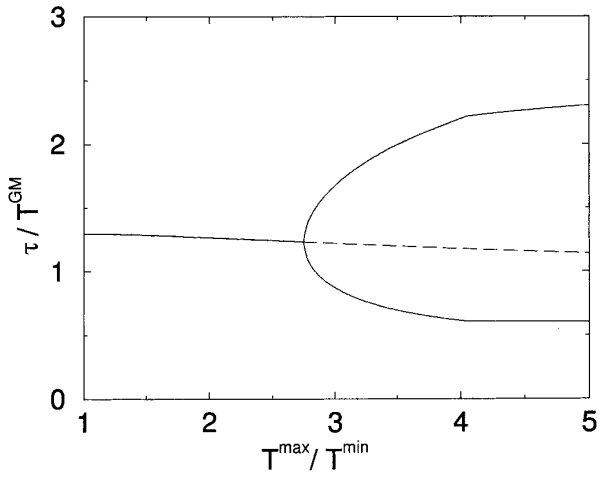

FIG. 3. Optimal sampling pattern for a range of values of T with five sample points. For details, see the legend to Fig. 1.

98%), and the efficiency falls off less rapidly after the bifurcation as a result of the greater flexibility in placing sample times. A five-point pattern is almost perfectly efficient for narrow ranges and remains at least 90% efficient as long as $T^{max} \leqslant 2\ T^{min}$. At wider ranges, the efficiency continues to fall, but remains at least 75% when $T^{max} \leqslant 5\ T^{min}$.

A potential weakness of these optimized sampling patterns is that they will not be very efficient at detecting deviations from exponential relaxation which may occur for a variety of reasons. Clearly any pattern involving only two distinct sample times cannot detect such deviations, while optimized patterns involving three or more sample times can detect them but may be less sensitive to these effects than more conventional patterns. In fact, this problem is less important than it may seem, as it is much more difficult to detect biexponential relaxation than is commonly realized: as discussed in (*7*), it is difficult to detect biexponential relaxation unless the two time constants differ by at least a factor of two, or the signal-to-noise ratio is at least 100.

Nevertheless, it is useful to consider some alternative sampling patterns involving a wide range of different sample times. A particularly simple sampling scheme is linear sampling, in which the sample times are spaced evenly between 0 and some maximum time, τ^{max}. The application of such sampling patterns to the measurement of a single relaxation time is discussed extensively in (*7*), and here I will only consider linear sampling with five sample times, which can be conveniently compared with the optimal five-point pattern, and linear sampling with an infinite number of sample times, which may serve as a model of conventional sampling schemes.

For a narrow range of values of T ($T^{max} \approx T^{min}$), the optimal five-point linear sampling pattern involves placing the five sample times between 0 and $\tau^{max} = 1.93\ T^{GM}$, and has an efficiency of 87% (the optimal five-point sampling pattern is almost perfectly efficient). As the range of T values is increased, the optimal value for τ^{max} increases very slightly, and the efficiency is decreased. For moderate ranges of T ($T^{max} \approx 3\ T^{min}$), the efficiency of the optimal five-point pattern decreases sharply, and the optimal linear pattern is only slightly less efficient (78%, compared with 80%). For wider ranges of T, however, the efficiency of the optimal five-point pattern decreases only slowly, and so this pattern is increasingly favored over the optimal linear sampling pattern: when $T^{max} = 5\ T^{min}$, the optimal five-point pattern has an efficiency of 75%, while that of the optimal linear pattern is only 70%.

As discussed in (*7*), linear sampling patterns with large numbers of sample times are significantly less efficient than optimized sampling patterns. For narrow ranges of T, the sample points should be evenly spaced between 0 and $\tau^{max} = 2.02\ T^{GM}$, resulting in an efficiency of only 74%. For wider ranges, τ^{max} increases very slightly and the efficiency is once again decreased: when $T^{max} = 5\ T^{min}$, the optimal linear pattern has an efficiency of only 59%.

In conclusion, it is possible to determine optimal sampling patterns for measuring relaxation times in complex systems such as proteins as long as the range of values of T is reasonably narrow. In general, the optimal five-point pattern, in which the sample times are chosen as shown in Fig. 3, is a simple and effective choice of sampling pattern for systems of this kind. For narrow ranges of T, this pattern is significantly more efficient than more conventional schemes, such as linear sampling, and these alternative schemes should only be used if nonexponential relaxation is likely to be a problem. For moderately wide

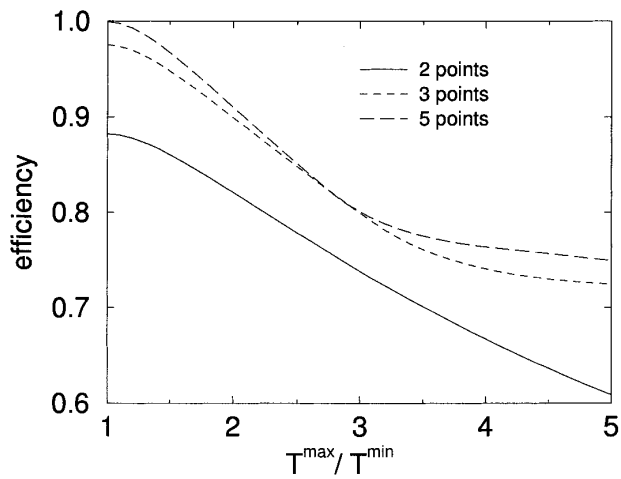

FIG. 4. Efficiency of various sampling patterns as a function of the width of the range of values of T.

ranges of T ($3\ T^{min} \leqslant T^{max} \leqslant 5\ T^{min}$), however, the loss of efficiency is less serious, and the use of these schemes is more appropriate. In particular, placing five sample points linearly between 0 and $1.96\ T^{GM}$ is fairly efficient over this range.

ACKNOWLEDGMENTS

I thank P. Hodgkinson and P. J. Hore for helpful discussions, and J. K. Elder and P. Biggs for help with coding the quadruple precision search routines. I am grateful to C. M. Dobson for his encouragement and support.

REFERENCES

1. L. E. Kay, D. A. Torchia, and A. Bax, *Biochemistry* **28,** 8972 (1989).
2. G. M. Clore, P. C. Driscoll, P. T. Wingfield, and A. M. Gronenborn, *Biochemistry* **29,** 7387 (1990).
3. J. W. Peng and G. Wagner, *Biochemistry* **31,** 8571 (1992).
4. A. van den Bos, *in* ''Handbook of Measurement Science'' (P. H. Sydenham, Ed.), Wiley, New York, 1982.
5. J. P. Norton, ''An Introduction to Identification,'' Academic Press, San Diego, 1986.
6. P. Hodgkinson and P. J. Hore, *Adv. Magn. Opt. Reson.,* in press.
7. J. A. Jones, P. Hodgkinson, A. L. Barker, and P. J. Hore, *J. Magn. Reson. B* **113,** 25 (1996).
8. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, ''Numerical Recipes in C,'' 2nd ed. Cambridge Univ. Press, Cambridge, U.K., 1992.