A COMPARISON OF VARIOUS MODELS FOR LIGAND RECOMBINATION KINETICS OF MYOGLOBIN

L. S. POWERS AND W. E. BLUMBERG
AT&T Bell Laboratories, Murray Hill, New Jersey 07974

INTRODUCTION

The dynamics of ligand binding to biological molecules is important to our understanding of biological processes. The experimental approach has classically been to measure the time course of the reaction rate by monitoring a spectroscopic marker. The data are then fit with an expression containing a multiplicity of exponentials because only simple cases, such as radioactive decay, can be adequately described by a single exponential process. The number of exponentials included is determined not only by the number of distinct processes which are sampled by the specific spectroscopic marker that is monitored but also by the quality of the data which determines the ability to separate similar rates in the fitting procedures. This methodology has the underlying assumption that, although each molecule is not identical to every other molecule at all points in time during the reaction, these differences do not have a significant or measurable effect on the reaction rates or, alternatively, that the distribution around each of the rates is within the error of the determination and analysis. This method was first used by Iizuka et al. (1) and later by Chance and Powers et al. (2, 3) to describe the low temperature recombination data of photolyzed carboxy-myoglobin. This method of analysis is compared with three other physical models of the recombination process.

If the assumptions underlying the exponential analysis (herein called model A) are not true for a particular reaction, i.e., there is a distribution of reaction rates, then the expression becomes more complicated. One such representation is an integral over an exponential term, which is a power law expression (model B) and was proposed by Frauenfelder et al. (4, 5) for the recombination of ligands with myoglobin using the optical absorption band at 436 nm and more recently by Frauenfelder et al. (6, 7) and others using other visible and infrared bands.

If the rate integral is weighted by a function which vanishes at zero and infinity and is otherwise well-behaved, then the computed time course of the reaction can have almost any arbitrary shape. Frauenfelder et al. (5), using an inverse Laplace transform technique, have proposed one such distribution (model C). This distribution was used as a fitting function, varying the scaling and center of gravity of the distribution as necessary for the best fit.

Lastly taking the suggestion of Stapleton et al. (8) that the phonon dynamics of a protein structure is best represented by the mathematics of fractals, one can hypothesize a similar nature for the recombination process. This leads to a time course similar to the power law expression mentioned above but, in this case, multiplied by a function periodic in the logarithm of time (9, 10) (model D).

The conceptual difference between model A and models B and C is significant. In the first, large numbers of molecules function similarly although they are not identical in every aspect, whereas in the latter two each molecule functions somewhat differently depending on its other properties such that a large number of molecules represent a distribution in function. Model D can be used to describe either physical situation. Typical fractal mathematics arises from the situation in which completion of a process, e.g., ligand recombination, requires the sequential completion of a large number of subprocesses at statistical rates. One can imagine a set of molecules with identical paths leading to ligand recombination. If each path is composed of many parts requiring varying amounts of time to complete, fractal mathematics will result. Likewise, a set of molecules, all different, with nonidentical paths leading to the same point will exhibit fractal behavior provided that the subprocesses comprising any one path proceed with a wide range of individual rates.

Understanding the differences in these models is pivotal in understanding the more complex biological processes, and it is crucial that the systematic differences be recognized. In this communication, the differences between the application of these models are examined with statistical techniques and comparisons using the data from the Frauenfelder group (5, 7).

METHODS

The data were taken from Frauenfelder et al. (5, 7) by computer digitizing methods where each data set for a given temperature was digitized five times. Comparison of these showed that $N(t)$ and $t(s)$ differed by <0.3% and individual analysis showed no differences within the fitting error.

All analysis procedures used a nonlinear least squares fit to the data. The two-parameter correlation matrix and the eigenvalues and eigenvectors of the Hessian matrix were calculated for each fit and the residuals were examined. The goodness of fit was judged by the average (rms) residual $\sqrt{\Sigma R^2}/N$, where N is the number of points in the data set.

The exponential model (model A) is given by

$$N(t) = N_0 \sum_q c_q e^{-k_q t},$$

where N_0 is the scaling factor which represents the degree of photolysis at $T = 0$ and c_q are the fractions having rate constants k_q, respectively. The maximum value of q was restricted to 4 because a well-defined minimum could not be obtained at higher values, and this behavior is indicative that the data cannot support additional parameters. The c_q values mathematically determine N_0: when q is large, a range of values for these parameters exist which cannot be discriminated by the rms residual or the respective k values.

The power law model (model B) is described by

$$N(t) = N_0 [1 + t/t_0]^n,$$

where N_0 is as just described and t_0 and n are temperature-dependent parameters.

Model C replaces this power law approximation by an actual integral over a continuous distribution of exponential terms.

$$N(t) = \int dE_{ba}\, g(E_{ba}) \exp(-k_{ba} t),$$

where $g(E_{ba})$ is the distribution function proposed by Frauenfelder et al. (5). This function was permitted some rescaling and shifting of the centroid to fit each individual data set.

For model D the form of model B was extended by including a factor of the form

$$(1 + m \sin [\ln(t)/\ln(t_s) + \phi]),$$

providing modulation depth m, scaling frequency $1/\ln(t_s)$, and phase ϕ as additional fitting parameters. The exponent of the power law function, given above as n, is then related to the Hausdorff dimension of the fractal system (11).

RESULTS

Some of the results for fitting the models to the digitized data of Frauenfelder et al. (5, 7) are given in Table I

TABLE I
BEST FIT PARAMETERS* FOR FOUR RECOMBINATION MODELS

					Model A‡				
T§	c_1	k_1	c_2	k_2	c_3	k_3	c_4	k_4	R_{rms}
K									
30	0.02	4.1	0.04	0.083	0.94	0.00016			0.0033
40	0.04	6.2	0.10	0.097	0.86	0.00040			0.0037
92	0.32	2200	0.35	93	0.22	2.2	0.11	0.011	0.0138
120	0.59	26500	0.23	1720	0.12	80	0.06	2.3	0.0052
139	0.47	49000	0.38	4000	0.15	100			0.0086
160	0.57	180000	0.29	13000	0.14	460			0.0144
					Models B and D				
T		n	t_0	m	$1/t_s$	ϕ	R_{rms}		
K									
30	B	0.013	0.44				0.0038		
	D	0.019	1.68	0.0066	0.57	5.0	0.0024		
40	B	0.033	0.31				0.0062		
	D	0.041	0.80	0.013	0.56	5.4	0.0012		
92	B	0.22	28e-5				0.0109		
	D	0.22	27e-5	0.0038	3.6	2.6	0.0108		
120	B	0.32	2.3e-5				0.0110		
	D	0.45	2.0e-5	0.037	3.3	0.50	0.0088		
139	B	0.44	1.5e-5				0.0056		
	D	0.33	3.0e-5	0.034	5.9	5.8	0.0038		
160	B	0.44	0.16e-5				0.0453		
	D	0.12	5.6e-5	0.077	7.2	1.3	0.0404		
					Model C				
T	Scale	Shift	R_{rms}						
K									
30	0.92	0.063	0.0032						
40	0.84	0.049	0.0025						
92	0.60	0.037	0.0060						
120	0.60	0.038	0.0074						
139	0.68	0.039	0.0050						
160	0.67	0.032	0.0392						

*The normalization N_0 is omitted as it does not help distinguish between models. ‡c_i is in fraction, k_q is in s^{-1}, t_0, t_s are in s. §Data for T = 30, 40, 92, and 139 K are from reference 7; data for T = 120 and 160 K are from reference 5.

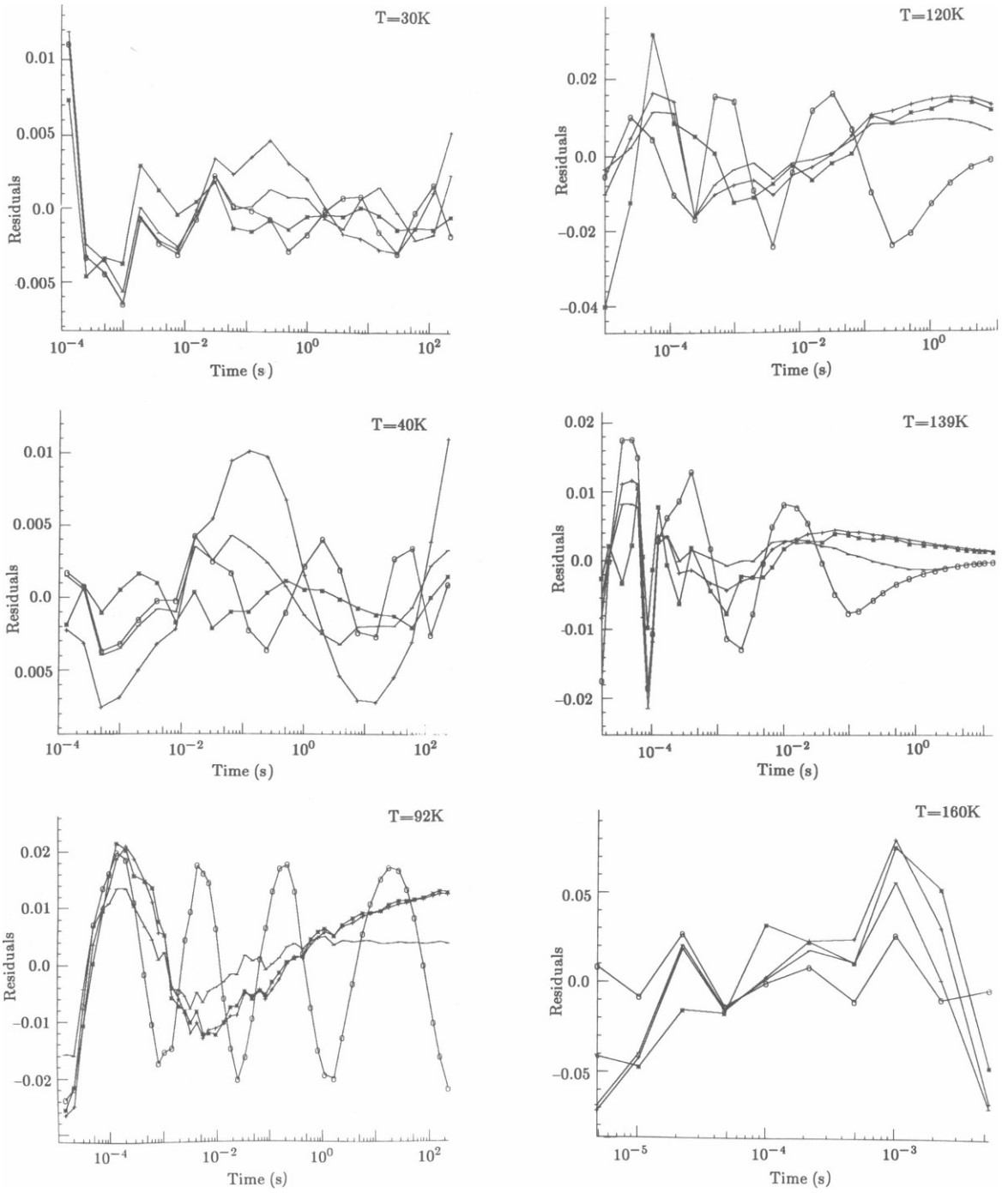

FIGURE 1 Residuals of the fits of the various models vs. time for six selected temperatures. Models A, B, C, and D are denoted by symbols O, +, −, and *, respectively.

together with the rms residuals, and Fig. 1 compares the residuals for the four methods. To insure that the digitized data and the fits to model B resemble those reported by Frauenfelder et al. (reference 5, Fig. 12), the effective activation energy E was calculated from the peak k values ($k_{pk} = n/t_0$), and the comparison is shown in Fig. 2. It is clear that they are similar within the error.

From the residuals of the four fits at each temperature, it is clear that the data can be fit almost equally well by all the models. At every temperature, the integral over the distribution (model C) and the fractal function (model D) fit the data better than the power law (model B), but in most cases only marginally better. Model D, of course, would be expected to fit better as it has six parameters whereas models B and C have only three each.

Fig. 3 is an Arrhenius plot of the respective k values and

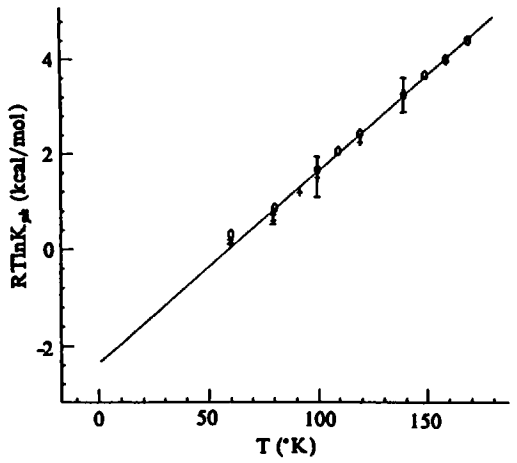

FIGURE 2 Comparison of the effective activation energy ($RT\ln K_{pk}$) vs. temperature for model B (+) and that reported by Frauenfelder et al. (5, Fig. 12) showing they are similar within the error.

errors using models A and B. It is important to note that in both models, the slope changes abruptly between 80 and 92 K, indicating a transition region.

Until now the error inherent in the data has not been explicitly considered. Certainly, the rms residuals are directly effected by this error. Frauenfelder et al. (5) give this error for the 160K data and it is in good agreement with the ±~8% difference between the two data sets at 140 K (5, 7). This difference is ±~4% for the two data sets at 40 K (5, 7). If this error is also valid for the data at other temperatures, then only a few points at the longer times at some temperatures have error that is smaller than the differences among any of the models.

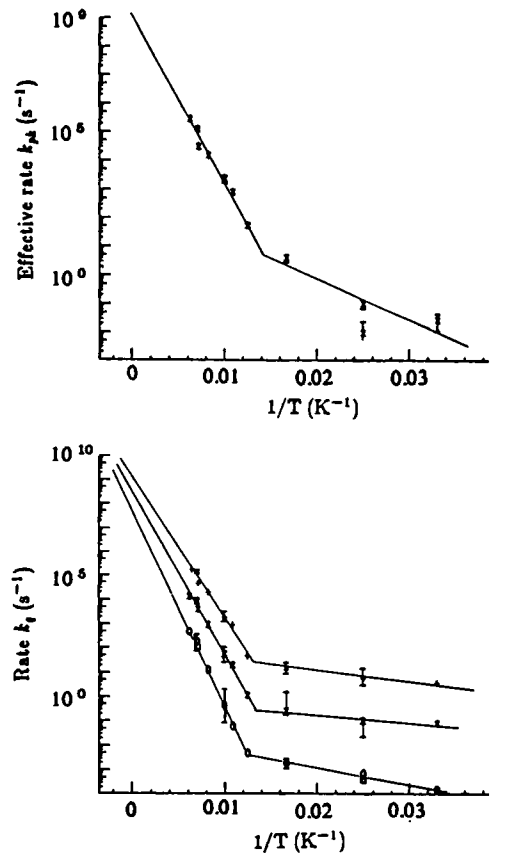

FIGURE 3 An Arrhenius plot of the results of fitting models A and B. (*Top*) Effective k for model B as defined in the text. (*Bottom*) k_q for model A using three exponentials. The transition region near T = 80K is exhibited regardless of choice of model.

DISCUSSION

It should be noted that all models express the major features of the data. All indicate that the recombination process involves more than one subprocess. However, only one of the fractions in the exponential model (model A) amounts to >10% at 30 and 40 K. Between 80 and 92 K, a transition region is indicated by both the exponential and power law models where effective k values can be easily obtained. It is important to note that the effects of the underlying physical processes exhibit themselves regardless of the model used for analysis.

It is often confusing to attempt to compare different models having different functional forms and different numbers of fitting parameters using only ΣR^2. One technique which is useful in this case is to compare the values of Mallows' statistic C_p (see, for example, reference 12). For this statistic an estimate of the fitting error due solely to random error in the data (s^2) independent of fitting bias is needed. Fitting bias describes the correlation among successive residuals and appears as a very low frequency noise component. Sometimes this effect is illustrated by plotting the autocorrelation of the residuals (13). An estimate of s^2 can be obtained with a so-called complete fitting function (a nonphysical flexible function with an embarrassingly large number of adjustable parameters), or it can be estimated from several of the best fits, making allowance for the presence of a small amount of fitting bias. The latter procedure leads to an estimate that the overall random noise component is about $\sqrt{s^2} = 0.001$. Fortunately, for comparison purposes, this method is not very sensitive to this value provided a single value applies to all the data. Then Mallows' statistic is given by

$$C_s = \frac{\Sigma R^2(N-p)}{s^2 N} - N + 2p,$$

where the fitting function has p parameters and the data has N points. If $\Sigma R^2/N$ is the same as s^2, then the fitting errors are truly random, and C_p values for different fits will cluster around the straight line $C_p = p$. When C_p falls significantly above this line, the fit is dominated by fitting bias. Two or more otherwise disparate models can then be compared by the distance that C_p lies above the line $C_p = p$. To compare fits having different numbers of points in the data set as well as different numbers of fitting parameters,

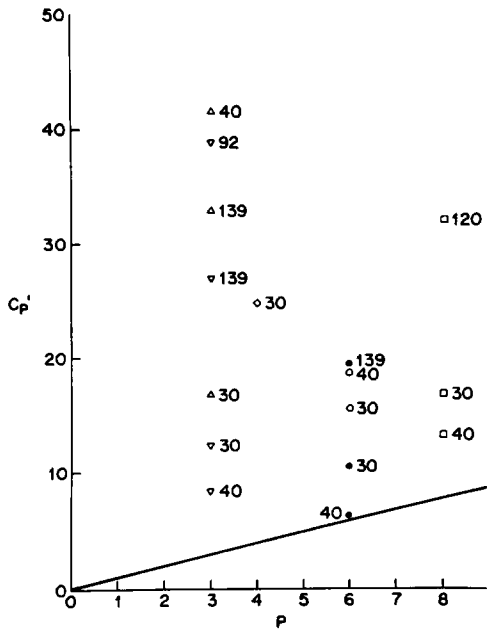

FIGURE 4 Mallows' method for comparison of fits using different models. The number of fitting parameters is p, and C_p' is defined in the text. The slanted line is the equation $C_p = p$. The lower the value of C_p', the better the model fits the data; any value not shown is off the top of the graph. The models are denoted by the symbols: (◇) two-exponential, (○) three-exponential, (□) four-exponential, (△) model B, (▽) model C, (●) model D. The numbers beside each point indicate the temperature at which the data was taken.

that part of C_p arising from bias (that part above the line) has been divided by $N - p$, and the modified statistic is called C_p'. Some of the resultant values of C_p' are given in Fig. 4. The fits at almost all temperatures for all models are seen to be dominated by bias. In addition one can see that, among the best fits at various temperatures, no one model clearly stands out as "the best."

It is unfortunate that published low-temperature myoglobin photolysis data cannot clearly distinguish the important physical differences among these models. The experimental data can be considered to be consistent with any of the physical models thus far proposed to describe ligand recombination in myoglobin.

Received for publication 9 December 1987 and in 21 March 1988.

REFERENCES

1. Iizuka, T., H. Yamamoto, M. Kotani, and T. Yonetani. 1974. Low-temperature photodissociation of hemoproteins: oxygenated cobalt-myoglobin and hemoglobin. *Biochim. Biophys. Acta.* 351:182–295.
2. Chance, B., R. Fischetti, and L. Powers. 1983. Structure and kinetics of the photoproduct of carboxymyoglobin at low temperatures: an x-ray absorption study. *Biochemistry.* 22:3820–3829.
3. Powers, L., B. Chance, M. Chance, B. Champbell, J. Friedman, S. Khalid, C. Kumar, A. Naqui, K. S. Reddy, and Y. Zhou. 1987. Kinetic, structural, and spectroscopic identification of geminate states of myoglobin: a ligand binding site on the reaction pathway. *Biochemistry.* 26:4785–4796.
4. Austin, R., K. Beeson, L. Eisenstein, H. Frauenfelder, I. Gunsalus, and V. Marshall. 1973. Dynamics of carbon monoxide binding by heme proteins. *Science (Wash. DC).* 181:541–543.
5. Austin, R. H., K. W. Beeson, L. Eisenstein, H. Frauenfelder, and I. C. Gunsalus. 1975. Dynamics of ligand binding to myoglobin. *Biochemistry.* 14:5355–5373.
6. Ansari, A., J. Berendzen, S. Bowne, H. Frauenfelder, I. Iben, T. Sauke, E. Shyamsunder, and R. Young. 1985. Protein states and proteinquakes. *Proc. Natl. Acad. Sci. USA.* 82:5000–5004.
7. Ansari, A., J. Berendzen, D. Braounstein, B. R. Cowen, H. Frauenfelder, M. K. Hong, I. E. T. Iben, J. B. Johnson, P. Ormos, T. B. Sauke, R. Scholl, A. Schulte, P. J. Steinbach, J. Vittitow, and R. D. Young. 1987. Rebinding and relaxation in the myoglobin pocket. *Biophys. Chem.* 26:337–355.
8. Stapleton, H. J., J. P. Allen, C. P. Flynn, D. G. Stinson, and S. R. Kurtz. 1980. Fractal form of proteins. *Phys. Rev. Lett.* 45:1456–1459.
9. Montroll, E. W., and M. F. Schlesinger. 1982. On 1/f noise and other distributions with long tails. *Proc. Natl. Acad. Sci. USA.* 79:3380–3383.
10. West, B. J., and A. L. Goldberger. 1987. Physiology in fractal dimensions. *Am. Sci.* 75:354–365.
11. Mandelbrot, B. B. 1982. The Fractal Geometry of Nature. W. H. Freeman, New York.
12. Daniel, C., and F. S. Wood. 1971. Fitting equations to data. Wiley-Interscience, New York. p. 87ff.
13. Grinvald, A., and I. Z. Steinberg. 1974. On the analysis of fluorescence decay kinetics by the method of least squares. *Anal. Biochem.* 59:583–598.