Two-Dimensional Distributions of Activation Enthalpy and Entropy from Kinetics by the Maximum Entropy Method

Peter J. Steinbach
Laboratory of Structural Biology, Division of Computer Research and Technology, National Institutes of Health, Bethesda, Maryland 20892 USA

ABSTRACT The maximum entropy method (MEM) is used to numerically invert the kinetics of ligand rebinding at low temperatures to obtain the underlying two-dimensional distribution of activation enthalpies and entropies, $g(H,S)$. A global analysis of the rebinding of carbon monoxide (CO) to myoglobin (Mb), monitored in the Soret band at temperatures from 60 to 150 K, is performed using a Newton-Raphson optimization algorithm. The MEM approach describes the data much better than traditional least-squares analyses, reducing χ^2 by an order of magnitude. The MEM resolves two barrier distributions suggestive of rebinding to different bound conformations of MbCO, the so-called A_1 and A_3 substates, whose activation barriers have been independently estimated from kinetics monitored in the infrared. The distribution corresponding to A_3 possesses higher activation entropies, also consistent with infrared measurements. Within an A substate, correlations of S and H are recovered qualitatively from simulated data but can be difficult to obtain from experimental data. When the rebinding measured at 60 K is excluded from the inversion, two peaks are no longer clearly resolved. Thus, data of very high quality are required to unambiguously determine the kinetic resolvability of subpopulations and the shape of the barrier distribution for a single A substate.

INTRODUCTION

For more than 20 years, the dynamics of heme proteins have been studied in great detail by a variety of techniques (Frauenfelder et al., 1988). In flash-photolysis experiments, the covalent bond between protein and ligand is broken by the absorption of light (e.g., from a laser flash), and the ensuing kinetics of ligand rebinding are monitored optically. Analysis of the rebinding of carbon monoxide (CO) to myoglobin (Mb) has contributed greatly to our understanding of proteins as dynamic systems (Austin et al., 1975; Steinbach et al., 1991). The measured kinetics support the view of protein conformation in which each protein can assume a variety of energetically similar configurations or conformational substates (Ansari et al., 1985).

In a 75% glycerol/water solvent near 300 K, rapid equilibrium fluctuation among these substates permits ligand escape into the solvent and creates an averaged, effective barrier to ligand rebinding as evidenced by approximately exponential kinetics (Austin et al., 1975). After flash photolysis at temperatures near 180 K, the proteins relax on the time scale of rebinding, resulting in rebinding that is neither exponential in time nor Arrhenius in temperature (Agmon and Hopfield, 1983; Steinbach et al., 1991; Panchenko et al., 1995). At temperatures below about 160 K, ligands can no longer escape to the solvent, and fluctuation among substates is frozen out. Proteins in different substates can present different barriers to ligand rebinding and the observed kinetics are highly nonexponential, with rebinding rates ranging over more than ten orders of magnitude.

This study addresses the analysis of this geminate rebinding measured in the absence of conformational relaxation. This regime shall be somewhat casually referred to as "low temperature," with the understanding that it may persist up to room temperature in some solvents (Hagen et al., 1995). Traditionally, these kinetics have been described by a temperature-independent distribution of activation enthalpies, with each protein possessing the same activation entropy (Austin et al., 1975). The enthalpy distribution inferred from CO rebinding to Mb in a 75% glycerol/water solvent below 160 K is peaked near 10 kJ/mol. For systems characterized by such large enthalpic barriers, the data can be described reasonably well by using a single value for the activation entropy.

Several heme-ligand systems possess smaller activation enthalpies, and for these, the effects of distributed entropy are more important. The enthalpy distribution is peaked at $H \approx 1$ kJ/mol or less for CO rebinding to protoheme (Alberding et al., 1976; Miers et al., 1991), at 4 to 5 kJ/mol for CO rebinding to separated α and β chains of hemoglobin (Alberding et al., 1978), and at 2.3 kJ/mol for CO rebinding to the β chain of hemoglobin Zürich (Dlott et al., 1983). Measured CO rebinding to protoheme from 20 to 80 K has been well described by assuming a linear correlation between activation enthalpy and entropy (Alberding et al., 1976). In this way, an entropy distribution was tied to an enthalpy distribution. At higher temperatures, the distribution of rebinding rates is less sensitive to the distribution in activation enthalpy, facilitating a more direct assessment of entropic effects, and an entropy distribution has been fit to

Received for publication 11 October 1995 and in final form 28 November 1995.

Address reprint requests to Dr. Peter J. Steinbach, National Institutes of Health, DCRT, LSB, Bldg. 12A, Rm. 2041, 12 South Drive, MSC 5626, Bethesda, MD 20892-5626. Tel.: 301-496-6520; Fax: 301-496-2172; E-mail: steinbac@helix.nih.gov.

0006-3495/96/03/1521/08 $2.00

ultrafast protoheme-CO kinetics at 200 K (Miers et al., 1991).

Even for CO rebinding to Mb, evidence for different entropic barriers comes from flash-photolysis experiments monitored in the infrared (Ansari et al., 1987). Three predominant bound conformations of MbCO, referred to as A_0, A_1, and A_3, are distinguished by different stretch frequencies (near 1950 cm^{-1}) of the ligand's C-O bond. Monitoring the three stretch bands individually after flash photolysis at low temperatures shows that the geminate rebinding to each substate is nonexponential, with rebinding to A_0 being the fastest and rebinding to A_3 being the slowest. The infrared kinetics indicate different activation entropies as well as different activation enthalpies for the three substates.

When rebinding is monitored in the Soret band near 440 nm, the absorption of the heme's π-electron system is probed. The Soret signal is much larger than the C-O stretch signals, but it cannot be used to monitor the A substates individually. Because CO rebinding to Mb monitored in the Soret represents a sum over the three A substates, an accurate determination of the temperature-independent distribution of barriers is made more complicated. Generally, the enthalpy distribution has been approximated as a parameterized function that is fit to the data, requiring only a few adjustable parameters but yielding values of $\chi^2 = \mathcal{O}(10^3)$.

A numerical determination of the distribution that does not force a functional form on the result is desirable. Such an approach would more fully exploit the information content of the data, converging to lower values of χ^2 and possibly resolving the kinetics corresponding to different A substates. The fundamental shape of the barrier distribution may also emerge from a stable numerical inversion of the kinetics. The shape has been of considerable interest in both experimental (Berendzen and Braunstein, 1990; Steinbach et al., 1992; Ehrenstein and Nienhaus, 1992) and theoretical (Agmon and Hopfield, 1983; Young and Bowne, 1984; Stein, 1985; Šrajer et al., 1988) studies. The theoretical models differ in the assumed shape of the protein energy landscape and in the modeled coupling between conformation and activation enthalpy. For example, Agmon and Hopfield assumed a harmonic landscape and an activation enthalpy that is nearly linear in the protein conformational coordinate, giving rise to a nearly gaussian enthalpy distribution. Stein has suggested a distribution that is the product of a gaussian and an exponential. The models proposed by Young and Bowne and Šrajer et al. involve distributions skewed to high H. A numerical inversion could assess the ability of the data to distinguish between such models by characterizing the sensitivity of the recovered distribution to the quality of the data (e.g., the time and temperature ranges measured, the signal-to-noise level, etc.).

A recent numerical approach partitions the enthalpy axis into 32 bins. Each H bin contributes two adjustable parameters to a fit to rebinding data; to each value of H is attributed a value of the probability density in H and a single activation entropy (Hagen et al., 1995). Consequently, this approach can resolve structure in the enthalpy distribution. However, in general, multiple S values may correspond to a given value of H, and use of a single S value can inhibit the recovery of S-H correlations.

Here, the maximum entropy method (MEM) is used to numerically invert the geminate rebinding data measured in the absence of protein relaxation to obtain a static two-dimensional distribution in activation enthalpy and entropy, $g(H,S)$. Thus, entropy and enthalpy are treated equally generally in an attempt to infer the barrier distribution directly from the data. The next section defines the $g(H,S)$ distribution to be recovered from the low-temperature kinetics. The MEM rationale is then summarized, followed by an overview of the algorithm developed for this work. The performance of the algorithm is characterized by numerical inversions of simulated data sets before being applied to experimental data. The two-dimensional distributions of activation barriers recovered from the simulated and actual data sets are presented and discussed.

LIGAND REBINDING AT LOW TEMPERATURES

After flash photolysis at low temperatures, protein relaxation contributes negligibly to changes in the absorption, ΔA, of the sample. The measured signal is attributable solely to ligand rebinding as governed by a time- and temperature-independent distribution of barriers. In general, both activation enthalpy and entropy are distributed:

$$\Delta A(t, T) = \Delta A_{max} \mathcal{N}(t, T) = \Delta A_{max} \int_{-\infty}^{\infty} dS \int_{0}^{\infty} dH g(H,S) e^{-k(H,S,T)t}, \quad (1)$$

where $\mathcal{N}(t, T)$ is the fraction of proteins at temperature T that have not yet rebound a ligand at time t and $g(H,S)\, dH\, dS$ is the fraction of proteins in conformational substates possessing an activation enthalpy between H and $H + dH$ and an activation entropy between S and $S + dS$. The normalization constant, ΔA_{max}, can be estimated reliably from the kinetics measured at a very low temperature, e.g., 10 K. The rate coefficient can be written as

$$k(H,S, T) = \nu(T/T_0)^{\xi} e^{-G/RT}, \quad (2)$$

where the activation Gibbs free energy is $G = H - TS$, ν is an attempt frequency taken to be 10^{12} s^{-1}, T_0 is taken to be 100 K, and ξ is a parameter often taken to be 0 or 1 (Dlott et al., 1983).

To compare the current results to traditional analyses in which only enthalpy is distributed, it is convenient to integrate the probability density over S and to average the logarithm of A, the pre-exponential of k:

$$g'(H) = \int_{-\infty}^{\infty} dS g(H,S) \quad (3)$$

and

$$\langle \log A \rangle(H) = \frac{\int_{-\infty}^{\infty} dS g(H, S) \log(\nu e^{S/R})}{g'(H)}, \quad (4)$$

where log connotes base 10.

NUMERICAL INVERSIONS VIA THE MAXIMUM ENTROPY METHOD

Consider a discretized data set that is linearly related to a distribution or image, $\vec{f}$:

$$\text{Data}_i = \sum_{j=1}^{M} c_{i,j} f_j + B_i + \eta_i. \quad (5)$$

B_i accounts for any background signal present. The noise, η_i, shall be assumed gaussian with standard deviation σ_i. For the current application, the $\Delta A(t, T)$ data are represented as a vector, the ith point having been measured at time t_i and temperature T_i. Similarly, $\vec{f}$ is $g(H,S)$ represented as a vector with the jth element associated with activation enthalpy H_j and activation entropy S_j. In the discretized H-S plane, enthalpies are separated by dH and entropies by dS. The physics of Eqs. 1 and 2 are encoded in the matrix c: $c_{i,j} = \Delta A_{max} dS\, dH e^{-k(H_j,S_j,T_i)t_i}$. The uncertainty σ_i is estimated by signal averaging (see below), and $B_i = 0$.

In numerical inversions, a representative distribution must be chosen from all those consistent with the noisy data. That is, when judged solely by a goodness-of-fit statistic such as χ^2,

$$\chi^2 = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\text{Fit}_i - \text{Data}_i}{\sigma_i} \right)^2, \quad (6)$$

many distributions, with different amounts of structure, produce equally acceptable fits to the N data points.

From all distributions yielding a particular value of χ^2, the MEM chooses that which maximizes the entropy given by (Skilling, 1989):

$$S(\vec{f}, \vec{F}) = \sum_{j=1}^{M} [f_j - F_j - f_j \ln(f_j/F_j)]. \quad (7)$$

$\vec{F}$ is the default distribution used to incorporate desirable characteristics in $\vec{f}$ based on previous experiments, e.g., the expectation that the distribution should be smooth. It defines maximum entropy; unconstrained maximization of S with respect to the M values of f_j yields $f_j = F_j$.

The MEM has been used to invert a variety of experimental measurements (Gull and Skilling, 1984). It is particularly applicable to linear transformations (Eq. 5) involving positive, additive distributions. For such transformations, there is a unique distribution $\vec{f}$ that maximizes S for a given value of χ^2 and default distribution $\vec{F}$. The natural logarithm in Eq. 7 ensures that all f_j are positive. Clearly, the distribution is biased in some way by any method that selects a single distribution from the many that yield a given value of χ^2. By biasing $\vec{f}$ toward $\vec{F}$, the MEM introduces no correlations in $\vec{f}$ that are not present in the data or in $\vec{F}$. In other words, it is "maximally noncommittal" with regard to unavailable information. Consequently, as the signal-to-noise ratio of the data decreases, structure in the distribution is washed out without the introduction of spurious peaks (Steinbach et al., 1992).

One way to maximize S for a given value of χ^2 is to maximize the function

$$Q \equiv S - \lambda \chi^2.$$

An iterative optimization is begun from $\vec{f} = \vec{F}$, $\lambda \approx 0$. The Lagrange multiplier λ is adjusted to constrain χ^2 to the desired value, say $\chi^2_{aim} \leq 1.0$, at the maximum of Q.

In a Newton-Raphson optimization, Q is approximated to second order in the neighborhood of the current point $\vec{f}$. The update to $\vec{f}$ is obtained as the solution of a linear system of M equations. This requires $\mathcal{O}(M^3/3)$ operations and is repeated at multiple points $\vec{f}$ as iterations converge to $\chi^2 = \chi^2_{aim}$. Because this approach is not feasible when recovering images made up of $M = \mathcal{O}(10^6)$ pixels, a rather complicated MEM algorithm has been developed that solves a linear system of equations in a small subspace of only a few dimensions (Skilling and Bryan, 1984).

Cornwell and Evans (1985) have used an efficient approximation to a true Newton-Raphson optimization of Q for the deconvolution of large astronomical images via the MEM. By approximating the second-derivative matrix, or Hessian, of Q as a diagonal matrix, they avoided the expensive $\mathcal{O}(M^3/3)$ operations. To compensate for this approximation, line searches were performed along the update to $\vec{f}$. The automated adjustment of λ was performed gradually to keep $|\vec{\nabla} Q|$ acceptably small.

Not all transformations are amenable to a diagonal approximation of the Hessian. Nonetheless, straightforward Newton-Raphson optimization may be expedient whenever modest-sized distributions are to be recovered from experiments that take a few hours or longer to complete. Such is the case when inverting $\Delta A(t,T)$ to obtain $g(H,S)$, a distribution requiring only $M = \mathcal{O}(10^3)$ pixels.

CALCULATING *g(H,S)*

For the recovery of $g(H,S)$, the MEM algorithm developed by Cornwell and Evans was generalized to perform a true Newton-Raphson optimization of Q. A diagonal approximation to the Hessian was used only at small values of λ. Line searches (Press et al., 1992) were performed, first attempting a full Newton-Raphson increment to $\vec{f}$ and, if necessary, reducing the length of the increment until Q increased sufficiently. The full step was generally always accepted. Matrix algebra was performed using the public-domain LINPACK routines (Dongarra et al., 1979).

In contrast to previous work (Steinbach et al., 1992), the noise was approximated as gaussian in ΔA, not in transmittance. The transmittance, $\mathcal{T}$, is the ratio of light intensity transmitted through the sample at time t to that transmitted just prior to the reaction-initiating laser pulse: $\mathcal{T}(t,T) \equiv I(t, T)/I(0^-, T) = 10^{-\Delta A(t,T)}$. Because intensities are actually measured, the standard errors in the mean were originally estimated in $\mathcal{T}$ space as $\sigma_{\mathcal{T}i}$. However, it is convenient to invert the data in ΔA space because the Hessian of χ^2 is independent of the current point $\vec{f}$ in this space. For data of sufficient signal-to-noise ratio, it is reasonable to approximate the uncertainties in ΔA as $\sigma_i = \log(e)\sigma_{\mathcal{T}i}/\mathcal{T}_i$.

Iteration was terminated when χ^2 stopped changing appreciably while keeping $|\vec{\nabla}Q|$ small. Distributions stored during convergence were then compared. The value of λ (i.e., χ^2) yielding the distribution that optimally trades smoothness for goodness of fit is not obvious. Gull has proposed a criterion that seeks an amount of structure in equal to the number of good, independent measurements $\vec{f}$ present in the data (Gull, 1989). Here, the presence of nonstatistical noise in the experimental data complicates the issue, and an admittedly subjective but reasonable choice of representative distribution was adopted. The g(H,S) distribution with the lowest value of χ^2 that resulted in a smooth $g'(H)$ function was chosen.

In many applications improved distributions have been found by convolving the distribution $\vec{f}$ corresponding to a uniform $\vec{F}$ with a gaussian to create a new $\vec{F}$ for a second calculation (Cornwell and Evans, 1985; Gull, 1989). In this way, spatial correlations of low frequency (broad features) are included in the definition of maximum entropy, reducing the bias incurred with a uniform $\vec{F}$. For the calculation of g(H,S), this approach was found to yield somewhat better S-H correlations within subpopulations. Relative to distributions obtained with a uniform $\vec{F}$, $\langle \log A \rangle(H)$ was more faithfully determined from synthetic data when a blurring gaussian with full width at half-maximum of 3 to 5 pixels was used to create $\vec{F}$ from a previous $\vec{f}$.

RESULTS FOR SIMULATED $\Delta A(t,T)$

Before a MEM algorithm can be applied with confidence to experimental data, it must successfully converge to $\chi^2 \leq 1$ when applied to stimulated data for which the transformation from distribution to data and the noise are known accurately. Failure to do so indicates algorithmic deficiencies, and convergence was achieved here only with true Newton-Raphson optimization using the full Hessian matrix to update $\vec{f}$.

The algorithm was tested with two data sets that were simulated using Eqs. 1 and 2 (with $\xi = 1$) and a random number generator (Press et al., 1992) to add gaussian noise. The times, temperatures, uncertainties, and ΔA_{max} from an experimental data set (Steinbach et al., 1991) were used to simulate $\Delta A(t,T)$ (Fig. 1). The other data set was extended at each temperature by five points before the first datum and five after the last datum. Consequently, the earliest time point in this extended set was ΔA(6.1 ns, 150 K) = 1.085 and the latest was ΔA(15.6 ks, 60 K) = 0.091. The g(H,S) distribution used to create the two data sets and the distributions recovered by the MEM are shown in Fig. 2, *a–c*.

The "true" g(H,S) distribution contains three cigar-shaped subpopulations (Fig. 2 *a*), in the spirit of the three A substates of MbCO, seen near 1950 cm^{-1}. One-tenth of the distribution is peaked at H = 8.2 kJ/mol and S/R = −7.6 (representing A_0), one-tenth at H = 19.5 kJ/mol and S/R = −3.7 (A_3), and 80% at H = 10.1 kJ/mol and S/R = −6.4 (A_1). These values of H and S at the peaks of the A-substate barrier distributions were estimated from enthalpy-distribution fits to flash photolysis experiments monitored in the infrared (Johnson et al., manuscript submitted for publication). Within each subpopulation, it was assumed that S decreased with increasing H.

As seen in Fig. 2, *b* and *c*, two of the simulated subpopulations, A_1 and A_3, are resolved by the MEM. The kinetic similarity of the A_0 and A_1 substates does not require two well-resolved subpopulations to fit the simulated data acceptably. From the MEM perspective, evidence for the kinetic separation of A_0 and A_1 has been lost in the process of acquiring a limited sampling of noisy data. The recovered probability densities do compare very well to the true density after integrating over S (Fig. 2 *d*). Because the MEM introduces structure conservatively, no spurious peaks were introduced into the g(H,S) distributions. The price paid for this desirable result is a modest reduction in the intensity at the maximum (Fig. 2 *d*) and a concomitant enhancement of the main peak's shoulder at large H. The agreement among the distributions in Fig. 2 *d* illustrates that for populations peaked at $H \approx 10$ kJ/mol or more, $\Delta A(t,T)$ is dominated by the effects of activation enthalpy; values for H are easier to recover than values for S, limiting any attempt to determine S-H correlations for such systems.

Still, much of the entropic character of the true g(H,S) is recovered from both the typical and the extended data sets. The correlation of S with H is plotted in terms of the

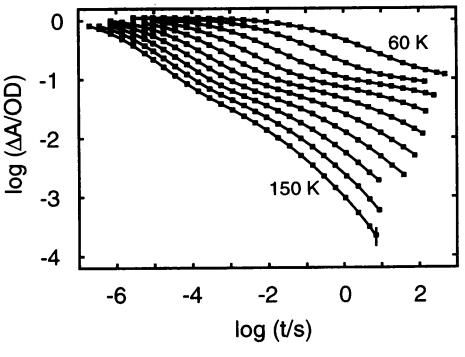

FIGURE 1 Ligand rebinding kinetics, simulated from 60 to 150 K in 10-K steps by adding gaussian noise to the kinetics generated using Eqs. 1 and 2 and the g(H,S) distribution shown in Fig. 2 *a*. Times, temperatures, uncertainties, and ΔA_{max} are as in Fig. 3. Lines depict the fit corresponding to the g(H,S) shown in Fig. 2 *c*.

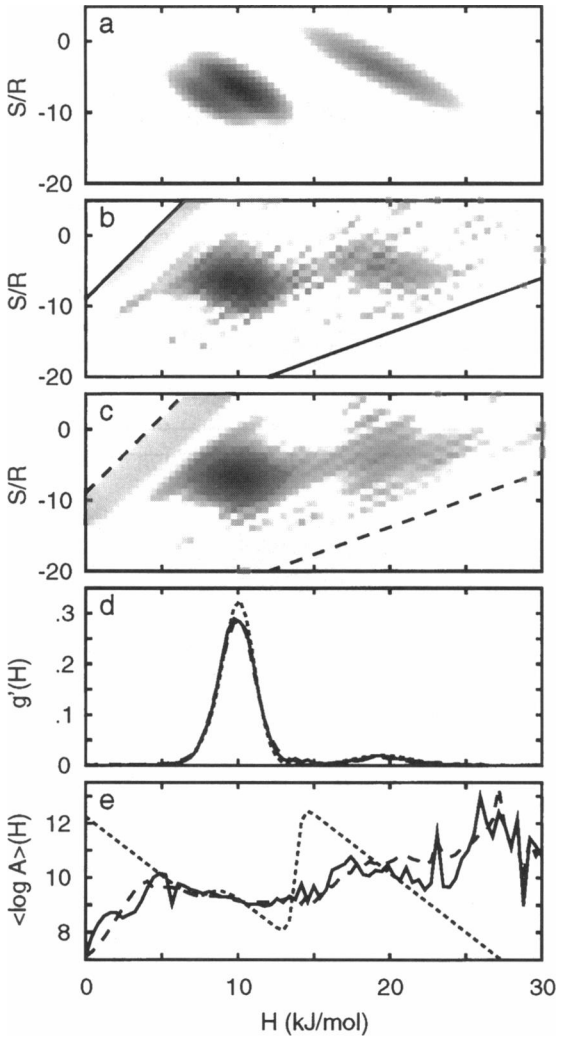

FIGURE 2 (*a*) "True" g(*H,S*) distribution used to synthesize two data sets: the data in Fig. 1 and an extended data set (sampled as in Fig. 1, with five points inserted before the shortest time and five appended after the longest time at each temperature). (*b, c*) *g(H,S)* distributions recovered by the MEM from the extended data set, stopping at $\chi^2 = 0.69$ (*b*) and from the data in Fig. 1, stopping at $\chi^2 = 0.57$ (*c*). Pixels included in the fit are bounded by the lines drawn in the upper left and lower right corners. A single logarithmic intensity scale was used for *a–c*, ignoring intensities below 10^{-4}. (*d*) *g(H,S)* integrated over *S* for the distributions in (*a*) (*dot*), (*b*) (*solid*), and (*c*) (*dash*). (*e*) log *A* averaged over *S* with lines drawn as in *d*.

pre-exponential *A* in Fig. 2 *e*. For both A_1 and A_3, the value of *S* at the peak density is correctly obtained from each data set. The correspondence of small *S* to large *H* within the dominant population is recovered qualitatively. Not too surprisingly, there is insufficient information at long times in the data set simulated at experimentally sampled times to determine the *S-H* correlation within the simulated A_3 substate. Because of its five additional points per temperature at long times, the extended data set yields this correlation more faithfully upon inversion.

Note also that the additional data at short times in the extended set lead to improved resolution of small activation barriers. In the upper left region of the *H-S* plane, g(*H,S*) is of lower intensity in Fig. 2 *b* than in 2 *c*.

RESULTS FOR EXPERIMENTAL $\Delta A(t,T)$

Convergence to $\chi^2 \leq 1$ for experimental data is impeded by systematic noise such as amplifier ringing at short times, variations in ΔA_{max} due to intensity fluctuations of the photolysis laser, and photolysis by the monitoring light source. At least the first two of these conditions hold for the experiment treated here. Furthermore, although Eq. 2 is a well-established physical approximation, it is not a mathematical identity; its limited validity may inhibit convergence even for data without systematic errors.

Three different g(*H,S*) distributions obtained from the rebinding of CO to Mb (Fig. 1 *b* of Steinbach et al., 1991) are shown in Fig. 4. Rebinding measured at ten temperatures, from 60 to 150 K, was inverted using $\xi = 0$ (Fig. 4 *a*) and $\xi = 1$ (Figs. 3 and 4 *b*). These distributions differ essentially by a shift to higher *H* and higher *S* for $\xi = 0$. Although the distributions are of comparable smoothness, χ^2 is 89.6 for $\xi = 1$ and 143 for $\xi = 0$, suggesting $\xi = 1$ is the more appropriate value.

The inversion represented in Fig. 4 *b* yielding $\chi^2 = 89.6$ is a much better representation of the data (Fig. 3) than is a more traditional four-parameter fit using a gamma distribution of enthalpies and a single activation entropy. This earlier fit yielding $\chi^2 = 922$ suggests a gamma distribution peaked at $H = 9.7 \pm 0.4$ kJ/mol and a pre-exponential with logarithm $\log[A(s^{-1})] = 8.8 \pm 0.2$ (table II of Steinbach et al., 1991). Although the MEM inversion does involve the determination of 1909 pixel intensities, the intensities are not varied independently. They are constrained by the entropy function (Eq. 7) to provide the distribution that is least altered from the slowly varying default distribution at a given value of χ^2. By inferring the distribution directly from the data, two subpopulations are clearly resolved, one peaked at $H = 10.9$ kJ/mol and a smaller one near 16.6 kJ/mol (Fig. 4 *d*). The latter population possesses a larger pre-exponential than the predominant population; $\langle \log A \rangle$ increases as *H* increases from 15 to 20 kJ/mol (Fig. 4 *e*). These results are consistent with the rebinding to A_1 and A_3 measured by flash photolysis monitored in the infrared.

However, these two peaks are resolved only by the introduction of two rather abrupt linear features in the g(*H,S*), one extending from *(H, S/R)* $\approx$ (6.3, −20.0) to *(H, S/R)* $\approx$ (13.8, −5.0) and the other from *(H, S/R)* $\approx$ (6.3, −20.0) to *(H, S/R)* $\approx$ (23.7, 10.0) (Fig. 4 *b*). The slope of such a line in the *H-S* plane is inversely proportional to the temperature corresponding to the data giving rise to the feature (see Discussion below). Thus, these features can be attributed to ΔA data at 60 and 70 K, respectively. The considerable intensity of these features suggests that the data are somehow incompatible and that a simultaneous description of the data at 60 K and the data at higher temperatures cannot be readily obtained with Eqs. 1 and

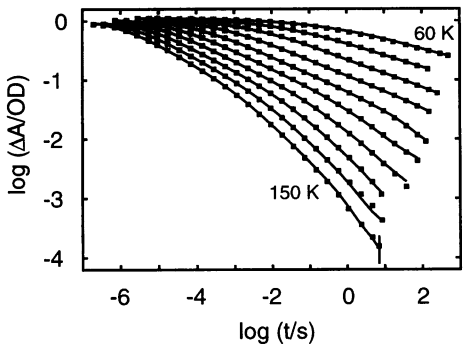

FIGURE 3 Kinetics of CO rebinding to Mb, measured from 60 to 150 in 10-K steps in a 75% glycerol/water solvent (figure 1 *b* of Steinbach et al., 1991). Lines depict the fit corresponding to the g(*H,S*) shown in Fig. 4 *b*.

2 and a smooth g(*H,S*) distribution. Indeed, when rebinding at 60 K is excluded from the inversion, these lines essentially disappear (Fig. 4 *c*), a more uniform distribution skewed to high H is obtained (Fig. 4 *d*), and a lower value of χ^2 is reached. Also, $\langle \log A \rangle$ no longer increases in the range $H = 15$ to 20 kJ/mol (Fig. 4 *e*).

When plotted on an expanded scale, $\Delta A(t, 60\text{ K})$ appears low relative to the data at 70 and 80 K, as might have been caused by a drop in photolysis-laser intensity when the data at 60 K were measured. Therefore, two additional inversions were performed, before which $\Delta A(t, 60\text{ K})$ was multiplied by 1.006 and 1.012. Although χ^2 dropped to 67.2 and 51.07, respectively, the g(*H,S*) distribution was not changed significantly from that shown in Fig. 4 *b*. Thus, the data at 60 K apparently call for two well-resolved peaks, whose positions agree well with infrared measurements of rebinding to A_1 and A_3. When the 60-K data are excluded, only a single peak with a shoulder at large H is recovered. The experimental data of Fig. 3 lack the obvious biphasic character that is seen easily at each temperature in the simulated data of Fig. 1. It appears better data are needed to determine unambiguously whether or not A_3 kinetics are resolvable in the Soret band.

Two other aspects of these results warrant comment. First, the upper left regions of the *H-S* plane in Fig. 4, *a–c*, are of rather uniform intensity, indicating the data at short times are insufficient to significantly alter $\vec{f}$ from $\vec{F}$. Thus, the gradual rise of the distributions at small H in Fig. 4 *d* should not be overinterpreted and the data's usefulness in assessing the symmetry of $g'(H)$ for the predominant subpopulation is limited. Better data at short times would better resolve small barriers and may result in a more sudden rise in g(*H,S*). Second, the increase in $\langle \log A \rangle$ for $H > 20$ kJ/mol (Fig. 4 *e*) should not be taken too seriously. Because the fraction of proteins possessing enthalpies in this range is very small (Fig. 4 *d*), these large pre-exponentials have little influence on the fit.

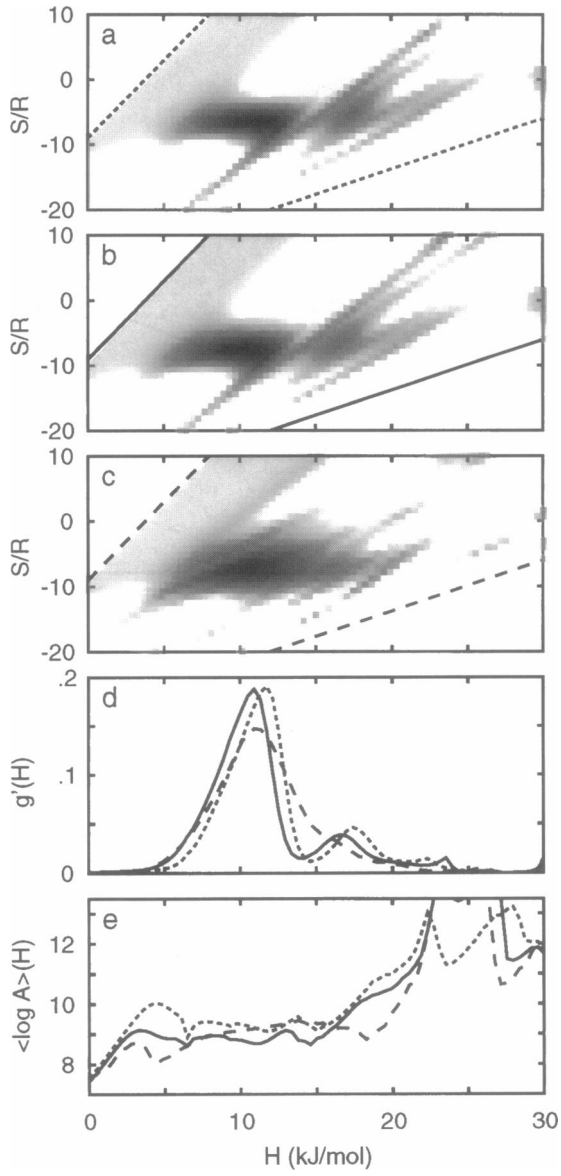

FIGURE 4 (*a–c*) g(*H,S*) distributions recovered by the MEM from the data of Fig. 3, by inverting all temperatures from 60 to 150 K with $\xi = 0$, stopping at $\chi^2 = 143$ (*a*), by inverting 60 to 150 K with $\xi = 1$, stopping at $\chi^2 = 89.6$ (*b*), and by inverting 70 to 150 K with $\xi = 1$, stopping at $\chi^2 = 18.9$ (*c*). Pixels included in the fit are bounded by the lines drawn in the upper left and lower right corners. A single logarithmic intensity scale was used, ignoring intensities below 10^{-4}. (*d*) g(*H,S*) integrated over S for the distributions in (*a*) (*dot*), (*b*) (*solid*), and (*c*) (*dash*). (*e*) log A averaged over S with lines drawn as in *d*.

DISCUSSION

Previous studies have applied the MEM in Laplace inversions of fluorescence (Livesey and Brochon, 1987; Bajzer and Prendergast, 1993; Kungl et al., 1994) and ligand-rebinding kinetics (Steinbach et al., 1992; Lavalette et al., 1991; Lambright et al., 1993), recovering one-dimensional distributions of lifetimes. By inverting several temperatures simultaneously, the current application embodies a physical model, not simply a mathematical relation. The g(*H,S*) distribution reflects the metastable configuration of the protein ensemble, in which each protein is trapped in a conformational substate with fixed activation enthalpy and en-

tropy. Therefore, the inversion of Eq. 1 is meaningful only at temperatures low enough to preclude substate interconversion on the time scale of rebinding.

The inversion of $\Delta A(t, T)$ to obtain $g(H,S)$ is underdetermined; there are typically a few hundred measured values of ΔA whereas 1000–2000 image pixels are needed for satisfactory resolution in H and S. The MEM addresses this shortage of data by spreading intensity in the H-S plane as much as is consistent with the measured $\Delta A(t,T)$. This conservative tendency results in intensity distributed along lines of positive slope in the H-S plane. The origin of these lines in $g(H,S)$ is easily explained. Consider data measured at two consecutive time points, t_i and t_{i+1}, at temperature T. Any measured decrease in ΔA from t_i to t_{i+1} results from proteins possessing values of H and S that correspond to a rebinding rate coefficient $k \approx 2/(t_i + t_{i+1})$. To fix k at temperature T, S must increase linearly with H (Eq. 2). Hence, all proteins rebinding between times t_i and t_{i+1} at temperature T lie close to a common, positively sloped line in the H-S plane. The slope of the line is determined by the temperature and the S intercept by the given value of k and the temperature (for $\xi \neq 0$). Thus, observed linear correlations of S and H do not necessarily imply chemical causation (Krug et al., 1976).

When visible in inversions of simulated $\Delta A(t,T)$, the linear features are relatively subtle (Fig. 2), but nonstatistical errors in experimental data can accentuate these lines (Fig. 4). The data of Fig. 1 were simulated with small S values correlated with large H values within the three subpopulations (Fig. 2 *a*) to assess the quality of the measurement required to override the MEM's noncommittal association of large S with large H. For data simulated over typical time and temperature ranges with realistic noise, the S-H correlation is recovered qualitatively for the main subpopulation, but data are required at longer times than are usually sampled to determine the correlation for the 10% component peaked at $H = 19.5$ kJ/mol (Fig. 2 *e*).

The MbCO data chosen as the first application for this analysis are exemplary in a number of respects. They range over nearly ten orders of magnitude in time and over four decades in ΔA (Fig. 3). Ten temperatures are sampled in the range that excludes the effects of tunneling measurable at low temperatures and the effects of protein relaxation near 180 K. Furthermore, a published enthalpy-distribution analysis of these data (Steinbach et al., 1991) facilitates evaluation of the current approach. Although they represent some of the best data available, they are not devoid of nonstatistical errors. Perhaps the most limiting is the variation in ΔA_{max} resulting from fluctuations in the intensity of the photolysis laser. In effect, laser fluctuations scale each ΔA trajectory up or down by a multiplicative constant relative to the curve that would be measured with a constant laser output. Averaging such curves together can wash out real undulations in the kinetics and result in relatively large error bars. In principle, trajectories could be individually rescaled when they are averaged together to correct for this effect. However, the current data set were not averaged in this way, and fluctuations in ΔA_{max} may consequently have been misinterpreted as evidence for temperature dependence in the rebinding, corrupting the $g(H,S)$ distribution and precluding $\chi^2 \approx 1$.

The MEM recovery of $g(H,S)$ is the most demanding analysis yet applied to this data set, and improved measurements and compensation for photolysis fluctuation can only lead to better results (e.g., lower values of χ^2, less pronounced linear features in $g(H,S)$). Increased resolution in time and temperature would surely help, providing a total number of data points comparable to the number of pixels used in $g(H,S)$. Increasing the temporal resolution would permit finer resolution in the H-S plane and requires nothing more than a re-averaging of the raw data, which were not available here. More importantly, an increase in temperature resolution amounts to a more complete scanning across the H-S plane by the data. The features in $g(H,S)$ should then be less sensitive to the exclusion of a single temperature from the inversion. Given the current results, it may be advantageous to measure $\Delta A(t,T)$ at 5-K intervals on a less concentrated sample. Reducing the sample concentration would lead to greater saturation of the sample by the photolysis laser, minimizing the systematic errors that accompany fluctuations in laser intensity.

CONCLUSIONS

MEM inversions of low-temperature reaction kinetics yield smooth distributions of activation enthalpy and entropy. For CO rebinding to Mb, the MEM reaches a value of χ^2 an order of magnitude smaller than traditional fits by a gamma distribution of enthalpies and a single value of entropy. By treating activation enthalpy and entropy on an equal footing and by permitting the data to influence the shape of the distribution, the MEM approach probes the data more closely than least-squares fitting with parameterized functions. Although the MEM $g(H,S)$ describes MbCO data much better than traditional analyses, χ^2 remains large enough to prevent unambiguous determination of the fraction of the signal due to rebinding to A_3 or the true shape of the $g(H,S)$ distribution for the individual A substates. Comparison of the MEM analyses of simulated and experimental data reported here indicates just how difficult these questions are to answer experimentally. Although $\chi^2 < 1$ is easily achieved by the MEM for simulations, it is more difficult for actual experiments. Data that are virtually free from fluctuations in photolysis and other systematic errors are needed for there to be a definitive characterization of the shape of $g(H,S)$ from this type of experiment.

The MEM recovery of $g(H,S)$ introduced here should prove especially valuable for biomolecular reactions characterized by lower activation enthalpies, for which a distribution in S would affect the kinetics more strongly. It would be instructive to apply the current analysis to low-temperature data of high quality for such systems (e.g., CO rebinding to protoheme or to isolated hemoglobin chains) to assess

the ranges of activation entropy and to identify multiple subpopulations that may exist.

I am thankful to Hans Frauenfelder and Kelvin Chu for providing experimental data and to G. Ulrich Nienhaus, V. Adrian Parsegian, and Srikanth Sastry for constructive comments on the manuscript.

REFERENCES

Agmon, N., and J. J. Hopfield. 1983. CO binding to heme proteins: a model for barrier height distributions and slow conformational changes. *J. Chem. Phys.* 79:2042–2053.

Alberding, N., R. H. Austin, S. S. Chan, L. Eisenstein, H. Frauenfelder, I. C. Gunsalus, and T. M. Nordlund. 1976. Dynamics of carbon monoxide binding to protoheme. *J. Chem. Phys.* 65:4701–4711.

Alberding, N., S. S. Chan, L. Eisenstein, H. Frauenfelder, D. Good, I. C. Gunsalus, T. M. Nordlund, M. F. Perutz, A. H. Reynolds, L. B. Sorensen. 1978. Binding of carbon monoxide to isolated hemoglobin chains. *Biochemistry.* 17:43–51.

Ansari, A., J. Berendzen, S. F. Bowne, H. Frauenfelder, I. E. T. Iben, T. B. Sauke, E. Shyamsunder, and R. D. Young. 1985. Protein states and proteinquakes. *Proc. Natl. Acad. Sci. USA.* 82:5000–5004.

Ansari, A., J. Berendzen, D. Braunstein, 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.

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.

Bajzer, Ž., and F. G. Prendergast. 1993. A model for multiexponential tryptophan fluorescence intensity decay in proteins. *Biophys. J.* 65: 2313–2323.

Berendzen, J., and D. Braunstein. 1990. Temperature-derivative spectroscopy: a tool for protein dynamics. *Proc. Natl. Acad. Sci. USA.* 87:1–5.

Cornwell, T. J., and K. F. Evans. 1985. A simple maximum entropy deconvolution algorithm. *Astron. Astrophys.* 143:77–83.

Dlott, D. D., H. Frauenfelder, P. Langer, H. Roder, and E. E. DiIorio. 1983. Nanosecond flash photolysis study of carbon monoxide binding to the β chain of hemoglobin Zürich [β63(E7)His → Arg]. *Proc. Natl. Acad. Sci. USA.* 80:6239–6243.

Dongarra, J. J., J. R. Bunch, C. B. Moler, and G. W. Stewart. 1979. LINPACK Users' Guide. SIAM, Philadelphia.

Ehrenstein, D., and G. U. Nienhaus. 1992. Conformational substates in azurin. *Proc. Natl. Acad. Sci. USA.* 89:9681–9685.

Frauenfelder, H., F. Parak, and R. D. Young. 1988. Conformational substates in proteins. *Annu. Rev. Biophys. Biophys. Chem.* 17:451–479.

Gull, S. F. 1989. Developments in maximum entropy data analysis. *In* Maximum Entropy and Bayesian Methods. J. Skilling, editor. Kluwer Academic, Norwell, MA. 53–71.

Gull, S. F., and J. Skilling. 1984. Maximum entropy method in image processing. *IEE Proc.* 131F:646–659.

Hagen, S. J., J. Hofrichter, and W. A. Eaton. 1995. Protein reaction kinetics in a room-temperature glass. *Science.* 269:959–962.

Krug, R. R., W. G. Hunter, and R. A. Grieger. 1976. Statistical interpretation of enthalpy-entropy compensation. *Nature.* 261:566–567.

Kungl, A. J., A. J. W. G. Visser, H. F. Kauffmann, and M. Breitenbch. 1994. Time-resolved fluorescence studies of dityrosine in the outer layer of intact yeast ascospores. *Biophys. J.* 67:309–317.

Lambright, D. G., S. Balasubramanian, and S. G. Boxer. 1993. Dynamics of protein relaxation in site-specific mutants of human myoglobin. *Biochemistry.* 32:10116–10124.

Lavalette, D., C. Tetreau, J.-C. Brochon, and A. Livesey. 1991. Conformational fluctuations and protein reactivity: determination of the rate constant spectrum and consequences in elementary biochemical processes. *Eur. J. Biochem.* 1956:591–598.

Livesey, A. K., and J. C. Brochon. 1987. Analyzing the distribution of decay constants in pulse-fluorimetry using the maximum entropy method. *Biophys. J.* 52:693–706.

Miers, J. B., J. C. Postlewaite, B. R. Cowen, G. R. Roemig, I-Y. S. Lee, and D. D. Dlott. 1991. Preexponential-limited solid state chemistry: ultrafast rebinding of a heme-ligand complex in a glass or protein matrix. *J. Chem. Phys.* 94:1825–1836.

Panchenko, A. R., J. Wang, G. U. Nienhaus, and P. G. Wolynes. 1995. Analysis of ligand binding to heme proteins using a fluctuating path description. *J. Phys. Chem.* 99:9278–9282.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. Numerical Recipes in Fortran, 2nd ed. Cambridge University Press, Cambridge.

Skilling, J. 1989. Classic maximum entropy. *In* Maximum Entropy and Bayesian Methods. J. Skilling, editor. Kluwer Academic, Norwell, MA. 45–52.

Skilling, J., and R. K. Bryan. 1984. Maximum entropy image reconstruction: general algorithm. *Mon. Notices R. Astron. Soc.* 211: 111–124.

Šrajer, V., L. Reinisch, and P. M. Champion. 1988. Protein fluctuations, distributed coupling, and the binding of ligands to heme proteins. *J. Am. Chem. Soc.* 110:6656–6670.

Stein, D. L. 1985. A model of protein conformational substates. *Proc. Natl. Acad. Sci. USA.* 82:3670–3672.

Steinbach, P. J., A. Ansari, J. Berendzen, D. Braunstein, K. Chu, B. R. Cowen, D. Ehrenstein, H. Frauenfelder, J. B. Johnson, D. C. Lamb, S. Luck, J. R. Mourant, G. U. Nienhaus, P. Ormos, R. Philipp, A. Xie, and R. D. Young. 1991. Ligand binding to heme proteins: connection between dynamics and function. *Biochemistry.* 30:3988–4001.

Steinbach, P. J., K. Chu, H. Frauenfelder, J. B. Johnson, D. C. Lamb, G. U. Nienhaus, T. B. Sauke, and R. D. Young. 1992. Determination of rate distributions from kinetic experiments. *Biophys. J.* 61:235–245.

Young, R. D., and S. F. Bowne. 1984. Conformational substates and barrier height distributions in ligand binding to heme proteins. *J. Chem. Phys.* 81:3730–3737.