Testing BR Photocycle Kinetics

John F. Nagle,* Laszlo Zimanyi,‡ and Janos K. Lanyi§

*Departments of Physics and Biological Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213 USA; ‡Institute of Biophysics, Biological Research Center of the Hungarian Academy of Sciences, Szeged, H-6701, Hungary; and §Department of Physiology and Biophysics, University of California, Irvine, California 92717 USA

ABSTRACT   An improved K absorption spectrum in the visible is obtained from previous photocycle data for the D96N mutant of bacteriorhodopsin, and the previously obtained M absorption spectrum in the visible and the fraction cycling are confirmed at 25°C. Data at lower temperatures are consistent with negligible temperature dependence in the spectra from 5°C to 25°C. Detailed analysis strongly indicates that there are two intermediates in addition to the first intermediate K and the last intermediate M. Assuming two of the intermediates have the same spectrum and using the L spectrum obtained previously, the best kinetic model with four intermediates that fits the time course of the intermediates is rather unusual, with two L's on a cul-de-sac. However, a previously proposed, more conventional model with five intermediates, including two L's with the same spectra and two M's with the same spectra, also fits the time course of the intermediates nearly as well. A new criterion that tests an individual proposed spectrum against data is also proposed.

1. INTRODUCTION

Quantifying the light activated photocycle of bacteriorhodopsin (bR) is a key ingredient in understanding the mechanism of proton pumping in the purple membrane of *Halobacterium salinarium* (formerly *H. halobium*). Accordingly, many studies of the photocycle using visible light absorption spectroscopy have been reported. However, the problem of determining the detailed kinetic photocycle has not been easy. This is due partly to the rather large number of chemical intermediates that seem to be required.

There is a fundamental problem in determining the kinetic photocycle from absorption spectroscopy that can be described in very simple terms. The time-resolved changes in absorption ΔA are due to a product of two factors,

$$\Delta A = \Delta\epsilon \otimes p. \tag{1}$$

The first factor $\Delta\epsilon$ denotes the difference spectra, relative to bR, of the chemical intermediates in the photocycle. The second factor p contains the time course of these intermediates. The multiplication $\otimes$ represents an inner product over all intermediates as will be written precisely in Eq. 5. The point of writing Eq. 1 is simply to emphasize that it is not possible to determine the factors by just knowing the product because an infinite number of pairs of factors will yield the same product.

Zimanyi and Lanyi (1993) have tried to resolve this fundamental problem by eliminating spectra that were judged as wrong or unlikely and using the average of what remained. Given spectra, the values of p in Eq. 1 can be obtained and the photocycle can be uniquely determined (Nagle, 1991a,b); this includes the rate constants for all possible

pathways and does not require assuming a particular kind of photocycle such as one not containing branches. Guessing the spectra, however, is nontrivial and involves assumptions. The first involves determining the number of intermediates. Second, each spectrum is an unknown function whose theoretical characterization by a small number of parameters is uncertain. However, there are a number of constraints, the best of which is that absorption spectra must be positive; this constitutes the first "filter" in a procedure (Zimanyi and Lanyi, 1993) that also involves two additional filters that are based on successively less well founded physical criteria. This procedure yields spectra that lie within fairly narrow ranges and that look quite reasonable. It is, however, important to test these results independently in so far as possible, and that is the first purpose of this paper.

Nagle (1991a) and Lozier et al. (1992) have developed a different method to resolve the fundamental problem posed by Eq. 1. This method is based on the assumption that the spectra ϵ of the chemical intermediates are the same as the temperature T is varied over small ranges ($\sim 30°C$) near ambient. In contrast to the spectra, the rate constants of the decays in the photocycle clearly vary substantially in these temperature ranges, so both ΔA and p vary with T. Therefore, each temperature would yield an independent Eq. 1 with the same, although unknown, $\Delta\epsilon$. Analysis (Nagle, 1991a) shows that when data ΔA at three or more temperatures are included, both $\Delta\epsilon$ and the values of p for the different temperatures can be obtained. A computer program has been written that performs this task successfully on simulated data that obey the assumption (Nagle, 1991a). However, this program has yet to be successful on real data. This raises the question whether the assumption of temperature independence of the spectra is valid. The second purpose of this paper is to test this assumption.

The first focus of this paper is to show that standard nonlinear least squares methods allow the relative, i.e., difference, spectrum of the first intermediate K to be determined at each temperature. This method will be applied to the D96N

mutant data obtained previously (Zimanyi and Lanyi, 1993) when it was emphasized that the later part of the photocycle of this particular mutant is greatly simplified, thereby making this a most promising photocycle for study. The first result is that these spectra are consistent with being temperature independent provided that one chooses appropriate values for the fractions FC of bR cycling at each temperature. To convert the difference spectrum to an absolute spectrum requires knowing the actual fraction FC of bR cycling. This is often a problematic quantity, but for the D96N mutant data, it and the M spectrum can be easily obtained at 25°C because of the long decay time of M. The ensuing M spectrum is essentially identical with the one obtained previously. The K spectrum is also rather similar, although there are real differences.

The second focus of this paper is to explore various models of the photocycle by using the K and M spectra obtained here, as well as the less certain L spectrum obtained earlier (Zimanyi and Lanyi, 1993). This determines the time course $p_i(t)$ of the intermediates i, but they noted discrepancies that seemed to indicate that the L spectrum cannot be quite right. Nevertheless, it is interesting to compare the consequences of assuming the L spectrum in order to compare with recent proposals for the photocycle.

2. DATA

The original data for the D96N mutant of bR consisted of 386 wavelengths spaced at 1 nm, but because so many wavelengths are unwieldy for analysis, the data set was previously reduced to 100 wavelengths (Zimanyi and Lanyi, 1993). In the present study the original data were reduced to data at 37 wavelengths spaced every 10 nm from 370 to 730 nm. The change in absorption ΔA for each of the reduced data points was obtained using a cubic polynomial fit to the original data at 19 contiguous wavelengths centered at the wavelength of the reduced data point. The original data are compared with the reduced data in Fig. 1 at two times, including the earliest time at which data were taken (70 ns) as well as a later time. As usual in visible spectroscopy (Xie et al., 1987) the data at the earliest times are rather noisier than data taken at later times and the data become noiser toward the blue end of the spectrum. The reduced data are somewhat smoother than the original data because of the cubic polynomial fit. Since the reduced data appear to be faithful to the original data, they will henceforth be called the data. Fig. 2 shows the time course of the data at several selected wavelengths for $T = 25°C$.

3. M SPECTRUM AND FRACTION CYCLING

Unlike for most photocycle data, the time course for each wavelength in Fig. 2 shows a remarkable plateau from 1 to 10 ms when $T = 25°C$. This suggests that the photocycling pigment is stuck in a single spectroscopic intermediate state during this time; this blue-shifted state is usually called the M intermediate. If FC is the fraction of bR cycling, then 1/FC times the spectrum at $t = 3$ ms added to the bR spectrum gives the M spectrum. Fig. 3 shows the relative absorption spectrum at $t = 3$ ms divided by FC = 0.152 together with the absorption spectrum of bR. Adding these two spectra together produces the spectrum labelled M in Fig. 3. If a larger value of FC were chosen, then there would be negative absorbance in M above 520 nm, which is unphysical. If a smaller value of FC were chosen, then there would be a second peak in M above 520 nm. These two criteria, de-
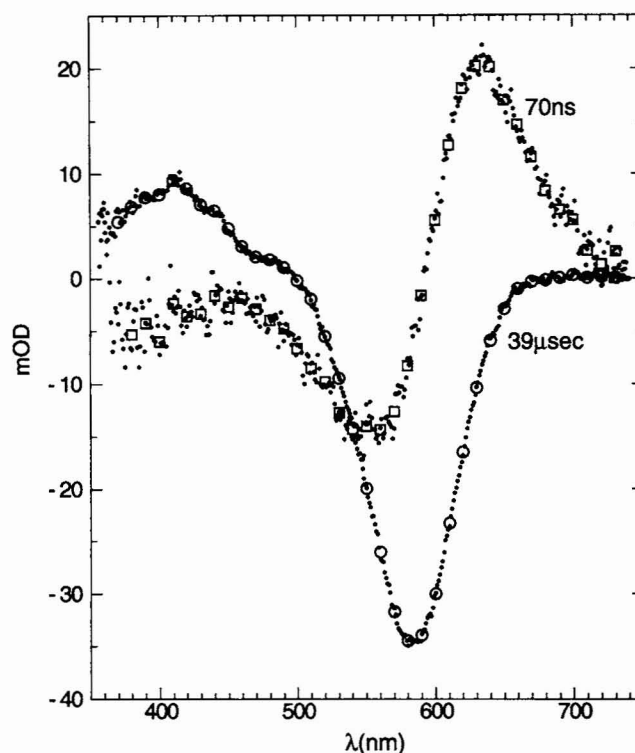

FIGURE 1 Absorbance in milli optical density (mOD) versus wavelength (nm) for two delay times, 70 ns and 39 μs. The raw data are plotted as small solid circles. The reduced and smoothed data are plotted as large open symbols. The maximum absorbance of D96N bR is near 377 mOD near 570 nm.

scribed as Propositions III and VIII by Nagle et al. (1982), are the first filter of Zimanyi and Lanyi (1993). The fraction cycling FC = 0.152 ± 0.004 is in good agreement with the FC ± 0.156 obtained previously (Zimanyi and Lanyi, 1993). What is so encouraging about the M spectrum in Fig. 3 compared with those in earlier studies (Nagle et al., 1982, Lozier et al., 1975) is that it simultaneously has negligible absorbance above 520 nm and no evidence of a secondary peak. The shape of this spectrum strongly supports the suggestion at the beginning of this paragraph that this is a single spectroscopic intermediate.

Also in Fig. 3 are plotted the relative spectra at four later times scaled by a factor that is larger for the later times. The fact that all the spectra at times later than $t = 3$ ms can be simply scaled to superimpose with the spectrum at $t = 3$ ms suggests that the final decay is a simple one straight to bR with no apparent additional N and O intermediates, as was concluded previously (Varo and Lanyi, 1991a and 1991b; Zimanyi and Lanyi, 1993).

At lower temperatures 5 and 15°C there are not such pronounced plateaux in the absorption curves. Relative to 25°C, this is consistent with the decay of the precursor of M, usually identified as L, being slowed more than the decay of M. This means that there will be no time at which all the photocycling bR is in M; rather, some will still be in L when some of the M has already decayed. Therefore, it is not possible to obtain as accurate values of the fraction cycling for these lower temperatures as for $T = 25°C$.
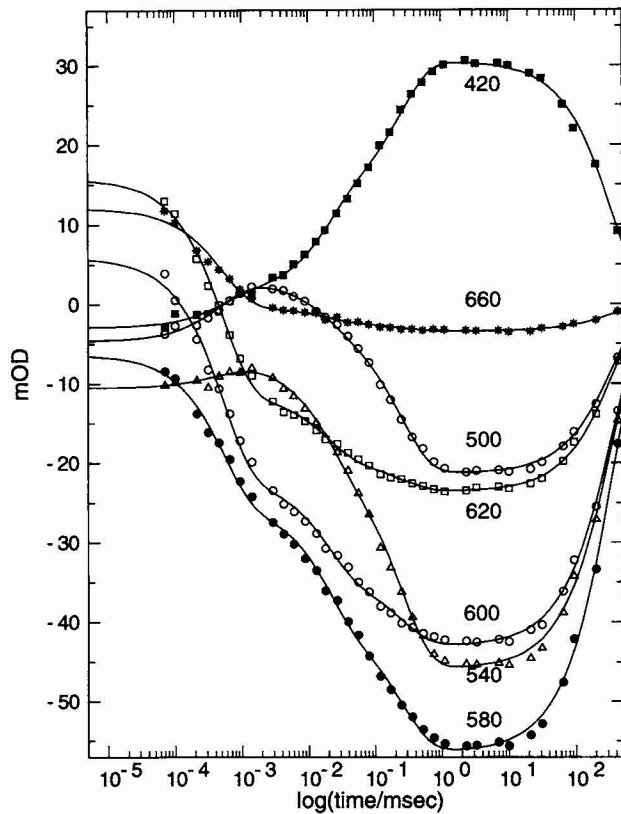
FIGURE 2   The time course of the absorbance (mOD) versus log(time) at selected wavelengths (in nanometers) indicated by number near each curve. The symbols are the reduced data as shown in Fig. 1, and the lines are the best fit using four exponentials as described in the text.

The difference spectra at the lower temperatures can be made to superimpose with each other at later times and with the difference spectrum at $T = 25°C$ as shown in Fig. 4. This implies that the decay of M at later times is also straight back to bR. Most importantly, it implies that the M spectrum is independent of temperature.

4. K SPECTRUM

The idea for obtaining the spectrum of the K intermediate is rather simple. If the data at the earliest times (the nanosecond range) report the decay of a single intermediate, which is usually called K, then extrapolation of the measured absorption back to zero time, $\Delta A(\lambda, t = 0)$, will give the K spectrum minus the bR spectrum times the fraction cycling FC. Therefore, one can obtain $\epsilon_K(\lambda)$ using

$$\epsilon_K(\lambda) = \epsilon_{bR}(\lambda) + \Delta A(\lambda, t = 0)/FC. \qquad (2)$$

If one further supposes that each bR passes through both K and M in the photocycle, then the FC in Eq. 2 is the same as derived in the previous section when $T = 25°C$.

The technical problem of extrapolating $\Delta A(\lambda, t)$ to $t = 0$ is addressed using a nonlinear least squares fitting technique (VARP) that simultaneously fits the data at multiple wavelengths and all times to a sum of exponentials (Golub and
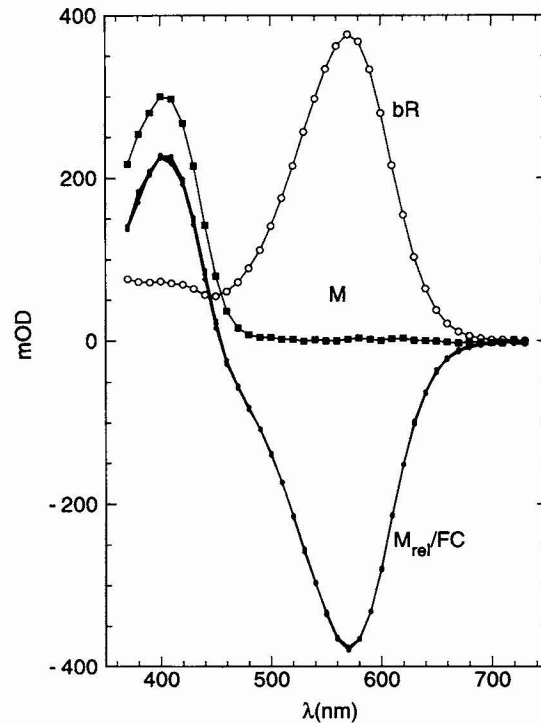

FIGURE 3   Determination of the M spectrum. The relative spectrum at 3.3 ms divided by FC = 0.152 is plotted as M_{rel}/FC (*small filled circles*). Addition of M_{rel}/FC to the spectrum of bR (*open circles*) yields the M spectrum (*filled squares*). Also plotted are the scaled relative spectra at times 10, 31, 94, and 200 ms (*small filled circles*) that accurately overlay the data at 3.3 ms.
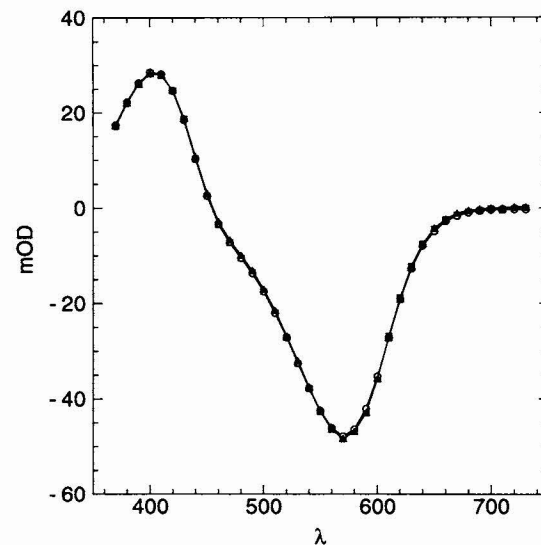

FIGURE 4   Comparison of relative M spectra for $T = 5°C$ (*solid circles*), 15°C (*open circles*), and 25°C (*solid triangles*). At each temperature the absorption data were added for the last seven times (see Fig. 2) and then scaled to overlay.

Leveque, 1979),

$$\Delta A(t, \lambda) = \sum_{j=1,n} \exp(-tk_j^*)b_j(\lambda), \qquad (3)$$

where the k_j^* are only apparent rate constants and the $b_j(\lambda)$ are the amplitudes; both are obtained by the fit. The

extrapolation of Eq. 3 to zero time simply yields

$$\Delta A(0, \lambda) = \sum_{j=1,n} b_j(\lambda). \qquad (4)$$

Although one could perform the extrapolation to $t = 0$ in simpler ways, the use of the VARP technique gives some intermediate results of considerable interest. As has been previously emphasized (Xie et al., 1987), fitting to sums of exponentials provides an estimate of the number of intermediates; this number must be the same as the number of terms n in the sum in Eq. 3. For these data the σ of the fits using $n = 3$ exceeds 1.1 milli optical density (mOD) at all three temperatures. When n is increased to 4, the fit greatly improves to $\sigma = 0.46$, 0.63, and 0.49 mOD for $T = 5$, 15, and 25°C, respectively. Comparing these results using the F_x test (Bevington, 1969) suggests that the fourth exponential is required at greater than the 99.9% level. When n is increased further, the fitting program does not converge or results in minimal improvement in σ. Assuming that the kinetics are first order and that there are no experimental artefacts, these results strongly indicate that there are four intermediates.

Fig. 2 shows the fit to the data at $T = 25°C$ and indicates the extrapolation to short times. Fig. 5 shows the relative K spectrum obtained from $\Delta A(\lambda, 0)/FC$ for 25°C, the bR spec-

trum and the K spectrum derived using Eq. 2. For $T = 5$ and 15°C one does not know the fraction cycling FC, but it is possible to scale the relative K spectra at these temperatures so that they nearly coincide with the relative K spectrum at $T = 25°C$. Fig. 5 shows the scaled relative K spectra and also the corresponding absolute K spectrum for $T = 25$ and 15°C.

Systematic variations in the relative K spectrum with temperature appear to be absent as shown in Fig. 5, suggesting that the K spectrum is largely independent of temperature. This encouraging result has not always been obtained. Earlier data (see Fig. 4 in Nagle et al., 1982) showed a rather stronger temperature dependence of the K spectrum, although it was suggested that this was due to experimental artefact. Also, the data obtained by Xie et al. (1987), when analyzed in this same way, yield a temperature dependent K spectrum. The present data would seem to be simpler and more plausible in this regard.

Fig. 5 also shows the range for the K spectrum determined previously (Zimanyi and Lanyi, 1993). While the new result in Fig. 5 is generally in agreement, there are real differences. The previous K spectrum has a somewhat higher peak and is red shifted by 5–10 nm from the one obtained here. One way to make the K spectrum derived here closer to the previous K spectrum would be to take a smaller fraction cycling FC than was derived in the preceding section. The analysis in the preceding section assumed that all the cycling bR is in the M state at 3 ms. If there are branches around M or if some M has decayed to the next state while some of the precursor has not yet decayed to M, then FC could be larger, but not smaller, and this would make the disagreement worse. Another way to make the K spectrum derived here closer to the previous K spectrum would be to change the extrapolations, thereby giving different relative K spectra than shown in Fig. 5. However, to obtain agreement at 600 nm would require the relative K spectrum to increase from ~6 to 15 mOD or larger; this would require the extrapolation curve for 600 nm in Fig. 2 to approach 15 mOD. To obtain agreement at 620 nm the extrapolation would have to go to 25 mOD in Fig. 2. Such altered extrapolations are not reasonable because the extrapolated curves must approach these values horizontally at 10 ns in order to be consistent with the fastest decay time required to fit the data at all wavelengths.

FIGURE 5 Determination of the K spectrum. The relative K spectrum divided by FC = 0.152 at 25°C (*open squares*), labeled K_{rel}/FC, is added to the bR spectrum (*solid squares*) to produce the K spectrum. For T = 5°C (*open circles*) and $T = 15°C$ (*open triangles*) the relative K spectra are scaled to the spectrum at 25°C and then added to the bR spectrum to give K spectra at 15°C (*open triangles*) and 5°C (*not shown*). The symbols (+) indicate the spread in the K spectrum at 15°C obtained previously by the three filter method of Zimanyi and Lanyi (1993). All spectra have been normalized to the maximum absorbance of bR.

5. SVD ANALYSIS

In the preceding section it was reported that fitting the data to sums of exponentials using VARP strongly indicates that there are $n = 4$ kinetic intermediates. An alternative way to obtain the number of intermediates is to use singular value decomposition (SVD) (Schrager, 1984; Hessling et al., 1993). The principal strength of the SVD method is that it does not make any assumption about the photocycle; it does not even require the assumption of first order kinetics. But because of this feature, the SVD method can not, in principle, distinguish between two intermediates with the same spectra (or, more generally, any intermediate whose spectrum is a linear combination of other spectra). Specifically, when decomposing the data into the wavelength and time matrices,
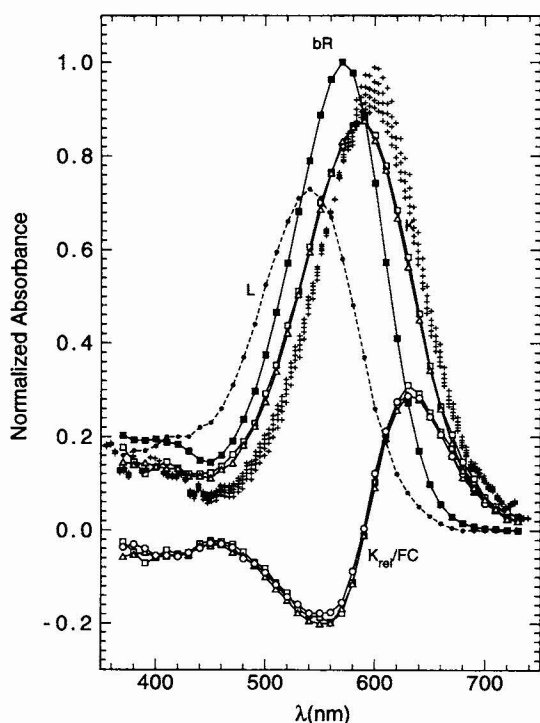
both the kinetic intermediates with the same spectrum can be represented by the same component in the wavelength dimension and the component in the time dimension can be the sum of the kinetics of the two intermediates. Of course, SVD further mixes these intermediates with the others to form abstract components, but the principle that the two intermediates with the same spectrum can be represented only as their sum continues to hold. Furthermore, with noise the method will not distinguish between two similar, but not identical, spectra either. Therefore, the SVD method generally tends to give too few intermediates.

The result of applying the SVD method to these data was that three abstract components are definitely indicated. The singular value of the fourth abstract component was very close to, but a little larger than, the noise level indicated by the $n = 4$ VARP fit; this indicates a fourth component. Also, for $T = 5$ and $15°C$, both the wavelength and time correlations exceeded 0.75, indicating that the fourth component was not random; for $T = 25°C$, the correlations were close to 0.5, suggesting that the fourth component was marginal. Therefore, the number of components obtained by SVD basically agrees with the $n = 4$ obtained from VARP. However, it should be emphasized that the three component fit to the data using SVD is very nearly as good as the four component fit. This feature, which contrasts with VARP, indicates that SVD is less discriminatory and therefore a poorer tool for diagnosing the number of intermediates when one believes that the kinetics are first order.

The results from SVD also have another implication that involves the observation made in the next-to-last paragraph. If two of the four intermediates have very similar spectra, then SVD should indicate only three abstract components. The fact that SVD indicates four components strongly suggests that no two of the intermediates have identical spectra. However, the fact that the fourth singular value is rather small is consistent with two of the spectra being similar, but not identical.

6. ESTIMATION OF TIME COURSE OF INTERMEDIATES

Given reliable spectra for the K and M intermediates, it might seem possible to deduce the L spectrum, but the methods of earlier analysis (Nagle, 1991a) prove that this is not possible using simple a priori methods at one temperature. Furthermore, the analysis in the preceding section strongly suggests that there are two additional intermediates, not a single L intermediate. Therefore, the derivation of an L spectrum will not be pursued in this paper. It will, nevertheless, be interesting to examine the consequences of the L spectrum proposed earlier by Zimanyi and Lanyi (1993). This course has been recently followed by Gergely et al. (1993).

It is straightforward to obtain the time course $p_i(t)$ of the intermediates i from the data and the spectra of the intermediates. The basic equation relating measured absorption changes to the spectra and the time course of the intermediates is

$$\Delta A(t, \lambda) = \sum_i \Delta\epsilon_i(\lambda)p_i(t) \equiv \Delta\epsilon(\lambda) \otimes \mathbf{p}(t), \qquad (5)$$

where $\Delta\epsilon(\lambda) = \{\Delta\epsilon_i(\lambda)\}$ are the relative spectra of the intermediates, i.e., $\Delta\epsilon_i(\lambda)$ is the spectrum of intermediate i minus the spectrum of bR. For each time t, Eq. 5 is an overdetermined set of 37 linear equations (one for each wavelength) for the $p_i(t)$, $i = 1$–3. The best solution is routinely obtained by standard linear least squares methods.

The $p_i(t)$ are shown in Fig. 6 for $T = 25°C$ for two cases. Both cases use the L spectrum of Zimanyi and Lanyi (1993) at $T = 25°C$ and the M spectrum derived in this paper, which is essentially identical to the previously derived M spectrum. The first case uses the K spectrum derived in this paper and the second case uses the K spectrum derived previously (Zimanyi and Lanyi, 1993). As would be expected, the derived $p_i(t)$ shown in Fig. 6 are similar for both cases, with the largest differences occurring at earlier times in the photocycle.

Fig. 6 also plots the sum $\sum p$ of the $p_i(t)$ for $i = K$, L, and M states derived from Eq. 5. One expects that $\sum p$ should be close to 1 until late in the photocycle when considerable decay of M to bR has occurred. Although this expectation is largely met in Fig. 6, there is a consistent deviation that was noted earlier (Zimanyi and Lanyi, 1993), namely, $\sum p$ is 5% smaller than 1 in the time interval from 0.5–6 μs and then increases to greater than 1 until after 20 ms. This deviation from 1 is larger than the estimated errors in the data ($\sim 2\%$ in this time interval). Inasmuch as the linear least squares determination of the $p_i(t)$ is blind to what $\sum p$ should be, this strongly suggests that the spectra are incorrect. Since the preceding analysis of the K and M spectra seems solid, this suggests that it is the L spectrum that is incorrect and/or that there is another intermediate.
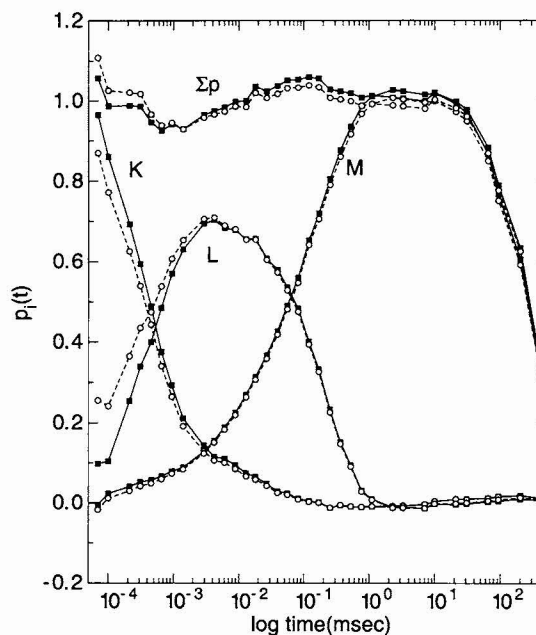

FIGURE 6 The calculated time course of the intermediates $p_i(t)$ for the K, L, and M states as indicated versus log(time/ms) for $T = 25°C$. The solid squares are for the case when the K spectrum determined in the previous section was used. The open circles are for the case when the K spectrum of Zimanyi and Lanyi (1993) was used. Also shown is the computed sum of the intermediates $\sum p$.

The problem in the preceding paragraph was fixed earlier (Zimanyi and Lanyi, 1993) by simply adding enough to $p_L(t)$ to make $\sum p$ equal to 1. Because this fix has no theoretical basis, a different fix is employed here. If it is assumed that the reason that $\sum p$ is not identically 1 is due to noise, say, in the flash intensity for different measuring times, then a valid fix is simply to normalize $\sum p$ to 1 (except for the late part of the photocycle). Also, $p_K(t)$ and $p_L(t)$ in Fig. 6 oscillate weakly around 0 for later times, which is clearly just noise, so it is appropriate to set these to 0. Fig. 7 shows the fixed result. These $p_i(t)$ are similar to those given earlier (Fig. 10 C in Zimanyi and Lanyi (1993) and Fig. 2 in Gergely et al. (1993); the latter study used the earlier spectra but only used three wavelengths to determine the $p_i(t)$).

7. INDEPENDENT CRITERION FOR GOODNESS OF SPECTRA

In the last section it was noted that $\sum p$ is not necessarily equal to 1 if the spectra are incorrect. This could be a useful criterion for improving the spectra, as suggested by Zimanyi and Lanyi (1993). It was attempted to use this simply by scaling the difference spectrum for L; this changed the time interval at which $\sum p$ was less than 1 but did not significantly improve its overall deviation from 1. As we also believe that there is an additional intermediate, further exploration in this direction was not pursued. It must be emphasized, however, that this criterion is not sufficient to determine the spectra uniquely (Nagle, 1991a), i.e., it is possible to have incorrect spectra that will obey this criterion even for the ideal case of no experimental noise.

In this section a similar criterion will be presented that has a distinct advantage. To apply the preceding $\sum p = 1$ criterion requires guessing all the spectra, and when the criterion is not met it is unclear which spectrum is incorrect. The criterion to be presented now allows each spectrum to be tested independently of the other spectra.

The basic idea is that the difference spectra $\Delta\epsilon$ can be written as linear combinations of the data ΔA with coefficients q:

$$\Delta\epsilon_i(\lambda) = \sum_j q_i(t_j)\Delta A(t_j, \lambda). \tag{6}$$

Of course, this requires that intermediate i be present at least at one of the times t_j; it also requires that the total number of times n_j be greater than or equal to the number of intermediates present at the chosen times t_j. If the number of chosen times n_j equals the number of intermediates n, then the coefficients $q_i(t_j)$ are uniquely determined unless the spectra are not linearly independent. If n_j exceeds n, then the $q_i(t_j)$ are not uniquely determined. The $q_i(t_j)$ matrix is essentially the inverse of the $p_i(t_j)$ matrix defined in Eq. 5; by multiplying Eq. 5 on the left by the $\mathbf{q}$ matrix one obtains

$$\sum_j q_i(t_j)\, p_k(t_j) = \delta_{ik}. \tag{7}$$

Next consider times t_j early enough that no decay back to bR has occurred, so that the sum $\sum_m p_m(t_j)$ is 1. Singling out

a particular intermediate k, this may be written as

$$p_k(t_j) = 1 - \sum_{m \neq k} p_m(t_j). \tag{8}$$

Substituting Eq. 8 into Eq. 7 when $k \neq i$ yields

$$\sum_j q_i(t_j) \left[1 - \sum_{m \neq k} p_m(t_j) \right] = 0. \tag{9}$$

For each term involving $p_m(t_j)$ the sum over j vanishes by Eq. 7 unless m equals i and it equals 1 when m equals i. Therefore, one has

$$\sum_j q_i(t_j) = 1 \tag{10}$$

for each intermediate i. Equation 10 is the criterion that must be satisfied by good spectra. The $q_i(t_j)$ are easily calculated using least squares analysis of Eq. 6. The key feature is that all the q_i for a given i are calculated from the data ΔA and from *only* the difference spectrum $\Delta\epsilon_i(\lambda_k)$ of intermediate i.

This test was applied to putative spectra in the following way. The data $\Delta A(t_j,\lambda_k)$ at the first five times were added, as were the data at subsequent blocks of five times. Only the first five blocks of times were used at $T = 25°C$ in order to guarantee that Eq. 8 was satisfied. This meant that the number of effective times was $n_j = 5$. Even though only $n = 3$ or 4 spectra are anticipated, making n_j larger than n guarantees that the spectral space is completely spanned. The nonuniqueness of the q_i does not affect the sum $\sum q_i$, which is required for the test in Eq. 10.

The result for the M spectrum was $\sum_j q_M(t_j) = 1.031$; this could be reduced to 1 by choosing a smaller fraction cycling of 0.148. Because the M spectrum is least controversial, this can be considered to be the benchmark for the other spectra. The results for the L and K spectra of Zimanyi and Lanyi (1993) are $\sum q_L = 0.941$ and $\sum q_K = 0.942$; both are significantly smaller than for the M spectrum. The result for the K spectrum proposed in this paper is $\sum q_K = 0.966$, which is somewhat closer to 1 than the previous result. This is an independent argument that the present K spectrum is better than the earlier one. However, there is still a 6% disagreement between $\sum q_M$ and $\sum q_K$, which suggests that there are other problems with these spectra or with the data or with the assumptions about the photocycle. Similar results were obtained when n_j was reduced below 5. Nevertheless, the test suggests that the L spectrum is the least reliable, as anticipated.

8. KINETIC MODELS FROM CONCENTRATIONS $P_i(T)$

From the time courses $p_i(t)$ of the intermediates, a relatively straightforward method exists to obtain the best kinetic model (Nagle, 1991a and 1991b), including the determination of which rate constants are nonzero as well as the determination of the values of the nonzero rate constants. The method begins by assuming nonzero rates for all possible transitions and the fitting program systematically drives rates successively to zero, thereby simplifying the model. This

method was applied for the $p_i(t)$ in Fig. 7. Detailed results will first be presented for the case with the K spectrum found in this paper, and differences for the earlier K spectrum will be described at the end of this section. When only three intermediates ($n3$) were allowed, the method led to the kinetic model.

$$K \leftrightarrow L$$
$$\downarrow \qquad (11)$$
$$M \rightarrow bR,$$

which may be more succintly and equivalently written as $K \leftrightarrow L + K \rightarrow M \rightarrow bR$. This model has four rate constants and places L on a dead-end side path. We will call this the $n3.4a$ model, where the final letter is different for other models with three intermediates and four rate constants. The fit to the $p_i(t)$ for this model is shown in Fig. 8 by the dotted-dashed lines. The σ_p of the fit for the $n3.4a$ model is 0.0354. The subscript p on σ emphasizes that this is the fit to the $p_i(t)$ and not to the data, so σ_p is not to be compared to the σ's in Section 4. It is also straightforward to calculate the fits to the data from the $p_i(t)$ and the relative spectra using Eq. 5. The sigma of this fit will be called σ_d. For the $n3.4a$ model $\sigma_d = 1.43$ mOD, which is somewhat poorer than the $\sigma = 1.1$ mOD obtained from fitting three exponentials in Section 4.

The $n3.4a$ model is somewhat unconventional in that L is not on the path from K to bR but lies on a cul-de-sac. It is therefore of interest to examine the model ($n3.5c$):

$$K \leftrightarrow L \leftrightarrow M \rightarrow bR. \qquad (12)$$

The fit for this model is shown in Fig. 8 by dashed lines. The σ_p for this $n3.5c$ model is 0.0559. Also, $\sigma_d = 2.04$ mOD for this model, so the $n3.5c$ is clearly a considerably poorer model than the $n3.4a$ model.
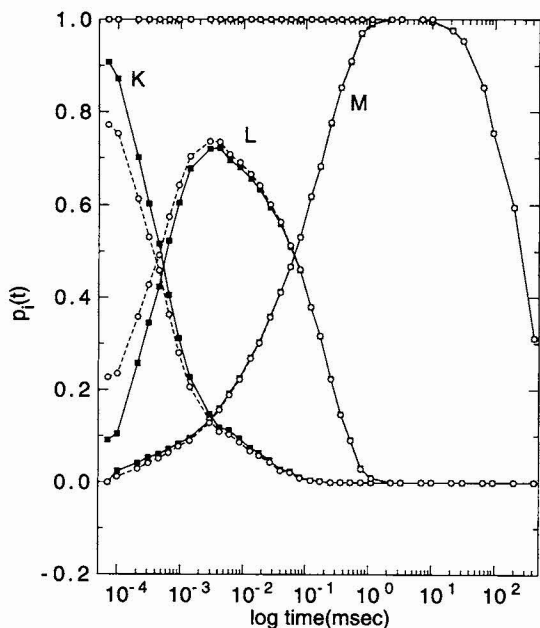

FIGURE 7 Same as Fig. 6 except that the $p_i(t)$ have been normalized so that their sum equals 1 for times less than 7 ms. Also, $p_K(t)$ has been set to 0 after 0.25 ms and $p_L(t)$ has been set to 0 after 2.3 ms.
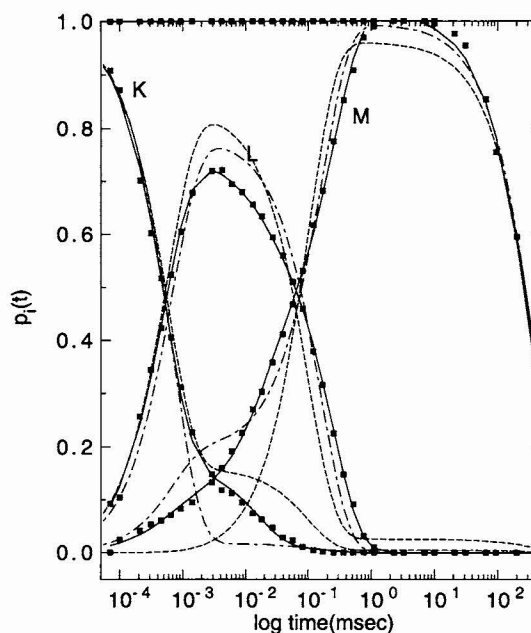

FIGURE 8 Fits to the time course of the intermediates at 25°C. The solid squares are the $p_i(t)$ from Fig. 7. The dotted-dashed lines show the fit from the $n3.4a$ model, and the dashed lines show the fit from the $n3.5c$ model described in the text. The solid lines show the simultaneous model independent nonlinear least squares fit using four exponentials.

In the ensuing analysis, one should remember that each kinetic model fits the p_i to special combinations of sums of n exponentials, so kinetic model fits can not have smaller σ_p than fits with general sums of exponentials using VARP. One way to determine the adequacy of the fit with the $n3.4a$ kinetic model is to compare the $\sigma_p = 0.035$ of the fit with the $\sigma_p = 0.030$ obtained by simultaneously fitting the $p_i(t)$ data with sums of three exponentials (with only three apparent rate constants), using VARP. The closeness of these two values of σ_p is consistent with $n3.4a$ being the best kinetic model with $n = 3$ intermediates.

Fig. 8 also shows the result of simultaneously fitting the $p_i(t)$ with sums of four exponentials with only four apparent rate constants, using VARP. The σ_p of this fit is 0.0095, considerably better than for either the $n3.4a$ model or the $n = 3$ VARP fit. This makes it clear that kinetic models with only $n = 3$ intermediates are not adequate, even though the $p_i(t)$ were biased toward three intermediates by choosing only three spectra. Increasing the number of exponentials to five only decreases σ_p to 0.0085, so most of our effort has been on models with $n = 4$ intermediates.

It is difficult to proceed with an analysis with $n = 4$ intermediates from three $p_i(t)$ unless it is assumed that the spectrum of the extra intermediate is the same as one of the three original spectra. If this is assumed, then the sum of two of the $p_i(t)$ determined from a kinetic model must add to one of the $p_i(t)$ shown in Fig. 8. This assumption and the corresponding procedure have been employed by Varo and Lanyi (1991a) and Zimanyi and Lanyi (1993) and by Gergely et al. (1993) for absorption data for the photocycle and by Nagle (1991b) for the analysis of Raman data. These

previous analyses have shown that models with two M states clearly fit the $p_i(t)$ better than models with only one M and one less intermediate.

With the assumption that there are $n = 4$ intermediates and two M states, two models have been found that fit the data nearly equally well, the $n4$–$2M.7b$ model

$$K \leftrightarrow L \leftrightarrow M1 \rightarrow M2 \rightarrow bR + K \leftrightarrow M1 \qquad (13)$$
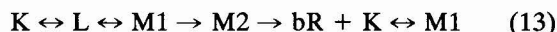
with $\sigma_p = 0.01781$ and $\sigma_d = 0.990$ mOD and the $n4$–$2M.7a$ model

$$K \leftrightarrow L \leftrightarrow M1 \leftrightarrow K + K \rightarrow M2 \rightarrow bR \qquad (14)$$
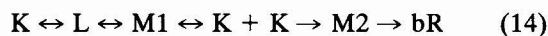
with $\sigma_p = 0.01778$ and $\sigma_d = 0.984$ mOD; these σ_d may be compared with the $\sigma = 0.49$ obtained from fitting four exponentials using VARP. Both models require only seven independent rates because the K $\leftarrow$ M1 backreaction is uniquely related to the other rates involving K, L, and M1. The dashed curves in Fig. 9 show the best fit for the $n4$–$2M.7b$ model. This fit is clearly superior to the fits for the $n = 3$ models in Fig. 8. The $n4$–$2M.7a$ model is rather unorthodox in having a cul-de-sac loop K $\leftrightarrow$ L $\leftrightarrow$ M1 $\leftrightarrow$ K involving two intermediates, L and M1, that are off the main pathway. The $n4$–$2M.7b$ model is rather more orthodox and would be identical, if the K $\leftrightarrow$ M1 path were removed, with the model preferred by Zimanyi and Lanyi (1993), which we will call the $n4$–$2M.6$ model,

$$K \leftrightarrow L \leftrightarrow M1 \rightarrow M2 \rightarrow bR. \qquad (15)$$
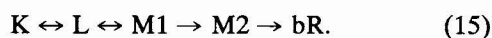
A fit was performed to the $n4$–$2M.6$ model and yielded σ_p of 0.02268 and $\sigma_d = 1.135$ mOD. Although the $n4$–

$2M.6$ model has one fewer independent rate constant than the $n4$–$2M.7b$ model, the K $\leftrightarrow$ M1 path is statistically significant using an F-test (Nagle, 1991b). Even the fit to the $n4$–$2M.7b$ model indicates systematic discrepancies from the $p_i(t)$, as seen in Fig. 9 and as indicated by the $\sigma_p = 0.0178$ being nearly twice the $\sigma_p = 0.0095$ of the fit by sums of four exponentials.

An attempt was also made to find a kinetic model with two L states. This was not expected to be successful because it had not helped much when dealing with Raman data (Nagle, 1991b); the result was rather surprising. The best fitting model with two L states, called $n4$–$2L.6$,

$$K \leftrightarrow L1 \leftrightarrow L2 + K \rightarrow M \rightarrow bR, \qquad (16)$$
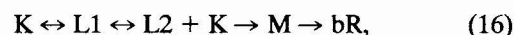
has $\sigma_p = 0.01126$, $\sigma_d = 0.742$ mOD, and only six independent rates. The fit to the $p_i(t)$ is shown by solid lines in Fig. 9 and is clearly superior to the fit from the $n4$–$2M.7b$ model. At this point this model is clearly the best one with fewer than five intermediates.

Further testing involves investigation of some of the choices not taken. The first of these is the use of the K spectrum from Zimanyi and Lanyi (1993) instead of the K spectrum derived in this paper and the alternative $p_i(t)$ shown in Fig. 7. The general conclusions regarding which is the best model continue to hold. The two M models $n4$–$2M.7a$ and $n4$–$2M.7b$ both have $\sigma_p = 0.0268$ (compared to 0.0178 previously). The favored two L model $n4$–$2L.6$ has $\sigma_p = 0.01948$ (compared to 0.01126 previously), and this was again the best fitting model. However, all kinetic models fit much more poorly to these $p_i(t)$ functions as seen by comparing the σ_p of the fits. It is even more remarkable that the simultaneous four exponential fit to these $p_i(t)$ has $\sigma_p = 0.00841$, slightly better than for the $p_i(t)$ analyzed above. The interpretation of these poorer fits using kinetic models in contrast to the equally good fit using general sums of exponentials follows from the fact that the sums of exponentials given by all possible kinetic models is a subset of general sums of exponentials. Therefore, an incorrectly derived set of $p_i(t)$ could be nicely fit with a general sum of exponentials but poorly fit with even the best kinetic model. Inasmuch as the only difference in deriving the two sets of $p_i(t)$ was the K spectrum used, this is additional evidence that the K spectrum derived in this paper is significantly superior to the one obtained previously.

At the lower temperatures $T = 5$ and $15°C$ the $p_i(t)$ were not so well behaved as at $25°C$, with clearly spurious nonzero values of p_K and p_L at late times. Perhaps as a consequence, none of the fits to any model was as good for these temperatures, and clear distinctions were not apparent. However, the $n4$–$2L.6$ model remained about as good as any other model, even though the best σ_p and σ_d were approximately twice as large as the σ_p and σ_d, respectively, obtained from fitting simultaneously to four exponentials. Also, the individual rates were reasonably consistent with Arrhenius behavior as a function of temperature.

Finally, we limited our tests of models with five intermediates to one specific model that has two L and two M
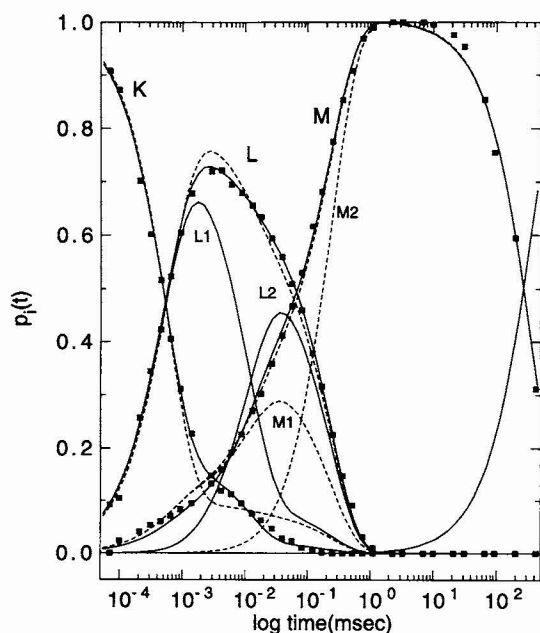

FIGURE 9 Fits to the time course of the intermediates at 25°C assuming four intermediate states. The solid squares are the $p_i(t)$ in Fig. 7. The dashed lines show the fit assuming two M states, model $n4$–$2M.7b$; p_{M1} and p_{M2} as well as $p_{M1} + p_{M2}$ are shown. The solid lines show the fit assuming two L states, model $n4$–$2L.6$; p_{L1} and p_{L2} as well as $p_{L1} + p_{L2}$ are shown.

states. This model, discussed previously (Zimanyi and Lanyi, 1993), will be called the n5–2L2M.8 model:

$$K \leftrightarrow L1 \leftrightarrow M1 \rightarrow M2 \rightarrow bR + M1 \leftrightarrow L2. \quad (17)$$
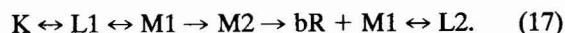
One unusual feature of this model is that the best L1 ↔ M1 rates given by the fit are very fast compared to all other rates. At 25°C the fitted σ_p is marginally smaller by 0.0001 than for the n4–2L.6 model, but for the lower temperatures σ_p is ~0.001 greater than for the n4–2L.6 model. While this makes the n5–2L2M.8 model slightly inferior, it is a very close alternative model.

9. ALTERNATIVE APPROACH TO KINETIC MODELS AND SPECTRA

It was clear from the process of obtaining the time course of the intermediates $p_i(t)$ that there are inherent difficulties in obtaining all the spectra involved, especially when there are clearly more than three intermediates. This suggests that it is worth considering the alternative approach to photocycle kinetics that does not try to obtain the spectra first (Nagle, 1991a; Lozier et al., 1992). Although this approach must assume that the spectra are independent of temperature, the K and M spectra derived earlier in this paper indicate that this is a reasonable assumption.

The difficulty with this alternative approach is that there are local minima in the global nonlinear least squares fitting space, so the result of the search depends upon the initial guess and therefore may not find the best solution. Also, as has been discussed (Nagle, 1991a), with even modest non-zero noise levels it may not be possible to distinguish the "true" solution from spurious solutions that fit the data about equally well. This is likely to be a serious problem for the present data, which have a signal-to-noise ratio of only ~100:1. This is some five times less than the signal-to-noise ratio of previous data (Xie et al., 1987) for which the method was not successful, especially in producing reasonable spectra, although it must be remembered that the full photocycle is complicated by having many more intermediates and that the K spectra produced independently from those data were not as independent of temperature as the present data. Nevertheless, this method should at least be able to determine whether certain models are distinctly inferior to other models.

Results of applying the alternative method to the T = 5, 15, and 25°C data simultaneously were obtained under the assumption that each fundamental reaction rate obeyed a simple Arrhenius relation. When applied to the n4–2L.6 model, the σ of the fit to the data was 0.8124 (which should be compared with VARP fits of 0.46–0.63 for different temperatures); this was the smallest σ of all the models tried. The two L spectra were blue shifted approximately 10–20 nm from bR, but the relative spectra were much too large near 540 nm and too strongly negative near 400 nm. For comparison, the n4–2M.6 model has a σ = 0.8128 with reasonable L and M2 spectra, but with an M1 spectrum that is much too large near 410 nm and much too negative near 570 nm. The model that gives the most reasonable looking spectra is

$K \leftrightarrow L1 \leftrightarrow M \rightarrow bR + L1 \leftrightarrow L2 +$ weak L2 ↔ M; the L2 spectrum, however, is more M-like than L-like, and σ = 0.8239. The alternative method therefore does not yield much new information about the D96N photocycle from these data, except to confirm that several models are equally consistent with the data.

Systematic experimental errors in the data could also account for difficulties in obtaining good results from this method. A number of systematic errors were discussed by Xie et al. (1987). One that was realized subsequently (Lewis and Kliger, 1991) concerns data taken with magic angle polarization, which are subject to an artifact not present in parallel plus two times perpendicular data. Although this artifact becomes negligible in the limit of dilute samples, perhaps the intermediate optical density for these D96N samples is sufficiently high to complicate the rather delicate analysis required by this method.

10. DISCUSSION

The D96N mutant of bR is particularly favorable for photocycle studies. An excellent M spectrum was obtained previously (Zimanyi and Lanyi, 1993). The somewhat simpler method in this paper confirms this M spectrum as well as the fraction cycling at 25°C. It is particularly pleasing, in contrast to earlier studies, that the M spectrum has a single peak in the blue and zero absorbance above 500 nm (see Fig. 3). The feasibility of this analysis is due to the slow decay of M and the lack of any other late intermediates, such as N and O.

This paper shows how the difference spectrum of the first intermediate, which is K for these data, can be obtained by global nonlinear least squares methods (VARP). Using the fraction cycling obtained by the determination of the M spectrum, an improved K spectrum is obtained (see Fig. 5). The previous K spectrum (Zimanyi and Lanyi, 1993) was considerably less well determined by the three filter method than the M spectrum, so it may not be too surprising that there are real differences compared to the new K spectrum.

An important result from the K and M difference spectra is that both are consistent with negligible temperature dependence between 5 and 25°C because the difference spectra at all temperatures can be scaled to the same difference spectrum for each intermediate. Complete proof of temperature independence would require obtaining the fraction cycling at the lower temperatures; this is not reliable even for the D96N photocycle. However, the values of the fraction cycling at 5 and 15°C required for temperature independence are reasonable; also, it would be unlikely that temperature dependence, which involves a multidimensional functional basis, would be duplicated by a one-dimensional variable such as the fraction cycling. Although the true rate constants and the kinetic photocycle change dramatically from 5 to 25°C, the spectra of the true chemical intermediates would not be expected to vary any more than the negligible temperature variation of noncycling bR. Thus, the K and M spectra presented here are likely to be the spectra of true chemical intermediates.

The temperature independence of the K and M spectra support the key assumption used in an ab initio method for photocycle analysis (Nagle, 1991a; Lozier et al., 1992). However, this method demands very accurate data, and a convincing derivation of the best kinetic model with acceptable spectra has not been achieved. Therefore, more attention has been paid in this paper to the alternative method of Zimanyi and Lanyi (1993).

A necessary step in any method of determining the photocycle is the determination of the number of chemical intermediates (Xie et al., 1987). Simple nonlinear least squares fitting of sums of exponentials to the data (VARP) strongly confirm that there are more than three intermediates, which was suspected previously (Zimanyi and Lanyi, 1993), and suggests that there are most likely four intermediates. Singular value decomposition is also consistent with the presence of a fourth component, although the fourth component is more marginal than for VARP fitting, consistent with the possibility that two of the spectra may be similar.

Even if there had been only three intermediates, rigorous solution of the photocycle is not possible because of uncertainty in the third, L, spectrum. With two intermediates in addition to the well-determined K and M, the problem is a good deal harder. However, since the three filter method worked very well for the M spectrum and not too badly for the K spectrum, it is reasonable to consider the L spectrum obtained previously and to suppose that the fourth spectrum is the same as L or M. Using these three spectra allows a straightforward calculation of the time course $p_i(t)$ of the intermediates from the data. As noted previously (Zimanyi and Lanyi, 1993), however, the sum of the $p_i(t)$, i.e., the fraction cycling at time t, did not decrease monotonically as required. This indicates some difficulties in the spectra. (See Section 6.) To try to pinpoint which spectrum is most likely to be the problem, a new criterion was developed (Section 7) that indicates that the L spectrum is the least reliable, although there were also inconsistencies between the M and K spectra.

From the time course $p_i(t)$ of the intermediates, it is computationally easy to obtain the best photocycle if it is assumed that there are three intermediates, K, L, and M, as shown in Section 8. This model is somewhat unusual in that L is on a cul-de-sac. However, the greatly improved fit to the data with four exponentials discussed above requires consideration of a photocycle with four intermediates; this conclusion is further reinforced by the great improvement in the nonlinear least squares fit to the $p_i(t)$ with four exponentials compared to three.

Both visible spectroscopy (Varo and Lanyi, 1991a) and Raman spectroscopy (Nagle, 1991b) indicate that the minimal photocycle for native bR has intermediates K, L, M1, M2, N, and O. Inasmuch as there are no N and O intermediates in the D96N mutant, the most likely suggestion would be that the fourth intermediate would be a second M state. It was therefore rather surprising that the kinetic model that gives the best fit for the D96N photocycle has two values of L both of which are on a single

cul-de-sac. This purely analytical result is not one that we find very appealing or understandable, although one might keep in mind that the early part of the photocycle of the D96N mutant could be less straightforward than for the wild type. Perhaps, in view of the assumptions and uncertainties in the spectra required to obtain the time course of the intermediates, this result should not be given much consideration. A close alternative is the $n5$–$2L2M.8$ model, which has a second extra intermediate and two extra rates but which is statistically almost as good as the $n4$–$2L.6$ model. Choosing between two such alternatives or more complicated models not considered in this paper may involve physical modeling of the mechanism of the photocycle. Even though a complete understanding of the photocycle of bR, even for this mutant, remains a challenge, the methods in this paper enable the choice to be better focused to a few models.

REFERENCES

Bevington, P. R. 1969. Data Reduction and Error Analysis for the Physical Sciences. McGraw-Hill, New York.

Gergely, C., C. Ganea, G. Groma, and G. Varo. 1993. Study of the photocycle and charge motions of the bacteriorhodopsin mutant D96N. *Biophys. J.* 65:2478–2483.

Golub, G. H., and R. J. Leveque. 1979. Extensions and uses of the variable projection algorithm for solving nonlinear least squares problems", *Proc. Army Numerical Analysis and Computing Conference, ARO Report.* 79–3:1–12.

Hessling, B., G. Souvignier, and K. Gerwert. 1993. A model-independent approach to assigning bacteriorhodopsin's intramolecular reactions to photocycle intermediates. *Biophys. J.* 65:1929–1941.

Lewis, J. W., and D. S. Kliger. 1991. Rotational diffusion effects on absorbance measurements: limitations to the magic angle approach. *Photochem. Photobiol.* 6:963–968.

Lozier, R. H., A. Xie, J. Hofrichter, and G. M. Clore. 1992. Reversible steps in the bacteriorhodopsin photocycle. *Proc. Natl. Acad. Sci. USA.* 89:3610–3614.

Lozier, R. H., R. A. Bogomolni, and W. Stoeckenius. 1975. Bacteriorhodopsin: a light-driven proton pump in *Halobacterium halobium. Biophys. J.* 15:955–962.

Nagle, J. F. 1991a. Solving complex photocycle kinetics: theory and direct method. *Biophys. J.* 59:476–487.

Nagle, J. F. 1991b. Photocycle kinetics: analysis of Raman data from bacteriorhodopsin. *Photochem. Photobiol.* 54:897–903.

Nagle, J. F., L. A. Parodi, and R. H. Lozier. 1982. Procedure for testing kinetic models of the photocycle of bacteriorhodopsin. *Biophys. J.* 38:161–174.

Schrager, R. I. 1984. Optical spectra from chemical titration. *Siam J. Alg. Disc. Meth.* 5:351–358.

Varo, G., and J. K. Lanyi. 1991a. Kinetic and spectroscopic evidence for an irreversible step between deprotonation and reprotonation of the Schiff base in the bacteriorhodopsin photocycle. *Biochemistry.* 30:5008–5015.

Varo, G., and J. K. Lanyi. 1991b. Thermodynamics and energy coupling in the bacteriorhodopsin photocycle. *Biochemistry.* 30:5016–5022.

Xie, A. H., J. F. Nagle, and R. H. Lozier. 1987. Flash spectroscopy of purple membrane. *Biophys. J.* 51:627–635.

Zimanyi, L., and J. K. Lanyi. 1993. Deriving the intermediate spectra and photocycle kinetics from time-resolved difference spectra of bacteriorhodopsin: the simpler case of the recombinant D96N protein. *Biophys. J.* 64:240–251.