# Time-Resolved Step-Scan FT-IR Spectroscopy: Focus on Multivariate Curve Resolution

C. Ruckebusch,*[,†,‡] L. Duponchel,[‡] B. Sombret,[‡] J. P. Huvenne,[‡] and J. Saurina[†]

Laboratoire de Spectrochimie Infrarouge et Raman, CNRS UMR 8516, Bât. C5, Ecole Polytechnique Universitaire de Lille, Université des Sciences et Technologies de Lille, 59655 Villeneuve d'Ascq Cedex, France, and Departement of Analytical Chemistry, University of Barcelona, Diagonal 647, E-08028 Barcelona, Spain

Received May 12, 2003

The present paper describes the application of step-scan FT-IR spectroscopy in combination with chemometric analysis of the spectral data for the study of the photocycle of bacteriorhodopsin. The focus is on the performance of this instrumentation for time-resolved experiments. Three-dimensional data—spectra recorded over time—are studied using various factor analysis techniques, e.g., singular values decomposition, evolving factor analysis, and multivariate curve resolution based on alternating least squares. Transient intermediates formed in the time domain ranging from 1 $\mu$s to 6.6 ms are clearly detected through reliable pure time evolving profiles. At the same time, pure difference absorbance spectra are provided. As a result, valuable information about transitions and dynamics of the protein can be extracted. We conclude first that step-scan FT-IR spectroscopy is a useful technique for the direct study of difficult photochemical systems. Second, and this is the essential motivation of this paper, chemometrics provide a step forward in the description of the photointermediates.

## INTRODUCTION

The specificity as well as the broadband nature of FT-IR spectroscopy makes it among the most effective ways to get information about the structure and properties of molecules. Moreover, dynamic vibrational spectroscopy[1] addresses new issues in the field of photoreactions. It probes transient photochemical systems and provides knowledge of intermediates along the reaction pathway.

Step-scan FT-IR[2] is a time-resolved spectroscopic technique combining laser excitation with fast IR differential detection. Since it detects small changes in the presence of large background absorption, it has extensively been used to study biological[3−5] and chemical systems.[6,7] Thus, step-scan FT-IR permits a detailed description of structures and dynamics in the time domain ranging from microsecond[8] to nanosecond.[9] Nevertheless, an optical trigger may be required to initiate cyclic processes on a fast time scale. It can be provided coupling the excitation of a pulsed Nd:YAG laser source to the interferometer of the spectrometer. The ensuing spectroscopic changes characterize the evolution of the intermediates along the reaction pathway. Unfortunately, this external optical triggering mechanism limits the application of the technique to reversible systems, although specific experimental setups allow the study of noncyclic reactions.

Among the potential photochemical systems, the photocycle of bacteriorhodopsin (bR) is of major importance and provides an ideal model of light-driven proton pump mechanisms,[10−12] providing results that generalize to many other protein systems. Upon absorption of light, the protein undergoes a cyclic reaction that is completed within the millisecond range. It involves several intermediate states that occur in a roughly linear order and have a time constant ranging from the picosecond to millisecond.[10,11,13] However that may be and despite all the work that has been done, the functioning of this protein is still subject to intense work and controversy.[13,14]

Time-resolved spectroscopic techniques provide two-way data sets: spectra are recorded as the system is evolving with time. Surprisingly, despite it being a very meaningful source of information, the state of the art for the detection of bR photointermediates relies on the analysis of the absorbance changes at one or a few single wavenumbers.[7,15−17] This approach of the data set is strongly wavenumber-dependent and often debatable since measured points always reflect a mixture of contribution from several intermediate species. Since the pioneered model-independent work[18] using factor analysis, namely, singular values decomposition, and iterative target transformation to obtain real factors, methodologies free of assumption about the kinetics have been sadly lacking.[19] The information that is attempted here is more comprehensive spectroscopic insights into the full reaction pathway as well as the multivariate and simultaneous characterization of all the transient species in terms of lifetime, appearance, and disappearance. Powerful data treatments should thus provide improvements and robustness in the extraction of spectroscopic and dynamic information.

Several approaches can be attempted in the qualitative analysis of two-way data sets. The present study takes advantage of the multivariate curve resolution based on an alternating least-squares optimization (MCR-ALS method).[20−22] In the curve resolution, an evolving mixture matrix is decomposed into the product of two matrices, which are related respectively to the time domain and to the spectral domain of the data set. These two small matrices contain

* Corresponding author Fax: 0033 3 20434902; e-mail: Cyril.ruckebusch@univ-lille1.fr.
† Université des Sciences et Technologies de Lille.
‡ University of Barcelona.

10.1021/ci034094i CCC: $25.00 
Published on Web 10/25/2003

the pure time profiles (e.g., pseudo-concentration profiles) and the pure spectra for each modeled component, that is to say, each transient species. To reduce the resolution ambiguities and to get meaningful biochemical solutions, several constraints are applied to the time profiles. Such constraints, based on knowledge or/and reasonable assumptions about the system, can be nonnegativity or unimodality for example. But first and foremost, since the number of intermediates is unknown, the complexity of the data set as well as reasonable initial estimates of the pure factors should be estimated through factor analysis. This is probably the most difficult task since the results are strongly data-set-dependent. Singular value decomposition (SVD) provides information about the rank of the data matrix. In the same manner, principal component analysis (PCA) can be used to give a first insight of pure profiles,[23] although principal components are only abstract factors resulting from specific mathematical constraints. On another hand, if pure information is suspected in the data set, several methods such as simple-to-use interactive self-modeling mixture analysis (SIMPLISMA)[24] or orthogonal projection approach (OPA)[25] aim at finding the purest rows (i.e., spectra) or purest columns (i.e., time profiles). At last, if measurements are time-organized in the data matrix, the appearance and disappearance of intermediates can be exploited to estimate pure factors. Evolving rank methods, in particular evolving factor analysis (EFA)[26] and fixed size window evolving factor analysis (FSWEFA),[27] are thus applicable.

We demonstrate here the potential and performance of the MCR-ALS methodology for the resolution of the bR photocycle, which is chosen as a performance test of the instrumentation. It is a powerful approach to provide spectral and temporal information about the structure and dynamics of all the intermediates that are accessible to the experiment.

## EXPERIMENTAL SECTION

**Sample Preparation.** Lyophylized wild-type bacteriorhodopsin as purple membrane was provided by Munich Innovative Biomaterials (MIB Gmbh). Aqueous suspension in 0.1 M KCl, pH 7, was prepared. To reduce the background absorption in the mid-infrared range but at the same time maintain the photocycle,[16] homogeneous hydrated films were allowed to dry on $CaF_2$ windows (13 mm × 2 mm) at room temperature.

**Step-Scan Difference Spectroscopy.** The step-scan FT-IR experiments were performed using a Nicolet Magna 860 FT-IR spectrometer with auxiliary external module in step-scan mode. It is combined to the second harmonic of a Q-switched Nd:Yag laser (Quantel, 532 nm, 4 ns·pulse$^{-1}$, 3 mJ·pulse$^{-1}$, 20 Hz). The synchronization of the laser source with the data acquisition system was carried out with a pulse generator (Standford, DG535) and a home-built electronic device.

The setup was positioned on a vibration isolation table, in a temperature-controlled room. The sample compartment was purged with dry air. These conditions increased the stability of the system.

The detector was a 20 MHz photovoltaic MCT detector with a built-in preamplifier that provided AC output and DC output. Both were digitized simultaneously ensuring phase matching. While the DC output provided the static interferogram, the AC signal measured small changes in the transient FT-IR signals. It was amplified by a factor of 10 (Standford, SR560) to reach a signal intensity running the full dynamic range of the digitizer. Besides, noise rejection was necessary in the step scan and was provided through low-pass (300 Hz) and high-pass (10 kHz) filters. Furthermore, a long-pass optical filter (cutoff at 5 $\mu$m) was set in front of the detector to prevent scattered photolysis light and to limit the bandwidth of the measurements. The amount of possible artifacts was thus reduced. The resulting lower and upper spectral limits were 950 and 2000 cm$^{-1}$, corresponding to the cutoff of the optical filter, the detector, and the $CaF_2$ window.

In step-scan FT-IR, difference interferograms are elaborated from successive retardation positions. In this experiment, the static signal was first accumulated during 30 ms at each mirror position. A single laser shot was then performed and—after a posttrigger delay here set to 1 $\mu$s—the intensity changes were recorded with respect to the ground state. A logarithmic time scale with a factor 1.25 and an initial dwell of 0.5 $\mu$s was chosen for the acquisition. The sequence consisted of 27 time sampling points. In practice, each of these points was averaged over 100 ns. At last, the trigger interval was set to 0.3 s, a large enough time to ensure that the bR has returned to its fundamental state. To achieve satisfactory signal-to-noise ratio, 16 scans were averaged per spectrum. The spectral resolution is also an important factor in step-scan since it is connected to the number of mirror positions to sample. We recorded spectra with 8 cm$^{-1}$ resolution, the experimental length being 3.5 h in these conditions. Two time domains were sampled under these conditions, from 1 $\mu$s to 660 $\mu$s and 100 $\mu$s to 6.6 ms.

The Nicolet Software OMNIC 5.1 package was used for parameter settings, instrument control, and Fourier transformation.

**Data.** The analysis of two complementary data sets is proposed in this study. Each one covers a specific time interval, but the range they sample slightly overlaps. It provides a way to internally check the consistency of the pure profiles extracted. Each data set contains 27 spectra and shares the same informative spectral range with upper and lower limits selected at 1866 and 1100 cm$^{-1}$.

Data set 1 contains time-resolved FT-IR difference spectra recorded with logarithmic time scale from 1 $\mu$s to 660 $\mu$s. The second experiment (data set 2) concerns absorbance changes from 100 $\mu$s to 6.6 ms. In each case, the spectral intensities are stored in a matrix **D** with one spectrum per row. The dimensions of **D** are the number of time positions (spectra) sampled during the transient experiment by the number of channels (wavenumber), that is to say, 27 lines of 199 numerical points each.

The data pretreatment consisted of a Savitsky−Golay smoothing[28] with second-order polynomial function and a 7 points moving window. This smoothing was performed successively in spectral and time directions providing, thus, an efficient data denoising without any significant loss of analytical information (see Figure 1a,b).

## DATA ANALYSIS METHODS

The purpose of multivariate curve resolution[29] techniques is to produce a factorization of a data matrix **D** into the

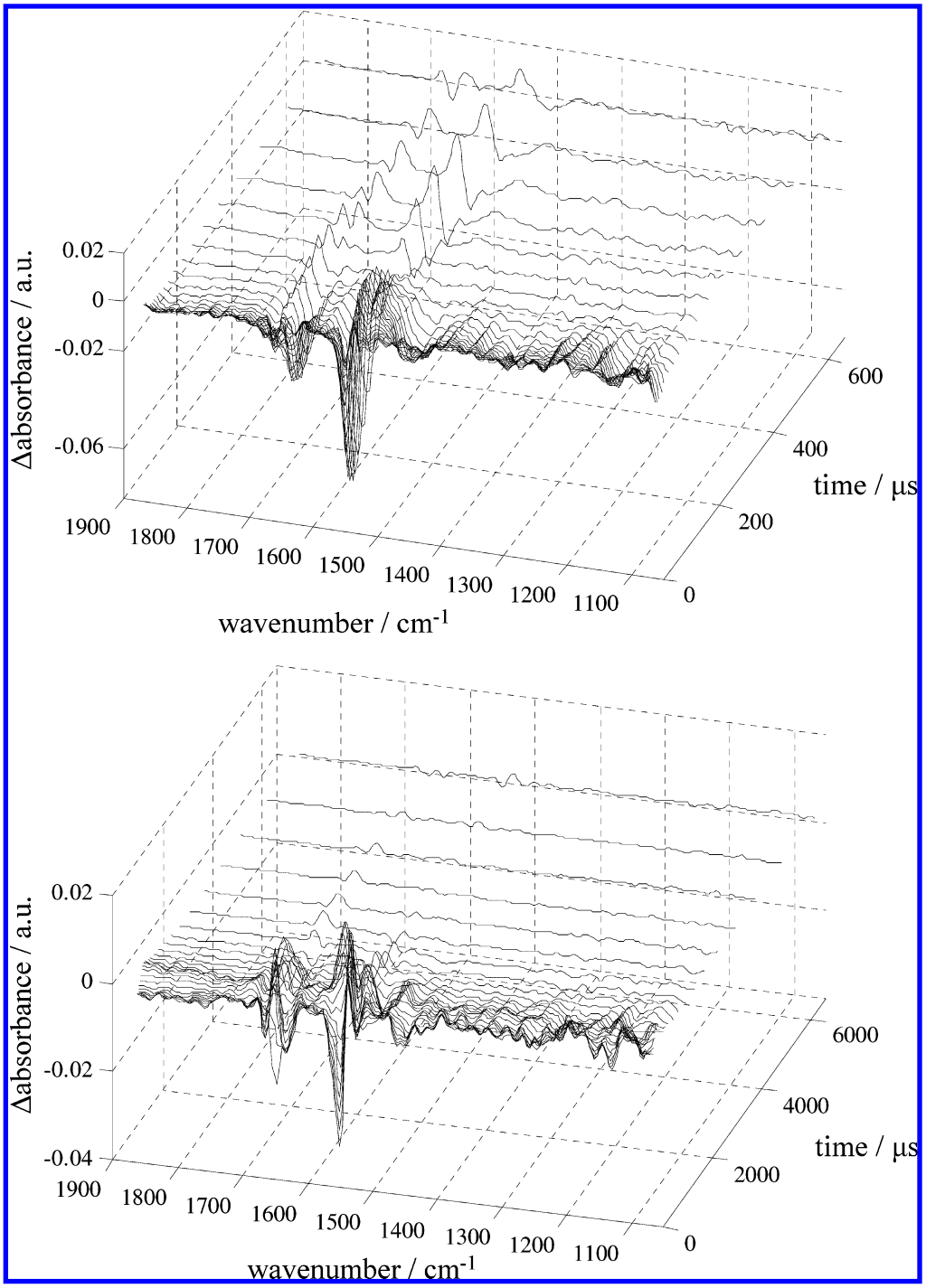


**Figure 1.** Step-scan FT-IR difference absorbance spectra: (a) data set 1 (1−660 $\mu$s); (b) data set 2 (0.1−6.6 ms).

product of two simpler matrices **C** and $\mathbf{S}^T$ (eq 1), namely, matrices of pure intensity (or concentration) and pure spectral profiles. **E** is the matrix of residuals: contributions left unexplained by the extracted components.

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad (1)$$

**Multivariate Curve Resolution−Alternating Least-Squares.** MCR-ALS is a data analysis method that has proved to be efficient in a wide variety of chemical applications. It relies on an iterative ALS optimization of the **C** and $\mathbf{S}^T$ matrices under applying constraints. Constraints force a given spectral or time profile to fulfill a defined feature reducing thus the ambiguity of the solution obtained.

Before starting the optimization procedure, the number of components in **D**, i.e. the number of chemical or physical contributions to the response, must be determined. It is a step of major importance for good resolution but seldom a trivial task in real analytical chemical situations. Singular value decomposition (SVD) of the data matrix is a good simple tool to estimate the rank of **D**, although the threshold that separates chemical contribution from noise contribution is most of the time hard to settle. In this case, the methodology proposed consists of attempting the different possible situations, the MCR-ALS optimization being then analyzed with respect to the lack of fit. Moreover, to be retained, a solution must above all present reasonable shapes

for the profiles and spectra as well as consistency with respect to initial estimates.

The next step consists of obtaining initial estimations for the species to be included in the optimization. For this purpose, as kinetic data are investigated, advantage might be taken from the intrinsic order of the spectra in evolutionary methods such as EFA or FSWEFA. The first method provides initial estimates of the time profiles, whereas FSWEFA detects the appearance of the compounds. The index of the spectra corresponding to the apparition of a new compound might then be used as the initial estimates of the pure spectra. It is thus possible to calculate the pure conjugated profiles in an iterative way solving eq 1 by alternating least squares. Conversely, in this study, pure variable (column or row) methods are not encouraged since we deal with protein FT-IR spectra presenting multiple overlapping bands. Here, pure variable as well as pure intermediates are seldom if ever detectable.

During this study, the following constraints have been applied: (1) nonnegativity[30] of the concentration profiles (all the elements in a profile have to be positive); (2) unimodality of concentration profiles (because the apparition of the intermediate is sequential, only one maximum per profile is allowed); (3) local selectivity, which imposes the presence of only one of the species over a specified range and might be useful in the field of proteins.[31] Constraints are usually applied with a certain tolerance level; up to 5% departure from the strict restriction have been allowed here. To estimate the model fit obtained, the lack of fit (lof) is calculated according to the expression as follows.

$$\%\text{lof} = \left(\sum (d_{ij} - d^*_{ij})^2 / \sum d_{ij}^2\right)^{1/2} \times 100 \qquad (2)$$

where $d_{ij}$ is the experimental absorbance at the $j$th wavenumber in the $i$th spectrum and $d^*_{ij}$ is the reconstructed value. The convergence criterion in the MCR-ALS is based on the relative change of the lack of fit during the optimization between two consecutive iterations. The default value is 0.1%. Furthermore, the maximum number of iterations in the optimization step is 50.

An attractive strategy for improving the resolution and minimizing ambiguities is based on the simultaneous analysis of various correlated runs sharing spectral or time information.[20] In most cases, spectral measurements in different repetitions of the experiment are performed on the same spectral domain. Nevertheless, a condition is that the spectrum of a given intermediate is the same whatever the run. It is fulfilled unless physical or chemical properties are changing in the different runs. Under this assumption, columnwise matrix augmentation on the spectral domain is possible for simultaneous analysis of, for example, two different runs $\mathbf{D}_1$ and $\mathbf{D}_2$ (eq 3).

$$D = \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{bmatrix} [\mathbf{S}^T] \qquad (3)$$

$\mathbf{C}_1$ and $\mathbf{C}_2$ matrices contain the time profiles associated with each run. Apart from those constraints implemented for the analysis of the individual data sets (see above), additional restrictions can be used in this simultaneous analysis as follows: an equal shape in the spectral profiles of species, i.e., the spectrum of a given species is unique whatever the run.

**Evolving Factor Analysis, Evolving Window Factor Analysis.** EFA and FSWEFA provide initial estimation of the profiles examining how the singular values evolve with time. Additionaly, selective regions might also be detected and constraints to be implemented may be inferred, eventually solving at least partially the ambiguity of the resolution. Both techniques are based on SVD of successive submatrices of the data set.

In forward EFA, rows are successively added to an initial submatrix. Plotting the eigenvalues at each step, it is thus possible to detect the appearance of a new compound. Since the analysis is also performed in the backward direction, it is conversely possible to detect the disappearance of species. At last, under the assumption that the first compound to appear is the first to disappear, concentration profiles are reconstructed.

In FSWEFA, small submatrices are based on a window of a user-defined number of rows (i.e., spectra), which is moved over the data set from the beginning to the end. In this study, a window of three consecutive spectra is used. For each moving window, the singular values are calculated and plotted as a function of the window position or, in a meaningful manner, as a function of time. From the variation observed for eigenvalue lines associated with noise, a threshold to decide on the significance of a component with respect to the noise level can be deduced, providing, thus, a powerful method to detect minor compounds.[32]

The data treatments were performed with Matlab using a set of functions, written by R. Tauler, which are freely available at http://www.ub.es/gesq/mcr/mcr.htm.

## RESULTS AND DISCUSSION

Parts a and b of Figure 1 show the evolution of the difference absorbance for data sets 1 and 2. While FT-IR probes conformations and dynamics of biological molecules, the strong overlap between the frequencies of the vibration modes makes it difficult to detect structural changes of individual groups. Moreover, because reproducibility is often challenging for FT-IR spectra of proteins, in particular in transient experiment with laser excitation,[5,17] it has been inspected by comparing the spectra obtained from the two data sets at a defined time. The infrared spectrum of a protein in the 1500−1700 $cm^{-1}$ region mainly presents contributions from all the amide groups as well as vibrational modes from side-chain vibrations of several amino acids residues. The strongest band is located around 1550 $cm^{-1}$ and corresponds to the C═C double bond of the retinal photolyzed states.

Concerning Figure 1a, on which the first intermediates of the bR can eventually be probed, it is observed that the initial spectra are already concerned with nonzero absorbance changes. It means that some phenomena are faster than the transient accessible to our experiment. One can also notice that changes are observed along the full time scale of this experiment as well as the latter. On the contrary, in Figure 1b, the difference spectra tend to go back to flat line at the end of the process as the system returns to a state similar to the one it had started from.

**Estimating the Number of Transient Compounds.** In many analytical situations, it is tricky to distinguish the relevant information from the noise contribution and to definitely estimate the total rank of an experimental data set.

**Table 1.** Singular Values Decomposition of the Experimental Data Matrices[a]

| SVD data set 1 | | SVD data set 2 | |
|---|---|---|---|
| all channels | selected channels | all channels | selected channels |
| 0.446 | 0.017 | 0.172 | 0.013 |
| 0.057 | 0.007 | 0.053 | 0.006 |
| 0.035 | 0.005 | 0.023 | 0.002 |
| 0.025 | 0.004 | 0.017 | 0.002 |
| 0.018 | 0.002 | 0.013 | 0.001 |
| 0.012 | 0.001 | 0.009 | 0.001 |
| 0.010 | | 0.006 | |
| 0.007 | | 0.004 | |
| 0.006 | | 0.003 | |
| 0.003 | | 0.002 | |

[a] All channels, full spectral region; selected channels, spectral region with no compound of interest.

Nevertheless, a first insight in the estimation of the number of chemical species is usually provided from SVD.[20,31,33] Table 1 gathers the results obtained from SVD of the two data sets. In both cases, the left-hand column of the table concerns the decomposition of the full-channel data sets, whereas the right-hand one presents the one performed on 10 channels selected where no band has been detected. The noise level is thus estimated through the highest contributions obtained, respectively 0.017 and 0.013 for data set 1 and data set 2. These values provide reasonable estimations of the thresholds to apply for the rejection of noise in the decomposition of the full spectral data sets. Hence, the number of significant contributions to the first data set is roughly estimated to four. At the same time, for data set 2, four contributions are detected as well. Furthermore, factor analysis methods such as evolutionary rank methods are tested in order to get a better insight into the numbers of intermediates.

**Evolving Factor Analysis.** In spectral data sets evolving with time, some factor analysis methods can be particularly interesting for getting initial estimates of the pure profiles or pure spectra of the different components. In particular, EFA is a useful technique for evaluating the total rank of a data matrix and for providing starting values for the time profile associated with each contribution. To illustrate this, Figure 2a shows the reconstructed abstract profiles from EFA of data set 1 when considering four contributions. Their analyses may help to detect and identify the different transient states in the interval from 1 $\mu$s to 660 $\mu$s. In this way, the first factor presents an almost instantaneous decay. The second profile reaches its maximum in a few tens of microseconds. Concerning the third factor, it presents a bell shape with a maximum around 200 $\mu$s. Finally, the last appearing intermediate is present almost over the full time range. Because the next extracted contributions (the fifth and higher) do not significantly differ from the noise contributions, the magnitude of four factors seems here significant, which is in agreement with the SVD results previously discussed.

At the same time, local rank information is obtained from FSWEFA. The index of the samples corresponding to the apparition of intermediates can also be inferred, providing initial values of the pure spectra. As an example, Figure 2b represents the lines describing the two first evolving factors obtained from FSWEFA of data set 1. They provide information on the index of the samples where intermediates

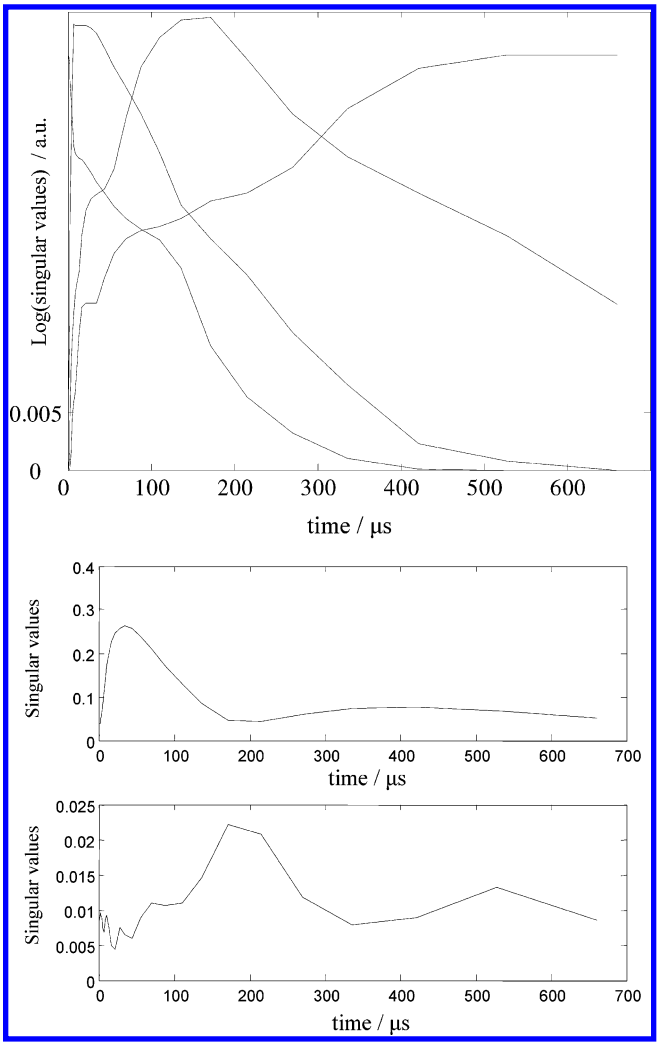


**Figure 2.** Evolving factor analysis of data set 1 (1−660 $\mu$s): (a) four first reconstructed profiles from forward and backward EFA; (b) factors from FSWEFA (top panel, first factor; bottom panel, second factor).

are detected. On the first factor extracted (see Figure 2b, top panel), a first compound is clearly detected around 40 $\mu$s, which is consistent with the EFA results. Another intermediate is detected later, although the information is broader. On the other hand, the analysis of the evolution of the second factor (see Figure 2b, bottom panel) shows a relevant feature at 200 $\mu$s, which should correspond to the third EFA profile. Another contribution—just above the noise level estimated from the analysis of the baseline zone—is detected slightly after 500 $\mu$s. This contribution is found significant as it will be detected from the analysis of data set 2 as well. FSWEFA appears here to be a performing method to extract minor information. At last, the third factor is not considered since it is mainly associated with noise.

The same methodology was repeated for data set 2 in which the time range under study was 0.1−6.6 ms. Results have shown an initial decaying profile and three successive peaks (approximately at 200 $\mu$s, 500 ms, and 1 ms or later).

**Alternating Least-Squares Optimization.** Further studies were conducted to resolve the spectra and the time profiles of the intermediates detected during the full time range. Concentration profiles from EFA and spectra corresponding to the time positions pointed by FSWEFA were indifferently used as initial estimates in the MCR-ALS optimization.

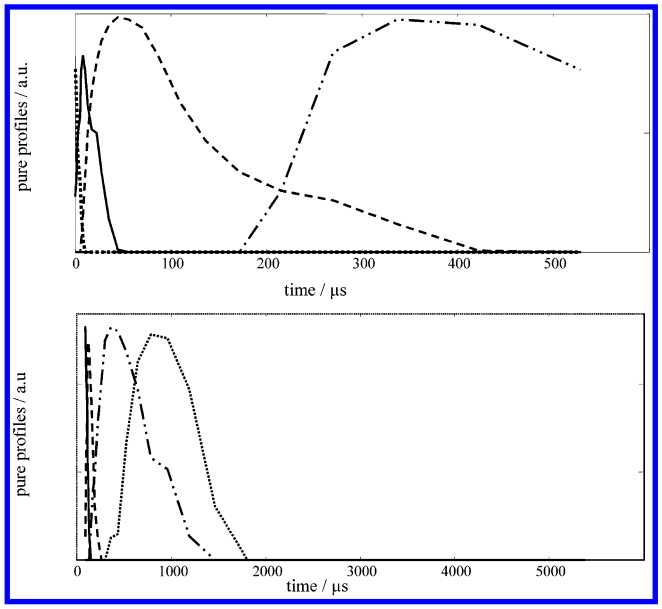


**Figure 3.** Pure time profiles obtained from MCR-ALS on individual experimental data sets: (a) data set 1 (dotted line, initial decay; solid line, first intermediate; dashed line, second intermediate; dashed−dotted line, third intermediate); (b) data set 2 (solid line, first intermediate; dashed line, second intermediate; dashed−dotted line, third intermediate; dotted line, last intermediate).

Moreover, to get meaningful results, the natural constraints previously described were applied to the concentration profiles (here nonnegativity, unimodality, and local zero-concentration ranges, in all cases applied to the time profiles).

**Individual Analysis of the Experimental Data Sets.** Figure 3a plots the concentration profiles obtained from the analysis of data set 1 (1−660 $\mu$s). The corresponding lack of fit was 5%. From the profile observed on this figure, a rapid initial decay is detected. Then, a second rather sharp emerging profile is observed in this range corresponding to an intermediate, which is detectable over about 40 $\mu$s. On the contrary, the third factor presents a wider shape. It reaches its maximum just before 100 $\mu$s and decays continuously until 200 $\mu$s or more. Concerning the fourth factor, it appears just before 200 $\mu$s, reaches its maximum value around 350 $\mu$s, and is subsisting over a length greater than the interval regarded here.

In the same manner, Figure 3b is a plot of the results obtained for the analysis of the second data set (0.1−6.6 ms). The first factor presents a decaying shape that lasts around a hundred microseconds. It corresponds to the second profile of the data set previously studied (see Figure 3a, solid line). The profile of the second intermediate, Figure 3b, approximately corresponds to a lifetime of 200 $\mu$s, and while disappearing, a third transient species is emerging which is centered around 500 $\mu$s. These two profiles are in good agreement with the last two profiles extracted from data set 1 as well (see Figure 3a, dashed and dashed−dotted lines). At last, a fourth intermediate is present in the millisecond scale, i.e., from 500 $\mu$s to 2 ms.

**Simultaneous Analysis of the Two Experimental Data Sets.** The analysis of the individual data sets provides information about the species present in the two overlapping intervals. As we have seen before, the time profiles obtained independently from the two repetitions of the experiment are consistent. To get better insight into the transient

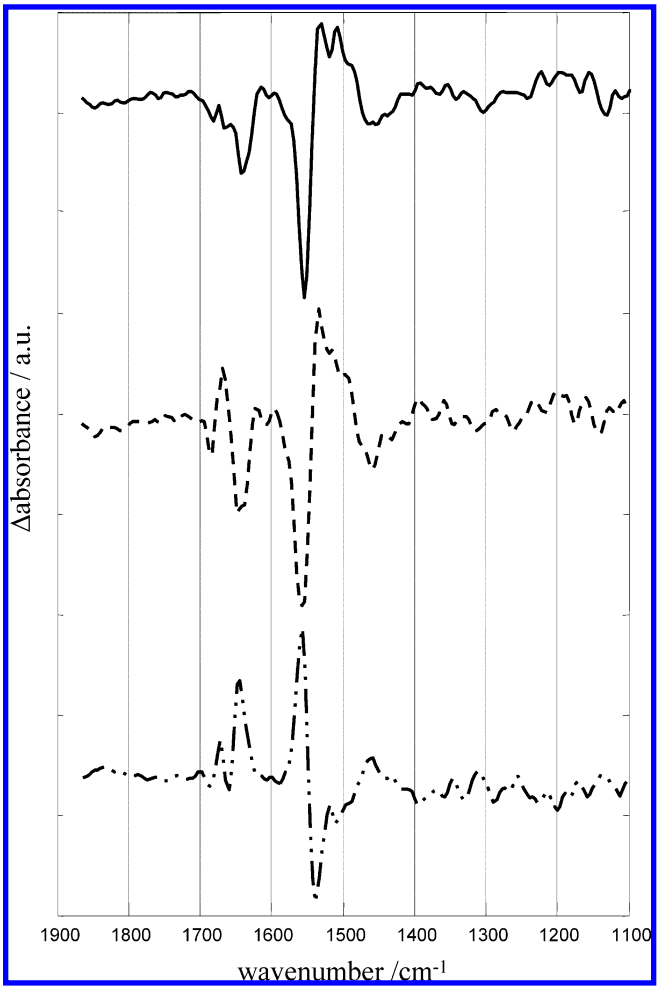


**Figure 4.** Pure difference absorbance spectra from MCR-ALS on the augmented data set (solid line, first intermediate; dashed line, second intermediate; dashed−dotted line, third intermediate).

phenomena, a columnwise augmented matrix was analyzed by starting from different initial conditions. It is also a way to ease interpretation through implementation of appropriate constraint, proposing unique pure spectrum solution for each of the intermediates detected. Under the conditions of the present work, the rank of the augmented matrix is equal or less than the sum of the ranks of the single matrices.[32] An additional constraint, namely, equal shapes in the spectral profiles, was applied to this simultaneous analysis of the two data sets.

A solution recovering five species was obtained providing a lack of fit of 9.6%. The resolution we get is still consistent with the ones obtained in the single matrix analyses, except for the last factor of data set 2 that was not recovered in the analysis of the augmented-matrix analysis. The resulting pure difference spectra are discussed in the next section.

**Interpretation of the Pure Profiles and Spectra of the bR Metastable States.** In this section, the pure difference spectra and pure intensity profiles of the detected photointermediates of bR are discussed. The pure spectra presented in Figure 4 have been obtained through MCR-ALS of the augmented data set, providing a multivariate (full-spectra) approach for the molecular description of the photointermediates. This is a convenient method to obtain—from raw data—pure bR spectra that are otherwise not accessible directly and to allow to trace the corresponding time

dynamics. It should be noticed that other experimental approaches exist, such as low-temperature spectroscopy, which enable trapping specific intermediates.[35,36]

Under absorption of light, bR first undergoes a rapid photoisomerization in the excited state.[13] This reaction is in the picosecond domain, i.e., definitely out of the range of the experiment here.

This state gives rise to the L intermediate of bR which is known to have a specific lifetime around 40−50 $\mu$s.[13,37] This transient state is clearly detected from our data analysis in the time range 1−660 $\mu$s. It can already be suspected for example from the first FSWEFA factor (see Figure 2b), and then it is confirmed through MCR-ALS optimization. The second profile of Figure 3a is thus corresponding to the appearance/disappearance of this intermediate. Moreover, we provide here a way to get molecular information describing this state because its corresponding difference spectrum is obtained as well (see Figure 4, solid line). It is important to note that this spectrum was assessed without any assumption about the time at which the L intermediate is present. We observe small structural changes of the protein, its environment, and the chromophore,[10] mostly in the amide I and amide II range. The 1600−1700 $cm^{-1}$ region contains amide I (C═O stretching and N−H bending of the protein backbone) as well as modes of the C═N and N−H stretching of the protonated Schiff base. In other respects, bands in the 1500−1600 $cm^{-1}$ region are due to amide II (C═N stretching and N−H bending vibration of the backbone) and C═C stretching vibration of the retinal. This latter band is located at 1526 $cm^{-1}$ in the unphotolyzed state and shifted around 1550 $cm^{-1}$ in the following L, M, and N states.[37] It should be noticed that the positive contribution around 1525 $cm^{-1}$ was not expected but has been previously observed in other light-driven proton pump systems.[38] Such difference spectra as well as molecular dynamic studies have been extensively discussed in the literature.[14,36,37,39] In this manner, the strong negative C═C stretching band corresponding to the bR depletion has been often chosen as univariate kinetic probe.[8,9,17]

The next step that is gone through by bR is the so-called M state in which the Schiff base is deprotonated. The L to M transition is of particular interest since (i) it is the most critical step in the proton transfer[13,15,40] and (ii) there is controversy about the number of photointermediates. Effectively, models in which the M state is a progression of two or more substates rather than a static entity are supported.[41] It is commonly accepted that the first substate ($M_1$) appears at 10 $\mu$s or more and is in equilibrium with the L state before reaching its maximum around 100 $\mu$s. In a second step, the substate ($M_2$) is defined with a maximum at 300 $\mu$s. Furthermore, this intermediate presents a rather longer lifetime, on the order of milliseconds. These two metastable configurations only differ from the point of view of their secondary structure. Focusing on the time profiles obtained from data set 1, they are both detected through the two last evolving profiles. But it is not as clear in the second interval. The second profile of Figure 3b presents a lifetime corresponding to $M_1$—although it is slightly shorter—but overall, its spectrum, previously extracted from the two first data sets is confirmed from the augmented data analysis. This spectrum (see Figure 4, dotted line) is mainly characterized by changes in the amide I region, where secondary structure modification can be observed. In agreement with other works, we notice the increasing intensity at 1650 $cm^{-1}$ and the two negative contributions around it, characterizing the first M state. The next intermediate can be completely followed only through the analysis of the second data set ranging through the millisecond time scale. One should note that its appearance around 200 $\mu$s has already been detected on Figure 2. This is here confirmed from the Figure 3 (dashed−dotted lines), and a lifetime shorter than 2 ms can be furthermore estimated. The corresponding pure difference absorbance spectrum is plotted in Figure 4 (dashed−dotted line). Several features of the spectra show that the dominant photointermediate present over this time range is M. A wide contribution is observable in the C═N stretching as well as a positive contribution in the amide I region. We also notice that the contributions at 1550 and 1530 $cm^{-1}$ are reversed with respect to the L difference spectrum. Despite that it could correspond to the N photointermediate,[17,42] we did not succeed in recovering the fourth profile extracted from data set 2 (Figure 3, dotted line) from the augmented data resolution. Besides, the next intermediate of the bR (namely, O) was not detected in this study. It is consistent with other observations in bR films.[43] Finally, the very fast decay at the beginning of the investigated time domain (see Figure 3a, dotted line) is left unexplained but might be attributed either to a nonchemical factor[15] or to an unusually long K−L transition time.[41]

## CONCLUSION

As a performance test of a step-scan FT-IR experimental setup, time-resolved difference absorbance spectra of the photocycle of bacteriorhodopsin are studied. Although this is a widely studied system presenting successive metastable intermediates, some aspects of the photophysical dynamics at the molecular level are not fully understood yet so that it is a field under intensive research. Methods are required for full-spectrum characterization of the intermediates of the photocycle, thus providing an essential step for evaluating the fundamental mechanisms of the protein.

In the present paper, multivariate curve resolution based on alternating least squares has been applied to the analysis of the whole spectral data set. The time domain under study, ranging from 1 $\mu$s to 6.6 ms, includes the formation of L and M transient species, and it is shown that time-resolved FT-IR spectroscopy is a suitable technique for generating valuable and consistent dynamic information about the intermediates. Moreover, it allows recovery of their pure spectra without any assumption about the photochemical mechanisms.

As a final remark, the whole procedure, which combines time-resolved step-scan FT-IR spectroscopy and chemometrics, results in a powerful way for obtaining kinetic and spectral information about (bio- or macro-) molecules so that this strategy can be applied to many other systems in different fields, such as environmental analysis.[43]

## REFERENCES AND NOTES

(1) Palmer, R. A. Step-scan FT-IR: A versatile tool for time-resolved and phase-resolved spectroscopy. *Spectroscopy* **1993**, *8*, 26−34.
(2) Palmer, R. A.; Chao, J. L.; Dittmar, R. M.; Gregoriou, V. S.; Plunkett, S. E. Investigation of time-dependent phenomena by use of step-scan FT-IR. *Appl. Spectrosc.* **1993**, *47*, 1297−1310.

(3) Rödig, C.; Siebert, F. Monitoring fast reaction of slow cycling systems. *Vib. Spectrosc.* **1999**, *19*, 271−276.
(4) Rammelsberg, R.; Boulas, S.; Chorongiewski, H.; Gewert, K. Molecular reaction mechanisms of proteins monitored by nanosecond step-scan FT-IR difference spectroscopy. *Vib. Spectrosc.* **1999**, *19*, 143−149.
(5) Gerwert, K. Molecular reaction mechanisms of proteins monitored by time-resolved FT-IR spectroscopy. *Biol. Chem.* **1999**, *380*, 931−935.
(6) Vasenkov, S.; Frei, H. Time-resolved study of acetyl radical in zeolite NaY by step-scan FT-IR spectroscopy. *J. Phys. Chem. A* **2000**, *104*, 4327−4332.
(7) Sun, X. S.; Nikiforov, S. M.; Yang, J.; Colley, C. S.; George, M. W. Nanosecond time-resolved step scan FT-IR spectroscopy in conventional and supercritical fluids using a four-window infrared cell. *Appl. Spectrosc.* **2002**, *56*, 31−41.
(8) Braiman, S. M.; Ahl, P. L.; Rothshild, K. J. Millisecond Fourier-transfor infrared difference spectra of bacteriorhodopsin's M412 photoproduct. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 5521−5225.
(9) Hage, W.; Kim, M.; Frei, H.; Mathies, R. A. Protein dynamics in the bacteriorhodopsin photocycle: A nanosecond step-scan FT-IR investigation of the KL to L transition. *J. Phys. Chem.* **1996**, *100*, 16026−16033.
(10) Rothschild, K. J. FT-IR difference spectroscopy of bacteriorhodopsin: Toward a molecular model. *J. Bioenerg. Biomembr.* **1992**, *24*, 147−166.
(11) Lanyi, J. K. Proton translocation mechanism and energetics in the light-driven pump bacteriorhodopsin. *Biochim. Biophys. Acta* **1993**, *1183*, 241−261.
(12) Oesterheldt, D.; Stoeckenius, W. Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*. *Nature (London), New Biol*. **1971**, *233*, 149−152.
(13) Neutze, R.; Pebay-Peyroula, E.; Edamn, K.; Royant, A.; Navarro, J.; Landau, E. M. Bacteriorhodopsin: A high-resolution structural view of vectorial proton transfer. *Biochim. Biophys. Acta* **2002**, *1565*, 144−168.
(14) Baudry, J.; Tajkhorshild, E.; Molnar, F.; Phillips, J.; Schulten, K. Molecular Dynamics study of bacteriorhodopsin and the purple membrane. *J. Phys. Chem. B* **2001**, *105*, 905−918.
(15) Rammelsberg, L.; Hebling, B.; Chorongiewski, H.; Gerwert, K. Molecular reaction mechanisms of proteins monitored by nanosecond step-scan FT-IR difference spectroscopy. *Appl. Spectrosc.* **1997**, *51*, 558−562.
(16) Uhmann, W.; Becker, A.; Taran, C.; Siebert, F. Time-resolved FT-IR absorption spectroscopy using a step-scan interferometer. *Appl. Spectrosc.* **1991**, *4*, 390−397.
(17) Hutson, M.; Alexiev, V.; Shilov, S. V.; Wise, K. J.; Braiman, M. S. Evidence of a perturbation of arginine-82 in the bacteriorhodopsin photocycle from time-resolved infrared spectra. *Biochemistry* **2000**, *39*, 13189−13200.
(18) Hessling, B.; Souvignier, G.; Gerwert, K. A model-independent approach to assigning bacteriorhodopsin's intramolecular reactions to photocycle intermediates. *Biophys. J.* **1993**, *65*, 1929−1941.
(19) Nagle, J. F.; Zimanyi, L.; Lanyi, J. K. Testing Br photocycle kinetics. *Biophys. J.* **1995**, *68*, 1490−1499.
(20) Tauler, R.; Kowalski, B.; Fleming, S. Multivariate curve resolution applied to spectral data from multiple run of an industrial process. *Anal. Chem.* **1993**, *65*, 2040−2047.
(21) Tauler, R. Multivariate curve resolution applied to second order data. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 133−146.
(22) Tauler, R. Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution. *J. Chemom.* **2001**, *15*, 627−646.
(23) Malinowski, E. R. *Factor Analysis in Chemistry*, 3rd ed.; Wiley Interscience: New York, 2002.
(24) Winding, W.; Guilment, J. Iterative self-modeling mixture analysis. *Anal. Chem.* **1991**, *63*, 1425−1432.
(25) Cuesta Sanchez, F.; Toft, J.; Van Den Bogaert, B.; Massart, D. L. Orthogonal projection approach applied to peak purity assessment. *Anal. Chem.* **1996**, *68*, 79−85.
(26) Keller, H. R.; Massart, D. L. Evolving factor analysis. *Chemom. Intell. Lab. Syst.* **1992**, *12*, 209−224.
(27) Keller, H. R.; Massart, D. L. Peak purity control in liquid chromatography with photodiode-array detection by a fixed size moving window evolving factor analysis. *Anal. Chim. Acta* **1991**, *246*, 379−390.
(28) Savitsky, A.; Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627−1639.
(29) Lawton, W. H.; Sylvestre, E. A. Self modelling curve resolution. *Technometrics* **1971**, *13*, 617−633.
(30) Bro, R.; De Jong, S. J. A fast non-negativity-constrained least squares algorithm. *J. Chemom.* **1997**, *11*, 393−401.
(31) Mendieta, J.; Diaz-Cruz, M. S.; Esteban, M.; Tauler, R. Multivariate curve resolution: A possible tool in the detection of intermediate structures in protein folding. *Biophys. J.* **1998**, *74*, 2876−2888.
(32) Vandengiste, B. G. M.; Massart, D. L.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J. *Handbook of chemometrics and qualimetrics*; Elsevier: Amsterdam, 1998; Chapter 34.
(33) Diewok, J.; De Juan, A.; Tauler, R.; Lendl, B. Quantitation of mixtures of diprotic organic acids by FT-IR flow titrations and multivariate curve resolution. *Appl. Spectrosc.* **2002**, *56*, 40−50.
(34) Tauler, R.; Smilde, A. K.; Henshaw, J. M.; Burgess, L. W.; Kowalsi, B. R. Multicomponent determination of chlorinated hydrocarbons using a reaction based chemical sensor. *Anal. Chem*. **1994**, *66*, 3337−3345.
(35) Engelhard, M.; Gewert, K.; Hess, B.; Kreutz, W.; Siebert, F. Light-driven protonation changes of internal aspartic acids of bacteriorhodopsin: An investigation by static and time-resolved infrared difference spectroscopy using [4-13C] aspartic acid labeled purple membrane. *Biochemistry* **1985**, *24*, 400−407.
(36) Maeda, A.; Balashov, S. P.; Lugtenburg, J.; Verhoeven, M. A.; Herzfeld, J.; Belenky, M.; Gennis, R. B.; Thomson, F. L.; Ebrey, T. G. Chromophore−protein−water interactions in the L intermediate of bacteriorhodopsin: FT-IR study of the photoreaction of L at 80 K. *Biochemistry* **2002**, *41*, 3803−3809.
(37) Heberle, J.; Fitter, J.; Sass, H. J.; Buldt, G. Bacteriorhodopsin: The functional details of a molecular machine are being resolved. *Biophys. Chem.* **2000**, *85*, 229−248.
(38) Friedrich, T.; Geibel, S.; Kalmbach, R.; Chizov, I.; Ataka, K.; Heberle, J.; Engelhard, M.; Bamberg, E. Proteorhodopsin is a light-driven proton pump with variable vectoriality. *J. Mol. Biol.* **2002**, *321*, 821−838.
(39) Chon, Y. S.; Kandori, H.; Sasaki, J.; Lanyi, J. K.; Needleman, R.; Maeda, A. Existence of two L photointermediates of halorhodopsin from *Halobacterium salinarum*, differing in their protein and water FT-IR bands. *Biochemistry* **1999**, *38*, 9449−9455.
(40) Rödig, C.; Siebert, F. Distortion of the L−M transition in the photocycle of the bacteriorhodopsin mutant D96N: A time-resolved step-scan FT-IR investigation. *FEBS Lett.* **1999**, *445*, 14−18.
(41) Betancourt, F. M. H.; Glaser, R. M. Chemical and physical evidence for multiple functional steps comprising the M state of the bacteriorhodopsin photocycle. *Biochim. Biophys. Acta* **2000**, *1460*, 106−118.
(42) Kumar, H. K. T.; Appaji Gowda, K. Absorption characteristics of bacteriorodopsin molecules. *Pramana−J. Phys.* **2000**, *3*, 447−452.
(43) Varo, G.; Lanyi, J. Distortion in the photocycle of bacteriorhodopsin at moderate dehydration. *Biophys. J.* **1991**, *39*, 313−322.
(44) Gossart, P.; Semmoud, A.; Ruckebusch, C.; Huvenne, J. P. Study of the interaction between humic acids and leads: Structural characterization by curve resolution of FT-IR spectra. *Anal. Chim. Acta* **2003**, *477*, 201−209.

CI034094I