Photoreaction of mutated LOV photoreceptor domains from *Chlamydomonas reinhardtii* with aliphatic mercaptans: Implications for the mechanism of wild type LOV.

Karin Lanzl, Madlene v. Sanden-Flohe, Roger-Jan Kutta and Bernhard Dick*
Institut für Physikalische und Theoretische Chemie,
Universität Regensburg, D 93040 Regensburg, Germany

ELECTRONIC SUPPLEMENTARY INFORMATION

Deconvolution of time sequences of spectra

We wish to analyze a sequence of spectra taken at various times. It is assumed that Beers law is valid, i.e, the data $A(\lambda, t)$ can be modelled as a linear combination of K spectra $S_k(\lambda)$ for each species k multiplied by a concentration-time profile $c_k(t)$,

$$A(\lambda, t) = \sum_{k}^{K} S_k(\lambda) c_k(t) \tag{1}$$

Measurements are performed at discrete wavelengths $\{\lambda_i, i = 1..N\}$ and discrete times $\{t_j, j = 1..M\}$. The data can be arranged into a $N \times M$ matrix with matrix elements $A_{ij} = A(\lambda_i, t_j)$. Equ. 1 translates then into the matrix equation

$$\mathbf{A} = \mathbf{S}\mathbf{C}^t \tag{2}$$

where

$$A_{ij} = A(\lambda_i, t_j) = \sum_{k}^{K} S_{ik} C^t_{kj} \tag{3}$$

In this expression, $\mathbf{S}$ is a $N \times K$ matrix whose columns represent the spectra of the individual species, and $\mathbf{C}$ is a $M \times K$ matrix whose columns represent the corresponding concentration time profiles. $\mathbf{C}^t$ denotes the transpose of this matrix. Our aim is to calculate $\mathbf{S}$ and $\mathbf{C}$ from the data matrix $\mathbf{A}$. We assume that $M < N$ and $K \ll M$. The case of $M > N$ can be treated in the same manner simply by exchanging the role of the matrices $\mathbf{S}$ and $\mathbf{C}$, i.e., replacing

---

[*] e-mail: Bernhard.Dick@chemie.uni-regensburg.de

$\mathbf{A}$ with the transpose matrix $\mathbf{A}^t$. In the following we represent the l-th column vector in a matrix $\mathbf{B}$ by $\mathbf{b}_l$. When $\mathbf{A}$ is exactly described by equ. 3, the rank of $\mathbf{A}$ has the upper limit K, unless the spectra and/or concentration-time profiles are linearly dependent. In a realistic measurement, a data matrix $\mathbf{D}$ will be obtained that differs from $\mathbf{A}$ by the presence of noise $\mathbf{R}$, i.e. $\mathbf{D} = \mathbf{A} + \mathbf{R}$. Our aim is to calculate the number of components K, their spectra $\mathbf{S}$ and concentration-time profiles $\mathbf{C}$ from such a noisy data set $\mathbf{D}$. The mathematical algorithm of singular value decomposition (SVD) can be used to split the matrix $\mathbf{D}$ into the product of three matrices $\mathbf{U}$, $\mathbf{W}$, and $\mathbf{V}$ in the following way

$$\mathbf{D} = \mathbf{U}\mathbf{W}\mathbf{V}^t \tag{4}$$

Where $\mathbf{U}$ is an $N \times M$ matrix, $\mathbf{W}$ a $M \times M$ diagonal matrix with positive diagonal elements w_j sorted in descending order, and $\mathbf{V}$ is a $M \times M$ matrix. The columns of $\mathbf{U}$ are mutually orthogonal, as are the columns of $\mathbf{V}$.

$$\sum_k^N U_{ki}U_{kj} = \delta_{ij} \tag{5}$$

$$\sum_k^N V_{ki}V_{kj} = \delta_{ij} \tag{6}$$

By using the first L of these vectors, a reconstruction of the data can be made that converges to the original data set when L approaches M.

$$D_{ij}^{(L)} = \sum_k^L U_{ik} w_k V_{kj}^t \tag{7}$$

It can be shown that for each L the standard deviation between the original data and the contribution of the first L principal components $\sigma^{(L)}$ defined by

$$\sigma^{(L)} = \left(\sum_{ij} \left(D_{ij} - D_{ij}^{(L)} \right)^2 / (NM) \right)^{1/2} \tag{8}$$

is the smallest standard deviation possible with a representation of the data by any L pairs of spectra and concentration-time profiles. Hence, a plot of $\sigma^{(L)}$ versus L can provide a criterion to estimate the rank of the significant part of the data: Each time a significant contribution is added to the reconstruction of the data, $\sigma^{(L)}$ will decrease by a significant amount. The first component that does not improve $\sigma^{(L)}$ can then be regarded as noise. The number L of principal components that is needed to represent the data within the accuracy of the noise is the effective rank of the data matrix and can be interpreted as the number of linearly independent spectral components K in equ. 1.

For a given number K of components a set of spectra and concentration-time profiles can be defined by

$$\mathbf{S} = \mathbf{U}\mathbf{X} \; ; \; \mathbf{C}^t = \mathbf{X}^{-1}\mathbf{W}^{(K)}\mathbf{V}^t \tag{9}$$

where $\mathbf{W}^{(K)}$ is the matrix $\mathbf{W}$ with all diagonal elements w_k with $k > K$ set to zero, and $\mathbf{X}$ is a non-singular $K \times K$ matrix. The product of these spectra and concentration-time profiles yields the best representation possible with the given number of components, i.e. $\mathbf{SC}^t = \mathbf{D}^{(K)}$. Obviously, this representation does not depend on the choice of the matrix elements of $\mathbf{X}$, which can hence be chosen to fulfill other conditions.

In all our experiments, initially only one compound is present. Hence the first column of $\mathbf{D}^{(K)}$ should be equal to the spectrum of this species. Comparison of coefficients in

$$\mathbf{d}_1^{(K)} = \sum_k^K \mathbf{u}_k w_k V_{k1}^t = \mathbf{s}_1 = \sum_k^K \mathbf{u}_k X_{k1} \tag{10}$$

yields

$$X_{k1} = w_k V_{k1}^t \tag{11}$$

In the case of two components ($K = 2$), two further conditions are required. An isosbestic point at wavelength λ_j means $S_{j1} = S_{j2}$, leading to

$$U_{j1}\left(X_{11} - X_{12}\right) = U_{j2}\left(X_{22} - X_{21}\right) \tag{12}$$

If it is known a-priori that spectrum $\mathbf{s}_2$ vanishes at a wavelength λ_l at which spectrum $\mathbf{s}_1$ does not vanish, the additional equation

$$S_{l2} = U_{l1}X_{12} + U_{l2}X_{22} = 0 \tag{13}$$

leads to a unique solution for all matrix elements of $\mathbf{X}$. If this is not the case, the ratio X_{12}/X_{22} may be adjusted manually in such a way that the characteristic bands of species 1 do not appear in spectrum 2. In the case of anaerobic photoreaction of LOV1-C57G with CH_3SH three components ($K = 3$) were detected. Hence, nine matrix elements of $\mathbf{X}$ need to be determined, three of which are already given by equ. 11. The postulate that the adduct does neither absorb at 615 nm nor at 480 nm yield two further conditions and results in a unique solution for the spectrum of the adduct. The postulate that the third spectrum (i.e. that of the radical) should not contain the characteristic bands of FMN or the adduct yields two further conditions. The spectrum thus obtained is identical to those found in the two-component systems with EDTA or βME as reducing agent. The two last conditions that suffice to make the solution unique are (i) the use of the isosbestic point between the spectra of FMN and the FMNH radical at 490.5 nm, and (ii) the postulate that the sum of the three concentrations must not increase in time but should be either constant or monotonously decaying.

There exists an independent method that checks this last postulate. The spectra of all three species have been determined in other experiments. Hence the equation $\mathbf{SC}^t = \mathbf{D}$ can be regarded as an overdetermined system of linear equations, with the components of $\mathbf{C}$ being the unknowns. The least squares solution of this problem

$$\left|\mathbf{SC}^t - \mathbf{D}\right|^2 = \mathrm{Min} \tag{14}$$

is given by

$$\mathbf{C}^t = (\mathbf{S}^t\mathbf{S})^{-1}\mathbf{S}^t\mathbf{D} \tag{15}$$

In this expression, the pseudo-inverse $(\mathbf{S}^t\mathbf{S})^{-1}\mathbf{S}^t$ of the rectangular matrix $\mathbf{S}$ is conveniently calculated by SVD.