BIOCHEMISTRY
including biophysical chemistry & molecular biology
Biochemistry, 1997, 36(9), 2439-2449, DOI:10.1021/bi9617419

1 Supplementary material to "Multichannel analysis of single turnover kinetics of cytochrome aa_3 reduction of O_2"

The experimental data that trace a chemical reaction are stored in a matrix (call it A), each column of which is a spectrum taken at some time t_j, and each row of which is a time course measured at some fixed wavelength w_i. Typically, there are tens or hundreds of times and wavelengths. Let $A_{i,j}$ denote that element of A in the ith row and jth column. Then $A_{i,j}$ is the spectral value measured at wavelength w_i and time t_j. The SVD-based matrix least-squares analysis of A attempts to factor A into two matrices D and Y such that:

$$A = DY + E$$

where the columns of D contain the spectra of individual chromophores, the rows of Y contain the kinetic (appearance-disappearance) curves of those chromophores, and E is a matrix of residuals $A - DY$, often interpreted as noise. The columns of D may be either absolute or difference spectra, plus a reference spectrum, usually representing either the initial or final total absolute spectrum. The fitting procedure minimizes the sum of squares of the elements of E. See Hendler and Shrager (1994) for details.

1.1 Working with exponentials

The models we will consider are of the linear compartmental type, which are known to produce time courses that are described by sums of exponentials (except in special cases that will not be met here). Therefore, regardless of the particular model we choose, we can start the modeling process with the following SVD-based decomposition:

$$A = CX + E$$

where X is a matrix, each row of which is a single exponential tabulated versus time, with a final row of ones to account for an unchanging background. C is a matrix, each column of which is the difference spectrum associated with the corresponding row of X, the final column being the absolute spectrum one would observe at infinite time. That is, the first columns of C are differences with respect to the final column. The SVD procedure provides clues as to the required number of exponentials, and the fitting procedure determines their decay rates $-k_i$, so that the elements of X are:

$$x_{i,j} = \exp(-k_i t_j)$$

The matrix C is produced as a by-product of the fitting procedure, but for completeness, C may be computed from A and X by the formula $C = AX^+$ where X^+ is the Moore-Penrose pseudoinverse of X.

Some of the exponentials in the rows of X may be spurious, in the sense that while they are required to accomplish the fit, they are not part of the physical process of interest (see text). These exponentials must be included in the fitting process. To omit them would cause

severe bias in the remaining exponentials. However, once the fit is accomplished, those parts of X and C that do not correspond to the model may be ignored. From this point onward, we will refer to edited versions of X and C, in which the final row of X and any rows containing spurious exponentials have been removed along with the corresponding columns of C.

If the chemical components of the system were in fact decaying independently and exponentially, then rows of X would contain their time courses, and C would contain their difference spectra, i.e., $Y = X$ and $D = C$ would be the desired decomposition. However, in the models of interest, the components are coupled, in that some molecules will pass through intermediate states before reaching the final state. This implies that some of the time courses will be sums of two or more exponentials, so that an additional model-dependent calculation is required to produce the proper sums. In brief, if the model differential equations and initial conditions can be expressed in matrix terms as

$$dy/dt \;=\; J\; y(t), \quad y(0) = y_0$$

where J is an n-by-n matrix, n being the number of species, $y(t)$ is the n-vector of appearance-disappearance values of the various species with elements $y_i(t)$, and dy/dt is the n-vector of derivatives $dy_i(t)/dt$. Then with certain exceptions which will not be met here, the solution curves $y_i(t)$ will be sums of exponentials. If the relevant exponentials are tabulated in the rows of X, then there is a mixing matrix M relating $y(t)$ to X by:

$$Y \;=\; MX$$

where the matrix Y contains tabulated solution curves $y_i(t)$ in its rows. The associated matrix of difference spectra can be computed as:

$$D \;=\; CM^{-1}$$

Note that the relation $DY = CM^{-1}MX = CX$ holds, since the product $M^{-1}M$ is the identity matrix. Thus, the model product DY is independent of the particular differential equations and their associated M matrix, implying that any model with the correct rates will fit as well as any other, a fact which precludes using goodness of fit as a basis for deciding between such models, In summary, the matrices X, C, M, Y, and D are computed from the data A, the differential equations described by J, and the initial conditions y_0 as follows, using commonly available software:

1. Deduce X and C from A using SVD-based least-squares.

2. Compute the eigensystem of J, consisting of a diagonal matrix K with eigenvalues $-k_i$ in the diagonal elements (the exponentials we use are decaying, yet we refer to k_i as positive, hence the minus signs), and a matrix W consisting of the corresponding column eigenvectors.

3. Form Z_0, a diagonal matrix whose diagonal elements are those of the vector $W^{-1}y_0$.

4. Finally, $M = WZ_0$, $Y = MX$, and $D = CM^{-1}$.

SVD-based least squares for computing X and C is explained in the references. The numerical procedure for computing M, Y, and D, coded in MATLAB, is:

```
% J, y0, C, and X are given.
% The row of ones and any spurious exponentials in X have been
% deleted at this point, along with corresponding columns of C.
[ W, K ] = eig( J );   % W now contains column eigenvectors.
Z0 = diag(W\y0);   % Z0 is diag. with diag. elements (W^-1)y0.
M = W * Z0;    % M = Mixing matrix.
Y = M * X;     % Y = Time courses for all species in the model.
D = C / M;     % D = Spectra for all species in the model.
```

The matrix M can be computed numerically from the above procedure, or it can be produced analytically e.g. using symbolic mathematics programs like Maple and Mathematica, both of which can provide analytic solutions to linear ordinary differential equations with initial conditions.

For information about the programs Maple, Mathematica, and MATLAB, contact the following:

Maple:

Waterloo Maple Software
160 Columbia Street West
Waterloo, Ontario,
Canada N2L 3L3
Tel: (519)747-2373
FAX: (519)747-5284
email: wmsi@daisy.uwaterloo.ca

Mathematica:

Wolfram Research, Inc.
100 Trade Center Drive
Champaign IL 61820-7237
Tel: (217)398-0700
FAX: (217)398-0747
email: info@wri.com

MATLAB:

The MathWorks, Inc.
24 Prime Park Way
Natick MA 01760
Tel: (508)647-7000
FAX: (508)647-7101
email: info@mathworks.com

The presence of four consistent exponential rates in our data suggests a model with four compartments, call them y_1, y_2, y_3, and y_4. (Actually, our models will imply a fifth state, but it is the final state, a sink that does not contribute to the kinetics.) Of the many models that can be generated from four rates, we will consider only two, which will be called the sequential model and the branched model.

1.2 The sequential model

The sequential model is frequently assumed, because it is conceptually simple:

$$y_1 \xrightarrow{k_1} y_2 \xrightarrow{k_2} y_3 \xrightarrow{k_3} y_4 \xrightarrow{k_4}$$

The associated differential equations and initial conditions are:

$$\begin{array}{lcllcl} dy_1/dt & = & -k_1 y_1, & y_1(0) & = & y_T \\ dy_2/dt & = & k_1 y_1 - k_2 y_2, & y_2(0) & = & 0 \\ dy_3/dt & = & k_2 y_2 - k_3 y_3, & y_3(0) & = & 0 \\ dy_4/dt & = & k_3 y_3 - k_4 y_4, & y_4(0) & = & 0 \end{array}$$

where y_T is the total concentration of y, $y_i(0)$ is the initial value of y_i at $t = 0$, and y_i and dy_i/dt without explicit arguments are both understood to be evaluated at some common time t. In matrix terms:

$$dy/dt \;=\; Jy, \quad y(0) \;=\; [y_T, 0, 0, 0]^T$$

where:

$$y \;=\; [\, y_1,\; y_2,\; y_3,\; y_4 \,]^T$$

$$dy/dt \;=\; [\, dy_1/dt,\; dy_2/dt,\; dy_3/dt,\; dy_4/dt \,]^T$$

$$\text{and } J = \begin{bmatrix} -k_1 & 0 & 0 & 0 \\ k_1 & -k_2 & 0 & 0 \\ 0 & k_2 & -k_3 & 0 \\ 0 & 0 & k_3 & -k_4 \end{bmatrix}$$

J is a lower-triangular matrix (i.e., all zeros above the main diagonal), so that the four exponential rates are on the main diagonal. The matrix M, computed by the procedure described above, is:

$$M = y_T \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{k_1}{k_2-k_1} & \frac{k_1}{k_1-k_2} & 0 & 0 \\ \frac{k_1 k_2}{(k_2-k_1)(k_3-k_1)} & \frac{k_1 k_2}{(k_1-k_2)(k_3-k_2)} & \frac{k_1 k_2}{(k_1-k_3)(k_2-k_3)} & 0 \\ \frac{k_1 k_2 k_3}{(k_2-k_1)(k_3-k_1)(k_4-k_1)} & \frac{k_1 k_2 k_3}{(k_1-k_2)(k_3-k_2)(k_4-k_2)} & \frac{k_1 k_2 k_3}{(k_1-k_3)(k_2-k_3)(k_4-k_3)} & \frac{k_1 k_2 k_3}{(k_1-k_4)(k_2-k_4)(k_3-k_4)} \end{bmatrix}$$

Note that the matrices J and M are completely specified by the rates k_i that were determined in the fit $A = CX$. This will not be the case in the branched model.

1.3 The branched model

A molecule in the branched model will follow one of two paths to the final state:

$$y_1 \xrightarrow{k_1} y_2 \xrightarrow{k_2 f_3} y_3 \xrightarrow{k_3}$$

$$y_1 \xrightarrow{k_1} y_2 \xrightarrow{k_2 f_4} y_4 \xrightarrow{k_4}$$

A fraction f_3 of the total material y_T will pass through the state y_3, while the remaining material, fraction f_4 of y_T, will pass through y_4, where $f_3 + f_4 = 1$. The matrix form of the differential equations is similar to that of the sequential model, except for the matrix J:

$$J = \begin{bmatrix} -k_1 & 0 & 0 & 0 \\ k_1 & -k_2 & 0 & 0 \\ 0 & k_2 f_3 & -k_3 & 0 \\ 0 & k_2 f_4 & 0 & -k_4 \end{bmatrix}$$

The corresponding M matrix is:

$$M = y_T \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{k_1}{k_2 - k_1} & \frac{k_1}{k_1 - k_2} & 0 & 0 \\ \frac{k_1 k_2 f_3}{(k_2 - k_1)(k_3 - k_1)} & \frac{k_1 k_2 f_3}{(k_1 - k_2)(k_3 - k_2)} & \frac{k_1 k_2 f_3}{(k_1 - k_3)(k_2 - k_3)} & 0 \\ \frac{k_1 k_2 f_4}{(k_2 - k_1)(k_4 - k_1)} & \frac{k_1 k_2 f_4}{(k_1 - k_2)(k_4 - k_2)} & 0 & \frac{k_1 k_2 f_4}{(k_1 - k_4)(k_2 - k_4)} \end{bmatrix}$$

As in the sequential model, the four exponential rates are seen on the main diagonal of J, and they are the same rates in both cases. As shown above, the fits of two such models to the data are identical, so that the models cannot be distinguished on that basis.

The branched model has an added difficulty, in that J and M contain the fractions f_3 and f_4, which cannot be deduced from the fitted k's, except for the limitation that neither fraction may be zero. (If either fraction were zero, the branched model would be reduced to a three-exponential model, because either y_3 or y_4 would be eliminated.) If other experimental evidence is not available for the values of f_3 and f_4, we may of course postulate some values, say, $f_3 = f_4 = 1/2$, to see what the result would be. The information one gets from the resulting D and Y matrices is as follows:

1. As in the sequential model, each jth column of D is the difference spectum of y_j with respect to the absolute spectrum at infinite time. The first two spectra (D columns 1 and 2) and their time courses (Y rows 1 and 2) are identical to those of the sequential case in both scale and shape.

2. The third and fourth spectra and their time courses differ from those of the sequential case. Although their shapes are unambiguous, their scale factors will depend on the choice of f_3, with $f_4 = 1 - f_3$. That is, Y row 3 is proportional to f_3, D column 3 is proportional to $1/f_3$, and similarly for Y row 4, D column 4, and f_4.

3. It follows that, unlike the sequential case, the shapes of sequential differences like the difference spectrum of $y_2 - y_3$, computed as

$$(D \text{ column } 2) - (D \text{ column } 3)$$

will depend on f_3 because D column 3 is subject to the scale factor $1/f_3$ whereas D column 2 is not. Therefore, while D columns 3 and 4 vary only in scale with the choice of f_3, the same cannot be said for differences between columns 3 and 4 and the other columns of D.

1.4 Observations on the parallel model

The parallel model, in which the four rate constants represent four independent decays, e.g., washout of a mixture of four non-interacting substances at distinct rates, is the "natural" exponential model, because the matrices C and X are already the desired description of the process. The matrix M is diagonal, containing scale factors, but no coupling. So it is very tempting to use the parallel model as an approximation even in cases where it does not apply. Indeed, when the time constants are widely separated, one often gets rather good sequential difference spectra directly from C in this manner, without committing oneself to any coupling scheme. But this lack of commitment reveals nothing about the mechanism beyond the fact that the underlying differential equations are linear or nearly so. In addition, the quality of the resulting spectra are not guaranteed *a priori*. The factors affecting the quality of such an approximation are: 1) the model being approximated, 2) the separation of the time constants, 3) the relative amplitudes of the spectra, and 4) the degree of linear independence of the spectra. Unless the separation of the time constants is, say, several orders of magnitude, the best way to be sure that one is getting an adequate approximation is to compare the spectra yielded by the parallel model to those of the model in question in some typical cases. Any assurance of a good approximation must involve a knowledge of the model one is approximating. In fact, without a model, one does not even know what difference spectra are being approximated by the spectra in C. Therefore, one must decide which model(s) to examine, and the avoidance of commitment is largely illusory. Considering that there is very little excess computation for small coupled models like those in this paper, it is best to use the proper model, and avoid the issues of parallel approximation entirely.