ELSEVIER

Chemometrics and Intelligent Laboratory Systems 59 (2001) 19–31

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemometrics

Multivariate curve resolution with non-linear fitting of kinetic profiles

Ernst Bezemer, Sarah C. Rutan *

Department of Chemistry, Virginia Commonwealth University, PO Box 842006, Richmond, VA 23284-2006, USA

Received 23 April 2001; received in revised form 15 June 2001; accepted 15 June 2001

Abstract

This paper describes the incorporation of a hard modeling step based on a kinetic model, into a soft modeling (multivariate curve resolution) technique. The soft modeling technique allows for the determination of the retention and spectral profiles from overlapped components while the hard modeling step allows for the simultaneous prediction of the rate constants of the various steps in the reaction pathway. The program uses standard MATLAB® functions for determining the solutions of the differential equations as well as for finding the optimal rate constants to describe the kinetic profiles. The kinetic model is entered by a set of command line parameters and can describe any order chemical reaction with multiple reaction pathways. This paper uses simulated first- and second-order reaction data as well as real data to characterize the performance of the program. The algorithm is able to resolve overlapped retention and spectral profiles and predict the rate constants for the reaction.

Keywords: Multivariate curve resolution; Non-linear fitting; Kinetic model

1. Introduction

When chemical species have overlapping responses, soft modeling chemometric methods may be used to isolate the individual species-dependent responses. Chemometric techniques such as multivariate curve resolution-alternating least squares (MCR-ALS) are very flexible and let the operator apply constraints that prevent the calculation from converging to a chemically invalid solution [1]. But in order to determine the kinetic parameters of the chemical system, the kinetic profiles resolved by this technique have to be fit to a chemical model separately [2,3].

The incorporation of a hard modeling step into a soft modeling method can decrease the rotational ambiguity of the soft modeling solution. Some researchers have done exactly this, by incorporating a chemical-kinetic model into their soft modeling methods. De Juan et al. [4] have used the analytical solution of the corresponding differential equation while applying MCR-ALS to a three-way data set in order to estimate rate constants for consecutive first-order reactions. Bijlsma et al. [5–8] also used the analytical solution of the corresponding differential equation for consecutive first- and second-order reactions in conjunction with the generalized rank annihilation method and with PARAFAC. This is an ex-

* Corresponding author. Tel.: +1-804-828-7517; fax: +1-804-828-8599.

0169-7439/01/$ - see front matter
PII: S0169-7439(01)00141-1

cellent and fast approach when the solution of the differential equations is known or when the kinetic profiles can be described by a sum of exponential functions [9 10]. However, this is not always possible for some reaction pathways. Haario and Taavitsainen [11] applied a numerical differential equation solver to determine the rate constants during indirect calibration methods such as principal component regression and partial least squares, and direct calibration methods such as generalized ridge regression. However, all these methods are fixed for a single reaction pathway. When a different reaction needs to be investigated, the program has to be partly rewritten.

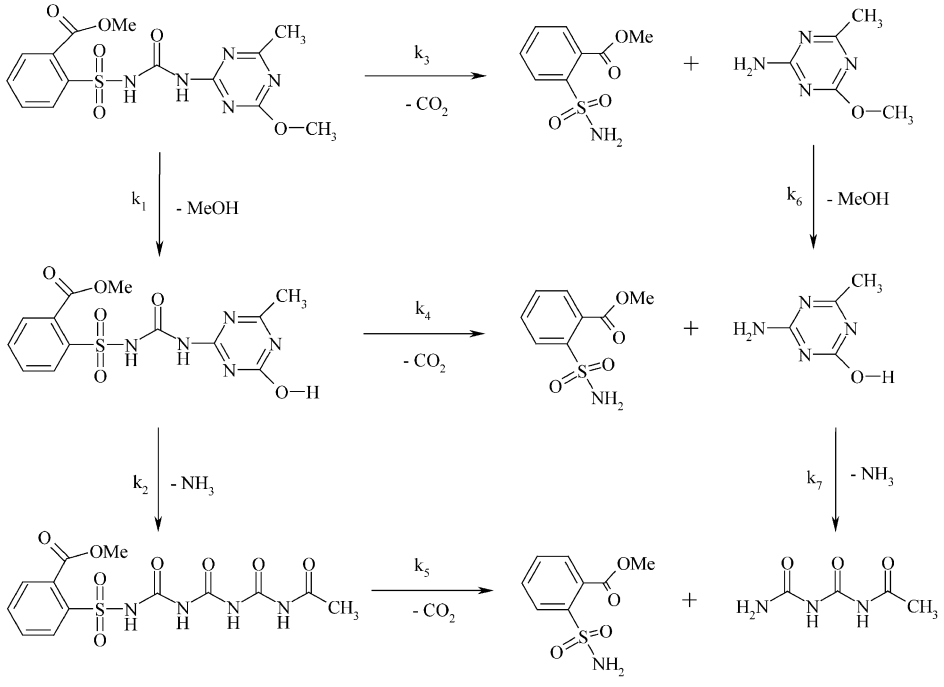
The method described in this paper (ALS-DQ) uses the standard MATLAB® ordinary differential equation solvers rather than the analytical solution of the corresponding differential equations. This paper also illustrates a convenient set of parameters to express the differential equations in such a way that it can be used by a differential equation solver without the need for rewriting part of the program when changing between different reaction models. The focus here is on fitting liquid chromatography-diode array detection (LC-DAD) data, but the curve resolution approach described here can be applied to a wide range of data types. Moreover, the kinetic fitting algorithm is a stand-alone subroutine that can be used to fit kinetic profiles from any type of experiment.

This paper examines the behavior of this non-linear kinetic fitting algorithm incorporated into an MCR method by analyzing synthetic data and the chemical degradation reaction of the herbicide Ally. The Ally degradation pathway is depicted in Scheme 1. The ability of the ALS-DQ program to fit the data representing this complex pathway demonstrates the applicability of this method for handling diverse kinetic models.

2. Theory

2.1. MCR-ALS

The MCR-ALS algorithm used here has been extensively described in an earlier paper and the gen-



Scheme 1. The degradation of Ally.

eral approach will be summarized here [12]. In this approach, a two-way data set can be described as the inner product of the concentration profiles and the pure component spectra according to Eq. (1) [13].

$$\mathbf{D} = \mathbf{C} \cdot \mathbf{S}^T + \mathbf{E} \tag{1}$$

where $\mathbf{D}$ $(T \times S)$ is the data matrix, $\mathbf{C}$ $(T \times N)$ contains the concentration profiles for each of the pure components, $\mathbf{S}$ $(S \times N)$ contains the pure component spectra and $\mathbf{E}$ $(T \times S)$ is the error matrix. The variables T, S and N represent the number of chromatographic time points, the number of measured wavelengths and the number of components, respectively.

With one set of the pure component profiles estimated initially (the concentration profiles in this case), the other set of profiles can be calculated using Eq. (2).

$$\mathbf{S} = (\mathbf{C}^T \cdot \mathbf{C})^{-1} \cdot \mathbf{C}^T \cdot \mathbf{D} \tag{2}$$

This process can be constrained and iterated upon until a set minimal improvement of the fit error is reached [14].

When analyzing three-way data using the generalized ALS program, the data is described by Eq. (3) [12].

$$\mathbb{D} = (\mathbb{R} \otimes \mathbf{K}) \otimes \mathbb{S} + \mathbb{E} \tag{3}$$

where $\mathbb{D}$ $(T \times S \times K)$ is the data tensor, $\mathbb{E}$ $(T \times S \times K)$ is the error, $\mathbb{R}$ $(T \times N \times K)$ represents the retention profiles, $\mathbb{S}$ $(S \times N \times K)$ contains the spectral profiles and $\mathbf{K}$ $(K \times N)$ holds the kinetic profiles. K is the number of kinetic time points.

The resolution occurs analogously to the two-way approach by multiplying the data set by the pseudo inverse of $(\mathbb{R} \otimes \mathbf{K})$ and $\mathbb{S}$ in sequence. $\mathbb{R} \otimes \mathbf{K}$ are subsequently separated by normalization.

The solutions are limited by chemically relevant constraints, such as non-negativity and unimodality and the kinetic model constraint. The kinetic profiles are fit to the chemical model and rate constants are obtained that can be used to simulate updated kinetic profiles. These new kinetic profiles result in different retention- and spectral profiles during the next step of the iteration and help the program to converge to a chemically sound solution.

2.2. Parameters describing a kinetic model

The method described in this paper uses a numerical approach to solve the differential equations by using an ordinary differential equation solver (ODE23) that comes with MATLAB® [15]. It requires simple parameters, not unlike state-space models, to describe the reaction pathway. This results in the flexibility to fit any chemical reaction pathway, including those with non-first-order reactions.

For a reaction pathway with N rate constants and M components, the differential equation describing the kinetic profile for each component can be expressed by Eq. (4).

$$\frac{\mathrm{d}[X_j]}{\mathrm{d}t} = \sum_{i=1}^{N} k_i \left(\prod_{j=1}^{M} [X_j]^{o_{ji}} \right) \tag{4}$$

where k_i is the rate constant and o_{ji} is the order of the reaction in species X_j in the ith step of the overall reaction pathway.

Consider for example the following hypothetical reaction: $A + B \rightarrow^{k_1} C$ and $2C \rightarrow^{k_2} D$ where the rate equations for the various components can be written as Eq. (5).

$$\begin{aligned}
\frac{\mathrm{d}[A]}{\mathrm{d}t} &= -k_1[A][B] \\
&= -\mathbf{1} \times k_1[A]^1[B]^1[C]^0[D]^0 \\
&\quad + \mathbf{0} \times k_2[A]^0[B]^0[C]^2[D]^0 \\
\frac{\mathrm{d}[B]}{\mathrm{d}t} &= -k_1[A][B] \\
&= -\mathbf{1} \times k_1[A]^1[B]^1[C]^0[D]^0 \\
&\quad + \mathbf{0} \times k_2[A]^0[B]^0[C]^2[D]^0 \\
\frac{\mathrm{d}[C]}{\mathrm{d}t} &= k_1[A][B] - k_2[C]^2 \\
&= +\mathbf{1} \times k_1[A]^1[B]^1[C]^0[D]^0 \\
&\quad - \mathbf{1} \times k_2[A]^0[B]^0[C]^2[D]^0 \\
\frac{\mathrm{d}[D]}{\mathrm{d}t} &= \frac{\mathbf{1}}{\mathbf{2}} k_2[C]^2 = +\mathbf{0} \times k_1[A]^1[B]^1[C]^0[D]^0 \\
&\quad + \frac{\mathbf{1}}{\mathbf{2}} \times k_2[A]^0[B]^0[C]^2[D]^0
\end{aligned} \tag{5}$$

A closer examination reveals that each rate constant is multiplied by the same component concentrations in all cases. This understanding is used in this algorithm in order to describe the chemical model. Each rate constant has a corresponding vector indicating the order of the components in each reactant. For k_1 this would be [1 1 0 0] for $[A]$, $[B]$, $[C]$ and $[D]$, respectively, while for k_2 this vector is [0 0 2 0]. The combination of these two vectors constitutes the reaction order matrix, **O**.

The next step is to indicate which rate constants are used to model each time dependent concentration for each species. These are the coefficients shown in bold in Eq. (5). For $\mathrm{d}[A]/\mathrm{d}t$ this would be $[-1\ 0]$ for k_1 and k_2, respectively. Each row in the matrix indicates a time-dependent concentration for a component and each column corresponds to a rate constant, resulting in the reaction pathway matrix,

$$\mathbf{R}: \begin{bmatrix} -1 & +0 \\ -1 & +0 \\ +1 & -1 \\ +0 & +\frac{1}{2} \end{bmatrix}$$

When **O** is the reaction order matrix, o_{ji} an element of this matrix, k_j is the rate constant, M is the number of reacting components and $[X_i]$ is the concentration of the ith component, for each reaction step, the rate law for each step in the reaction mechanism can be calculated according to Eq. (6).

$$p_j = k_j \prod_{i=1}^{M} \left([X_i]^{o_{ji}} \right) \tag{6}$$

Then when **R** is the reaction pathway matrix, r_{mj} is an element in this matrix, N is the number of rate constants and m is the component number, it follows that each time derivative of each component concentration as shown in Eq. (4) is given by Eq. (7).

$$\frac{[X_m]}{\mathrm{d}t} = \sum_{j=1}^{N} \left(r_{mj} \times p_j \right) \tag{7}$$

Eqs. (6) and (7) are converted into a Matlab function ('kinfun') that requires the input parameters of the reaction order and reaction pathway matrix as well as a vector containing the rate constants. The output is a vector containing the derivatives for each component. The ODE function call is given by the following command line

$$[t, y] = \mathrm{ODE23}(@\mathrm{kinfun}, \mathrm{reaction\_times}, \mathrm{initial\_conc}, [\,], \mathrm{model});$$

where y contains the concentration profiles for each reaction time in t, kinfun is the name of the function described above, reaction_times is a vector containing the various time points, initial_conc is a vector containing the starting concentrations for each of the components and model is a structure array containing the reaction order and pathway matrix and the vector with the rate constants.

The kinetic profiles of all four species in this example are depicted in Fig. 1 for the case where all the rate constants are 0.5. When equal starting concentrations are used, the concentration profiles of A and B will be identical in this kinetic model and are overlapped in Fig. 1.

2.3. Optimization of the fit

A second function (kinetic_fit) is used to find the rate constants that fit the kinetic profiles using the multi-parameter minimum search (FminSearch) in MATLAB® based on the Nelder–Mead algorithm [16]. This function uses the modeled rate constants, simulates the kinetic profiles and calculates the difference between the simulated kinetic profile and the

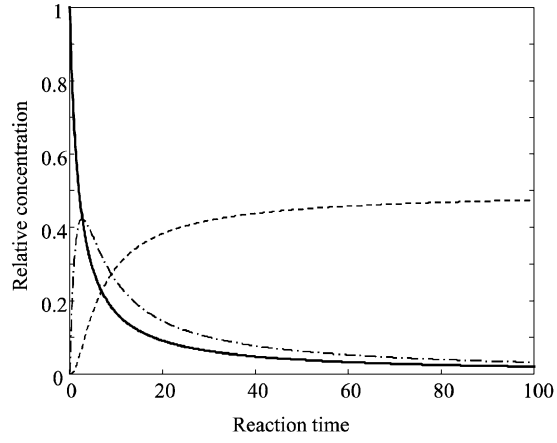


Fig. 1. Simulated kinetic profile for consecutive second order reactions for A (—), B (· · ·), C (·–·–) and D (–––).

chemometrically resolved kinetic profiles. The search algorithm finds the rate constants that result in the best fit between the simulated profiles and the resolved kinetic profiles during each iteration of the ALS-DQ procedure. The FminSearch function call is given by the following command line:

```
opt_rates = FminSearch(@kinetic_fit, old_rates, [],
                       k_profiles, model);
```

where opt_rates is a vector with the optimal rate constants, kinetic_fit is the name of the function described above, old_rates is a vector with the length equal to the number of rate constants that contain the initial estimates for the rate constants, k_profiles contains the matrix with the kinetic profiles resolved by ALS and model is a structure array containing the reaction order and pathway matrix and a vector containing the reaction times.

3. Experimental

3.1. Generation of synthetic data

All the simulated data were generated in a similar fashion to that described previously [12]. Components with equal height and Gaussian peak shapes were used for the retention profiles and are shown in Fig. 2a. The spectral profiles were generated by adding a Gaussian function to an exponential decay function and are shown in Fig. 2b. The kinetic profiles were generated by numerically solving the differential equations for the specific reaction pathway. A summary of the simulated data sets is given in Table 1. Table 1 also shows the parameters that were used to describe the reaction pathways as well as the parameters to describe the Ally degradation pathway.

For the simple first- and second-order reactions, the data set was generated with either 1% or 7% normally distributed random noise, 5, 10 and 15 reaction time points and to 99%, 75% and 50% completion of the reaction as indicated by the product concentration. The rate constant was 0.5 for each experiment.

For the multiple step reactions, the data were generated with 0.1% and 0.5% normally distributed noise, 5, 10 and 15 reaction time points and 99%, 75% and 50% completion of the reaction as indicated by the product concentration. For the consecutive first-order reactions, the second rate constant was varied between 0.25, 0.5 and 1.0 while the first was set to be constant at 0.5. For the consecutive first-order reaction with a reversible side reaction, k_1 and k_2 were set constant at 0.5, while k_3 was varied between 0.25, 0.5 and 1 and the equilibrium constant was either 5 or 0.2.

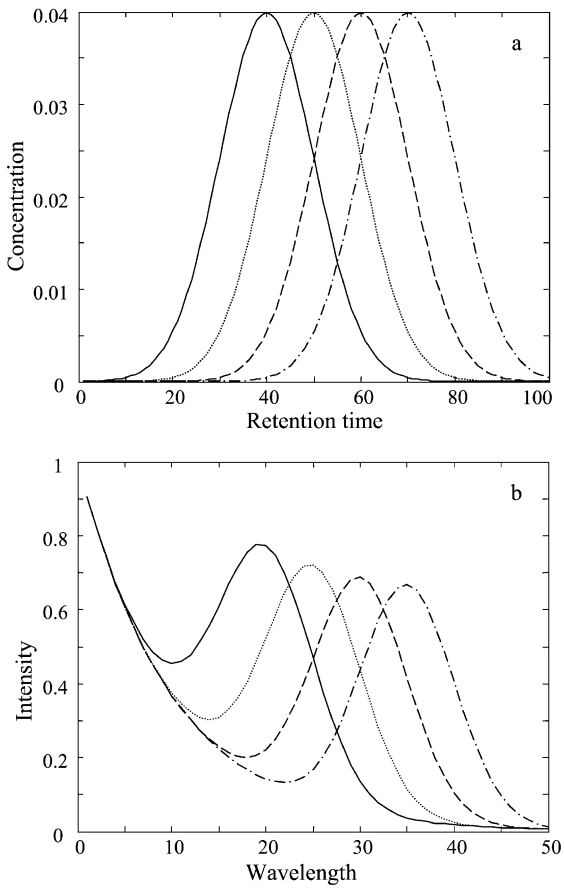


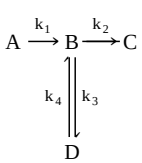
Fig. 2. Simulated retention (a) and spectral (b) profiles for components A (—), B (· · ·), C (·–·–) and D (–––).

The ALS fit quality was calculated according to Eq. (8).

$$\%\text{lack of fit} = 100\sqrt{\frac{\sum_{t,s,k}\left(d_{t,s,k}^{\text{optimal}} - d_{t,s,k}^{\text{input}}\right)^2}{\sum_{t,s,k}\left(d_{t,s,k}^{\text{input}}\right)^2}} \qquad (8)$$

The actual pattern of the noise is different for every simulation. Therefore, all simulations and the

Table 1
Different chemical models used to characterize the ALS-DQ program

Chemical reaction	Differential equations	Reaction pathway matrix	Reaction order matrix
$A \xrightarrow{k_1} B$	$\frac{dA}{dt} = -k_1[A]$	$\begin{bmatrix} -1 \\ 1 \end{bmatrix}$	$[1 \quad 0]$
$2A \xrightarrow{k_1} B$	$\frac{dA}{dt} = -k_1[A]^2$	$\begin{bmatrix} -1 \\ \frac{1}{2} \end{bmatrix}$	$[2 \quad 0]$
	$\frac{dB}{dt} = +\frac{1}{2}k_1[A]^2$		
$A \xrightarrow{k_1} B \xrightarrow{k_2} C$	$\frac{dA}{dt} = -k_1[A]$	$\begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$
	$\frac{dB}{dt} = +k_1[A] - k_2[B]$		
	$\frac{dC}{dt} = k_2[B]$		
	$\frac{d[A]}{dt} = -k_1[A]$	$\begin{bmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
	$\frac{d[B]}{dt} = k_1[A] - k_2[B] - k_3[B] + k_4[D]$		
	$\frac{d[C]}{dt} = k_2[B]$		
	$\frac{d[D]}{dt} = k_3[B] - k_4[D]$		
$A \xrightarrow{k_1} B \xrightarrow{k_2} C$	$\frac{dA}{dt} = -k_1[A] - k_3[A]$	$\begin{bmatrix} -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$
$A \xrightarrow{k_3} D + E$	$\frac{dB}{dt} = +k_1[A] - k_2[B] - k_4[B]$		
$B \xrightarrow{k_4} D + F$	$\frac{dC}{dt} = +k_2[B] - k_5[C]$		
$C \xrightarrow{k_5} D + G$	$\frac{dD}{dt} = +k_3[A] + k_4[B] + k_5[C]$		
$E \xrightarrow{k_6} F$	$\frac{dE}{dt} = +k_3[A] - k_6[E]$		
$F \xrightarrow{k_7} G$	$\frac{dF}{dt} = +k_4[B] + k_6[E] - k_7[F]$		
	$\frac{dG}{dt} = +k_5[C] + k_7[F]$		

analysis of the simulated data were done in replicates of 10 and the resolution results were averaged and the standard deviations were calculated.

3.2. Chemical data

Methyl-2-[[(4-methoxy-6-methyl-1,3,5-triazin-2-yl)aminocarbonyl]-aminosulfonyl]benzoate is known under the trade name Metsulfuron methyl or Ally [17]. This compound was generously donated by Dupont de Nemours. All other chemicals were purchased from Aldrich and used without further purification. The water was purified using a nanopure, ultrapure water system from Barnstead.

The LC-DAD instrument was a Hewlett-Packard 1090 system. The mobile phase consisted of 50%

Table 2
Fit results of simple first- and second-order reactions using the correct kinetic model (upper half) and the incorrect kinetic model (lower half). The synthetic rate constant is 0.5 time^{-1} in all cases

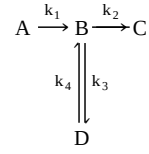
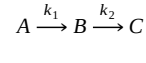
$A \xrightarrow{k_1} B$						$2A \xrightarrow{k_1} B$					
Noise (%)	Number of data points	Completion (%)	Kinetic model	k_1^a (time^{-1})	ALS fit error (%)[a]	Noise (%)	Number of data points	Completion (%)	Kinetic model	k_1^a (time^{-1})	ALS fit error (%)[a]
1	15	99	$A \rightarrow B$	0.499(0.003)	5.1(0.04)	1	5	99	$2A \rightarrow B$	0.52(0.01)	6.1(0.2)
1	10	99	$A \rightarrow B$	0.49(0.004)	5.2(0.1)	1	10	99	$2A \rightarrow B$	0.52(0.01)	6.0(0.1)
1	5	99	$A \rightarrow B$	0.49(0.01)	5.2(0.1)	1	15	99	$2A \rightarrow B$	0.51(0.004)	6.0(0.1)
1	15	75	$A \rightarrow B$	0.54(0.004)	6.4(0.2)	1	5	75	$2A \rightarrow B$	0.59(0.02)	7.5(0.3)
1	10	75	$A \rightarrow B$	0.54(0.005)	6.3(0.2)	1	10	75	$2A \rightarrow B$	0.59(0.02)	7.2(0.2)
1	5	75	$A \rightarrow B$	0.54(0.01)	6.5(0.1)	1	15	75	$2A \rightarrow B$	0.60(0.01)	7.2(0.2)
1	15	50	$A \rightarrow B$	0.61(0.01)	7.0(0.4)	1	5	50	$2A \rightarrow B$	0.69(0.07)	7.9(0.4)
1	10	50	$A \rightarrow B$	0.61(0.01)	7.5(0.7)	1	10	50	$2A \rightarrow B$	0.66(0.02)	7.0(0.5)
1	5	50	$A \rightarrow B$	0.59(0.01)	7.2(0.3)	1	15	50	$2A \rightarrow B$	0.66(0.01)	7.1(0.5)
7	15	99	$A \rightarrow B$	0.38(0.02)	34.3(0.2)	7	5	99	$2A \rightarrow B$	0.30(0.06)	46.5(0.4)
7	10	99	$A \rightarrow B$	0.38(0.04)	34.4(0.4)	7	10	99	$2A \rightarrow B$	0.33(0.04)	46.2(0.2)
7	5	99	$A \rightarrow B$	0.35(0.04)	35.1(0.4)	7	15	99	$2A \rightarrow B$	0.29(0.03)	46.4(0.2)
7	15	75	$A \rightarrow B$	0.44(0.02)	34.3(0.2)	7	5	75	$2A \rightarrow B$	0.36(0.03)	45.8(0.3)
7	10	75	$A \rightarrow B$	0.44(0.02)	34.3(0.3)	7	10	75	$2A \rightarrow B$	0.36(0.05)	45.9(0.2)
7	5	75	$A \rightarrow B$	0.44(0.03)	34.5(0.4)	7	15	75	$2A \rightarrow B$	0.36(0.04)	45.9(0.2)
7	15	50	$A \rightarrow B$	0.47(0.03)	33.8(0.2)	7	5	50	$2A \rightarrow B$	0.47(0.03)	33.8(0.2)
7	10	50	$A \rightarrow B$	0.46(0.06)	34.0(0.3)	7	10	50	$2A \rightarrow B$	0.47(0.05)	33.6(0.2)
7	5	50	$A \rightarrow B$	0.45(0.04)	34.1(0.2)	7	15	50	$2A \rightarrow B$	0.47(0.05)	33.7(0.1)
1	15	99	$2A \rightarrow B$	0.78(0.01)	8.1(0.1)	1	5	99	$A \rightarrow B$	0.32(0.01)	8.6(0.3)
1	10	99	$2A \rightarrow B$	0.78(0.02)	8.1(0.1)	1	10	99	$A \rightarrow B$	0.31(0.01)	7.9(0.4)
1	5	99	$2A \rightarrow B$	0.78(0.01)	8.0(0.1)	1	15	99	$A \rightarrow B$	0.31(0.01)	7.8(0.3)
1	15	75	$2A \rightarrow B$	0.78(0.01)	6.6(0.1)	1	5	75	$A \rightarrow B$	0.41(0.01)	8.0(0.3)
1	10	75	$2A \rightarrow B$	0.78(0.01)	6.7(0.1)	1	10	75	$A \rightarrow B$	0.41(0.01)	7.9(0.3)
1	5	75	$2A \rightarrow B$	0.78(0.01)	7.0(0.04)	1	15	75	$A \rightarrow B$	0.42(0.004)	7.9(0.1)
1	15	50	$2A \rightarrow B$	0.80(0.02)	7.1(0.3)	1	5	50	$A \rightarrow B$	0.50(0.01)	7.7(0.7)
1	10	50	$2A \rightarrow B$	0.86(0.07)	7.3(0.7)	1	10	50	$A \rightarrow B$	0.50(0.01)	7.2(0.5)
1	5	50	$2A \rightarrow B$	0.78(0.02)	7.3(0.2)	1	15	50	$A \rightarrow B$	0.50(0.01)	7.1(0.5)
7	15	99	$2A \rightarrow B$	0.65(0.08)	34.8(0.3)	7	5	99	$A \rightarrow B$	0.16(0.03)	46.7(0.4)
7	10	99	$2A \rightarrow B$	0.66(0.11)	34.7(0.1)	7	10	99	$A \rightarrow B$	0.17(0.03)	46.4(0.2)
7	5	99	$2A \rightarrow B$	0.56(0.05)	35.1(0.3)	7	15	99	$A \rightarrow B$	0.16(0.01)	46.7(0.3)
7	15	75	$2A \rightarrow B$	0.67(0.03)	34.4(0.2)	7	5	75	$A \rightarrow B$	0.24(0.01)	45.8(0.3)
7	10	75	$2A \rightarrow B$	0.65(0.07)	34.4(0.3)	7	10	75	$A \rightarrow B$	0.23(0.02)	45.8(0.2)
7	5	75	$2A \rightarrow B$	0.67(0.06)	34.7(0.4)	7	15	75	$A \rightarrow B$	0.24(0.01)	45.9(0.2)
7	15	50	$2A \rightarrow B$	0.65(0.07)	33.9(0.2)	7	5	50	$A \rightarrow B$	0.36(0.02)	33.7(0.2)
7	10	50	$2A \rightarrow B$	0.64(0.11)	34.1(0.3)	7	10	50	$A \rightarrow B$	0.35(0.03)	33.6(0.2)
7	5	50	$2A \rightarrow B$	0.59(0.06)	34.1(0.3)	7	15	50	$A \rightarrow B$	0.36(0.03)	33.6(0.2)

[a] The mean result for 10 fits with the standard deviation given in parenthesis.

acetonitrile and 50% of a pH 2.00, 0.005 M aspartate aqueous buffer mixed by the instrument. Particulates in the mobile phase were removed by 0.45 μm membrane vacuum filtration, and solvents were degassed by a helium purge. The flow rate was 0.4 ml/min and the column temperature was controlled at 38 °C. The eluate spectra were observed from 212 to 330 nm. Samples were injected using a 20 μl sample loop.

A 3 mg sample of Ally was dissolved in 0.5 ml of acetonitrile. This solution was diluted to 100 ml with 0.005 M aspartate buffer, pH 2.00. Aliquots of the Ally solution were stored in closed vials and kept in a waterbath held at 45 °C. At each time point, a sample was taken, after which the vial was discarded. Further discussions of these experimental conditions are given in Ref. [2].

Table 3
Influence of reaction completion, number of data points and noise level on the fit results of multiple step reactions. The rate constants used are 0.5, 0.5, 0.5 and 2.5 for k_1, k_2, k_3 and k_4, respectively

Fit results					Synthetic data parameters			
ALS fit (%)	k_1 (time^{-1})	k_2 (time^{-1})	k_3 (time^{-1})	k_4 (time^{-1})	Completion (%)	Number of data points	Noise (%)	Model
2.2	0.50	0.50	0.53	2.51	99	15	0.1	
3.4	0.50	0.52	0.53	2.39	99	15	0.5	
2.2	0.50	0.50	0.52	2.55	99	10	0.1	
3.4	0.50	0.53	0.53	2.47	99	10	0.5	
1.6	0.50	0.50	0.52	2.46	99	5	0.1	
3.4	0.50	0.52	0.52	2.28	99	5	0.5	
9.3	0.50	0.51	0.47	2.63	75	15	0.1	
9.9	0.49	0.53	0.47	2.72	75	15	0.5	
9.5	0.50	0.52	0.45	2.69	75	10	0.1	
10.4	0.49	0.53	0.45	2.73	75	10	0.5	
10.3	0.50	0.52	0.45	2.74	75	5	0.1	
11.2	0.50	0.53	0.44	2.72	75	5	0.5	
12.5	0.50	0.63	0.37	2.61	50	15	0.1	
12.8	0.50	0.62	0.36	2.39	50	15	0.5	
12.6	0.50	0.59	0.35	2.62	50	10	0.1	
13.3	0.50	0.61	0.35	2.69	50	10	0.5	
13.6	0.50	0.57	0.35	2.89	50	5	0.1	
14.2	0.50	0.57	0.37	3.02	50	5	0.5	
5.5	0.51	0.47	-	-	99	15	0.1	
5.0	0.52	0.45			99	15	0.5	
5.4	0.51	0.49			99	10	0.1	
5.4	0.51	0.47			99	10	0.5	
3.5	0.50	0.50			99	5	0.1	
4.3	0.50	0.49			99	5	0.5	
9.9	0.50	0.50			75	15	0.1	
10.9	0.50	0.47			75	15	0.5	
9.8	0.50	0.52			75	10	0.1	
10.8	0.50	0.48			75	10	0.5	
10.3	0.50	0.50			75	5	0.1	
11.0	0.50	0.48			75	5	0.5	
10.6	0.50	0.65			50	15	0.1	
11.9	0.49	0.51			50	15	0.5	
11.0	0.50	0.66			50	10	0.1	
11.9	0.50	0.53			50	10	0.5	
12.2	0.50	0.67			50	5	0.1	
12.4	0.50	0.61			50	5	0.5	

4. Data analysis

All data sets were analyzed using an ALS program based on a version described in a previous paper [12]. The previous version did not include the kinetic model constraint in the algorithm that is described in the present paper. The program was written and the data analysis performed using MATLAB® on various x86 computers [18].

4.1. Construction of initial estimates for the simple first- and second-order reaction

The initial guess tensor was generated by evolving factor analysis (EFA) with two components for each kinetic time point [19]. The EFA results were refined by applying two-way ALS with non-negativity constraints on the retention- and spectral profiles. These refined results were used as the initial estimates for the three-way ALS-DQ.

4.2. Construction of initial estimates for the multiple step reactions

Direct trilinear decomposition (DTD) [20,21] was used in the automatic generation of the initial estimates for the consecutive first-order reaction and the same reaction with a reversible side reaction. The retention profiles that resulted from DTD were normalized and matched to the corresponding species in the kinetic model. These retention profiles were not further refined, and the same profiles were used for each kinetic time point.

For all simulated data sets, three-way ALS was performed with non-negativity on the retention and spectral profiles and trilinearity. None of the other common constraints such as unimodality or local rank restrictions were used. The kinetic model constraint was also implemented while fitting the data.

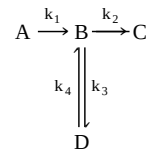
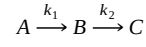
For a complete description of the generation of the initial estimates for the Ally data set refer to an earlier paper [2]. The three-way chemometric analysis of the Ally hydrolysis data-set benefited from the use of local rank restrictions. Some of the components were baseline separated and thus had a local rank of 1.

5. Results and discussion

5.1. Simple first- and second-order synthetic data

The fit results of the simulated simple first- and second-order reactions are shown in Table 2. The results from fitting the data to the correct model are given in the upper half of the table. When comparing the fit quality of the simple first- and second-order reactions, the biggest determinant of the fit quality is the noise level present in the data. The fit error typically increases fivefold when increasing the noise level from 1% to 7% while the estimate for the rate constant is typically biased low at the lower signal-

Table 4
Influence of rate constant on the fit results of multiple step reactions. The reaction was observed to 99% completion, with 15 data points and 0.1% noise was added to the data-set

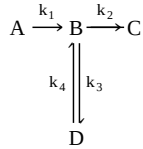
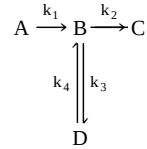
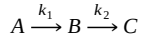
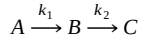
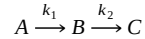
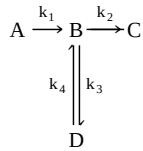
Fit results					Synthetic data parameters				
ALS fit (%)	k_1 (time^{-1})	k_2 (time^{-1})	k_3 (time^{-1})	k_4 (time^{-1})	k_3 (time^{-1})	k_4 (time^{-1})	k_1 (time^{-1})	k_2 (time^{-1})	Model
24.4	0.50	0.52	0.26	0.06	0.25	0.05	0.50	0.50	
2.4	0.50	0.50	0.26	1.24	0.25	1.25			
56.8	0.50	0.52	0.53	0.11	0.50	0.10			
2.2	0.50	0.50	0.53	2.51	0.50	2.50			
83.5	0.58	0.42	0.79	0.21	1.00	0.20			
2.1	0.50	0.50	1.07	4.90	1.00	5.00			
6.1	0.50	0.25	-	-	-	-	0.50	0.25	
5.5	0.50	0.47					0.50	0.50	
3.1	0.50	0.99					0.50	1.00	

to-noise ratio. The number of kinetic points does not have a significant impact on the fit quality or the prediction of the rate constants. By increasing the extent of reaction completion, local rank and selectivity constraints can be applied to help in achieving unambiguous resolution results. Therefore, the estimated rate constant is optimal when the reaction is followed to completion.

5.2. Multiple step synthetic data

As seen with the simple first- and second-order reactions, the noise level is of great influence in the prediction of the rate constants. This is even more the case for the more complex reaction pathways. Even at a signal-to-noise ratio of 100, the algorithm has significant difficulty in achieving a minimum while

Table 5
Influence of using the wrong kinetic model in multiple step reactions. The rate constants are 0.5, 0.5, 0.5 and 2.5 for k_1, k_2, k_3 and k_4, respectively, and 0.1% noise was added to the data-set

Fit results						Synthetic data parameters		
ALS fit (%)	k_1 (time^{-1})	k_2 (time^{-1})	k_3 (time^{-1})	k_4 (time^{-1})	Fit by model	Synthetic model	Completion (%)	Data points
2.2	0.50	0.50	0.53	2.51			99	15
2.2	0.50	0.50	0.52	2.55			99	10
1.6	0.50	0.50	0.52	2.46			99	5
9.2	0.50	0.51	0.47	2.63			75	15
9.4	0.50	0.52	0.45	2.69			75	10
10.3	0.50	0.52	0.45	2.74			75	5
12.4	0.50	0.63	0.37	2.61			50	15
12.6	0.50	0.59	0.35	2.62			50	10
13.6	0.50	0.57	0.35	2.89			50	5
16.9	0.59	0.32	-	-			99	15
16.8	0.62	0.33					99	10
15.8	0.61	0.34					99	5
27.8	0.66	0.26					75	15
26.7	0.63	0.29					75	10
24.8	0.58	0.34					75	5
26.1	0.51	0.46					50	15
26.0	0.50	0.50					50	10
26.5	0.51	0.48					50	5
5.5	0.51	0.47	-	-			99	15
5.4	0.51	0.49					99	10
3.5	0.50	0.50					99	5
9.9	0.50	0.50					75	15
9.8	0.50	0.52					75	10
10.3	0.50	0.50					75	5
10.6	0.50	0.65					50	15
11.0	0.50	0.66					50	10
12.2	0.50	0.67					50	5
23.1	0.73	0.32					99	15
20.1	0.69	0.34					99	10
21.0	0.80	0.32					99	5
31.1	0.68	0.24					75	15
31.2	0.71	0.30					75	10
33.7	0.78	0.23					75	5
30.9	0.74	0.25					50	15
27.6	0.53	0.61					50	10
31.2	0.74	0.35					50	5

searching for the rate constants. Therefore, the data was simulated with 0.1% and 0.5% noise (signal-to-noise ratios of 1000 and 200, respectively). This results in reasonable fit quality as can be seen in Table 3.

The algorithm is again influenced by the duration of the experiment. When the reaction is observed up to 99% completion, the rate constants are estimated with good accuracy as the availability of data enables local rank and selectivity constraints to be applied. If the entire reaction is not monitored, the selectivity and local rank conditions are not valid constraints, the ambiguity of the resolution is increased and the predicted rate constants are less accurate. The number of data points does not have a significant influence on the quality of fit.

Especially when the second component has a maximum intensity of less than 5% of the first component, the overall ALS fit error increases dramatically as can be seen in Table 4. This is the case when $k_2 \gg k_1$ or when the equilibrium constant is either 5 or 0.2. However, the found optimal rate constants are still within one standard deviation of the value used for the simulation of the data set. This shows a limitation of the ALS curve resolution technique, but not the fitting of the kinetic model. The ALS method is not finding the global minimum, which is evidenced by the fact that fitting using the true profiles as initial estimates of the true profiles gives fit errors between 0.17% and 2.3% depending on the noise level and the degree of reaction completion. Further research needs to focus on the implementation of constraints that will drive the resolution to the global minimum.

5.3. Incorrect chemical model

Fitting the data to the incorrect model in the case for simple first- and second-order reactions results in a worse fit until 50% or less of the reaction is observed. This is shown in the lower part of Table 2. At that point, the fit error for both models is the same within the standard deviation, resulting in an inability to confirm the reaction mechanism. This is not a limitation of the program but rather the fundamental

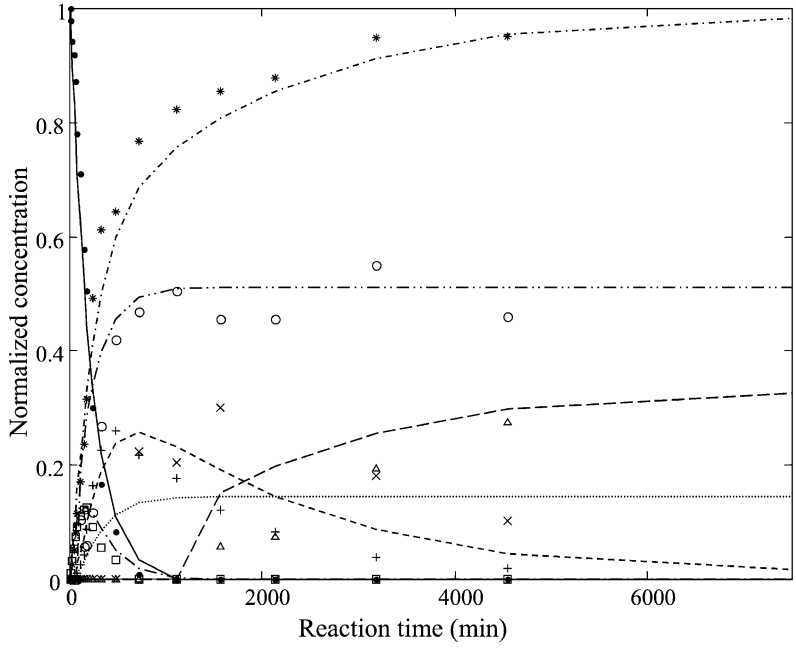


Fig. 3. The time dependence at pH 2 and 45 °C of Ally and its degradation products as the ALS kinetic profiles (symbols) and the chemical model with the rate constants fit by this program (lines) for Ally (——•), OH-Ally (–·–■), ring-opened Ally (·––+), benzenesulfonamide (·–·–·–∗), triazine (–··–○), OH-triazine (–––△) and ring opened triazine (···×).

observability of the model with respect to the amount of noise and the time range of the data set [22].

When more complex reaction pathways are analyzed, the ALS fit error is typically smaller when the correct model is used as can been seen in Table 5. However, this may be a result of the higher signal-to-noise ratio for these data.

5.4. Ally hydrolysis data

When analyzing the real data generated by the hydrolysis of Ally, the data are treated in identical manner to that mentioned in an earlier paper [2]. The difference is the application of the kinetic model to the data within the ALS iterations rather than ALS resolution followed by a separate program (Gepasi) [23–25] to fit the determined kinetic profiles to a chemical model. Fig. 3 shows the kinetic profiles of the components of the Ally hydrolysis. The previously found kinetic profiles are scaled to match the concentrations found by the kinetic model in the current program implementation. This shows the intensity ambiguity of the earlier methods where the retention and kinetic profiles were multiplied by a scalar determined by the application of the closure constraint. With the implementation of the kinetic model constraint, this scalar is completely specified and corresponds to a molar absorptivity coefficient.

The rate constants also differ from those found previously, as can be seen in Table 6. They have the same order of magnitude, but are significantly different. This is also due to the arbitrary scaling, as well as the necessity of using a closure constraint before subsequent fitting by a kinetic modeling program. Application of an additional closure constraint to the data results in very poor fit quality. The kinetic model already includes inherent mass balance assumptions and adding a more general mass balance to the fitting procedure disrupts the optimization process.

6. Conclusion

This is the first reported flexible incorporation of a differential equation solver into a multivariate curve resolution technique. It allows for the analysis of complex reaction pathways without the need for rewriting the program when exploring different reaction pathways. This method is slower in comparison to methods that use the analytical solutions of the corresponding differential equations, but with faster computers this is not a significant problem. In addition some complex reaction pathways do not have analytical solutions to the corresponding differential equations and can therefore not be analyzed with the other methods.

In extreme cases, the ALS-DQ program does not let the operator distinguish whether the correct chemical model is applied to the data by means of the overall fit error, as in some cases the fit error is the same when the data is fit to the wrong model. However, this is not a limitation of the method, but a fundamental limitation of the data characteristics as well as the basic similarities between some models. The noise level is a very important factor in determining the accuracy and precision of the prediction of rate constant from unresolved data. The high signal-to-noise ratio and long-term stability of UV–VIS measurements are more likely to result in better predictions of the rate constants than experiments such as mass spectrometry that may be subject to more drift and have typically lower signal-to-noise ratios.

The MCR-ALS technique resolves the retention and spectral profiles from overlapped components while the kinetic model constraint allows the user to fit the multivariate data to the appropriate kinetic model, directly calculating the rate constants for the various steps of the reaction. The inclusion of a kinetic model limits the rotational ambiguity found in the methods that do not use a kinetic model con-

Table 6
Rate constants for the degradation of Ally at pH 2 and 45 °C determined by previous methods and ALS-DQ

Rate constant 10^{-4} min^{-1}	k_1	k_2	k_3	k_4	k_5	k_6	k_7	ALS fit error (%)	Gepasi fit error (%)
ALS + Gepasi	61	33	43	5.7	4.9	1.3	0.16	24	11
ALS-DQ	46	48	29	9.8	0.7	4.8	1.7	19	–

straint. This program is also an improvement over previous versions where the kinetic profiles had to be fit independently by a separate program in order to determine the rate constants or where the kinetic models were limited to those with analytical solutions to the corresponding differential equations.

References

[1] R. Gargallo, R. Tauler, F. Cuesta Sanchez, D.L. Massart, Trends Anal. Chem. 15 (1996) 279–286.
[2] E.C. Bezemer, S.C. Rutan, Anal. Chem. (2001) (still in press).
[3] Y. Tan, J.-H. Jiang, H.-L. Wu, H. Cui, R.-Q. Yu, Anal. Chim. Acta 412 (2000) 195–202.
[4] A. de Juan, M. Maeder, M. Martínez, R. Tauler, Chemom. Intell. Lab. Syst. 54 (2000) 123–141.
[5] S. Bijlsma, A.K. Smilde, J. Chemom. 14 (2000) 541–560.
[6] S. Bijlsma, A.K. Smilde, Anal. Chim. Acta 396 (1999) 231–240.
[7] S. Bijlsma, D.J. Louwerse, A.K. Smilde, J. Chemom. 13 (1999) 311–329.
[8] S. Bijlsma, H.F.M. Boelens, A.K. Smilde, Appl. Spectrosc. 55 (2001) 77–83.
[9] W. Windig, B. Antalek, L.J. Sorriero, S. Bijlsma, D.J. Louwerse, A.K. Smilde, J. Chemom. 13 (1999) 95–110.
[10] S. Bijlsma, D.J. Louwerse, W. Windig, A.K. Smilde, Anal. Chim. Acta 376 (1998) 339–355.
[11] H. Haario, V.-M. Taavitsainen, Chemom. Intell. Lab. Syst. 44 (1998) 77–98.
[12] E.C. Bezemer, S.C. Rutan, Chemom. Intell. Lab. Syst. (2001) (still in press (Chemometrics and Environmetric special issue)).
[13] T.S. Blyth, E.F. Robertson, Matrices and Vector Spaces. Chapman & Hall, London, UK, 1986.
[14] R. Tauler, A. Smilde, B. Kowalski, J. Chemom. 9 (1994) 31–58.
[15] L.F. Shampine, Numerical solutions of ordinary differential equations. Chapman & Hall, London, UK, 1994.
[16] J.A. Nelder, R. Mead, Comput. J. 7 (1965) 308–313.
[17] M. Rodriguez, D.B. Orescan, Anal. Chem. (1998) 2710–2717.
[18] MATLAB 6.0. Mathworks, So. Natick, MA, 1999.
[19] M. Maeder, A. Zilian, Chemom. Intell. Lab. Syst. 3 (1988) 205–213.
[20] E. Sanchez, B.R. Kowalski, J. Chemom. 4 (1990) 29–45.
[21] K. Johnson, A. de Juan, C. Rutan, J. Chemom. 13 (1999) 1–11.
[22] S.C. Rutan, C.P. Fitzpatrick, J.W. Skoug, W.E. Weiser, H.L. Pardue, Anal. Chim. Acta 224 (1989) 243–261.
[23] P. Mendes, Comput. Appl. Biosci. 9 (1993) 563–571.
[24] P. Mendes, Trends Biochem. Sci. 22 (1997) 361–363.
[25] P. Mendes, D.B. Kell, Bioinformatics 14 (1998) 869–883.