Aachen Institute for Advanced Study in Computational Engineering Science

Preprint: AICES-2007-3

17/January/2007

Application of the incremental identification method to the formate oxidation using formate dehydrogenase

C. Michalik, T. Schmidt, M. Zavrel, M. Ansorge-Schumacher,

Antje Spiess and W. Marquardt

RWTH AACHEN UNIVERSITY

AICES

Financial support from the Deutsche Forschungsgemeinschaft (German Research Association) through grant GSC 111 is gratefully acknowledged.

List of AICES technical reports: http://www.aices.rwth-aachen.de/preprints

ELSEVIER

Chemical Engineering Science 62 (2007) 5592–5597

Chemical Engineering Science

www.elsevier.com/locate/ces

Application of the incremental identification method to the formate oxidation using formate dehydrogenase

Claas Michalik[a], Thomas Schmidt[c], Michael Zavrel[b], Marion Ansorge-Schumacher[c], Antje Spiess[b], Wolfgang Marquardt[a,*]

[a]*Lehrstuhl für Prozesstechnik, RWTH Aachen University, Turmstr. 46, D-52064 Aachen, Germany*
[b]*Lehrstuhl für Bioverfahrenstechnik, RWTH Aachen University, Worringer Weg 1, D-52074 Aachen, Germany*
[c]*Lehrstuhl für Biotechnologie, RWTH Aachen University, Worringer Weg 1, D-52074 Aachen, Germany*

Received 5 July 2006; received in revised form 17 November 2006; accepted 24 December 2006
Available online 17 January 2007

Abstract

In this paper we apply the incremental method for the identification of reaction kinetics [Brendel, M., Bonvin, D., Marquardt, W., 2006. Incremental identification of complex reaction kinetics in homogeneous systems. Chemical Engineering Science 61, 5404–5420] to the oxidation of formate by dehydrogenase. For the identification task little a priori knowledge is assumed and thus a high number of model candidates have to be considered. In order to identify the best model from a set of candidates the number of candidates is efficiently reduced in a first step using the incremental method. In a second step the best model and its parameters are determined based on the reduced candidate set in a statistically sound manner using the classical simultaneous method.

Keywords: Biocatalysis; Biochemical engineering; Enzyme; Reaction engineering; Model identification; Parameter estimation; Formate dehydrogenase

1. Introduction

Though frequently required, models for processes including (bio)chemical steps are often hard to identify (Berger et al., 2001).

A standard approach is to perform a few experiments and develop model assumptions based on the observed data. This approach suffers two major drawbacks, firstly making assumptions on the mechanisms requires expert knowledge and secondly the process of making assumptions is strongly subjective.

A method which avoids these drawbacks is the so-called incremental method (Brendel et al., 2006; Bardow and Marquardt, 2004a). This method is linked to the principal differential analysis, proposed by Ramsay and Munhall (1996). A batch reaction example is considered below to briefly describe the incremental and the simultaneous method.

A batch reactor operated at constant volume and temperature is modeled by the mass balance equation

$$\frac{\mathrm{d}\mathbf{c}(t)}{\mathrm{d}t} = \mathbf{r}(\mathbf{c}, \boldsymbol{\Theta})\mathbf{N}, \quad \mathbf{c}(t_0) = \mathbf{c}_0. \tag{1}$$

Here, $\mathbf{c}(t)$ is the vector of concentrations at time t, $\boldsymbol{\Theta}$ the vector of kinetic parameters, $\mathbf{N}$ is the stoichiometric matrix and $\mathbf{r}(\mathbf{c}, \boldsymbol{\Theta})$ the rate of reaction. The task of model identification is to find a suitable model structure and a set of parameter values $\boldsymbol{\Theta}$ for the reaction rates $\mathbf{r}(\mathbf{c}, \boldsymbol{\Theta})$.

1.1. Simultaneous (integral) model identification

Based on a priori knowledge or experimental data candidate kinetic laws are proposed for any reaction involved. One kinetic model is developed for each combination of kinetic laws (Holland and Rayford, 1990). Subsequently, any of these models is fitted to the available measurements, using dynamic parameter estimation techniques (Schittkowski, 2002). For the

* Corresponding author. Tel.: +49 241 8096712; fax: +49 241 8092326.
E-mail address: Marquardt@lpt.rwth-aachen.de (W. Marquardt).

0009-2509/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ces.2006.12.072

batch example considered above, a simple least squares approach results in

$$\min_{\mathbf{c}_0,\boldsymbol{\Theta}} \quad \sum_{i=1}^{N_c}\sum_{j=1}^{N_{m,i}}(c_i(\boldsymbol{\Theta},t_j)-\tilde{c}_i(t_j))^2, \tag{2}$$

$$\text{s.t.} \quad \frac{\mathrm{d}\mathbf{c}}{\mathrm{d}t}=\mathbf{f}(\mathbf{c},\boldsymbol{\Theta}), \quad \mathbf{c}(t_0)=\mathbf{c}_0. \tag{3}$$

N_c is the number of components present in the system, $N_{m,i}$ is the number of measurements of component i, $\tilde{c}_i(t_j)$ is the measurement of component i at time t_j and Eq. (3) represents the dynamic process model.

When following this approach one should be aware of the following downsides:

(1) The number of candidate models grows exponentially with the number of candidate kinetic laws (Milstein, 1980).
(2) Dynamic parameter estimation problems can be cumbersome to solve, especially for strongly non-linear models. In addition, the solution is computationally expensive, because either the model has to be integrated repeatedly (Bock, 1980) or it has to be transformed by discretization so that a non-linear programming problem has to be solved (Biegler et al., 2002). These problems grow rapidly with the number of model parameters and discretization points.
(3) The source of a potential misfit is hard to determine. In particular it is hard to distinguish between an unsuitable model structure, only locally optimal parameter values or even insufficient information content in the data.

1.2. Incremental model identification

Using the incremental method the integration of the kinetic model is avoided and thus the computational cost is strongly reduced (Brendel et al., 2006). Model (1) is split into a sequence of less complex sub problems represented by the equations

$$\frac{\mathrm{d}\mathbf{c}(t)}{\mathrm{d}t}=\mathbf{f}_\mathbf{r}(t), \tag{4}$$

$$\mathbf{f}_\mathbf{r}(t)=\mathbf{r}(t)\mathbf{N}, \tag{5}$$

$$\mathbf{r}(t)=\mathbf{f}(\mathbf{c},\boldsymbol{\Theta}). \tag{6}$$

The three major steps in the identification procedure follow these equations:

(1) The mole balances—Eq. (4)—are used to estimate the reaction fluxes $\mathbf{f}_\mathbf{r}$. This step requires the differentiation of the measurement data, which is an ill-posed problem in the sense of Hadamard and will introduce a bias to the estimation procedure (Bardow and Marquardt, 2004a). A thorough discussion of the selection of measurements to guarantee identifiability has been presented by Brendel et al. (2006).
(2) The stoichiometric relations—Eq. (5)—are used to estimate the reaction rates $\mathbf{r}(t)$, based on the estimated reaction fluxes $\mathbf{f}_\mathbf{r}$. Knowledge of the stoichiometric matrix $\mathbf{N}$ is assumed, here. Note, that target factor analysis (Bonvin and Rippin, 1990) can be used to identify those stoichiometries that are consistent with the measurement data from a set of proposed stoichiometries.
(3) The kinetic laws—Eq. (6)—are fitted to the estimated reaction rates, one at a time, using for instance a least squares approach and standard non-linear programming techniques (Nocedal and Wright, 1999).
(4) Finally a classical simultaneous parameter estimation is performed for the best model to obtain statistically optimal parameter values (Brendel et al., 2006). Please note that this step is computationally cheap, because almost optimal initial guesses are known from step 3.

This way the best model can be identified and its optimal parameter values can be determined. The complexity of the identification task is reduced, since any major assumption on the model structure can be analyzed in a separate step as shown above. In addition, the combinatorial explosion of model candidates is avoided, since any estimated reaction rate can be correlated with its candidates individually.

Here, the incremental method was briefly explained using an isothermal batch reactor. The method can, however, be applied to a variety of more complex problems, as for example non-isothermal reaction conditions, semi-batch reactors and multiphase reaction systems (Brendel, 2006).

In the following sections the benefits of the above described method will be demonstrated for an enzyme catalyzed reaction. In contrast to the examples presented so far (Brendel et al., 2006) this is the first application of the method to a reaction example with real measurement data. All examples presented so far are based on simulated measurements and do not comprise enzyme catalyzed reactions. These reactions tend to have more complex rate expressions and are thus harder to solve. Please note that the method has already been used for the identification of concentration dependent binary diffusion coefficients (Bardow and Marquardt, 2004b).

2. FDH example

The above described method has been applied to an enzymatic reaction system, where the enzyme formate dehydrogenase from *Candida boidinii* (EC 1.2.1.2) catalyzes the oxidization of formate to carbon dioxide by concomitantly reducing NAD^+ to NADH. Many enzymatic reductions need NADH as a cofactor, which is oxidized to NAD^+. The cost for such a process can be reduced if NADH is regenerated in situ by a second enzymatic reaction, such as the one described above (Chenault and Whitesides, 1987). This idea has already been used in industrial practice (Kula, 1994; Liese et al., 2000). In the remainder of this work a model for the formate dehydrogenase will be identified, which is valid for near equimolar conditions as well as at the technically relevant high surplus of formate.

2.1. The reaction system

The reactions considered are

$$\mathrm{HCOO^-} + \mathrm{NAD^+} \overset{r_1}{\rightleftharpoons} \mathrm{CO_2} + \mathrm{NADH}, \tag{7}$$

$$\mathrm{NADH} \overset{r_2}{\rightarrow} \text{Decomposition}. \tag{8}$$

These reactions only include the reactions for which we do not assume equilibrium concentrations at all times.

Reaction rate r_2 is known to be of first order (Alivasatos and Ungar, 1964) and can be described as

$$r_2 = \Theta_{r_2} \cdot [\mathrm{NADH}]. \tag{9}$$

The structure of the reaction rate r_1 is to be identified from measurement data. The reaction is known to follow an ordered bi–bi mechanism (Kato et al., 1979). Firstly NAD^+ and secondly formate binds to the enzyme. The first released product is CO_2, the second is NADH (Kato et al., 1979; Blanchard and Cleland, 1980). This mechanism can be described by the following set of micro-kinetic reactions:

$$\mathrm{NAD^+} + E \overset{r_{m1}}{\rightleftharpoons} \mathrm{EA}, \tag{10}$$

$$\mathrm{EA} + \mathrm{HCOO^-} \overset{r_{m2}}{\rightleftarrows} \mathrm{EPQ}, \tag{11}$$

$$\mathrm{EPQ} \overset{r_{m3}}{\rightleftharpoons} \mathrm{EQ} + \mathrm{CO_2}, \tag{12}$$

$$\mathrm{EQ} \overset{r_{m4}}{\rightleftharpoons} E + \mathrm{NADH}, \tag{13}$$

E denotes the enzyme and EA, EPQ and EQ the enzyme-substrate and enzyme-product complexes.

Following Cleland (1963), a so-called macro-kinetic model can be derived for r_1 based on Eqs. (10)–(13). In order to reduce the number of parameters to be estimated in such a macro-kinetic model, irreversible reactions are frequently assumed for part of or all the micro-kinetic reactions.

However, assuming little a priori knowledge, irreversible and reversible reactions are taken into account for any micro-kinetic reaction, here. This way 12 candidate macro-kinetic models based on the 16 possible micro-kinetic models can be derived as shown in Table 1. Here r_{m1} through r_{m4} are the 4 micro-kinetic reactions, 'R' indicates a reversible reaction, 'I' indicates an irreversible reaction and 'R/I' means that either a reversible or an irreversible reaction can be assumed to result in the same macro-kinetic model.

2.2. Measurement technique and experimental setting

All chemicals have been purchased from Fluka, Buchs, Switzerland except for NAD^+-specific formate dehydrogenase from *Candida boidinii* (EC 1.2.1.2) and NAD^+, which have been obtained from Julich Chiral Solutions, Jülich, Germany.

NAD^+ and sodium formate are dissolved in 200 mM sodium phosphate buffer pH 7.5 for all experiments. The temperature is kept constant at 25 °C. The reaction is started by adding 22 μM formate dehydrogenase and the number of active sites is determined by the method of Felber (2001). Absorption of NADH is measured at 340 nm in a 300 μl micro cuvette with 0.5 mm optical path using a uv/vis spectrometer (SpectraMax Plus, Molecular Devices, Sunnywale, California, USA). Data are obtained with an interval of 30 or 15 s[1] over a time period of 1 to 20 h.

Table 1
Regression results for all candidate models

No.	r_{m1}	r_{m2}	r_{m3}	r_{m4}	RSoS	No. of parameters
1	R	R	R	R	7.0978e4	10
2	R	R	R	I	4.0978e6	8
3	R	R/I	I	R	7.3556e4	5
4	R	R/I	I	I	4.0932e6	4
5	R	I	R	R	7.0982e4	8
6	R	I	R	I	4.1066e6	5
7	I	R	R	R	5.6352e6	6
8	I	R	R	I	5.6367e6	5
9	I	R/I	I	R	1.6758e6	4
10	I	R/I	I	I	6.2184e6	3
11	I	I	R	R	5.6357e6	5
12	I	I	R	I	6.1085e6	4

In columns r_{m1} through r_{m4} reversibility (R) or irreversibility (I) of the micro-reaction steps are denoted. RSoS is the residual sum of squares, which serves as objective function. The last column gives the number of parameters that have to be estimated for each model.

Table 2
Results of the dynamic parameter estimations of selected model candidates (identified in Table 1)

No.	RSoS
1	2.4843e5
3	2.5245e5
5	2.4844e5

RSoS is the residual sum of squares, which serves as objective function.

Table 3
Parameter values and relative standard deviations for the best model candidate, No. 3 as given in Eq. (14)

Parameter	Value	$\frac{\text{STDV}}{\text{Value}}$
k_{cat_f}	176.7e − 3 s^{-1}	4.32e−4
K_{mA}	36.04 μM	4.36e−4
K_{mB}	381.8e − 3 μM	6.66e1
K_{iA}	848.0e2 mM	6.65
K_{iQ}	101.6 μM	3.68e−3

Table 4
Parameter values and relative standard deviations for the modified model as given in Eq. (15)

Parameter	Value	$\frac{\text{STDV}}{\text{Value}}$
k_{cat_f}	176.7e − 3 s^{-1}	4.31e−4
K_{mA}	36.42 μM	3.27e−3
P_1	32.38 $(\mathrm{mM})^2$	1.49e−3
K_{iQ}	101.6 μM	1.56e−3

[1] For experiments running for more than 5 h, 30 s intervals are used, 15 s otherwise.

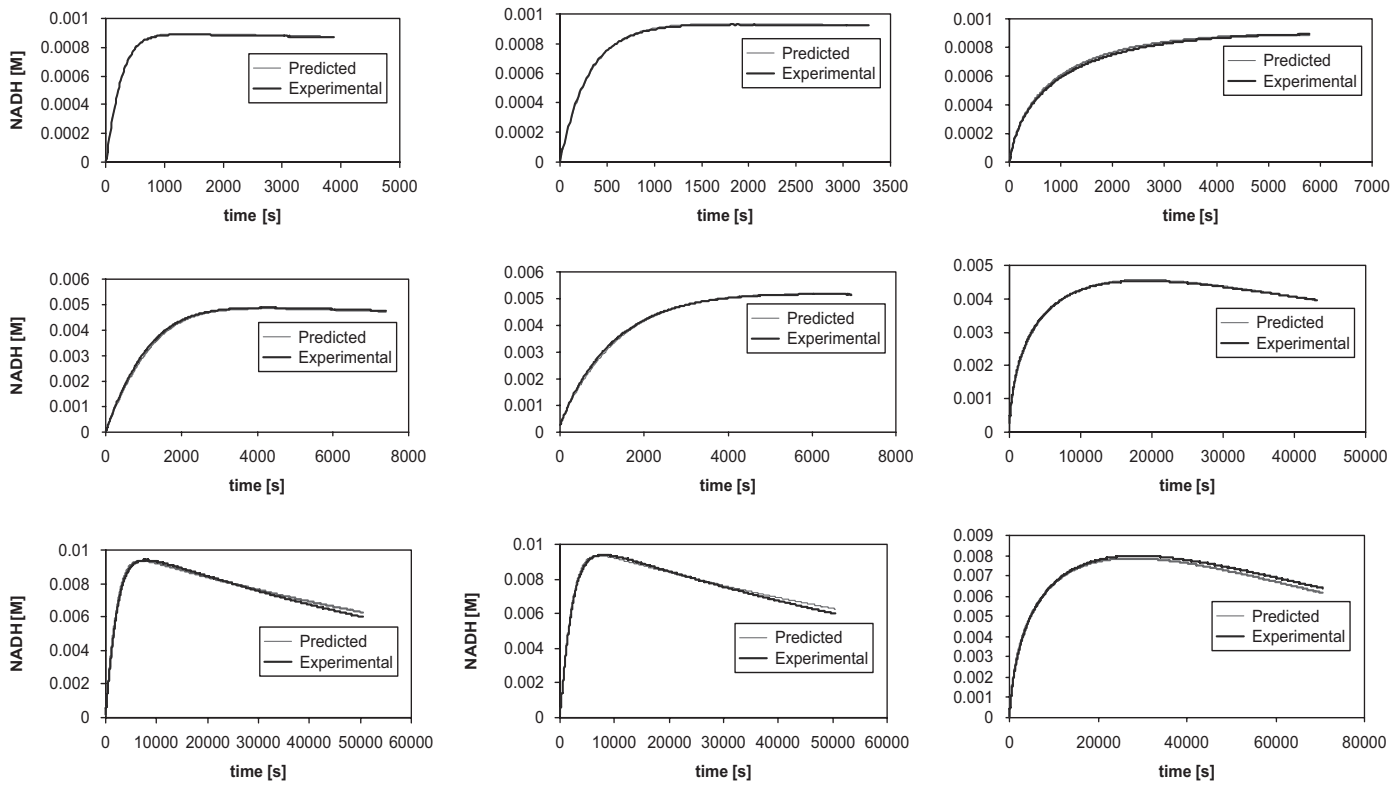

Fig. 1. Model prediction and measurements for the 9 factorial design experiments column 1 through 3 correspond to an initial concentration of formate of 1500, 500 and 50 mM; row 1 through 3 correspond to an initial concentration of NAD^+ of 1, 5 and 10 mM.

With this experimental setting a 2^3 factorial design of experiments is performed. The degrees of freedom considered are the initial concentration of $HCOO^-$ and the initial concentration of NAD^+. The applied initial concentration settings are 50, 500 and 1500 mM for formate and 1, 5 and 10 mM for NAD^+. In addition, one independent experiment with only NADH present in the system was performed to determine the rate constant for the decomposition of NADH. Two experiments were performed to confirm that no significant enzyme deactivation occurs. Hence, 12 experiments were performed in total.

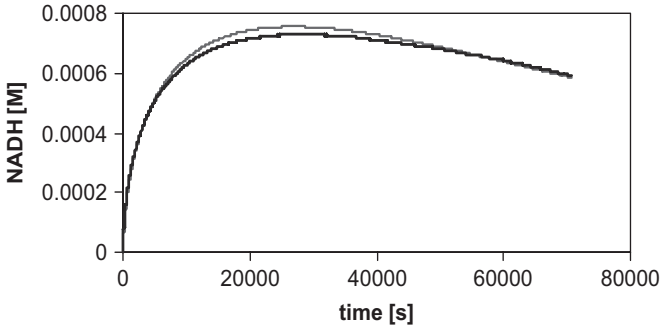

Fig. 2. Model prediction and measurements for the additional experiment.

2.3. Model identification

In the above described reaction system, two reactions are involved. To decrease the complexity of the identification task it is favorable to decompose it as far as possible. In this case a decomposition of the two reactions is possible. Due to knowledge of a rate expression for r_2 (Alivasatos and Ungar, 1964) Θ_{r2}, the only unknown parameter in the rate expression (9), can be estimated. This can be achieved using one independent experiment, as mentioned above. Knowing the rate constant of (9) the reaction rate r_2 can be calculated. Based on the estimate of r_2 and the reaction flux of NADH caused by r_1 and r_2, it is possible to estimate r_1. Hence, the candidate models can be fitted to the estimated reaction rate r_1 using a least squares approach and non-linear programming techniques (Nocedal and Wright, 1999).

Especially for the models with a high number of parameters (namely model one, two and five) the existence of many local minima can be observed. In order to find the global minimum (or at least a "good" local minimum) each regression is run 200 times with different, randomly chosen starting points. The results of this step are summarized in Table 1.

Since three models (model one, three and five) perform almost equally well, we apply the final simultaneous (dynamic) parameter estimation (step 4) to those three models instead of only taking the best one into account. This way the identification of the statistically optimal model is assured. The results are summarized in Table 2.

It can be seen that the first model (10 parameters) gives the lowest residual sum of squares, the fifth model (8 parameters) performs slightly worse and the third model (5 parameters) is, again, slightly worse. The third model, however, has got the lowest number of parameters. If models with a different number of parameters perform practically equally well, the model with

the lowest number of parameters can be considered best (Hastie et al., 2001), which is model No. 3 as given by

$$r_1 = \frac{(k_{\mathrm{cat}_f}/K_{iA} \cdot K_{mB}) \cdot A \cdot B}{1 + A/K_{iA} + K_{mA} \cdot B/K_{iA} \cdot K_{mB} + A \cdot B/K_{iA} \cdot K_{mB} + Q/K_{iQ} + K_{mA} \cdot B \cdot Q/K_{iA} \cdot K_{mB} \cdot K_{iQ}} \cdot E, \qquad (14)$$

A refers to the concentration of NAD^+, B to $HCOO^-$, P to CO_2, Q to NADH and E to the enzyme concentration.

The parameters and relative standard deviations are listed in Table 3. In contrast to the other parameters, the standard deviations for K_{iA} and K_{mB} are unacceptably high. In order to determine those parameters more accurately, experiments at very low formate and very high NAD^+ concentrations are necessary. However, the above described measurement technique does not allow for good measurement data in such regions. To get a well defined model with a high confidence in all parameters, the model given in Eq. (14) is slightly modified. The term (A/K_{iA}) is dropped[2] and a new parameter $P_1 = K_{iA} \cdot K_{mB}$ is introduced, yielding

$$r_1 = \frac{(k_{\mathrm{cat}_f}/P_1) \cdot A \cdot B}{1 + K_{mA} \cdot B/P_1 + A \cdot B/P_1 + Q/K_{iQ} + K_{mA} \cdot B \cdot Q/P_1 \cdot K_{iQ}} \cdot E. \qquad (15)$$

Parameter values and relative standard deviations for this model are given in Table 4.

Finally, one additional experiment was executed to test the extrapolation ability of the model (15). For this experiment initial concentrations of 5 mM for NAD^+ and 1 mM for formate were used. Plots showing the measurement data and the model predictions for the 9 factorial designed experiments as well as for the additional experiment are shown in Figs. 1 and 2. It can be seen that both, the original and the additional data are predicted well by the model, even though the additional data lie outside of the range used previously.

3. Conclusions

Applying the incremental method for the identification of reaction kinetics it was possible to identify a model for the formate dehydrogenase, valid for a wide range of settings. The method proved to be computationally efficient, allowing to take a high number of candidate models into account. Due to the low computational cost of the incremental method it was possible to estimate the parameters for the 12 model candidates solving altogether 2400 randomly initialized non-linear programming problems. This approach drastically lowered the risk of getting stuck in bad, locally optimal parameter values. Finally, it was shown that the identified model possesses good extrapolation capabilities.

Future work should focus on the temperature dependence of the reaction, which was not considered in this study and on the exact determination of K_{iA} and K_{mB} in the mechanistic model given in Eq. (14).

[2] K_{iA} is large, so that (A/K_{iA}) is effectively zero for all relevant concentrations of NAD^+.

Acknowledgments

We gratefully acknowledge the financial support by the Deutsche Forschungsgemeinschaft (DFG) via the collaborative research center 540 (SFB 540). We would also like to thank the reviewers as well as André Bardow for their fruitful comments that helped increasing the quality of this manuscript.

References

Alivasatos, S.G.A., Ungar, F.G.A., 1964. Nonenzymatic interactions of reduced coenzyme I with inorganic phosphate and certain other anions. Nature 203, 973–974.

Bardow, A., Marquardt, W., 2004a. Incremental and simultaneous identification of reaction kinetics: methods and comparison. Chemical Engineering Science 59 (13), 2673–2684.

Bardow, A., Marquardt, W., 2004b. Identification of diffusive transport by means of an incremental approach. Computers and Chemical Engineering 28 (5), 585–595.

Berger, R.J., Stitt, E.H., Marin, G.B., Kapteijn, F.A., Moulijn, J., 2001. Eurokin - chemical reaction kinetics in practice 5 (1), 30–60.

Biegler, L.T., Cervantes, A.M., Wächter, A., 2002. Advances in simultaneous strategies for dynamic process optimization. Chemical Engineering Science 57, 575–593.

Blanchard, J., Cleland, W.W., 1980. Kinetic and chemical mechanisms of yeast formate dehydrogenase. Biochemistry 19, 3543–3550.

Bock, H.G., 1980. Numerical treatment of inverse problems in chemical reaction kinetics. In: Ebert, K., Deuflhard, P., Jäger, W. (Eds.), Modelling of Chemical Reaction Systems. pp. 102–125.

Bonvin, D., Rippin, D.W.T., 1990. Target factor analysis for the identification of stoichiometric models. Chemical Engineering Science 43 (12), 3417–3426.

Brendel, M., 2006. Incremental identification of complex reaction systems. Ph.D. Thesis, Aachen, Germany.

Brendel, M., Bonvin, D., Marquardt, W., 2006. Incremental identification of complex reaction kinetics in homogeneous systems. Chemical Engineering Science 61, 5404–5420.

Chenault, H.K., Whitesides, G.M., 1987. Regeneration of nicotinamide cofactors for use in organic synthesis. Applied Biochemistry and Biotechnology 14, 147–197.

Cleland, W.W., 1963. The kinetics of enzyme-catalyzed reactions with two or more substrates or products. Biochimica et Biophysica Acta (BBA) 67, 104–137.

Felber, S., 2001. Optimierung der NAD-abhängigen Formiatdehydrogenase aus Candida Boidinii für den Einsatz in der Katalyse. Ph.D. Thesis, Düsseldorf, Germany.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning. Springer, New York.

Holland, D., Rayford, G., 1990. Fundamentals of Chemical Reaction Engineering. Prentice-Hall, New Jersey.

Kato, N., Sahm, H., Wagner, F., 1979. Steady-state kinetics of formaldehyde dehydrogenase and formate dehydrogenase from a methanol-utilizing yeast, candida boidinii. Biochimica et Biophysica Acta, Enzymology 566, 12–20.

Kula, M.-R., 1994. Enzyme catalysed reductions of carbonyl groups. In: Gani, R., Jorgensen, S. (Eds.), Proceedings of the Chiral Europe 1994 Symposium. pp. 27–33.

Liese, A., Seelbach, K., Wandrey, C., 2000. Industrial Biotransformations. Wiley-VCH, Weinheim, Germany. pp. 79–80, 103, 125.

Milstein, J., 1980. The inverse problem: estimation of kinetic parameters. In: Ebert, K., Deuflhard, P., Jäger, W. (Eds.), Modelling of Chemical Reaction Systems. pp. 92–101.

Nocedal, J., Wright, S., 1999. Numerical Optimization. Springer, New York.

Ramsay, J., Munhall, K., 1996. Functional data analysis of lip motion. Journal of the Acoustical Society of America 99, 3718–3727.

Schittkowski, K., 2002. Numerical Data Fitting in Dynamical Systems. Kluwer Academic Publishers, Dordrecht, The Netherlands.