



IDENTIFICATION OF RATE CONSTANTS AND NONOBSERVABLE ABSORPTION SPECTRA IN NONLINEAR BIOCHEMICAL REACTION DYNAMICS

D. MARAUN, W. HORBELT, H. RUST and J. TIMMER*

*Freiburger Zentrum für Datenanalyse und Modellbildung
and Department of Physics, University of Freiburg,
Eckerstr. 1, 79104 Freiburg, Germany
jeti@fdm.uni-freiburg.de

H. P. HAPPERSBERGER

*Scil Biomedicals GmbH,
Fraunhoferstr. 15, 82152 Martingried, Germany*

F. DREPPER

*Institute for Biology II, University of Freiburg,
Schänzlestr. 1, 79104 Freiburg, Germany*

Received February 5, 2002; Revised April 15, 2002

On analyzing data of biochemical reaction dynamics monitored by time-resolved spectroscopy, one faces the problem that the concentration time courses of the involved components are not directly observed, but the superposition of their absorption spectra. Furthermore the single spectra are often unknown, because the corresponding reagents cannot be isolated. We propose a method based on Bock's multiple shooting algorithm to estimate the rate constants and individual spectra simultaneously. Applying this procedure to a biochemical reaction we identify the specific rate constants characterizing the reaction dynamics as well as the nonobservable absorption spectra. The results lead to a better understanding of the kinetics of a novel modification reaction which was used as trapping reaction in disulfide bond mediated protein folding reactions.

Keywords: Modeling; parameter estimation; identification; biochemical reaction.

1. Introduction

One of the aims of time series analysis is to gain information about a process by fitting a model of ordinary differential equations to the experimentally measured data. If the dynamical variables are not directly observed, one also has to model a suitable observation function connecting the dynamical state variables to the measured data. The estimation of this observation function can provide new

insight into additional characteristics of the investigated system.

Analysis of data measured by time-resolved optical absorption spectroscopy of a biochemical process illustrates this situation: given some nonlinear reaction dynamics modeled by a system of ordinary differential equations, one measures the temporal evolution of the absorption spectra of the solution. These data are a superposition of the

* Author for correspondence.

spectral contributions of the reaction partners. While the spectra of the involved educts are usually known from isolated measurements, those of intermediate products often cannot be measured isolatedly. In this context, the nonobservable concentrations represent the dynamical variables and the superposition of the spectra represents the non-parametric observation function connecting the reaction dynamics to the measured absorption data. The task is to estimate the parameters of the differential equation as well as the unknown absorption spectra.

A standard method to analyze such data is to use information at specific wavelengths around the absorption peaks of the individual components. A first approach linearizes the differential equations at $t = 0$ and uses data of the first time steps to estimate an effective rate constant of the linearized model. An extension uses the full nonlinear model. Both methods discount all the data at other wavelengths. For linear reaction dynamics, Hayashi and Sakamoto [1986] suggested to use data at more wavelengths than components taking part in the reaction. We extend this idea and develop a method for nonlinear dynamics using information at a subset of all measured data. Doing this we are able to estimate not only the underlying dynamical parameters, but also the nonobservable spectra.

In this work we present a detailed analysis of a biochemical reaction which is used for the study of disulfide bond-mediated folding pathways of proteins. The trapping agent melarsen oxide (MEL) is used to trap folding intermediates of proteins which can then be analyzed by mass spectrometry and peptide mapping [Happersberger *et al.*, 1998a; Happersberger & Glocker, 1998]. This “static” approach can be extended to a dynamical analysis of the formation or breaking of disulfide bonds by time-resolved optical spectroscopy [Happersberger *et al.*, 1998b]. Here we use experimental data of the reaction with the model peptide glutathione (GSH). We show that our approach is well suited to determine rate constants of individual reaction steps and the optical spectra of transient species which are not observable individually.

The paper is organized as follows: in Sec. 2 we describe the proposed method in general. It is applied to the introduced biochemical model reaction in Sec. 3. The results are presented in Sec. 4.

2. Methods

Consider a time-continuous, dynamical process de-

scribed by a nonlinear ordinary differential equation $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{k})$, with $\mathbf{x}(t) \in \mathbb{R}^K$, $t \in [0, T]$. (1)

Each component of \mathbf{x} represents the concentration of one chemical substance taking part in the reaction dynamics. The setting is characterized by an unknown parameter vector $\boldsymbol{\theta} = (\mathbf{k}, \mathbf{x}(0))$ including the dynamical parameters \mathbf{k} , and the initial values $\mathbf{x}(0)$.

Information about the process is gained by time-resolved spectroscopy, i.e. at each time t_i ($i = 1, \dots, N$) the state vector $\mathbf{x}(t_i, \boldsymbol{\theta})$ is related to the noisy measured multivariate time series $\{(y_{ij})\}$ by the product with the spectra at wavelength λ_j ($j = 1, \dots, M$) of each substance

$$\begin{aligned} y_{ij} &= \sum_{k=1}^K s_k(\lambda_j) x_k(t_i, \boldsymbol{\theta}) + c(\lambda_j) + \eta_{ij} \\ &= \mathbf{s}^T(\lambda_j) \mathbf{x}(t_i, \boldsymbol{\theta}) + c(\lambda_j) + \eta_{ij}. \end{aligned} \quad (2)$$

Each row $\mathbf{s}(\lambda_j)$ of the $K \times M$ matrix \mathbf{S} contains the spectral contributions of the K components of $\mathbf{x}(t_i)$ to the data points y_i at the j th wavelength. $c(\lambda_j)$ is an offset not yet specified and η_{ij} denotes independent normally distributed noise with zero mean and variance σ_{ij}^2 .

The aim is to determine those parameters $\hat{\boldsymbol{\theta}}$ and spectra $\hat{\mathbf{S}}$ for which $\hat{\mathbf{S}}\mathbf{x}(t, \hat{\boldsymbol{\theta}}) + \mathbf{c}$ is closest to the observed data in a least squares sense. A first method to solve this optimization problem is the initial value approach: the objective function

$$\begin{aligned} \chi^2(\mathbf{S}, \boldsymbol{\theta}) &= \sum_{i=1}^N \sum_{j=1}^M \left(\frac{y_{ij} - \sum_{k=1}^K s_k(\lambda_j) x_k(t_i, \boldsymbol{\theta}) - c(\lambda_j)}{\sigma_{ij}} \right)^2 \end{aligned} \quad (3)$$

is minimized with respect to $\boldsymbol{\theta}$ and \mathbf{S} using a Gauss–Newton method [Gill *et al.*, 1981]. Given starting guesses for $\boldsymbol{\theta}$ and \mathbf{S} , one integrates the differential equations numerically and calculates $\mathbf{S}\mathbf{x}(t_i, \boldsymbol{\theta})$. A numerical optimization algorithm calculates an update step $(\Delta p, \Delta S)$ for the parameters by solving the linearized problem. This procedure is iterated with the new vector until a predefined convergence criterion is satisfied. Applied to nonlinear ordinary differential equations, this naive approach often converges to local minima or even diverges,

because the nonlinear dependency of $\mathbf{x}(t_i)$ on $\boldsymbol{\theta}$ lets the estimated trajectories diverge from the data.

The multiple shooting algorithm, developed by [Bock, 1981], circumvents this problem by reformulating the task as a multipoint boundary value problem. The fitting interval $[0, T]$ is partitioned into L subintervals:

$$0 = T_0 < T_1 < \dots < T_L = T. \quad (4)$$

For each subinterval $[T_l, T_{l+1}]$, local initial values $\mathbf{x}_l = \mathbf{x}(T_l)$ are introduced as additional parameters. The dynamical equations are integrated piecewise and the objective functional $\chi^2(\mathbf{S}, \boldsymbol{\theta})$ is evaluated and minimized as in the initial value approach. While the dynamical parameters \mathbf{k} are unique over the entire interval, the local initial values are optimized separately in each subinterval. This approach leads to initially discontinuous trajectories which are, however, always near the data. The finally estimated trajectory must of course be continuous, i.e. the computed solution at the end of the subinterval l must finally equal the local initial value of the following subinterval $l + 1$:

$$\lim_{t \rightarrow T_l} \mathbf{x}(t) = \mathbf{x}(T_l) = \mathbf{x}_l, \quad l = 1, \dots, L - 1. \quad (5)$$

Equation (5) are taken into account as equality constraints in the optimization procedure. Since only their linearizations are imposed on the update step, the iterates will generally be discontinuous trajectories. This freedom allows the method to stay close to the observed data, prevents divergence of the numerical solution and reduces the problem of local minima.

More details of the mathematical and implementational aspects of the method are given in [Bock, 1981, 1983]. Some applications are given in [Horbelt *et al.*, 2001; Timmer *et al.*, 2000].

3. Application to a Biochemical Model Reaction

In this section the presented methods are applied to time-resolved measurements of a biochemical reaction: the trapping agent melarsen oxide [p-(4, 6-diamino-1, 3, 5-triazin-2-yl)aminophenylarsonous acid (MEL)] is able to bridge two cystein residues of a protein which can adopt a disulfide bond in the native structure. The basic reaction steps of this process, i.e. the formation and the stability of the first and the second thiobonds of the bridge are already visible in the reaction of MEL with the tripeptide glutathione (GSH, cf. Fig. 1) according to

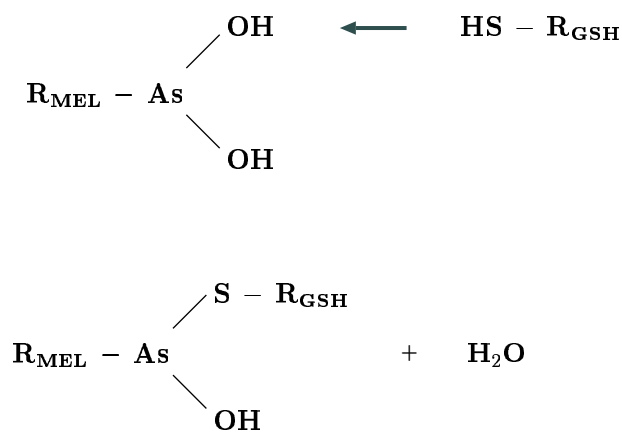
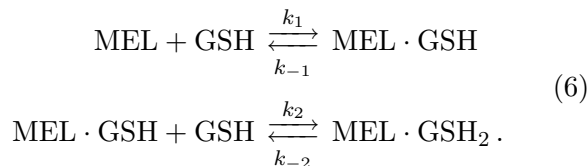


Fig. 1. The reaction: When MEL reacts with GSH, one of the hydroxyl groups is replaced by the sulfur, and water is produced.

the model



The reaction is described by the system of nonlinear ordinary differential equations

$$\begin{aligned} \dot{x}_1 &= -k_1 x_1 x_2 + k_{-1} x_3 \\ \dot{x}_2 &= -k_1 x_1 x_2 + k_{-1} x_3 - k_2 x_2 x_3 + k_{-2} x_4 \\ \dot{x}_3 &= +k_1 x_1 x_2 - k_{-1} x_3 - k_2 x_2 x_3 + k_{-2} x_4 \\ \dot{x}_4 &= +k_2 x_2 x_3 - k_{-2} x_4, \end{aligned} \quad (7)$$

where x_1 represents the MEL-concentration [MEL], $x_2 = [\text{GSH}]$, $x_3 = [\text{MEL} \cdot \text{GSH}]$ and $x_4 = [\text{MEL} \cdot \text{GSH}_2]$. MEL and GSH are below called educts, MEL · GSH and MEL · GSH₂ products. The vector $\mathbf{k} = (k_1, k_{-1}, k_2, k_{-2})^T$ comprises the rate constants, a positive index denotes an on-rate, a negative an off-rate.

Data were obtained by time-resolved optical spectroscopy. The reaction was measured in a stopped-flow device by a single-beam photo meter (see Fig. 2). Experiments with initial concentrations of $[\text{MEL}]_0 = 45 \mu\text{mol/l}$, different initial GSH concentrations of around $60 \mu\text{mol/l}$, $120 \mu\text{mol/l}$, $240 \mu\text{mol/l}$ and $720 \mu\text{mol/l}$ and vanishing initial concentrations of MEL · GSH and MEL · GSH₂ were recorded. The length of each measurement is $T = 2055 \text{ ms}$, sampled in 5 ms steps.

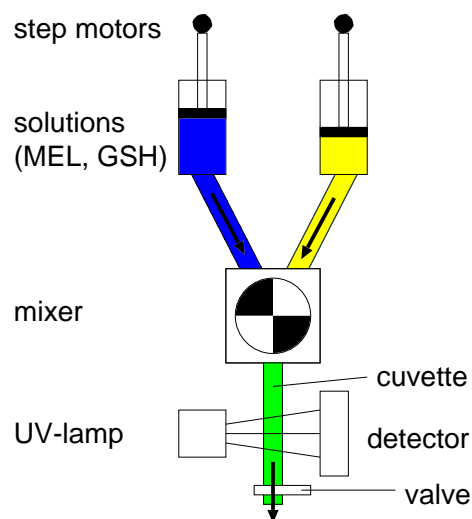


Fig. 2. Stopped-flow device: the two educts are injected by step-motors into the mixer, where the reaction starts. The mixture flows through the cuvette until the residua of former experiments are washed out. At $t = 0$ the flow is stopped by a valve and the reaction proceeds in the cuvette. The optical pathlength in the cuvette is 1 cm.

The measured wavelengths range from 240 nm to 350 nm with a resolution of 1 nm. Since the absorption spectra are smooth, they can be described sufficiently by a small subset of wavelengths. Because of observational noise, one should use as much information as possible also from the wavelengths not taken into account. This can be achieved by smoothing the data by a cubic spline. We observed that a spline with 16 knots results in an approximation of the underlying curve which is accurate to within the error bars of the data. Ten of these knots were at equidistant wavelengths from 247 nm to 327 nm. These were evaluated for the dynamical fit. The remaining six knots had to be placed around a knee at 300 nm to obtain an unbiased approximation. The reduction affects an increased numerical efficiency because only a ten-dimensional time series has to be analyzed instead of one with 111 dimensions. Wavelengths with $\lambda > 320$ nm were discarded because they were corrupted by additional effects, e.g. bubbles with a different density of the soluted components causing a lower or higher absorption. Moreover data at $t > 2055$ ms were ignored because they showed systematic drifts not captured by the suggested model. Figure 3 shows a typical dataset.

The transmission I of the UV-light passing a substance with extinction coefficient ε and concentration c in a cuvette with optical pathlength d is

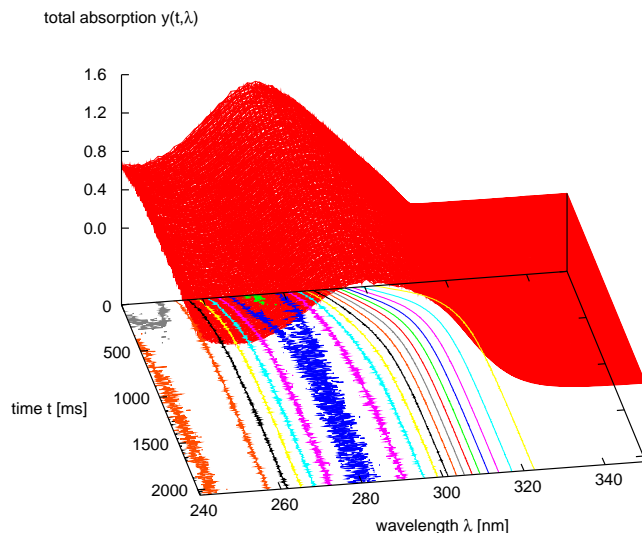


Fig. 3. Measured absorption spectra ($[\text{GSH}]_{t=0} = 720 \mu\text{mol/l}$): sampling rate: 5 ms, spectral resolution: 1 nm. The maximum absorption shifts from 270 nm to 300 nm due to a decreasing concentration of the educts MEL and GSH and an increase of the concentration of the products MEL · GSH and MEL · GSH₂. Contours are plotted in the λ - t -plane to visualize the three-dimensional shape of the data.

described by Beer's law of absorption:

$$I(\lambda) = I_0(\lambda)10^{-\varepsilon cd}. \quad (8)$$

The absorption is defined as $A(\lambda) := \log_{10}(I_0(\lambda)/I(\lambda))$, where I_0 is a reference measurement without absorbing reagents. This leads to the observation function Eq. (2). The spectra of MEL and GSH, i.e. the columns \mathbf{s}_1 and \mathbf{s}_2 of the matrix \mathbf{S} , are known with high precision from independent measurements of the respective components. The spectra of MEL · GSH and MEL · GSH₂ have to be estimated, because these substances cannot be isolated. Since we have only heuristic ideas of the shape of these spectra, we investigated the dependency of our results on the starting guesses for the spectra. Simulation studies showed that the algorithm converges even without any *a priori* information about the spectra, i.e. taking a wavelength independent constant as initial guesses. To optimize the computational effort we used the resulting spectra of these fits as starting guesses for the product spectra for all following fits.

The initial concentrations $[\text{MEL}]_0$ and $[\text{GSH}]_0$ are not known exactly and thus have to be estimated. The initial concentrations of the products MEL · GSH and MEL · GSH₂ vanish in the beginning of the reaction. However we decided to fit the first of them: a delay between the beginning

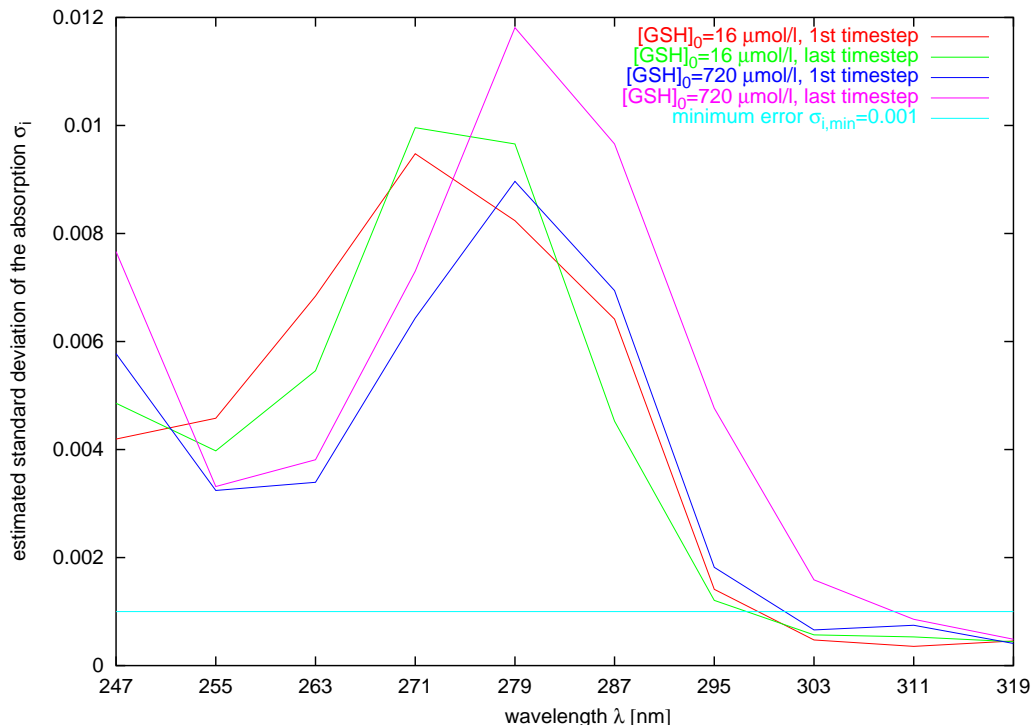


Fig. 4. Standard deviation of the data estimated with Rice's method for different datasets at $t = 0$. The cyan line shows the minimum error that is chosen, when the estimated value drops below this bound. The horizontal tic marks denote the values of the evaluated wavelengths.

of the reaction in the mixer and the first observed data point in the cuvette has to be taken into account. Thus the starting point of the reaction has to be estimated. We circumvented this non-trivial problem by fitting the initial MEL · GSH concentration: the time of the first observed data point $t = 0$ corresponds to the beginning of a reaction with slightly lower concentrations of the educts and a nonvanishing initial concentration of the product MEL · GSH. Therefore the initial value $[\text{MEL} \cdot \text{GSH}]_0$ is included as a fit variable. The concentration $[\text{MEL} \cdot \text{GSH}_2]$ increases quadratically in the beginning of the reaction and thus can be neglected for short times. Therefore its initial concentration is fixed to $[\text{MEL} \cdot \text{GSH}_2]_0 = 0$. Treating this parameter in the same way as $[\text{MEL} \cdot \text{GSH}]_0$ led to unreasonable results.

The offset $c(\lambda_j)$ accounts for small variations of the true intensity I_0 during every experiment in relation to its reference measurement recorded only once for all experiments. It is modeled by a quadratic function $c(\lambda_j) = a_0 + a_1 \cdot \lambda_j + a_2 \cdot \lambda_j^2$.

The standard deviation of the data σ_{ij} is estimated using Rice's method [Rice, 1984]. In order not to overemphasize data with very low estimated

errors, we defined a fixed minimum error bound of around 10% of the maximum estimated standard deviation. Figure 4 shows some examples for the estimated standard deviation.

We aim to estimate the underlying rate constants k_1 , k_{-1} , k_2 , and k_{-2} and the nonobservable spectra of MEL · GSH and MEL · GSH₂. To reduce the estimation error and to cover a sufficiently large area of the state space, we performed a multi-experiment fit. In this context, the initial concentrations and offset parameters are fitted individually for each experiment, while spectra and rate-constants are forced to be the same for each group of experiments.

The concentration of the product $[\text{MEL} \cdot \text{GSH}_2]$ scales with the initial concentration of GSH. Therefore one should use initial GSH concentrations of 240 $\mu\text{mol/l}$ and higher to estimate the $[\text{MEL} \cdot \text{GSH}_2]$ spectra accurately. However under these conditions the product MEL · GSH appears only for a short time. Thus, one has to obtain information about the MEL · GSH spectra by using additional experiments with $[\text{GSH}]_0$ around 60 $\mu\text{mol/l}$, where $[\text{MEL} \cdot \text{GSH}]$ stays on a high level.

Twelve experiments, i.e. three datasets (a–c) for each of four different initial GSH concentrations (60, 120, 240, 720 $\mu\text{mol/l}$), were analyzed. We fitted groups of three different concentrations (*triples*) simultaneously. Using quadruples would have forbidden to discover systematic errors resulting from different initial concentrations. All $4 \cdot 3^3 = 108$ possible triples were fitted. To reduce the probability of converging to local minima further, each triple was fitted eight times with different initial guesses of the rate constants. From these replicates the one with the minimum χ^2 was evaluated. An appropriate selection criterion discriminating between reliable results and local minima will be defined in the following section.

For the optimization process we used the multiple shooting algorithm. A simulation study showed, that this method converges to the global minimum twice as often as the initial value approach.

4. Results

This section presents the estimated rate constants and absorption spectra. As local minima of the

objective function always occur when dealing with nonlinear optimization, we had to define a criterion for the goodness of fit to select reliable results for further analysis. For this purpose we plotted the rate constants versus the χ^2 values of the objective function for all 108 fits. Several clusters are noticeable in these plots, as shown in Fig. 5 for the case of k_{-1} . Two of these can be identified as local minima, one with $\chi^2 \approx 40000$, the other one with $\chi^2 > 50000$. These were excluded from further analysis. The other clusters each have a consistent estimate of k_{-1} , while their χ^2 is between 11000 and 17000. Thus we considered these clusters to describe the global minimum. The splitting into four clusters for different initial GSH-concentrations is caused by a slight misspecification of the model that is discussed later.

4.1. Estimated time courses

Figure 6 shows the time courses of four representative wavelengths of a fit with a medium χ^2 value of 16466. The first time course shows an increasing absorption during the whole measurement, starting

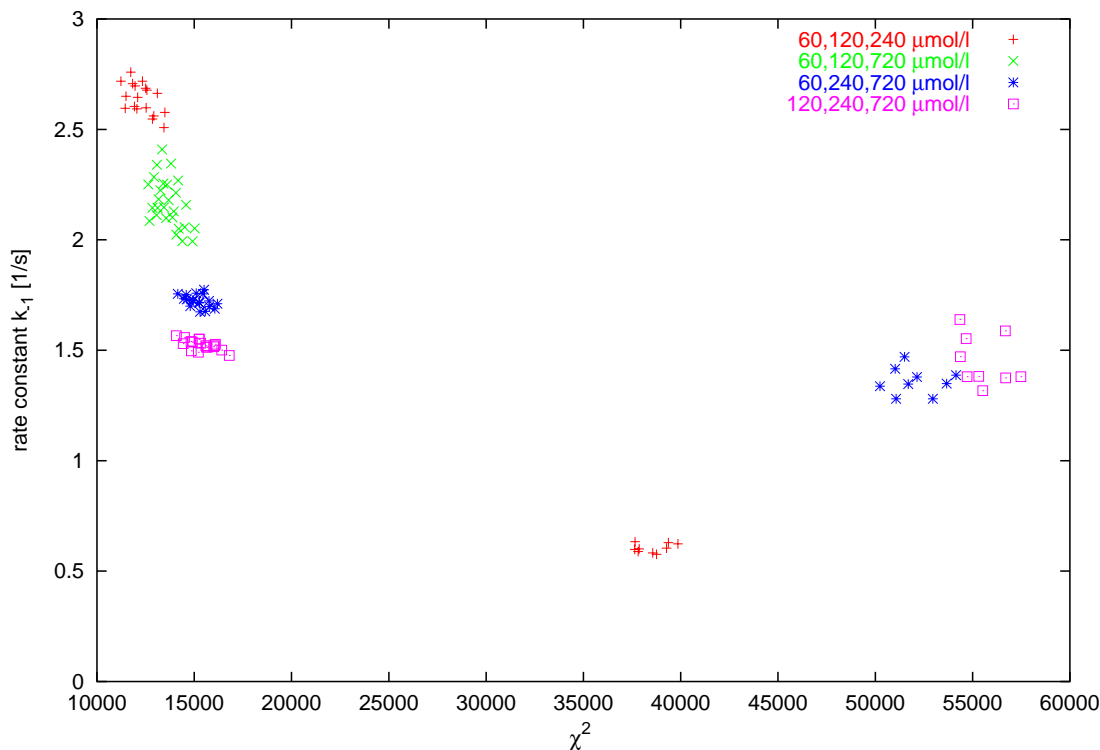


Fig. 5. Rate constant k_{-1} versus χ^2 value for all fits. The combinations of different initial concentrations are coded into the symbol types and colors. Several clusters are revealed. Two of them can be identified as local minima, one with $\chi^2 \approx 40000$, the other one with $\chi^2 > 50000$. The other clusters are considered to reflect the global minimum.

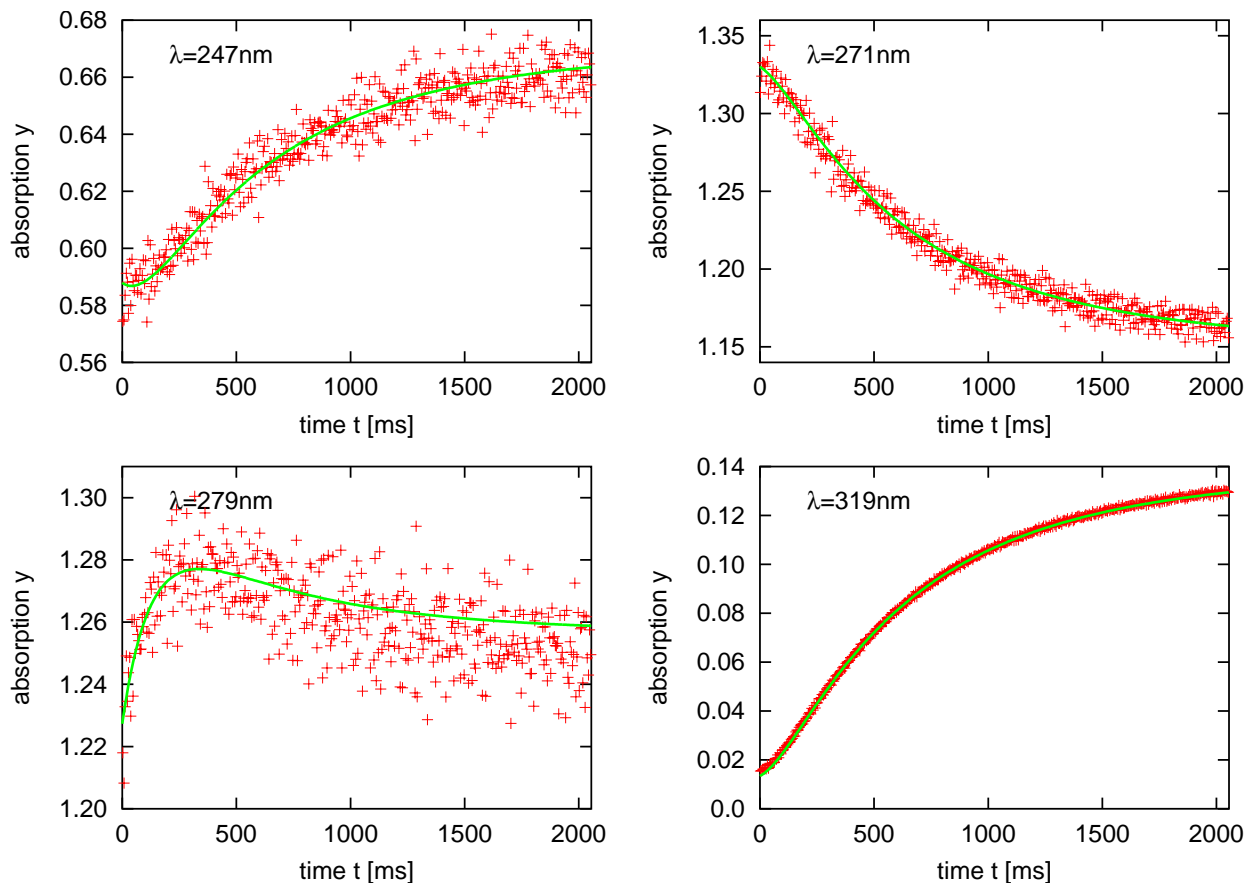


Fig. 6. Measured absorption (red crosses) and estimated time courses (solid lines) of one experiment at four representative wavelengths. The initial GSH-concentration was estimated to be $225.6 \mu\text{mol/l}$. For details refer to the text.

with a zero slope at $t = 0$ and ending in saturation. The absorption only increases by less than 15%. This fact points to a superposition of the spectra of the educts and products cancelling their contributions at this wavelength: the initial concentrations of the educts decline overlaid by the increasing concentrations of the products. The upper right time course corresponds to a wavelength near the maximum of the MEL spectra. This suits to the monotonous decline of the absorption. The third time series shows a steep ascent at $t = 0$, a saturation after 400 ms and a slight decrease afterwards suggesting the maximum absorption of the first product MEL · GSH lying near the wavelength of this time course. At the fourth time series, the small slope at $t = 0$ and the steady increase by a factor of seven correspond to the expected course of the MEL · GSH₂ concentration. Thus the spectral contributions of the other components seem to be negligible at this wavelength.

4.2. Estimated spectra of the individual components

The form of the estimated spectra can be used to check the quality of our method: our nonparametric model of the spectra does not assume any special shape whereas physics postulates smooth spectra. Also, the estimated spectra should be consistent for all different fits. To check this consistency requirement, all spectra are displayed in Fig. 7. The algorithm yields smooth spectra varying only in a narrow band for different experiments. Furthermore they correspond with the assumptions deduced from the time courses (cf. Sec. 4.1).

With these results it is easy to estimate the shape of the spectra in the original resolution: The observation function Eq. (2) is linear in the spectral parameters \mathbf{S} , i.e. when the time courses $\mathbf{x}(t)$ are known, the minimization of the objective function Eq. (3) can be solved analytically. This leads

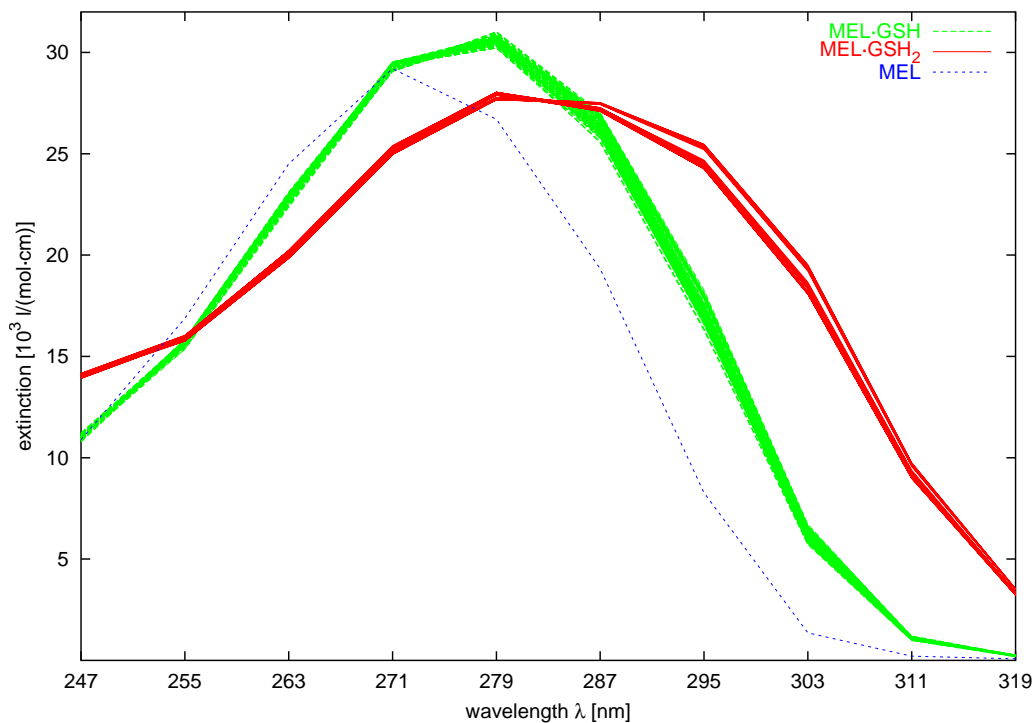


Fig. 7. Estimated spectra of all multiexperiment fits with $\chi^2 \leq 2\chi_{\min}^2$: MEL · GSH (green), MEL · GSH₂ (red). The scale was gauged by other experiments yielding an extinction coefficient for MEL (blue) at the maximum: $\varepsilon_{\text{MEL}}(\lambda = 272 \text{ nm}) = 29300 \text{ l}/(\text{mol}\cdot\text{cm})$ [Happersberger *et al.*, 1998a]. The algorithm yields smooth spectra varying only in a narrow band for different experiments.

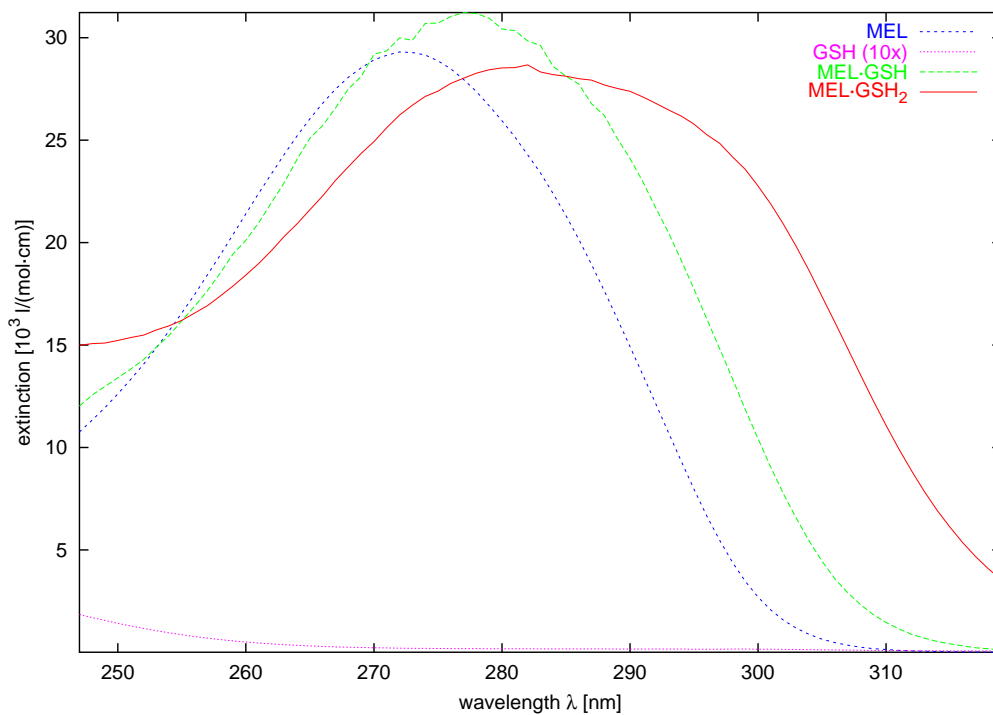
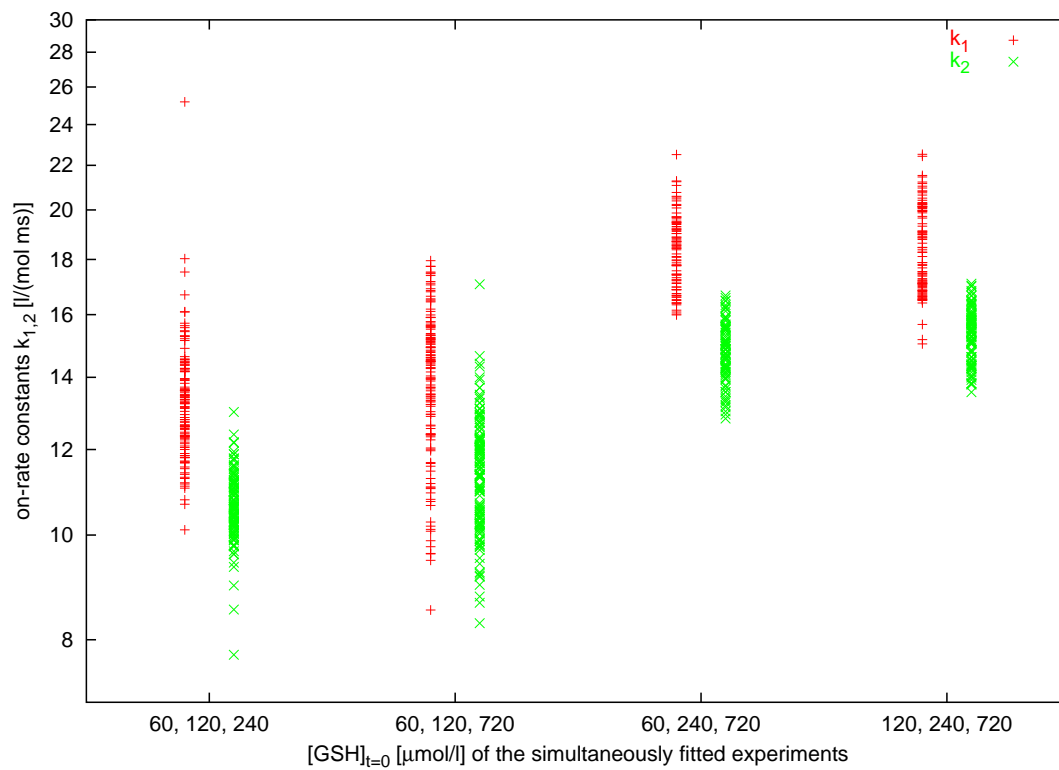
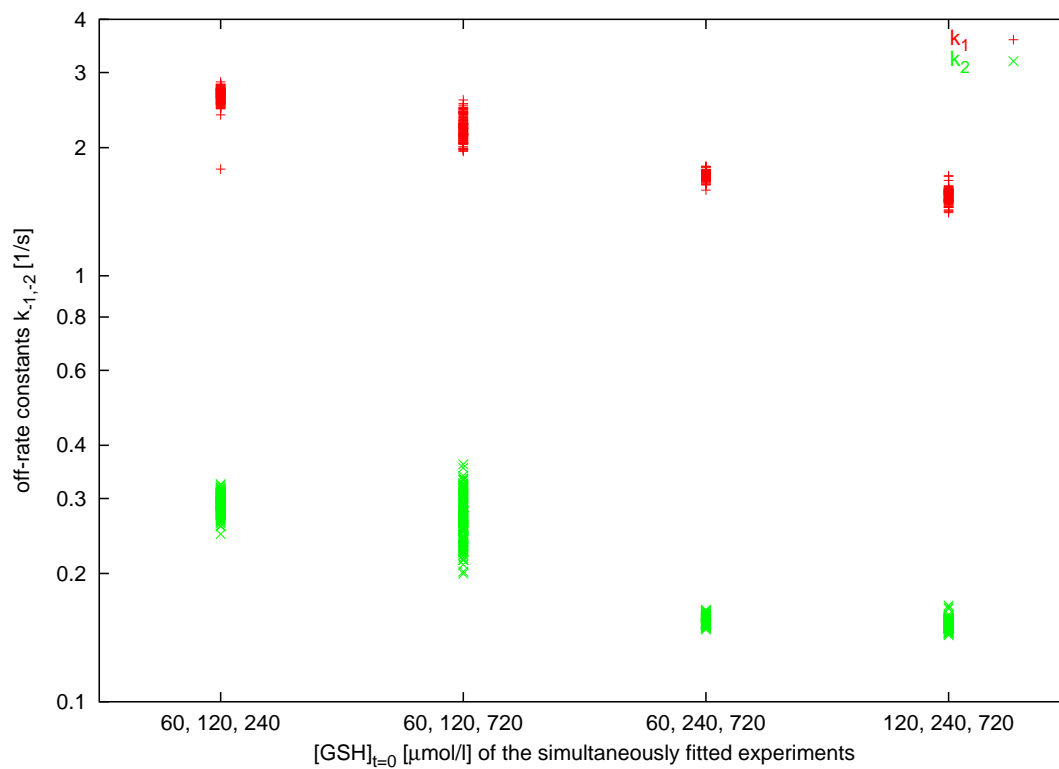


Fig. 8. High resolution spectra obtained as described in the text. The binding of GSH molecules to MEL (blue) causes two consecutive red-shifts to MEL · GSH (green) and MEL · GSH₂ (red).



(a)



(b)

Fig. 9. Estimated values of the (a) on-rates k_1 and k_2 , (b) off-rates k_{-1} and k_{-2} . For each rate constant, the estimated values do not vary by more than a factor of two. The variance is small enough to reveal a slight systematic dependency on the initial concentration of GSH.

to a simple matrix inversion for each wavelength. Based on the estimated rate constants, initial concentrations and offsets of the fit with the minimum χ^2 -value, we simulated the time courses of the concentrations of all reaction partners and estimated the spectra. Figure 8 shows these high resolution spectra.

It is of note that the maximum of the MEL spectrum appears to be red-shifted in two consecutive steps, from 272 nm to 278 nm upon binding of the first and to \approx 284 nm upon binding of the second GSH molecule. Furthermore, both shifts are accompanied by a broadening of the absorption band. These effects on the MEL chromophore upon formation of the mono- and bis-cysteinyl products MEL · GSH and MEL · GSH₂ may be interpreted tentatively. Both oxygen ligands of the arsen atom, when replaced by the more electropositive sulphur ligand, contribute to a red-shift by about 6 nm and to a slight broadening of the absorption band.

4.3. Estimated rate constants

Figure 9 shows the estimated (a) on-rate and (b) off-rate constants plotted versus the combination of simultaneously fitted datasets. For each rate constant, the estimated values do not vary by more than a factor of two. The variance is small enough to reveal a slight systematic dependence on the initial concentration of GSH, that was already shown for k_{-1} in Fig. 5.

This bias is related to $[\text{GSH}]_0$ in the following way: For experiments with high concentrations of GSH, the initial GSH-concentration tends to be underestimated in relation to the assumed value. However the dynamical model Eq. (7) contains only products of the on-rate constants and the educt concentrations. Thus an underestimation of $[\text{GSH}]_0$ leads to an overestimation of the on-rate constants k_1 and k_2 . As a consequence, Fig. 9(a) shows a slight increase of the on-rate constants for fits containing experiments with high $[\text{GSH}]_0$. The overestimation of the on-rates may lead to the slight decrease in the off-rates k_{-1} and k_{-2} which is visible in Fig. 9(b). This effect may necessitate extensions to the suggested model and remains to be analyzed in more detail.

For the left-most set of data in Fig. 9 the rate constants are not affected by this bias. Therefore we use the parameters obtained from these 18 fits as best estimates in the following discussion. Table 1

Table 1. Mean and sample standard error of all estimated parameters for selected multiexperiment fits with $[\text{GSH}]_0 = 60, 120, 240 \mu\text{mol/l}$.

Rate Constant	[Unit]	Mean	Standard Deviation
k_1	$[10^4 \text{ l}/(\text{mol}\cdot\text{s})]$	1.36	0.11
k_2	$[10^4 \text{ l}/(\text{mol}\cdot\text{s})]$	1.08	0.09
k_{-1}	$[\text{s}^{-1}]$	2.62	0.09
k_{-2}	$[\text{s}^{-1}]$	0.288	0.009

shows the mean estimates and the standard deviations calculated from their sample variance. The confidence limits calculated from the covariance matrix at the convergence point were much smaller. However, the former are supposed to reflect more realistically the uncertainties arising from varying experimental conditions.

From the rate constants we can draw a more detailed description of the overall reaction. The formation of the first MEL · GSH bond (k_1) is slightly faster than that of the second (k_2) by a factor of about 1.3. This small effect may reflect a steric hindrance of binding of the second molecule of GSH by the first one. The difference in the dissociation rates of the first and the second GSH is more pronounced: the stability of the bis-cysteinyl is tenfold higher than that of the mono-cysteinyl complex. This reflects the selectivity of MEL in modifying bis-cysteinyl residues over modification of single cysteines. This stabilization may originate from non-covalent interactions between the two GSH peptides which contribute to a chelating effect. The stabilization may be even more pronounced in proteins when the tertiary structure provides geometric restriction for the disulfide bond.

5. Discussion

For the analysis of nonlinear biochemical reaction dynamics by time-resolved optical spectroscopy, a standard method, which is often used, regards the information at specific wavelengths around the absorption peaks of the single components. This approach is intrinsically not able to reveal information about nonobservable absorption spectra. The simple version of this method linearizes the differential equations at $t = 0$. Using the first time steps of the data this method estimates an effective over-all rate constant of the linearized model, but does not regard the individual on- and off-rate constants.

Table 2. Reagents applied for the chemical modification of cysteinyl thiol groups in proteins and their respective reaction rates, if known. The reaction rates were determined at the following pH values: VP, pH 8 (pseudo first order) [Griffith, 1980]; NTCB, pH 7.4 [Degani *et al.*, 1970]; IAM, pH 8 [MacQuarrie & Bernhard, 1971; Nelson & Creighton, 1994; Oesterhelt *et al.*, 1977]; NEM, pH 6 [Brubacher & Glick, 1974; Gorin *et al.*, 1966; Oesterhelt *et al.*, 1977].

Reagent	Reaction Rate
4-Vinylpyridine (VP) (pseudo 1st order)	0.04 s^{-1}
2-Nitro-5-thiocyanobenzoic acid (NTCB)	$3 \text{ M}^{-1}\text{s}^{-1}$
Iodoacetamide (IAM)	$5 \text{ M}^{-1}\text{s}^{-1}$
Iodoacetic acid (IAA)	not known
N-Ethylmaleinimide (NEM)	$154 \text{ M}^{-1}\text{s}^{-1}$
1-Cyano-4-dimethylaminopyridine tetrafluoroborat (CDAP)	not known

Starting from the multiple shooting algorithm, we developed a method integrating the full system of nonlinear differential equations during the optimization process. This approach is able to identify the entire set of on- and off-rate constants characterizing the detailed nonlinear model of the reaction dynamics. Based on the idea of considering the absorption spectra as a multivariate nonparametric observation function, we can use nearly all spectral information of the data. This procedure reduces the estimation error in comparison to the standard method. But first of all it allows to estimate the absorption spectra of reaction products, that cannot be isolated and measured individually.

Understanding of protein folding pathways requires the determination of thermodynamic and kinetic properties of folding reactions and the characterization of intermediate structures. In order to analyze intermediates in disulfide bond mediated protein folding reactions it is necessary to modify irreversibly the cysteine thiol groups (*trapping*) in otherwise transient folding intermediates. Mono-thiol selective modification reagents (Table 2) are widely used in folding [Creighton *et al.*, 1995; Weissman & Kim, 1991] and unfolding [Li *et al.*, 1995] reactions as they quantitatively form stable thiol modified protein derivatives. In order to distinguish reduced cysteinyl residues from such cysteine residues involved in disulfide bonds the reaction rate of the trapping reaction should be significantly higher than the reaction rate of the thiol/disulfide exchange reaction. Reaction rates of intermolecular thiol/disulfide exchange reactions

between (low molecular weight) mono thiols, such as GSH, and cysteinyl thiol groups in proteins were assumed in the range of $10\text{--}20 \text{ M}^{-1}\text{s}^{-1}$ [Creighton & Goldenberg, 1984; Gilbert, 1997]. Both competitive reactions, thiol/disulfide exchange reaction and cysteinyl thiol modification (Table 2), showed reaction rates in the same order of magnitude. Therefore, in the absence of a high surplus of trapping reagent disulfide bond scrambling was possible during the trapping reaction [Weissman & Kim, 1991]. In contrast to the mono-thiol selective modification reagents, MEL exhibited several advantageous features that make it particularly useful as a trapping reagent.

We have determined the rates of both the formation and the dissociation of the complexes between MEL and GSH. Our results show that the stability of the bis-cysteinyl complex (dissociation rate k_{-1}) is about tenfold higher than that of the mono-cysteinyl complex (dissociation rate k_{-2}). This further emphasizes the ability of MEL to rapidly and selectively modify those cysteines of a protein which are arranged in pairwise close contact during a certain folding state of a protein. The approach of dynamical estimation is sensitive enough to discover a slight bias of the results for high GSH concentrations. This misspecification of the model remains to be studied in terms of slight extensions either of the spectral or the dynamical model.

A further result of this work is the identification of the absorption spectra of MEL · GSH and MEL · GSH₂, i.e. transient species that cannot be isolated and analyzed as pure compounds as they are not stable. This observation also opens the door to distinguish between mono- and bis-cysteinyl modifications of proteins during *in vitro* folding reactions.

Acknowledgment

We thank Drs. M. Glocker and W. Haehnel for their support.

References

- Bock, H. G. [1981] "Numerical treatment of inverse problems in chemical reaction kinetics," in *Modelling of Chemical Reaction Systems*, eds. Ebert, K., Deuffhard, P. & Jäger, W. (Springer, Heidelberg-NY), pp. 102–125.
- Bock, H. G. [1983] "Recent advances in parameter identification techniques for O.D.E.," in *Numerical*

- Treatment of Inverse Problems in Differential and Integral Equations*, eds. Deuffhard, P. & Hairer, E. (Birkhäuser, Basel), pp. 95–121.
- Brubacher, L. J. & Glick, B. R. [1974] “Inhibition of papain by N-ethylmaleimide,” *Biochemistry* **13**, 915–920.
- Creighton, T. E. & Goldenberg, D. P. [1984] “Kinetic role of a meta-stable native-like two-disulphide species in the folding transition of bovine pancreatic trypsin inhibitor,” *J. Mol. Biol.* **179**, 497–526.
- Creighton, T. E., Zapun, A. & Darby, N. J. [1995] “Mechanisms and catalysts of disulfide bond formation in proteins,” *Trends Biotechnol.* **13**, 18–23.
- Degani, Y., Neumann, H. & Patchornik, A. [1970] “Selective cyanylation of sulfhydryl groups,” *J. Am. Chem. Soc.* **92**, 6969–6971.
- Gilbert, H. F. [1997] “Thiol/disulfide exchange and redox potentials of proteins,” in *Bioelectrochemistry of Biomacromolecules*, eds. Lenaz, G. & Milazzo, G. (Birkhäuser, Basel), pp. 256–324.
- Gill, P. E., Murray, W. & Wright, M. H. [1981] *Practical Optimization* (Academic Press, London).
- Gorin, G., Martic, A. & Doughty, G. [1966] “Kinetics of the reaction of N-ethylmaleimide with cysteine and some congeners,” *Arch. Biochem. Biophys.* **115**, 593–597.
- Griffith, O. W. [1980] “Determination of glutathione disulfide using glutathione reductase and 2-vinylpyridine,” *Anal. Biochem.* **106**, 207–212.
- Happersberger, H. P. & Glocker, M. O. [1998] “A mass spectrometric approach to the characterization of protein folding reactions,” *Eur. Mass. Spectrom.* **4**, 209–214.
- Happersberger, H. P., Przybylski, M. & Glocker, M. O. [1998a] “Selective bridging of bis-cysteinyl residues by arsonous acid derivatives as an approach to the characterization of protein tertiary structures and folding pathways by mass spectrometry,” *Anal. Biochem.* **264**, 237–250.
- Happersberger, H. P., Stapleton, J., Cowgill, C. & Glocker, M. O. [1998b] “Characterization of the folding pathway of recombinant human macrophage-colony stimulating-factor beta (rhM-CSF β) by bis-cysteinyl modification and mass spectrometry,” *Proteins, Suppl.* **2**, 50–62.
- Hayashi, K. & Sakamoto, N. [1986] *Dynamic Analysis of Enzyme Systems — An Introduction* (Japan Scientific Societies Press, Tokyo).
- Horbelt, W., Timmer, J., Büchner, M., Meucci, R. & Ciofini, M. [2001] “Identifying physical properties of CO₂ laser by dynamical modeling of measured time series,” *Phys. Rev.* **E64**, 016222.
- Li, Y. J., Rothwarf, D. M. & Scheraga, H. A. [1995] “Mechanism of reductive protein unfolding,” *Nature Struct. Biol.* **2**, 489–494.
- MacQuarrie, R. A. & Bernhard, S. A. [1971] “Mechanism of alkylation of rabbit muscle glyceraldehyde 3-phosphate dehydrogenase,” *Biochemistry* **10**, 2456–2466.
- Nelson, J. W. & Creighton, T. E. [1994] “Reactivity and ionization of the active site cysteine residues of DsbA, a protein required for disulfide bond formation *in vivo*,” *Biochemistry* **33**, 5974–5983.
- Oesterhelt, D., Bauer, H., Kresze, G.-B., Steber, L. & Lynen, F. [1977] “Reaction of yeast fatty acid synthetase with iodoacetamide. 1. Kinetics of inactivation and extent of carboxamidomethylation,” *Eur. J. Biochem.* **79**, 173–180.
- Rice, J. [1984] “Bandwidth choice for nonparametric regression,” *Ann. Statist.* **12**, 1215–1230.
- Timmer, J., Rust, H., Horbelt, W. & Voss, H. U. [2000] “Parametric, nonparametric and parametric modelling of a chaotic circuit time series,” *Phys. Lett.* **A274**, 123–134.
- Weissman, J. S. & Kim, P. S. [1991] “Reexamination of the folding of BPTI: Predominance of native intermediates,” *Science* **253**, 1386–1393.