Computers and Chemical Engineering 187 (2024) 108734

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/cace

Mechanistic soft-sensor design for protein refolding processes based on intrinsic fluorescence measurements

Chika Linda Igwe [a,b], Florian Gisperg [b,c], Matthias Kierein [b], Eva Přáda Brichtová [b,c], Oliver Spadiut [b,c], Don Fabian Müller [a,b,*]

[a] Competence Center CHASE GmbH, Hafenstraße 47-51, Linz, 4020, Austria
[b] Research Area Biochemical Engineering, Institute of Chemical, Environmental and Bioscience Engineering, Technische Universität Wien, Gumpendorferstraße 1a, Vienna, 1060, Austria
[c] Christian Doppler Labor für Inclusion Body Prozessierung 4.0, Gumpendorferstraße 1a, Vienna, 1060, Austria

ARTICLE INFO

Keywords:
Intrinsic fluorescence
Soft-sensor
Protein refolding
Bioprocess monitoring
Process analytical technology (PAT)
Particle filter

ABSTRACT

In protein refolding processes the scarce availability of online measurements hampers effective process monitoring. In this work we developed a mechanistic soft-sensor for protein refolding based on online intrinsic fluorescence measurements of tryptophan and tyrosine. In validation experiments using two model proteins, lactate dehydrogenase (LDH) and galactose oxidase, the soft-sensor showed accurate estimates for the prediction of the total sum of folding products (NRMSE < 6.1%) by calculating the changing rate of the average emission wavelength. For refolding of the enzyme LDH it was possible to obtain separate predictions of native protein and insoluble aggregates. The soft-sensor design was further extended by a model-based observer approach using particle filtering to incorporate kinetic formulations as well as physical constraints.

The novel approach enabled the analysis of kinetic mechanisms during rapid reaction dynamics and can therefore be seen as an enabler to achieve a better understanding of kinetic refolding mechanisms.

1. Introduction

In recent years the bioprocessing industry has been encouraged by regulatory authorities to implement Quality by design (QbD) and PAT guidelines into their manufacturing processes (FDA, 2004). Model-based methods such as soft-sensors and digital twins have been identified as key enablers as they can enhance the information space, guide decision making and optimize the process control in real-time (Narayanan et al., 2020; Mears et al., 2017). Whereas many applications of advanced monitoring and control platforms have emerged in upstream processing, the downstream, including the inclusion body (IB) refolding process, lags significantly behind due to a lack of informative real-time measurements leading to reduced process knowledge (Pauk et al., 2021).

The vast majority of mechanistic refolding models in literature describe the *in vitro* protein folding mechanism based on few representative folding states including solubilized protein (S), intermediate conformations (I), native protein (N) and aggregates (A) (Kiefhaber et al., 1991; Dong et al., 2004b; Pan, 2015; Pauk et al., 2021; Igwe et al., 2023a). Despite the almost infinite number of different possible intermediate conformations, this serves as an efficient model simplification to represent the essential process dynamics in order to calculate meaningful yields and reaction rates (Pan, 2015). In more complex models further definitions of additional states are added for more detailed descriptions of folding intermediates (Cleland et al., 1992), the consideration of misfolded structures (M) for disulfide-bond containing proteins (Ryś et al., 2015), or to include states describing the cofactor assembly (IC or NC) if required for the bioactivity of a target protein (Cao and Li, 2011).

As the transition from the solubilized protein happens instantaneously after initialization of the refolding process, the transition rate from solubilized protein to the intermediates as well as the state S are often neglected (Dong et al., 2004a; de Bernardez Clark et al., 1998). The kinetics are illustrated as a flow diagram in Fig. 1. However, the reaction rates describing the transition between the other model states such as the refolding rate k_N or the aggregation rate k_A are (i) very complex, (ii) dependent on several factors such as the denaturant concentration (de Bernardez Clark et al., 1998) the buffer components and chemical additives (Dong et al., 2004a), and (iii) highly specific depending on the protein structure itself (Igwe et al., 2023a). The description of rates regarding cofactor interactions or the formation of misfolded protein states is even a greater challenge and hardly

* Corresponding author at: Competence Center CHASE GmbH, Hafenstraße 47-51, Linz, 4020, Austria.
E-mail address: donfabian.mueller@chasecenter.at (D.F. Müller).

https://doi.org/10.1016/j.compchemeng.2024.108734
Received 14 January 2024; Received in revised form 5 April 2024; Accepted 11 May 2024
Available online 14 May 2024

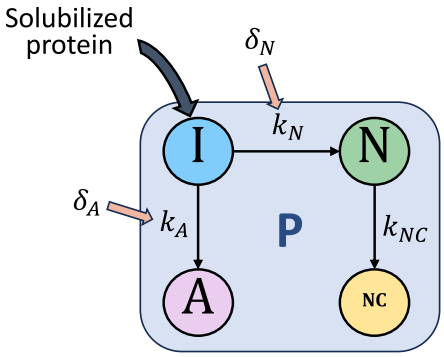


Fig. 1. Reaction scheme of protein refolding processes. The solubilized protein instantaneously reacts to folding intermediates (I), followed by a competitive reaction between aggregate (A) formation and folding to native protein (N) and NC in the case of cofactor bindings. The reaction rates towards the protein state are denoted as k (adapted from Kiefhaber et al. (1991)). The total protein concentration P is the sum of all folding states. δ_N and δ_A denote the modeled uncertainty on the reaction rates which is later utilized by the state-estimator.

applicable in a generic manner across different proteins (Ryś et al., 2015; Pauk et al., 2021).

In IB processing, in particular in protein refolding, information on the underlying reaction are mainly based on off-line or at-line measurements resulting in a significant time delay between sampling and retrieval of the processed results (Pauk et al., 2021; Igwe et al., 2023b) which poses challenges for real-time state estimation (Kager et al., 2018). For the monitoring of refolding processes those methods are usually based on chromatographic separation, e.g. size exclusion chromatography (SEC) or reversed-phase high-performance-liquid chromatography (RP-HPLC) (Pizarro et al., 2009; Igwe et al., 2023a), photometric assays (Humer et al., 2020) or spectroscopic techniques (e.g. circular dichroism or fluorescence measurements) (Sharma et al., 2022). Monitoring strategies solely based on off-line sampling often do not have the sufficient resolution to capture fast changing process dynamics (Wechselberger et al., 2013) as they are present in protein refolding. For a reliable system identification, informative measurements have to be available in a sufficient frequency in order to represent the fast mechanisms correctly. In signal processing, this principle is known as the *Shannon-theorem* (Jerri, 1977). Furthermore, the off-line analytics for quantification of protein conformations exhibit a noticeable amount of measurement uncertainty that might lead to insufficient signal quality (Igwe et al., 2023a).

Due to the aforementioned obstacles, soft-sensor algorithms pose an attractive tool to predict fast refolding process dynamics in the absence of sufficiently time-resolved measurements (Luttmann et al., 2012; Mohd Ali et al., 2015). However, reports on their successful implementation in protein refolding are scarce. Published attempts of real-time state estimation strategies applied to refolding processes had to deal with sparse and highly time-delayed measurements leading to the need of recalculation methods in order to correct the state estimate obtained in the past (Pauk et al., 2024).

Soft-sensors can be distinguished into *direct* calculation based algorithms, which directly apply functions to the measured data, and *indirect* model-based observers, that perform state-estimation by recursive prediction and correction steps (Luttmann et al., 2012). As an observer approach particle filtering (PF) is often applied, which is a well-established tool for the estimation of nonlinear systems (Mohd Ali et al., 2015; Müller et al., 2023). Informative online measurements are crucial for the effective application of these techniques.

Recently, we showed the potential of online intrinsic fluorescence measurements as an analytical tool for process development (Igwe et al., 2024). Intrinsic fluorescence based on the amino acids tryptophan (Trp) and tyrosine (Tyr) is a prominent tool to observe structural changes during protein folding pathways or ligand binding (Michaux et al., 2016; Lakowicz, 2006; Hellmann and Schneider, 2019). Its utilization has also been applied to IB processing chains especially to the monitoring of refolding dynamics using offline sampling (Sharma et al., 2022). Investigating the batch refolding of three different model enzymes we demonstrated that by using the changes in average emission wavelength (AEW) and the integral of the fluorescence intensity (F) valuable insights could be gathered regarding the underlying process dynamics posing a method of gathering sound process understanding. However, relative changes are often challenging to associate with actual process kinetics (Rüdt et al., 2017), thus we proposed the combination of online intrinsic Trp and Tyr measurements with mechanistic soft-sensing techniques for state estimation in protein refolding processes.

Here, we derived two soft-sensor approaches, a direct soft-sensing technique and an indirect PF observer in order to predict the folding states from the changing rate of AEW and the intensity F. We performed experimental validation using the two model enzymes lactate dehydrogenase (LDH) and galactose oxidase (GalOx). Under the assumption of functional relationships between (i) the changing rate of AEW and the total reaction rate ($\frac{dI}{dt}$) and (ii) the intensity F and the soluble protein concentration, the soft-sensors were applied to separately predict the states I, N and A. The validity of both assumptions are discussed with regard to the protein and process characteristics. We compared the prediction performance of the two soft-sensor approaches based on the normalized root mean square error (NRMSE) as a measure of accuracy. Our contribution shows that the resulting estimates pose a valuable data basis for a deeper analysis and modeling of kinetic mechanisms that would not be possible with off-line sampling strategies.

2. Material and methods

2.1. Processing of LDH IBs

Lactate dehydrogenase (LDH) from *Lactobacillus plantarum* was produced as IBs in *Escherichia coli* fed-batch fermentations, followed by isolation and washing as described by Igwe et al. (2023b). Isolated IBs were resuspended in solubilization buffer (0.15 M NaH_2PO_4, pH 6.0, 4 M guanidine hydrochloride (GuHCl)) and incubated under slight agitation for 2 h at room temperature (RT). After removal of insoluble substances by centrifugation (13,000 xg, 10 min, 4 °C) the solubilized protein was used for protein refolding in batch dilution approaches. Refolding was conducted using pre-cooled refolding buffer (0.15 M NaH_2PO_4, pH 6.0) at different protein concentrations for 2.5 h. The reaction was carried out in a volume of 1.5 mL or 3.0 mL at 5 °C under constant stirring.

2.2. Processing of GalOx IBs

Galactose oxidase originating from *Fusarium graminearum* was expressed as IBs in *E. coli* fed-batch cultivations as described by Igwe et al. (2024). Following isolation and washing, the isolated IBs were resuspended at a concentration of 100 g wet IB L^{-1} using solubilization buffer: 0.1 M NaH_2PO_4, pH 7.0, 6 M GuHCl. Then, 25 mM of dithiothreitol was added to initiate the rupture of disulfide bonds. After an incubation period of 2 h at RT, the solubilized protein was separated from insoluble substances by centrifugation (13,000 xg, 10 min, 4 °C). Refolding was initiated by dilution of the solubilized protein into pre-cooled refolding buffer (100 mM NaH_2PO_4, 5 mM cystamine, 1 M L-arginine at pH 7.4). Then, copper(II) was added as the cofactor to reach a concentration of 1 mM. The reaction was carried out in a volume of 1.5 mL for 2.5 h under constant stirring.

2.3. Online intrinsic fluorescence measurements

The measurement of intrinsic Trp and Tyr fluorescence was performed utilizing an FP-8550 Spectrofluorometer (Jasco, Tokyo, Japan) equipped with a multi-cuvette holder (Jasco, Tokyo, Japan) that allowed for temperature control and stirring of the cuvettes. Refolding procedures were carried out in 3-mL quartz fluorescence cuvettes (Starna GmbH, Germany) with magnetic stirrers, using volumes of either 1.5 mL or 3.0 mL. The cuvette holder temperature was set to 5 °C, and the stirring speed ranged between 500 and 800 rpm. Excitation of the sample occurred at 280 nm, and the emission spectrum was recorded within the range of 310 nm to 370 nm (λ_0 to λ_1). Photos and further information about the experimental setup are included in the supplementary information (SI figure 1-4). The intensity integral F was processed as the integral of the peak signal f from λ_0 to λ_1 as described by Eq. (1).

$$F(t) = \int_{\lambda_0}^{\lambda_1} f(t) d\lambda \tag{1}$$

The average emission wavelength (AEW) was calculated by the sum of $\lambda \cdot f$ from λ_0 to λ_1 (Eq. (2)).

$$AEW(t) = \frac{\sum_{i=\lambda_0}^{\lambda_1} (\lambda_i \cdot f_i(t))}{\sum f_i(t)} \tag{2}$$

The two extracted variables F and AEW were used to feed the soft-sensor.

2.4. Offline measurements of protein states

For both proteins (LDH and GalOx), the total soluble protein concentration and aggregate concentration were measured using different analytical tools. The definition as well as the calculation of protein states and their respective propagated state errors was conducted as described by Igwe et al. (2023a). The soluble protein concentration was measured by RP-HPLC using a Polyphenyl BioResolve-RP-mAb 2.7 µm 3.0 x 100 mm column (Waters Corporation, Milford, USA) on an UltiMate 3000 HPLC system (Thermo Fisher Scientific, Waltham, MA, United States of America). Bovine serum albumin at concentrations between 0.05–2.0 g L^{-1} was used to obtain a calibration curve (Kopp et al., 2020). The aggregated fraction was calculated based on the difference between soluble and insoluble protein concentration.

For the LDH the native protein concentration was determined via enzymatic activity measurements and correlations to SEC measurements for absolute values using LDH from porcine muscle (Cas No.: 9001-60-9) at concentrations between 0.05–1.0 g L^{-1} as reference standards. SEC was measured using an UltiMate 3000 HPLC system (Thermo Fisher Scientific, Waltham, MA, United States of America) and a BEH 200 A SEC 1.7 µm, 4.6 x 300 mm, 3.5 µm column (Waters Corporation, Milford, USA). Separation was carried out using isocratic elution (80 mM KH_2PO_4, pH 6.8, 250 mM KCl) for 18 minutes at a flow rate of 0.5 mL min^{-1} and a sample injection volume of 2 µL. The column oven was controlled at 25 °C and the absorbance was monitored at a wavelength of 214 nm and 280 nm.

The enzymatic activity was determined using a photometric assay measured in a TECAN Spark® microplate reader (Tecan Trading AG, Männedorf, Switzerland). The temperature was set to 30 °C and absorbance at a wavelength of 340 nm was recorded for 3 minutes. The sample was diluted in the reaction buffer (100 mM NaH_2PO_4, 0.425 mM nicotinamide adenine dinucleotide (NADH), 0.45 mM pyruvate) at a ratio of 30 % (v/v). Calculation was conducted using the extinction coefficient of NADH of 6.22 mM^{-1} cm^{-1} (Vanderlinde, 1985) as shown in Eq. (3).

$$vAc = \frac{V_t \cdot \frac{\Delta A}{\Delta t}}{V_s \cdot l \cdot \epsilon} \tag{3}$$

Table 1
Process parameters for refolding experiments of LDH and GalOx: initial protein concentration (P_0) and final protein concentration (P_{final}) in g L^{-1}, denaturant concentration (D) in mol L^{-1}, initial fluorescence intensity (F_0) in a.u. and the total shift in AEW (ΔAEW) in nm.

Process ID	P_0 [g L^{-1}]	P_{final} [g L^{-1}]	D [M]	F_0 [a.u.]	ΔAEW [nm]
LDH 1	0.15	0.45	0.2	2.4E5	0.67
LDH 2	0.3	0.44	0.32	4.2E5	0.82
LDH 3	0.05	0.26	0.13	1.0E5	0.62
LDH 4	0.57	0.57	0.10	5.3E5	0.97
LDH 5	0.57	0.57	0.10	5.6E5	0.67
LDH 6	0.20	0.20	0.10	2.8E5	0.83
LDH 7	0.21	0.21	0.10	2.8E5	0.93
LDH 8	0.21	0.21	0.10	2.9E5	0.75
LDH 9	0.21	0.21	0.10	2.8E5	0.69
GalOx 1	0.22	0.22	0.12	2.1E5	0.47
GalOx 2	0.22	0.22	0.12	2.7E5	0.85
GalOx 3	0.36	0.36	0.20	2.6E5	1.57
GalOx 4	0.36	0.36	0.20	3.0E5	1.70
GalOx 5	1.09	1.09	0.60	3.7E5	4.38
GalOx 6	1.09	1.09	0.60	3.7E5	4.08

The calculation is based on the time-dependent change in absorbance ($\frac{\Delta A}{\Delta t}$), where V_t represents the total volume of the reaction mixture in mL, V_s denotes the volume of the enzyme solution in mL, l signifies the length of the optical path in cm, and ϵ represents the extinction coefficient in mM^{-1} cm^{-1}. A single Unit of volumetric activity (vAc) is defined as the necessary enzyme for the conversion of 1 µmol of NADH per minute.

2.5. Refolding data-sets for experimental validation

For the experimental validation we conducted a set of experiments with the two model proteins LDH and GalOx. The parameters for the considered experiments are listed in Table 1. The processes LDH 1-3 were conducted in a pulse-based batch processing mode, where the solubilized protein was incrementally added over time such that the total protein concentration (P) was increased to a final concentration of P_{final}. The other processes were conducted in batch dilution mode and therefore contained constant P over time.

3. Models and soft-sensor design

3.1. Mechanistic reference model

As the reference model for protein refolding of LDH we consider the following nonlinear state space system as proposed by Kiefhaber et al. (1991) based on the concentration balances

$$\frac{dI(t)}{dt} = -k_N \cdot I(t) - k_A \cdot I(t)^n = -k_I \cdot I(t) \tag{4}$$

$$\frac{dN(t)}{dt} = k_N \cdot I(t) \tag{5}$$

$$\frac{dA(t)}{dt} = k_A \cdot I(t)^n \tag{6}$$

and algebraic equations for the definition of the reaction kinetics

$$k_N(t) = a_N \cdot (1 + D)^{b_N} + \delta_N(t) \tag{7}$$

$$k_A(t) = a_A \cdot (1 + D)^{b_A} + \delta_A(t) \tag{8}$$

where:

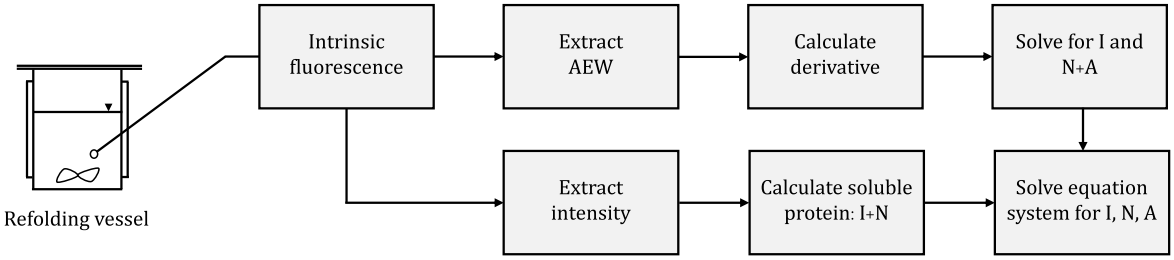


Fig. 2. Overview of the *direct soft-sensor* principle to estimate folding states I, N and A from intrinsic fluorescence measurements. In this study the data was obtained from a thermostated and stirred cuvette (SI figure 1-2). The concept, however, can be applied to larger reaction vessels.

$I(t)$ = intermediates in $\mathrm{g\,L^{-1}}$
$N(t)$ = native protein in $\mathrm{g\,L^{-1}}$
$A(t)$ = aggregates in $\mathrm{g\,L^{-1}}$
k_N = specific refolding rate in $\mathrm{h^{-1}}$
k_A = specific aggregation rate in $\mathrm{h^{-1}}$
k_I = specific total reaction rate in $\mathrm{h^{-1}}$
$\delta_N(t)$ = uncertainty for k_N in $\mathrm{h^{-1}}$
$\delta_A(t)$ = uncertainty for k_A in $\mathrm{h^{-1}}$
D = denaturant concentration in $\mathrm{g\,L^{-1}}$
n = order of aggregation

and a_N, a_A, b_N, and b_A being the kinetic model parameters. Identified kinetic parameter values were taken from Pauk et al. (2024) as initial guesses to perform open-loop model simulations for LDH and GalOx refolding. On the basis of the differential Eq. (4) an explicit formula for k_I can be formulated as follows:

$$k_I = k_N - I \cdot k_A \cdot I^n \tag{9}$$

For state estimation using the model-based PF observer described in Section 3.3, δ_N and δ_A were augmented into the state vector for the estimation of reaction rates.

In contrast to the other time dependent states, D was defined as a time invariant model parameter since it is assumed to be constant in a batch process. For the simulation of the pulsing experiments (LDH 1-3) however, D and I were altered through a discrete callback function when pulses are applied. The total protein concentration P is the sum of all different folding configurations of the protein. In this case:

$$P(t) = I(t) + N(t) + A(t) \tag{10}$$

For the GalOx refolding experiments we extended the model to include the cofactor binding mechanism. Therefore, another state equation was introduced describing the reaction from correctly folded protein (N) to the protein cofactor complex (NC)

$$\frac{dNC(t)}{dt} = k_{NC} \cdot N(t) \tag{11}$$

and Eq. (5) has to be extended to

$$\frac{dN(t)}{dt} = k_N \cdot I(t) - k_{NC} \cdot N(t) \tag{12}$$

to respect the conservation laws.

3.2. Soft-sensor principle 1: Direct calculation

A basic scheme of the *direct soft-sensor* operating principle is shown in Fig. 2. In general, the direct soft-sensor algorithm is based on calculations of the average emission wavelength (AEW) and the integral of fluorescence intensity (F).

The soft-sensor builds up on the following assumptions.

1. The total reaction rate ($\frac{dI}{dt}$, Eq. (4)) is functionally correlated to the changing rate of AEW.
2. The concentration of soluble proteins ($I + N$) is proportional to the intensity (F, Eq. (1)). When proteins aggregate they precipitate out of solution causing F to decrease.
3. The initial concentrations at $t = 0$ for N and A are 0 and the sum of all folding states, the total protein concentration (P, Eq. (10)) is known.

If the first assumption holds, the pace of the folding reaction can be obtained by numerically calculating the time derivative of AEW. In this work we used the *Savitzky–Golay differentiation filter* (Savitzky and Golay, 1964) with a window size of 15 and an order of 5 as well as *locally weighted scatterplot smoothing* (Loess) (Cleveland and Devlin, 1988). After obtaining the value for β_1 by linear regression of the changing rate of average emission wavelength ($\frac{dAEW}{dt}$) to the total reaction rate ($\frac{dI}{dt}$), the obtained function can be integrated to solve for the state $I(t)$.

$$I(t) = \int_{t0}^{t} (\beta_1 \cdot \frac{dAEW}{dt}) dt = \int_{t0}^{t} \frac{dI}{dt} dt \tag{13}$$

The value of the correlation coefficient β_1 was identified to be $\beta_1 = 0.174 \pm 0.025$ for the LDH and the same correlation could be used for the GalOx (SI figure 5). If the second and third assumption hold, we can additionally describe the intensity integral F as a function of soluble protein ($I + N$) by

$$I(t) + N(t) = \beta_2 \cdot F(t) = \frac{P_0}{F_0} \cdot F(t) \tag{14}$$

Since the initial concentrations of N and A are 0 the soluble protein concentration is equal to I, which is the total protein concentration at the start of the process (P_0). Therefore, the initial intensity F_0 can be correlated to P_0 by

$$\beta_2 = \frac{P_0}{F_0} \tag{15}$$

which allows Eq. (14) to be solved for the sum of I and N. Finally, we can calculate all three state variables I, N and A by solving the linear equation system composed of Eqs. (10), (13) and (14), compactly expressed in matrix form as

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} I \\ N \\ A \end{bmatrix} = \begin{bmatrix} P \\ \beta_2 \cdot F \\ \int_{t0}^{t} (\beta_1 \cdot \frac{dAEW}{dt}) dt \end{bmatrix}. \tag{16}$$

In summary, the time-invariant parameters of the direct soft-sensor formulation are

1. the initial protein concentration (P_0)
2. the correlation coefficient (β_1)

and the input variables of the soft-sensor are $F(t)$ (Eq. (1)) and $AEW(t)$ (Eq. (2)). The method does not rely on kinetic model parameters and is therefore easy to transfer between different proteins.

3.3. Soft-sensor principle 2: Model-based particle filter

The second option to calculate unknown state variables is an *indirect* estimation through a model-based observer. To estimate the model states from the intrinsic fluorescence measurements, we applied a particle filter observer (PF). The PF requires the definition of a process

model and was combined with the mechanistic model defined in Section 3.1. The algorithm employs a set of particles, each representing a potential state hypothesis, and assigns weights to these particles based on their likelihood given observed measurements. After initialization, it consists of a recursive iteration of prediction, correction and resampling steps. In the prediction step, the process model is predicted one time step into the future. After that, the correction step is performed, where the true measurements ($\frac{dAEW}{dt}$ and F) are compared with the simulated measurements from the prediction step. Based on this deviation the model states are corrected which enables state estimation. The implementation of the PF was adapted from Müller et al. (2023), where the algorithm is presented in more detail.

According to the model (Section 3.1) the augmented state vector included 5 states and was defined as

$$x = [I, N, NC, A, \delta_N, \delta_A]^T. \tag{17}$$

The condensed form of the nonlinear state-space equation can be formulated as

$$\frac{d}{dt}x = \begin{bmatrix} dI/dt \\ dN/dt \\ dNC/dt \\ dA/dt \\ \delta_N \\ \delta_A \end{bmatrix} = \begin{bmatrix} -k_N \cdot I(t) - k_A \cdot I(t)^n \\ k_N \cdot I(t) - k_{NC} \cdot N(t) \\ k_{NC} \cdot N(t) \\ k_A \cdot I(t)^n \\ 0 \\ 0 \end{bmatrix}. \tag{18}$$

The output function (Eq. (19)) to calculate the two measured outputs from the model states was defined by rearranging the direct calculation formulas Eqs. (13) and (14) containing the assumptions about the system defined in Section 3.2.

$$y = \begin{bmatrix} \beta_2 \cdot (I(t) + N(t)) \\ \beta_1 \cdot (-k_N \cdot I(t) - k_A \cdot I(t)^n) \end{bmatrix} \tag{19}$$

Here, the first entry corresponds to the fluorescence intensity integral F and the second entry corresponds to AEW. In principle, the state-estimation works by repeated iterations of predicting the future system state using Eq. (18) and subsequently correcting the state prediction by the two measurements using Eq. (19).

The PF was configured as follows. For initialization, 500 particles were initialized using a Gaussian initial state distribution with a mean of

$$\mu(x_0) = [P_0, 0, 0, 0, 0]^T \tag{20}$$

and a covariance matrix

$$cov(x_0) = 10^{-5} \cdot \mu(x_0) \tag{21}$$

assuming that the initial protein concentration is well known with a relative uncertainty of 10^{-5} and solely in the form of intermediates. At the start of the process, solubilized protein is added to the refolding buffer, where it immediately transitions into the intermediate folding state I, since the denaturant GuHCl is diluted. Therefore, the initial mass of I is entirely defined by the initial protein concentration P_0. All other folding states (N, NC, A) as well as the uncertainty values on the folding rates (δ_N, δ_A) are assumed to be 0 at the start of the process.

Gaussian distributions were also used for additive state transition noise with a covariance matrix of

$$cov(x) = diag[0, 0, 0, 0, 0.002, 0.002] \tag{22}$$

and for the measurement noise with a covariance matrix of

$$cov(y) = diag[3 \cdot 10^4, 0.001]. \tag{23}$$

State transition noise was only defined for the propagation of δ_1 and δ_2 (Eq. (22)) in order to restrict the filter to the estimation of only the reaction rate kinetics. Although there is no direct additive noise defined for the concentration states (I, N, A, NC), the adaptations of k_N and k_A propagate through the model and in turn influence the dynamics of the concentrations. Since the measurements F and $\frac{dAEW}{dt}$ exhibit varying orders of magnitudes, the measurement noise had to be tuned accordingly (Eq. (23)).

Table 2
Mean NRMSE values compared to the measured AEW for each applied methods of numerical differentiation after reintegration of the derivative.

Differentiation method	Mean error for LDH	Mean error for GalOx
Finite differences	1.32%	0.78%
Savitzky–Golay	1.11%	1.03%
Loess	1.31%	3.15%

3.4. NRMSE calculation as soft-sensor performance measure

As a measure of soft-sensor and open-loop model performance the normalized root mean square error (NRMSE) was calculated using normalization towards the maximum value of the respective measurement vector y_p (Eq. (24)).

$$\mathrm{NRMSE}(x_p) = \frac{\sqrt{\frac{1}{d}\sum_{k=1}^{d}(y_{p,k} - x_{p,k})^2}}{\max(y_p)} \tag{24}$$

where:

d	total number of measured points
k	iterator to sum over all measured points
x_p	vector of predicted concentrations
y_p	vector of measured concentrations
$x_{p,k}$	predicted concentration at time k
$y_{p,k}$	measured concentration at time k

3.5. Software implementation

All scientific computations were conducted in Julia version 1.9. The mechanistic models were implemented using ModelingToolkit.jl (Ma et al., 2022) and the numerical simulations of the resulting differential equation systems were conducted using DifferentialEquations.jl (Rackauckas and Nie, 2017). Nonlinear uncertainty propagation of the soft-sensor estimation was performed using MonteCarloMeasurements.jl (Carlson, 2020) and particle filtering was carried out using LowLevelParticleFilters.jl. The developed algorithms, the code for data analysis and the raw data used in this paper is publicly available under the following repository: https://github.com/dfabianus/FLUMO.jl.git.

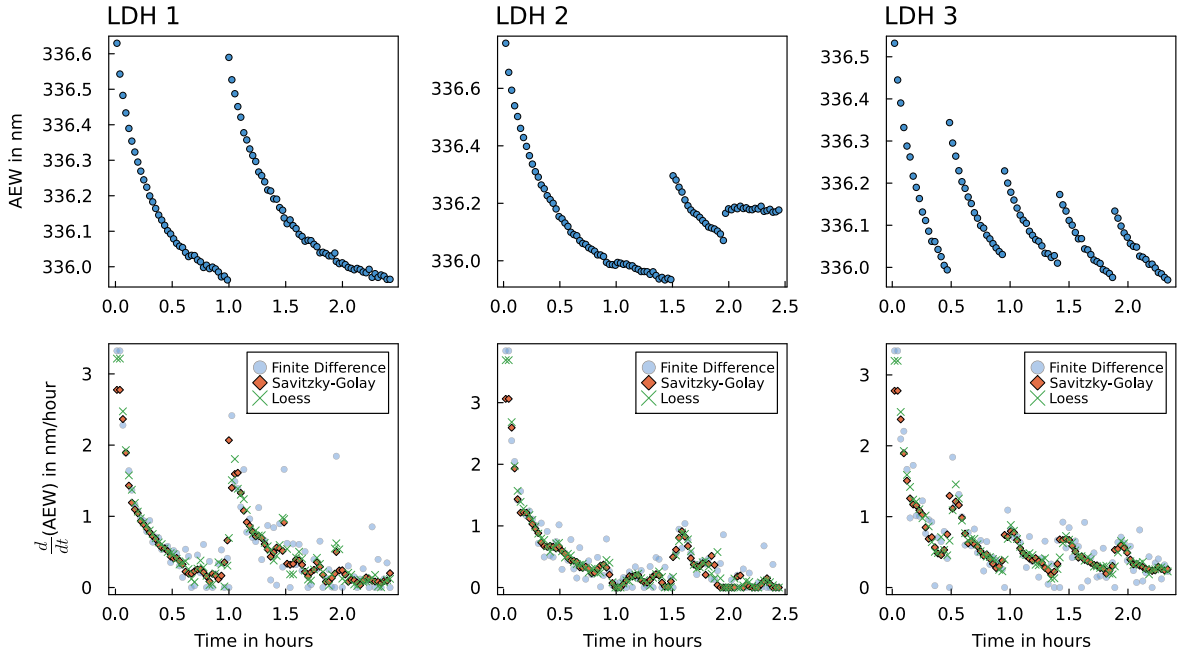
4. Results and discussion

4.1. Calculation of the total reaction rate using the AEW

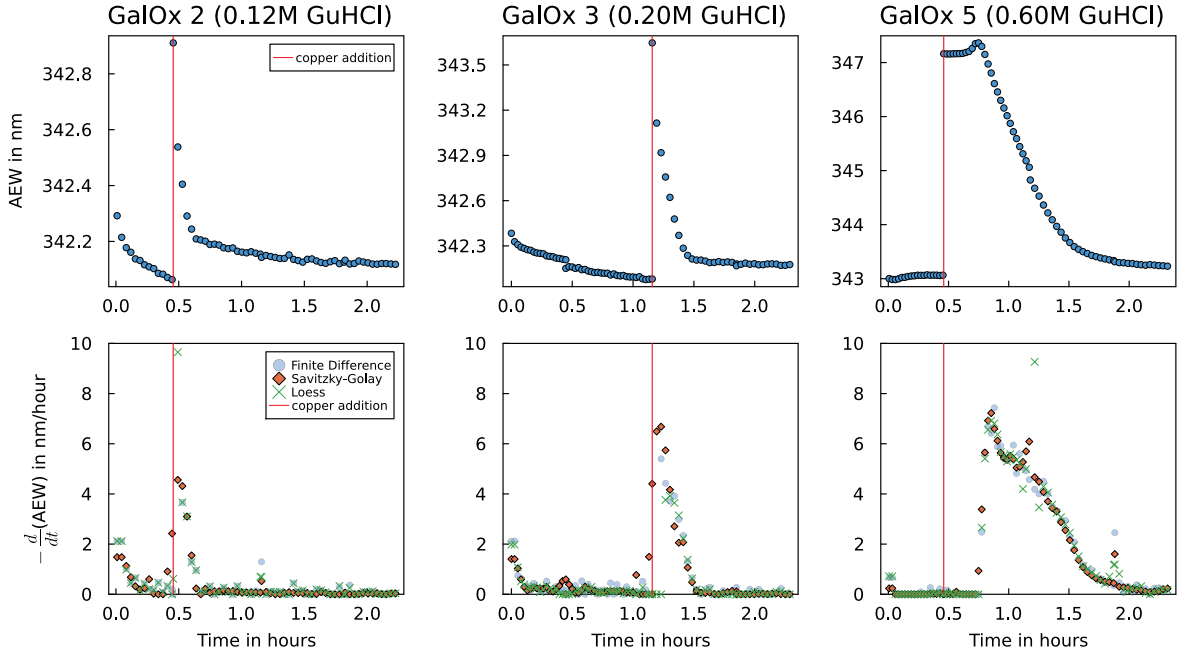
In order to obtain the total folding rate of the reactive intermediate species (I) we assumed a functional relationship between the changing rate of AEW and the reaction rate. In Fig. 3(a) the numerical differentiation of AEW is illustrated for the processes LDH 1-3 and in Fig. 3(b) for the processes GalOx 2, GalOx 3 and GalOx 5, respectively, which were executed with varying amounts of denaturant (GuHCl). The numerical derivative in $\mathrm{nm\,h^{-1}}$ was calculated using a finite-difference approximation. Since numerical differentiation is a noise amplifying operation (Bayer et al., 2020), the derivatives obtained by finite-difference approximation show a higher noise than the raw AEW data (Fig. 3(a)).

Therefore, we applied two methods for derivative smoothing, (i) the *Savitzky-Golay* differentiation filter (Savitzky and Golay, 1964) and (ii) the *Loess regression* (Cleveland and Devlin, 1988), which significantly reduced the noise levels. In order to check for the correctness of the approach we re-integrated the obtained derivatives for all considered experiments (Table 1) by means of a cumulative sum, determined the NRMSE with respect to the raw AEW data and calculated the mean over all experiments. The results are displayed in Table 2.

For the LDH refolding they show the highest error for the finite-difference approximation (1.32%) and the lowest error for the *Savitzky–Golay* differentiation filter (1.11%). Although, for GalOx refolding the



(a) Differentiation of AEW in LDH refolding processes (LDH 1-3).



(b) Differentiation of AEW in GalOx refolding processes (GalOx 2,3,5).

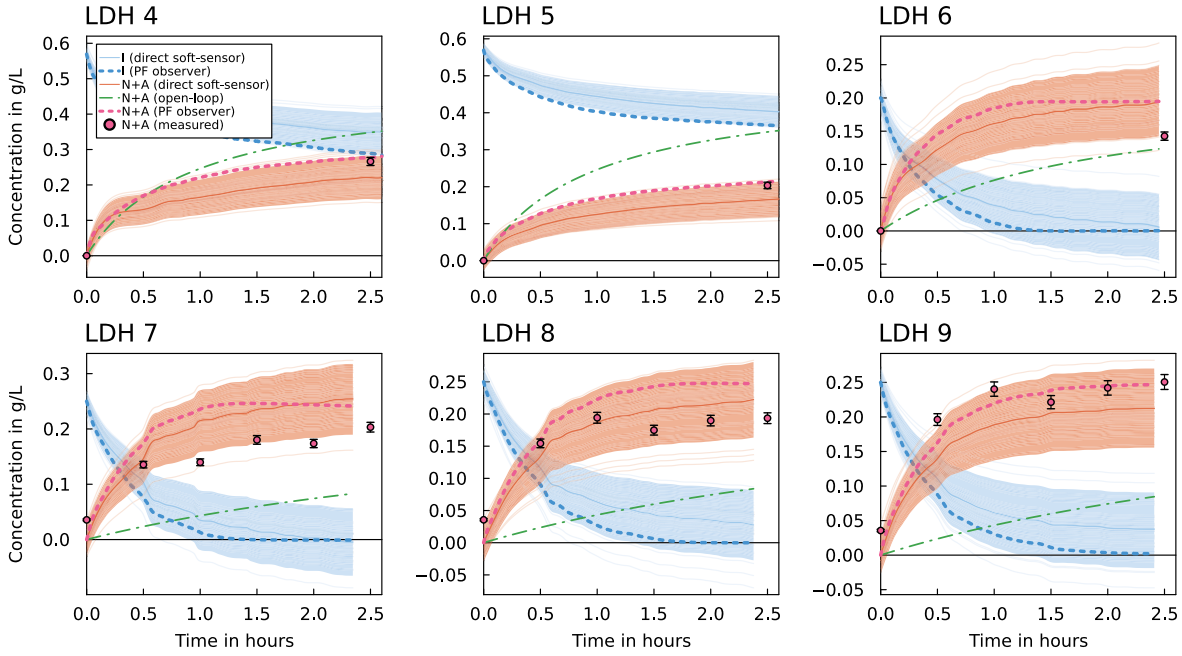
Fig. 3. Numerical differentiation of the AEW using finite difference approximation, a *Savitzky–Golay* differentiation filter and *Loess* regression for three distinct refolding processes per protein to obtain a measure of total refolding reaction rate $\frac{dI}{dt}$. The data points for AEW were obtained by applying Eq. (2) to the fluorescence spectra.

raw finite-difference approximation were the lowest (0.78%) for better comparability we also used the derivatives obtained by the *Savitzky–Golay* differentiation filter for further processing of the GalOx data and calculation of folding reaction rates for all experiments.

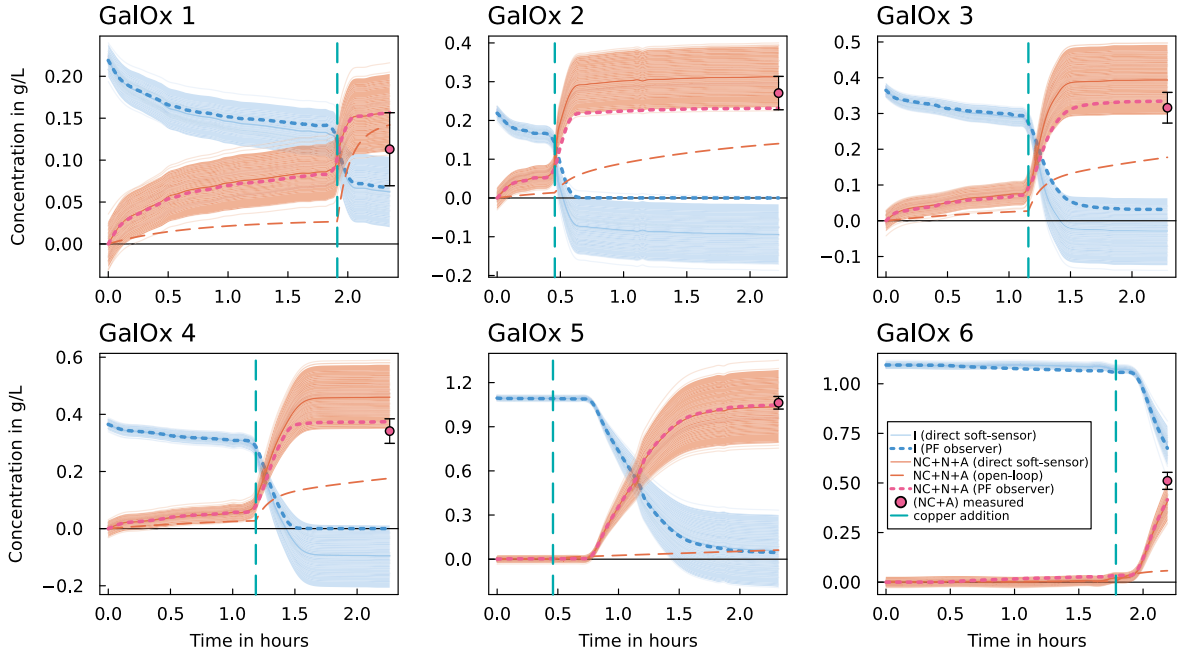
4.2. Soft-sensing of the total folding reaction

In this section we analyze the prediction accuracy of the soft-sensor based on direct calculation derived in Section 3.2 (the direct soft-sensor) as well as the prediction accuracy of the model-based particle filter derived in Section 3.3 (the PF observer) and compared them to the open-loop reference model and to the measured data of folding states.

Fig. 4(a) shows the time-resolved estimates for $I(t)$ and the sum of $N(t)$ and $A(t)$ that is representing the total sum of folding products for the LDH refolding processes 4-9. The results show that for the processes with single end-point measurements (LDH 4-6) the direct soft-sensor as well as the PF observer estimates ($N + A$) predicted the final state measurement better than the open-loop model with the exception of LDH 6, where the final state was more accurately predicted by the open-loop model. However, in all processes with time-resolved measurements (LDH 7-9) the direct soft-sensor and the PF observer followed the state trajectories more accurately compared to the prediction of the open-loop model. An overview of the NRMSE values is displayed in Table 3.



(a) LDH validation processes 4-9. For the processes 7-9 time-resolved offline measurements were measured every 30 min. For the processes 4-6 only the end-point was measured. The displayed measurement points were obtained by summing up the distinct measurements for N and A.



(b) GalOx validation processes (GalOx 1-6). *NC* represents the native protein cofactor complex. The main reaction only takes place after copper addition as the cofactor depicted by the vertical dashed line.

Fig. 4. Soft-sensing of the total folding reaction for (a) the LDH folding products (N+A) and (b) the GalOx folding products (NC+N+A). The ribbons display 1σ standard deviation obtained by nonlinear uncertainty propagation as a result of the correlation function uncertainty shown in SI figure 5.

For the GalOx refolding processes, Fig. 4(b) shows the estimates for $I(t)$ and the sum of $A(t)$, $N(t)$ and $NC(t)$, which represents the native protein cofactor complex that is being formed after the addition of copper as the cofactor (Whittaker, 2005), here indicated by the vertical dashed line. The end-point measurements mostly show accurate predictions by the direct soft-sensor (NRMSE = 6.1%) and the PF observer (NRMSE = 5.3%).

Overall, both soft-sensor algorithms show good prediction accuracy for the total folding dynamics of both LDH refolding (NRMSE < 6.1%) and GalOx refolding (NRMSE < 8.7%).

However, as can be seen in Fig. 4 the direct soft-sensor estimate for I can reach concentrations below $0\,\mathrm{g\,L^{-1}}$ for the LDH and the GalOx, which is unfeasible. This constraint is not respected by the direct soft-sensor since the turnover rate $\frac{dI}{dt}$ is based on a direct static relationship to the derivative of AEW (Eq. (13)) and not described by kinetics as derived for the open-loop model (Eq. (4)).

Comparing the direct soft-sensor with the PF observer we see that the PF observer estimates for I do not reach values below $0\,\mathrm{g\,L^{-1}}$, since in contrast to the direct soft-sensing approach, the kinetic model of the PF observer prevents the state estimate from reaching unfeasible states. The results of Table 3 also show that the PF observer in general exhibit

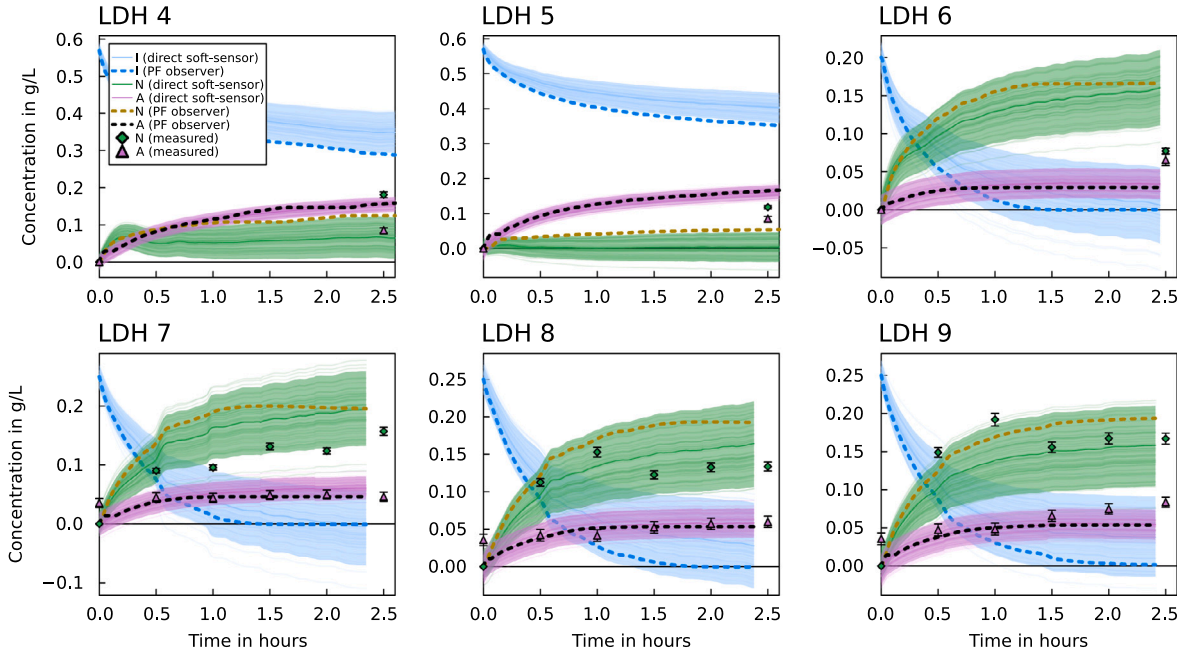


Fig. 5. Soft-sensing of separate folding states in LDH refolding processes. Separating $N + A$ into N and A of LDH 4-9 by incorporation of the fluorescence intensity (F). Compare to Fig. 4(a) for the sum of N and A.

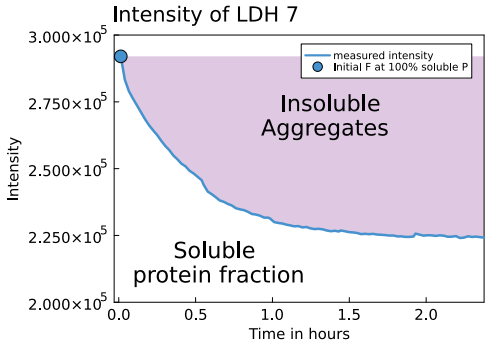


Fig. 6. Intensity of process LDH 7 declines over time which is assumed to be predominantly due to the formation of insoluble aggregates. LDH 7 was taken for illustration. The effect was observed in the other LDH processes as well.

slightly lower NRMSE values than the direct soft-sensor which could be due to the prevention of unfeasible states by the additional constraints. This extension however comes with an increased complexity since additional kinetic model parameters have to be considered to define the PF observer, which has to be weighed up against the gain of accuracy. The direct soft-sensor in contrast does not rely on any kinetic parameters and therefore requires less prior knowledge of the system making it more transferable to other proteins.

4.3. Soft-sensing of separate folding states

Under the assumption that the measured fluorescence intensity (F) is proportional to the soluble protein concentration, the total folding products ($N + A$) can be further differentiated into native protein (N) and insoluble aggregates (A) as stated in Section 3.2.

Fig. 6 shows, that during refolding of LDH, F tends to decrease over time which is assumed to be partially due to the formation of aggregates that are precipitating out of solution and thus are only partially contributing to the fluorescence signal (Igwe et al., 2024). To obtain the correlation between fluorescence signal and protein concentration (β_2) the known protein concentration was divided by the fluorescence at the initial time point t_0 (Eq. (15)) where 100% of the added solubilized protein is assumed to be dissolved. The soluble fraction of the protein concentration ($I + N$) is then calculated by Eq. (14). By incorporating this equation into the soft-sensor equation system, it can be solved for all three states I, N and A by only providing the initial total protein concentration P_0, the correlation coefficient between $\frac{dAEW}{dt}$ and $\frac{dI}{dt}$ (β_1) and the fluorescence measurement signals $F(t)$ and $AEW(t)$.

Fig. 5 shows the separated estimation of N and A for the LDH processes 4–9 in contrast to the total estimation $N + A$ as shown in Fig. 4(a). Although LDH 4-6 show a relatively high uncertainty in the direct soft-sensor estimations for N and A the time-resolved measurements of LDH 7-9 are still accurately captured as can be seen in Fig. 5 leading to an overall NRMSE for state N of 8.7% and for state A of 4.5% (Table 3). The PF observer overall shows a similar prediction accuracy compared to the direct soft-sensor but prevents state I to fall below $0\,\mathrm{g\,L^{-1}}$ (Fig. 5). The NRMSE values for the PF observer are comparable to the direct soft-sensor (Table 3).

The result shows that the hypothesis of F being functionally correlated to soluble protein $I + N$ can indeed be used in LDH refolding processes to distinguish the prediction of $N + A$ into the single folding states. However, the high prediction error in some of the experiments (LDH 4-7) also indicate that there might be more complex effects which influence the fluorescence intensity and thus deteriorate the state estimation.

Fig. 7 shows that it is not possible to directly transfer this assumption to other refolding reactions for proteins including cofactors such as shown here for the GalOx. After the addition of the cofactor, the intensity suddenly drops, which in turn leads to a jump of the estimates to $0.3\,\mathrm{g\,L^{-1}}$ for A and to below $0\,\mathrm{g\,L^{-1}}$ for $NC + N$. Here, the above mentioned hypothesis does not hold, which is probably due to additional quenching effects that are induced by the addition of copper into the solution leading to infeasible solutions for $NC + N$ and A (Weiner et al., 1977). In contrast to LDH refolding (Fig. 6) the intensity is not predominantly influenced by aggregate formation anymore but mainly affected by quenching (Fig. 7). The same effect was observed for all other experiments of GalOx refolding that are depicted in Table 1.

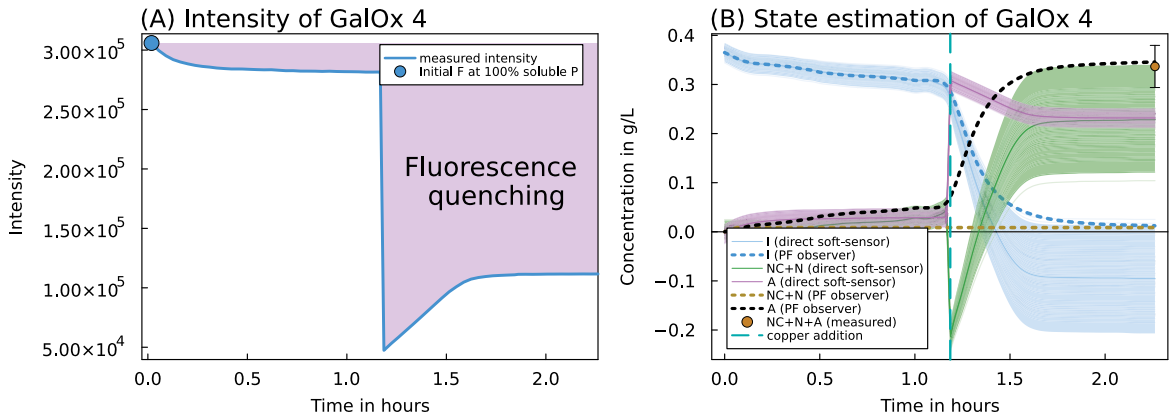


Fig. 7. Soft-sensing of separate folding states in GalOx refolding processes. Separating $NC+N+A$ into $NC+N$ and A by incorporation of the intrinsic fluorescence intensity (A) is exemplarily shown for GalOx 4 which leads to infeasible results probably due to quenching effects induced by the cofactor addition (B).

Table 3
NRMSE values of the soft-sensor and open-loop model estimates. Mean and standard deviation of NRMSE are specified for LDH and GalOx experiments respectively.

LDH: NRMSE values in %			
State	Open-loop model	Direct soft-sensor	PF observer
N+A	12% ± 6.1%	6.1% ± 3.9%	5.3% ± 4.6%
N	18% ± 6.6%	8.7% ± 6.4%	8.2% ± 6.4%
A	15% ± 5.9%	4.5% ± 2.2%	4.7% ± 2.8%
GalOx: NRMSE values in %			
State	Open-loop model	Direct soft-sensor	PF observer
NC+N+A	83% ± 24%	12% ± 9.2%	8.7% ± 7.3%
NC+N	n.a.[a]	n.a.	n.a.
A	n.a.	n.a.	n.a.

[a] For the GalOx there is no reliable estimation data available for separate prediction of NC+N and A.

These quenching effects would need to be correctly modeled to use the intensity for soft-sensing of cofactor induced refolding processes such as the GalOx refolding. This, however, is not a trivial task since the aggregate formation and quenching effects both influence the intensity in a similar way leading to an underdetermination of the solution space. It is practically unfeasible to predict the amount of fluorescence quenching induced by the cofactor *a priori*. Further experiments would need to be conducted to identify the functional relationship between fluorescence intensity (F) and cofactor concentration using varying concentrations of copper. This, in turn, could be incorporated in the soft-sensor calculations in order to decouple the effects of aggregation and quenching. This, however, was outside the scope of this work. In addition, it must be assumed that the influence on the intensity shift could also have other origins and thus is even more compex as previously assumed.

4.4. Kinetic knowledge based on soft-sensor estimates

As the results of Fig. 4 showed, the open-loop reference model with kinetic parameters from literature does not provide a good prediction of the total measured reaction dynamics. Even though the soft-sensors developed in this work are not perfectly accurate, especially when considering the separated state estimation of N and A (Fig. 5), it generally represents the process dynamics much better than the open-loop reference model (NRMSE values in Table 3). In this respect, we might use the soft-sensor estimates as a time-resolved measurement signal in order to gain more insights into the kinetic mechanisms of the refolding process. However, it has to be noted that the prediction power of such relationships is highly dependent on the accuracy of the soft-sensor estimates and thus might be prone to errors in cases where the soft-sensor does not deliver good estimates. Still, the soft-sensor estimates provide a valuable data basis for the analysis of major process dynamics.

In this section we specifically consider the effect of the denaturant concentration (D) on the specific total reaction rate (k_I) in the GalOx refolding process. As proposed in literature (Tsumoto et al., 2003; Pan, 2015) we assumed that the folding reaction is mainly driven by the amount of reactive intermediates I, which is in turn depending on the concentration of denaturing agent. Thus, we calculated $k_I(t)$ for each GalOx process by dividing the soft-sensor estimates for $\frac{dI}{dt}$ by the respective estimates for $I(t)$. Fig. 8 shows k_I before and after the addition of copper in a form of violin plots and box plots. Even though the specific rates exhibit an increased amount of noise, it can be noticed from the data that a low amount of GuHCl (0.12M) leads to a stronger total reaction whereas a higher amount of GuHCl leads to a weaker reaction (0.6M). This trend is observed prior to the addition of copper as well as afterwards, whereas the absolute levels of k_I significantly differ depending on the presence of copper. After copper addition k_I exhibits around 10–15 times higher values than prior to the addition, which is convincing since copper is required as cofactor for the formation of a catalytically active GalOx. These effects can also be seen in the raw AEW data shown in Fig. 3, where the time span in which the AEW signal reaches a steady state after copper addition is negatively correlated to the concentration of GuHCl suggesting a strong fast reaction in case of low GuHCl concentrations (0.12 M) and a milder but longer lasting reaction in case of higher GuHCl concentrations (0.6 M).

Finally, it has to be noted that this kind of analysis of fast process dynamics would hardly be possible without the time-resolved online estimates as provided by the presented soft-sensors. Since the reaction is very fast after copper addition compared to upstream processing (k_I up to $15\,h^{-1}$ as shown in Fig. 8 vs. $q_{Smax} \approx 1.2\,h^{-1}$ in cultivations of *Escherichia coli* (Neubauer et al., 2003)), samples would have to be required in a relatively high frequency (probably under 10 min) in order to represent the fast reaction after copper addition. Given the relatively high workload of offline sampling and analytics in protein refolding processes, this would not be feasible especially when considering the low signal quality of offline analytics obtained in short sampling intervals (Igwe et al., 2023a). Therefore, the proposed soft-sensor can be seen as an enabler for the identification and analysis of fast reaction dynamics in protein refolding.

5. Conclusion

Two soft-sensing approaches for protein refolding processes based on online intrinsic fluorescence measurements were developed and

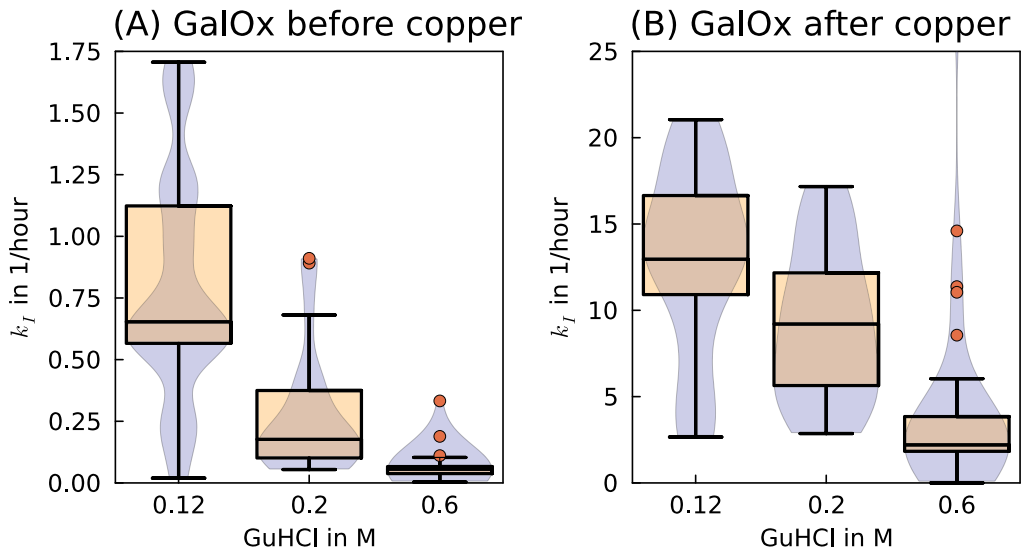


Fig. 8. Specific total reaction rate k_I obtained from the soft-sensor estimates through divison of $\frac{dI}{dt}$ by I prior to copper addition (A) and after copper addition (B). The violin plot shows the underlying distribution of k_I by kernel density estimation. The overlaid box plot visually represents the statistical properties of the distribution using quartiles with the red dots being outliers.

experimentally validated on LDH and GalOx refolding processes. The direct soft-sensor is easily applicable since it does not require any kinetic model parameters and is solely based on online measurements of intrinsic fluorescence. Prerequisites for the application only include a known initial protein concentration and a correlation function between the derivative of AEW and total reaction rate. Thereby, the direct soft-sensor was able to accurately predict the total folding products of the LDH (NRMSE = 6.1%) and the GalOx (NRMSE = 12%) refolding experiments.

Due to its simplicity the direct soft-sensor has some limitations. Since it is based on a correlation model for the calculation of the reaction rate $\frac{dI}{dt}$, it might violate physical constraints such as non-negative concentrations. By using an indirect model-based observer such as the particle filter used in this work, we were able to introduce reaction kinetics and further constraints into the estimation. Thereby, the PF observer did not reach concentrations below $0\,\mathrm{g\,L^{-1}}$.

This novel approach also serves as an enabler for advanced process analysis. In particular, the soft-sensor supports the identification of kinetic mechanisms under fast reaction dynamics, which is still unfeasible with a state-of-the-art offline sampling approach due to a low sampling frequency and consequently low signal quality. The investigation of the effects, that the denaturing agent (GuHCl) exerts on the specific reaction rate k_I, was illustrated as an example for the analysis of process dynamics in GalOx refolding experiments.

Furthermore, a separation of the total prediction into native protein and insoluble aggregates was possible for LDH refolding under the assumption of the fluorescence integral (F) being a measure for soluble protein concentration ($I + N$). This, in turn, enables the online calculation of quantifiable metrics such as the refolding yield over the course of the process. However, we found that for proteins which require a cofactor binding such as the GalOx, the assumption of F being proportional to the soluble protein concentration does not hold anymore since fluorescence quenching effects are often induced by the addition of the cofactor.

Still, the results show that online monitoring of Trp and Tyr fluorescence is a valuable tool for soft-sensor development in refolding processes. We propose further investigation of the prediction capabilities for other relevant proteins, which include disulfide bridges and cofactor binding. Especially the direct soft-sensor poses a high potential for method transfer to other proteins since it does not rely on any kinetic model parameters. Instead, a simple linear correlation between $\frac{dAEW}{dt}$ and the reaction rate needs to be defined, which can be obtained by a few experiments in small scale.

CRediT authorship contribution statement

Chika Linda Igwe: Writing – original draft, Investigation, Data curation, Conceptualization. **Florian Gisperg:** Software, Methodology. **Matthias Kierein:** Investigation. **Eva Přáda Brichtová:** Writing – review & editing. **Oliver Spadiut:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization. **Don Fabian Müller:** Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have shared the link to my data in the material & methods section in the manuscript.

Acknowledgments

The authors CLI and DFM acknowledge financial support through the COMET Centre CHASE, funded within the COMET Competence Centers for Excellent Technologies program by the BMK, the BMDW and the Federal Provinces of Upper Austria and Vienna. The COMET program is managed by the Austrian Research Promotion Agency (FFG).

The financial support by the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and Development and the Christian Doppler Research Association is gratefully acknowledged by FG, EPB and OS.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compchemeng.2024.108734.

References

Bayer, B., Sissolak, B., Duerkop, M., von Stosch, M., Striedner, G., 2020. The shortcomings of accurate rate estimations in cultivation processes and a solution for precise and robust process modeling. Bioprocess Biosyst. Eng. 43 (2), 169–178. http://dx.doi.org/10.1007/s00449-019-02214-6.

Cao, Y., Li, H., 2011. Dynamics of protein folding and cofactor binding monitored by single-molecule force spectroscopy. Biophys. J. 101 (8), 2009–2017. http://dx.doi.org/10.1016/j.bpj.2011.08.051.

Carlson, F.B., 2020. MonteCarlomeasurements.jl: Nonlinear propagation of arbitrary multivariate distributions by means of method overloading. arXiv:2001.07625.

Cleland, J.L., Hedgepeth, C., Wang, D.I., 1992. Polyethylene glycol enhanced refolding of bovine carbonic anhydrase B. Reaction stoichiometry and refolding model. J. Biol. Chem. 267 (19), 13327–13334. http://dx.doi.org/10.1016/S0021-9258(18)42214-4.

Cleveland, W.S., Devlin, S.J., 1988. Locally weighted regression: An approach to regression analysis by local fitting. J. Amer. Statist. Assoc. 83 (403), 596–610. http://dx.doi.org/10.1080/01621459.1988.10478639.

de Bernardez Clark, E., Hevehan, D., Szela, S., Maachupalli-Reddy, J., 1998. Oxidative renaturation of hen egg-white lysozyme. Folding vs aggregation. Biotechnol. Prog. 14 (1), 47–54. http://dx.doi.org/10.1021/bp970123w.

Dong, X.-Y., Huang, Y., Sun, Y., 2004a. Refolding kinetics of denatured-reduced lysozyme in the presence of folding aids. J. Biotechnol. 114 (1), 135–142. http://dx.doi.org/10.1016/j.jbiotec.2004.06.012.

Dong, X.-Y., Shi, G.-Q., Li, W., Sun, Y., 2004b. Modeling and simulation of fed-batch protein refolding process. Biotechnol. Prog. 20 (4), 1213–1219. http://dx.doi.org/10.1021/bp0499597.

FDA, 2004. Guidance for industry PAT - A framework for innovative pharmaceutical development, manufacturing, and quality assurance. p. 19.

Hellmann, N., Schneider, D., 2019. Hands on: Using tryptophan fluorescence spectroscopy to study protein structure. In: Kister, A.E. (Ed.), Protein Supersecondary Structures: Methods and Protocols. In: Methods in Molecular Biology, Springer, New York, NY, pp. 379–401. http://dx.doi.org/10.1007/978-1-4939-9161-7_20.

Humer, D., Ebner, J., Spadiut, O., 2020. Scalable high-performance production of recombinant horseradish peroxidase from E. coli inclusion bodies. Int. J. Mol. Sci. 21 (13), 4625. http://dx.doi.org/10.3390/ijms21134625.

Igwe, C.L., Müller, D.F., Gisperg, F., Pauk, J.N., Kierein, M., Elshazly, M., Klausser, R., Kopp, J., Spadiut, O., Přáda Brichtová, E., 2024. Online monitoring of protein refolding in inclusion body processing using intrinsic fluorescence. Anal. Bioanal. Chem. http://dx.doi.org/10.1007/s00216-024-05249-1.

Igwe, C.L., Pauk, J.N., Hartmann, T., Herwig, C., 2023a. Quantitative analytics for protein refolding States. Process Biochem. S1359511323003811. http://dx.doi.org/10.1016/j.procbio.2023.11.022.

Igwe, C.L., Pauk, J.N., Müller, D.F., Jaeger, M., Deuschitz, D., Hartmann, T., Spadiut, O., 2023b. Comprehensive evaluation of recombinant lactate dehydrogenase production from inclusion bodies. J. Biotechnol. http://dx.doi.org/10.1016/j.jbiotec.2023.11.006.

Jerri, A., 1977. The Shannon sampling theorem—Its various extensions and applications: A tutorial review. Proc. IEEE 65 (11), 1565–1596. http://dx.doi.org/10.1109/PROC.1977.10771, Conference Name: Proceedings of the IEEE.

Kager, J., Herwig, C., Stelzer, I.V., 2018. State estimation for a penicillin fed-batch process combining particle filtering methods with online and time delayed offline measurements. Chem. Eng. Sci. 177, 234–244. http://dx.doi.org/10.1016/j.ces.2017.11.049.

Kiefhaber, T., Rudolph, R., Kohler, H.-H., Buchner, J., 1991. Protein aggregation in vitro and in vivo: A quantitative model of the kinetic competition between folding and aggregation. Bio/Technology 9 (9), 825–829. http://dx.doi.org/10.1038/nbt0991-825, Number: 9 Publisher: Nature Publishing Group.

Kopp, J., Zauner, F.B., Pell, A., Hausjell, J., Humer, D., Ebner, J., Herwig, C., Spadiut, O., Slouka, C., Pell, R., 2020. Development of a generic reversed-phase liquid chromatography method for protein quantification using analytical quality-by-design principles. J. Pharm. Biomed. Anal. 188, 113412. http://dx.doi.org/10.1016/j.jpba.2020.113412.

Lakowicz, J.R. (Ed.), 2006. Protein fluorescence. In: Principles of Fluorescence Spectroscopy. Springer US, Boston, MA, pp. 529–575. http://dx.doi.org/10.1007/978-0-387-46312-4_16.

Luttmann, R., Bracewell, D.G., Cornelissen, G., Gernaey, K.V., Glassey, J., Hass, V.C., Kaiser, C., Preusse, C., Striedner, G., Mandenius, C.-F., 2012. Soft sensors in bioprocessing: A status report and recommendations. Biotechnol. J. 7 (8), 1040–1048. http://dx.doi.org/10.1002/biot.201100506, roam.

Ma, Y., Gowda, S., Anantharaman, R., Laughman, C., Shah, V., Rackauckas, C., 2022. ModelingToolkit: A composable graph transformation system for equation-based modeling. http://dx.doi.org/10.48550/arXiv.2103.05244, arXiv:2103.05244 [cs].

Mears, L., Stocks, S.M., Albaek, M.O., Sin, G., Gernaey, K.V., 2017. Mechanistic fermentation models for process design, monitoring, and control. Trends Biotechnol. 35 (10), 11.

Michaux, C., Roussel, G., Lopes-Rodrigues, M., Matagne, A., Perpète, E., 2016. Unravelling the mechanisms of a protein refolding process based on the association of detergents and co-solvents. J. Peptide Sci. 22 (7), 485–491. http://dx.doi.org/10.1002/psc.2893, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/psc.2893.

Mohd Ali, J., Ha Hoang, N., Hussain, M., Dochain, D., 2015. Review and classification of recent observers applied in chemical process systems. Comput. Chem. Eng. 76, 27–41. http://dx.doi.org/10.1016/j.compchemeng.2015.01.019.

Müller, D.F., Wibbing, D., Herwig, C., Kager, J., 2023. Simultaneous real-time estimation of maximum substrate uptake capacity and yield coefficient in induced microbial cultures. Comput. Chem. Eng. 173, 108203. http://dx.doi.org/10.1016/j.compchemeng.2023.108203.

Narayanan, H., Luna, M.F., Stosch, M., Cruz Bournazou, M.N., Polotti, G., Morbidelli, M., Butté, A., Sokolov, M., 2020. Bioprocessing in the digital age: The role of process models. Biotechnol. J. 15 (1), 1900172. http://dx.doi.org/10.1002/biot.201900172.

Neubauer, P., Lin, H.Y., Mathiszik, B., 2003. Metabolic load of recombinant protein production: Inhibition of cellular capacities for glucose uptake and respiration after induction of a heterologous gene inEscherichia coli. Biotechnol. Bioeng. 83 (1), 53–64. http://dx.doi.org/10.1002/bit.10645.

Pan, S., 2015. Engineering batch and pulse refolding with transition of aggregation kinetics_ An investigation using green fluorescent protein (GFP). Chem. Eng. Sci..

Pauk, J.N., Igwe, C.L., Herwig, C., Kager, J., 2024. An all-in-one state-observer for protein refolding reactions using particle filters and delayed measurements.

Pauk, J.N., Raju Palanisamy, J., Kager, J., Koczka, K., Berghammer, G., Herwig, C., Veiter, L., 2021. Advances in monitoring and control of refolding kinetics combining PAT and modeling. Appl. Microbiol. Biotechnol. http://dx.doi.org/10.1007/s00253-021-11151-y, ni.

Pizarro, S.A., Dinges, R., Adams, R., Sanchez, A., Winter, C., 2009. Biomanufacturing process analytical technology (PAT) application for downstream processing: Using dissolved oxygen as an indicator of product quality for a protein refolding reaction. Biotechnol. Bioeng. 104 (2), 340–351. http://dx.doi.org/10.1002/bit.22382.

Rackauckas, C., Nie, Q., 2017. DifferentialEquations.jl – A performant and feature-rich ecosystem for solving differential equations in Julia. J. Open Res. Softw. 5 (1), 15. http://dx.doi.org/10.5334/jors.151.

Rüdt, M., Briskot, T., Hubbuch, J., 2017. Advances in downstream processing of biologics – spectroscopy: An emerging process analytical technology. J. Chromatogr. A 1490, 2–9. http://dx.doi.org/10.1016/j.chroma.2016.11.010.

Ryś, S., Muca, R., Koł odziej, M., Piątkowski, W., Dürauer, A., Jungbauer, A., Antos, D., 2015. Design and optimization of protein refolding with crossflow ultrafiltration. Chem. Eng. Sci. 130, 290–300. http://dx.doi.org/10.1016/j.ces.2015.03.035.

Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36 (8), 1627–1639. http://dx.doi.org/10.1021/ac60214a047, Publisher: American Chemical Society.

Sharma, R., Anupa, A., Kateja, N., Rathore, A.S., 2022. Optimization of the in-vitro refolding of biotherapeutic Fab Ranibizumab. Biochem. Eng. J. 187, 108601. http://dx.doi.org/10.1016/j.bej.2022.108601.

Tsumoto, K., Ejima, D., Kumagai, I., Arakawa, T., 2003. Practical considerations in refolding proteins from inclusion bodies. Protein Expr. Purif. 28 (1), 1–8. http://dx.doi.org/10.1016/S1046-5928(02)00641-1.

Vanderlinde, R.E., 1985. Measurement of total lactate dehydrogenase activity.

Wechselberger, P., Sagmeister, P., Herwig, C., 2013. Model-based analysis on the extractability of information from data in dynamic fed-batch experiments. Biotechnol. Prog. 29 (1), 285–296. http://dx.doi.org/10.1002/btpr.1649.

Weiner, R.E., Ettinger, M.J., Kosman, D.J., 1977. Fluorescence properties of the copper enzyme galactose oxidase and its tryptophan-modified derivatives. Biochemistry 16 (8), 1602–1606. http://dx.doi.org/10.1021/bi00627a012, Publisher: American Chemical Society.

Whittaker, J.W., 2005. The radical chemistry of galactose oxidase. Arch. Biochem. Biophys. 433 (1), 227–239. http://dx.doi.org/10.1016/j.abb.2004.08.034.