Journal of Chromatography A, 1217 (2010) 7723–7737

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Size-exclusion chromatographic protein refolding: Fundamentals, modeling and operation

Esteban J. Freydell[a], Luuk A.M. van der Wielen[a], Michel H.M. Eppink[b,1], Marcel Ottens[a,*]

[a] Department of Biotechnology, Delft University of Technology, Delft, The Netherlands
[b] MSD, Oss, The Netherlands

ARTICLE INFO

Article history:
Received 27 May 2010
Received in revised form 4 September 2010
Accepted 6 October 2010
Available online 13 October 2010

Keywords:
Size-exclusion chromatography
On-column refolding
Modeling

ABSTRACT

Size-exclusion chromatography (SEC) has proven its capability to refold a variety of proteins using a range of gel filtration column materials, demonstrated in the growing body of experimental evidence. However, little effort has been allocated to the development of mechanistic models describing size-exclusion chromatographic refolding reactors (SECRR). Mechanistic models are important since they provide a link between process variables like denatured and reduced protein feed concentration ($C_{f,D\&R}$), flow rate, column length, etc., and performance indicators like refolding yield (Y_N), thereby opening the possibility for in silico design of SECRRs. A critical step, in the formulation of such models, is the selection of an adequate reaction mechanism, which provides the direct link between the separation and the refolding yield. Therefore, in this work we present a methodology using a SEC refolding reactor model, supported by a library of reaction mechanisms, to estimate a suitable reaction scheme using experimental SEC refolding data. SEC refolding data is used since it provides information about the mass distribution of monomers and aggregates after refolding, information not readily available from batch dilution refolding data alone. Additionally, this work presents (1) a systematic analysis of the reaction mechanisms considered using characteristic time analysis and Damköhler maps, revealing (a) the direct effect of a given reaction mechanism on the shape of the SEC refolding chromatogram (number of peaks and resolution) and (b) the effect that the competition between convection, refolding and aggregation is likely to have on the SEC refolding yield; (2) a comparison between the SECR reactor and the batch dilution refolding reactor based on mechanistic modeling, quantitatively showing the advantages of the former over the latter; and (3) the successful application of the modeling based strategy to study the SEC refolding data of an industrially relevant protein. In principle, the presented modeling strategy can be applied to any protein refolded using any gel filtration material, providing the proper mass balances and activity measurements are available.

1. Introduction

Size-exclusion chromatography (SEC) is a widely used chromatographic technique for the purification and characterization of protein mixtures. As a characterization tool, SEC is used to deduce information about molecular weights or sizes of the various proteins purified, provided reasonable calibration standards are used. For purification, SEC is commonly applied as a desalting and/or buffer exchange step, and throughout the years, this application has been used for the purpose of protein refolding. Protein refolding using SEC was first reported in the early nineties in the work of Werner et al., and subsequently followed by other groups [1–3]. Basically, the procedure of SEC refolding works as follows. A feed pulse, composed of denaturant, reducing agent, and denatured and reduced (D&R) protein is injected to the SEC column, pre-equilibrated with refolding buffer. As a consequence of the differences in distribution coefficient, the concentration waves of denaturant, reducing agent and the denatured and reduced protein, separate as they migrate through the column. This separation, leads to a decrease on the local concentration of the denaturant and reducing agent around the protein, inducing protein refolding.

Size-exclusion refolding studies have been conducted using a variety of model proteins and a range of gel filtration materials. The majority of these studies have been summarized in several review papers [4–6]. Despite the abundance of size-exclusion refolding data, at this time little effort has been allocated to modeling this type of chromatographic refolding reactors. Probably SEC refolding was first modeled in the work presented by Endo et al. [7]. This work used an equilibrium reaction mechanism and the plate model to analyze SEC protein denaturation data. The reaction mechanism coupled to the column model, did not consider protein

* Corresponding author. Tel.: +31 15 278 2151; fax: +31 15 278 2355.
E-mail address: m.ottens@tudelft.nl (M. Ottens).
[1] Current address: Synthon, Nijmegen, The Netherlands.

0021-9673/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2010.10.038

aggregation and it was constituted by equilibrium reactions. An interesting aspect of the work presented by Endo was the coupling of the denaturant concentration to the refolding kinetics, using a mathematical relation adopted from the work done by Creighton on electrophoretic analysis of protein denaturation with urea [8]. These equations have been also applied to study the renaturation of Thioredoxin in SEC [9]. More recently, Ding et al studied the effect of axial dispersion on SEC refolding [10], modeling the SEC refolding reactor as a dispersive plug-flow reactor, omitting mass transfer and incorporating the refolding mechanism as a first-order refolding, competing with a third-order aggregation.

Modeling allows the development of a quantitative relation between the process indicators (e.g., refolding yield, resolution, etc.) and the operational variables (e.g. flow rate, denatured and reduced protein feed concentration, etc.). To properly link the performance indicators with the process variables, the different mechanisms involved should be accounted for by the model. In the case of size-exclusion chromatographic refolding (SECR), these mechanisms include: convection, axial dispersion, mass transfer and the competition between refolding and aggregation. The fundamental challenge, when it comes to the formulation of such a model, is the selection of an adequate reaction scheme that properly captures the competition between refolding and protein aggregation. Such reaction scheme is of paramount importance, as it provides the link between the separation and the refolding yield of the SEC refolding reactor.

Up to this point, there seems to be no reaction scheme that can describe the competition between refolding and aggregation for all proteins. On the contrary, most likely the reaction scheme depends on the type of protein and on the chemical composition of the protein solution. A situation that exemplifies the difficulty of finding a unified scheme is the case of lysozyme. On one side, Hevehan et al. [11] showed that the increase of the native lysozyme concentration with time, during batch dilution refolding, is best described by a first-order refolding competing with a third-order aggregation. In contrast, Buswell and Middelberg [12] showed that such mechanism was not adequate to describe their fed-batch refolding data of lysozyme, as a consequence a new mechanism was proposed. Strictly speaking, to decipher a fundamental reaction mechanism, the transient concentration changes of the products, reactants and any measurable intermediate, have to be measured. In the case of protein refolding from inclusion bodies, this endeavor seems impractical. Most likely for that reason, the common approach to determine kinetic constants is to follow the transient profile of the native protein, during batch dilution refolding, and thereafter fit a kinetic mechanism to it [11,13]. A shortcoming of this approach though, is that it omits the formation of other by-products (i.e., aggregates, misfolded species), as it only considers the native protein, limiting therefore its capability to represent the real system.

This work presents a novel modeling approach to obtain a suitable reaction scheme, which captures the competition between folding and aggregation, using SEC refolding data. The modeling approach is presented as a rational strategy, supported by a library of reaction schemes and a modeling tool, both developed and presented in this work. The modeling tool is constituted by a column model to describe the separation behavior, and a reaction scheme to couple the separation to the refolding yield. SEC refolding data, particularly the chromatogram, is used since it provides information about the amount and type of by-products formed, extracted using activity assays and mass balances. Together, the rational strategy, the library of reaction schemes, and the modeling tool, give the ability to discern between reaction schemes, on a quantitative and qualitative basis and to choose the scheme that describes the elution chromatogram best. Our results show (1) the capability of the modeling tool to describe complex SEC refolding chromatograms, (2) the effect that reactor topography has on the refolding yield, exemplified by a comparison between batch and SEC refolding simulations, and (3) the application of the methodology to study the SEC refolding of an industrially relevant protein. In principle, the presented methodology can be applied to study the SEC refolding data of any protein using any gel filtration material, providing that proper mass balances and activity data are available.

2. Theory and modeling

2.1. Column model

The packed bed column was modeled using the equilibrium-transport-dispersive model of chromatography, in combination with the solid-film linear driving force kinetic equation [14–16]. This model is described by the two differential mass balances presented in Eqs. (1) and (2).

$$\frac{\partial C_{b,i}}{\partial t} = D_{L,i}\frac{\partial^2 C_{b,i}}{\partial x^2} - u\frac{\partial C_{b,i}}{\partial x} - Pk_{ov,i}(C_{EqS,i} - C_{S,i}) + r_{b,i} \quad (1)$$

$$\frac{\partial C_{S,i}}{\partial t} = k_{ov,i}(C_{EqS,i} - C_{S,i}) + r_{S,i} \quad (2)$$

where t represents time, x represent the axial distance, $D_{L,i}$ is the axial dispersion coefficient, u is the interstitial velocity, P is the phase ratio, $C_{b,i}$ is the bulk liquid phase concentration and $C_{S,i}$ is the solid phase concentration. $C_{EqS,i}$ is the solid phase concentration in equilibrium with the bulk concentration $C_{b,i}$, as given by the isotherm equation. $r_{b,i}$ and $r_{s,i}$ represent the net concentration change due to reaction, for the bulk and solid phases, respectively. These two terms will be hereinafter called source terms and they will depend on the reaction mechanism chosen. To solve the two column equations, the following initial and boundary conditions (Danckwerts boundary conditions [15]) are required:

$$x = 0 \quad \frac{\partial C_{b,i}}{\partial x} = \frac{u}{D_{L,i}}(C_{b,i} - C_{f,i}(t)) \quad (3a)$$

$$x = L_C \quad \frac{\partial C_{b,i}}{\partial x} = 0 \quad (3b)$$

$$t = 0 \quad C_{b,i}(0, 0 < x < L_C) = 0 \quad (3c)$$

$$t = 0 \quad C_{S,i}(0, 0 < x < L_C) = 0 \quad (3d)$$

where L_C represent the column length. The operation of the column, unless otherwise specified, was modeled as a loading-elution operation, which is represented as follows:

$$C_{f,i}(t) = 0 \quad \text{for} \quad t_{\text{pulse}} < t \quad (4a)$$

$$C_{f,i}(t) = C_{\text{feed},i} \quad \text{for} \quad t < t_{\text{pulse}} \quad (4b)$$

where t_{pulse} is the feed duration, determined by the volume and flow rate of injection. $C_{\text{feed},i}$ is the feed concentration for component i. $C_{f,i}(t)$ is the concentration of component i just before entering the column. Eqs. (4a) and (4b) represent the elution and loading, respectively.

Eqs. (1)–(4) are solved for each component present in the feed or formed during a given reaction. The partial differential equations were solved using the method of lines (MOL) [17]. The second and first order derivatives, with respect to space, were discretized using fourth-order finite difference equations, transforming the PDE into a system of ordinary differential equations (ODEs). This system, in vector and matrix form, is presented in Eqs. (5a) and (5b).

$$\frac{d\mathbf{c}_b}{dt} = \mathbf{A}\cdot\mathbf{c}_b - Pk_{ov}(\mathbf{c}_s^* - \mathbf{c}_s) + \mathbf{b} + \mathbf{r}_b \quad (5a)$$

$$\frac{d\mathbf{c}_s}{dt} = k_{ov}(\mathbf{c}_s^* - \mathbf{c}_s) + \mathbf{r}_s \quad (5b)$$

Table 1
Model parameters and correlations.

Mass transfer parameters	Notation	Name of the correlation	Reference
Free diffusivities	$D_{f,\text{Proteins}}$	Young	[21,22]
	$D_{f,\text{Urea}}$	Wilke–Chang	
Pore diffusivity	D_p	Satterfield	[19]
Film mass transfer coefficient	k_f	Wilson–Geankoplis	[23]
Overall mass transfer coefficient	k_{ov}	Glueckauf	[24]
Axial Péclet	Pe_L	Chung–Wen	[25]

where is $\mathbf{c}_b$ is the $N \times 1$ vector of bulk concentrations, $\mathbf{A}$ is an $N \times N$ matrix, P is the phase ratio, $\mathbf{c}_s$ is the $N \times 1$ vector of solid phase concentrations, $\mathbf{b}$ is the $N \times 1$ boundary vector and $\mathbf{c}_S^*$ is the $N \times 1$ vector of equilibrium concentrations. $\mathbf{r}_b$ and $\mathbf{r}_s$ are the $N \times 1$ source vectors.

The total number of equations to be solved is equal to $N \times 2 \times nc$, being N the number of grid-points in which the axial direction is discretized, 2 stands for the two state variables (i.e., bulk and solid phase concentrations) and nc is the number of components considered. These equations were numerically integrated using the stiff solver ode15s of MATLAB R2007b.

2.2. Model parameters and correlations

Mass transfer parameters include free diffusivities and effective diffusivities, among others. These parameters are ultimately used to calculate the overall mass transfer coefficient. All these parameters can be estimated using well-established correlations (Table 1). The correlations used have successfully been applied for the modeling of size-exclusion band profiles [18–20], supporting their use in this work.

The axial dispersion coefficient was calculated using the axial Péclet number as follows:

$$D_L = \frac{uL_c}{Pe_L} \tag{6}$$

2.3. Gel filtration media

Superdex 75 and the Sephacryl S-100 are gel filtration materials widely used in industry and academia for standard protein purifications. In addition, these gel filtration materials have been widely used for on-column size-exclusion refolding studies [26–30]. The characteristics of these gel filtration materials are presented in Table 2.

2.4. Distribution coefficient

In size-exclusion chromatography (SEC), solutes are separated on the basis of the different fractions of the pore volume that, for steric reasons, are available for solutes of different sizes [32,33]. In other words, smaller molecules have greater access to the pores and larger molecules have less. Hence, the distribution coefficient in SEC is a function of the dimensions of both, the solute and the pores. Strictly speaking, in SEC the distribution coefficient represents the fraction of the intraparticle void volume accessible to a molecule of certain size. Herein the distribution coefficient is referred to as the average distribution coefficient K_{av}. The average distribution coefficient was determined using the extended Ogston model (Eq. (7)). Originally introduced by Ogston [34], this model considers the gel filtration material as an arrangement of randomly oriented, infinitely thin and long fibers, and the solutes as rigid spheres. The model was years after modified by Bosma and Wesselingh [35], to account for the fraction of fibers and for a finite fiber thickness, which are parameters that depend on the gel filtration material. These modifications gave birth to the so-called extended-Ogston model presented in Eq. (7).

$$K_{\text{av}_i} = \exp\left(-\ln\left(\frac{1}{1-\phi_f}\right)\left(1+\frac{r_i}{r_f}\right)^2\right) \tag{7}$$

where K_{av} and r_i represent the average distribution coefficient and the hydrodynamic radius of the solute, respectively. ϕ_f represent the volume fraction of the fibers and r_f the fiber radius. The solute radius (r_i) was calculated, assuming an spherical shape using Eq. (8) [33,36].

$$r_i = 0.81 \times 10^{-10}(\text{MW}^{1/3}) \tag{8}$$

where MW is the molecular weight of the protein.

Table 2
Properties of the gel filtration media.

Gel filtration material	Fractionation range (Da)	[a] d_{pore} (nm)	[b] d_p (μm)
Superdex 75	3000–70 000	6.00	13
Sephacryl S-100	1000–100 000	6.60	47

[a] Ref. [31].
[b] Ref. [32].

2.5. Reaction mechanisms

Protein refolding is a process in which the formation of the native product (i.e., correctly folded protein), competes against the formation of misfolded species and the formation of aggregates. Accordingly, the refolding yield is a function of the competition of these reaction rates. Of the two competing side reactions, aggregation affects the refolding yield the most. Probably for this reason, most research has been targeted towards preventing it [37,38].

In terms of reactions rates, the rate of refolding has been described as a first order reaction rate, because it is considered to be a uni-molecular reaction. In contrast, protein aggregation is a multimolecular reaction [39] and therefore its reaction rate is a function of the protein concentration prone to aggregate, to some power higher than 1.

The competition between folding and aggregation was probably first presented in the work of Kiefhaber et al. [40], who described the refolding rate as a first order reaction rate and the aggregation rate as a second-order reaction. This mechanism gained popularity throughout the years and it has been applied, with slight modifications, in the study of lysozyme refolding by both batch [11,13,41] and fed-batch [42]. Buswell and Middelberg [12] reported that this simplified model was not suitable to represent their lysozyme refolding experiments on a fed-batch system. As a result, the authors proposed an alternative mechanism to describe the competition between aggregation and folding. Following the work of Speed et al. [43], Buswell et al. represented the aggregation process by a condensation mechanism. The modification improved the model prediction, especially with respect to the transient aggregation profile measured by dynamic light-scattering. The previous mechanisms have been mainly used to study batch refolding and fed-batch refolding, but they have not yet being applied to analyze SEC refolding data.

This work presents and compares three different reaction mechanisms, which describe the competition between folding and aggregation. These mechanisms include: (1) first order refolding competing against a higher-order aggregation, as this is a popular representation, (2) first-order refolding competing against aggregation via sequential polymerization, and (3) first-order refolding competing against aggregation via cluster–cluster polymerization. The effect of the competition between refolding and aggregation, on the SEC chromatogram and the SEC refolding yield, was systematically studied using Damköhler maps.

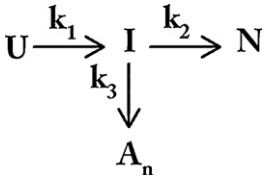

Fig. 1. Schematic representation of the reaction mechanism 1. U: unfolded protein, N: native protein, I: on-pathway intermediate. k_1 and k_2, are the kinetic constants of the first order reactions (s^{-1}). k_3 is the rate constant of aggregation ($l^{(p-1)}\,mmol^{(1-p)}\,s^{-1}$) and p is the reaction order.

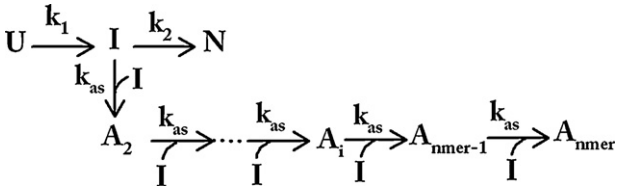

Fig. 2. Schematic representation of the reaction mechanism 2. Aggregation occurs via sequential polymerization. U: unfolded state, I: intermediate state, $A_{2\ldots nmer}$: aggregates class 2 to *nmer*, N: native state. k_1 and k_2, are the kinetic constants of the first order reactions (s^{-1}). k_{as} is the association constant ($l\,mmol^{-1}\,s^{-1}$).

2.5.1. Mechanism 1: first-order refolding and higher-order aggregation

As previously mentioned this mechanism was introduced by Kiefhaber et al. [40] and for the sake of further discussion we will describe it herein. This mechanism considers four species: the unfolded protein (U), one on-pathway intermediate (I), the native protein (N) and the aggregated protein (A_n). The model assumes: (1) the transition from unfolded (U) to intermediate (I) occurs instantaneously, meaning that the rate constant is considered to approach infinity (i.e., $k_1 \to \infty$), (2) the reaction rate of folding (i.e., the rate from I to N) is considered as a first-order reaction rate with respect to the concentration of I, (3) the rate of aggregation is considered of the order p with respect to the concentration of I, and (4) the intermediate reacts stoichiometrically to form an aggregate of class n (n moles of I form one mol of A_n). Fig. 1 presents the schematic depiction of this mechanism and Table 3 presents the different reactions and their corresponding reaction rates.

Using the equations presented in Table 3, the net concentration change of the product and reactants, as a function of the reaction rates and thereby time can be derived. In addition, by substituting the corresponding rate law, the net concentration change becomes a function of the active concentration of the intermediate (I) and the reaction rate constants. The net concentration change due to reaction, for the intermediate, the native and the aggregate is given by Eqs. (9a)–(9c).

$$r_I = \left.\frac{dC_I}{dt}\right|_{net} = -\upsilon_2 - n\upsilon_3 = -k_2C_I - nk_3{C_I}^P \tag{9a}$$

$$r_N = \left.\frac{dC_N}{dt}\right|_{net} = \upsilon_2 = k_2C_I \tag{9b}$$

$$r_{A_n} = \left.\frac{dC_{A_n}}{dt}\right|_{net} = \upsilon_3 = k_3{C_I}^P \tag{9c}$$

where υ_2 is the folding reaction rate, υ_3 is the aggregation reaction rate, k_2 is the folding rate constant and k_3 is the aggregation rate constant. p and n are the aggregation order and aggregation number, respectively.

2.5.2. Mechanism 2: first-order refolding and aggregation by sequential polymerization

Compared to the previous mechanism, this one considers the aggregation to be the result of a polymerization reaction. This is a more accurate representation of protein aggregation from a mechanistic point of view, compared to the previous description. Aggregation of proteins via this mechanism was probably first reported in the work of Oosawa et al. [44], who studied the aggregation of G-actin protein and concluded that the aggregation resembled a condensation process (i.e., a subsequent monomer addition process). Ever since, the mathematical description of this mechanism has been further improved and used to study the aggregation of several proteins, e.g., sickle cell hemoglobin, actin and P22 tailspike [43,45,46].

Fig. 2 presents the schematic representation of mechanism 2, showing how the aggregation is driven by the active concentration of I. The scheme includes formation and depletion of multimers, of different class sizes, based on the addition of molecules of I. Such considerations will lead to the set of irreversible reactions and reaction rates presented in Table 4.

In essence, this aggregation mechanism has an infinite set of coupled ordinary differential equations, and as a consequence an infinite number of reaction kinetic constants. Therefore, key assumptions needed to be made in order to solve it. (1) It is assumed that all reactions of the aggregation pathway are irreversible, (2) that the pathway will terminate at an aggregate of class size *nmer*, and (3) that all reactions can be described by one kinetic constant (i.e., k_{as}) [43]. The net concentration change of the products and reactants for this mechanism is given by Eqs. (10a)–(10e).

$$r_I = \left.\frac{dC_I}{dt}\right|_{net} = -\upsilon_2 - 2\upsilon_3 - \cdots - \upsilon_i - \cdots - \upsilon_{nmer}$$
$$= -k_2C_I - 2k_{as}{C_I}^2 - k_{as}C_I\sum_{j=2}^{nmer}C_J \tag{10a}$$

$$r_N = \left.\frac{dC_N}{dt}\right|_{net} = \upsilon_2 = k_2C_I \tag{10b}$$

$$r_{A_2} = \left.\frac{dC_{A_2}}{dt}\right|_{net} = \upsilon_3 - \upsilon_4 = k_{as}C_IC_I - k_{as}C_IC_{A_2} \tag{10c}$$

$$r_{A_i} = \left.\frac{dC_{A_i}}{dt}\right|_{net} = \upsilon_{i+1} - \upsilon_{i+2} = k_{as}C_IC_{A_{i-1}} - k_{as}C_IC_{A_i} \tag{10d}$$

Table 3
Reactions and reaction rates for mechanism 1.

Reaction	Reaction rate	Rate law
$I \xrightarrow{\upsilon_2} N$	$\upsilon_2 = -\frac{1}{1}\frac{dC_I}{dt} = \frac{1}{1}\frac{dC_N}{dt}$	$\upsilon_2 = k_2C_I$
$nI \xrightarrow{\upsilon_3} A_n$	$\upsilon_3 = -\frac{1}{n}\frac{dC_I}{dt} = \frac{1}{1}\frac{dC_{A_n}}{dt}$	$\upsilon_3 = k_3C_I^p$

Table 4
Reactions and reaction rates of mechanism 2.

Reaction	Reaction rate	Rate law
$I \xrightarrow{\upsilon_2} N$	$\upsilon_2 = -\frac{1}{1}\frac{dC_I}{dt} = \frac{1}{1}\frac{dC_N}{dt}$	$\upsilon_2 = k_2C_I$
$I + I \xrightarrow{\upsilon_3} A_2$	$\upsilon_3 = -\frac{1}{2}\frac{dC_I}{dt} = \frac{1}{1}\frac{dC_{A_2}}{dt}$	$\upsilon_3 = k_{as}C_IC_I$
$I + A_2 \xrightarrow{\upsilon_4} A_3$	$\upsilon_4 = -\frac{1}{1}\frac{dC_I}{dt} = -\frac{1}{1}\frac{dC_{A_2}}{dt} = \frac{1}{1}\frac{dC_{A_3}}{dt}$	$\upsilon_4 = k_{as}C_IC_{A_2}$
⋮	⋮	⋮
$I + A_{i-1} \xrightarrow{\upsilon_i} A_i$	$\upsilon_i = -\frac{1}{1}\frac{dC_I}{dt} = -\frac{1}{1}\frac{dC_{A_{i-1}}}{dt} = \frac{1}{1}\frac{dC_{A_i}}{dt}$	$\upsilon_i = k_{as}C_IC_{A_{i-1}}$
⋮	⋮	⋮
$I + A_{nmer-1} \xrightarrow{\upsilon_{nmer}} A_{nmer}$	$\upsilon_{nmer} = -\frac{1}{1}\frac{dC_I}{dt} = -\frac{1}{1}\frac{dC_{A_{nmer-1}}}{dt} = \frac{1}{1}\frac{dC_{A_{nmer}}}{dt}$	$\upsilon_{nmer} = k_{as}C_IC_{A_{nmer-1}}$

$$r_{A_{nmer}} = \left.\frac{dC_{A_{nmer}}}{dt}\right|_{net} = \upsilon_{nmer} = k_{as} C_I C_{A_{nmer-1}} \quad (10e)$$

where υ_2 is the folding reaction rate, υ_i is the reaction rate of reaction i, k_2 is the folding rate constant and k_{as} is the association rate constant.

2.5.3. Mechanism 3: first-order refolding and aggregation via cluster–cluster polymerization

Aggregation studies of glutamate dehydrogenase resulted in the introduction, by Thusius et al. [47], of the random association mechanism. This mechanism proposes that two monomeric or polymeric units, of any size, can associate to form a larger polymer. During the work of Speed et al. [43] this mechanism was simplified, by assuming, instead of reversible associations irreversible ones, reducing the number of kinetic parameters. In addition, the mechanism was given the name of cluster–cluster mechanism. An interesting feature of this mechanism, is that the polymerization (i.e. aggregation) is not exclusively regulated by the active concentration of the on-pathway intermediate (I), distinguishing it from the sequential polymerization. In principle, the cluster–cluster polymerization model accounts for the formation and depletion of a multimer by both, monomer–multimer interactions and multimer–multimer interactions. To simplify this system, all the aggregation rates are assumed to have the same rate constant (k_{as}). Eqs. (11a)–(11c) present the net concentration change due to reaction for the intermediate (I), the native protein (N), and the different aggregates formed. Eq. (11c) applies for i equal to an even number. In the case of i being an odd number, the index of the first summation in the right hand side of Eq. (11c) changes to $(i-1)/2$.

$$r_I = \left.\frac{dC_I}{dt}\right|_{net} = -k_2 C_I - 2k_{as} {C_I}^2 - k_{as} C_I \sum_{j=2}^{nmer} C_{A_j} \quad (11a)$$

$$r_N = \left.\frac{dC_N}{dt}\right|_{net} = k_2 C_N \quad (11b)$$

$$r_{A_i} = \left.\frac{dC_{A_i}}{dt}\right|_{net} = k_{as} \sum_{j=1}^{i/2} C_{A_j} C_{A_{i-j}} - k_{as} C_{A_i} C_{A_i} - k_{as} C_{A_i} \sum_{j=1}^{nmer} C_{A_j} \quad (11c)$$

2.6. Rational strategy to select a reaction mechanism

The step preceding the determination of the kinetic constants is the selection of a suitable reaction mechanism. This is because the mechanism defines the number of reactions, the rate expressions and thus the number of kinetic constants. Once coupled to the column model (Eqs. (1) and (2)), the mechanism provides the link between separation and refolding yield. It is thus evident the impact that the selection step has on the proper modeling of the size-exclusion refolding reactor, as this selection step defines the reaction mechanism. To deal with this challenge, this work presents a rational strategy to select a suitable reaction mechanism, based on quantitative and qualitative criteria derived from the SEC refolding data. The strategy is presented in Fig. 3. The approach presented is divided in 4 major sections: data generation, data processing, systematic analysis and output.

The 'Data generation' phase deals with all the experimental measurements. These include: column calibration and characterization, SEC refolding experiments and the offline analysis (e.g., protein determination, activity measurements). '*Data processing*' involves: (1) mass balances to determine the refolding yield, the fraction monomers (active and non-active) and the fraction of aggregates, (2) determination from the SEC refolding chromatogram, of the elution volume and the number of species that the mechanism needs to account for, (3) estimation of the average distribution coefficient (K_{av}) of the various species, and (4) preselection, from the library of mechanisms, of the plausible reaction schemes that will be taken to the 'Systematic analysis' phase. The latter step is based on the qualitative information derived from the SEC chromatogram, thus is highly likely than more than one mechanism is selected.

At the stage 'Systematic analysis' simulations using the SECR model are conducted. The first aim is to reduce the number of the mechanisms selected in the 'Data processing' phase. This is done using Damköhler maps, which are generated for each mechanism. The maps provide a diversity of simulated chromatograms, obtained under different competitive scenarios. The simulated chromatograms are compared, on a qualitative basis, to the experimental SEC refolding chromatogram. This comparison is done by looking at how well a selected mechanism reproduces the peak shapes and the resolution displayed by the experimental SEC refolding chromatogram. Such comparison is supported by the fact that peak splitting, merging, fronting and tailing in SEC, are strongly affected by the reaction mechanism and the type of reactions included, whether irreversible or equilibrium [18].

Once the number of mechanisms has been reduced using the previous analysis, the promising ones are taken to the quantitative selection phase. The quantitative selection phase is based on a nonlinear optimization, which employs as objective function the sum of squared errors (SSE), calculated using the model SEC refolding chromatogram and the experimental chromatogram. The mechanism yielding the minimum SSE is then selected to assemble the function describing the SEC refolding reactor. Lastly, the 'Output' phase returns (1) the set of kinetic parameters, as they are the fitting parameters used in the optimization, (2) the suitable reaction scheme and (3) the yield function, relating the refolding yield with the operational variables of the size-exclusion refolding reactor.

2.7. Systematic analysis of the reaction mechanisms

Proteins often undergo unwanted aggregation or conformational changes during standard chromatography. Such reactions often lead to the formation of new species resulting in additional peaks, merging of peaks or oddly shaped peaks. Perhaps as early as the nineties, modeling tools had already proven their usefulness in the study of complex reaction–separations systems, particularly for chromatographic fractionations [48–50]. A good example of such studies was presented by Whitley et al. [49]. Their work analyzed and described, using the general rate model coupled to a reaction mechanism, the breakthrough curves of myoglobin during an immobilized metal affinity run and the elution profile of the fractionation of β-lactoglobulin on a weak hydrophobic resin [49]. Their work showed that dimensionless groups are a good predictor of peak resolution and peak shape. Dimensionless groups are defined as the ratio of the characteristic times of two mechanisms. Accordingly, they give information about the relative rates of different processes (e.g., convection vs. axial-dispersion, convection vs. reaction, etc.). These relative rates are important to understand reaction–separation systems.

In contrast to standard purifications, in on-column refolding the conversion of the injected denatured and reduced protein is desired. Thus, the shape and number of peaks of the chromatogram are the result of a reaction (e.g., folding, aggregation) occurring in conjunction with the separation (convection and mass transfer). Since all these processes occur concomitantly, it is important to assess the effect that their competition has on the SEC chromatogram. To systematically study the effect of this competition, Damköhler maps are used. Damköhler maps have been shown to successfully explain the elution behavior (peak shape, peak merging, peak tailing, peak fronting) in SEC with a dimerization [18], here this tool is applied to study SEC refolding. Damköhler num-

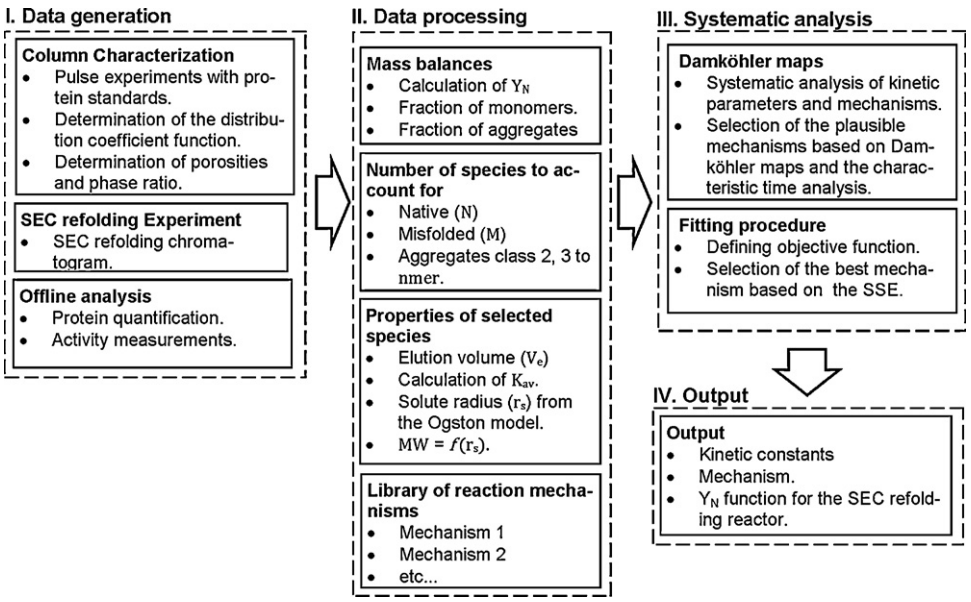

Fig. 3. Strategy to determine a suitable competitive reaction scheme, capturing the competition between refolding and aggregation, using SEC refolding data.

bers compare the characteristic time of reaction, that is refolding and aggregation, against the characteristic time of convection. A Damköhler greater than 1 means that the reaction is faster than the convection, whereas a Damköhler number less than 1 indicates that convection is faster than the reaction. The mechanisms 1–3 (Section 2.5) can be represented by two Damköhler numbers namely Da_I and Da_{II}, presented in Table 5. Da_I compares the characteristic times of folding (τ_{fold}) and convection (τ_{conv}), and Da_{II} compares the characteristic times of aggregation (τ_{agg}) and convection.

2.8. Performance criteria

2.8.1. Refolding yield

Refolding yield is defined as the amount of native product formed per amount of denatured and reduced protein loaded.

$$Y_{N,SECR,model} = (V_{inj}C_{f,D\&R})^{-1}\int_{V_1}^{V_2} C_{N,mod}(V)dV \quad (12a)$$

$$Y_{N,SECR,exp} = M_N(C_{f,D\&R}V_{inj})^{-1} \quad (12b)$$

$$Y_{N,Batch,model} = \frac{C_{N,model}}{C_{f,D\&R}} \quad (12c)$$

where $Y_{N,SECR,model}$ represent the modeled size-exclusion refolding yield and $Y_{N,SECR,exp}$ represent the experimental refolding yield. M_N is the mass of native protein. $C_{f,D\&R}$ is the feed concentration of denatured and reduced (D&R) protein. In the case of batch refolding this concentration represents the total protein concentration after dilution. $C_{N,mod}(V)$ is the modeled native concentration profile. V_{inj} represent the injection size to the column. $Y_{N,Batch,model}$ represent the modeled refolding yield of a batch reactor.

Table 5
Damköhler numbers of folding (DaI) and aggregation (DaII), for each reaction mechanism.

Mechanism	Da_I	Da_{II}
Mechanism 1	$\frac{k_2 L_C}{u}$	$\frac{L_C k_3 (C_{f,D\&R})^{(p-1)}}{u}$
Mechanism 2	$\frac{k_2 L_C}{u}$	$\frac{L_C k_{as} C_{f,D\&R}}{u}$
Mechanism 3	$\frac{k_2 L_C}{u}$	$\frac{L_C k_{as} C_{f,D\&R}}{u}$

3. Material and methods

3.1. Materials

The model protein used in this study was a fusion protein (FP) that has 127 amino acids, three disulfide bonds, no free cysteines and a theoretical isoelectric point of 7.64 (based on its primary sequence). The protein was obtained in the form of inclusion bodies and was provided by Schering-Plough (Oss, The Netherlands)

All chemicals used were at least reagent grade purity or higher. Urea and DL-Dithiothreitol (DTT) were purchased from Sigma–Aldrich (Zwijndrecht, The Netherlands). Sodium hydrogen carbonate, sodium hydroxide, sodium chloride, tris(hydroxymethyl)aminomethane were purchased from JT. Baker (Mallinckrodt, Deventer, The Netherlands). Acetone, ethylenediamine tetraacetic acid (EDTA), and Hydrochloric acid were purchased from Merck (Schiphol-Rijk, The Netherlands). All solutions were prepared using water purified by a Milli-Q Ultrapure Water Purification System from Millipore (Amsterdam, The Netherlands), and were vacuum filtered through a 0.22 μm pore size membrane filter from Pall (Portsmouth, Hampshire, United Kingdom).

All chromatographic separations, and on-column refolding experiments, were performed on an ÄKTA explorer 10 equipped with the UNICORN software version 5.01 from GE Healthcare (Uppsala, Sweden). Two pre-packed columns were used, a Sephacryl S-100 HR HiPrep 16/60 column (1CV = 120.6 ml, i.d. 1.6 cm, L_C 60 cm) and a Superdex 75 10/300 (1CV = 23.56 ml, i.d. 1.0 cm, L_C 30 cm).

3.2. Protein quantification

The concentration of soluble protein was estimated using the BCA protein assay, purchased from Fisher Scientific (Landsmeer, The Netherlands). Bovine serum albumin was used as the reference to build the calibration line.

3.3. Quantification of the native protein

The refolded fusion protein was digested using trypsin to obtain the mature monomer. The digestion was done using an enzyme

to substrate ratio, of 1:300 (mg:mg) [51,52]. The samples were incubated for 30 min and 25 °C using a thermomixer comfort from Eppendorf (Amsterdam, The Netherlands). The reaction was quenched by diluting the samples using a 100 mM HCL solution. The concentration of active protein was determined using RPHPLC and a calibration line constructed using human insulin as standard [53].

3.4. Inclusion bodies solubilization

The inclusion bodies were solubilized in the solubilization buffer (4 M Urea/25 mM DTT/10 mM $NaHCO_3$/0.1 mM EDTA; pH 10.5) [54]. Solutions of various protein concentrations were prepared by diluting the protein stock solution with solubilization buffer.

3.5. Size-exclusion refolding

Size-exclusion refolding was done using the Sephacryl S-100 HR HiPrep 16/60 column. The column was equilibrated, prior to the injection of the protein pulse, using a mobile phase composition of 50 mM Tris/10 mM $NaHCO_3$/0.1 mM EDTA, pH 10.10. After equilibration, a pulse of the denatured and reduced protein of 1.2 ml was injected to the column. The protein concentration in the feed pulse was varied in the range 2.0–5.0 mg/ml. The flow rate for all SEC refolding experiments was fixed at 1 ml min^{-1}.

3.6. Column calibration and characterization

The Sephacryl S-100 HR HiPrep 16/60 column was calibrated using the following set of protein standards: Aprotinin (6500), Ribonuclease A (13 700), Carbonic Anhydrase (29 000), Ovalbumin (43 000), Conalbumin (75 000). The Superdex 75 10/300 column was calibrated using Aprotinin, Ribonuclease A, Carbonic Anhydrase and Conalbumin.

The column porosities were determined by pulse experiments with blue dextran to establish the inter-particle porosity (ε_b) and with acetone to determine the total porosity (ε_t).

3.7. Generation of the Damköhler maps

The systematic analysis was conducted for different scenarios generated varying Da_I on the range of 0.1–100 and Da_{II} on the range of 1.0–200. The simulations were done using the Sephacryl S100 1.6/60 column, a feed concentration of the denatured and reduced ($C_{f,D\&R}$) protein of 1.0 mg/ml, an injection volume of 1 ml and a flow rate of 1.0 ml min^{-1}. Lysozyme was used as model protein, with a molecular weight as a monomer of 14 700 g/mol. Fixing the column geometry, the gel filtration material and the flow rate, fixes the partition behavior of the various components. Accordingly, the changes on the SEC chromatogram are solely due to the competition between folding, aggregation and convection.

3.8. Size-exclusion distribution coefficient

The experimental SEC distribution coefficient (K_{av}) was calculated using the following equation:

$$K_{av} = \frac{V_e - V_0}{V_t - V_0} \tag{13}$$

where V_e is the elution volume, V_0 is the interparticle void volume and V_t is the total void volume. The interparticle void volume was determined from pulse experiments using blue dextran, and the total void volume was determined using acetone. These volumes were corrected by subtracting the contributions of the system volumes (i.e., tubing, valves, etc.).

3.9. Optimization and parameter estimation

The parameters of the selected kinetic model were determined by minimizing the sum of squared errors (SSE) using the MATLAB function fmincon. The SSE was set as the objective function and the experimental refolding yield was set as the nonlinear constraint for the optimization. The latter was defined as a nonlinear equality constraint. The objective function is presented in Eq. (14) and the yield constraint is presented in Eq. (15).

$$\min_x \mathrm{SSE}(x) = \sum_{i=1}^{n_t} (C_{\exp_i} - C_{\mathrm{mod}_i}(x, t))^2 \tag{14}$$

where x is the vector of kinetic constants, C_{exp} is the experimental concentration vector, C_{mod} is the simulated vector of concentration and n_t is the total number of elements of the vectors.

$$Y_{N,\mathrm{mod}}(x) - Y_{N,\mathrm{exp}} = 0 \tag{15}$$

where $Y_{N,mod}(x)$ and $Y_{N,exp}$ represent the modeled and experimental refolding yield, respectively.

4. Results and discussion

4.1. Distribution coefficient

The extended Ogston model (Eq. (7)) is a mechanistic model that relates the distribution coefficient, on SEC, to the solute dimensions. The distribution coefficient is an important parameter, as it influences the elution time of a solute. The extended Ogston model is a useful tool to model SEC refolding, since it allows for the estimation of the solute dimensions, that later can be used to estimate various mass transfer parameters (Table 1), from the experimentally determined distribution coefficient. But before using it, the model needs to be calibrated. Fig. 4 presents the calibration results, obtained for the gel filtration materials Superdex 75 and Sephacryl S100. The model parameters, fiber fraction and fiber radius, are presented in the inset of Fig. 4. The data presented shows that the model is well suitable to predict the behavior of the distribution coefficient. Compared to an empirical model, such as the one described by Eq. (16), the extended Ogston model limits the predictions to reasonable values. By reasonable it is meant $K_{av} > 0$. This is not the case for the empirical model which may yield a $K_{av} < 0$ (Fig. 4B), specially for large molecules. Fig. 4B presents a comparison between the predicted distribution coefficients and the experimentally measured ones. The distribution coefficients were estimated with the extended Ogston model (■) and the empirical model (♦). The data show that (1) the extended Ogston model predicts reasonably well the experimental distribution coefficient within the range $0 < K_{av} < 0.6$, and (2) the linear model (Eq. (16)) lead to a $K_{av} < 0$ as the solute size increases. It is interesting to point out that neither model is capable to predict the distribution coefficient of Aprotinin (Apr) on the Sephacryl S100 material. One plausible explanation for this discrepancy might be that the shape of Aprotinin cannot be represented by a spherical shape, as it was done using Eq. (8) for the other protein standards, whose distribution coefficient was accurately predicted.

$$K_{av} = s \log(\mathrm{MW}) + b \tag{16}$$

4.2. Column model validation

The size-exclusion refolding chromatogram has contributions of (1) the separation (i.e. mass transport and transfer) and (2) the reaction (i.e., refolding and aggregation). Accordingly, before any kinetic information can be derived, the contribution of the separation to the chromatogram must be accurately predicted. The ability

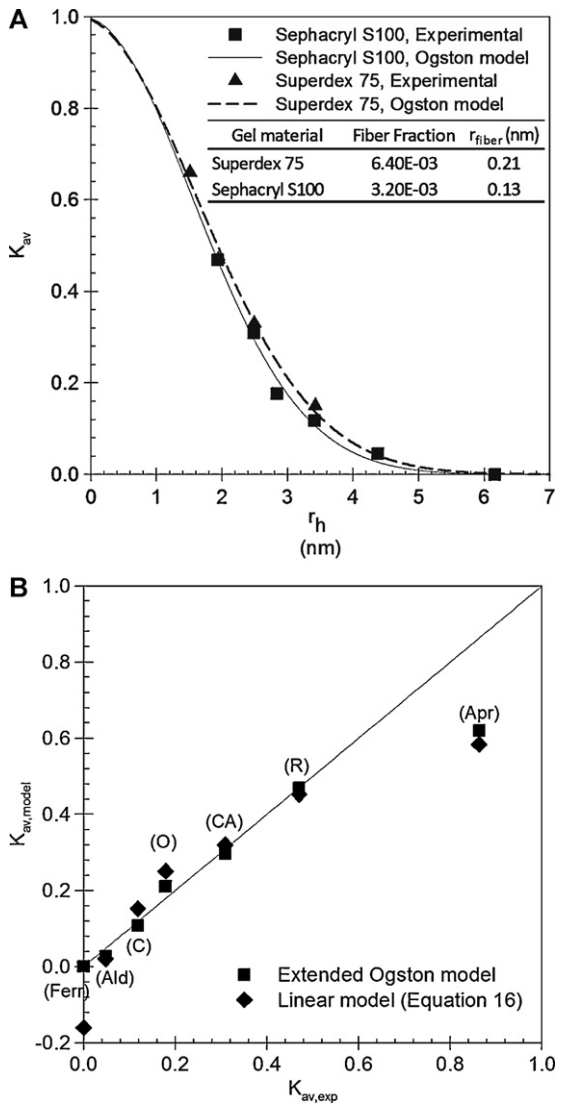

Fig. 4. (A) Average distribution coefficient experimental and model predictions. Solid line: distribution coefficient, predicted by Eq. (7), for the Sephacryl S100 16/600 column. (■) Experimental distribution coefficient, determined with the Sephacryl S100 16/600 column and Eq. (13). Dashed-line: distribution coefficient, predicted by Eq. (7) for the Superdex 75 10/300 column. (▲) Experimental distribution coefficient, determined with the Superdex 75 10/300 column. (B) Distribution coefficient parity plot for the Sephacryl S100 16/600 column. (■) Extended Ogston model. (♦) Eq. (16). (Apr) Aprotinin, (R) Ribonuclease A, (CA) Carbonic Anhydrase, (O) Ovalbumin, (C) Conalbumin, (Ald) Aldolase, (Ferr) Ferritin.

of the model, to predict the separation, was evaluated by comparing the modeled and experimental chromatogram of a set of protein standards. Fig. 5A and B presents this comparison, for the Sephacryl S100 16/600 and the Superdex 75 10/300, respectively. The modeled chromatograms were calculated with Eqs. (1)–(4b), omitting the source terms. The data in Fig. 5 show that the model is well capable to describe the separation behavior using the properties of the column material (e.g., d_p, d_{pore}), the column (e.g., L_c, D_c), the packing (i.e., ε_b, ε_t), the solutes (i.e., MW, D_f, etc.) and the operational variables (i.e., flow rate, injection volume, feed concentration). Having shown that the model can describe the contribution of the separation to the SEC chromatogram, opens the possibility to assess the contribution of the reaction to the SEC refolding chromatogram, using the column model (Eqs. (1)–(4)) together with a selected reaction mechanism (Section 2.5).

4.3. Systematic analysis of the reaction mechanisms

The systematic analysis was done to evaluate the effect that the competition between convection, refolding and aggregation, has on the SEC refolding chromatogram and the refolding yield. Since this competition is also dependent on the reaction mechanism, the analysis was conducted for the three mechanisms presented in Section 2.5, exemplifying this dependency. The effect on the SEC refolding chromatogram was studied with the aid of a Damköhler map. This map was built for each reaction mechanism using the Damköhler numbers for folding (Da_I) and aggregation (Da_{II}), presented in Table 5. The maps were built such that the rate of folding increases from left to right, whilst the rate of aggregation increases from top to bottom. To assess the competition between folding and aggregation Eq. (17) is used, in conjunction with the Damköhler map. This equation relates the characteristic times of the two reactions. A $\Psi > 1$ indicates that aggregation is slower than folding, $\Psi < 1$ indicates that aggregation is faster than folding, and $\Psi = 1$ indicates that both processes have equal characteristic times. The chromatograms presented on the Damköhler maps were

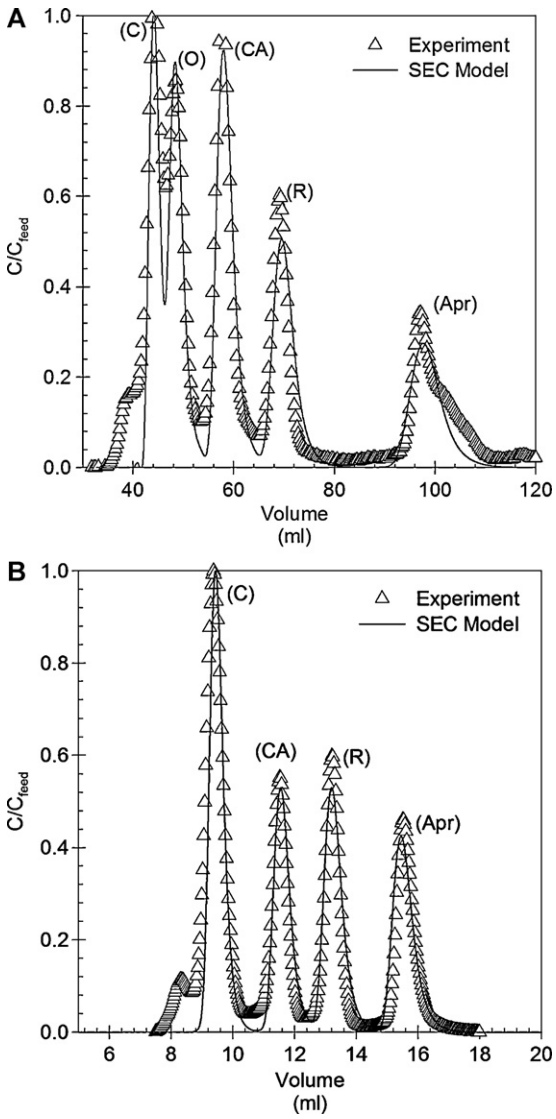

Fig. 5. Modeled and experimental chromatograms. (A) Chromatograms corresponding to the Sephacryl S100 16/600. (B) Chromatograms corresponding to the Superdex 75 10/300. Protein standards: C: Conabumin, O: Ovalbumin, CA: Carbonic Anhydrase, R: Ribonuclease A, Apr: Aprotinin.

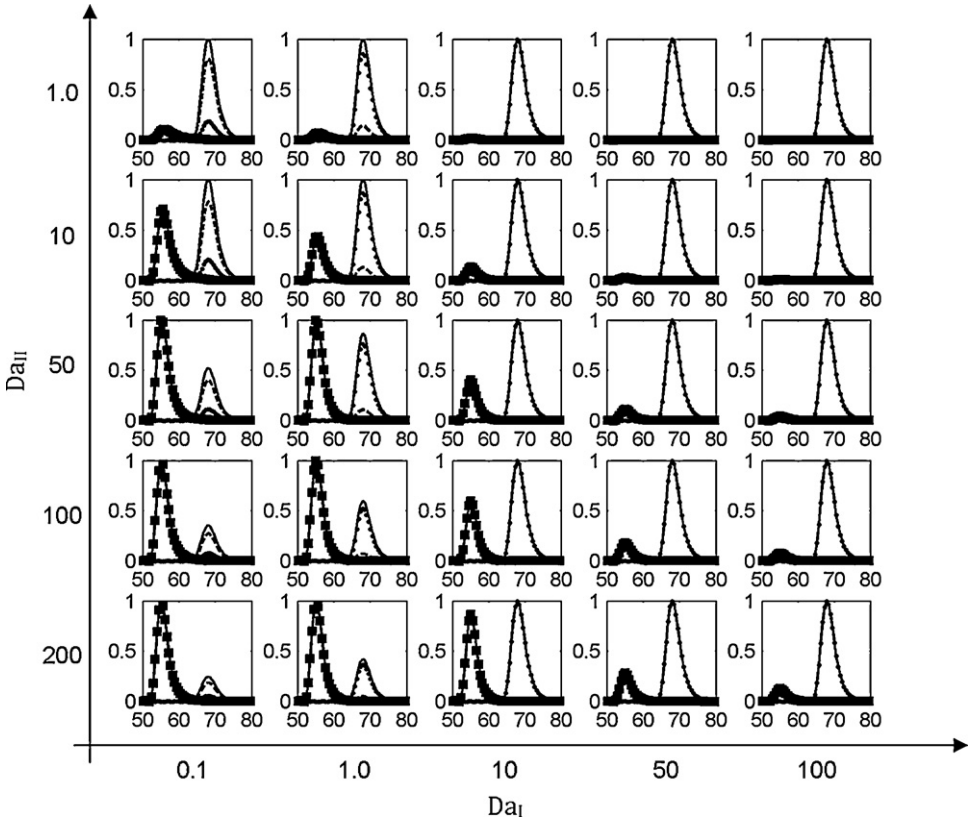

Fig. 6. Damköhler map for mechanism 1. Solid-line: total protein, dots (·): native protein (N), dash-line: intermediate (I), (■): aggregate class 3 (A_3), i.e. trimer. The insets have y-axis C/C_{max}, and x-axis volume (ml). ψ (Eq. (17)) increases from left to right and from bottom to top. Simulations done with Eqs. (1)–(4) and Eqs. (9a)–(9c).

obtained by numerical simulations using the settings described in Section 3.7 and Eqs. (1)–(4), (9a)–(9c), (10a)–(10c), and (11a)–(11c). The molecular weight of the aggregates (e.g., dimer, trimer) was obtained using, as basis of calculation, the molecular weight of the monomer model protein (i.e., lysozyme 14 700 g/mol). The hydrodynamic radius and distribution coefficient, of each component contemplated in a given mechanism, were estimated with Eqs. (8) and (7), respectively.

$$\Psi = \frac{Da_I}{Da_{II}} = \frac{\tau_{Agg}}{\tau_{fold}} \quad (17)$$

Fig. 6 shows the Damköhler map for mechanism 1. The simulations were carried out using an aggregation number (n) equal to 3 and an aggregation order (p) equal to 3 [11]. It is evident from the simulated chromatograms, presented in Fig. 6, that the separation between the aggregate class 3 (A_3) and the monomers (native and misfolded) is likely to occur during SEC refolding of lysozyme, using a Sephacryl S100 gel filtration material. This result is in line with experimental SEC refolding data of lysozyme, as presented by Batas et al. [26]. Their experiments showed the presence of a single aggregated peak and a single monomer peak, after SEC refolding of denatured lysozyme. These results are clearly captured by mechanism 1, which assumes the aggregates as a single component, either A_3 or A_2 according to the aggregation number. This comparison is pertinent because it exemplifies the usefulness of the Damköhler map, which provides information (the simulated chromatograms) that can be directly compared to experimental data, aiding the decision of whether a given mechanism is suitable or not. Figs. 7 and 8 present the Damköhler map for mechanisms 2 and 3, respectively. It is important to reiterate that the major difference between the mechanisms (i.e., 1, 2, and 3) is primarily the representation of protein aggregation. That is higher-order for mechanism 1, chain polymerization for mechanism 2, and cluster–cluster polymerization for mechanism 3. The Damköhler maps of both the chain polymerization and the cluster–cluster polymerization display two aggregate peaks (Figs. 7 and 8), corresponding to an aggregate class 2 (A_2) and 3 (A_3), which are formed as a result of monomer–monomer and monomer–multimer interactions, accounted by mechanisms 2 and 3. This trait opens the possibility to use these mechanisms for the analysis of SEC refolding data displaying unresolved chromatograms or multiple peaks associated with aggregates. It is relevant to point out that as the aggregation rate of mechanism 3 increases, the mass of the aggregates formed (i.e., A_2, A_3) decreases (Fig. 8). This is because protein aggregation in this mechanism proceeds not only by monomer–multimer and monomer–monomer interactions, but also by multimer–multimer interactions. Thus, as the mass of aggregate class 2 (A_2) and 3 (A_3) decreases, the mass of higher order aggregates (e.g. A_4, A_5) increases (data not shown).

The effect of the competition between folding and aggregation (ψ, Eq. (17)) on the refolding yield (Y_N) is presented in Fig. 9, for the three reaction mechanisms considered. The simulations show, that the SEC refolding yield (Y_N) increases as ψ increases. This is because as ψ increases, so does the τ_{Agg}/τ_{fold} ratio, which translates in a faster folding rate and a slow aggregation rate. It is important to notice that the increase on the refolding yield is sharper for mechanism 1, than for either mechanism 2 or 3 that follow practically equal trends (Fig. 9). The sharp increase for mechanism 1 is explained by the not so aggressive aggregation. Meaning, that the intermediate (I) is only consumed by one reaction on the aggregation pathway, which is not the case for either mechanism 2 or 3 (Fig. 9). For these mechanisms, the intermediate (I) is consumed by multiple reactions in the aggregation pathway, resulting in a relatively higher mass flow committed to the formation of aggregates, adversely affecting the refolding yield. The fact that the change in refolding yield follows practically the same trend for both mech-

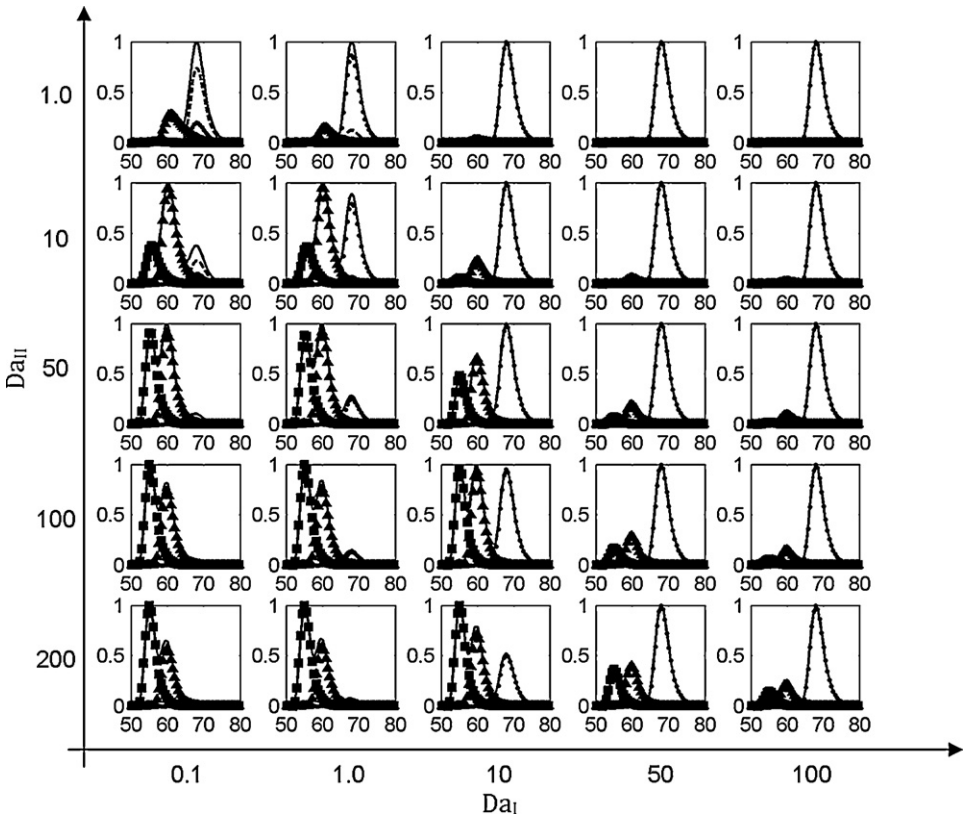

Fig. 7. Damköhler diagram for mechanism 2. Solid-line: total protein, dash-line: intermediate (I), dots (·): native (N), (▲): aggregate class 2, i.e. dimer (A_2), (■): aggregate class 3, i.e. trimer (A_3). The insets have y-axis C/C_{max}, and x-axis volume (ml). ψ (Eq. (17)) increases from left to right and from bottom to top. Simulations done with Eqs. (1)–(4) and Eqs. (10a)–(10e).

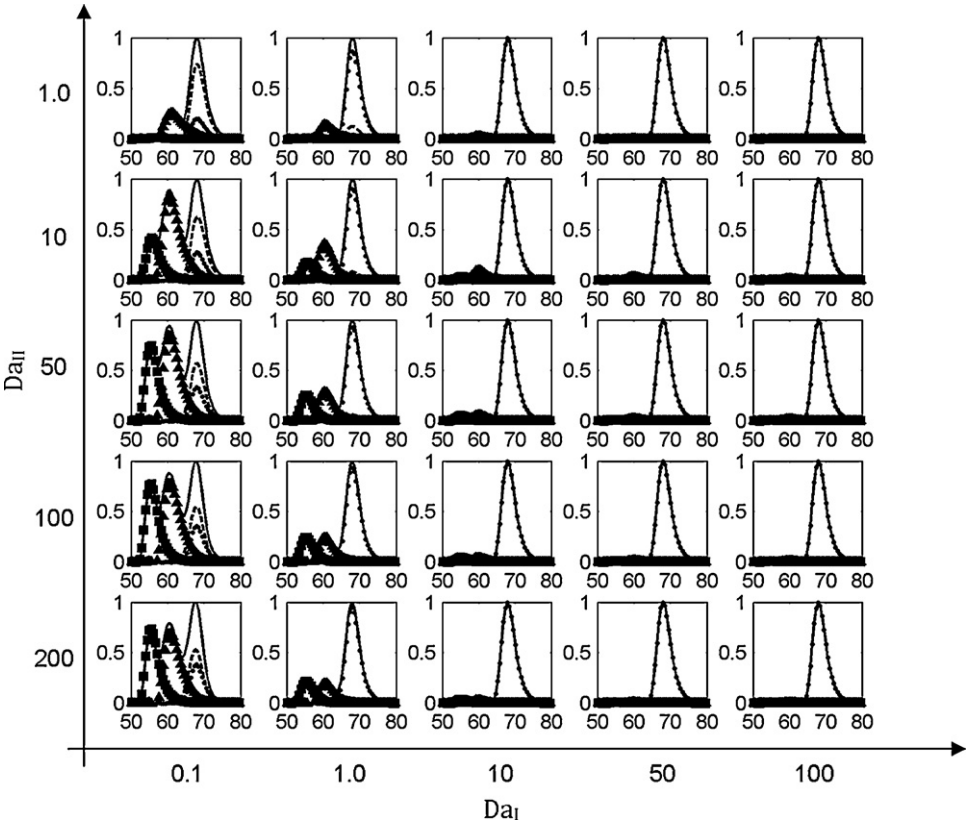

Fig. 8. Damköhler map for mechanism 3. Solid-line: total protein. Dots (·): native protein (N), dash-line: intermediate (I), (▲): aggregate class 2, i.e. dimer (A_2), (■): aggregate class 3, i.e. trimer (A_3). The insets have y-axis C/C_{max}, and x-axis volume (ml). ψ (Eq. (17)) increases from left to right and from bottom to top. Simulations done with Eqs. (1)–(4) and Eqs. (11a)–(11c).

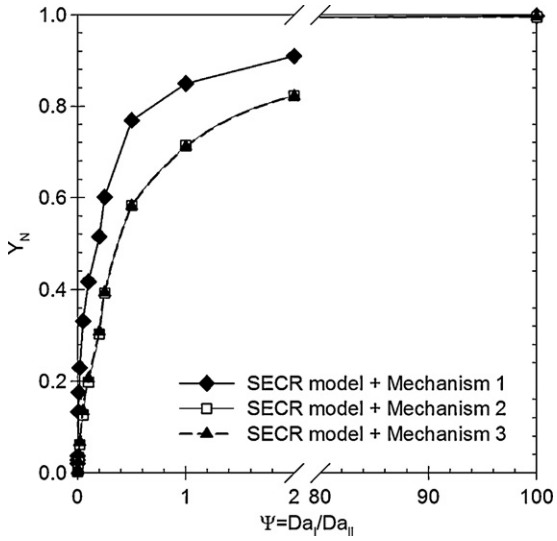

Fig. 9. Effect of the competition between folding and aggregation (ψ) on the SEC refolding yield (Y_N).

anisms 2 and 3, shows that the multimer–multimer interactions accounted by mechanism 3 only, do not affect the rate of folding, as they do not consume the intermediate (I).

So far the discussion has revolved around the competition between folding and aggregation (ψ) and its effect on the refolding yield, but what about the role of convection? To answer this question the data presented in Table 6 is used. Table 6 presents those situations were equal values of ψ are obtained for different combinations of Da_I and Da_{II}. The data show, that changes in the SEC refolding yield are expected for equal values of ψ, mainly as a consequence of the competition between (1) convection (τ_{conv}) and folding (τ_{fold}), and (2) convection and aggregation (τ_{Agg}), rather than the competition between folding and aggregation (ψ) alone, as this is constant for the same value of ψ. Moreover, the data also shows that the magnitude of the change in refolding yield will depend on the type of aggregation mechanism followed by the protein (Table 6). It is important to point out though, from the data in Table 6, that (1) the lowest refolding yield is expected when $\tau_{fold} > \tau_{conv} > \tau_{Agg}$, as in this situation folding is the slowest process; (2) a significant increase in refolding yield is predicted when $\tau_{conv} > \tau_{fold} > \tau_{Agg}$ or $\tau_{fold} = \tau_{conv} = \tau_{Agg}$, and (3) the highest refolding yield is expected when $\tau_{conv} > \tau_{Agg} > \tau_{fold}$, as in this situation folding is the fastest process.

Overall, from the previous analysis (characteristic time analysis) it can be concluded that changing the residence time ($\tau_{conv} = L_C/u$), either by varying the velocity or the bed height, will result in a change in the SEC refolding yield. And whether this change is positive or adverse depends on how the residence time compares with the characteristic times of folding and aggregation. Experimental evidence showing the correlation between residence time and the SEC refolding yield of lysozyme [55], supports the conclusions derived from the data in Table 6. The fact that the experimental observations corroborate the theoretical ones, demonstrates the usefulness of the characteristic time analysis. This analysis provides information that can be directly compared to experimental data and that primarily serves to study the effect of the residence time on the SEC refolding yield.

In summary, the systematic analysis, composed of the Damköhler map and the characteristic time analysis, yields information on two levels: (1) simulated SEC refolding chromatograms, in the form of Damköhler maps (Figs. 6–8), provide information about the effect that a given reaction mechanism (e.g., mechanisms 1, 2 and 3) has of the peak shape, mass distribution, and resolution of the different components; and (2) the characteristic time analysis provides information about the effect that the competition between convection τ_{conv}, folding τ_{fold} and aggregation τ_{Agg}, has on the SEC refolding yield (Fig. 9 and Table 6). Together these two information streams comprise the foundations of the 'Systematic analysis' presented in Fig. 3, and they are use to select the suitable mechanism that captures best the behavior of the SEC refolding data.

4.4. Application of the rational strategy: fusion protein case

Until this point, it has been clearly shown what the rational strategy to determine a suitable reaction mechanism is all about, in terms of the information offered and how this information can be directly compared to SEC refolding experimental data. The question now is how can this strategy be applied to a practical case? To answer this question SEC refolding data of the model protein (i.e., fusion protein) was used to determine, using the rational strategy (Fig. 3), a suitable reaction mechanism that describes the experimental SEC refolding data best.

Fig. 11A presents a typical chromatogram from the on-column refolding experiments showing three distinct peaks. These peaks are the result of the reaction(s) happening concurrently with the separation during the SEC refolding experiment. The three peaks were collected, analyzed by BCA to determine the total protein concentration and by RPHPLC to determine the concentration of native protein (i.e., fusion protein). Peak 1 (V_e ~35.40 ml) and 2 (V_e ~47.34 ml) are constituted by high molecular weight aggregates, and peak 3 (V_e ~55.40 ml) contains monomers, native and non-native protein. The mass of non-native protein was estimated by mass balances and the RPHPLC measurements.

Before any simulations could be performed, or any kinetic information could be derived, a reaction mechanism needed to be selected. This selection process is done during the phases 'Data processing' and 'Systematic analysis' (Fig. 3), with the aid

Table 6
Effect of the competition between convection, folding and aggregation on the SEC refolding yield.

Da_{II}	Da_I	$\psi = Da_I/Da_{II}$	Mechanism 1 Y_N	Mechanism 2 Y_N	Mechanism 3 Y_N	Competition
10	0.1	0.01	0.07	0.03	0.03	$\tau_{fold} > \tau_{conv} > \tau_{Agg}$
100	1		0.18	0.03	0.04	$\tau_{conv} > \tau_{Agg} \wedge \tau_{conv} = \tau_{fold}$
1	0.1	0.1	0.14	0.11	0.11	$\tau_{fold} > \tau_{conv} \wedge \tau_{conv} = \tau_{Agg}$
10	1		0.41	0.22	0.22	$\tau_{conv} > \tau_{Agg} \wedge \tau_{conv} = \tau_{fold}$
100	10		0.42	0.20	0.21	$\tau_{conv} > \tau_{fold} > \tau_{Agg}$
100	50	0.5	0.70	0.53	0.53	$\tau_{conv} > \tau_{fold} > \tau_{Agg}$
200	100		0.77	0.58	0.58	$\tau_{conv} > \tau_{fold} > \tau_{Agg}$
1	1	1	0.72	0.65	0.65	$\tau_{conv} = \tau_{Agg} = \tau_{fold}$
10	10		0.76	0.64	0.64	$\tau_{conv} > \tau_{Agg} \wedge \tau_{Agg} = \tau_{fold}$
50	50		0.79	0.67	0.67	$\tau_{conv} > \tau_{Agg} \wedge \tau_{Agg} = \tau_{fold}$
100	100		0.85	0.71	0.71	$\tau_{conv} > \tau_{Agg} \wedge \tau_{Agg} = \tau_{fold}$
1	10	10	0.96	0.94	0.94	$\tau_{conv} > \tau_{fold} \wedge \tau_{conv} = \tau_{Agg}$
10	100		0.98	0.96	0.96	$\tau_{conv} > \tau_{Agg} > \tau_{fold}$

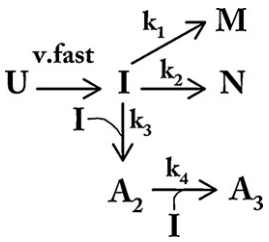

Fig. 10. Mechanism describing the competition between folding, misfolding and aggregation, during SEC refolding of the model fusion protein. U: unfolded protein, I: intermediate, M: misfolded protein, N: native protein, A_2: aggregate class 2, A_3: aggregate class 3. k_1, k_2, k_3, and k_4 are the reaction kinetic constants.

of the information extracted from the experimental SEC refolding data (Fig. 11A). From the SEC experimental data the following observations were made: (1) the presence of two peaks (i.e., peaks 1 and 2) carrying high molecular weight aggregates indicated that, representing protein aggregation as a higher-order reaction was not applicable to the presented case, as this representation captures aggregation as a single component, leaving thus as suitable options either the chain-polymerization or the cluster–cluster polymerization. The cluster–cluster polymerization was ruled-out since substantial aggregation was not observed and the chain-polymerization is sufficient to capture the aggregation displayed by the model protein; and (2) the mass of monomers in peak 3 contains native and non-native protein, hence the mechanism should be able to account for the formation of the misfolded material (i.e. non-native protein). Taking these observations into account, lead to the creation of the mechanism presented in Fig. 10. This mechanism has the following features (1) the mass flow committed to the formation of aggregated protein is captured by the chain polymerization mechanism, representing the mass committed to peaks 1 and 2 by an aggregate class size 2 (A_2) and 3 (A_3), respectively; and (2) the mass flow committed to the formation of native (N) and misfolded protein (M) is represented by two irreversible reactions. The kinetic parameters of the mechanism were determined using a constrained non-linear optimization algorithm (Section 3.9), during the 'Systematic analysis' phase of the rational strategy (Fig. 3). Fig. 11A presents the modeled and the experimental SEC refolding chromatograms. The data shows that the experimental SEC refolding chromatogram is well represented by the SEC reaction model with $k_1 = 1.03 \times 10^{-4} \pm 9.7 \times 10^{-5}\ s^{-1}$, $k_2 = 0.08 \pm 8 \times 10^{-3}\ s^{-1}$, $k_3 = 1.69 \pm 8.2 \times 10^{-2}\ l\,mmol^{-1}\ s^{-1}$ and $k_4 = 2.42 \pm 0.19\ l\,mmol^{-1}\ s^{-1}$. These set of kinetic parameters complete the refolding yield function describing the SEC refolding reactor ('Output phase' Fig. 3). The yield function is the equation that links the operational variables of the size-exclusion refolding reactor with the refolding yield. In other words, this equation links the separation to the refolding yield. Fig. 11B presents the comparison between the predicted and the experimental refolding yield as a function of the concentration of denatured and reduced protein fed ($C_{f,D\&R}$). The data shows that the SEC refolding model captures well the trend displayed by the experimental data, as the model predicts the decrease of the refolding yield as $C_{f,D\&R}$ increases. However, the model prediction is not 100% accurate, which may plausibly be explained by the effect that the change in chemical composition has on the kinetic constants, not accounted for in the presented model. The change in chemical composition is caused, during SEC refolding, by the separation of the protein concentration wave from the concentration waves of DTT and urea. This hypothesis is supported by observations obtained from denaturation–renaturation studies [8,56] and from refolding studies [11,12], as these have shown the dependency of the kinetic constants on the concentration of urea and DTT. Including such dependency into the SEC reaction model is not difficult, the challenge lies in determining (1) the appropriate relations that capture the effect of the transient chemical composition on the kinetics and (2) those reactions, within the competitive reaction scheme, that are being affected.

4.5. Theoretical comparison between batch dilution refolding and SEC refolding: effect of the reactor type

Having develop a modeling tool, capable of describing a size exclusion refolding reactor (SECR) raised the following question, how does the theoretical (modeled) refolding yield of the SECR compares to the refolding yield of a batch refolding reactor, if for both reactors the same reaction scheme holds? To answer this question the refolding yield of both reactors was estimated as a function of the concentration of denatured and reduced protein ($C_{f,D\&R}$). This concentration corresponds, in the case of the batch reactor, to the total protein concentration attained right after the dilution of the concentrated denatured and reduced protein solution, with refold-

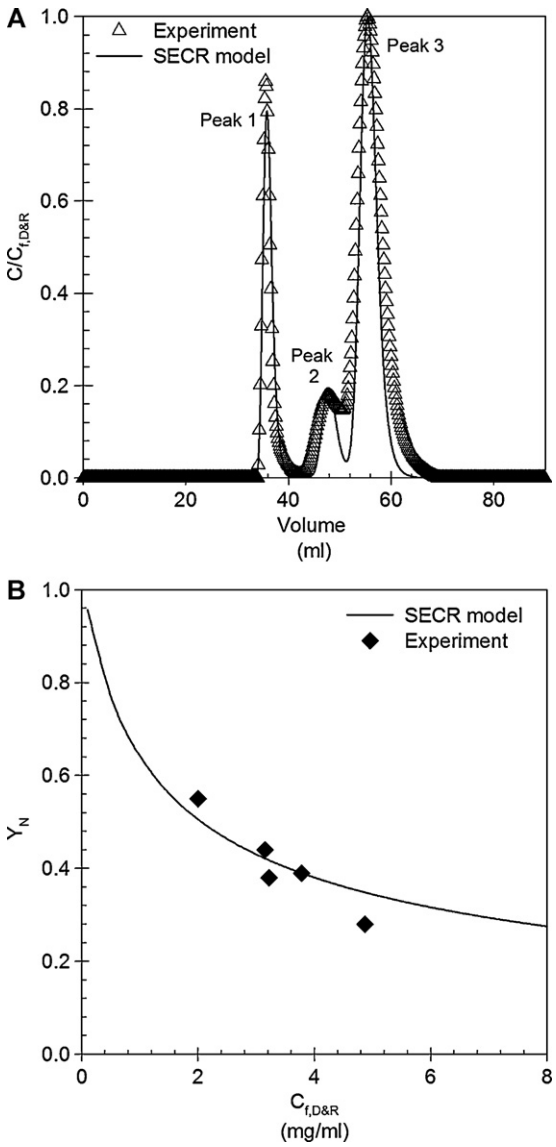

Fig. 11. Application of the model based strategy. (A) Experimental and modeled SEC refolding chromatograms. The simulated SEC refolding chromatogram was obtained with Eqs. (1)–(4b) and the reaction mechanism presented in Fig. 10. (B) Experimental and predicted SEC refolding yield (Y_N). The modeled SEC refolding yield was obtained using Eqs. (1)–(4b) and the reaction mechanism presented in Fig. 10.

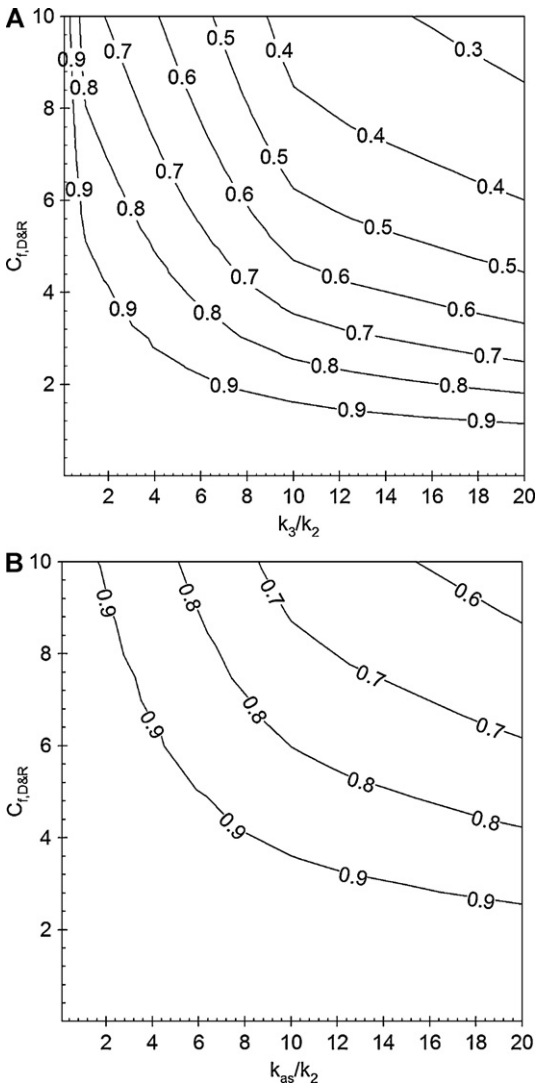

Fig. 12. Contour plots for the refolding yield (Y_N) of mechanism 1, as a function of $C_{f,D\&R}$ and the kinetic competition between aggregation and folding (k_3/k_2 or k_{as}/k_2). (A) Batch refolding reactor. (B) Size-exclusion refolding reactor. The contour lines represent the refolding yield.

ing buffer. In the case of the SECR, this concentration corresponds to the total protein concentration fed to the size-exclusion column. The simulations were done using lysozyme as the model protein, a $C_{f,D\&R}$ ranging from 0.01 to 10 mg ml^{-1} and the kinetic competition between aggregation and folding, represented by the ratio of their kinetic constants (k_3/k_2 or k_{as}/k_2). The SECR was modeled using Sephacryl S100 1.6/60 column, operating at 1 ml min^{-1} and an injection volume of 1 ml.

Figs. 12 and 13 present the refolding yield contour plots as a function of $C_{f,D\&R}$ and the kinetic competition between aggregation and folding, for mechanisms 1 and 2, respectively. The contour lines represent the expected refolding yield. Mechanism 3 was not included in this analysis because its refolding yield follows the same trend as the one displayed by mechanism 2. Figs. 12 and 13(A and B) present the simulation results belonging to the batch reactor and the SECR, respectively. From the figures, it is evident that the SECR is expected to perform better than the batch reactor. This claim is supported by the relatively high refolding yields expected for the SECR, operating at relative high $C_{f,D\&R}$ and facing relative unfavorable kinetic scenarios (high k_3/k_2 or k_{as}/k_2 ratio). Accordingly based on these observations, it can be concluded that the advantage of the SECR comes from primarily two things (1) the relatively low local protein concentration, which primarily slows down the aggregation rates and (2) the reduction of the multimer–monomer interactions as a result of their separation, own to their differences in distribution coefficients, which decreases the rate of aggregation. In a nut-shell, the expected differences in refolding yield between the two reactors boil down to a difference in the aggregation rates, which are relatively high for the batch reactor compared to the SECR.

The previous conclusions, based on sound mechanistic modeling, are now compared to actual experimental data. Experimental evidence, comparing the refolding yield of lysozyme refolded by batch dilution and size-exclusion [55], confirm the model predictions, as the data also shows that the SEC refolding reactor gives a higher refolding yield compared to the batch reactor as $C_{f,D\&R}$ increases. Although, the experimental evidence corroborates the model predictions, there is still a discrepancy between the modeled and the experimental data. The model predicts a difference in refolding yield between the two reactors (for a given $C_{f,D\&R}$) higher than the one exhibit by the experimental data [55]. This discrepancy may plausibly be explained by the effect that the change in chemi-

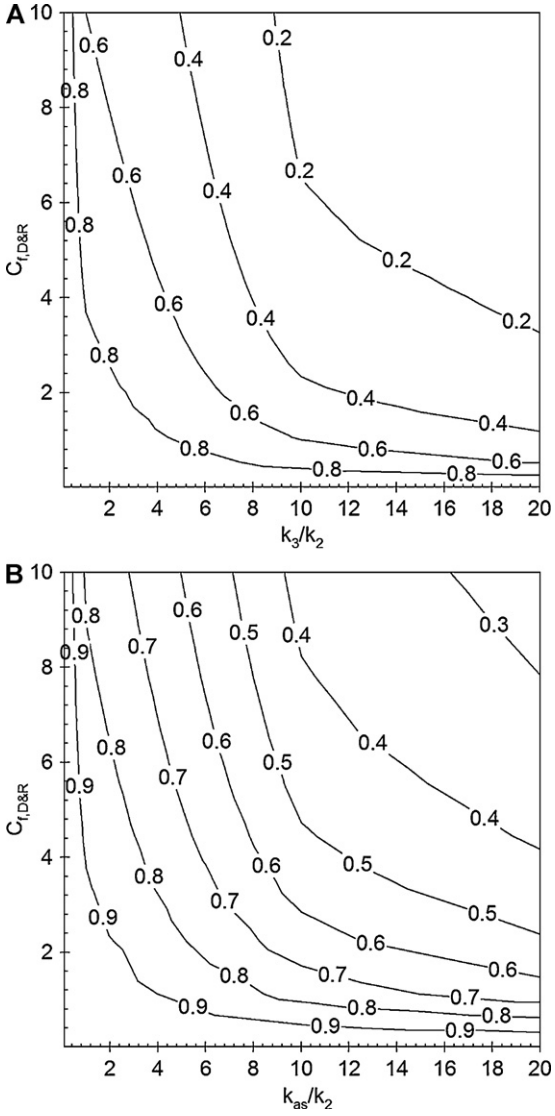

Fig. 13. Contour plots for the refolding yield (Y_N) of mechanism 2, as a function of $C_{f,D\&R}$ and the kinetic competition between aggregation and folding (k_3/k_2 or k_{as}/k_2). (A) Batch refolding reactor. (B) Size-exclusion refolding reactor. The contour lines represent the refolding yield.

cal composition has on the refolding and aggregation kinetics. This claim is supported by the following (1) both, $C_{f,D\&R}$ and the chemical composition have been shown to be contributors to the changes in refolding yield [11,12], and (2) the model accounts only for the effect of the former (i.e., $C_{f,D\&R}$) but not of the latter. Accordingly, it can be concluded that accounting only for the contribution $C_{f,D\&R}$ alone, represented by the effect it has on the reaction rates, is necessary but not sufficient to completely explain the variation of the SEC refolding yield. Therefore, the effect that the change in chemical composition has, specially the change in the concentrations of urea and DTT, on the refolding kinetics, should be incorporated in the future modeling efforts.

5. Further discussion

The presented methodology has been build to deal with an inverse problem, meaning that it starts from the SEC chromatogram obtained after on-column refolding, and goes back, using a mathematical representation of the SEC refolding reactor, to infer a plausible competitive reaction scheme. As shown in the text, this approach gives an estimate of the kinetic mechanism and its parameters. Using SEC refolding data offers an advantage over using batch refolding data alone. This advantage lies on the SEC refolding chromatogram, as this provides information about the mass distribution towards aggregates and monomers after refolding. As shown, this information, combined with mass balances and activity measurements, aids in deciding the number of species a given mechanism should account for. This sort of rational cannot be done only with the transient concentration profile of the native protein, which is the type of data commonly derived from batch refolding studies.

Deriving SEC refolding data only with a size-exclusion column and a UV detector is straightforward; however it suffers from an inherent limitation. This limitation is connected to the fact that the elution volume of a solute in size-exclusion chromatography not only depends on its molecular mass but also on its shape (e.g., coil, rod, sphere, etc.). In the presented work this shortcoming was overcome by either one of the following alternatives: (1) assuming a spherical shape (Eq. (8)) in order to estimate the solute dimensions and thereafter its distribution coefficient (Eq. (7)), and (2) Determine the distribution coefficient from the measured elution volume (Eq. (15)), as this quantity contains the contribution of shape and size. Considering the previous discussion, it becomes evident that the presented methodology will benefit from the use of an absolute size-exclusion (ASEC) experimental set-up [57]. Such experimental set-up consist of a size-exclusion, a UV-detector, a refractive index (RI) detector and a dynamic light scattering (DLS) detector. ASEC allows the direct measurement of protein size and shape [58,59], that can be directly put into the presented methodology and modeling tool, increasing its capability to tease out information from the SEC refolding data. This is the next step on the development of a more robust modeling based strategy to study size-exclusion protein refolding.

6. Conclusions

The work presented in this paper describes a novel methodology to select a suitable reaction scheme, describing the competition between refolding and aggregation, using SEC refolding data. The methodology supported by a library of reaction mechanisms and a modeling tool describing the size-exclusion refolding reactor, was successfully applied to the analysis of SEC refolding data of an industrially relevant protein. This work further achieved: (1) the development of a modeling tool to describe a size-exclusion refolding reactor, based on mechanistic relationships including: mass transport, mass transfer, and reaction kinetics (i.e. rate law and mass action formalisms), linking the separation to the refolding yield (Y_N); (2) the development of a library of reaction mechanisms including the classical first-order folding competing against a higher-order aggregation, as well as more elegant mechanisms representing protein aggregation by chain polymerization and by cluster–cluster polymerization; (3) an assessment of the effect that the competition between folding (τ_{fold}), aggregation (τ_{Agg}) and convection (τ_{conv}) has on the SEC refolding chromatogram and the refolding yield, using a characteristic time analysis. From this analysis became evident that the residence time affects the SEC refolding yield, however whether this effect is positive or adverse depends on how the residence time compares to the characteristic times of folding and aggregation; (4) a comparison, based on mechanistic modeling, of the conventional batch dilution refolding and the SEC refolding leading to the following conclusion. Compared to batch dilution refolding, SEC refolding is expected to give a higher refolding yield as $C_{f,D\&R}$ increases owing to (1) a relative low local protein concentration, which primarily slows down the aggregation rates and (2) the mitigation of monomer–multimer interactions. Comparison with literature data confirmed these claims and further show that the modeling tool captures the trend of the SEC refolding yield well, based solely on the effect that $C_{f,D\&R}$ has on the reaction rates. Furthermore, it also made evident that the model predictions are highly likely to improve if the effect of the dynamic change in chemical composition on the reaction kinetics is accounted for. This, however falls beyond the scope of the presented work and is left for future work.

Acknowledgements

The authors would like to thank Dr. Emrah Nikerel for the valuable discussions about modeling, parameter estimation and reaction kinetics, which were extremely beneficial for the culmination of the presented work.

References

[1] M.H. Werner, G.M. Clore, A.M. Gronenborn, A. Kondoh, R.J. Fisher, FEBS Lett. 345 (1994) 125.
[2] B. Batas, J.B. Chaudhuri, Biotechnol. Bioeng. 50 (1996) 16.
[3] B. Batas, H.R. Jones, J.B. Chaudhuri, J. Chromatogr. A 766 (1997) 109.
[4] A. Jungbauer, W. Kaar, R. Schlegl, Curr. Opin. Biotechnol. 15 (2004) 487.
[5] X. Geng, C. Wang, J. Chromatogr. B 849 (2007) 69.
[6] M. Li, Z.-G. Su, J.-C. Janson, Protein. Expr. Purif. 33 (2004) 1.
[7] S. Endo, Y. Saito, A. Wada, Anal. Biochem. 131 (1983) 108.
[8] T.E. Creighton, J. Mol. Biol. 129 (1979) 235.
[9] W. Shalongo, R. Ledger, M.V. Jagannadham, E. Stellwagen, Biochemistry 26 (1987) 3135.
[10] Y. Ding, L. He, A.P.J. Middelberg, Chem. Eng. Sci. 63 (2008) 4333.
[11] D.L. Hevehan, E.D.B. Clark, Biotechnol. Bioeng. 54 (1997) 221.
[12] A.M. Buswell, A.P.J. Middelberg, Biotechnol. Bioeng. 83 (2003) 567.
[13] A.M. Buswell, A.P.J. Middelberg, Biotechnol. Prog. 18 (2002) 470.
[14] L. Lapidus, N.R. Amundson, J. Phys. Chem. 56 (1952) 984.
[15] G. Guiochon, D.G. Shirazi, A. Felinger, A.M. Katti, Fundamentals of Preparative and Nonlinear Chromatography, Academic Press, 2006.
[16] S. Golshan-Shirazi, G. Guiochon, J. Chromatogr. 603 (1992) 1.
[17] W.E. Schiesser, The Numerical Method of Lines: Integration of Partial Differential Equations, Academic Press Limited, London, 1991.
[18] C.-M. Yu, S. Mun, N.-H.L. Wang, J. Chromatogr. A 1132 (2006) 99.
[19] Z. Li, Y. Gu, T. Gu, Biochem. Eng. J. 2 (1998) 145.
[20] B. Zelic, B. Nesek, Eng. Life Sci. 6 (2006) 163.
[21] C.R. Wilke, P. Chang, AIChE J. 1 (1955) 264.
[22] M.E. Young, P.A. Carroad, R.L. Bell, Biotechnol. Bioeng. 22 (1980) 947.
[23] E.J. Wilson, C.J. Geankoplis, Ind. Eng. Chem. Fund. 5 (1966) 9.
[24] Z. Ma, N.H.L. Wang, AIChE J. 43 (1997) 2488.
[25] S.F. Chung, C.Y. Wen, AIChE J. 14 (1968) 857.
[26] B. Batas, C. Schiraldi, J.B. Chaudhuri, J. Biotechnol. 68 (1999) 149.
[27] B. Batas, J. Chaudhuri, Bioprocess Biosyst. Eng. 24 (2001) 255.
[28] B.-J. Park, C.-H. Lee, S. Mun, Y.-M. Koo, Process Biochem. 41 (2006) 1072.
[29] S.S.S. Wang, C.-K. Chang, H.-S. Liu, Biochem. Eng. J. 29 (2006) 2.
[30] Z. Gu, X. Zhu, S. Ni, H. Zhou, Z. Su, J. Biochem. Biophys. Methods 56 (2003) 165.
[31] L. Hagel, M. Ostberg, T. Andersson, J. Chromatogr. A 743 (1996) 33.
[32] L. Hagel, Curr. Protoc. Protein. Sci. 8 (3) (1998).
[33] G. Sofer, L. Hagel, Handbook of Process Chromatography: A Guide to Optimization, Scale-up and Validation, Academic Press, San Diego, 1997.

[34] A.G. Ogston, Trans. Faraday Soc. 54 (1958) 1754.
[35] J.C. Bosma, J.A. Wesselingh, J. Chromatogr. B 743 (2000) 169.
[36] L. Hagel, H. Lundstrom, T. Andersson, H. Lindblom, J. Chromatogr. 476 (1989) 329.
[37] E.D.B. Clark, Curr. Opin. Biotechnol. 12 (2001) 202.
[38] E.D.B. Clark, E. Schwarz, R. Rudolph, Methods Enzymol. 309 (1999) 217.
[39] W. Weiss, T.M. Young, C.J. Roberts, J. Pharm. Sci 98 (2009) 1246.
[40] T. Kiefhaber, R. Rudolph, H.-H. Kohler, J. Buchner, Nat. Biotechnol. 9 (1991) 825.
[41] J. Maachupalli-Reddy, B.D. Kelley, E. De Bernardez Clark, Biotechnol. Prog. 13 (1997) 144.
[42] X.Y. Dong, G.Q. Shi, W. Li, Y. Sun, Biotechnol. Prog. 20 (2004) 1213.
[43] M.A. Speed, J. King, D.I.C. Wang, Biotechnol. Bioeng. 54 (1996) 333.
[44] F. Oosawa, S. Asakura, K. Hotta, N. Imai, T. Ooi, J. Polym. Sci. 37 (1959) 323.
[45] L.S. Tobacman, E.D. Korn, J. Biol. Chem. 258 (1983) 3207.
[46] F.A. Ferrone, J. Hofrichter, H.R. Sunshine, W.A. Eaton, Biophys. J. 32 (1980) 361.
[47] D. Thusius, P. Dessen, J.-M. Jallon, J. Mol. Biol. 92 (1975) 413.
[48] K.E. Van Cott, R.D. Whitley, N.H.L. Wang, Sep. Technol. 1 (1991) 142.
[49] R.D. Whitley, K.E.V. Cott, J.A. Berninger, N.H.L. Wang, AIChE J. 37 (1991) 555.
[50] J.A. Berninger, R.D. Whitley, X. Zhang, N.H.L. Wang, Comput. Chem. Eng. 15 (1991) 749.
[51] J. Winter, H. Lilie, R. Rudolph, Anal. Biochem. 310 (2002) 148.
[52] J. Winter, P. Neubauer, R. Glockshuber, R. Rudolph, J. Biotechnol. 84 (2000) 175.
[53] USP27-NF26, US Pharmacopeia (2007) 2.
[54] E.J. Freydell, M. Ottens, M. Eppink, G.V. Dedem, L.V.D. Wielen, Biotechnol. J. 2 (2007) 678.
[55] Z. Gu, Z. Su, J.-C. Janson, J. Chromatogr. A 918 (2001) 311.
[56] T.E. Creighton, J. Mol. Biol. 137 (1980) 61.
[57] J. Wen, T. Arakawa, J.S. Philo, Anal. Biochem. 240 (1996) 155.
[58] E. Folta-Stogniew, in: H. Press (Ed.), Methods in Molecular Biology, Totowa, NJ, 2006, p. 97.
[59] H. Ye, Anal. Biochem. 356 (2006) 76.