

# Liquid Chromatography and its Applicability to Protein Refolding

Proefschrift

ter verkrijging van de graad van doctor  
aan de Technische Universiteit Delft,  
op gezag van de Rector Magnificus Prof. ir. K.C.A.M. Luyben  
voorzitter van het College voor Promoties,  
in het openbaar te verdedigen op maandag 20 september 2010 om 15:00 uur

door

**Gabriel Esteban Jaramillo Freydell**  
Technologisch Ontwerper in Bioprocess Technologie  
geboren te Medellín, Colombia

Dit proefschrift is goedgekeurd door de promotor:

Prof. Dr.ir. L.A.M. van der Wielen

Copromotor: Dr.ir. M. Ottens

Samenstelling promotiecommissie:

Rector Magnificus	Voorzitter
Prof.dr.ir. L.A.M. van der Wielen	Technische Universiteit Delft, promotor
Dr.ir. M.Ottens	Technische Universiteit Delft, copromotor
Prof.dr. A. Jungbauer	University of Natural Resources and Applied Life Sciences, Austria
Dr. M.H.M. Eppink	Synthon, The Netherlands
Prof.dr. W.R. Hagen	Technische Universiteit Delft
Prof.dr. G.J. Witkamp	Technische Universiteit Delft

Reservelid:

Prof. dr. J.H. de Winde                      Technische Universiteit Delft

The studies presented in this thesis were carried out at the Bioseparations Technology section, Department of Biotechnology, faculty of Applied Sciences, Delft University of Technology. The research was financially supported by Organon (nowadays Merk Sharp & Dohme (MSD)), the CASIMIR foundation and the Delft University of Technology.

ISBN 978-90-9025676-4

*-“Yesterday is history. Tomorrow is a mystery. Today is a gift. That is why we call it the Present”-*

*- Eleanor Roosevelt-*

*-“Although nature commences with reason and ends in experience it is necessary for us to do the opposite, which is to commence with experience and from this to proceed to investigate the reason”-*

*-Leonardo da Vinci-*

*To the memory of  
Sonia Gomez  
(2010)*

# Summary

---

The advent of DNA recombinant technology during the last century opens the possibility to practically express any given gene sequence in any given expression system (e.g., prokaryotes, yeast, mammalian cells, etc.). Currently, approximately 30 gene recombinant bio-pharmaceuticals with an estimated market value of USD 50 to 60 billion, have already been introduced into therapy, and roughly 300 more are projected to be under development worldwide [1].

When it comes to the production of recombinant proteins that do not require post-translational modifications, the prokaryotic system is the expression system of choice. Within this system, *Escherichia coli* (*E. coli*) is the preferred organism mainly because of reasons such as: (a) high capacity to accumulate recombinant proteins (up to 80% of its dry weight [2]); (b) high fermentation yields are attainable using high cell density cultures (HCDC) [3], and (c) its genome can be modified with ease, as its molecular genetics are well understood. One of the main limitations associated with the expression of a gene sequence in *E. coli* is the formation of inclusion bodies (IBs). Inclusion bodies are insoluble protein aggregates, often formed as a result of the over-expression of either homologous or heterologous gene sequences. IBs accumulate inside the cell and contain the recombinant protein product in an inactive, insoluble form. To obtain the protein product in an active, soluble form, the IBs must be first solubilized and thereafter, the soluble often denatured and reduced (D&R) protein product must be refolded. These two steps, -IBs solubilization- and -protein refolding- pose a significant technical challenge. This thesis focuses on these two topics because improvements in either one, should undoubtedly, have a positive impact on the current production strategies of recombinant proteins in *E. coli*.

The first topic addressed in this thesis is the solubilization of the IBs. This topic is the core of chapter 2. Chapter 2 presents a study where quantitative criteria, namely the fraction of soluble monomer and the total protein solubility, are used in combination with a design of experiments (DOE) methodology to define the composition of the optimal solubilization buffer. The study shows that the efficient solubilization of the IBs depends on the concentration of the denaturant, the concentration of the reducing agent and the solution's pH. Furthermore, the study shows that relatively low concentrations of denaturant and reducing agent can be employed to achieve an efficient IBs solubilization step.

After defining the solubilization conditions for the IBs, the next question is how to refold (renature) the D&R soluble protein. Protein refolding occurs basically as a result of a buffer exchange step. This step decreases the concentration of the solubilization additives, inducing protein folding and disulfide bond formation. In general, this buffer exchange is done by dilution. Dilution refolding occurs after mixing a volume of the concentrated, D&R protein solution, with a volume of refolding buffer. Dilution is a relatively simple approach, and this is most likely the reason why it is preferred at an industrial scale. Arguably, the major limitation of dilution refolding is the large dilution factor(s), needed to obtain a relatively high refolding yield ( $Y_N$ ). As a result of the large dilution, large vessels and subsequently extensive concentration steps are needed, driving the production cost upward and setting a limit to the production scale. Probably owing to these reasons, alternatives to dilution refolding are being sought [4-5]. This thesis puts forward liquid chromatography, and its applicability for protein refolding.

Compared to dilution refolding, liquid chromatography offers a relatively high degree of process intensification, which is represented by the capability of performing the reaction (protein refolding) and the separation in one unit operation. Buffer exchange in liquid chromatography occurs as a result of the separation of the solubilization additives from the D&R protein. This separation can be done while the protein is bound to the adsorbent (adsorption chromatography) or flowing free down the column (gel filtration or size-exclusion chromatography). These two options are discussed in this thesis. Chapter 3 discusses ion-exchange (adsorption) chromatographic refolding and chapters 4 and 5 discuss size-exclusion chromatographic refolding.

The application of ion-exchange (IEx) chromatography to protein refolding is very promising mostly because of two main reasons: (1) The concentration of the solubilization additives can be decreased in a controllable fashion, using gradients, and (2) it has been suggested that the spatial isolation of the bound protein molecules (proteins bound to the adsorbent) decreases protein aggregation, positively affecting the refolding yield. Chapter 3 presents a quantitative study that assesses the effect of various process variables on the performance of IEx protein refolding. The process variables included: protein load, loading state, adsorbent type, and slope of the urea gradient. The performance of IEx refolding was judged on the basis of fractional mass recovery ( $f_{\text{Prot,Rec}}$ ) and refolding yield. Additionally, the study evaluates the possibility of using fractional surface coverage ( $\theta$ ) and the slope of the denaturant gradient ( $\xi$ ) as numerical descriptors of the spatial isolation of the bound molecules and the denaturant gradient, respectively. The study shows: (1) that the protein load has a direct effect on  $f_{\text{Prot,Rec}}$  and that its magnitude depends on the loading state of the

protein solution and on the adsorbent type; (2) that regardless of the IEx adsorbent used, the saturation capacity of a denatured protein is less than the native protein, which is reasonably explained by their differences in accessible binding surface area; (3) that there is a clear correlation between fractional surface coverage ( $\theta$ ) and  $f_{\text{Prot,Rec}}$ , indicating that the former could serve as a good descriptor to assess spatial isolation, and (4) that the urea gradient has a direct link with the variations on the refolding yield, which can be quantitatively estimated using the slope of the gradient ( $\xi$ ) as numerical descriptor. Lastly, the work presented in chapter 3 strongly suggests that understanding the role played by the physicochemical properties of the IEx adsorbent, is the next immediate challenge.

In contrast to IEx, the application of size-exclusion chromatography (SEC) to protein refolding is relatively simple, mostly because the interactions of the protein with the column material are negligible. Additionally, SEC is a very attractive option for protein refolding, especially as it has been shown to be successful at denatured and reduced protein feed concentrations of up to  $80 \text{ mg ml}^{-1}$ [6]. Chapter 5 presents a study where protein refolding is done using batch and continuous SEC. Continuous SEC was implemented using Simulated Moving Bed (SMB) technology. Additionally, chapter 5 presents the development and application of a modeling tool to analyze the SMBSEC refolding data. The study shows: (1) that refolding yields of up to 50% are attainable by tuning pH and the denatured and reduced protein feed concentration ( $C_{f,D\&R}$ ), and that the positive effect of pH is strongly dependent on  $C_{f,D\&R}$ ; (2) that the mechanistic tool captures the behavior of the SMBSEC refolding reactor data well, based solely on the effect that  $C_{f,D\&R}$  has on the reaction rates and (3) operating the SEC refolding reactor in a continuous mode resulted in a roughly 53 fold increase of the volumetric productivity, approximately 10 fold decrease in the solvent consumption and a roughly 4.5 fold increase in the product concentration leaving the reactor.

Besides the possibility of being used as a refolding reactor, SEC has the potential to be used as an analytical tool to study protein refolding. This is because the SEC refolding chromatogram contains important information about the distribution of monomers and aggregates after refolding. To exploit its potential however, a mathematical model(s) is required. Such model(s) should account for the contributions of the mass transport and transfer mechanisms, to isolate the contribution of the reaction pathway to the shape of the SEC refolding chromatogram. The development of such modeling tool is the main subject of chapter 4. Chapter 4 presents: (1) a systematic analysis of various refolding reaction mechanisms using Damköhler maps. The refolding mechanisms incorporate the competition between folding and aggregation. Moreover, aggregation is modeled using the conventional

higher-order model, as well as the more mechanistically adequate polymerization formalisms. (2) A comparison based on mechanistic modeling, between the size-exclusion refolding reactor (SECR) and the batch dilution refolding reactor; and (3) the application of the modeling strategy to the analysis of SEC refolding data of an industrially relevant protein. In principle, the methodology described in chapter 4 can be applied to any protein refolded using any gel filtration material, providing that the proper mass balances and activity measurements are available.

The last chapter of this thesis, chapter 6, focuses on the economic aspect of the IBs and protein refolding operations. In this chapter, four (4) distinct process options, all capable of performing IBs solubilization and protein refolding, are compared on the basis of the direct fixed capital (DFC) and the contribution to the production cost per gram of protein product ( $P_C$ ). The four alternatives considered include two strategies for IBs solubilization and four alternatives to achieve protein refolding. Overall, the data show that based on technical and economical criteria SMBSEC refolding reactors are a viable alternative to batch dilution refolding.

## References

- [1] F.R. Schmidt, *Appl. Microbiol. Biotechnol.* 65 (2004) 363.
- [2] A.L. Demain, P. Vaishnav, *Biotechnology Advances* 27 (2009) 297.
- [3] J.H. Choi, K.C. Keum, S.Y. Lee, *Chem. Eng. Sci.* 61 (2006) 876.
- [4] A. Jungbauer, W. Kaar, *J. Biotechnol.* (2007) 587.
- [5] A.P.J. Middelberg, *Trends Biotechnol.* 20 (2002) 437.
- [6] B. Batas, J.B. Chaudhuri, *Biotechnol. Bioeng.* 50 (1996) 16.

# Samenvatting

---

Met de komst van de recombinante DNA technologie in de afgelopen eeuw is het mogelijk geworden om bijna elk gen tot expressie te brengen in elk willekeurig expressiesysteem (bijv. prokaryote-, gist-, of zoogdier cellen). Op dit moment zijn ongeveer dertig biofarmaceutische producten gebaseerd op deze recombinante DNA technologie reeds opgenomen in therapieën, met een gezamenlijke marktwaarde van tussen de vijftig en zestig miljard US dollar. Volgens schattingen zijn er wereldwijd nog zo'n driehonderd in ontwikkeling [1].

Bacteriële expressiesystemen zijn populair voor de productie van eiwitten die geen post-translationele modificaties vereisen. En van alle bacteriële expressiesystemen gaat de voorkeur vaak uit naar *Escherichia coli* vanwege (a) de hoge capaciteit om recombinante eiwitten te produceren (tot wel 80% van het eigen drooggewicht [2]), (b) de hoge fermentatie efficiënties die behaald kunnen worden door het gebruik van hoge dichtheidsculturen [3], en tot slot (c) het gemak waarmee modificaties aan het genoom kunnen worden uitgevoerd. Dit laatste omdat de moleculair genetische technieken voor *E.coli* ver ontwikkeld zijn.

Een nadeel echter van genexpressie in *E.coli* is de vorming van zogenaamde *inclusion bodies* (IBs). IBs zijn onoplosbare eiwitaggregaten die vaak worden gevormd als gevolg van het tot overexpressie brengen van een homoloog of heteroloog gen. IBs stapelen zich op in de cel en bevatten het recombinante eiwit in een inactieve en onoplosbare vorm. Om het eiwit actief en oplosbaar te krijgen, moeten de IBs worden opgelost en vervolgens moet het gedenateerde en gereduceerde eiwit hervouwen worden. Deze twee stappen, het oplossen van de IBs en het hervouwen van het eiwit, vormen samen een grote technische uitdaging. Dit proefschrift behandelt deze twee onderwerpen, omdat verbeteringen in elk van de twee ongetwijfeld een positieve invloed zullen hebben op de huidige strategieën die ontwikkeld worden voor de productie van recombinante eiwitten in *E. coli*.

Het eerste onderwerp dat besproken wordt in dit proefschrift is het oplossen van de IBs. Dit onderwerp is het hoofdthema van hoofdstuk 2. Hoofdstuk 2 presenteert een studie waarin kwantitatieve criteria, namelijk de fractie oplosbaar monomeer en de totale fractie opgelost eiwit, gecombineerd worden via een 'Ontwerp van Experiment' methodologie waarmee de samenstelling van de meest optimale buffer voor het oplossen van de IBs bepaald wordt. De



studie laat zien dat het efficiënt oplossen van de IBs sterk afhangt van de concentratie van de denaturant, het reducerend reagens en de pH van de oplossing. Bovendien laat de studie zien dat er slechts relatief lage concentraties denaturant en reducerend reagens nodig zijn om de IBs efficiënt op te lossen.

Na het vaststellen van de optimale condities voor het oplossen van de IBs, rijst de volgende vraag: hoe wordt het gereduceerde en gedatureerde eiwit hervouwen? Eiwithervouwing vindt plaats door middel van een bufferwisseling. In deze stap worden de concentraties van de denaturant en het reducerende reagens verlaagd, waardoor eiwitvouwing en disulfide bond plaats kunnen vinden. Over het algemeen wordt deze bufferwisseling bereikt door middel van verdunning. Hervouwing via een verdunningsstap is een relatief simpel proces en dit is waarschijnlijk de reden waarom het veel wordt toegepast op industriële schaal. Een nadeel is echter de hoge verdunningsfactor die nodig is om een efficiënte hervouwing ( $Y_N$ ) te bewerkstelligen. Als gevolg van deze hoge verdunning zijn grote tanks en veel opeenvolgende concentratiestappen nodig, waardoor de productiekosten omhoog gaan en dit stelt een limiet aan de productieschaal. Om deze reden is het belangrijk dat er naast hervouwing door middel van verdunning, ook andere procesalternatieven onderzocht worden [4-5]. In dit proefschrift wordt de toepassing van vloeistofchromatografie in dit perspectief besproken.

Vergeleken met het hervouwings proces via verdunning, biedt vloeistofchromatografie de mogelijkheid tot intensivering van het proces door het koppelen van de eiwithervouwing en -scheiding in één unit operatie. De bufferwisseling vindt plaats als gevolg van de scheiding die optreedt tussen de denaturant en het reducerende reagens en het gedatureerde en gereduceerde eiwit. Deze scheiding kan bewerkstelligd worden wanneer het eiwit bindt aan het kolom materiaal (adsorptie chromatografie) of wanneer het eiwit vrij door de kolom kan bewegen (gelfiltratie of size-exclusion chromatografie). Deze twee chromatografie technieken worden besproken in dit proefschrift. Hoofdstuk 3 beschrijft de toepassing van ion-exchange (IEx, adsorptie) chromatografie voor het hervouwen van eiwitten en de hoofdstukken 4 en 5 richten zich op size-exclusion chromatografie (SEC) voor dezelfde toepassing.

De toepassing van IEx chromatografie op het hervouwen van eiwitten is veelbelovend vanwege twee redenen: (1) de concentratie van de denaturant en het reducerende reagens kunnen op een gecontroleerde manier verlaagd worden door gebruik te maken van een gradiënt, en (2) de ruimtelijke afzondering van eiwitmoleculen die binden aan het kolom materiaal kan bijdragen aan het voorkomen van eiwitaggregatie, wat vervolgens de

efficiëntie van het hervouwingsproces vergroot. Hoofdstuk 3 beschrijft het effect van verschillende procesvariabelen op de prestatie van de ion-exchange chromatografische hervouwing. De procesvariabelen bevatten: de hoeveelheid geladen eiwit, de staat van het geladen eiwit, het type kolommateriaal, en de helling van de urea gradiënt. De prestatie van de ion-exchange chromatografische hervouwing is bepaald op basis van de fractionele massaopbrengst ( $f_{Prot,Rec}$ ) en de efficiëntie van het hervouwingsproces. Daarnaast is gekeken naar mogelijkheid om de fractionele oppervlaktebedekking ( $\theta$ ) en de helling van de urea gradiënt ( $\xi$ ) te gebruiken om een numerieke beschrijving op te stellen voor de ruimtelijke afzondering van eiwitmoleculen die binden aan het kolommateriaal en de gradiënt van de urea. De studie laat zien: (1) dat de hoeveelheid geladen eiwit een direct effect heeft op de  $f_{Prot,Rec}$  en dat de mate van het effect afhangt van de staat van het geladen eiwit en van het kolommateriaal; (2) dat een verzadiging onafhankelijk van het type kolommateriaal sneller optreedt wanneer het geladen eiwit gedenateerd is dan wanneer het in een natieve vorm geladen wordt (dit kan verklaard worden door een verschil in beschikbaar oppervlak waarmee het eiwit kan binden aan het kolommateriaal); (3) dat er een duidelijke correlatie is tussen de fractionele oppervlaktebedekking ( $\theta$ ) en  $f_{Prot,Rec}$ , waardoor  $\theta$  gebruikt zou kunnen worden om de ruimtelijke afzondering van de eiwitmoleculen te beschrijven, en (4) dat de urea gradiënt direct gekoppeld is aan de efficiëntie van het hervouwingsproces en daarom kan de helling van de urea gradiënt ( $\xi$ ) gebruikt worden om hier een numerieke beschrijving aan te geven. Tot slot suggereert het werk in hoofdstuk 3 dat het belangrijk is om een beter begrip te krijgen van de physicochemische eigenschappen van het ion exchange kolommateriaal.

In tegenstelling tot IEx chromatografie is het proces van SEC voor de hervouwing van eiwitten relatief simpel, vooral omdat de interacties tussen de eiwitmoleculen en het kolommateriaal verwaarloosd kunnen worden. Daarnaast is SEC een aantrekkelijk alternatief voor eiwithervouwing omdat het heeft bewezen effectief te zijn voor geladen eiwitconcentraties tot wel  $80 \text{ mgml}^{-1}$  [6]. Hoofdstuk 5 presenteert de resultaten van een studie waarin eiwithervouwing bewerkstelligd is door gebruik te maken van SEC in een batch en een continue opstelling. Continue SEC is uitgevoerd met behulp van Simulated Moving Bed (SMB) technologie. Daarnaast beschrijft hoofdstuk 5 de ontwikkeling en toepassing van een model waarmee de SMBSEC hervouwingsdata geanalyseerd kunnen worden. De studie laat zien dat: (1) een hervouwingsefficiëntie van 50% bereikt kan worden door de pH en de concentratie van het gedenateerde en gereduceerde geladen eiwit ( $C_{f,D\&R}$ ) te variëren, waarbij het effect van de pH sterk afhangt van de  $C_{f,D\&R}$ ; dat (2) het model de SMBSEC hervouwings reactor goed kan beschrijven, gebaseerd op het effect dat de  $C_{f,D\&R}$  heeft op de reactiesnelheden, en dat (3) een verschuiving van batch naar continue de volumetrische

productiecapaciteit ongeveer 53 maal vergroot, het gebruik van oplosmiddelen ongeveer 10 maal verlaagt, en de eindconcentratie van het product 4.5 maal verhoogt.

SEC kan naast het gebruik als hervouwingsreactor, ook worden ingezet als analytische techniek om eiwithervouwing te bestuderen. Het SEC chromatogram bevat namelijk belangrijke informatie over de verdeling monomeren en aggregaat na de hervouwingsreactie. Wiskundige modellen kunnen hierbij helpen. Deze modellen houden rekening met het transport en transfer van massa, waardoor ze de bijdrage van het reactiemechanisme aan de vorm van het chromatogram kunnen isoleren. De ontwikkeling van een dergelijk model is het hoofdthema van hoofdstuk 4. Hoofdstuk 4 presenteert: (1) een systematische analyse van verschillende reactiemechanismen voor hervouwing door gebruik te maken van zogenaamde Damköhler maps. Deze reactiemechanismen voor hervouwing bevatten ook de competitie tussen hervouwing en aggregatie, waarin aggregatie niet alleen gemodelleerd wordt met het conventionele hogere orde model, maar ook met het mechanistische, meer adequate polymerizatie formalisme. (2) Een vergelijking gebaseerd op mechanistische modellen tussen de SEC hervouwingsreactor en de conventionele batch verdunning hervouwingsreactor. En (3) de toepassing van de modeleringsstrategie op de SEC hervouwingsdata van een industrieel relevant eiwit. De methode die beschreven wordt in hoofdstuk 4 kan in principe worden toegepast op elk te hervouwen eiwit, onafhankelijk van het gebruikte gel filtratie kolom materiaal, mits de juiste massabalansen en de activiteits assays beschikbaar zijn.

Het laatste hoofdstuk van dit proefschrift, hoofdstuk 6, richt zich op de economische aspecten van de IBs en de eiwithervouwingsoperaties. In dit hoofdstuk worden vier onafhankelijke processen met elkaar vergeleken op basis van direct vast kapitaal (DFC) en de bijdrage aan de productiekosten per gram of eiwitproduct (Pc). Al deze processen omvatten het oplossen van IBs en het hervouwen van het eiwit. De vier processen bevatten twee verschillende methoden voor het oplossen van de IBs en vier alternatieve processen om hervouwing te bewerkstelligen. Over het geheel genomen laat de studie zien dat op basis van de technische en economische criteria de SMBSEC hervouwingsreactie een goed alternatief is voor de conventionele batch verdunningshervouwing.

# Table of contents

---

Summary/Samenvatting	4
<b>Chapter 1:</b> Introduction	13
<b>Chapter 2:</b> Efficient Solubilization of Inclusion Bodies	26
<b>Chapter 3:</b> Ion-Exchange Chromatographic Protein Refolding	43
<b>Chapter 4:</b> Size-Exclusion Chromatographic Protein Refolding: Fundamentals, Modeling and Operation	68
<b>Chapter 5:</b> Size-Exclusion Simulated Moving Bed Chromatographic Protein Refolding	108
<b>Chapter 6:</b> Techno-Economic Evaluation of an Inclusion Body Solubilization and Recombinant Protein Refolding Process	140
<b>Chapter 7:</b> Outlook	176
Curriculum Vitae	182
Publications and Conferences	183
Acknowledgements/Dankwoord	184

# 1

## Introduction

---

## 1.1 Introduction

DNA recombinant technology, is the technology that has permitted the expression of any given gene sequence in any given expression system (e.g., prokaryotes, yeast, insect cells, mammalian cells). This technology significantly contributed to the industrial production of recombinant biopharmaceuticals. When it comes to the production of recombinant proteins that do not require post-translational modifications, the prokaryotic organism *Escherichia coli* (*E. coli*) is the preferred organism, most likely owing to the following reasons: (a) its ability to grow rapidly and at a high density on inexpensive substrates; (b) its well characterized genetics, which allows the modification of its genome with ease; (c) its high capacity to accumulate the recombinant protein (up to 80% of its dry weight); and (d) the availability of an increasingly large number of cloning vectors and mutant host strains [1-2]. *E. coli* has, however, some limitations which include: (a) it is not suitable for the production of many large, complex proteins containing multiple disulfide bonds, and (b) the over-expression of the gene sequence of interest may result in the formation of inclusion bodies (IBs). Inclusion body formation happens mostly as a consequence of the over expression of heterologous gene products, but it is not restricted to it. The propensity of forming IBs depends on several factors related to the particular protein being expressed, including primarily: cysteine content, size, hydrophobicity, secondary structure content [3]. Among these, the only one that correlates well with the tendency to forming IBs is the content of cysteines or the ability to form disulfide bridges. The reason for this good correlation has been attributed to the reducing environment of the cytoplasm. Such environment prevents the formation of disulfide bonds *in vivo*, by keeping the cysteines in a reduced state. Disulfide formation *in vivo*, and under normal expression levels, takes place in the periplasm and the endoplasmic reticulum. Upon increasing the expression level of the recombinant protein, the folding machinery (chaperons, foldases) gets saturated. Accordingly, the polypeptide chains can interact via their reduced cysteines forming inter-chain disulfide bonds, leading to the formation of inclusion bodies.

Inclusion bodies are dense electron-refractile particles that can be found in either the cytoplasmic or peri-plasmic spaces of the *E. coli* cell, and they contain the recombinant protein product in an inactive, insoluble form. They have a spherical shape with a diameter that ranges from 0.5-1.3  $\mu\text{m}$  and a density of around 1.3  $\text{mg ml}^{-1}$  [4]. Fig. 1.1 shows a scanning electron micrograph of *E. coli* IBs; notice the spherical shape (left panel) and their porous surface (right panel). Having a protein expressed in the form of inclusion bodies offers several advantages, including: (a) protein expression at relatively high levels; (b) after cell disruption, the IBs can be readily separated from the cell debris owing to their relatively high density; (c) a lower or negligible degradation of the protein product by proteases; (d) protection of the host when expression a toxic product; and (e) the recombinant protein is present at a relatively high purity (>70% of the total protein). The main challenge associated with the expression of the protein product in IBs is the recovery of the

product in an active (native), soluble form. To do this the IBs must first be solubilized and thereafter the soluble, often denatured and reduced (D&R), protein product must be refolded (renatured). Fig. 1.2 presents a block flow diagram (BFD) depicting the major processing steps involved in the production process of a recombinant product in *E. coli*, and indicating the approximate location of the inclusion bodies solubilization and subsequent protein refolding steps. For most production processes utilizing IBs, the solubilization and refolding steps set the limit for the overall downstream processing yield ( $Y_{ov,DSP}$ ), and provide the highest contribution to the product production cost. Therefore, any improvement in either the solubilization or the refolding steps would have a positive impact on the overall production process. And this is one of the major aims of this thesis.

This thesis is divided in two major parts. The first focuses on the study of the inclusion bodies solubilization operation. The second studied the application of liquid chromatography, primarily ion-exchange and size-exclusion chromatography, for the refolding of the soluble D&R protein.

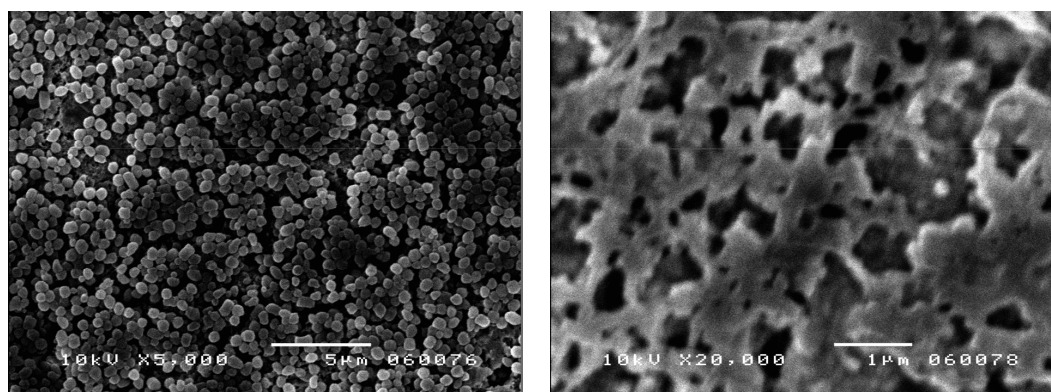


Figure 1.1: Scanning electron micrographs (SEM) of *E. coli* inclusion bodies magnified 5000x and 20000x. Diameter of the particles is around 0.5-0.8 $\mu$ m

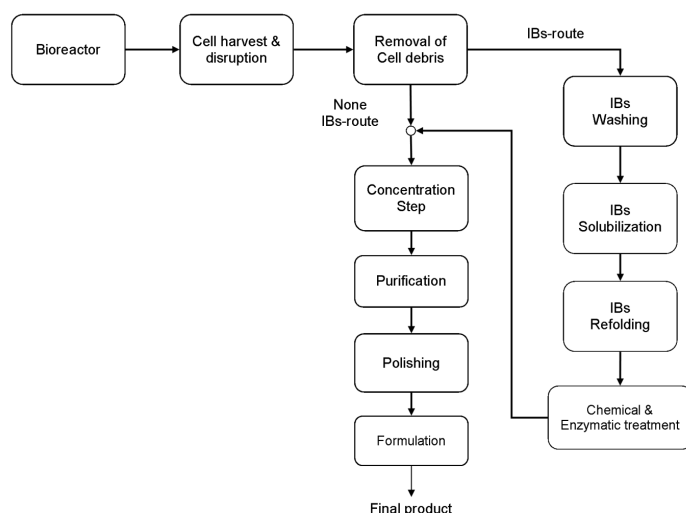


Figure 1.2: Block-flow diagram illustrating the process of recovery of a recombinant product produced as an inclusion body

## 1.2 Inclusion bodies solubilization

Inclusion bodies are, in essence, insoluble protein aggregates enriched in the target product. These aggregates are held together mainly by hydrophobic interactions and inter-molecular disulphide bonds. Though disulfide bonds are hardly formed in the reducing environment of the cytosol *in vivo*, random formation of these bonds either within the same protein (scrambled disulfides) or between different polypeptide chains, may occur by air oxidation during cell lysis and inclusion body isolation [3].

Generally speaking, to solubilize the IBs the solution used should have the proper chemical composition to disrupt covalent and non-covalent interactions. Ideally, an efficient IBs solubilization step should result in a solution containing a mono dispersed distribution of the soluble protein, enriched in denatured and reduced monomers. Attaining such chemical composition requires, typically, chaotrope agents (e.g., urea, guanidine hydrochloride) at a concentration in the range of 6 to 8 M [5]. Furthermore, if the protein expressed possesses cysteines, a reducing agent should be also present in the solution, because disulfide bonds will not be disrupted only by high denaturant concentrations [6]. Typical reducing agents include dithiothreitol (DTT), dithioerythritol (DTE), or 2-mercaptoethanol [7]. These agents are used at concentrations ranging from 5 up to a 100 mM [6]. Table 1.1 presents a series of examples of solubilization conditions used in refolding studies.

Because most of the studies involving IBs solubilization seem to set the concentrations of the solubilization additives (chaotrope and reducing agent) arbitrarily to high levels, raised the following questions: (a) what are appropriate performance indicators to assess the efficiency of the solubilization step? and (b) how are the performance indicators affected by the operational



variables of the step? These and other relevant questions about IBs solubilization are addressed in chapter 2.

Table 1.1: Examples of solubilization conditions

Protein	Classification	Solubilization conditions	C <sub>p</sub> (mg/ml)	Reference
Human IL-15	Inclusion bodies	8MUrea/1mMDTT/10%glycerol/20mMimidazole/ 20mM sodium phosphate/ pH 7.4	n.r.	[9]
Prochymosin	Inclusion bodies	8MUrea/ 50mM KH <sub>2</sub> PO <sub>4</sub> /50mM NaCl/ 1mM EDTA/ pH 11	2	[10]
Lysozyme	Commercial	8MUrea/150mM DTT/100mMTris-HCL/ pH 8.6	10-80	[11]
Carbonic Anhydrase B	Commercial	5MGnd-HCL/ 50mM Tris-HCL/ pH 7.5	2-63	[12]
<sup>a</sup> )PDGF -A/B	Inclusion bodies	6MGnd-HCL/ 100mM DTT/ 100mM Tris-HCL/ 1mM EDTA/ pH 8.5	1.2	[13]
<sup>b</sup> )U-PA	Inclusion bodies	6MGnd-HCL/ 5mMDTT / 50mMTris-HCL/ 5mMEDTA / pH 7.5	n.r.	[14]
$\alpha$ -lactalbumin	Commercial	8MUrea / 200mM monoethioglycerol	8.5	[15]
Bovine Serum Albumin	Commercial	8MUrea/ 100mMDTT / 50mMTris-HCL/ 3mM EDTA /pH 8.5	20-100	[16]
<sup>c</sup> )Fe-SOD	Inclusion bodies	10MUrea/ 0.05M sodium phosphate/ pH 7.4	20	[17]
<sup>d</sup> )r-hGH	Inclusion bodies	2M Urea/ 100mM Tris-HCL/ pH 12.5	6	[17]

<sup>a</sup>) Platelet-derived growth factor AB

<sup>b</sup>) Urokinase plasminogen activator fragment

<sup>c</sup>) Iron Superoxide dismutase

<sup>d</sup>) Recombinant Human Growth Hormone

n.r. : Not reported

### 1.3 Protein refolding using liquid chromatography

#### 1.3.1 Ion-exchange chromatographic protein refolding

Protein refolding by means of ion-exchange chromatography (IExC) was most likely introduced by Creighton. In his work the non-covalent adsorption of denatured as well as denatured and reduced proteins to ion-exchange media was demonstrated using five model proteins, including: Cytochrome C (MW 11.8 kDa; pI 10.2), Ovalbumin (MW 45kDa, pI 4.7),  $\alpha$ -Lactalbumin (MW 14.17 kDa; pI 4.2), Lysozyme (MW 14.4kDa, pI 11) and Aprotinin (MW 6.5kDa, pI 10.5) [17]. Additionally, his work also showed that while bound, the protein could refold and could thereafter be eluted using a salt gradient. The work of Creighton made the advantages of the application of IExC to protein refolding evident, and these are: (1) the concentration of the solubilization additives can be changed, in a controllable fashion using gradients, while the protein remains bound to the adsorbent, and (2) the protein can be purified and even concentrated by applying a salt gradient. Additionally, more recently it has been suggested that the spatial isolation of the protein molecules, provided upon binding to the adsorbent's surface, may reduce protein aggregation[18].

Alterations to the method presented by Creighton have been introduced through the years, including a concomitant urea, pH and salt gradient [15,19-20]. Compared to the single denaturant gradient, these modifications resulted in improvements for the ion-exchange refolding of iron-superoxide dismutase and lysozyme [15,19-20]. However, the application of pH gradients to ion-exchange is limited to those cases where the change on pH will not induce unwanted elution or protein precipitation. Supported by current data (table 1.2) it is safe to say that the successful application of IEx for protein refolding has been demonstrated. However, mechanistic understanding of the process remains evasive and thus this opens the possibility to formulate a significant number of interesting questions. These include: (1) what role does the nature of the adsorbent's backbone play? (2) how can spatial isolation be quantitatively described? (3) what is the contribution of the controllable change in chemical composition to the performance of the ion-exchange refolding reactor?. These and other questions are addressed in chapter 3.

Table 1.2: Examples of ion-exchange refolding

Protein	Resin	Bonds #	RSH [mM]	RSSR [mM]	pH	C <sub>Urea0</sub> [M]	C <sub>UreaE</sub> [M]	<sup>a)</sup> ξ	<sup>b)</sup> Load	<sup>c)</sup> Rec (%)	Ref
Cytochrome C	CM Cellulose	1	0	0	8.7	8	0	-0.10	3.7	100	
Hen Ovalbumin	DEAE-Cellulose	1	0	0	8.7	8	0	-0.11	3.7	50	[18]
BPTI	CM Cellulose	3	0.1	1	8.7	8	0	-0.05	3.7	90	
α-Latalbumin	DEAE-Cellulose	4	0	0	8.7	8	0	-0.11	3.7	10	
SOD - <i>E.Coli</i>	Q-Sepharose	n.r	0	0	7.4	6	1	-0.20	0.8	n.r	[16]
Lysozyme	SP-Sepharose	4	3	0.3	8.7	6	1	-0.4	1.6	n.r	[20]
Lysozyme	SP-Sepharose	4	3	0.3	8	6	1	-0.21	1.14	n.r	[21]
BSA	Q-Sepharose	17	2.2	1.1	8.5	8	0.079	∞	2-20	67-25	[15]
TGase	SP-Sepharose	n.r	0	0	6	6	0.5	-0.15	0.4	49.7	[22]
rhG-CSF	Q-Sepharose.FF	2	0	0	8.0	3	1.2	-0.08	n.r	43	[23]
LK-68	Q-Sepharose	9	0	0	9.0	8	0	-0.22	2.5	n.r	[24]
NS3	SP-Sepharose	n.r	10	0	4	8	1	-0.08	0.059	n.r	[25]
	DEAE	n.r	10	0	9	8	1	-0.09	0.059	n.r	

n.r: Not reported.

∞: Step change.

<sup>a)</sup> ξ: Urea slope (molUreal<sup>-1</sup>min<sup>-1</sup>).

<sup>b)</sup> Load: milligrams of denatured protein per milliliter of packed bed

<sup>c)</sup> Recovery: Mass recovered (protein eluted / protein injected).

#### 1.4 Size-exclusion chromatographic protein refolding

Size-exclusion chromatographic (SEC) refolding was probably first introduced in the work of Werner et al [25]. SEC refolding is basically a buffer exchange step that works as follows: a column equilibrated with renaturation buffer (buffer free of denaturant and reducing agents) is fed with a pulse containing the denatured and reduced protein solution. As the feed pulse migrates down the column the D&R protein separates from the solubilization additives (denaturant and the reducing agent). This separation occurs as a result of the large difference in partition coefficient between the D&R protein (large molecules) and the solubilization additives (small molecules). In essence, the large molecules migrate faster, because they travel through the inter-particle void. In contrast, since the small molecules can penetrate a higher fraction of the pore volume, they face a relatively larger traveling distance. Accordingly, they are delayed and exit the column later than the large molecules. This separation translates into a decrease of the local concentration of the solubilization additives around the protein, inducing folding and disulfide bond formation.

Table 1.3 presents a list of proteins that have been successfully refolded by SEC, as well as the operational conditions used. Most likely, one of most attractive features of SEC refolding is its ability to operate using a concentrated feed of D&R protein. This was shown in the work of Batas et al [10], where denatured and reduced lysozyme was fed at concentrations up to  $80\text{mgml}^{-1}$ . Compared to batch dilution refolding SEC refolding has several advantages, including: (a) operating at relatively high concentrations of D&R protein, (b) relatively high degree of process intensification, as refolding and separation can be done in one unit operation. Compared to IEx refolding, SEC refolding has several advantages including (a) the non-specific interactions with the gel-matrix are negligible, (b) operating at relatively high D&R feed concentrations, (3) the feed to the SEC refolding reactor may contain urea or guanidine hydrochloride, depending on the solubilization requirements.

SEC has one main limitation, which strictly related to the mode of operation. When operated in batch mode, SEC requires low loading volumes and relatively long columns to keep the resolution. These requirements lead to a non-optimal utilization of the column material, high consumption of eluent and long processing times. These limitations, however, can be overcome by changing the mode of operation from batch to continuous; which can be done with the aid of Simulated Moving Bed (SMB) technology. The topic of SEC refolding is discussed in chapters 4 and 5 of this thesis.

Table 1.3: Examples of size-exclusion protein refolding

Protein name	Gel type	$V_{inj}$ (%BV)	<sup>b)</sup> D&R (mg/ml)	$\dot{Q}$ (ml/min)	Ref
rETS-1	Superdex 75	4.2-8.5	1 - 10	0.2	[26]
RNase A	Sephacryl S100	0.18-0.37	1 - 10	0.5	
Lysozyme	Sephacryl S100	0.32	9.6	2.5	
Lysozyme	mixed gel	0.67	n.r.	2.5	[11]
Carbonic anhydrase	Sephacryl S100	0.67	11	1.1	
Lysozyme	Sephacryl S100	4.24	10	0.3	[27]
<sup>a)</sup> u-PA	Sephacryl S300	0.20-3.31	2 - 16	0.5 - 4.5	[28]
Lysozyme	Sephacryl S100	0.32	40	2.5	[29]
	Superdex 75	0.70	40	2	
u-PA	Sephacryl S300	0.42	8	0.5	[13]
Bovine Carbonic Anhydrase B	Superdex 75	4.24	5.2	0.4	[30]
Lysozyme	Superdex 75	0.08	5	1	[31]
Lysozyme	Sephacryl S100	0.25	11.8	0.8	[32]
Beta-Lactamase	Sephacryl S300	0.47-0.94	8	0.83-4	[33]
Lysozyme	Sephacryl S200	0.62	8.3	2	[34]
alpha-lactalbumin	Superdex 75 PrepGrade	1.50	10	1.01	[35]

<sup>a)</sup>Urokinase Plasminogen Activator

<sup>b)</sup>D&R: Denatured and Reduced protein

## 1.5 Thesis Outline

The work described in this thesis uses liquid chromatography for the refolding of a crude, industrially relevant protein expressed in inclusion bodies. The model protein, a fusion protein of 127 amino acids and 6 cysteines, is a rather sticky protein prone to aggregate.

The topic of inclusion bodies solubilization is presented in chapter 2. The approach followed, made use of a statistical design of experiments (DOE) methodology to optimize the concentrations of denaturant, reducing agent, and solution's pH. The efficiency of the solubilization was judged on the basis of the total protein solubility and the mass fraction of soluble monomer. The study showed, that by simultaneously optimizing the three selected parameters, a high protein solubility and a high fraction of soluble monomer are attainable, at relatively low concentrations of denaturant and reducing agent.

Once the conditions for solubilization were defined, the research moved on to study different chromatographic modes that could be potentially used as refolding reactors. These modes included: (1) hydrophobic interaction chromatography (HIC), quickly ruled out due to unfeasible binding conditions, (2) ion-exchange chromatography (IEX), and (3) size-exclusion chromatography (SEC). The latter two techniques were more suitable and therefore were further studied.

Chapter 3 presents the results on ion-exchange refolding. The scientific challenges in this work included low protein mass recoveries, and relatively low refolding yields. The questions posed include: (1) What are the process parameters that have the major influence on the unit's performance? (2) How do the loading conditions affect the protein mass recovery? (3) What is the influence of the nature of the adsorbent's backbone on the protein mass recovery? And, (4) how does fractional surface coverage relate to protein recovery and the spatial isolation?

After the study on ion-exchange refolding, size-exclusion chromatographic refolding became the focus of attention. This work evaluated two potential applications of SEC. The first, discussed in chapter 4, studied the possibility of using SEC as an analytical tool to study protein refolding. The work in chapter 4 presents the development of a modeling tool that can be used to extract valuable information from the SEC refolding chromatogram. The second application evaluated was the use of SEC as a refolding reactor. This application is discussed in chapter 5. Chapter 5 presents a study that includes the optimization of the operational conditions of the SEC refolding reactor using a DOE methodology. Additionally, this chapter

discusses the effect of changing the mode of operation of the SEC refolding reactor from batch, to continuous with the aid of SMB technology.

Finally, chapter 6 presents an economic analysis that evaluates the economic performance of four different process options, all suitable to perform IBs solubilization and protein refolding. These process options include the pH induced IBs solubilization concomitantly with protein refolding, chaotrope induced IBs solubilization followed by batch SEC refolding, chaotrope induced IBs solubilization followed by SMBSEC refolding, and chaotrope induced IBs solubilization followed by batch dilution refolding.



## 1.6 References

- [1] F. Baneyx, *Curr. Opin. Biotechnol.* 10 (1999) 411.
- [2] A.L. Demain, P. Vaishnav, *Biotechnology Advances* 27 (2009) 297.
- [3] J. Buchner, T. Kiefhaber, in *Protein Folding Handbook*, WILEY-VCH, Weinheim, 2005, p. 1251.
- [4] S. Singh, A. Panda, *J. Biosci. Bioeng.* 99 (2005) 303.
- [5] E.D.B. Clark, *Curr. Opin. Biotechnol.* 12 (2001) 202.
- [6] L.D. Cabrita, S.P. Bottomley, *Biotechnology annual review* 10 (2004) 31.
- [7] E.D.B. Clark, *Curr. Opin. Biotechnol.* 9 (1998) 157.
- [8] M. Matsumoto, S. Misawa, K. Tsumoto, I. Kumagai, H. Hayashi, Y. Kobayashi, *Protein. Expres. Purif.* 31 (2003) 64.
- [9] C. Wei, B. Tang, Y. Zhang, K. Yang, *Biochem. J.* 340 (1999) 345.
- [10] B. Batas, J.B. Chaudhuri, *Biotechnol. Bioeng.* 50 (1996) 16.
- [11] C. Muller, U. Rinas, *J. Chromatogr. A.* 855 (1999) 203.
- [12] E.M. Fahey, J.B. Chaudhuri, P. Binding, *J. Chromatogr B.* 737 (2000) 225.
- [13] C. Machold, R. Schlegl, W. Buchinger, A. Jungbauer, *J. Biotechnol.* 117 (2005) 83.
- [14] M. Langenhof, S.S.J. Leong, L.K. Pattenden, A.P.J. Middelberg, *J. Chromatogr. A.* 1069 (2005) 195.
- [15] M. Li, Z.-G. Su, *Biotechnol. Lett.* 24 (2002) 919.
- [16] A.K. Patra, R. Mukhopadhyay, R. Mukhija, A. Krishnan, L.C. Garg, A.K. Panda, *Protein. Expres. Purif.* 18 (2000) 182.
- [17] T.E. Creighton, *UCLA Symp. Mol. Cell. Bio., New Ser.* 39 (1986) 249.
- [18] A.P.J. Middelberg, *Trends Biotechnol.* 20 (2002) 437.
- [19] M. Li, G. Zhang, Z. Su, *J. Chromatogr. A.* 959 (2002) 113.
- [20] M. Li, Z. Su, *Chromatographia* 56 (2002) 33.
- [21] X.-Q. Liu, X.-Q. Yang, F.-H. Xie, L.-Y. Song, G.-Q. Zhang, S.-J. Qian, *Protein. Expres. Purif.* 51 (2007) 179.
- [22] C.Z. Wang, J.F. Liu, X.D. Geng, *Chin. Chem. Lett.* 16 (2005) 389.
- [23] W.C. Choi, M.Y. Kim, C.W. Suh, E.K. Lee, *Process Biochem.* 40 (2005) 1967.
- [24] M. Li, A. Poliakov, U.H. Danielson, Z. Su, J.-C. Janson, *Biotechnol. Lett.* 25 (2003) 1729.
- [25] M.H. Werner, G.M. Clore, A.M. Gronenborn, A. Kondoh, R.J. Fisher, *FEBS Lett.* 345 (1994) 125.
- [26] B. Batas, C. Schiraldi, J.B. Chaudhuri, *J. Biotechnol.* 68 (1999) 149.
- [27] E.M. Fahey, J.B. Chaudhuri, P. Binding, *Sep. Sci. Technol.* 35 (2000) 1743
- [28] B. Batas, J. Chaudhuri, *Bioprocess Biosyst. Eng.* 24 (2001) 255.
- [29] Z. Gu, X. Zhu, S. Ni, H. Zhou, Z. Su, *J. Biochem. Biophys. Methods* 56 (2003) 165.
- [30] S.S.S. Wang, C.-K. Chang, H.-S. Liu, *Biochem. Eng. J.* 29 (2006) 2.
- [31] B.-J. Park, C.-H. Lee, S. Mun, Y.-M. Koo, *Process Biochem.* 41 (2006) 1072.
- [32] S.R. Harrowing, J.B. Chaudhuri, *J. Biochem. Biophys. Methods* 56 (2003) 177.
- [33] H. Lanckriet, A.P.J. Middelberg, *J. Chromatogr. A.* 1022 (2004) 103.
- [34] R. Schlegl, G. Iberer, C. Machold, R. Necina, A. Jungbauer, *J. Chromatogr. A.* 1009 (2003) 119.

# 2

## Efficient solubilization of Inclusion Bodies

---

### Abstract

The over-expression of recombinant proteins in *Escherichia coli* leads in most cases to their accumulation in the form of insoluble aggregates referred to as inclusion bodies. In order to obtain active product, the inclusion bodies must be solubilized and thereafter the soluble monomeric protein needs to be refolded. In this work we studied the solubilization behavior of a model-protein expressed as IBs at high protein concentrations, using a statistically designed experiment; in order to determine which of the process parameters, or their interaction, have the greatest impact on the amount of soluble protein and the fraction of soluble monomer. The experimental methodology employed pointed out an optimum balance between maximum protein solubility and minimum fraction of soluble aggregates. The optimized conditions solubilized the IBs without the formation of insoluble aggregates; moreover the fraction of soluble monomer was roughly 75% while the fraction of soluble aggregates was approximately 5%. Overall this approach guarantees a better use of the solubilization reagents, which brings an economical and technical benefit, at both large and lab scale and may be broadly applicable for the production of recombinant proteins.

Key words: Inclusion bodies; Solubilization; Experimental design; urea; reducing agents

## 2.1 Introduction

The introduction of recombinant DNA technology made possible the transfer and efficient expression of many pools of desired genes in a cell. Hence unlimited and inexpensive sources of otherwise rare proteins were thought to become accessible. The host selected for the first industrial trials was the bacterium *E.coli* due to available insight in its genetic machinery, short duplication times and a low fermentation cost. However, the high level of expression required to match the productivity targets often results in the accumulation of the target protein as insoluble aggregates known as inclusion bodies (IBs).

The IBs are dense particles of aggregated protein that accumulates either in the cytoplasmic or periplasmic spaces of *E.coli*. Their size varies between 0.5-1.3 $\mu\text{m}$ ; they have either an amorphous or para-crystalline nature depending on their localization and an average density of  $\sim 1.3\text{mg}\cdot\text{ml}^{-1}$  [1]. Regarding their composition, they normally contain very little host proteins, ribosomal components or DNA/RNA fragments [2]; instead they contain the over-expressed recombinant protein in a concentration of up to 70-90% of the total protein[3].

The recovery of the protein from the IBs encompasses four steps: isolation of the IBs from the *E. coli* cells; solubilization of the IBs; renaturation of the soluble protein and purification of the native or bioactive target. Among these steps the solubilization and renaturation have the greatest impact on the overall process performance. An inefficient solubilization leads to the formation of soluble aggregates, which directly affects the performance of the subsequent renaturation step, the overall process yield and the economical performance of the production of recombinant proteins expressed as IBs[1, 4].

Generally in order to achieve high protein solubility and a high monomer fraction, either at large scale or lab scale, the IBs are solubilized using high concentrations (6-8M) of chaotropic agents such as urea and guanidine hydrochloride [5, 6] and also low concentration of detergents, such as SDS [7, 8]. These co-solutes diminish the protein-protein interactions and increase the protein-solvent interactions. In the case of proteins with cysteines, the co-solutes should also prevent the formation of inter-chain disulfides by keeping the cysteine residues in a reduced state, thus leading to higher yields of monomer soluble protein. The reducing agents (RA) most commonly used are dithiothreitol (DTT) and  $\beta$ -mercaptoethanol (5-100mM). To aid the reduction step, alkaline pH is also use to maintain the cysteines in a reduced state especially in highly concentrated protein solutions [9-11]. Finally, in order to obtain the native (bioactive) product, the concentrations of the solubilization agents have to

be decreased in a controllable manner; minimizing then the loss of product due to the competing side reactions[12-14], namely misfolding and aggregation.

As previously mentioned, there are two important issues in the production of recombinant proteins expressed in IBs; solubilization of the IBs and renaturation of the soluble protein. Both steps need careful consideration if one wants to achieve an economical and productive operation. In the last decade many novel refolding methods have been developed [15], while less attention has been paid to the solubilization step which is the focus of this work. Even though there have been studies on how to optimize the solubilization of inclusion bodies [1, 16-21], the approaches taken are either “trial and error” or “one factor at the time experimentation”; the last approach being the most frequently used as presented by Patra et al,[18] and Singh et al [1] . The drawback of last approach is that the information obtained overlooks the combined effect the experimental factors considered. In this work we studied the solubilization behavior of a model-protein expressed as IBs using a statistically designed experiment; in order to determine which of the process parameters, or their interaction, have the greatest impact on the amount of soluble protein and the fraction of soluble monomer. Overall this approach will guarantee a better use of the solubilization reagents, which brings an economical benefit and technical benefit, at both large and lab scale and may be broadly applicable for the production of recombinant proteins.

## 2.2 Materials and Methods

The *E. coli* inclusion bodies containing the fusion protein were kindly provided by Diosynth B.V (The Netherlands). The model protein used in this study is a construct made of an insulin-like polypeptide containing 6 cysteines and a fragment of an *E. coli* homologous protein. The fusion protein has a total length of 117 amino acids and a molecular mass of 12710 Daltons.

The following additives were obtained accordingly: Urea, sodium hydroxide (NaOH), Trometamol, sodium bicarbonate (NaHCO<sub>3</sub>) and sodium chloride (NaCl) were purchased from J.T Baker; ethylenediaminetetraacetic acid (EDTA) was from Calbiochem; Guanidine hydrochloride was obtained from Pierce, dithiothreitol (DTT) was procured from Sigma, β-mercaptoethanol and Hydrochloric acid (HCl) were from Merck. All reagents were of analytical grade or of higher purity.

All chromatographic separations were performed using an ÄKTA explorer 10 from GE Healthcare. The elution profiles were analyzed with UNICORN software version 5.01 from GE Healthcare. Native and non native electrophoresis was done using the X-Sure-Mini cell system from Invitrogen. Quantitative densitometry was done using a GEL-DOC XR densitometer in combination with the software package Quantity-One; both acquired from Biorad.

### 2.2.1 Gel electrophoresis

Electrophoresis of the protein fractions was done using NuPAGE (4-12%) Bis-Tris pre cast gels. The running conditions were 37min and 200V. Reduced electrophoresis was done by incubating the samples with 65mM DTT for 10 min at 85°C. As molecular weight marker, Mark-12 was used and the gels were stained with Simply-Blue safe stain (Invitrogen). The protein concentration of the samples was adjusted to 0.4 mg/ml and the volume loaded per well was 25µl. When the samples were treated with DTT the running buffer contained NuPAGE Antioxidant (Invitrogen), to prevent the formation of disulfide bridges in the gel.

### 2.2.2 Size exclusion chromatography

The fractionation of the samples by size exclusion chromatography was done using a Superdex 75-10/300 (GE-Healthcare). The injection volume was 0.2ml (0.8% CV) and the flow rate was set to 0.5ml/min. The column was calibrated using a low molecular mass calibration kit for gel filtration from GE Healthcare. The standards used were Aprotinin (6.5kDa), Ribonuclease A (13.7kDa), Carbonic Anhydrase (29kDa) and Conalbumin (75kDa). The elution volume was normalized to a column-independent partition coefficient  $K_{av}$ , obtained from the expression:

$$K_{av} = \frac{V_R - V_0}{V_T - V_0} \quad (2.1)$$

Where,  $V_R$  is the retention volume of the unknown compound, and  $V_0$  and  $V_T$  are the void and total liquid volumes, which were determined from pulse experiments with Blue dextran and acetone as 7.61ml and 19.63ml respectively. Thus the column has a particle porosity of 0.83 and a bed porosity of 0.323. The calibration curve of the column was calculated using a linear regression of the plot of  $K_{av}$  vs.  $\log M_w$ ; the calibration line for the column used in this study was: The hydrodynamic radius ( $R_h$ ) of the standards was calculated according to the following expression: and it was used to estimate the radius of the eluting compounds from the plot of  $R_h$  vs.  $1/K_{av}$  [22, 23].

### 2.2.3 Estimation of the total soluble protein

The protein concentration in the samples was estimated using the absorbance at 280nm. The measurements were done using an ND1000 spectrophotometer from NanoDrop. The mass of protein per mass of inclusion body was measured using the Bradford Assay (SIGMA) with bovine serum albumin as standard. The measurements were done taking samples from the supernatant recovered after the centrifugation of the samples at 10000 RPM for 20 min right after the dissolution of the inclusion bodies.

### 2.2.4 Determination of the solids fraction and the protein content per mass of inclusion bodies

The solids fraction present in the inclusion bodies suspension was determined by dry weight measurements. Aliquots of a well-homogenized suspension were dried at 60°C for 8 hrs. The weight of the samples was checked every 2 hrs till a constant weight was reached. The final weight was equal to  $8 \pm 0.1\%$  of the wet weight, which means that per gram of suspension 80mg of inclusion bodies are present.

The analysis of the protein concentration in the inclusion bodies prior to solubilization was carried out by dissolving  $6.59 \pm 0.78$ mg of inclusion bodies (Dry weight basis) in 1ml of 4M Urea/25mM DTT/10mM NaHCO<sub>3</sub>/0.1mM EDTA/20mM Tris-HCL at a pH of 10.5. The clear samples were diluted based on the urea concentration to  $\sim 1$ M urea and assayed by the Bradford-Reagent (Sigma-Aldrich). The mass balance indicated that the inclusion bodies consist of 100% protein.

### 2.2.5 Reversed Phase -HPLC Analysis

The concentration of the native monomer was determined using RPHPLC. The protocol used can be found in [24]. To obtain the native molecule the fractions collected from the size exclusion experiments were digested by adding trypsin in an enzyme to substrate mass ratio of 1:100. The samples were incubated for 16 hours at room temperature. The concentration

of the native precursor was determined from the peak area of human insulin which was used as a standard.

### 2.2.6 Experimental design

A Box-Behnken experimental design was used [25], to study the combined effects of Urea, dithiothreitol (DTT) and pH on the solubilization of inclusion bodies and the formation of soluble monomers. The solid concentration was kept constant at  $3.5 \pm 0.34$  mg/ml. The order of the experiments was fully randomized, which statistically provides protection against the external lurking variables to bias the results. All factors were evaluated at two levels which resulted in 15 experimental runs. The experimental design was analyzed using a statistical software package Statgraphics version XV (StatPoint).

### 2.2.7 Solubilization of the inclusion bodies

The solubilization behavior of the model IBs was studied using aliquots of 100-112mg of protein (in dry weight basis) which was suspended in the different buffer conditions and all the solutions contained as basic composition 10mM NaHCO<sub>3</sub> and 0.1mM EDTA. To evaluate the pH effect on the solubility, the inclusion bodies were incubated for 30 minutes at pH values ranging from 9 to 12. After incubation the samples were centrifuged at 10000 RPM for 10min. The supernatant was transferred to fresh tubes, the amount of soluble protein was then measured by absorbance at 280nm and the percentage of monomer in solution was determined by SDS-PAGE followed by quantitative densitometry.

Parallel studies were carried out to evaluate the combined effect of Urea, DTT and pH on the total protein solubility and the formation of soluble monomer. The study was conducted according to the experimental design described in the previous section. All the solutions contained as basic buffer composition 10mM NaHCO<sub>3</sub> and 0.1mM EDTA, to which the different additives were added. The fraction of soluble monomer was quantified by SDS-PAGE combined with quantitative densitometry, and also by size exclusion chromatography combined with peak integration. Size exclusion was also employed to measure the degree of opening of the monomer, which was related to the its changes on its partition over the column media.

### 2.2.8 Equations used

Calculation of the solubilization yield, refolding yield and overall yield

The solubilization efficiency of the product was defined as:

$$Y_{s-product} = \frac{W_{SP} \cdot x_{product}}{W_{DIBs} \cdot x_{Product}^0} \quad (2.2)$$

while the total protein solubilization was:

$$Y_{s-Protien} = \frac{W_{SP}}{W_{DIBs} \cdot x_{Protien}} \quad (2.3)$$

where WSP refers to the mass of soluble protein after solubilization and centrifugation,  $x_{Protien}$  is the fraction of the product present in the total soluble protein, WDIBS is the mass of dry inclusion bodies added to the system and  $x_{Protien}$  is the product fraction contained in the dry inclusion bodies.

The refolding yield was defined as:

$$Y_R = \frac{W_N}{W_{SP} \cdot x_{product}} \quad (2.4)$$

where  $W_N$  refers to the mass of native product recovered after refolding. The overall yield will be then given by the product of the solubilization and refolding yield.



## 2.3 Results and discussion

### 2.3.1 Combined effect of urea, dithiothreitol and pH over the solubilization of the IBs

It is known that the two major forces responsible for protein aggregation in solutions from inclusion bodies are hydrophobic interactions and for some proteins the formation of inter-chain disulfides, thus there is a clear interrelation between pH, RedOx potential and chaotrope concentration [26-28]. The response variables chosen were: the total soluble protein and the fraction of soluble monomer. The experimental factors were: pH at 9.0 and 12, DTT at 1 and 50mM and Urea at 0 and 4M.

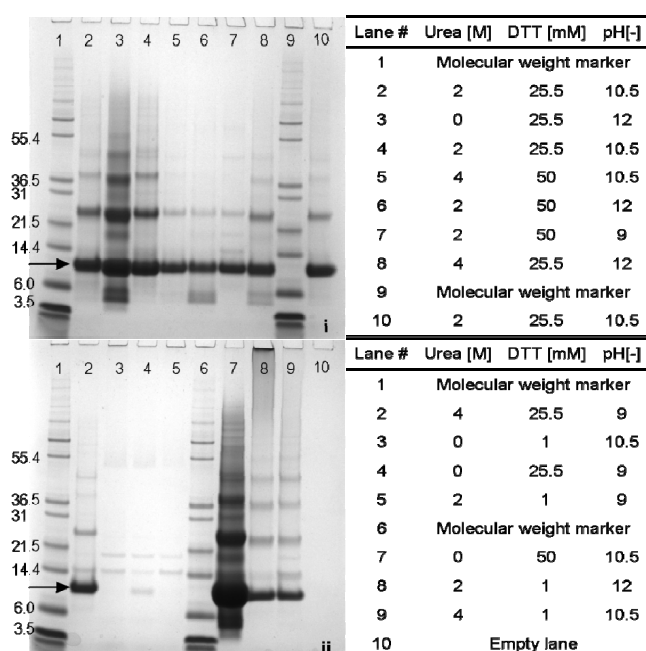


Figure 2.1: SDS-PAGE (4-12%) analysis of solubilized fusion protein from the model inclusion bodies, stained with Simply-Blue safe stain. Between 100-112 mg of IBs (based on dry weight) was suspended in the buffers prepared based on the various experimental combinations. The monomer fusion protein is indicated by the arrow. The molecular markers (from bottom to top): 3.5, 6.0, 14.4, 21.5, 31, 36.5 and 55.4 kDa

The experimental results presented in Figure 2.1 show that the aggregation during solubilization was prevented by adding urea in combination with dithiothreitol at alkaline pH. Urea increases the solubility of those hydrophobic intermediates prone to aggregate and affects the structure of water by hydrogen bonding increasing the protein solvation [29-31]. Nevertheless, the presence of Urea will not prevent, even at high concentrations (6-8M), the formation of soluble aggregates which is a major drawback of inclusion bodies solubilization, and has a direct effect on the performance of the subsequent refolding step [32].

The reducing environment is of paramount importance for an efficient solubilization of proteins expressed as inclusion bodies, especially for those that contain cysteines. This is

because the proteins are trapped in the inclusion bodies in a reduced state; once dissolved they will rapidly oxidize forming soluble aggregates by inter-chain cystines. Regarding the interaction between pH and Urea, we concluded that urea solubilization is a function of the pH of the solution; which is in agreement with the review paper from Vallejo et al,[33]. Finally it was observed that the exposure of the protein to extreme alkaline pH and a reduced environment lead to product degradation (Figure 2.1-Gel i-lanes 3, 6 and 8).

The statistical analysis of the experimental runs was done using an analysis of variance (ANOVA) with a confidence interval of 95%. The analysis indicated that regarding the protein solubilization, changes in the concentration of urea, dithiothreitol and in the pH of the buffer have a positive effect on the amount of soluble protein recovered and on the fraction of soluble monomer. The experimental methodology allowed us to find an optimum balance between maximum protein solubility and minimum fraction of soluble aggregates. The optimal solubilization conditions were: 4M Urea, 25mM DTT and pH 10.5.

Regarding solubilization of IBs with Urea or guanidine HCL, in most cases the common practice is to use 6-8M urea and 6-7M guanidine to achieve an efficient solubilization [34]. However we have shown that a more economical and efficient solubilization could be achieved at urea concentrations relatively low and relatively mild alkaline pH, by considering combined effect of the chosen process parameters.

Further experiments were performed varying one parameter at the time (Table 2.1). The results showed that the formation of soluble aggregates can be controlled, even at high protein concentrations ( $>1\text{mg}\cdot\text{ml}^{-1}$ ), by controlling the concentrations of the chaotrope and reducing agents. These process parameters influence the two forces responsible for protein aggregation, namely hydrophobic and inter-chain disulfides. Therefore by modulating their concentrations, by a controlled buffer exchange, it will then be possible to favor the formation of native monomer during the renaturation step, leading to higher process yields [35-39].

Table 2.1 Effect of urea, dithiothreitol and pH on the performance of the IBs solubilization

UREA [M]	DTT [mM]	pH*	Soluble protein [mg-BSA·ml <sup>-1</sup> ]	Aggregates <sup>a)</sup> fraction	Monomers <sup>a)</sup> fraction	$Y_{s-Protein}$ [mg Solub.protein/ mg Dry-IBs]
0	0	12	3.9 ± 0.28	0.65	0.34	0.73 ± 0.14
1,5	0	12	5.3 ± 0.17	0.59	0.39	0.97
4	0	10.5	1.52±0.28	0.89	0.09	0.28
4	25	10.5	7.7±1.02	0.25	0.75	1

<sup>a)</sup>The fractions of monomers and aggregates were determined by peak area integration of the elution patterns obtained by SEC. The samples were incubated at the conditions presented and the column was equilibrated using the same conditions.

To validate the optimal conditions obtained, 106 mg of dry inclusion bodies were dissolved in 19.33 ml of buffer containing 4M Urea, 25mM DTT and pH 10.5. The dissolution took less than 5 min and the resulting solution had no detectable turbidity at 350nm, thus no insoluble aggregates were formed. The solubilization efficiency was a 100%, which means that all the suspended solids in dry weight end up as soluble protein. Subsequently the soluble protein mixture was fractionated by size exclusion chromatography using a Superdex-75 10/300 in order to determine the fraction of soluble aggregates. The column was equilibrated with the same buffer composition as the feed (no buffer exchange was done). The estimated mass distribution at equilibrium was 24.8%±0.02 as soluble aggregates and 75.2%±0.04 as monomer soluble monomer (Figure 2.2), for a total protein concentration of 7.7±1.02mg·ml<sup>-1</sup>; which is by a factor of 3.85 higher than what has been reported in similar work [1].

### 2.3.2 Solubilization of the Inclusion bodies by alkaline pH

Inclusion bodies were solubilized over a pH range from 9 to 12 using an initial solid concentration of 5.6mg Dry-weight IBs·ml<sup>-1</sup>. The results indicate that solubility of this inclusion bodies has a high dependency on the pH; which results in an 80% decrease of the soluble protein when the pH was 9 compared with pH 12 (Data not shown). Maximum solubility of the inclusion bodies was observed at pH 12 with an efficiency of 73±10%, which means that approximately 27% of the mass of solids suspended resulted in macro aggregates observed as a cloudy solution. The solubilization resulted in fraction of soluble aggregates of 59±14% and 34±10% of soluble monomers. These percentages were calculated by peak area integration of the elution pattern of the fractionated sample, obtained using a Superdex 75 10/300 SEC column (Figure 2.2, thicker solid line). The separate peaks were collected and analyzed by SDS-PAGE to determine their purity. The soluble aggregates (A) eluted at

retention volume 7.70 ml, whereas the monomers (PF and FM) eluted at a retention volume of 10.55 and 11.28ml, respectively (Figure 2.2). The native monomer was estimated after the tryptic digestion of the pool of monomers and the subsequent analysis by RPHPLC, using human insulin as standard.

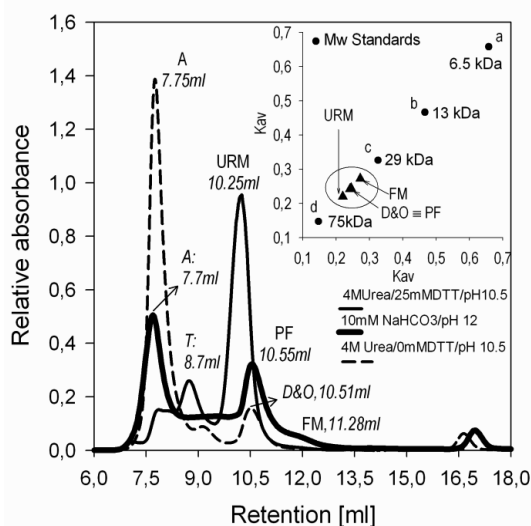


Figure 2.2: SEC-Chromatograms Superdex 75. The y-axis has the units [peak area / mAU]. The three chromatograms correspond to 3 different solubilization conditions (See figure). The species identified were: aggregates (A), trimer (T), unfolded and reduced monomer (URM), partially folded monomer (PF), denatured and oxidized monomer (D&O) and folded monomer (FM). The inset ( $K_{av}$  vs.  $K_{av}$ ) compares the different monomer conformation against the globular standards (a-d)

### 2.3.3 Limitations of pH induced solubilization

The dissolution of IBs by means of alkaline pH does not need a subsequent renaturation, because in this protocol both processes take place simultaneously. This gives an apparent advantage over the use of chaotropes and therefore we decided to quantify for this system the refolding yield and the overall process yield. The inclusion bodies were suspended in solubilization buffer at a pH approximately of 12 for a period of 30 to 40 min, and then the pH was decreased to approximately 11 for 30 hours. The monomers (native and non-native) were obtained by SEC fractionation and their amount was estimated by peak area integration (Figure 2.2 pool from 10.55 to 11.28 ml).

The solubilization efficiency of this system was calculated as 0.734 ( ), which means that a solid concentration of 5.47mg/ml will result in 3.99 mg/ml of soluble protein. The resulting soluble monomer was estimated as ~34% and a refolding yield of 28%±1.9% over the product loaded to the reaction vessel; which confirmed that pH induced solubilization and refolding systems are limited, even at low protein concentrations (~0.13mg·ml<sup>-1</sup>), by the formation of soluble aggregates and by the formation of intermediate species which elute between the front of aggregates and the front of monomers.

### 2.3.4 Identification of different monomer conformations

Size exclusion chromatography is a non-invasive analytical tool and therefore is perfectly suitable for the study of the solubilization of IBs at different incubation conditions. In addition by SEC the separation of multimers and monomers is easily achievable; this is because the elution fronts of bigger particles travel faster. During solubilization the proteins get denatured, reduced and oxidized, resulting in variations of shape and size [22, 40, 41]. These changes in the structure of the monomers can be measure by the changes in the average partition coefficient ( $K_{av}$ ). In theory the native conformations have the smallest stokes's radius, and therefore a larger partition coefficient compared to denature molecules, partially folded molecules and aggregates.

The studied monomer has a molecular weight of 12.70kDa (based on its structure) and contains 3 disulfide bonds. Depending on the buffer composition, the soluble monomer can be present as denatured and reduced (URM), denatured and oxidized (D&O), partially folded (PF) and folded (FM) (Figure 2.2). The URM was obtained after the solubilization of the IBs in 4M urea / 25mM DTT and pH 10.5; the PF and the FM were obtained in 10mM NaHCO<sub>3</sub>/0.1mM EDTA and pH 12; the D&O monomer was obtained in 4M urea, 0mM DTT and pH 10.5. All these conformations have the same molecular mass but they partition different over the size exclusion material ( $0.22 < K_{av} < 0.3$ ).

The results presented in Figure 2.2 showed that the monomer is trapped in the IBs in a reduced form. This explains why when the IBs are dissolved in 4M urea with no RA approximately 77% of the product ended up as soluble aggregates and only ~10% was recovered as monomer (D&O); whereas by including 25mM of DTT approximately 75% of the ended up as monomer (URM) and approximately 5% ended up as soluble aggregates.

Regarding the changes in size and shape, the Superdex column was first calibrated using globular shape standards of known Stokes radii. The data obtained was used to generate two calibration plots. The first was  $K_{av}$  vs. Log Mw, which was used to estimate the molecular weight and the shape of the eluting compounds observed. The second plot related the average partition coefficient  $K_{av}$  to the Stoke's radius [22, 42]; which was used as a calibration plot for this column to estimate the stoke's radius (hydrodynamic radius) of the eluting compounds. Using the known molecular mass of the monomer and the calibration line ( $K_{av}$  vs. Log Mw) assuming a globular shape, the theoretical  $K_{av}$  was calculated as 0.50, equivalent to an elution volume of 13.67ml. Experimentally an average partition coefficient of the folded monomer (FM) was estimated  $0.27 \pm 0.04$  (Figure 2). Thus we concluded that the monomer has a rod-like shape and therefore size exclusion chromatography could not

be used to determine its molecular weight, unless it is coupled with SDS-PAGE analysis of the eluting fronts.

To summarize, the solubilization of IBs by means of alkaline pH can be achieved without the need of a chaotropic agent; which apparently makes this an economic option; nevertheless using pH alone will not prevent the formation of soluble aggregates (Figure 2.2). At the end the process yield is limited by the solubilization step, to a maximum which will not surpass the fraction of soluble monomer. Furthermore the solubilization by means of alkaline pH it is unfortunately highly protein dependent, which affects its general applicability.[10, 18, 43]; this is a clear disadvantage compared to the method develop during this work.

## 2.4 Conclusions

We have described a novel experimental methodology to develop a suitable solubilization strategy for recombinant proteins expressed as IBs. Our results show us that via this approach one can find an optimal balance between maximum protein solubility and minimum formation of soluble aggregates. We also compared our system with a solubilization done without chaotrops or reducing agents for high protein concentrations (>2 mg·ml<sup>-1</sup>). The conclusion was that relatively mild conditions will not prevent nor reverse the formation of soluble aggregates, reaffirming then the applicability of chaotrops and reducing agents. Based on the presented evidence, the way to go in order to develop efficient solubilization strategies is via statistically designed experiments. They allow the optimization of the concentrations of the reagents by considering their combined effect over the performance parameters (i.e. protein solubility and percentage of soluble monomer), thus making better use of them which leads then to efficient processes. This work also showed that via a careful study of the solubilization of the IBs using reducing agents, chaotrops and a non invasive analytical technique (e.g. SEC) will give much information about how to control the formation of partially folded monomers, unfolded and reduced monomers, folded monomers and aggregates. This information can then be used for the careful design of the renaturation step, resulting then in a complete optimization of the solubilization and renaturation of proteins expressed as IBs. This will be the focus of our future work.

## 2.5 References

- [1] Singh, S., Panda, A., Solubilization and refolding of bacterial inclusion body proteins. *J. Biosci. Bioeng.* 2005, 99, 303-310.
- [2] Valax, P., Georgiou, G., Molecular characterization of  $\beta$ -lactamase inclusion bodies produced in *Escherichia coli*. 1. Composition. *Biotechnol. Prog.* 1993, 9, 539-547.
- [3] Clark, E. D. B., Schwarz, E., Rudolph, R., Inhibition of Aggregation side reactions during in vitro protein folding. *Methods Enzymol.* 1999, 309, 217-236.
- [4] Petrides, D., Cooney, C. L., Evans, L. B., Field, R. P., Snoswell, M., Bioprocess simulation: An integrated approach to process development. *Comput. Chem. Eng.* 1989, 13, 553-561.
- [5] Cabrita, L. D., Bottomley, S. P., Protein expression and refolding - A practical guide to getting the most out of inclusion bodies *Biotechnology annual review* 2004, 10, 31-50.
- [6] Yasuda, M., Murakami, Y., Sowa, A., Ogino, H., Ishikawa, H., Effect of Additives on Refolding of a Denatured Protein. *Biotechnol. Prog.* 1998, 14, 601-606.
- [7] Cardamone, M., Puri, N. K., Brandon, M. R., Comparing the Refolding and Reoxidation of Recombinant Porcine Growth Hormone from a Urea Denatured State and from *E. Coli* IBs. *Biochemistry* 1995, 34, 5773-5794.
- [8] Kurucz, I., Titus, J. A., Jost, C. R., Segal, D. M., Correct disulfide pairing and efficient refolding of detergent-solubilized single-chain Fv proteins from bacterial inclusion bodies. *Mol. Immunol.* 1995, 32, 1443-1452.
- [9] Clark, E. D. B., Refolding of recombinant proteins. *Curr. Opin. Biotechnol.* 1998, 9, 157-163.
- [10] Khan, R. H., Rao, K. B. C. A., Eshwari, A. N. S., Totey, S. M., Panda, A. K., Solubilization of Recombinant Ovine Growth Hormone with Retention of Native-like Secondary Structure and Its Refolding from the Inclusion Bodies of *Escherichia coli*. *Biotechnol. Prog.* 1998, 14, 722-729.
- [11] Winter, J., Lilie, H., Rudolph, R., Renaturation of human proinsulin--a study on refolding and conversion to insulin. *Anal. Biochem.* 2002, 310, 148-155.
- [12] Clark, E. D. B., Protein refolding for industrial processes. *Curr. Opin. Biotechnol.* 2001, 12, 202-207.
- [13] Middelberg, A. P. J., Preparative protein refolding. *Trends Biotechnol.* 2002, 20, 437-443.
- [14] Rudolph, R., Lilie, H., In vitro folding of inclusion body proteins. *The FASEB Journal* 1996, 10.
- [15] Jungbauer, A., Kaar, W., Current status of technical protein refolding. *J. Biotechnol.* 2007, 587-596.



- [16] Avena, S. M., Pusey, M. L., Bogle, D. L., Protein Solubility Modeling. *Biotechnol. Bioeng.* 1999, 64, 144-150.
- [17] Okumura, S., Saitoh, H., Wasano, N., Katayama, H., et al., Efficient solubilization, activation, and purification of recombinant Cry45Aa of *Bacillus thuringiensis* expressed as inclusion bodies in *Escherichia coli*. *Protein Expression Purif.* 2006, 47, 144-151.
- [18] Patra, A. K., Mukhopadhyay, R., Mukhija, R., Krishnan, A., et al., Optimization of Inclusion Body Solubilization and Renaturation of Recombinant Human Growth Hormone from *Escherichia coli*. *Protein Expression Purif.* 2000, 18, 182-192.
- [19] Tsumoto, K., Umetsu, M., Kumagai, I., Ejima, D., Arakawa, T., Solubilization of active green fluorescent protein from insoluble particles by guanidine and arginine. *Biochem. Biophys. Res. Commun.* 2003, 312, 1383-1386.
- [20] Umetsu, M., Tsumoto, K., Ashish, K., Nitta, S., et al., Structural characteristics and refolding of in vivo aggregated hyperthermophilic archaeon proteins. *FEBS Lett.* 2004, 557, 49-56.
- [21] Vinogradov, A. A., Kudryashova, E. V., Levashov, A. V., Dongen, W. M. A. M. v., Solubilization and refolding of inclusion body proteins in reverse micelles. *Anal. Biochem.* 2003, 320, 234-238.
- [22] Batas, B., Jones, H. R., Chaudhuri, J. B., Studies of the hydrodynamic volume changes that occur during refolding of lysozyme using size-exclusion chromatography. *Journal of Chromatography A* 1997, 766, 109-119.
- [23] Sofer, G., Hagel, L., *Handbook of process chromatography: A guide to optimization, scale up, and validation* ACADEMIC PRESS 1999.
- [24] *Methods*, U. S. P.
- [25] Douglas, M., *Design and Analysis of Experiments*, Wiley 2000.
- [26] WEI, C., TANG, B., ZHANG, Y., YANG, K., Oxidative refolding of recombinant prochymosin. *Biochem. J.* 1999, 340, 345-351.
- [27] Anderson, D. E., Becktel, W. J., Dahlquist, F. W., pH-Induced Denaturation of Proteins: A Single Salt Bridge Contributes 3-5 kcal/mol to the Free Energy of Folding of T4 Lysozyme. *Biochemistry* 1990, 29, 2403-2408.
- [28] Schein, C. H., Solubility as a function of protein structure and solvent components. *Nat. Biotechnol.* 1990, 8, 308-317.
- [29] Makhatadze, G. I., Thermodynamics of Protein Interactions with Urea and Guanidinium Hydrochloride. *The journal of Physical Chemistry B* 1999, 103, 4781-4785.
- [30] Schellman, J. A., Gassner, N. C., The enthalpy of transfer of unfolded proteins into solutions of urea and guanidinium chloride. *Biophys. Chem.* 59, 59, 259-275.
- [31] Tandford, C., Isothermal unfolding of globular proteins. *J. Am. Chem. Soc.* 1964, 86, 2050-2059.

- [32] Schlegl, R., Iberer, G., Machold, C., Necina, R., Jungbauer, A., Continuous matrix-assisted refolding of proteins. *Journal of Chromatography A* 2003, 1009, 119-132.
- [33] Vallejo, L. F., Rinas, U., Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. *Microbial Cell Factories* 2004, 3-11.
- [34] Tsumoto, K., Ejima, D., Kumagai, I., Arakawa, T., Practical considerations in refolding proteins from inclusion bodies. *Protein Expression Purif.* 2003, 28, 1-8.
- [35] Creighton, T. E., Electrophoretic analysis of the unfolding of proteins by urea. *J. Mol. Biol.* 1979, 129, 235-264.
- [36] Fahey, E. M., Chaudhuri, J. B., Binding, P., Refolding and purification of a urokinase plasminogen activator fragment by chromatography. *Journal of Chromatography B* 2000, 737, 225-235.
- [37] Gu, Z., Su, Z., Janson, J.-C., Urea gradient size-exclusion chromatography enhanced the yield of lysozyme refolding. *Journal of Chromatography A*, 2001, 918, 311-318.
- [38] Li, M., Su, Z., Refolding Human Lysozyme produced as an inclusion body by urea concentration and pH gradient ion exchange Chromatography. *Chromatographia* 2002, 56, 33-38.
- [39] Li, M., Zhang, G., Su, Z., Dual gradient ion-exchange chromatography improved refolding yield of lysozyme. *Journal of Chromatography A* 2002, 959, 113-120.
- [40] Batas, B., Chaudhuri, J. B., Considerations of sample application and elution during size-exclusion chromatography-based protein refolding. *Journal of Chromatography A* 1999, 864, 229-236.
- [41] Batas, B., Schiraldi, C., Chaudhuri, J. B., Inclusion body purification and protein refolding using microfiltration and size exclusion chromatography. *J. Biotechnol.* 1999, 68, 149-158.
- [42] Uversky, V. N., Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry* 1993, 32, 13288-13298.
- [43] Umetsu, M., Tsumoto, K., Nitta, S., Adschiri, T., et al., Nondenaturing solubilization of [beta]2 microglobulin from inclusion bodies by L-arginine. *Biochem. Biophys. Res. Commun.* 2005, 328, 189-197.

# 3

## Ion-Exchange Chromatographic Protein Refolding

---

### Abstract

The application of ion-exchange (IEX) chromatography to ion-exchange protein refolding (IEXR), has been successfully proven, as supported by the various studies using different model proteins, ion-exchange media and flow configurations. Compared to batch dilution refolding, ion-exchange refolding offers a relatively high degree of process intensification, represented by the possibility of performing protein refolding, product purification and product concentration, in one unit operation. Besides the relatively high degree of process intensification offered, its additional key advantages include: spatial isolation of the bound protein molecules and the controllable change in chemical composition, attained using gradients. Despite of the acknowledgement of its main advantages, the lack of mechanistic understanding on how these influence the process performance on the ion-exchange refolding reactor, limits the ability to exploit them in order to optimize the performance of the unit. This paper presents a quantitative analysis that assesses the effect that, the spatial isolation and the urea gradient, have on the IEXR performance, judged on the basis of the refolding yield ( $Y_N$ ) and the fractional mass recovery ( $f_{\text{Prot,Rec}}$ ). Additionally, this work also discusses the effect of the protein load, the protein loading state (i.e., native, denatured, denatured and reduced (D&R)) and the adsorbent type on  $f_{\text{Prot,Rec}}$ . The presented work show: (1) that the protein load has a direct effect on  $f_{\text{Prot,Rec}}$ , and the magnitude of this effect depends on the loading state of the protein solution and the adsorbent type; (2) that irrespectively of the type of adsorbent used, the saturation capacity of a denatured protein is less than the native protein and that this difference can be linked to differences in accessible binding surface area; (3) that there is a good correlation between fractional surface coverage ( $\theta$ ) and  $f_{\text{Prot,Rec}}$ , indicating that the former could serve as a good descriptor to assess spatial isolation, and (4) the urea gradient has a direct link with the variations on the refolding yield, and this link can be quantitatively estimated using as descriptor the urea gradient slope ( $\xi$ ). Overall, the information provided in this paper aims at the eventual development of rational design or selection strategies of ion-exchange media for the satisfactory and successful refolding of a target protein.

Keywords: Ion-exchange refolding; matrix assisted refolding; surface coverage

---

This chapter has been submitted as E. Freydell, L.A.M van der Wielen, M. Eppink, M. Ottens to J. Chrom. A (2010) (under review)

### 3.1 Introduction

Developments attained in recombinant DNA technology have significantly changed the way valuable proteins are produced today. Proteins that used to be purified from human fluids, animal or plant tissue can now be produced in large quantities using for example *Escherichia coli* (*E. coli*). As an expression system *E. coli* offers several advantages including (1) its molecular genetics are well understood, meaning that its genome can be modified with easy; (2) relatively inexpensive culturing procedures and (3) high fermentation yields [1-3]. One important limitation of this expression system though, is that the over expression of certain gene sequences leads to the accumulation of the product in an inactive, insoluble aggregate known as inclusion body (IB). To obtain the soluble and bio-active (native) product, two process steps are required and these are (1) inclusion bodies (IBs) solubilization and (2) protein refolding. IBs solubilization is usually done using a solution containing chaotropes (e.g., urea, guanidine hydrochloride), reducing agents (e.g., dithiothreitol,  $\beta$ -mercaptoethanol, etc) and alkaline pH. This cocktail disrupts the intermolecular interactions holding the aggregated protein, releasing the product in a denatured and reduced (D&R) soluble form. Optimal solubilization conditions should provide maximum protein solubility, minimizing the fraction of soluble aggregates formed and maximizing the fraction of soluble denatured and reduced monomer product [4-7]. Protein refolding is achieved by decreasing the concentration of chaotropes and reducing agents in the concentrated protein solution, allowing the soluble protein to refold and to form its disulphide bonds. This change in chemical composition is basically a buffer exchange step that can either be attained by direct dilution or using liquid chromatography (chromatographic refolding).

Compared to batch dilution, chromatographic protein refolding offers a relatively high degree of process intensification, represented by the possibilities of performing: (a) protein refolding, (b) a controllable change in chemical composition, attained using gradients, (c) product purification and (d) product concentration, in one single unit operation. Chromatographic refolding has been successfully conducted using hydrophobic interaction chromatography (HIC) [8-10], size-exclusion chromatography (SEC) [11-12], immobilized metal affinity chromatography (IMAC) [13-14], and ion-exchange chromatography (IEX) [15-17].

Ion-exchange refolding (IExR) was probably first reported by Creighton [17], who showed that (1) a urea denatured protein could be reversibly bound to an ion-exchange resin, (2) the denaturant (i.e. urea) could be removed using a gradient, while keeping the protein bound inducing refolding and disulfide bond formation and (3) using a salt gradient, the refolded

bound protein could be eluted, obtaining it in a relatively pure denaturant free fraction. Ever since, IExR has gained interests in the scientific community most likely due to reasons such as: (1) ion-exchange chromatography is widely used as a purification unit in the biopharmaceutical industry, thus IExR could be easily implemented into current processes; (2) compared to SEC, IEX chromatography offers high loading capacities and the possibility to concentrate the product during elution, using a salt gradient; (3) compared to HIC, the urea denatured protein can be easily bound to the IEX resin at relatively low salt concentrations, which is certainly not the case for HIC refolding and (4) protein-protein interactions are presumably mitigated as a result of the spatial isolation obtained upon the binding of the protein to the IEX resin, decreasing protein aggregation, positively affecting the refolding yield. The approach presented by Creighton has been, through the years, modified to include for example a concomitant pH, urea and salt gradient that has successfully been applied to refold lysozyme on a cation-exchanger [18-19] and iron superoxide dismutase (Fe-SOD) on an anion-exchanger [20]. More recently, using a no-flow incubation period IExR has successfully been applied to refold bovine serum albumin (BSA) [15] and  $\alpha$ -fetoprotein (AFP) [21], demonstrating that IExR is well suitable to refold relatively complex proteins, as the latter two examples have 16 and 17 disulfide bonds, respectively. Despite the proven versatility of IExR, represented by its successful application to the refolding of a range of proteins (e.g., Lysozyme, BSA, Fe-SOD) using different flow configurations (e.g., single gradient, dual gradient, no-flow incubation), the contributions of its key advantages (the spatial isolation of the bound protein molecules and the controlled change in chemical composition) to the unit's performance have not yet been quantitatively assessed.

This paper presents a quantitative analysis that assesses the effect of the fractional surface coverage ( $\theta$ ), the slope of the urea gradient ( $\xi$ ) and the load of denatured and reduced protein on the performance of IExR, judged on the basis of refolding yield ( $Y_N$ ) and fractional mass recovery ( $f_{\text{Prot,Rec}}$ ). Additionally, this study also discusses the effect of the resin backbone and the influence of the spacer length. Fractional surface coverage and the slope of the urea gradient quantitatively describe the degree of spatial isolation and urea gradient, respectively. Accordingly, these variables provide the possibility to assess the contribution of these key advantages to the unit's performance and to determine whether their contribution is positive or adverse. As distinct from the approach presented by Li et al [18-20] the flow configuration used in this study, decouples the salt and the denaturant gradient, allowing the independent study of the effect of the chaotrope gradient. Our results show: (1) the strong correlation between  $f_{\text{Prot,Rec}}$  and fractional surface coverage and how this correlation is influenced by the loading state of the protein (e.g., native, denatured, denatured and reduced) and the backbone of the adsorbent; (2) the effect that the mass

loaded of denatured and reduced (D&R) protein has on  $Y_N$  and  $f_{\text{Prot,Rec}}$ , and how the loading state of the protein affects the magnitude of this effect; and (3) the effect of slope of the urea gradient ( $\xi$ ) on  $Y_N$  and  $f_{\text{Prot,Rec}}$ .

## 3.2 Materials and Methods

### 3.2.1 Materials

Two model proteins were used in this study. The first was a fusion protein (FP) that has 117 amino acids, a molecular weight of 12 710 g mol<sup>-1</sup>, three disulfide bonds, no free cysteines and a theoretical isoelectric point (pI) of 7.64 (based on its primary sequence). The protein was obtained in the form of inclusion bodies and was provided by Shering-Plough (Oss, the Netherlands). Based on its isoelectric point this model protein was used for the anion-exchange refolding experiments. The second model protein was Lysozyme, purchased as L6876 from Sigma-Aldrich (Zwijndrecht, The Netherlands). Lysozyme has 129 amino acids, a molecular weight of 14 307 g mol<sup>-1</sup>[22], four disulfide bonds [23], no free cysteines and a theoretical isoelectric point of 11.35 [24]. Based on its pI this model protein was used for the cation-exchange refolding experiments.

All the columns used for the ion-exchange refolding experiments were pre-packed columns, with a packed bed volume ( $V_c$ ) of 1 ml and were purchased from GE-Healthcare (Uppsala, Sweden), these columns were: RESOURCE Q ( $V_c$ = 1ml), RESOURCE 15S ( $V_c$ = 1ml), Hi-Trap SPFF and Hi-Trap SPXL. The SOURCE ion-exchange media is based on porous particles with a hydrophilic matrix made from polystyrene/divinyl benzene (PS/DVB) and substituted with quaternary ammonium (Q) or methyl sulfonate (S) groups. The Hi-Trap columns are packed with Sepharose Fast Flow (FF) media and Sepharose XL media. The Sepharose FF media is made of particles with a macroporous gel structure, with a neutral hydrophilicity, based on chains of agarose that are arranged in bundles [25]. The matrix is made of 6% agarose, highly cross-linked. Sepharose XL media is based on the same structure as the Sepharose FF media, however for this media the ionic ligands are bound to a long, approximately 40 kDa in molecular size [26], dextran chain grafted onto the agarose matrix prior to its functionalization, resulting in the functionalization of both the agarose matrix and the dextran chain [26].

Analytical size-exclusion of the fractions collected during the IExR experiments, was done using a Superdex 75 10/300 pre-packed gel filtration column, purchased from GE-Healthcare (Uppsala, Sweden). All chromatographic separations were performed on an ÄKTA explorer 10 equipped with the UNICORN software version 5.01 from GE Healthcare (Uppsala, Sweden).

All chemicals used were at least reagent grade purity or higher. Urea, DL-Dithiothreitol (DTT) and Guanidine-HCL (GuHCL) were purchased from Sigma-Aldrich (Zwijndrecht, The

Netherlands). Sodium hydrogen carbonate, sodium hydroxide, sodium chloride and tris-(hydroxymethyl)-aminomethane (Tris) were purchased from JT.Baker (Deventer, The Netherlands). Acetone, ethylenediamine tetra acetic acid (EDTA), and Hydrochloric acid were purchased from Merck (Schiphol-Rijk, The Netherlands). All solutions were prepared using water purified by Milli-Q Ultrapure Water Purification System from Millipore (Amsterdam, The Netherlands) and were vacuum filtered through a 0.22  $\mu\text{m}$  pore size membrane filter from Pall (Portsmouth, Hampshire, United Kingdom).

### **3.2.2 Protein quantification**

The concentration of soluble protein was estimated using a BCA protein assay purchased from Fisher Scientific (Landsmeer, The Netherlands). Bovine serum albumin (BSA) was employed as a standard to estimate the total protein concentration of the fractions from the anion-exchange refolding experiments. Lysozyme was used as the standard to estimate the total protein concentration in the fractions collected from the cation-exchange experiments.

### **3.2.3 Quantification of the native fusion protein**

The refolded fusion protein was digested using trypsin in order to obtain the mature monomer. The digestion was done using an enzyme to substrate ratio, of 1:300 (mg:mg) [27-28]. 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 with a 100mM HCL solution. The concentration of active protein was determined using reversed phase HPLC and a calibration line constructed using human insulin as standard [29].

### **3.2.4 Inclusion bodies solubilization**

The inclusion bodies were solubilized in solubilization buffer (4M Urea/ 25mM DTT/ 10mM  $\text{NaHCO}_3$ /0.1mM EDTA; pH 10.5) [4]. Solutions of various concentrations were prepared diluting the concentrated denatured and reduced protein solution with solubilization buffer.

### **3.2.5 Preparation of lysozyme solutions**

Lysozyme in a native form was prepared in a solution of 50mM Tris-HCL, pH 8.7. Denatured not reduced (DNR) lysozyme was prepared in a solution of 8M Urea/ 50mM Tris, pH 8.7. Denatured and reduced (D&R) lysozyme was prepared in a solution of 8M Urea/ 100mM DTT/ 50mM Tris /pH 8.7 [18].

### **3.2.6 Analytical size-exclusion chromatography**

Analytical SEC was used to estimate the amount of monomers and high-molecular weight aggregates present in the fractions collected from the IExR experiments. The analysis was



conducted using a Superdex 75 10/300, at a flow rate of 0.5 ml min<sup>-1</sup> and an injection volume of 0.2 ml. Prior to the analysis the column was calibrated using Aprotinin (6 500 Da), Ribonuclease A (13 700 Da), Carbonic anhydrase (29 000 Da) and conalbumin (75 000 Da). The columns void volumes, inter-particle volume and total volume, were determined with pulses on blue-dextran and acetone, respectively.

### 3.2.7 Isotherm measurements

Adsorption isotherms were determined for lysozyme under native and denaturing conditions (i.e., lysozyme in 8M urea), for the Hi-Trap SP and the RESOURCE 15S columns. The isotherms were measured by frontal analysis (FA) [30] and the breakthrough curves (BTCs) were analyzed according to the approach presented by Gritti et al [31]. The adsorbed protein concentration ( $q^*$ ) in equilibrium with the feed concentration ( $C$ ) was determined with Eq. (3.1).

$$q^* = \frac{C(V_{\text{eq}} - V_0)}{V_C} \quad (3.1)$$

where  $V_{\text{eq}}$  represent the volume of the equivalent area [31],  $V_0$  is the inter-particle void volume,  $V_C$  represent the packed bed volume, and  $C$  is the concentration of the feeding solution.

### 3.2.8 Anion-exchange refolding

Anion-exchange refolding (AExR) experiments were done using the Resource-Q anion-exchanger and the fusion protein. These experiments evaluated the effect of the slope of the denaturant gradient and the load of denatured and reduced protein on the refolding yield and the fractional mass recovery. Table 1 presents the list of buffers used and figure 3.1 presents the schematic representation of the flow scheme, indicating the different stages (e.g., equilibration, loading, etc). The AExR experiments were performed as follows. The column was first equilibrated for 10 column volumes (CVs) with buffer B1 (table 3.1), at 2 mlmin<sup>-1</sup>. Subsequently, a feed pulse of 1ml, containing the denatured and reduced protein, at a concentration of 1 mgml<sup>-1</sup> was injected. The column was thereafter washed for 5 CVs with buffer B1. Next, the urea concentration was decreased using a linear gradient, exchanging buffer B1 for buffer A11 in 20CVs. This gradient was operated at a flow rate selected to achieve the desired gradient slope ( $\xi$ ). The slopes used were 0.35, 0.10 and 0.03 mol l<sup>-1</sup> min<sup>-1</sup>, obtained using the flow rates ( $\phi_V$ ) 2, 0.57 and 0.17 mlmin<sup>-1</sup>, and Eq. (3.2).

$$\xi = \frac{(C_{\text{Urea,Init}} - C_{\text{Urea,End}})}{\Delta CV} \cdot \frac{\phi_V}{V_C} \quad (3.2)$$

where  $\xi$  represents the gradient slope,  $C_{\text{Urea,Init}}$  represents the urea concentration at the beginning of the gradient,  $C_{\text{Urea,End}}$  represents the urea concentration at the end of the

gradient,  $\Delta CV$  represent the gradient's length,  $\phi_V$  represents the flow rate and  $V_C$  represents the packed bed volume.

Following the urea gradient, the column was washed for another 5CVs with buffer A11. Then, the bound protein was eluted using a sodium chloride (NaCl) gradient. This gradient was done at  $2 \text{ mlmin}^{-1}$ , in 10CVs, exchanging buffer A11 for buffer B2 (table 3.1). The column was then washed for another 5CVs with buffer B2. Finally, the column was rinsed with milli-Q water for 5CVs, regenerated using a 3M guanidine hydrochloride (GnHCL) solution for 5CVs and finally rinsed for 5CVs with milli-Q water; all these steps were done at  $2 \text{ mlmin}^{-1}$ .

During the AExR experiments the effect of the gradient slope ( $\xi$ ), on the refolding yield ( $Y_N$ ) and the fractional mass recovery, was studied at a fixed denatured and reduced protein load of 1mg. The effect of the denatured and reduced protein load was studied at a fixed gradient slope of  $0.10 \text{ mol l}^{-1} \text{ min}^{-1}$ , and for a D&R protein load ranging from 1-4 mg of protein.

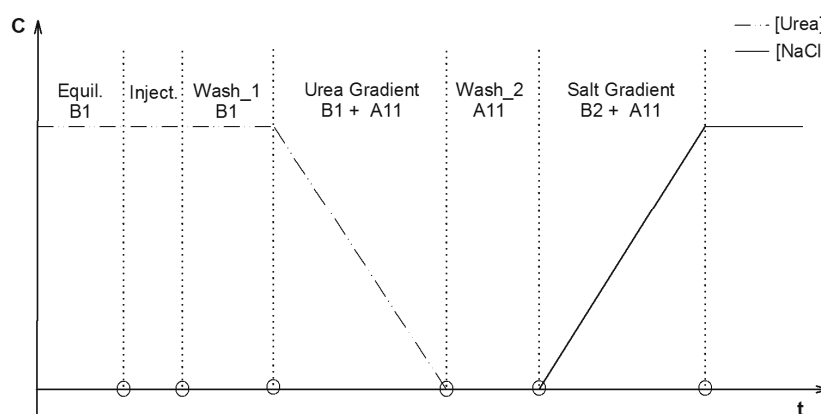


Figure 3.1 Schematic representation of the flow scheme used for the anion-exchange refolding (AExR) experiments. Dash-dot line: Urea concentration. Solid thick line: Sodium chloride concentration. Equil: Equilibration block; Inject: Injection block; B1: Buffer B1 (table 3.1); A11: Buffer A11 (table1); B2: Buffer B2 (table3.1)

Table 3.1: Solutions used for the anion-exchange refolding (AExR) experiments

Buffer	Urea (M)	NaCl (mM)	Tris (mM)	NaHCO <sub>3</sub> (mM)	EDTA (mM)	pH (-)
A11	0.5	-	50	20	0.2	10.5
B1	4	-	50	20	0.2	10.5
B2	0.5	1000	50	20	0.2	10.5

### 3.2.9 Cation-exchange refolding

Cation-exchange experiments were done to study the effect of the denatured and reduced protein load, the backbone of the adsorbent and the spacer length on the fractional mass recovery. These experiments were done using the following strong cation-exchangers: RESOURCE 15S, Hi-trap SPFF and Hi-trap SPXL. The effect of the backbone was assessed by comparing the data obtained using the relatively hydrophilic backbone of the SOURCE adsorbent, based on PS/DVB, against the data obtained using the Sepharose adsorbents (SPFF, SPXL), whose backbone is based on cross-linked agarose. The effect of the spacer or the lack thereof was studied by comparing the data obtained from the experiments with the SPFF and SPXL adsorbents, as both adsorbents have the same backbone and the same functional group. All these experiments were done using lysozyme as the model protein. Additionally, these experiments also evaluated the effect of the loading state of the protein. By loading state it is meant: native lysozyme, denatured lysozyme or denatured and reduced lysozyme. These loading states were generated dissolving lysozyme on buffers with or without urea and with urea and DTT (section 2.5). All these experiments were done using a lysozyme feed concentration ranging from 0.5-4.0 mgml<sup>-1</sup> and a fixed loading volume of 1 ml.

Fig. 3.2 presents the schematic representation of the flow scheme used for the CExR experiments, indicating the various blocks (e.g., equilibration, injection, etc) and table 3.2 presents the list of buffers used. The CExR experiments were done as follows. The column was first equilibrated for 12CVs with buffer A13 (table 3.2), for the experiments with either denatured or D&R lysozyme, or with buffer A11 (table 3.1) for the experiments with native lysozyme. Subsequently, a feed pulse of 0.5 ml, containing either native, denatured or D&R lysozyme, was injected. Next, the column was washed with the equilibration buffer for 5CVs. Thereafter, the mobile phase composition was changed on a step fashion (Wash\_2, Fig. 3.2), going from buffer A13 to A11 in the experiments done with denatured or D&R lysozyme. For those experiments with native lysozyme no such mobile phase change took place, as for these experiments buffer A11 was used in the blocks equilibration, injection, Wash\_1 and Wash\_2 (Fig. 3.2). Right after block Wash\_2 (Fig. 3.2), the bound protein was eluted with a linear sodium chloride gradient, changing the mobile phase composition from A13 or A11 to buffer B1 in 10CVs. Then the column was washed with buffer B1 for 5CVs. Finally, the column was rinsed with milli-Q water for 5CVs, regenerated using a 3M GmHCL solution for 5CVs and finally rinsed for 5CVs with milli-Q water. All the blocks on the flow scheme were executed at 1 mlmin<sup>-1</sup>.

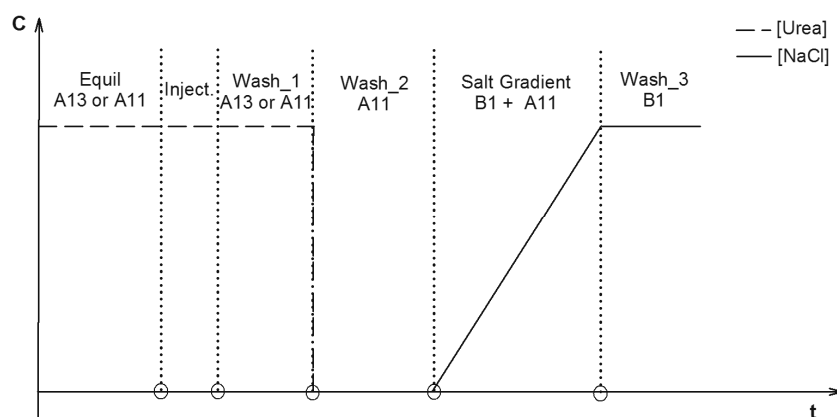


Figure 3.2 Schematic representation of the flow scheme used for the cation-exchange refolding (CEXR) experiments. Dash-line: Urea concentration. Solid-thick line: Sodium chloride concentration. Equil: Equilibration block; Inject: Injection block. A13, A11, B1 are the buffers used (see table 2)

Table 3.2: Solutions used for the cation-exchange refolding experiments

Buffer	Urea (M)	NaCl (M)	Tris (mM)	pH (-)
A13	8	-	50	8.7
<sup>a)</sup> A11	0	-	50	8.7
B1	0	1	50	8.7

<sup>a)</sup>Buffer A11 was used, in place of buffer A13, for the experiments using native lysozyme .

### 3.2.10 Isotherm modeling

The isotherm data determined for lysozyme, under native and denaturing conditions, was analyzed using two mechanistic models, namely the Langmuir model and the Fowler model. The Langmuir model is commonly used to analyze adsorption data obtained on homogeneous surfaces and preferably under conditions where the fractional surface coverage ( $\theta$ ) is less than 0.10 [30]. The model assumptions include: the solute gives monolayer coverage and the bound solutes do not interact with each other in the monolayer. The Langmuir model is presented in Eq. (3.3).

$$q^* = \frac{bq_s C}{1 + bC} \quad (3.3)$$

where  $C$  represents the liquid concentration in equilibrium with the adsorbed concentration  $q^*$ ,  $q_s$  is the saturation capacity of the stationary phase and  $b$  is a numerical coefficient related to the affinity of the solute for the binding surface.

The Fowler model was conceived as a model to correct for the first-order deviations of the Langmuir model, assuming that the bound solutes may interact while bond to the homogeneous surface [30]. Eqs. (3.4a) and (3.4b) describe the Fowler model.

$$bCe^{-\chi\theta} = \frac{\theta}{1-\theta} \quad (3.4a)$$

$$\theta = \frac{q^*}{q_s} \quad (3.4b)$$

where  $\chi$  represents the interaction energy and  $\theta$  represents the fractional surface coverage. The interaction parameter ( $\chi$ ) describes weak interactions between bound and incoming solute molecules.

### 3.2.11 Parameter estimation

The parameters of the isotherm models, i.e., Langmuir and Fowler, were estimated using a non-linear least squares algorithm programmed and solved in MATLAB R2007b. The objective function, minimized by the algorithm, was based on the method of Marquardt, as presented in Guiochon et al [30]. The objective function is presented in Eq. (3.5). To solve Fowler's model, the optimization routine was constraint for  $1 < \chi < 4$ , because in this range the model has physical meaning.

$$f(x) = \sqrt{\frac{1}{N_D - P} \sum_{i=1}^n \left( \frac{q_{\text{exp},i}^* - q_{\text{mod},i}^*(x)}{q_{\text{exp},i}^*} \right)^2} \quad (3.5)$$

where  $N_D$  represents the number of data points,  $P$  represents the number of parameters of the model,  $q_{\text{exp},i}^*$  represent the experimental adsorbed concentration,  $q_{\text{mod},i}^*$  represent the modeled adsorbed concentration,  $n$  is the number of elements of the concentration vectors, and  $x$  represent the vector of the model parameters.

### 3.2.12 Performance indicators

#### 3.2.12.1 Fractional mass recovery

The fractional mass recovery ( $f_{\text{Prot,Rec}}$ ) is defined as the total mass of protein recovered in the elution, divided by the total mass of protein injected to the ion-exchanger and it was calculated using:

$$f_{\text{Prot,Rec}} = \frac{V_{\text{EPool}} C_{\text{Prot,E}}}{V_{\text{inj}} C_{\text{Prot,feed}}} \quad (3.6)$$

where  $f_{\text{Prot,Rec}}$  represent the fractional mass recovery,  $V_{\text{EPool}}$  represent the volume of the elution pool,  $V_{\text{inj}}$  represent the injection volume,  $C_{\text{Prot,E}}$  is the total protein concentration in the elution pool and  $C_{\text{Prot,feed}}$  is the total protein concentration fed to the column. For those experiments, were denatured and reduced (D&R) protein was used  $C_{\text{Prot,feed}} = C_{f,D\&R}$ . Where  $C_{f,D\&R}$  represent the feed concentration of denatured and reduced protein.

#### 3.2.12.2 Refolding yield

The refolding yield ( $Y_N$ ) is defined as the amount of active product formed per amount of denatured and reduced protein loaded to the column and it was calculated as follows:

$$Y_{N,\text{IEXR}} = \frac{V_{\text{MPool}} C_{\text{NPool}}}{V_{\text{inj}} C_{f,D\&R}} \quad (3.7)$$

where  $V_{\text{MPool}}$  represent the volume of the pool carrying the native product,  $C_{\text{NPool}}$  is the concentration of native protein in the pool.

#### 3.2.12.3 Volumetric productivity

The volumetric productivity of a chromatographic refolding reactor is defined as the amount of product (i.e., native protein) produced per unit time per volume of chromatographic medium

$$\text{Pr}_{\text{IEXR}} = \frac{C_{f,D\&R} V_{\text{inj}}}{t_{\text{cycle}} V_C} Y_{N,\text{IEXR}} \quad (3.8)$$

where  $C_{f,D\&R}$  is the feed concentration of denatured and reduced protein,  $V_{\text{inj}}$  is the injection volume,  $t_{\text{cycle}}$  is the cycle time,  $V_C$  is the packed bed volume and  $Y_{N,\text{IEXR}}$  is the ion-exchange refolding yield.

### 3.3 Results and Discussion

#### 3.3.1 Adsorption behavior under denaturing and native conditions

Changes in the protein structure, such as the disruption of the tertiary and/or secondary structures, have been shown to have a strong influence in the adsorption behavior of proteins onto surfaces [32], and thus these changes will have a direct influence in the saturation capacity of the ion-exchanger and the affinity of a given protein for the surface of the adsorbent. Changes in protein structure may occur as a result of, for example, the disruption of hydrogen bonding occurring due to the action of a chaotropic agent (e.g., urea, GuHCL). To assess the effect of such structural changes on the adsorption behavior, adsorption experiments using native lysozyme, denatured lysozyme and the ion-exchangers Source 15S and Sepharose SPFF were done. The results of these experiments are presented in figure 3.3. Figure 3.3A and 3.3B present the adsorption isotherm of lysozyme under denaturing and non-denaturing (native) conditions, respectively. The isotherms were measured by frontal analysis (FA). From the data on figures 3.3A-3.3B the following observations are evident: (1) the Sepharose SPFF has a higher capacity for lysozyme than the Source 15S media, for both native and denaturing conditions and (2) the saturation capacity under native conditions is higher than under denaturing conditions, regardless the type of adsorbent use. The latter behavior seems to be protein independent, as similar observations using native and denatured and reduced  $\alpha$ -lactalbumin, adsorbed onto a Source Q IEX adsorbent, have been reported [16]. The relatively low saturation capacity, exhibited by the denatured lysozyme compared to the native lysozyme, might be reasonably explained by a change in the accessible binding surface area, as this has been shown to be directly related to the molecule's dimensions [33-34]. In essence, molecules of different sizes have access to different fractions of the pore volume of a given adsorbent, affecting their accessibility to the binding surface. Basically, a big molecule will have access to less binding surface area than a small molecule. Denatured lysozyme is a larger molecule than native lysozyme [35], hence it has less access to the binding surface, explaining its relatively low saturation capacity. The saturation capacities of denatured and reduced lysozyme, for both the Source 15S and the Sepharose SPFF, are presented in table 3. The data on table 3.3 was derived from the analysis of the adsorption data with the Fowler model (Eqs. (3.4a)-(3.4b)) and the Langmuir model (Eq.(3.3)). It is interesting to point out from the data in table 3.3, that both models predict practically a change in saturation capacity (denatured vs. native) of equal magnitude. This finding supports the hypothesis that the change in saturation capacity ( $q_s$ ) is caused by the differences in accessibility of denatured and native lysozyme to the binding surface, as both ion-exchange materials have been shown to have very close mean pore radius and phase ratio functions [34].

Lastly, although both models are in principle suitable for the analysis of the adsorption data, as shown by their ability to represent the experimental trend (Fig 3.3A,3.3B), the Langmuir model is not adequate for the presented case, as the fractional surface coverage ( $\theta$ ) in these experiments goes beyond 0.10 [30]. Therefore, the Fowler model will be used in follow up discussions.

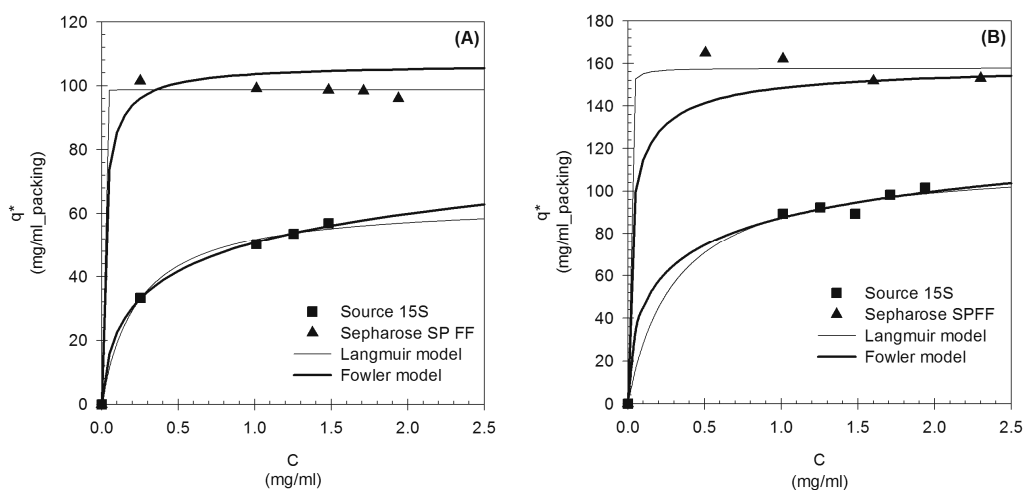


Figure 3.3 Adsorption isotherms of lysozyme loaded in a denatured (A) and native (B) state, to the adsorbents Source 15S (■) and Sepharose SPFF (▲). Solid thin line: Langmuir model (Eq.(3.3)); Solid thick line: Fowler model (Eqs. (3.4a)-(3.4b)). The parameters of the models, estimated with the aid of Eq. (3.5), are presented in table 3.3

### 3.3.2 Effect of the protein load and the loading state on the fractional mass recovery ( $f_{\text{Prot,Rec}}$ ) and the refolding yield ( $Y_N$ )

Protein load is an important operational variable in ion-exchange refolding (IExR) because it directly affects the volumetric productivity of the reactor (Eq.(3.8)), which is an important performance indicator. And thus, its effect on the fractional mass recovery (Eq.(3.6)) and the refolding yield (Eq.(3.7)) of the IExR unit was studied using both the cation- (CExR) and the anion-exchange (AExR) model systems. Additionally, the effect of the loading state (i.e., denatured, D&R) was also investigated, since during ion-exchange refolding the protein might be fed either as denatured protein or as denatured and reduced protein, depending on whether a reducing agent (e.g., DTT,  $\beta$ -mercaptoethanol) was used or not during the solubilization of the inclusion bodies.

Fig. 3.4 presents the results from the CExR experiments. These series of experiments were done using native lysozyme (▲), denatured lysozyme (●), D&R Lysozyme (■) and the cation



exchangers Source 15S (Fig. 3.4A), Sepharose SPFF (Fig. 3.4B) and Sepharose SPXL (Fig. 3.4C). The first question addressed, using the data from the CExR experiments, was whether the protein load had an effect on the fractional mass recovery ( $f_{\text{Prot,Rec}}$ ). The data in figure 3.4 show that the protein load does affect the  $f_{\text{Prot,Rec}}$  and this effect seems to be adverse. Furthermore, these data show that this effect depends on the type of cation exchanger and on the loading state of the protein (i.e., native, denatured, D&R). The effect of the type of cation exchanger seems to be linked to the adsorbent's backbone; this claim is supported by the comparison of figs. 3.4A and 3.4B. This comparison shows that the effect of the protein load on the  $f_{\text{Prot,Rec}}$  is more pronounced on the hydrophilic backbone of the Source 15S media (Fig. 3.4A) than on the neutrally hydrophilic backbone of the Sepharose media (Figs. 3.4B, 3.4C). It is important to point out from the data in Fig. 3.4A, that the effect of the protein load is relatively more pronounced for the native lysozyme, than for the denatured lysozyme. This behavior may be reasonably explained by the hydrophilic nature of the Source's backbone, since the native protein binds more tightly than the denatured protein, as the latter should have a higher hydrophobic surface area than the former, as a result of the denaturation process.

The effect of the loading state is evident from the comparison of the experimental data corresponding to the D&R lysozyme (■) and the data from the native (▲) and denatured (●) lysozyme (figs. 3.4A, 3.4B, 3.4C). From this comparison, it can be concluded that loading under D&R conditions results in the minimum fractional mass recovery, irrespectively of the amount loaded or the type of adsorbent use. To explain this finding, the following hypothesis is formulated, loading under D&R conditions results in the formation of a tightly bound layer of protein, insensible to the action of sodium chloride. Loading under D&R conditions means that the protein being fed comes with a high concentration of free thiolate anions ( $-S^-$ ), which are reactive species in nature. Since these species have a negative charge, their interaction with the negatively charged surface of the cation exchanger, is improbable. As a result, interactions between bound protein molecules and bound protein molecules with incoming protein molecules are likely to occur via the thiolate anion, forming inter-chain disulfide bonds. Additionally, hydrophobic interactions between protein molecules may also occur, as under D&R conditions the core of the protein molecules is fully exposed. As a result, a multilayer of protein forms on top of the adsorbent's surface, blocking the accessibility of the  $Na^+$  (counter-ion) to the first layer of protein allegedly bound solely via ionic interactions with the adsorbent's ligands. This hypothesis is supported by the data presented in fig. 3.4D. Figure 3.4D presents a typical elution chromatogram obtained using the Hitrap SPFF column and a protein load of 0.5 mg. The data show that when lysozyme is loaded under native (fig. 3.4D Dash-dot line) or denatured (fig. 3.4D solid thin line)

conditions, elution with sodium chloride (NaCl) is feasible. However, elution with NaCl is inefficient when lysozyme is injected in a denatured and reduced form (fig. 4D solid thick line), resulting in the elution of the protein during the regeneration with the 3M guanidine hydrochloride solution.

Lastly, a comparison between the data on figures 3.4B and 3.4C, indicated that having a long spacer does not seem to bring an advantage, as the behavior of the data obtained with both the Sepharose SPFF and the Sepharose SPXL follows practically equal trends

Up to this point the data have made clear the effect of the protein load on the fractional mass recovery, but what is its effect on the refolding yield? To address this question the AExR model system was used. Figure 5 presents the results of the anion-exchange refolding experiments, obtained using the Source Q anion-exchange media, a fixed urea gradient slope ( $\xi$ ) of  $0.10 \text{ mol l}^{-1} \text{ min}^{-1}$  and the fusion protein. The fusion protein was fed to the column in a denatured and reduced state, as both urea and DTT were used for the efficient solubilization of the inclusion bodies [4]. Figure 5A presents the AExR chromatograms, showing the three fractions collected (i.e., F1, F2, F3). The data show that the increase in loading resulted mostly in an increase of the peak area corresponding to fraction F3. Analysis of the AExR fractions using analytical SEC (Fig. 3.5B), indicated that fraction F3 is predominantly composed of aggregated protein. The SEC analysis also revealed that the percentage of aggregated protein in fraction F3 correlates well with the protein load used during the AExR experiments (inset Fig. 3.5B), increasing as the protein load increases. This finding explains the adverse effect that the protein load had on the refolding yield of the AExR system, which is evident from the data on table 3.4. Table 3.4 summarizes the results from the AExR experiments, indicating the effect of the protein load on the refolding yield and the fractional mass recovery of the AExR system. It is important to point out from the data on table 3.4, that the effect of the protein load on the  $f_{\text{Prot,Rec}}$  of the AExR system, is very similar to the effect measured during the CExR experiments, as in both systems increasing the protein load decreases the  $f_{\text{Prot,Rec}}$ . Thus, it is safe to say that protein load adversely affects the fractional mass recovery of ion-exchange refolding (IExR) systems. This claim is further supported by similar trends reported for a different model protein and using a different flow scheme configuration [15].

Finally, a comparison between the findings obtained from the AExR and CExR data and the reported data of the concomitant salt, pH, urea system [18-19], strongly suggested that endowing the elution buffer with a relatively high urea concentration will likely increase the  $f_{\text{Prot,Rec}}$  of the IExR system, loaded with D&R protein. This hypothesis was confirmed using

the AExR model system by adding 4M urea to the elution buffer, leading to a  $f_{\text{Prot,Rec}} \geq 0.80$  (data not shown). This hypothesis was also confirmed by Langenhof et al [15], who reported a  $f_{\text{Prot,Rec}} \approx 0.90$  when the elution buffer of his AExR model system was endowed with a 8M urea concentration. Thus, it can be concluded that adding urea to the elution buffer will likely increase the  $f_{\text{Prot,Rec}}$ , however the concentration of urea to be used should be determined for each system, as this value should minimally comply with the following requirements: (1) the concentration of urea in the protein fractions should be less than the critical urea concentration, as being close or above this value will likely result in a decrease in protein activity; and (2) the concentration of urea leaving the IExR unit should be compatible with the subsequent processing steps.

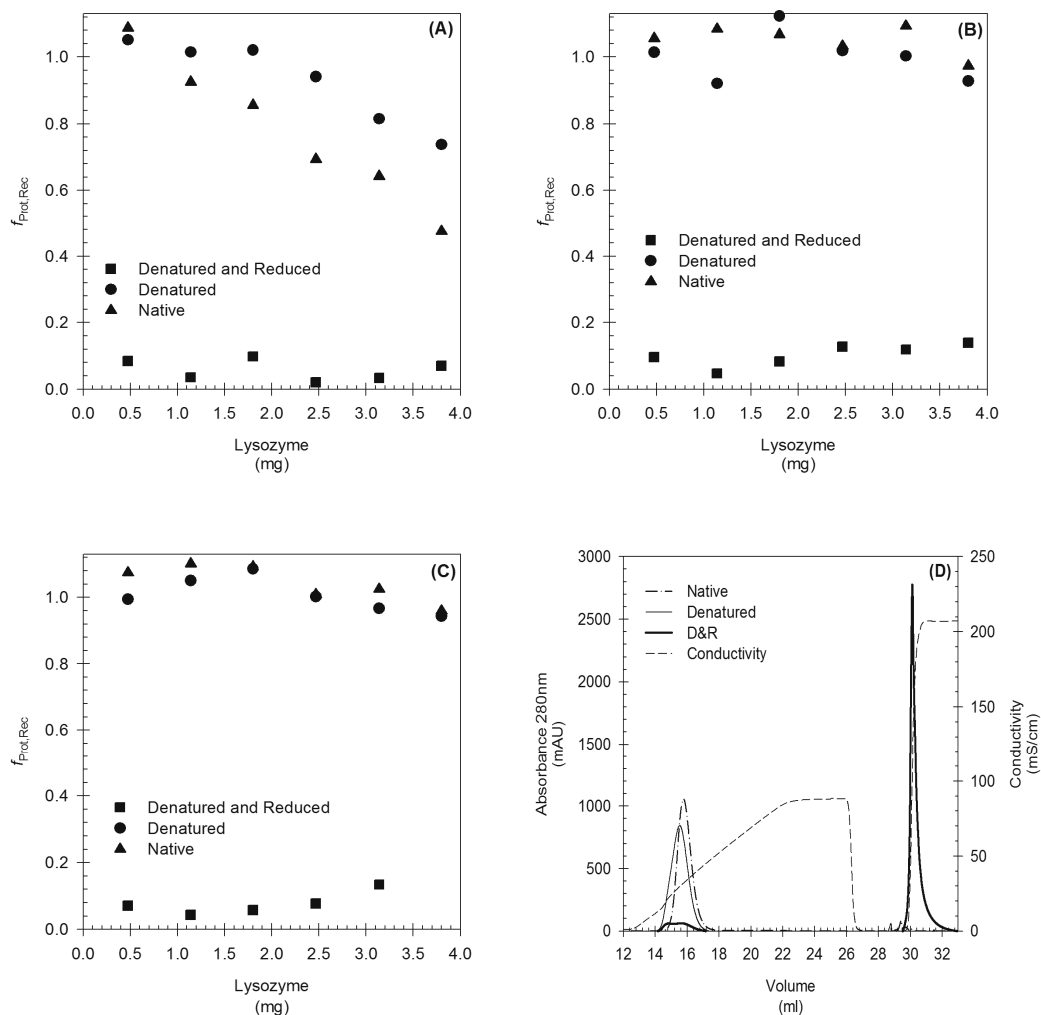


Figure 3.4: Effect of the protein load and the protein feeding state (i.e., native, denatured, D&R) on the fractional mass recovery ( $f_{Prot,Rec}$ ). The experiments were done using the cation-exchangers Source 15S (A), Sepharose SPFF (B), Sepharose SPXL (C) and lysozyme on a native (▲), denatured (●) and denatured and reduced (D&R) (■) state. (D) Typical elution chromatogram obtained during the CExR experiments; Column: HiTrap SPFF, load = 0.5 mg. Dash-dot-dash line: Native lysozyme; Solid thin line: Denatured lysozyme; Solid thick line: D&R lysozyme; Dashed line: Conductivity signal.

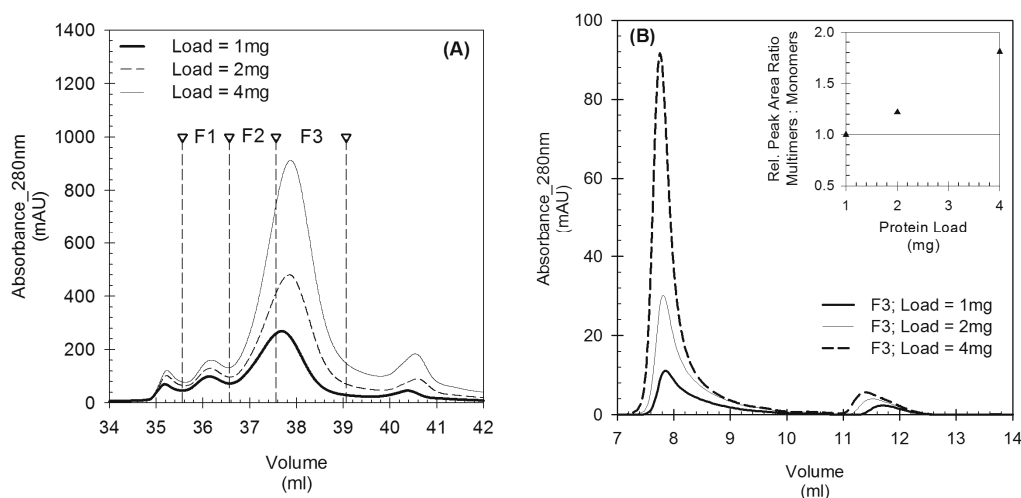


Figure 3.5: Effect of the denatured and reduced (D&R) protein load on the elution chromatogram of the AExR model system. (A) Elution chromatogram of the AExR experiments. Solid thick line: load = 1mg; Short-dashed line: load= 2 mg; Solid thin line: Load = 4 mg. All AExR experiments were done with a urea gradient slope ( $\xi$ ) = 0.10 mol l<sup>-1</sup> min<sup>-1</sup>. (B) SEC analysis of the AExR fractions. Short-dash line: Fraction F3 from the AExR experiment done at a load of 4 mg; Solid thin line: Fraction F3 from the AExR experiment done at a load of 2mg; Solid thick line: Fraction F3 from the AExR experiment done at a load of 1mg. The inset presents the correlation between relative peak area ratio, determined as the peak area of the monomer divided as the peak area of the aggregates, and the AExR protein load.

Table 3.4 Effect of the protein load on the refolding yield and the fractional mass recovery of the anion-exchange model system

<sup>a)</sup> Load (mg protein / ml bed)	<sup>b)</sup> $Y_N$ (g/g)	$f_{Prot,Rec}$ (g/g)
1	0.18	0.57
2	0.08	0.41
4	0.06	0.35

<sup>a)</sup>The protein was loaded as denatured and reduced protein

<sup>b)</sup>Urea slope ( $\xi$ ) fixed at 0.1mol l<sup>-1</sup> min<sup>-1</sup>

### 3.3.3 Fractional surface coverage ( $\theta$ ) and its relation to the fractional mass recovery ( $f_{Prot,Rec}$ )

Spatial isolation, resulting from the binding of the protein molecules to the adsorbent's surface, has been suggested to be one key advantages of IExR systems [36]. In principle, a high spatial isolation should prevent protein-protein interactions, hence decreasing the chance of protein aggregation. Although theoretically accepted, attempts to quantitatively

assess spatial isolation, in the context of ion-exchange protein refolding, have not been reported yet. Here, the fractional surface coverage ( $\theta$ ) (Eq.(3.4b)) has been chosen as the variable to quantify the degree of spatial isolation. Figure 6 presents the correlation between fractional surface coverage ( $\theta$ ) and fractional mass recovery ( $f_{\text{Prot,Rec}}$ ), for native lysozyme ( $\blacktriangle$ ) and denatured lysozyme ( $\bullet$ ), estimated for the Source 15S (Fig.3.6A) and the Sepharose SPFF (Fig. 3.6B) ion-exchange media. Overall, the data indicates a good correlation between the two variables, showing how as the fractional surface coverage increases, beyond a certain threshold, the fractional mass recovery will commence to descend. A reasonable interpretation of this behavior is that as  $\theta$  increases, the spatial isolation of the bound molecules decreases, increasing the chance of the bound molecules to interact with each-other and with incoming molecules. These interactions may lead then to the formation of a tightly bound protein layer, of variable thickness, that blocks the access of  $\text{Na}^+$  (counter-ion) to the first layer of protein, allegedly bound to the surface of the adsorbent solely via ionic interactions with the ligands; resulting in a decrease of the  $f_{\text{Prot,Rec}}$  owing to an inefficient protein elution during the salt gradient. Lastly, from the data in figs. 3.6A-3.6B is evident that the correlation between  $\theta$  and  $f_{\text{Prot,Rec}}$  is more pronounced for the adsorbent Source 15S than for the adsorbent Sepharose SPFF. The reason for this difference is at this point unknown, however it does suggest that the physicochemical properties of the adsorbent may have a contribution to the behavior of the data, and raises the following questions (1) what are the properties of the adsorbent involved? And (2) what is their contribution to the behavior of the data? These questions, fall out of the scope of the presented work and are material for follow up research.

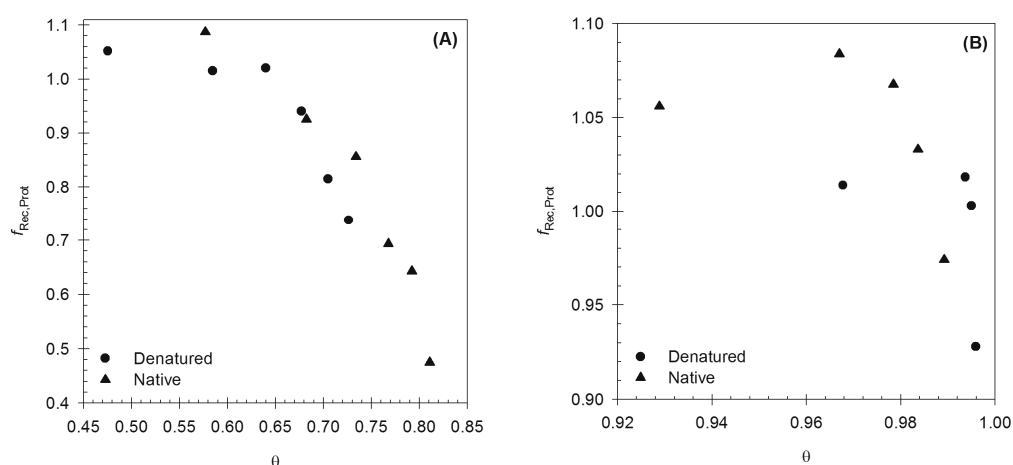


Figure 3.6: Correlation between fractional surface coverage ( $\theta$ ) and fractional mass recovery ( $f_{\text{Rec,Prot}}$ ), for native (▲) and denatured (●) lysozyme loaded to the adsorbents Source 15S (A) and Sepharose SPFF (B).

### 3.3.4 Effect of the urea gradient on the refolding yield ( $Y_N$ ) and the fractional mass recovery ( $f_{\text{Rec,Prot}}$ )

In addition to the spatial isolation of the bound molecules, the ability to change the urea concentration around the protein, in a controllable manner using a gradient, is also a key advantage of ion-exchange refolding (IExR). It has been suggested that the ion-exchange refolding yield is positively influenced by a gradual change in the urea concentration [17], because such gradual change allegedly prevents the rapid collapse of the protein, favoring the formation of the native product [18-19]. Despite the different suggestions and hypotheses, a quantitative analysis indicating if there is a direct link between the urea gradient and the performance of ion-exchange refolding unit, has not been reported yet. So far, the effect of the urea gradient has been reported using a concomitant urea, pH and salt gradient [18-19]. The data on these studies however, cannot be used to tease out the contribution of the urea gradient to the change in refolding yield, since the concentration of urea changes simultaneously with the salt concentration and the pH. In this study, the urea and the salt gradient were decoupled and the pH of the mobile phase was fixed, to investigate the effect of the urea gradient alone. As descriptor of the urea gradient, the urea gradient slope ( $\xi$ ) was used and its definition is presented in Eq. (3.2). The experiments were done using the AExR model system and the conditions for these experiments are presented in section 3.2.8. It is important to point out that these experiments were done at a fixed D&R protein load of 1mg; this relatively low load was chosen as such because a low load

decreases the fractional surface coverage, increases the fractional mass recovery and the refolding yield, as has been shown in previous sections of this paper.

Fig. 3.7A presents the elution chromatograms of the AExR experiments as a function of the urea gradient slope ( $\xi$ ), showing the three distinct fractions that were collected. These fractions were analyzed by BCA to determine the total protein concentration, by RPHPLC to determine the amount of native protein, and by SEC to estimate the fraction of monomers and aggregates. Figure 3.7B presents a typical SEC chromatogram resulting from the analysis of the AExR fractions, showing their composition in terms of monomers ( $10.50 < V_e < 13.50$ ) and aggregates ( $V_e < 10.50$ ). Table 3.5 presents the summary of the results, and from these data the following conclusions could be made: (1) the AEx refolding yield ( $Y_N$ ) of the model protein is inversely proportional to the urea gradient slope ( $\xi$ ), meaning that as  $\xi$  increases the  $Y_N$  decreases and vice versa; and (2) the contribution of the urea gradient slope to the fractional mass recovery ( $f_{Rec,Prot}$ ) is negligible. The fact that the  $f_{Rec,Prot}$  remained practically constant was anticipated, as this indicator strongly depends on the protein load (table 3.4, Figs.3.4A-3.4C), which was fixed during these experiments. The change in the AExR refolding yield as a function of  $\xi$  is reasonably explained by the dependency of the fraction of monomers to aggregates ratio ( $f_M/f_A$ ) on  $\xi$ . This dependency is presented in table 3.5 and the inset of fig.3.7B. The data show that  $f_M/f_A$  is inversely proportional to  $\xi$ , explaining the refolding yield variation. Overall, based on the data gathered it can be concluded that there is a direct link between the AExR yield and the urea gradient, and this link can be quantitatively described by the gradient slope ( $\xi$ ). Finally, it is important to mention that these experiments do not capture the individual contributions to the magnitude of change in the AExR refolding yield, from the speed of change in urea concentration and the change in residence time of the bound protein, as both are changing with  $\xi$ . Eq. 3.9 shows the relation between the residence time, represented by the duration of the urea gradient block ( $t_{UreaGradient}$ ) (Fig. 3.1), and  $\xi$ . Basically, a shallower (low  $\xi$ ) or steeper (high  $\xi$ ) urea gradient not only results in a change in the speed at which the urea concentration is changing, but it also results in a change on the time the protein is bound to the column (a shallower gradient means a longer time and a steeper gradient means a short time). Understanding these individual contributions is important and it is the aim of future work.

$$t_{UreaGradient} = \frac{V_C}{\phi_{V,UreaGradient}} = \frac{1}{\xi} \frac{(C_{Urea,Init} - C_{Urea,End})}{\Delta CV} \quad (3.9)$$



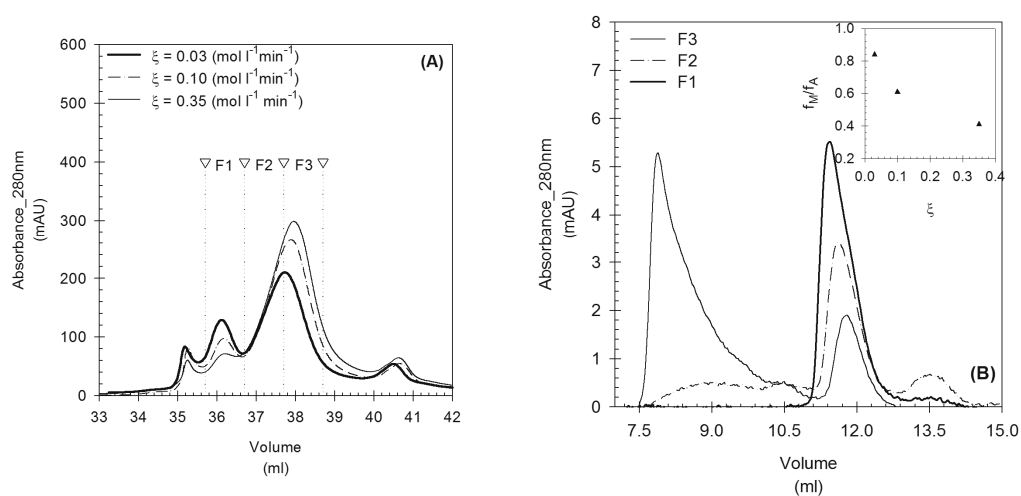


Figure 3.7: AExR refolding experiments as a function of the urea gradient slope ( $\xi$ ). (A) Typical elution profiles of the anion-exchange model system as a function of the urea gradient slope. Solid thick line:  $\xi=0.03$  mol l<sup>-1</sup> min<sup>-1</sup>; Dash-dot line:  $\xi=0.10$  mol l<sup>-1</sup> min<sup>-1</sup>; Solid thin line:  $\xi=0.35$  mol l<sup>-1</sup> min<sup>-1</sup>. The D&R protein load in these experiments was fixed at 1 mg. (B) SEC analysis of the fractions collected during the AExR experiments. Solid thin line: AExR fraction F3; Dash-dot line: AExR fraction F2; Solid thick line: AExR fraction F1. Inset: correlation between the fraction of monomers to aggregates ratio ( $f_M/f_A$ ) and  $\xi$

Table 3.5: Performance of the AExR model system as a function of the urea gradient slope ( $\xi$ )

<sup>a)</sup> $\xi$	$Y_N$	$f_{Prot,Rec}$	<sup>b)</sup> $f_M/f_A$
0.35	0.14	0.55	0.41
0.10	0.18	0.57	0.61
0.03	0.20	0.57	0.84

<sup>a)</sup> Load = 1 mg protein / ml of packed bed.

<sup>b)</sup> This indicator was determined from the SEC analysis of the AExR fractions

### 3.4 Conclusion

The work presented in this paper described a quantitative analysis of two of the main acknowledged advantages of ion-exchange refolding (IExR), namely the spatial isolation and the controllable change in the chemical composition, attained using a urea gradient. These advantages were described using two descriptors, the fractional surface coverage ( $\theta$ ) and the slope of the urea gradient ( $\xi$ ). Using these descriptors, the following observations were made: (a) there is a good correlation between fractional mass recovery ( $f_{\text{Rec,Prot}}$ ) and  $\theta$ . This correlation indicated that as  $\theta$  approaches a certain threshold, the  $f_{\text{Rec,Prot}}$  will begin to decrease. Basically, as the  $\theta$  increases the spatial isolation between bound protein molecules decreases, leading to an increase in protein-protein interactions between the bound protein molecules and incoming molecules, resulting in the formation of a tightly bound layer of protein; and (b) this work showed that there is a direct link between the urea gradient and the ion-exchange refolding yield ( $Y_N$ ), and that this link can be quantitatively described by the urea gradient slope ( $\xi$ ). Additionally this study shows: (1) That the difference in saturation capacity of an adsorbent ( $q_s$ ) for a denatured protein and a native protein, can be reasonably explained by their difference in accessible binding surface area; (2) That the protein load does affect the fractional mass recovery ( $f_{\text{Rec,Prot}}$ ) and that the magnitude of this effect strongly depends on the loading state of the protein solution (i.e., native, denatured, denatured and reduced (D&R)), and the adsorbent type. Furthermore, the effect of the adsorbent is predominantly related to the nature of the adsorbent's backbone; and (3) that increasing the denatured and reduced (D&R) protein load adversely affect the ion-exchange refolding yield, and this adverse effect could be linked to an increase in the mass fraction of aggregated protein.

Finally, as a consequence of the answers given to the initial questions posed at the beginning of this work, it became evident that significant progress could be gain if future studies on ion-exchange protein refolding set to investigate what are the contributions of the physicochemical properties of a given ion-exchange media, to the performance of the ion-exchange refolding reactor. Such knowledge, would pave the way towards the rational design or selection of an ion-exchange media suitable for the efficient refolding of a target protein.

### 3.5 References

- [1] F.R. Schmidt, *Appl. Microbiol. Biotechnol.* 65 (2004) 363.
- [2] K. Terpe, *Appl. Microbiol. Biotechnol.* 72 (2006) 211.
- [3] A.L. Demain, P. Vaishnav, *Biotechnology Advances* 27 (2009) 297.
- [4] E.J. Freydell, M. Ottens, M. Eppink, G.v. Dedem, L.v.d. Wielen, *Biotechnol. J.* 2 (2007) 678.
- [5] K. Tsumoto, D. Ejima, I. Kumagai, T. Arakawa, *Protein. Expres. Purif.* 28 (2003) 1.
- [6] M. Umetsu, K. Tsumoto, S. Nitta, T. Adschiri, D. Ejima, T. Arakawa, I. Kumagai, *Biochem. Biophys. Res. Commun.* 328 (2005) 189.
- [7] S. Singh, A. Panda, *J. Biosci. Bioeng.* 99 (2005) 303.
- [8] X. Geng, X. Chang, *J. Chromatogr.* 599 (1992) 185.
- [9] Q. Bai, Y. Kong, X.-d. Geng, *J. Liq. Chromatogr. Relat. Technol.* 26 (2003) 683
- [10] C. Wang, X. Geng, D. Wang, B. Tian, *J. Chromatogr B.* 806 (2004) 185.
- [11] E.J. Freydell, Y. Bultink, S.H. Van Hateren, L.A.M. van der Wielen, M. Eppink, M. Ottens, *Chem. Eng. Sci.* Article in Press (2010).
- [12] B. Batas, J.B. Chaudhuri, *Biotechnol. Bioeng.* 50 (1996) 16.
- [13] H. Rogl, K. Kosemund, W. Kühlbrandt, I. Collinson, *FEBS Lett.* 432 (1998) 21.
- [14] R. Zahn, C. von Schroetter, K. Wüthrich, *FEBS Lett.* 417 (1997) 400.
- [15] M. Langenhof, S.S.J. Leong, L.K. Pattenden, A.P.J. Middelberg, *J. Chromatogr. A.* 1069 (2005) 195.
- [16] C. Machold, R. Schlegl, W. Buchinger, A. Jungbauer, *J. Biotechnol.* 117 (2005) 83.
- [17] T.E. Creighton, *UCLA Symp. Mol. Cell. Bio., New Ser.* 39 (1986) 249.
- [18] M. Li, G. Zhang, Z. Su, *J. Chromatogr. A.* 959 (2002) 113.
- [19] M. Li, Z. Su, *Chromatographia* 56 (2002) 33.
- [20] M. Li, Z.-G. Su, *Biotechnol. Lett.* 24 (2002) 919.
- [21] Y. Chen, S.S.J. Leong, *J. Chromatogr. A.* 1216 (2009) 4877.
- [22] R.E. Canfield, *J. Biol. Chem.* 238 (1963) 2698.
- [23] E.D.B. Clark, D. Hevehan, S. Szela, J. Maachupalli-Reddy, *Biotechnol. Prog.* 14 (1998) 47.
- [24] L.R. Wetter, H.F. Deutsch, *J. Biol. Chem.* 192 (1951) 237.
- [25] J.-C. Janson, *Korean J. Chem. Eng.* 18 (2001) 149.
- [26] B.D. Bowes, H. Koku, K.J. Czymbek, A.M. Lenhoff, *J. Chromatogr. A.* 1216 (2009) 7774.
- [27] J. Winter, H. Lilie, R. Rudolph, *Anal. Biochem.* 310 (2002) 148.
- [28] J. Winter, P. Neubauer, R. Glockshuber, R. Rudolph, *J. Biotechnol.* 84 (2000) 175.
- [29] *USP27-NF26, US Pharmacopeia* (2007) 2.
- [30] G. Guiochon, D.G. Shirazi, A. Felinger, A.M. Katti, *Fundamentals of Preparative and Nonlinear Chromatography*, Academic Press, 2006.
- [31] F. Gritti, W. Piatkowski, G. Guiochon, *J. Chromatogr. A.* 978 (2002) 81.
- [32] W. Norde, *Clin. Mater.* 11 (1992) 85.
- [33] P. DePhillips, A.M. Lenhoff, *J. Chromatogr. A.* 883 (2000) 39.
- [34] Y. Yao, A.M. Lenhoff, *J. Chromatogr. A.* 1126 (2006) 107.
- [35] B. Batas, H.R. Jones, J.B. Chaudhuri, *J. Chromatogr. A.* 766 (1997) 109.
- [36] A.P.J. Middelberg, *Trends Biotechnol.* 20 (2002) 437.

# 4

## Size-Exclusion Chromatographic Protein Refolding: Fundamentals, Modeling and Operation

---

### 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 ( $C_{i,D\&R}$ ), flow rate, column length, etc, and performance indicators like refolding yield ( $Y_N$ ), thereby opening the possibility for in-silico design of SECRR reactors. 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.

Key words: Size-exclusion; refolding; modeling.

---

This chapter has been submitted to J. Chrom. A. as E. Freydel, L.A.M. van der Wielen, M.H.M. Eppink, M. Ottens, 2010 (under review)

#### 4.1 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 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, Hevenhah 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 et al [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.

## 4.2 Theory and Modeling

### 4.2.1 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. (4.1) and (4.2).

$$\frac{\partial C_{b,i}}{\partial t} = D_{L,i} \frac{\partial^2 C_{b,i}}{\partial x^2} - u \frac{\partial C_{b,i}}{\partial x} - P k_{ov,i} (C_{EqS,i} - C_{S,i}) + r_{b,i} \quad (4.1)$$

$$\frac{\partial C_{S,i}}{\partial t} = k_{ov,i} (C_{EqS,i} - C_{S,i}) + r_{S,i} \quad (4.2)$$

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 are required:

$$x = 0 \quad \frac{\partial C_{b,i}}{\partial x} = \frac{u}{D_{L,i}} (C_{b,i} - C_{f,i}(t)) \quad (4.3a)$$

$$x = L_c \quad \frac{\partial C_{b,i}}{\partial x} = 0 \quad (4.3b)$$

$$t = 0 \quad C_{b,i}(0, 0 < x < L_c) = 0 \quad (4.3c)$$

$$t = 0 \quad C_{S,i}(0, 0 < x < L_c) = 0 \quad (4.3d)$$

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:

$$C_{f,i}(t) = 0 \quad \text{for } t_{\text{pulse}} < t \quad (4.4a)$$

$$C_{f,i}(t) = C_{\text{feed},i} \quad \text{for } t < t_{\text{pulse}} \quad (4.4b)$$



where  $t_{pulse}$  is the feed duration, determined by the volume and flow rate of injection.  $C_{feed,i}$  is the feed concentration for component  $i$ .  $C_{f,i}(t)$  is the concentration of component  $i$  just before entering the column. Eqs. (4.4a) and (4.4b) represent the elution and loading, respectively.

Eqs. (4.1)-(4.4) are solved for each component present in the feed or formed during a given reaction. The partial differential equation was 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. (4.5a) and (4.5b).

$$\frac{d\mathbf{c}_b}{dt} = \mathbf{A} \cdot \mathbf{c}_b - Pk_{ov} (\mathbf{c}_s^* - \mathbf{c}_s) + \mathbf{b} + \mathbf{r}_b \quad (4.5a)$$

$$\frac{d\mathbf{c}_s}{dt} = k_{ov} (\mathbf{c}_s^* - \mathbf{c}_s) + \mathbf{r}_s \quad (4.5b)$$

where  $\mathbf{c}_b$  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.

#### 4.2.2 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 4.1). The correlations used have successfully being applied for the modeling of size-exclusion band profiles [18-20], supporting their use in this work.

Table 4.1: 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]

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

$$D_L = \frac{uL_c}{Pe_L} \quad (4.6)$$

#### 4.2.3 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 4.2.

Table 4.2: Properties of the gel filtration media

Gel filtration material	Fractionation range (Da)	<sup>a)</sup> $d_{\text{pore}}$ (nm)	<sup>b)</sup> $d_p$ ( $\mu\text{m}$ )
Superdex 75	3 000 – 70 000	6.00	13
Sephacryl S-100	1 000 – 100 000	6.60	47

<sup>a)</sup> Reference [31]

<sup>b)</sup> Reference [32]

#### 4.2.4 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. 4.7). 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 et al [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. (4.7).

$$K_{av_i} = \exp\left(-\ln\left(\frac{1}{1-\phi_f}\right)\left(1 + \frac{r_i}{r_f}\right)^2\right) \quad (4.7)$$

where  $K_{av}$  and  $r_i$  represent the average distribution coefficient and the hydrodynamic radius of the solute, respectively.  $\phi_f$  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. (4.8) [33,36].

$$r_i = 0.81 \times 10^{-10} \left(MW^{1/3}\right) \quad (4.8)$$

where  $MW$  is the molecular weight of the protein.

#### 4.2.5 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 et al [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 been 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.

#### 4.2.5.1 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.,  $k_1 \rightarrow \infty$ ), (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  $A_n$ ). Figure 4.1 presents the schematic depiction of this mechanism and table 4.3 presents the different reactions and their corresponding reaction rates.

Table 4.3: Reactions and reaction rates for mechanism 1

Reaction	Reaction rate	Rate law
$I \xrightarrow{\nu_2} N$	$\nu_2 = -\frac{1}{1} \frac{dC_I}{dt} = \frac{1}{1} \frac{dC_N}{dt}$	$\nu_2 = k_2 C_I$
$nI \xrightarrow{\nu_3} A_n$	$\nu_3 = -\frac{1}{n} \frac{dC_I}{dt} = \frac{1}{1} \frac{dC_{A_n}}{dt}$	$\nu_3 = k_3 C_I^p$

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. (4.9a)-(4.9c).

$$r_I = \left. \frac{dC_I}{dt} \right|_{\text{net}} = -v_2 - nv_3 = -k_2 C_I - nk_3 C_I^p \quad (4.9a)$$

$$r_N = \left. \frac{dC_N}{dt} \right|_{\text{net}} = v_2 = k_2 C_I \quad (4.9b)$$

$$r_{A_n} = \left. \frac{dC_{A_n}}{dt} \right|_{\text{net}} = v_3 = k_3 C_I^p \quad (4.9c)$$

where  $v_2$  is the folding reaction rate,  $v_3$  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.

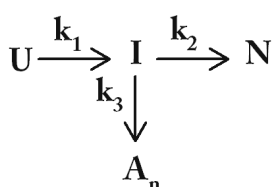


Figure 4.1: Schematic representation of the reaction mechanism 1. U: Unfolded protein, N: Native protein, I: on-pathway intermediate.  $k_1$  and  $k_2$ , are the kinetic constants of the first order reactions ( $\text{s}^{-1}$ ).  $k_3$  is the rate constant of aggregation ( $\text{l}^{(p-1)} \text{mmol}^{(1-p)} \text{s}^{-1}$ ) and  $p$  is the reaction order

#### 4.2.5.2 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].

Figure 4.2 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 4.4.

Table 4.4: Reactions and reaction rates of mechanism 2

Reaction	Reaction rate	Rate law
$I \xrightarrow{\nu_2} N$	$\nu_2 = -\frac{1}{1} \frac{dC_1}{dt} = \frac{1}{1} \frac{dC_N}{dt}$	$\nu_2 = k_2 C_1$
$I + I \xrightarrow{\nu_3} A_2$	$\nu_3 = -\frac{1}{2} \frac{dC_1}{dt} = \frac{1}{1} \frac{dC_{A_2}}{dt}$	$\nu_3 = k_{as} C_1 C_1$
$I + A_2 \xrightarrow{\nu_4} A_3$	$\nu_4 = -\frac{1}{1} \frac{dC_1}{dt} = -\frac{1}{1} \frac{dC_{A_2}}{dt} = \frac{1}{1} \frac{dC_{A_3}}{dt}$	$\nu_4 = k_{as} C_1 C_{A_2}$
$\vdots$	$\vdots$	$\vdots$
$I + A_{i-1} \xrightarrow{\nu_i} A_i$	$\nu_i = -\frac{1}{1} \frac{dC_1}{dt} = -\frac{1}{1} \frac{dC_{A_{i-1}}}{dt} = \frac{1}{1} \frac{dC_{A_i}}{dt}$	$\nu_i = k_{as} C_1 C_{A_{i-1}}$
$\vdots$	$\vdots$	$\vdots$
$I + A_{nmer-1} \xrightarrow{\nu_{nmer}} A_{nmer}$	$\nu_{nmer} = -\frac{1}{1} \frac{dC_1}{dt} = -\frac{1}{1} \frac{dC_{A_{nmer-1}}}{dt} = \frac{1}{1} \frac{dC_{A_{nmer}}}{dt}$	$\nu_{nmer} = k_{as} C_1 C_{A_{nmer-1}}$

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. (4.10a)-(4.10e).

$$r_1 = \left. \frac{dC_1}{dt} \right|_{\text{net}} = -\nu_2 - 2\nu_3 - \dots - \nu_i - \dots - \nu_{nmer} = -k_2 C_1 - 2k_{as} C_1^2 - k_{as} C_1 \sum_{j=2}^{nmer} C_j \quad (4.10a)$$

$$r_N = \left. \frac{dC_N}{dt} \right|_{\text{net}} = \nu_2 = k_2 C_1 \quad (4.10b)$$

$$r_{A_2} = \left. \frac{dC_{A_2}}{dt} \right|_{\text{net}} = \nu_3 - \nu_4 = k_{as} C_1 C_1 - k_{as} C_1 C_{A_2} \quad (4.10c)$$

$$r_{A_i} = \frac{dC_{A_i}}{dt} \Big|_{\text{net}} = v_{i+1} - v_{i+2} = k_{as} C_1 C_{A_{i-1}} - k_{as} C_1 C_{A_i} \quad (4.10d)$$

$$r_{A_{nmer}} = \frac{dC_{A_{nmer}}}{dt} \Big|_{\text{net}} = v_{nmer} = k_{as} C_1 C_{A_{nmer-1}} \quad (4.10e)$$

where  $v_2$  is the folding reaction rate,  $v_i$  is the reaction rate of reaction  $i$ ,  $k_2$  is the folding rate constant and  $k_{as}$  is the association rate constant.

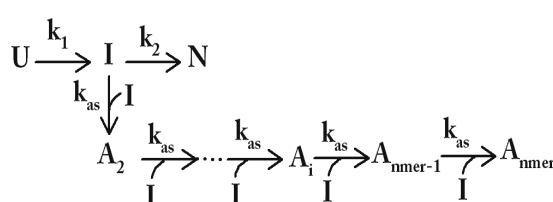


Figure 4.2: Schematic representation of the reaction mechanism 2. Aggregation occurs via sequential polymerization. U: Unfolded state, I: Intermediate state,  $A_{2...nmer}$ : Aggregates class 2 to  $nmer$ , N: Native state.  $k_1$  and  $k_2$ , are the kinetic constants of the first order reactions ( $s^{-1}$ ).  $k_{as}$  is the association constant ( $l \text{ mmol}^{-1} s^{-1}$ )

#### 4.2.5.3 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. (4.11c) 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. (4.11c) changes to  $(i - 1)/2$ .

$$r_1 = \left. \frac{dC_1}{dt} \right|_{\text{net}} = -k_2 C_1 - 2k_{\text{as}} C_1^2 - k_{\text{as}} C_1 \sum_{j=2}^{\text{nmer}} C_{A_j} \quad (4.11a)$$

$$r_N = \left. \frac{dC_N}{dt} \right|_{\text{net}} = k_2 C_N \quad (4.11b)$$

$$r_{A_i} = \left. \frac{dC_{A_i}}{dt} \right|_{\text{net}} = k_{\text{as}} \sum_{j=1}^{i/2} C_{A_j} C_{A_{i-j}} - k_{\text{as}} C_{A_i} C_{A_i} - k_{\text{as}} C_{A_i} \sum_{j=1}^{\text{nmer}} C_{A_j} \quad (4.11c)$$

#### 4.2.6 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 (Eq. 4.1-4.2), 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 figure 4.3. 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 ( $K_{\text{av}}$ ) of the various species, and (4) pre-selection, 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 non-linear 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.

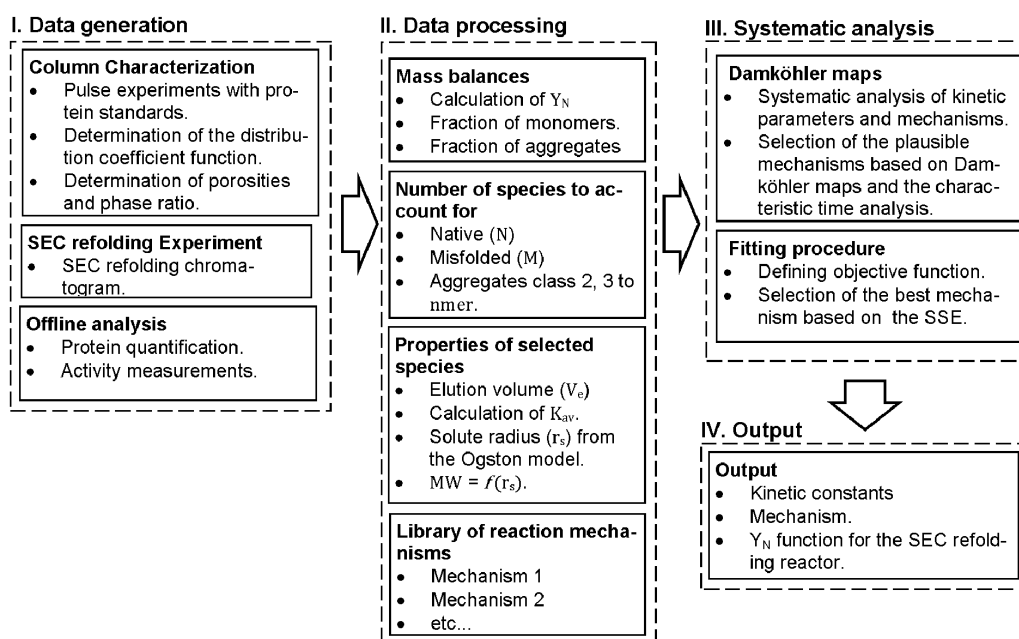


Figure 4.3: Strategy to determine a suitable competitive reaction scheme, capturing the competition between refolding and aggregation, using SEC refolding data.

#### 4.2.7 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  $\beta$ -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 numbers 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 to 3 (section 4.2.5) can be represented by two Damköhler numbers namely  $Da_I$  and  $Da_{II}$ , presented in table 4.5.  $Da_I$  compares the characteristic times of folding ( $\tau_{\text{fold}}$ ) and convection ( $\tau_{\text{conv}}$ ), and  $Da_{II}$  compares the characteristic times of aggregation ( $\tau_{\text{agg}}$ ) and convection.

Table 4.5: Damköhler numbers of folding ( $Da_I$ ) and aggregation ( $Da_{II}$ ), 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	$\frac{k_2 L_c}{u}$	$\frac{L_c k_{as} C_{f,D\&R}}{u}$

#### 4.2.8 Performance criteria

##### *Refolding yield*

Refolding yield is defined as the amount of active product (i.e., native) formed per amount of denatured and reduced protein loaded.

$$Y_{N,SECR,model} = (V_{inj} C_{f,D\&R})^{-1} \int_{V_1}^{V_2} C_{N,model}(V) dV \quad (4.12a)$$

$$Y_{N,SECR,exp} = M_N (C_{f,D\&R} V_{inj})^{-1} \quad (4.12b)$$

$$Y_{N,Batch,model} = \frac{C_{N,model}}{C_{f,D\&R}} \quad (4.12c)$$

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.

### 4.3 Material and Methods

#### 4.3.1 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  $\mu$ m 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=120.6ml, i.d. 1.6cm,  $L_c$  60cm) and a Superdex 75 10/300 (1CV=23.56ml, i.d. 1.0cm,  $L_c$  30cm).

#### 4.3.2 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.

#### 4.3.3 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 100mM HCL solution. The concentration of active protein was determined using RPHPLC and a calibration line constructed using human insulin as standard [53].

#### 4.3.4 Inclusion bodies solubilization

The inclusion bodies were solubilized in the solubilization buffer (4M Urea/ 25mM DTT/ 10mM NaHCO<sub>3</sub>/0.1mM EDTA; pH 10.5) [54]. Solutions of various protein concentrations were prepared by diluting the protein stock solution with solubilization buffer.

#### 4.3.5 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 50mM Tris /10mM NaHCO<sub>3</sub>/0.1mM 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 1mlmin<sup>-1</sup>.

#### 4.3.6 Column calibration and characterization

The Sephacryl S-100 HR HiPrep 16/60 column was calibrated using the following set of protein standards: Aprotinin (6 500), 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 ( $\epsilon_b$ ) and with acetone to determine the total porosity ( $\epsilon_t$ ).

#### 4.3.7 Generation of the Damköhler maps

The systematic analysis was conducted for different scenarios generated varying  $Da_I$  on the range of 0.1 to 100 and  $Da_{II}$  on the range of 1.0 to 200. The simulations were done using the Sephacryl S100 1.6/60 column, a feed concentration of the denatured and reduced ( $C_{i,D\&R}$ ) protein of 1.0 mg/ml, an injection volume of 1ml and a flow rate of 1.0 ml/min. 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.

#### 4.3.8 Size-exclusion distribution coefficient

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

$$K_{av} = \frac{V_c - V_0}{V_t - V_0} \quad (4.13)$$

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).

#### 4.3.9 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 equation 15 and the yield constraint is presented in equation 4.16.

$$\min_x SSE(x) = \sum_{i=1}^{n_t} (C_{\text{exp}_i} - C_{\text{mod}_i}(x, t))^2 \quad (4.14)$$

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

$$Y_{N,\text{mod}}(x) - Y_{N,\text{exp}} = 0 \quad (4.15)$$

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

## 4.4 Results and Discussion

### 4.4.1 Distribution coefficient

The extended Ogston model (Eq.4.7) 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 4.1), from the experimentally determined distribution coefficient. But before using it, the model needs to be calibrated. Figure 4.4 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. 4.4. 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 equation 4.16, 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. 4.4B), specially for large molecules. Figure 4.4B 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.4.16) 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. (4.8) for the other protein standards, whose distribution coefficient was accurately predicted.

$$K_{av} = s \log(MW) + b \quad (4.16)$$



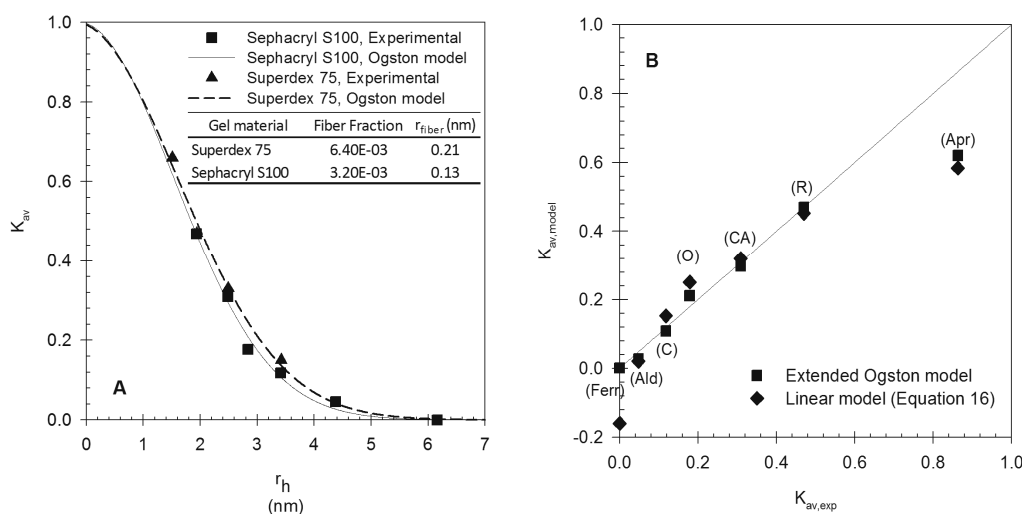


Figure 4.4: (A) Average distribution coefficient experimental and model predictions. Solid line: Distribution coefficient, predicted by equation 7, for the Sephacryl S100 16/600 column. (■) Experimental distribution coefficient, determined with the Sephacryl S100 16/600 column and equation 15. Dashed-line: Distribution coefficient, predicted by equation 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. (◆) Equation 16. (Apr): Aprotinin, (R): Ribonuclease A, (CA): Carbonic Anhydrase, (O): Ovalbumin, (C): Conalbumin, (Ald): Aldolase, (Ferr): Ferritin

#### 4.4.2 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 of the model, to predict the separation, was evaluated by comparing the modeled and experimental chromatogram of a set of protein standards. Figure 4.5A and 4.5B present this comparison, for the Sephacryl S100 16/600 and the Superdex 75 10/300, respectively. The modeled chromatograms were calculated with Eqs. (4.1)–(4.4b), omitting the source terms. The data on figure 4.5 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.,  $\epsilon_b$ ,  $\epsilon_t$ ), the solutes (i.e.,  $MW$ ,  $D_i$ , 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 (Eq. 4.1-4.4) together with a selected reaction mechanism (section 4.2.5).

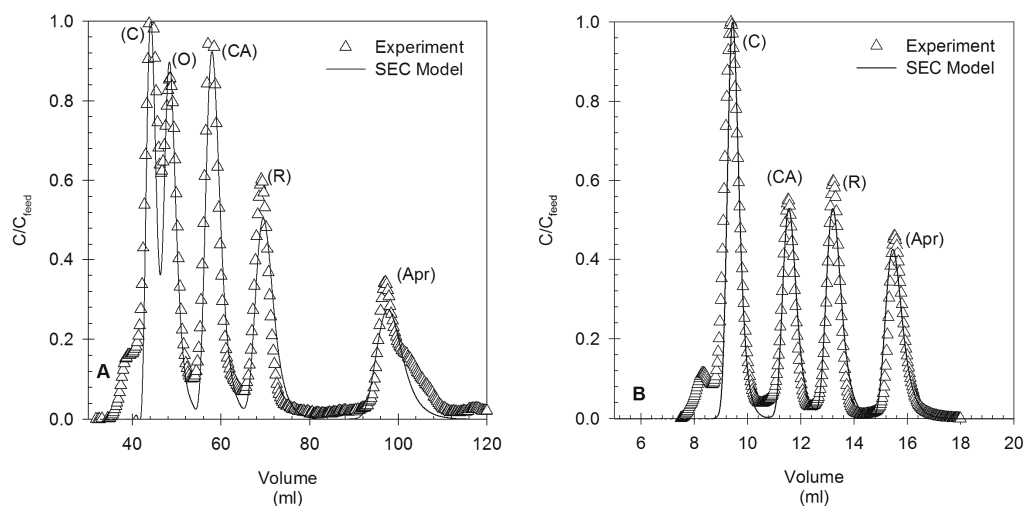


Figure 4.5: Modeled and experimental chromatograms. (A) Chromatograms corresponding to the Sephacryl S100 16/600. (B) Chromatograms corresponding to the Superdex 75 10/300. Protein standards: C: Conalbumin, O: Ovalbumin, CA: Carbonic Anhydrase, R: Ribonuclease A, Apr: Aprotinin

#### 4.4.3 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 4.2.5, 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 4.5. 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. (4.17) 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 obtained by numerical simulations using the settings described in section 4.3.7 and Eqs. (4.1)-(4.4), (4.9a)-(4.9c), (4.10a)-(4.10c), and (4.11a)-(4.11c). 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. (4.8) and (4.7), respectively.

$$\Psi = \frac{Da_I}{Da_{II}} = \frac{\tau_{Agg}}{\tau_{fold}} \quad (4.17)$$

Figure 4.6 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 figure 4.6, 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. Figures 4.7 and 4.8 present the Damköhler map for mechanism 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 (Fig 7-8), 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.4.8). 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 ( $\psi$ , Eq. 4.17) on the refolding yield ( $Y_N$ ) is presented in figure 4.9, for the three reaction mechanisms considered. The simulations show, that the SEC refolding yield ( $Y_N$ ) increases as  $\psi$  increases. This is because as  $\psi$  increases, so does the  $\tau_{Agg}/\tau_{fold}$  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.4.9). 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. 4.9). 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 mechanisms 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 ( $\psi$ ) and its effect on the refolding yield, but what about the role of convection? To answer this question the data presented in table 4.6 is used. Table 4.6 presents those situations where equal values of  $\psi$  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  $\psi$ , mainly as a consequence of the competition between (1) convection ( $\tau_{conv}$ ) and folding ( $\tau_{fold}$ ), and (2) convection and aggregation ( $\tau_{Agg}$ ), rather than the competition between folding and aggregation ( $\psi$ ) alone, as this is constant for the same value of  $\psi$ . 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 4.6). It is important to point out though, from the data in table 4.6, 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.

Table 4.6: 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		0.07	0.03	0.03	$\tau_{fold} > \tau_{conv} > \tau_{Agg}$
100	1	0.01	0.18	0.03	0.04	$\tau_{conv} > \tau_{Agg} \wedge \tau_{conv} = \tau_{fold}$
1	0.1		0.14	0.11	0.11	$\tau_{fold} > \tau_{conv} \wedge \tau_{conv} = \tau_{Agg}$
10	1	0.1	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		0.70	0.53	0.53	$\tau_{conv} > \tau_{fold} > \tau_{Agg}$
200	100	0.5	0.77	0.58	0.58	$\tau_{conv} > \tau_{fold} > \tau_{Agg}$
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	1	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		0.96	0.94	0.94	$\tau_{conv} > \tau_{fold} \wedge \tau_{conv} = \tau_{Agg}$
10	100	10	0.98	0.96	0.96	$\tau_{conv} > \tau_{Agg} > \tau_{fold}$

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 on table 4.6. 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 (Fig 4.6, 4.7, 4.8), provide information about the effect that a given reaction mechanism (e.g., mechanism 1, 2, 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 ( $\tau_{conv}$ ), folding ( $\tau_{fold}$ ) and aggregation ( $\tau_{agg}$ ), has on the SEC refolding yield (Fig 4.9, table 4.6). Together these two information streams comprise the foundations of the 'Systematic analysis' presented in figure 4.3, and they are used to select the suitable mechanism that captures best the behavior of the SEC refolding data.

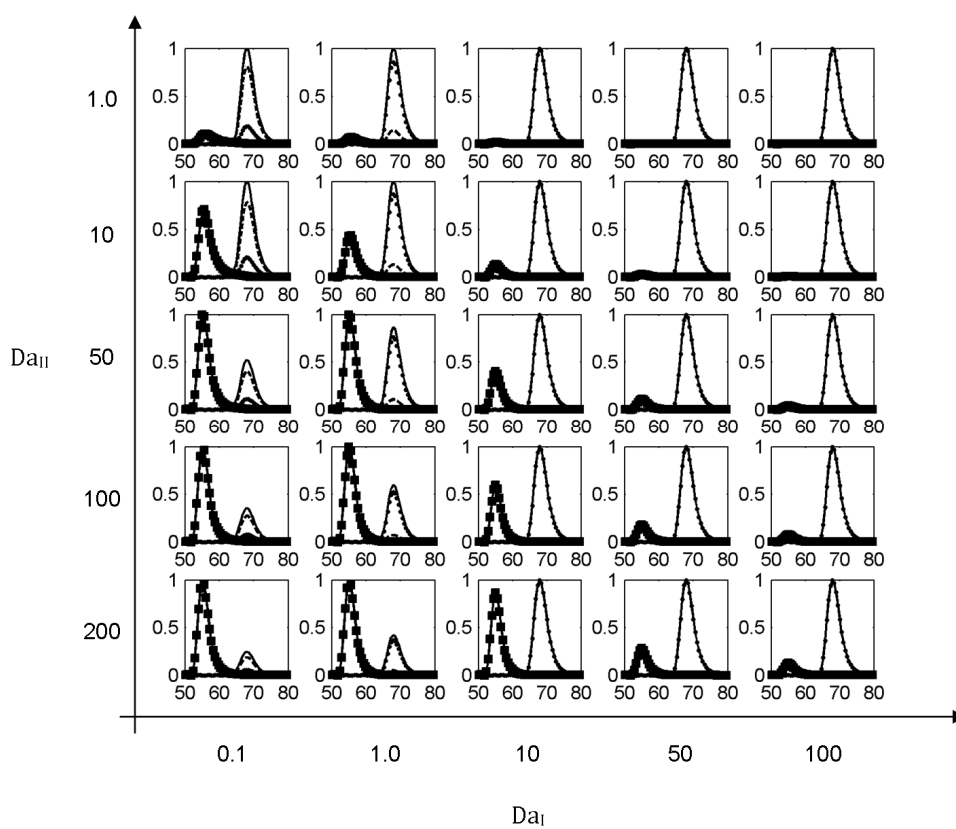


Figure 4.6: Damköhler map for mechanism 1. Solid-line: Total protein, Dots ( $\cdot$ ): Native protein (N), Dash-line: Intermediate (I), ( $\blacksquare$ ): Aggregate class 3 ( $A_3$ ), i.e. trimer. The insets have y-axis  $C/C_{\max}$ , and x-axis volume (ml).  $\psi$  (Eq. 17) increases from left to right and from bottom to top. Simulations done with Eqs. (1)-(4) and Eqs. (9a)-(9c)

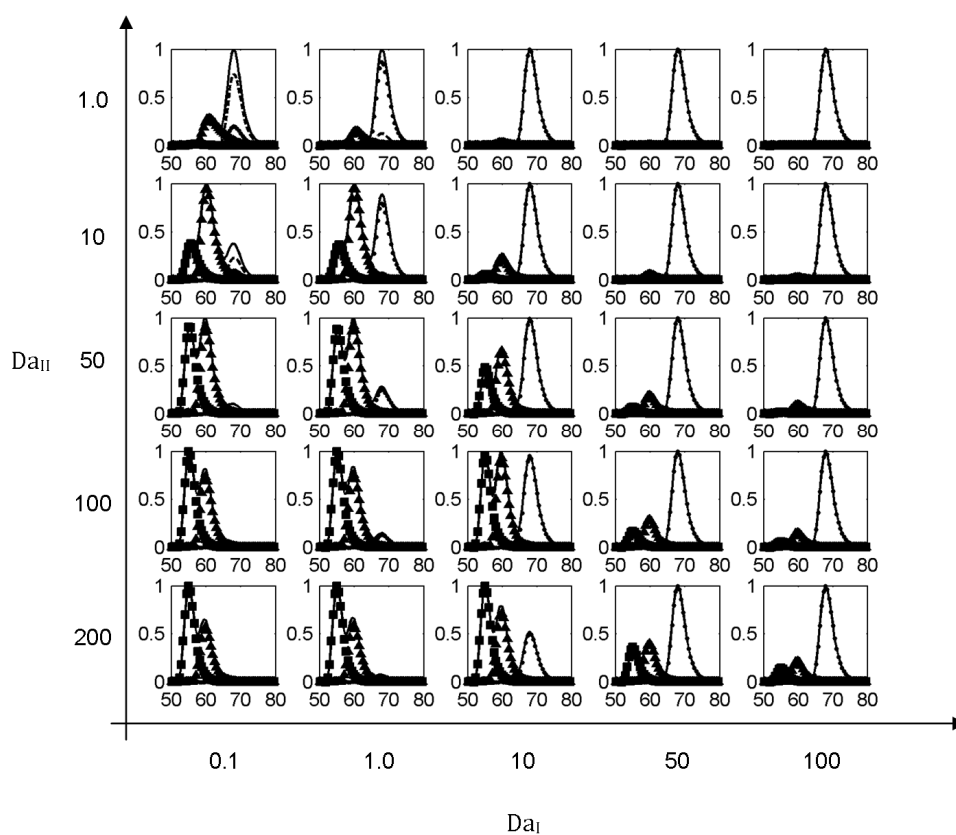


Figure 4.7: Damköhler diagram for mechanism 2. Solid-line: Total protein, Dash-line: Intermediate (I), Dots ( $\cdot$ ): Native (N), ( $\blacktriangle$ ): Aggregate class 2, i.e. dimer ( $A_2$ ), ( $\blacksquare$ ): Aggregate class 3, i.e. trimer ( $A_3$ ). The insets have y-axis  $C/C_{\max}$ , and x-axis volume (ml).  $\psi$  (Eq. 4.17) increases from left to right and from bottom to top. Simulations done with Eqs. (4.1)-(4.4) and Eqs. (4.10a)-(4.10e)



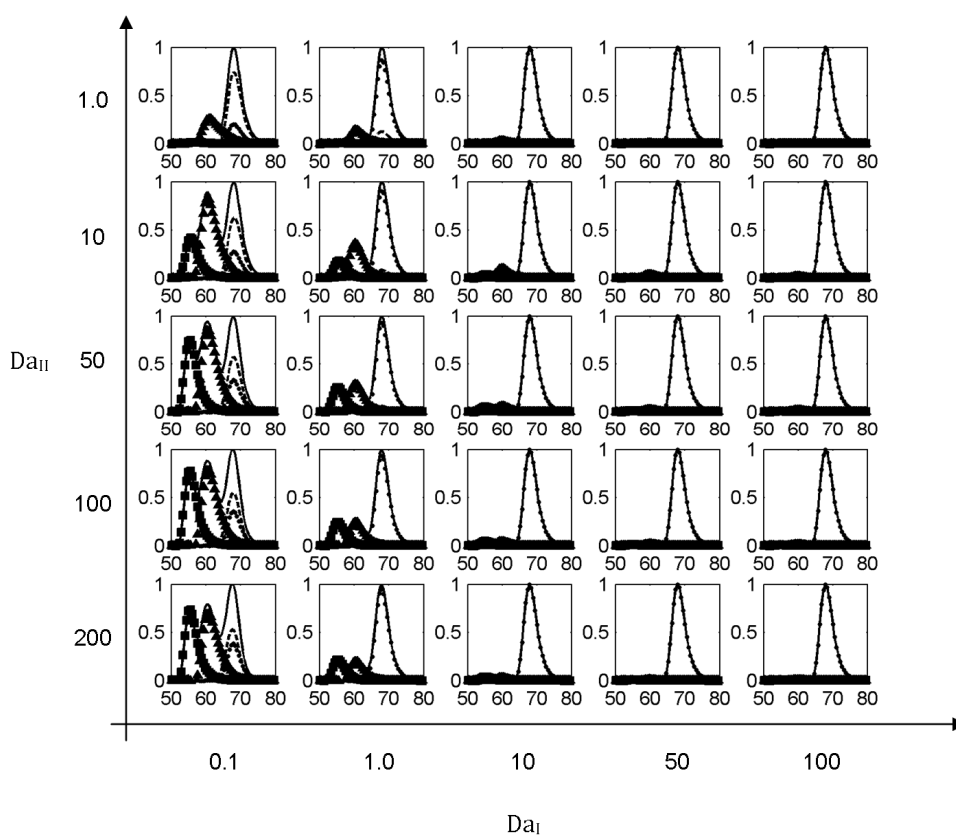


Figure 4.8: Damköhler map for mechanism 3. Solid-line: Total protein. Dots ( $\cdot$ ): Native protein (N), Dash-line: Intermediate (I), ( $\blacktriangle$ ): Aggregate class 2, i.e. dimer ( $A_2$ ), ( $\blacksquare$ ): Aggregate class 3, i.e. trimer ( $A_3$ ). The insets have y-axis  $C/C_{\max}$ , and x-axis volume (ml).  $\psi$  (Eq. 4.17) increases from left to right and from bottom to top. Simulations done with Eqs. (4.1)-(4.4) and Eqs. (4.11a)-(4.11c)

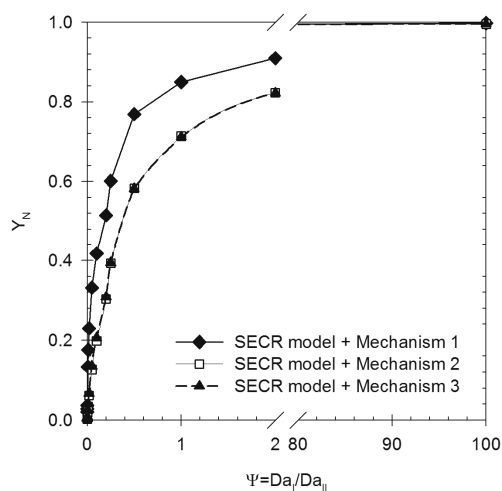


Figure 4.9 Effect of the competition between folding and aggregation ( $\psi$ ) on the SEC refolding yield ( $Y_N$ ).

#### 4.4.4 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. 4.3), a suitable reaction mechanism that describes the experimental SEC refolding data best.

Figure 4.11A 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 ( $V_e \sim 35.40$  ml) and 2 ( $V_e \sim 47.34$  ml) are constituted by high molecular weight aggregates, and peak 3 ( $V_e \sim 55.40$  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 4.3), with the aid of the information extracted from the experimental SEC refolding data (Fig 4.11A). From the SEC experimental data the following observations were made: (1) The presence of two peaks (i.e., peak 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 figure 4.10. 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 peak 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 4.3.9), during the ‘Systematic analysis’ phase of the rational strategy (Fig 4.3). Figure 4.11A 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  $k_1=1.03\cdot 10^{-4}\pm 9.7\cdot 10^{-5} \text{ s}^{-1}$ ,  $k_2=0.08\pm 8\cdot 10^{-3} \text{ s}^{-1}$ ,  $k_3=1.69\pm 8.2\cdot 10^{-2} \text{ l mmol}^{-1} \text{ s}^{-1}$  and  $k_4=2.42\pm 0.19 \text{ l mmol}^{-1} \text{ s}^{-1}$ . These set of kinetic parameters complete the refolding yield function describing the SEC refolding reactor (‘Output phase’ fig 4.3). 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. Figure 4.11B 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.

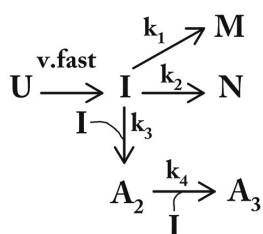


Figure 4.10: Mechanism describing the competition between folding, misfolding and aggregation, during SEC refolding of the model fusion protein. U: Unfolded protein, I: Intermediate, M: misfolded protein, N: Native protein,  $A_2$ : Aggregate class 2,  $A_3$ : Aggregate class 3.  $k_1$ ,  $k_2$ ,  $k_3$ , and  $k_4$  are the reaction kinetic constants

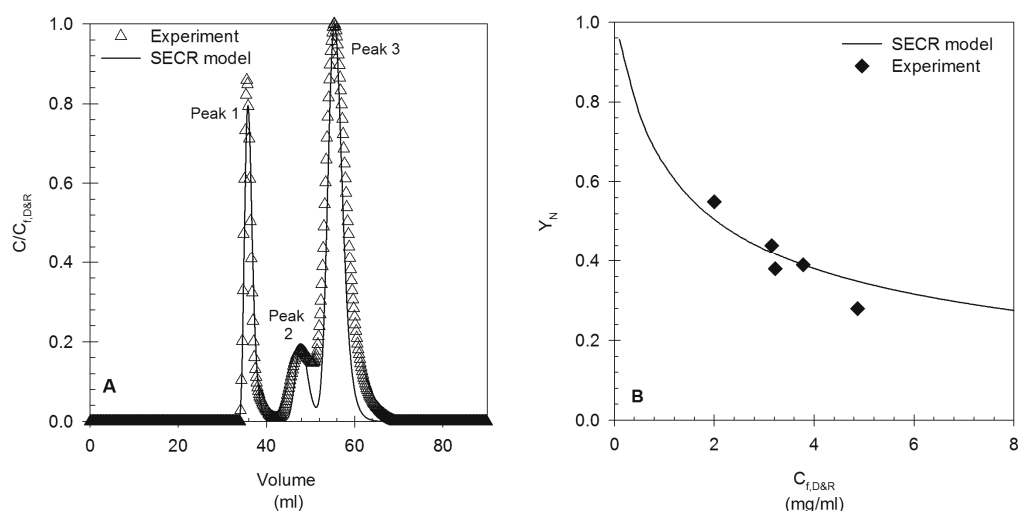


Figure 4.11 Application of the model based strategy. (A) Experimental and modeled SEC refolding chromatograms. The simulated SEC refolding chromatogram was obtained with Eqs. (4.1)-(4.4b) and the reaction mechanism presented in figure 4.10. (B) Experimental and predicted SEC refolding yield ( $Y_N$ ). The modeled SEC refolding yield was obtained using Eqs (4.1)-(4.4b) and the reaction mechanism presented in figure 4.10

#### 4.4.5 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 ( $C_{i,D\&R}$ ). 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 refolding 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  $C_{i,D\&R}$  ranging from 0.01 to 10  $\text{mgml}^{-1}$  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  $1\text{mlmin}^{-1}$  and an injection volume of 1ml.

Figure 4.12 and 4.13 present the refolding yield contour plots as a function of  $C_{i,D\&R}$  and the kinetic competition between aggregation and folding, for mechanism 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. Figures 4.12A-4.13A and 4.12B-4.13B 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 chemical 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

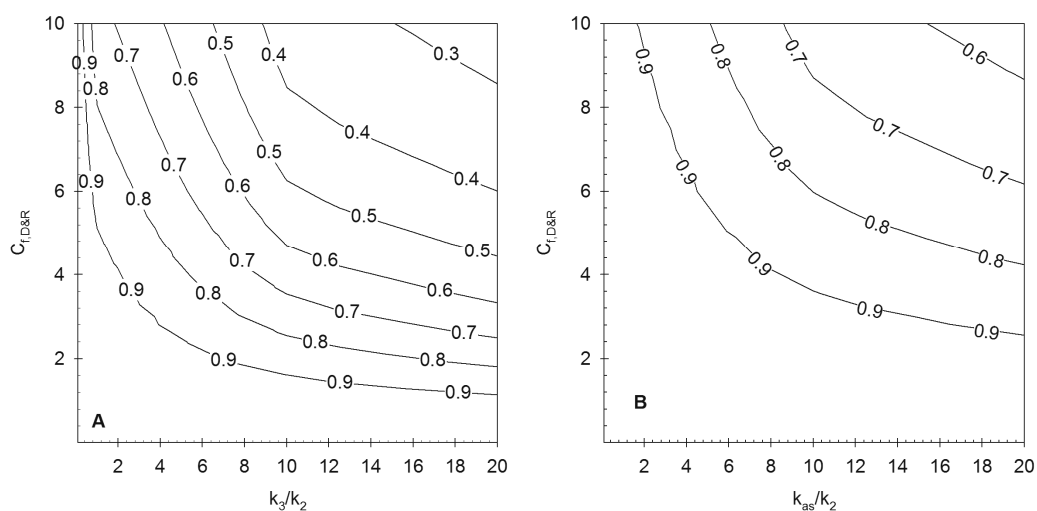


Figure 4.12 contour plots for the refolding yield ( $Y_N$ ) of mechanism 1, as a function of  $C_{f,D\&R}$  and the kinetic competition between aggregation and folding ( $k_3/k_2$  or  $k_{as}/k_2$ ). (A) Batch refolding reactor. (B) Size-exclusion refolding reactor. The contour lines represent the refolding yield

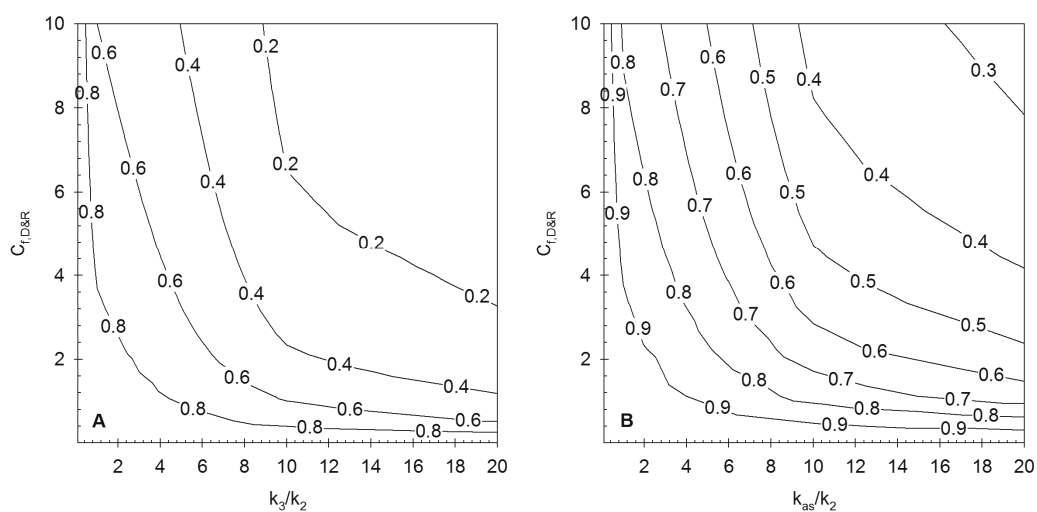


Figure 4.13 Contour plots for the refolding yield ( $Y_N$ ) of mechanism 2, as a function of  $C_{f,D\&R}$  and the kinetic competition between aggregation and folding ( $k_3/k_2$  or  $k_{as}/k_2$ ). (A) Batch refolding reactor. (B) Size-exclusion refolding reactor. The contour lines represent the refolding yield

#### 4.5 Further discussion

The presented methodology has been built 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. 4.8) in order to estimate the solute dimensions and thereafter its distribution coefficient (Eq. 4.7), and (2) Determine the distribution coefficient from the measured elution volume (Eq. 4.15), 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.

## 4.6 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 ( $\tau_{fold}$ ), aggregation ( $\tau_{Agg}$ ) and convection ( $\tau_{conv}$ ) 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  $C_{i,D\&R}$  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  $C_{i,D\&R}$  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.



#### **4.7 Acknowledgements**

The authors will 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

#### 4.8 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. Expres. 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. A.* 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. Neseck, *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. A.* 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.F.W. IV, 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, *Separations Technology* 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 (Editor), *Methods in molecular Biology*, Totowa;Nj, 2006, p. 97.
- [59] H. Ye, *Anal. Biochem.* 356 (2006) 76

# 5

## Size-Exclusion Simulated Moving Bed Chromatographic Protein Refolding

---

### Abstract

Size-exclusion chromatographic refolding (SECR) has successfully proven its capability to refold a variety of proteins using a range of gel filtration column materials. Several approaches have also been undertaken to improve the refolding yield of these systems, mostly under batch operation. Although, these approaches may lead to an increase on refolding yield, it is not expected that they will lead to significant increases in other important process indicators, such as volumetric productivity, specific eluent consumption and product concentration in the product stream; as these indicators are strongly dependent on the mode of operation. To overcome the shortcomings of batch chromatography, the size-exclusion refolding reactor may be operated in a continuous mode, with the aid of Simulated Moving Bed (SMB) technology. Albeit, SMB technology has inherent advantages over batch chromatography, these have been proven mainly in the context of conventional purifications and are still to be addressed in the context of chromatographic refolding reactors. In this work we report the on-column refolding of an industrially relevant protein, produced in inclusion bodies, by batch size-exclusion chromatography (SEC) and simulated moving bed size-exclusion chromatography (SMBSEC). The presented study encompasses: (1) a statistical design of experiments (DOE) to study the combined effect of the mobile phase pH (9.0-11.20) and the feed concentration of denatured and reduced protein ( $C_{f,D\&R}$  2.50-7.50mgml<sup>-1</sup>) on the refolding yield of the model protein, (2) a mechanistic analysis of the SMBSECR data, using a detailed model that accounts for both separation and refolding, and (3) a detailed comparison of the SMBSECR against the SECR, based both on quantitative and qualitative criteria. Our work showed that: (1) refolding yields of 50% are attainable by tuning pH and  $C_{f,D\&R}$ , and that the positive effect of pH is strongly dependent on the  $C_{f,D\&R}$ ; (2) the modeling tool captured well the SMBSECR behavior, based solely on the effect that  $C_{f,D\&R}$  has on the reaction rates; and (3) the volumetric productivity of the SMBSEC is about 53 times higher than that of the SECR, the specific solvent consumption is approximately 1/10<sup>th</sup> of that of the SECR, and the concentration of the product (i.e. native protein), leaving the SMBSECR, is roughly 4.5 fold higher than the one leaving the SECR. Accordingly, the comparison revealed the advantages that SMB technology has to offer to the design of chromatographic refolding reactors.

Keywords: Simulated moving bed; protein refolding; size-exclusion chromatography; modeling.

## 5.1 Introduction

The deep genetic and physiological characterization, ease of handling and well established fermentation processes, are among the reasons why, the production of recombinant pharmaceuticals using Prokaryotic systems such as *Escherichia Coli* (*E. coli*), remain a process of choice for the biopharmaceutical industry. Examples of commercial biotherapeutics, produced using *Escherichia Coli*, include: tumor necrosis factor, recombinant plasminogen-activator, Interferon alpha, Insulin, Growth factors (GCSF, GMCSF), among others (Schmidt 2004). Often, the over-expression of the protein product in *E. coli* is accompanied by the concomitant formation of inclusion bodies (IBs). Inclusion bodies are basically insoluble refractile bodies that accumulate inside the cell, and contain mostly the over-expressed protein product in an aggregated, inactive form. To obtain an active (i.e., native) soluble protein the IBs are isolated, solubilized and finally the soluble protein is refolded (renatured). IBs solubilization is regularly done using a denaturant (e.g., urea, guanidine HCL), a reducing agent (e.g., dithiothreitol (DTT), $\beta$ -mercaptoethanol), and alkaline pH. This step yields a soluble denatured and reduced (D&R) protein. To refold the soluble protein, the concentration of the solubilization additives must be decreased, this can be attained by either dilution or by buffer exchange using liquid chromatography (chromatographic refolding).

Chromatographic refolding has been a topic of research for at least three decades. It's well accepted advantages include: (1) easy automation using commercially available systems, and (2) a relatively high degree of process intensification, owed to the possibility of performing a reaction and separation in one single unit operation. Various model proteins have been successfully refolded using various types of chromatography, including: size-exclusion chromatography (SEC) (Werner, Clore et al. 1994; Batas and Chaudhuri 1996), hydrophobic interaction chromatography (HIC) (Geng and chang 1992; Bai, Kong et al. 2003), ion-exchange chromatography (IEC) (Creighton 1986; Machold, Schlegl et al. 2005), and immobilized metal affinity chromatography (IMAC) (Zahn, von Schroetter et al. 1997; Lemercier, Bakalara et al. 2003). Among the various types of chromatography used for on-column refolding, size-exclusion offers various advantages, which include: (1) relatively economic column material, (2) the unwanted interactions with the column material are negligible, as compared to adsorption refolding, and (3) in principle it can be applied to any protein, which is not the case for IMAC refolding.

Size-exclusion chromatographic refolding works, basically, based on a buffer exchange mechanism. This mechanism takes place as a result of the large differences in distribution

coefficients, between the small solutes (i.e., denaturant and reducing agent) and the denatured and reduced protein. As the concentration waves progress through the column, they separate. Accordingly, the local concentration of the solubilization additives decreases, which induces protein folding and disulfide bond formation. SEC refolding has been shown to be capable of working with feed concentration of denatured and reduced protein, ranging from  $0.5 \text{ mgml}^{-1}$  (Shalongo, Ledger et al. 1987), up to  $80 \text{ mgml}^{-1}$  (Batas and Chaudhuri 1996). Most likely, as a result of this flexibility, SEC refolding has been studied for different model proteins, and using a variety of gel filtration materials (Jungbauer, Kaar et al. 2004). Among the various studies, a common objective has been to optimize the SEC refolding yield. The different strategies tested include: (1) influence of the gel filtration material (Fahey, Chaudhuri et al. 2000; Gu, Zhu et al. 2003), (2) the introduction of a urea gradient, or a pH and urea gradient (Gu, Su et al. 2001; Gu, Weidenhaupt et al. 2002), and (3) the introduction of an artificial chaperone system (Dong, Wang et al. 2002). Although, these different strategies may lead to increases in refolding yield, it is not expected to have significant improvements on process performance indicators such as: specific eluent consumption, volumetric productivity, and product concentration after refolding; as these indicators are strongly dependent on the mode of operation, which has been for most studies batch chromatography.

Size-exclusion batch refolding, suffers of the same shortcomings as conventional batch SEC. By conventional it is meant SEC use solely for purification. These shortcomings include: (1) long columns, and (2) large product dilution, owed to a large column volume to loading volume ratio. These shortcomings have several consequences, which include: large column material requirement, long processing times, and large consumption of eluent(s). Presumably, own to the awareness of the aforementioned limitations, more recent studies have been focused on size-exclusion continuous chromatographic refolding. Continuous size-exclusion refolding has been demonstrated using preparative continuous annular chromatography (P-CAC) for  $\alpha$ -lactalbumin (Schlegl, Iberer et al. 2003) and lysozyme (Lanckriet and Middelberg 2004), and simulated moving bed (SMB) chromatography for lysozyme (Park, Lee et al. 2005; Park, Lee et al. 2006)

Simulated moving bed technology is a well established technology, which dates back to the 1960s (Broughton and Gerhold 1961; Broughton 1968), and was originally design for the petrochemical and sugar industries. The application of SMB technology has been already demonstrated for the purification of biomolecules. Examples include: insulin (Xie, Mun et al. 2002), human bone morphogenetic protein-2 (Gueorguieva, Vallejo et al. 2006),  $\alpha$ -

lactalbumin and  $\beta$ -lactoglobulin (Lucena, Rosa et al. 2002), immunoglobulin G (Horneman, Ottens et al. 2007), and viral clearance (Horneman, Ottens et al. 2007). SMB technology has inherent advantages over conventional batch chromatography. Such advantages include: less column material requirement, less product dilution and lower eluent consumption. These advantages are more pronounced when SMB is applied to size-exclusion chromatography, as has already been shown in the context of conventional separations (Horneman, Ottens et al. 2007). However, this still needs to be address in the context of size-exclusion chromatographic refolding reactors.

In this work the size-exclusion chromatographic refolding (SECR) of an industrially relevant protein, produced in IBs, is presented. On column refolding was done using batch SEC and simulated moving bed SEC (SMBSEC), utilizing as the gel filtration material Sephacryl S-100. A statistical design of experiments was used to (1) optimize the operational parameters, pH of the mobile phase and feed concentration of denatured and reduced protein, and (2) to study the combined effect of these parameters on the refolding yield of the model protein. A column model was then used in combination with size-exclusion refolding data, to estimate a competitive reaction mechanism and its kinetics parameters. This reaction mechanism was coupled to the single column model to build a modeling tool, which was then used to analyze the size-exclusion simulated moving bed refolding (SMBSECR) data. Finally, a detailed comparison between the two chromatographic refolding reactors is given. This comparison was based on quantitative (e.g., refolding yield, volumetric productivity, etc), as well as qualitative criteria. Our work shows, that (1) using statistical analysis we were able to determine when, the positive effect that pH has on refolding yield, is masked by the effect of the feed concentration of denatured and reduced protein, (2) the modeling tool, developed to analyze the SMBSECR data is well suitable for the data analysis of this type of refolding reactors and thus valuable conclusions can be drawn with its aid, and (3) the comparison of the SMBSECR against the SECR, show the clear advantages that the SMBSECR offers, together with its limitations. The latter issue, opened the possibility to finalize the discussion with what would be the next step on the development of SMBSEC refolding reactors.

## 5.2 Theory and modeling

### 5.2.1 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 (Lapidus and Amundson 1952; Guiochon, Shirazi et al. 2006). This model is described by the two differential as shown below

$$\frac{\partial C_{b,i}}{\partial t} = D_{L,i} \frac{\partial^2 C_{b,i}}{\partial x^2} - u \frac{\partial C_{b,i}}{\partial x} - P k_{ov,i} (C_{EqS,i} - C_{S,i}) + r_{b,i} \quad (5.1)$$

$$\frac{\partial C_{S,i}}{\partial t} = k_{ov,i} (C_{EqS,i} - C_{S,i}) + r_{S,i} \quad (5.2)$$

Where  $t$  is time,  $x$  is the axial coordinate,  $D_{L,i}$  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. In order to solve the two column equations, the following initial and boundary conditions are required:

$$x = 0 \quad \frac{\partial C_{b,i}}{\partial x} = \frac{u}{D_{L,i}} (C_{b,i} - C_{f,i}(t)) \quad (5.3a)$$

$$x = L_c \quad \frac{\partial C_{b,i}}{\partial x} = 0 \quad (5.3b)$$

$$t = 0 \quad C_{b,i}(0, 0 < x < L_c) = 0 \quad (5.3c)$$

$$t = 0 \quad C_{S,i}(0, 0 < x < L_c) = 0 \quad (5.3d)$$

where  $L_c$  is the column length. The operation of the column, unless otherwise specified, was model as a loading-elution operation, which is represented as follows:

$$C_{f,i}(t) = 0 \quad \text{for } t_{\text{pulse}} < t \quad (5.4a)$$

$$C_{f,i}(t) = C_{\text{feed},i} \quad \text{for } t < t_{\text{pulse}} \quad (5.4b)$$

where  $t_{\text{pulse}}$  is the feed duration, determined by the volume and flow rate of injection.  $C_{\text{feed},i}$  is the feed concentration for component  $i$ . Eqs (5.4a) and (5.4b) represent the elution and loading, respectively.



Eqs. (5.1)-(5.4) are solved for each component present in the feed or formed during a given reaction scheme. The partial differential equation was solved using the method of lines (MOL) (Schiesser 1991). The second and first order derivatives, with respect to space, were discretized using a fourth-order finite difference equation thereby transforming the PDE into a system of ordinary differential equations.

### 5.2.2 Model parameters and correlations

Mass transfer parameters include free diffusivities, effective diffusivities, among others. These parameters are ultimately used to calculate the overall mass transfer coefficient. All parameters can be estimated using well-established correlations (Table 5.1). The correlations used have successfully being applied for the modeling of size-exclusion band profiles (Li, Gu et al. 1998; Yu, Mun et al. 2006; Zelic and Neseek 2006), supporting their usage in this work.

Table 5.1: Model parameters and correlations

Mass transfer parameter	Parameter	Correlation	Reference
Free diffusivity	$D_{f,Proteins}$	Young	(Guiochon, Shirazi et al.
	$D_{f,Urea}$	Wilke-Chang	2006)
Pore diffusivity	$D_p$	Satterfield	(Li, Gu et al. 1998)
Film mass transfer coefficient	$k_{f,i}$	Wilson-Geankoplis	(Wilson and Geankoplis 1966)
Overall mass transfer coefficient	$k_{ov,i}$	Glueckauf	(Ma and Wang 1997)
Axial Péclet	$Pe_L$	Chung-Wen	(Chung and Wen 1968)

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

$$D_L = \frac{uL_c}{Pe_L} \quad (5.5)$$

### 5.2.3 Simulated moving bed model

A simulated moving bed unit consists basically of a certain number of columns that are connected in series. A schematic representation of such arrangement is presented in Fig. 5.1. To model the coupling of the columns a periodic boundary condition(s) is required. The

periodic boundary conditions are presented in Eqs. (6a)-(6d). They were derived from the mass balances around the external nodes, represented in figure 1 as Desorbent ( $D$ ), Extract ( $E$ ), Feed ( $F$ ) and Raffinate ( $R$ ).

$$C_{f_i}(t, L_1^+) = \frac{D}{Q_I} C_{D,i} \quad (5.6a)$$

$$C_{f_i}(t, L_2^+) = C_{f_i}(t, L_1^-) \quad (5.6b)$$

$$C_{f_i}(t, L_3^+) = \frac{FC_{feed,i} + Q_{II}C_{f_i}(t, L_2^-)}{Q_{III}} \quad (5.6c)$$

$$C_{f_i}(t, L_4^+) = C_{f_i}(t, L_3^-) \quad (5.6d)$$

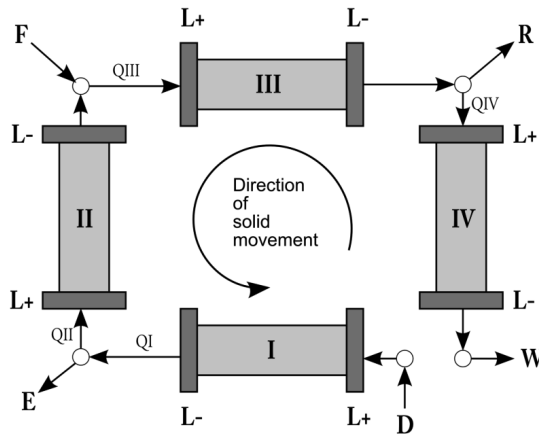


Figure 5.1: Schematic representation of a four columns, four zones open-loop SMB unit.  $L^+$  and  $L^-$  represent the entry and exit of a column, respectively.  $Q_{I...IV}$  represent the zone flow rates.  $D, E, F, R, W$  represent the external or pump flow rates Desorbent, Extract, Feed, Raffinate, and Waste

The last important step in the construction of the SMB model is the incorporation of the port switching. This feature of the model represents the counter-current movement of the packed beds, as occurs in the SMB unit. Port switching was implemented in the model as a periodic movement of the concentration vector. This means, the final state of the state variables is moved, by a step size of one column length, in the negative direction of the fluid flow at the end of the switching interval.

#### 5.2.4 Size-exclusion distribution coefficient

The size exclusion distribution coefficient ( $K_{av,i}$ ) was calculated using

$$K_{av,i} = \frac{V_{e,i} - V_0}{V_i - V_0} \quad (5.7)$$

where  $V_{e,i}$  is the elution volume, determined for example from the peak apex.  $V_0$  represents the inter-particle void volume and  $V_t$  the total void volume. The inter-particle void volume was determined by pulse experiments using blue dextran. The total void volume was determined using acetone. These volumes were corrected by subtracting the contributions of the system volumes.

### 5.2.5 Refolding mechanism

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. As a consequence, the refolding yield is clearly a function of the competition of these reaction rates. Of the two competing side reactions, often the aggregation reaction(s) are the ones affecting significantly the refolding yield. Hence, most research has been targeted towards preventing them (Clark, Schwarz et al. 1999; Clark 2001).

Conventionally the folding versus aggregation has been represented as a competition between a first-order refolding and a higher-order aggregation (Kiefhaber, Rudolph et al. 1991). This competitive reaction scheme gained popularity through the years and it has been applied, with slight modifications, to study for example the refolding of lysozyme in batch (Hevehan and Clark 1997; Maachupalli-Reddy, Kelley et al. 1997; Buswell and Middelberg 2002) and fed-batch (Dong, Shi et al. 2004) systems. From a mechanistic point of view, representing protein aggregation as a higher-order reaction may, in some cases, not be adequate. Recent work on protein refolding has shown the successful application of polymerization mechanisms to describe the protein aggregation reaction(s) competing with refolding (Speed, King et al. 1996; Buswell and Middelberg 2003). One of these mechanisms is the so-called sequential polymerization, which has been employed to study the aggregation of several proteins, e.g., sickle cell hemoglobin, actin and P22 tailspike (Ferrone, Hofrichter et al. 1980; Tobacman and Korn 1983; Speed, King et al. 1996).

In this work a hybrid mechanism was constructed and is presented in Fig. 5.2. This mechanism considers the formation of misfolded species and native product as first order reactions. The competing aggregation is captured by a sequential polymerization mechanism that accounts for the formation of aggregates of different size-classes. The competitive scheme of Fig. 5.2 is described by Eqs. (5.8a)-(5.8e). This set of equations was incorporated to the column model (i.e. Eqs. (5.1) and (5.2)) as the source terms, coupling the separation and the refolding.

$$r_{j,M} = k_1 C_I \quad (5.8a)$$

$$r_{j,N} = k_2 C_I \quad (5.8b)$$

$$r_{j,I} = -k_1 C_{j,I} - k_2 C_{j,I} - 2k_3 (C_{j,I})^2 - k_4 C_{j,I} C_{j,A_2} \quad (5.8c)$$

$$r_{j,A_2} = k_3 (C_{j,I})^2 - k_4 C_{j,I} C_{j,A_2} \quad (5.8d)$$

$$r_{j,A_3} = k_4 C_{j,I} C_{j,A_2} \quad (5.8e)$$

where  $r_{j,M...A_3}$  are the net source terms of phase  $j$  (i.e., bulk or solid) for misfolded ( $M$ ), native ( $N$ ), intermediate ( $I$ ), aggregate class 2 ( $A_2$ ) and aggregate class 3 ( $A_3$ ).  $k_{1...4}$  are the reaction kinetic constants.  $C_{j,I}$  represent the intermediate concentration in phase  $j$ .  $C_{j,A_2}$  represent the concentration of aggregate class 2 in phase  $j$ .

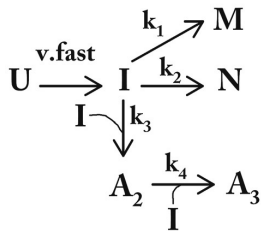


Figure 5.2. Schematic representation of the kinetic model describing the competition between refolding, misfolding and aggregation.  $U$ , unfolded protein (denatured and reduced);  $I$ , transition intermediate;  $A_2$  aggregate class size 2 (i.e., dimer);  $A_3$ , aggregate class size 3 (i.e., trimer)

## 5.2.6 Performance criteria

### 5.2.6.1 Refolding yield

The refolding yield is defined as the amount of active product formed per amount of denatured and reduced protein loaded to the size-exclusion reactor. Eqs (5.9a)-(5.9c) present the definition of the experimental and modeled refolding yield

$$Y_{N,SECR,exp} = M_N (C_{f,D\&R} V_{inj})^{-1} \quad (5.9a)$$

$$Y_{N,SMBSECR,exp} = C_{N,Raff}^{CSS} R(FC_{f,D\&R})^{-1} \quad (5.9b)$$

$$Y_{N,SMBSECR,model} = \bar{C}_{N,Raff}^{CSS} R(FC_{f,D\&R})^{-1} \quad (5.9c)$$

where  $Y_{N,SECR,exp}$  represent the experimental (exp) refolding yield for the size-exclusion reactor (SECR).  $Y_{N,SMBSECR}$  represent the modeled (mod) or experimental (exp) refolding yield for the simulated moving bed size-exclusion reactor (SMBSECR).  $M_N$  is the mass of native protein.  $C_{f,D\&R}$  is the feed concentration of denatured and reduced (D&R) protein.  $C_{N,Raff}^{CSS}$  is the concentration of native protein, in the raffinate stream, at cyclic steady state

(CSS).  $\bar{C}_{N,Raff}^{CSS}$  is the average concentration of native protein, in the raffinate stream, at cyclic steady state.  $V_{inj}$  is the injection volume.  $R$  and  $F$  represent the raffinate and feed streams from the SMB.

### 5.2.6.2 Volumetric productivity

The volumetric productivity of a chromatographic reactor is defined as the amount of product (i.e. Native protein) produced per unit time per volume of chromatographic medium.

$$Pr_{SECR} = \frac{C_{f,D\&R} V_{inj}}{t_{cycle} V_C} Y_{N,SECR} \quad (5.10a)$$

$$Pr_{SMBSECR} = \frac{FC_{f,D\&R}}{N_{Col} V_C} Y_{N,SMBSECR} \quad (5.10b)$$

where  $N_{Col}$  represent the number of columns on the SMB unit.  $V_C$  is the packed bed volume of a single column.  $t_{cycle}$  represent the cycle time of the batch column.  $Pr_{SECR}$  and  $Pr_{SMBSECR}$ , represent the volumetric productivity of the SECR and the SMBSECR, respectively.

### 5.2.6.3 Specific eluent consumption

The specific eluent consumption of a chromatographic reactor is defined as the volume of eluent used per amount of product produced.

$$Ec_{SECR} = \frac{\dot{Q} t_{cycle}}{V_{inj} C_{f,D\&R} Y_{N,SECR}} \quad (5.11a)$$

$$Ec_{SMBSECR} = \frac{D}{FC_{f,D\&R} Y_{N,SMBSECR}} \quad (5.11b)$$

where  $\dot{Q}$  represent the flow rate used in the batch SEC experiment.  $Ec_{SECR}$  and  $Ec_{SMBSECR}$ , represent the specific eluent consumption for the SECR and the SMBSECR, respectively.

### 5.3 Materials and Methods

#### 5.3.1 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).

Size-exclusion batch experiments were done using a HiPrep 16/60 (i.e. 1.6 cm i.d. & 60cm length) column, pre-packed with Sephacryl S-100 high resolution gel filtration material. The column was obtained from GE-Healthcare (Uppsala, Sweden). All chromatographic batch separations were performed using an ÄKTA explorer 10, equipped with the UNICORN software version 5.01 from GE-Healthcare (Uppsala, Sweden). Simulated moving bed size-exclusion was done using omnifit glass columns of 1.0cm i.d, purchased from Alltech Nederland BV (Breda, 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  $\mu$ m pore size membrane filter from Pall (Portsmouth, Hampshire, United Kingdom).

#### 5.3.2 Protein quantification

The concentration of soluble protein was estimated using the BCA protein assay purchased from Fisher Scientific (Landsmeer, The Netherlands). The protein concentration was calculated from the calibration line built using bovine serum albumin as the standard.

#### 5.3.3 Quantification of the native fusion protein

The refolded fusion protein was digested using trypsin in order to obtain the mature monomer. The digestion was done using an enzyme to substrate ratio, of 1:300 (mg:mg) (Winter, Neubauer et al. 2000; Winter, Lillie et al. 2002). The samples were incubated for 30 min at 25 °C, using a thermomixer comfort from Eppendorf (Amsterdam, The Netherlands). The reaction was stopped by diluting the samples with a 100mM HCL solution. The

concentration of active protein was determined using reversed phase HPLC (RPHPLC) and a calibration line constructed using human insulin as standard.

#### 5.3.4 Inclusion bodies solubilization

The inclusion bodies were solubilized in 4M Urea/ 25mM DTT/ 10mM NaHCO<sub>3</sub>/ 0.1mM EDTA, pH 10.5 (Freydell, Ottens et al. 2007). Solubilization was less than 3 minutes. Solutions of various concentrations were prepared by diluting the concentrated protein solution using the solubilization buffer. This protocol yields 75% denatured and reduced monomer and the remainder percentage as high molecular weight aggregates.

#### 5.3.5 Analytical size-exclusion chromatography

Analytical size-exclusion was used to assess the composition of the outlet streams (i.e. raffinate, extract and waste) of the SMB. The analysis was done using a Superdex-75 10/300 GL, at a flow rate of 1ml/min and an injection size of 0.2ml. Aprotinin (6500), Ribonuclease A (13 700), Carbonic anhydrase (29 000) and Conalbumin (75 000) were used to calibrate the column.

#### 5.3.6 Batch size-exclusion refolding

Size-exclusion refolding was done using a Sephacryl S-100 HR HiPrep 16/60 (1.6cm i.d. x 60cm) column. Sephacryl S-100 has a fractionation range of 1 000-100 000 Da, an average pore diameter of 6.60 nm (Hagel, Ostberg et al. 1996), and average particle diameter of 47  $\mu$ m (Hagel 1998). The porosities of the column (i.e. bed and total porosities) were determined using sodium chloride and blue dextran pulses. The porosities determined were a total porosity ( $\epsilon_t$ ) of 0.88 and a bed porosity of 0.29. These porosities indicate a phase ratio ( $P$ ) of 2.04. The phase ratio was calculated as  $P = (\epsilon_t - \epsilon_b)/\epsilon_b$ . The column was calibrated using Aldolase (MW 158 000), Conalbumin (MW 75 000), Carbonic Anhydrase (MW 29 000) and Ribonuclease (MW 13 700). All experiments were done on an ÄKTA explorer 10.

Prior to the injection of the protein pulse the column was equilibrated using a mobile phase composition of 50mM Tris /10mM NaHCO<sub>3</sub>/0.1mM EDTA. The pH of the mobile phase was adjusted according to the experiment being conducted. The pH was varied on the range of 9-11.2. After equilibration, a pulse of the denatured and reduced protein of 1.2 ml (~1%CV, 1CV=120.6ml) was injected to the column. The protein concentration in the feed pulse was varied in the range 2.5-7.5 mgml<sup>-1</sup>. The flow rate for all SEC refolding experiments was fixed at 1mlmin<sup>-1</sup>.

#### 5.3.7 Simulated moving bed size-exclusion refolding

SMB experiments were done using a custom made carousel SMB set-up, equipped with a rotating valve to generate the clock-wise movement of the columns (Figure 5.3). The set-up

has four LC-8A pumps from Shimadzu, to pump in the feed (F) and the desorbent (D) and the pump out the extract (E) and raffinate (R) streams. Two UV-vis SPD-10AV detectors from Shimadzu were placed in the raffinate and waste outlets to monitor the protein. A refractive index (RI) detector 1200-series from Agilent, was placed at the extract outlet to monitor the urea.

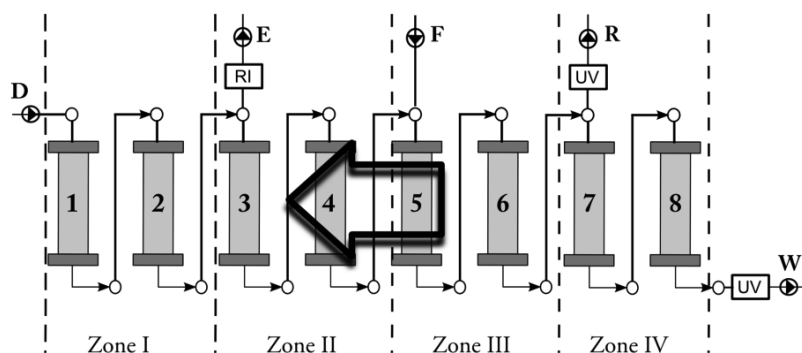


Figure 5.3: Schematic representation of the SMB experimental set-up. Columns are switched from right to left, according to the representation. *D*, *E*, *F*, *R*, *W* represent the desorbent, extract, feed, raffinate, and waste streams, respectively. RI: Refractive index detector, to monitor the urea signal. UV: UV-vis detector, to monitor the protein signal

The SMB was operated in a 2-2-2-2 (i.e., 2 columns per zone) open-loop configuration. This configuration results in a total of 8 columns and four-zones. The columns used in the SMB were Omnifit glass columns of 1 cm i.d. with one fixed and one adjustable ends. These columns were packed with Sephacryl S-100 HR at  $260 \text{ cmhr}^{-1}$ , to a bed height of approximately 8 cm. The packing quality was evaluated using pulses of sodium chloride. The columns used in the SMB had number of theoretical plates (NTP) within 8 000-10 000 and an asymmetry number in the range of 0.9 to 1.2.

To monitor the progress of the experiment various detectors and offline assays were used. Protein concentration was monitor on the raffinate (R) and waste (W) streams using on-line UV detectors. Protein was monitor off-line on samples taken from the collection bottles receiving the effluent from the raffinate, extract and waste streams. These samples were analyzed using BCA and analytical size-exclusion. The samples from the raffinate bottle were further analyzed using RPHPLC to determine the level of native protein. Urea was monitor only using the refractive index detector placed on the extract line.



### 5.3.8 Flow selection for the SMB

The operational flow rates were determined using the triangle theory, constraint for linear isotherms (Storti, Mazzotti et al. 1993; Mazzotti, Storti et al. 1997). The flow rate ratios (*m-values*) were obtained from the triangle diagram. The diagram was built using the distribution coefficients of urea ( $K_{av,urea} = 1$ ) and the fusion protein ( $K_{av,FP} = 0.27$ ) (Freydell, Ottens et al. 2007).  $K_{av,urea} > K_{av,FP}$  means that urea is the more retained component (i.e. extract product) and the fusion protein is the less retained component (i.e., the raffinate product). Eqs. (5.12a)-(5.12c) represent the linear inequalities which define the zone of complete separation in the ( $m_2, m_3$ ) plane (Fig. 5.4A).

$$K_{av,urea} < m_1 < m_{max} \quad (5.12a)$$

$$K_{av,FP} < m_2 < m_3 < K_{av,urea} \quad (5.12b)$$

$$\frac{-\varepsilon_p}{(1-\varepsilon_p)} < m_4 < K_{FP} \quad (5.12c)$$

where  $m_{1...4}$  represent the flow rate ratios of zones 1 through 4.  $m_{max}$  represents the maximum flow rate ratio, which is defined by the packing velocity.  $\varepsilon_p$  is the particle porosity defined as  $\varepsilon_p = (\varepsilon_t - \varepsilon_b)/(1 - \varepsilon_b)$ . Satisfying the linear constraints leads to the separation of urea and the fusion protein, which initiates the refolding reaction. Fig. 5.4A presents the triangle of complete separation and the operational point selected to run the SMBSEC refolding experiment. The operational point is located at the coordinates  $m_3=0.80$  and  $m_2=0.50$ . The flow rate ratios of zone 1 and 4 were selected as  $m_1=1.30$  and  $m_4=0.24$ , respectively.

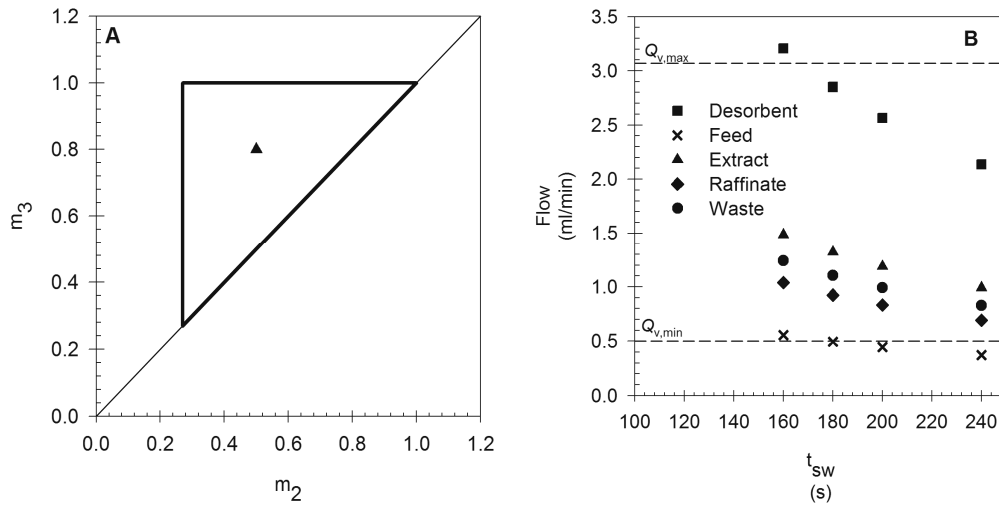


Figure 5.4. (A) SMB triangle of complete separation.  $m_3$  and  $m_2$  represent the flow rate ratios of zone 3 and zone 2. (B) Practical flow rate constraints used to select the switching time ( $t_{sw}$ ).  $Q_{v,max}$ : maximum operational flow rate, derived from the packing flow rate.  $Q_{v,min}$ : minimum operational flow rate

After selecting the  $m$ -values, the next step is the calculation of the internal or zone flow rates ( $Q_{I-IV}$ ). The internal flow rates are calculated using

$$Q_j = \frac{m_j V_C (1 - \varepsilon_b) + V_C \varepsilon_b}{t_{sw}} \quad (5.13)$$

where  $Q_j$  and  $m_j$  are the zone flow rate and flow rate ratio of zone  $j$ , respectively.  $V_C$  and  $\varepsilon_b$  represent the bed volume and the bed porosity, respectively.  $t_{sw}$  represent the switching time. To operate the SMB the external flow rates are required. These flows are calculated using the internal zone flow rates and the continuity balances around the external nodes

$$D = Q_I \quad (5.14a)$$

$$E = Q_I - Q_{II} \quad (5.14b)$$

$$F = Q_{III} - Q_{II} \quad (5.14c)$$

$$R = Q_{III} - Q_{IV} \quad (5.14d)$$

$$W = Q_{IV} \quad (5.14e)$$

To summarize, the selection of the external flow rates was based on three criteria, (1) they must be derived from  $m$ -values within the triangular region of full separation (Fig. 5.4A); (2)

the highest flow rate must be less than  $Q_{v,max}$  (Fig. 5.4B, upper dash line); and (3) the lowest flow rate must higher than  $Q_{v,min}$  (Fig. 5.4B, lower dash line). Using the selected  $m$ -values and the selection criteria, a switching time ( $t_{sw}$ ) of 180 seconds was obtained from Fig. 5.4B. This switching time corresponds to the following operational flows:  $D=2.80\text{mlmin}^{-1}$ ,  $E=1.30\text{mlmin}^{-1}$ ,  $F=0.50\text{mlmin}^{-1}$ ,  $R=0.90\text{mlmin}^{-1}$ , and  $W=1.10\text{mlmin}^{-1}$ .

### 5.3.9 Experimental design

An experimental design methodology was employed to evaluate the effect of the mobile phase pH and the feed concentration of denatured and reduced protein ( $C_{f,D\&R}$ ), on the size-exclusion refolding yield of the model protein. A 3-level factorial design was constructed using pH at 9.0, 10.10 and 11.2; and  $C_{f,D\&R}$  at 2.50, 5.0 and 7.50  $\text{mgml}^{-1}$ . A total of 13 experiments were carried out, including 4 center points used to assess the random experimental error. The experimental design was constructed and analyzed using STATGRAPHICS Centurion XV.

### 5.3.10 Estimation of the reaction rate constants

The kinetic constants of the competing reaction mechanism (Fig.5.2) were obtained using a non-linear constraint optimization. The objective function used in the optimization routine is presented in Eq. 5.15.

$$\min_x SSE(x) = \sum_{i=1}^n (C_{exp,i} - C_{mod,i}(x,t))^2 \quad (5.15)$$

where  $SSE$  is the sum of squared error and  $x$  is the vector of kinetic constants.  $C_{exp,i}$  represents the experimental concentration profile, which was obtained from the on-column batch refolding experiments.  $C_{mod,i}$  represents the modeled concentration profile obtained using Eqs. (5.1)-(5.4b) and Eqs. (5.8a)-(5.8e).

## 5.4 Results and discussion

### 5.4.1 Competitive reaction mechanism and kinetic parameters

To model a refolding reactor, whether a batch tank or a continuous chromatographic reactor (e.g., SMBSECR), a reaction mechanism is required. In the case of protein refolding, such reaction mechanism should minimally include the competition between the formation of native protein and the formation of aggregated protein. The incorporation of the competitive reaction scheme, to the overall differential mass balance, provides the link between the operational variables of the reactor and the refolding yield.

Conventionally, the competition between folding and aggregation has been viewed as a competition between a first-order folding, competing against a higher-order aggregation (Kiefhaber, Rudolph et al. 1991; Hevehan and Clark 1997). These schemes typically have three fitting parameters: two kinetic constants and the aggregation order. The aggregation order is normally fixed and the kinetic constants are thereafter determined using a fitting algorithm, and batch dilution refolding experimental data. A shortcoming of this approach is that it does not take into account the extent of aggregation, since it only considers the concentration profile of the active protein. As distinct from the aforementioned approach, in this work size-exclusion refolding data were used to construct the competitive scheme and to determine its kinetic parameters. The size-exclusion chromatogram, obtained after an on-column refolding experiment, provides information about the degree of protein aggregation, which is important because it aids to decide how many species the polymerization mechanism needs to account for. But before the on-column refolding chromatogram can be used, the contribution of the separation and the reaction to the signal need to be taking into consideration.

Figure 5.5B presents a typical chromatogram from the on-column refolding experiments showing three distinct peaks. These peaks are the result of a reaction(s) and a separation. Peak 1 and peak 2 include the higher molecular weight aggregates, and peak 1 contains the monomers, containing the native product. The presence of two peaks carrying high molecular weight aggregates indicated that representing aggregation as a higher-order reaction was not applicable for the presented case. That kind of representation considers the aggregates as a single component and thus it will not capture the aggregation behavior of the model protein, justifying the usage of the polymerization chain mechanism.

To extract the kinetic information from the chromatogram the column model (i.e., Eqs. (5.1)-(5.4b)) was used in combination with the reaction mechanism (i.e., Eqs. (5.8a)-(5.8e)). Prior

to the non-linear regression process however, the ability of the model to predict the separation of a set of protein standards on the Sephacryl S100 1.6/60 column was verified. Fig 5.5A shows that the model is well capable to describe this 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.,  $\epsilon_b$ ,  $\epsilon_t$ ), the solutes (i.e.,  $MW$ ,  $D_f$ , etc) and the flow rate. Having been able to account for the contribution of the separation, leaves the reaction mechanism adopted (Fig. 5.2) as the single contributor to the shape and number of peaks of the on-column refolding chromatogram. Figure 5.5B shows that the number of peaks and the shape of the on-column refolding data is well represented by the competitive mechanism used, with  $k_1=1.03\pm 9.7\cdot 10^{-5} \text{ s}^{-1}$ ,  $k_2=0.08\pm 8\cdot 10^{-3} \text{ s}^{-1}$ ,  $k_3=1.69\pm 8.2\cdot 10^{-2} \text{ l mmol}^{-1} \text{ s}^{-1}$  and  $k_4=2.42\pm 0.19 \text{ l mmol}^{-1} \text{ s}^{-1}$ .

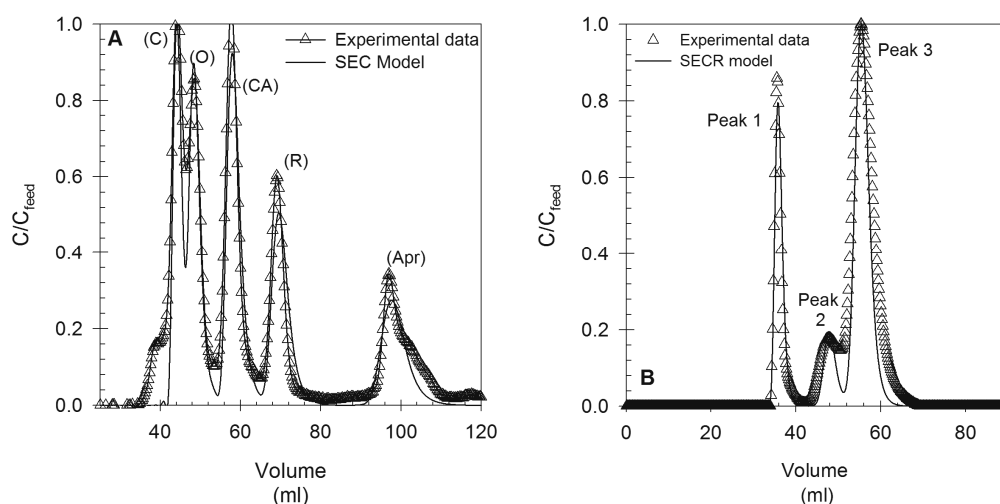


Figure 5.5 (A) Column model validation. Column type Sephacryl S100 16/60. Flow rate  $1 \text{ ml min}^{-1}$ . C: Conalbumin, O:Ovalbumin, CA: Carbonic anhydrase, R: Ribonuclease, Apr: Aprotinin. (B) Size-exclusion reactor model and experimental data. Peak 1 and peak 2: high molecular weight aggregates. Peak 3: monomers, including the native protein

#### 5.4.2 Effect of pH and denatured and reduced protein feed concentration on refolding yield

Variations on the oxidative refolding yield, of pure proteins or protein expressed in inclusion bodies, have been strongly correlated to changes on (1) the redox potential of the system and (2) the total denatured and reduced protein concentration at the beginning of the refolding process (Buswell and Middelberg 2002; Winter, Lilie et al. 2002). Own to the strong correlations displayed by these two factors, the refolding yield of the model protein was studied as a function of (1) the pH of the mobile phase, which influences the redox potential of the system, and (2) the feed concentration of denatured and reduced protein.

Figure 5.6 shows the response surface obtained from the experimental design (section 5.3.8) and the batch size-exclusion experimental measurements. The estimated response surface was obtained using Eq. (5.16). This equation is purely empirical and it explains quantitatively the variability of the refolding yield due to changes on pH and the concentration of denatured and reduced protein.

$$Y_{N,SECR} = -3.06 + 0.48pH + 0.30C_{fD\&R} - 0.014(pH)^2 - 0.026pHC_{fD\&R} - 0.007(C_{fD\&R})^2 \quad (5.16)$$

The data in Fig. 5.6 indicate that the on-column refolding yield, of the model protein, is strongly dependent on the pH of the mobile phase and on the concentration of denatured and reduced protein fed ( $C_{fD\&R}$ ) to the size-exclusion column. Higher refolding yields were attained at the lowest protein concentration tested and the highest pH. This is presumably because of a net positive effect of the operational parameters; at low protein concentration the aggregation rates decrease which favors refolding. On the other hand, an operational pH (i.e., 11.2) far above the protein's isoelectric point (i.e., 7.64) has two positive effects: (1) protein aggregation is decreased presumably due to electrostatic repulsion and to increased protein solubility, and (2) in general alkaline pH aids on the formation of disulfide bonds (Li, Zhang et al. 2002). Fig. 5.6 additionally shows that a decrease of the refolding yield is to be expected when (1) the pH decreases or (2) the protein feed concentration increases. The analysis of variance (ANOVA) of the refolding yield showed that pH, denatured and reduced protein feed concentration ( $C_{fD\&R}$ ) and their interaction, have  $p$ -values of less than 0.05; indicating that they are statistically significant at the 95% confidence interval. Moreover, a Pareto analysis of the data showed that  $C_{fD\&R}$  has a higher impact on the variability of the refolding yield, compared to the pH or the interaction of the two. These findings support the hypothesis that protein aggregation explains the strong dependence of the refolding yield on  $C_{fD\&R}$ .

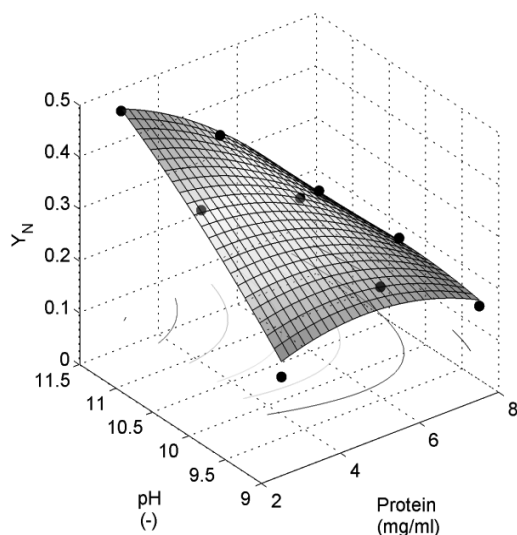


Figure 5.6: Estimated response surface (RS) for the refolding yield as a function of the pH and the denatured and reduced protein feed concentration. Experimental points showed as black circles. pH: pH of the mobile phase.  $C_{f,D\&R}$ : Feed concentration of denatured and reduced protein

### 5.4.3 Size-exclusion refolding

Size-exclusion chromatographic protein refolding is based on the separation of the chaotrope (i.e., urea) and reducing agent (i.e., DTT) from the unfolded and reduced protein. The small components (i.e., urea and DTT) can penetrate a larger fraction of the pore volume of the gel filtration material (e.g., Sephacryl S100), increasing their traveling distance and thus delaying their migration through the column. The larger solute (i.e., proteins) can penetrate less the pore volume, facing a shorter traveling distance. As a result, they migrate faster than the small solutes. The separation of the small solutes (i.e., urea and DTT) from the protein, results in a decrease of their local concentration around the protein, which induces protein folding and the formation of disulfide bonds.

The on-column refolding of the model protein (Fig 5.5B) resulted in the formation of three protein peaks (i.e., peak 1, peak 2 and peak 3). The urea and DTT eluted in the total volume of the column (not shown in Fig.5.5B). The three protein containing peaks were collected, analyzed by BCA to determine their total protein concentration, and by RPHPLC to determine the concentration of native protein (i.e., fusion protein). Fig. 5.7 shows the experimental yield of the on-column refolding experiments, together with the estimation of the random experimental error (error bars), estimated using the center points of the experimental design of experiments. The data presented in Fig. 5.7 show that practically for all pH values evaluated, the refolding yield decreases as the feed concentration of denatured and reduced protein increases. This claim holds, except for the point at a  $C_{f,D\&R}$  of  $2.5\text{mgml}^{-1}$

and pH 9.0. The reason why this point diverges from the tendency displayed by the other data series (i.e., pH 10.10 and 11.20) is unknown. It is interesting to point out though, that for those experiments conducted at  $C_{fD\&R}$  of  $7.50\text{mgml}^{-1}$ , the refolding yield converged to an approximate value of 0.20, disregarding the operational pH used. Such behavior is plausibly explained by a significant increase on the aggregation rates due to the high  $C_{fD\&R}$ , which mask the positive effect of pH, observed at low  $C_{fD\&R}$  where protein aggregation rates are relatively low.

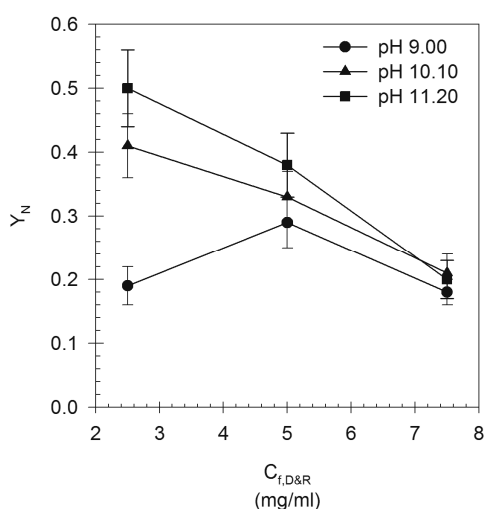


Figure 5.7: Refolding yield as a function of the denatured and reduced protein feed concentration and the pH of the mobile phase. The error bars represent the random error estimated from the center points of the experimental design

#### 5.4.4 Simulated moving bed size-exclusion refolding

The operational variables of the SMBSECR (i.e.,  $t_{sw}$ , external flows) were intentionally tuned to guarantee the complete separation of the denatured and reduced protein from the chaotrope agent (i.e., urea) and the reducing agent (i.e., DTT). This was done as such because complete separation between these components is the minimum criterion that needs to be satisfied in order to have refolding. The operational conditions, pH and protein feed concentration, were selected based on the results obtained from the batch size-exclusion refolding experiments. The pH of the mobile phase was set to 10.50 and the concentration of denatured and reduced protein was adjusted to approximately  $3\text{mgml}^{-1}$ .

Fig. 5.8 presents the internal concentration profiles of the SMBSECR. The profiles of urea, native (N), intermediate (I), aggregate class 2 ( $A_2$ ) and aggregate class 3 ( $A_3$ ), were predicted using Eqs. (5.1)-(5.3b), (5.6a)-(5.6d), and (5.8a)-(5.8e). The concentration profile of the misfolded (M) species is absent in Fig. 5.8 because this species is formed at negligible amounts, according to the reaction mechanism. The data presented in Fig. 5.8 shows that



the protein front moved to the raffinate port as predicted by the flow rate selection criteria. This claim is supported by the good agreement between the predicted internal total protein concentration profile and the profile measured during the experiment. In addition, in Fig. 5.8 the total protein profile is divided into the individual signals of the different protein components consider by the competitive reaction mechanism presented in Fig. 5.2 and Eqs. (5.8a)-(5.8e). It is interesting to point out, that based on the model predictions, the aggregate class 3 ( $A_3$ ) concentration wave will reach the waste outlet intermittently under the flow rate settings used in the experiment. This prediction however is not 100% accurate, since the analysis of the samples collected from the waste stream, indicated no protein in the waste (data not shown).

Further analysis of the SMBSECR experimental data, conducted with the aid of the SMBSECR model, is presented in Figs. 5.9A and 5.9B. Fig. 5.9A presents the data corresponding to the raffinate port. Samples from the raffinate port were monitored at two places: (1) On-line measurements were done using the UV detector placed on the flow path of the raffinate stream and (2) off-line measurements were done on samples taken from the raffinate bottle. The behavior of both measurements is displayed in Fig. 5.9A. The data presented in Fig. 5.9A, show that there is a good agreement between the breakthrough of the protein signal predicted by the SMBSECR model and the one registered by the UV detector (Fig.5.9A dotted line vs. dash line). The same claim holds for the total protein concentration in the raffinate bottle, since both the experimental measurement and the model prediction are in good agreement (Fig.5.9A solid line vs. ▲). The delay time between on-line and offline measurements is explained by the fed-batch behavior of the raffinate bottle. This was taken into account in the SMBSECR model by a fed-batch equation that describes the raffinate bottle, allowing the prediction of both on-line and off-line signals.

Fig. 5.9B presents the data recorded by the refractive index detector, placed on the flow path of the extract stream. This detector was placed in this line to monitor the breakthrough of urea. The signal of the detector was considered a qualitative measurement because of two reasons (1) the detector reached a saturation limit at low urea levels which was not expected and (2) it was noted that the signal was affected by pressure pulsations, most probably caused by the piston pumps used on the SMB and the absence of a pneumatic dampener (Gupta and Nikelly 1985). Despite its qualitative nature, the RI signal does confirm the breakthrough of urea through the extract line (Fig. 5.9B dotted line). This observation is supported by the prediction of the SMBSECR model (Fig. 5.9B solid line) and the off-line analysis of the extract bottle and line (data not shown). The offline analysis of the extract bottle, using analytical SEC (data not shown), confirmed the presence of DTT. The presence

of this compound in the extract stream is expected, since both DTT and urea have approximately the same distribution coefficient.

It can be concluded that the SMBSECR model describes the separation behavior of the SMB unit well, represented by both the internal and the outlet concentration profiles (Figs. 5.8, 5.9A and B). On the other hand, the model is yet not able to fully capture the formation of native product, which is evident from Fig. 5.9A. Fig. 5.9A presents the measured profile of the native protein and the profile predicted by the SMBSECR model (Fig. 9A ■ vs. dash-dot line). It is relevant to point out that the model captures the initial breakthrough of the native protein quite well, as this is connected to its distribution coefficient and the flow settings of the SMB. However, as time progresses the model prediction begins to diverge from the trend of the experimental measurements, ultimately predicting a concentration of native product approximately three times higher than the measured value.

Deviations of the SMBSECR model predictions from experiment occur in two situations (1) a predicted protein leakage in the form of an aggregate class 3 in the waste stream, was not measured during the SMBSEC experiment, and (2) the predicted concentration of native product, is approximately three fold higher compared to the measured one (table 2). This discrepancy between the model and experiment might be explained by the effect that the dynamic change in chemical environment, occurring in zones II and III of the SMB, has on the kinetics of the competitive mechanism assumed (Fig.5.2). This hypothesis is supported by the evidence presented in the study by Buswell et al (Buswell and Middelberg 2003), where a dependence of the kinetic constants, of their assumed mechanism, on the concentrations of DTT and urea was observed. Coupling the dynamic change of the chemical environment to the SMBSECR model is *per se* not difficult, the challenge lies though in finding out which reactions are the ones affected and to define a function that properly represent such influence. Table 5.2 presents a summary of the model's output, compared to the experimentally determined values.

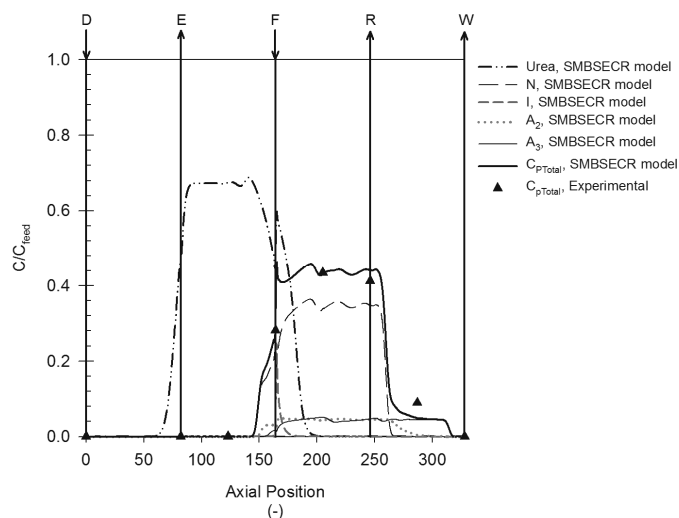


Figure 5.8. SMBSECR internal concentration profiles. ( $\blacktriangle$ ) Total protein concentration measured at cyclic steady state (CSS) using the BCA assay. Solid thick line: Predicted total protein profile. Dash-dot-dot line: Predicted urea profile. Dashed line: Predicted native profile. Red short dash line: predicted intermediate profile. Cyan dotted line: Predicted aggregate class 2 profile. Solid thin line: Predicted aggregate class 3 profile. All predicted profiles were calculated using the SMBSECR model. N: Native protein; I: intermediate;  $A_2$ : Aggregate class 2;  $A_3$ : Aggregate class 3. D: Desorbent; E: Extract; F: Feed; R: Raffinate; W: Waste. Simulated profiles correspond to half switching time. Axial position correspond to  $N \cdot N_{col}$ , where  $N$  is the number of grid points per column and  $N_{col}$  the number of columns on the SMB unit

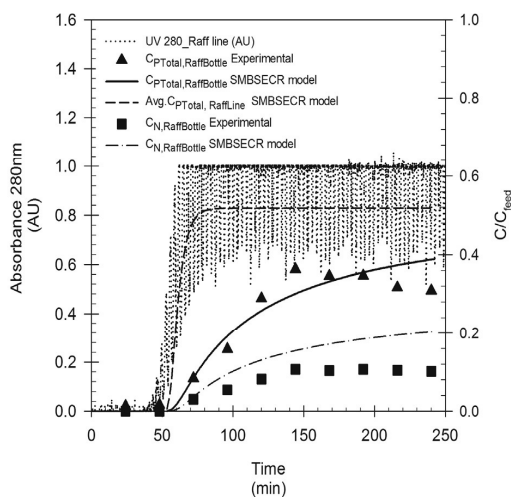


Figure 5.9A: Raffinate concentration profiles. Dotted line: Signal of the UV detector located in the raffinate stream, the increasing UV-280 signal indicates the increase in protein concentration. Dash-line: Total protein concentration in the raffinate stream as predicted by the SMBSECR model. ( $\blacktriangle$ ) Total protein concentration on the raffinate bottle. ( $\blacksquare$ ) Concentration of native protein on the raffinate bottle. Solid line: Total protein concentration on the raffinate bottle, as predicted by the SMBSECR model. Dash-Dot line: Concentration of native product on the raffinate bottle, as predicted by the SMBSECR model

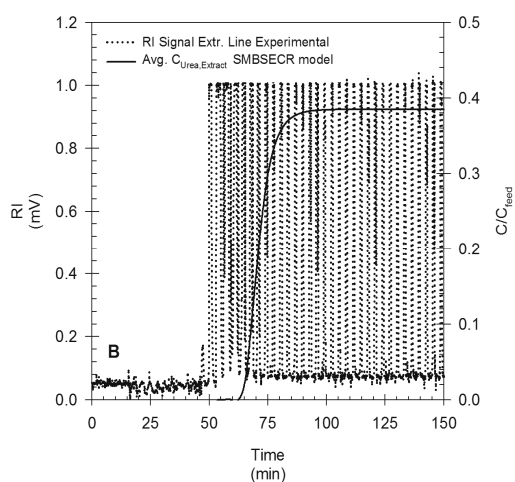


Figure 5.9B. Extract concentration profiles. Dotted line: Signal of the refractive index detector (RI) located in the extract line. Solid line: Urea concentration profile in the extract line, as predicted by the SMBSECR model

#### 5.4.5 Comparison between SEC batch refolding and SMBSEC refolding

Simulated moving bed technology has inherent advantages over conventional batch chromatography. Such advantages include (1) less column material requirement compared to the batch operation and (2) a significant decrease in the amount of eluent use in contrast to the batch operation. The aforementioned advantages become more pronounced when the type of chromatographic separation is based on size-exclusion. This is because the amount of column material, in conventional batch size-exclusion chromatography, is directly related to the sample loading volume, which is not the case for size-exclusion simulated moving bed chromatography (SMBSEC). These advantages have been well demonstrated in the context of conventional purifications (Chin and Wang 2004), however their validity is yet to be tested in the case of size-exclusion refolding reactors. Thus, the first comparison will be done on the basis of four performance indicators. These are (1) refolding yield (Eqs. (5.9a)-(5.9b)), (2) volumetric productivity (Eqs.(5.10a)-(5.10b)), (3) specific eluent consumption (Eqs. (5.11a)-(5.11b)), and (4) concentration of native product in the product stream. These four indicators are listed in table 5.3, for both the size-exclusion refolding reactor (SECR) and the simulated moving bed size-exclusion refolding reactor (SMBSECR).

From the data on table 5.3 the following conclusions are evident: (1) the refolding yield of the SMBSECR is more or less half of that from the SECR, (2) the volumetric productivity of the SMBSEC is about 53 times higher than that of the SECR, (3) the specific solvent consumption of the SMBSECR is approximately  $1/10^{\text{th}}$  of that of the SECR, and (4) the

concentration of the product (i.e. native protein), leaving the SMBSECR, is roughly 4.5 fold higher than the product concentration leaving the SECR.

The differences in volumetric productivity and specific solvent consumption, between the two chromatographic reactors, are quite significant (i.e., SMBSECR and SECR). But what is the actual contribution from the operational conditions to the differences on these two performance indicators? The answer to this question is not so evident from the data on table 5.3, and it was extracted using Eqs. (5.17a)-(5.17b). These two equations were obtained from Eqs. (5.10a)-(5.10b) and (5.11a)-(5.11b). Assuming that both, the refolding yield and the feed concentration of denatured and reduced protein are equal for both chromatographic reactors (i.e., SMBSECR and SECR), isolates the contribution of the operational variables of the reactors, to the productivity ratio (Eq. 5.17a) and to the specific eluent consumption ratio (Eq. 5.17b). The results of this analysis indicate that (1) the productivity of the SMB unit is about 117 times higher than the batch unit, and (2) the specific eluent consumption of the SMBSEC unit is  $1/20^{\text{th}}$  of that of the batch SEC unit. The increase of these indicators is anticipated, owing to the fact that an SMBSEC unit is in principle more efficient than the batch SEC unit. Thus, operating the SMBSEC unit as a SEC refolding reactor decreases its inherent advantage, in terms of productivity and specific eluent consumption, over the batch SEC reactor. Despite of the decrease, the remainder difference remains significantly high.

$$\frac{Pr_{\text{SMBSECR}}}{Pr_{\text{SECR}}} = \underbrace{\left( \frac{F}{N_{\text{Col}}^{\text{SMBSECR}} V_C^{\text{SMBSECR}}} \right)}_{\text{Operational variables SMBSECR}} \left( \frac{C_{f,D\&R}^{\text{SMBSECR}}}{C_{f,D\&R}^{\text{SECR}}} \right) \left( \frac{Y_{N,\text{SMBSECR}}}{Y_{N,\text{SECR}}} \right) \underbrace{\left( \frac{t_{\text{Cycle}} V_C^{\text{SECR}}}{V_{\text{inj}}} \right)}_{\text{Operational variables SECR}} \quad (5.17a)$$

$$\frac{Ec_{\text{SMBSECR}}}{Ec_{\text{SECR}}} = \underbrace{\left( \frac{D}{F} \right)}_{\text{Operational variables SMBSECR}} \left( \frac{C_{f,D\&R}^{\text{SECR}}}{C_{f,D\&R}^{\text{SMBSECR}}} \right) \left( \frac{Y_{N,\text{SECR}}}{Y_{N,\text{SMBSECR}}} \right) \underbrace{\left( \frac{V_{\text{inj}}}{\dot{Q}t_{\text{cycle}}} \right)}_{\text{Operational variables SECR}} \quad (5.17b)$$

Having discussed the differences in volumetric productivity and specific eluent consumption, between the SMBSECR and the SECR, brings us to the differences in refolding yield and product concentration.

As previously mentioned, the refolding yield of the SECR was approximately 2 fold higher than the refolding yield of the SMBSECR (table 5.3). This difference might be reasonably explained by either one of the following scenarios. Scenario (1) the dynamic change in the chemical composition, occurring in zones II and III of the SMB, has a negative effect on the rate of product formation. In these zones, the standing waves of protein, denaturant (i.e.,

urea) and reducing agent (i.e., DTT), are moving back a forth. The change in chemical composition becomes more severe as the waves move into zone III, because they are mixed with fresh feed. Compared to what happens on the SECR, this is a rather complex situation. On the SECR, once injected, the concentration waves of protein, denaturant, and reducing agent separate quite fast and never meet again during the rest of the process. Thus, the protein faces the change in chemical composition only once, which may be positive assuming that this scenario explains the higher refolding yield obtained with the SECR.

Scenario (2) the low local protein concentration at which refolding proceeds in the SECR, favors refolding, explained by a decrease on the rate(s) of protein aggregation, owing to the low local protein concentration. This scenario is supported by the large dilution occurring during the SECR operation, owing to the large column volume to sample ratio. During SEC refolding this dilution factor has been estimated to be on the area of 40 to 120 fold (Gu, Su et al. 2001; Park, Lee et al. 2005). Furthermore, high dilution has been linked to an increase on the SEC batch refolding yield (Ding, He et al. 2008). In contrast, during the SMBSECR operation such large dilution does not occur. Actually, the largest dilution occurring during the SMB operation is given by Eq. (5.18). During the SMBSECR experiment this dilution factor was roughly 4, which is relatively low. Thus, the SMBSECR operates at relatively high local protein concentration, as compared to the SECR. Therefore, if the dilution on the SMBSECR is not as high as the one on the SECR, and operating at low local protein concentrations slow down the aggregation rates, then this scenario may explain why the SECR has a higher yield compared to the SMBSECR.

Scenario (3) a higher pH of operation on the SECR (i.e., 11.20), compared to the SMBSECR (i.e., pH 10.50), explains the approximately two fold refolding yield difference, between the SECR and the SMBSECR (see table 3). This scenario is evaluated with the aid of Eq. (5.16). This equation predicts the refolding yield of the SECR as a function of the pH of the mobile phase and the feed concentration of denatured and reduced protein ( $C_{ID\&R}$ ). Using Eq. (5.16) to estimate the refolding yield of the SECR at a pH of 10.50, same as the one used for the SMBSECR experiment, produces an expected refolding yield for the SECR which is still roughly twice as high as the one for the SMBSECR. Therefore, the pH difference does not explain the difference in refolding yield.

Finally, the difference in the concentration of the native protein in the product stream coming out of the refolding reactors (i.e., SECR and SMBSECR), can be reasonably be explained by the large difference in dilution occurring in the units. As previously discussed, the large dilution on the SECR reactor is explained by the large column volume to sample

loading volume ratio. This condition is important in order to preserve a good resolution during the separation of the solutes in the column. In contrast, dilution on the SMBSECR is relatively low, owing to the use of smaller columns.

$$D_{f,\text{SMBSECR}} = \frac{Q_{\text{III}}}{F} = \frac{D-E}{F} + 1 \quad (5.18)$$

Up to this point, the comparison based on quantitative criteria undoubtedly showed the significant advantages that a SMBSEC refolding reactor offers, as compared to the conventional batch SEC refolding reactor. The separation of monomers from aggregates formed during refolding is also an important question to be addressed (very important in most biopharmaceutical processes, since aggregates have been linked to adverse effects in patients, such as unexpected immune responses (Mahler, Friess et al. 2009)), since clearly both reactors can easily separate the small solutes (i.e., DTT and urea) from the proteins. To address this point, Figs. 5.5B and 5.8 are needed. Fig. 5.5B shows that the SECR is capable of fractionating, with good resolution, the monomers from the large molecular weight aggregates. On the other hand, Fig. 5.8 shows that the SMBSECR is not capable to do so using a classical 4-zone 2-2-2-2 configuration. What is important to point out though, is that the lack of ability of the current SMBSECR to separate the monomers from the aggregates after refolding, is strictly related to the configuration of the unit and to the design criteria used for the flow rate selection. Considering the number of degrees of freedom, for its design, and the flexibility offered by the SMB technology (Seidel-Morgenstern, Keßler et al. 2008), will allow an altered design of the SMBSEC refolding reactor. This design will (1) take into consideration the distribution coefficients of urea, monomers and aggregates, to design the flow selection criteria, and (2) it will have specific zones in charge of: column regeneration, protein refolding, and product purification, that is native product free of aggregated material. This separation boils down to a three component continuous chromatographic separation, which can be handled by system configurations like e.g., Morbidelli et al MCCC system (Ströhlein, Aumann et al. 2006) or two sequential 4 zone binary separation SMBs (Keßler and Seidel-Morgenstern 2006).

## 5.5 Conclusions

The work presented in this paper described the successful refolding, of an industrially relevant protein, by means of size-exclusion. Size-exclusion refolding was done using the conventional batch chromatographic refolding reactor (SECR) and the novel simulated moving bed refolding reactor (SMBSECR). This work achieved:

- (1) the estimation of a suitable reaction mechanism, to capture the competition between refolding, misfolding and protein aggregation. This was done using a novel approach that employed a mechanistic column model and size-exclusion batch refolding data. This mechanism was then used to link the separation to the refolding yield of the SMBSECR;
- (2) a statistical design of experiments (DOE) methodology to determine the effect of the selected operational variables, mobile's phase pH and feed concentration of denatured and reduced protein, on the size-exclusion refolding yield of the model protein. Using the DOE, the combined effect of both process parameters on the refolding yield, was determined. The data indicated that (a) high refolding yields are expected at low protein feed concentration and high pH, and (b) that the positive effect of the alkaline pH is more pronounced at low protein concentrations. The statistical analysis finally revealed, that the positive effect of pH lessens, as the protein feed concentration increases, presumably owing to an increase of the aggregation rates;
- (3) a mechanistic analysis of the SMBSECR data, with the aid of a detailed model developed during this work, leading to the following conclusions: (a) the designed model is capable of predicting the separation behavior of the reactor well, based solely on the effect that protein concentration has on the reaction rates; and (b) the model might be improved by incorporating the dynamic change in chemical environment, occurring in zones II and III. This slight modification to the model is not difficult. The challenge, however, lies in finding the proper relations, describing the influence of the chemical environment on the reaction mechanism; and
- (4) a detailed comparison between the two refolding reactors (i.e., SECR and SMBSECR) based on four performance criteria: refolding yield, volumetric productivity, specific eluent consumption, and the product concentration leaving the reactor. This quantitative comparison, clearly showed the advantages of the SMBSECR, compared to a batch SEC reactor. A further qualitative comparison, showed a drawback of the current SMBSECR over the SECR. However, this limitation of the SMBSECR is strictly related to the SMB configuration used in this work (i.e., classical 4 zone 2-2-2-2 columns for binary separation) to achieve refolding. A more appropriate three component separation SMB design would additionally achieve purification of the generated monomers. This, however, falls beyond the scope of the present paper and is material for future work.



### **5.6 Acknowledgements**

The authors would like to thank: Dirk Geerts (Delft University of Technology) for his assistance in getting the SMB unit up and running, and Danny Lagarde (Schering-Plough) for his help with the RPHPLC analysis.

## 5.7 References

- [1] F.R. Schmidt, *Appl. Microbiol. Biotechnol.* 65 (2004) 363.
- [2] M.H. Werner, G.M. Clore, A.M. Gronenborn, A. Kondoh, R.J. Fisher, *FEBS Lett.* 345 (1994) 125.
- [3] B. Batas, J.B. Chaudhuri, *Biotechnol. Bioeng.* 50 (1996) 16.
- [4] X. Geng, X. Chang, *J. Chromatogr.* 599 (1992) 185.
- [5] Q. Bai, Y. Kong, X.-d. Geng, *J. Liq. Chromatogr. Relat. Technol.* 26 (2003) 683
- [6] T.E. Creighton, *UCLA Symp. Mol. Cell. Bio., New Ser.* 39 (1986) 249.
- [7] C. Machold, R. Schlegl, W. Buchinger, A. Jungbauer, *J. Biotechnol.* 117 (2005) 83.
- [8] G. Lemercier, N. Bakalara, X. Santarelli, *J. Chromatogr B.* 786 (2003) 305.
- [9] R. Zahn, C. von Schroetter, K. Wüthrich, *FEBS Lett.* 417 (1997) 400.
- [10] W. Shalongo, R. Ledger, M.V. Jagannadham, E. Stellwagen, *Biochemistry* 26 (1987) 3135.
- [11] A. Jungbauer, W. Kaar, R. Schlegl, *Curr. Opin. Biotechnol.* 15 (2004) 487.
- [12] E.M. Fahey, J.B. Chaudhuri, P. Binding, *J. Chromatogr B.* 737 (2000) 225.
- [13] Z. Gu, X. Zhu, S. Ni, H. Zhou, Z. Su, *J. Biochem. Biophys. Methods* 56 (2003) 165.
- [14] Z. Gu, Z. Su, J.-C. Janson, *J. Chromatogr. A.* 918 (2001) 311.
- [15] Z. Gu, M. Weidenhaupt, N. Ivanova, M. Pavlov, B. Xu, Z.-G. Su, J.-C. Janson, *Protein. Expres. Purif.* 25 (2002) 174.
- [16] X.-Y. Dong, Y. Wang, J.-H. Shi, Y. Sun, *Enzyme Microb. Technol.* 30 (2002) 792.
- [17] R. Schlegl, G. Iberer, C. Machold, R. Necina, A. Jungbauer, *J. Chromatogr. A.* 1009 (2003) 119.
- [18] H. Lanckriet, A.P.J. Middelberg, *J. Chromatogr. A.* 1022 (2004) 103.
- [19] B.-J. Park, C.-H. Lee, S. Mun, Y.-M. Koo, *Process Biochem.* 41 (2006) 1072.
- [20] B.-J. Park, C.-H. Lee, Y.-M. Koo, *Korean J. Chem. Eng.* 22 (2005) 425.
- [21] D.B. Broughton, C.G. Gerhold, in *U.S. Patent 2985589*, 1961.
- [22] D.B. Broughton, *Chem. Eng. Prog.* 64 (1968) 60.
- [23] Y. Xie, S. Mun, J. Kim, N.-H.L. Wang, *Biotechnol. Prog.* 18 (2002) 1332.
- [24] L. Gueorguieva, L.F. Vallejo, U. Rinas, A. Seidel-Morgenstern, *J. Chromatogr. A.* 1135 (2006) 142.
- [25] S. Lucena, P. Rosa, L. Furlan, C. Santana, in *Engineering and Manufacturing for Biotechnology*, 2002, p. 325.
- [26] D.A. Horneman, M. Ottens, J.T.F. Keurentjes, L.A.M. van der Wielen, *J. Chromatogr. A.* 1157 (2007) 237.
- [27] D.A. Horneman, M. Ottens, L.A.M.v.d. Wielen, J.T.F. Keurentjes, *AIChE J.* 53 (2007) 1441.
- [28] G. Guiochon, D.G. Shirazi, A. Felinger, A.M. Katti, *Fundamentals of Preparative and Nonlinear Chromatography*, Academic Press, 2006.
- [29] L. Lapidus, N.R. Amundson, *J. Phys. Chem.* 56 (1952) 984.
- [30] W.E. Schiesser, *The numerical method of lines: integration of partial differential equations*, Academic Press Limited, London, 1991.
- [31] C.-M. Yu, S. Mun, N.-H.L. Wang, *J. Chromatogr. A.* 1132 (2006) 99.
- [32] Z. Li, Y. Gu, T. Gu, *Biochem. Eng. J.* 2 (1998) 145.
- [33] B. Zelic, B. Neseck, *Eng. Life Sci.* 6 (2006) 163.
- [34] E.J. Wilson, C.J. Geankoplis, *Ind. Eng. Chem. Fund.* 5 (1966) 9.
- [35] Z. Ma, N.H.L. Wang, *AIChE J.* 43 (1997) 2488.
- [36] S.F. Chung, C.Y. Wen, *AIChE J.* 14 (1968) 857.

- [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] T. Kiefhaber, R. Rudolph, H.-H. Kohler, J. Buchner, *Nat. Biotechnol.* 9 (1991) 825.
- [40] D.L. Hevehan, E.D.B. Clark, *Biotechnol. Bioeng.* 54 (1997) 221.
- [41] A.M. Buswell, A.P.J. Middelberg, *Biotechnol. Prog.* 18 (2002) 470.
- [42] J. Maachupalli-Reddy, B.D. Kelley, E. De Bernardez Clark, *Biotechnol. Prog.* 13 (1997) 144.
- [43] X.Y. Dong, G.Q. Shi, W. Li, Y. Sun, *Biotechnol. Prog.* 20 (2004) 1213.
- [44] A.M. Buswell, A.P.J. Middelberg, *Biotechnol. Bioeng.* 83 (2003) 567.
- [45] M.A. Speed, J. King, D.I.C. Wang, *Biotechnol. Bioeng.* 54 (1996) 333.
- [46] L.S. Tobacman, E.D. Korn, *J. Biol. Chem.* 258 (1983) 3207.
- [47] F.A. Ferrone, J. Hofrichter, H.R. Sunshine, W.A. Eaton, *Biophys. J.* 32 (1980) 361.
- [48] J. Winter, H. Lilie, R. Rudolph, *Anal. Biochem.* 310 (2002) 148.
- [49] J. Winter, P. Neubauer, R. Glockshuber, R. Rudolph, *J. Biotechnol.* 84 (2000) 175.
- [50] E.J. Freydell, M. Ottens, M. Eppink, G.v. Dedem, L.v.d. Wielen, *Biotechnol. J.* 2 (2007) 678.
- [51] L. Hagel, M. Ostberg, T. Andersson, *J. Chromatogr. A.* 743 (1996) 33.
- [52] L. Hagel, *Curr. Protoc. Protein. Sci.* 8.3 (1998).
- [53] G. Storti, M. Mazzotti, M. Morbidelli, S. Carrà, *AIChE J.* 39 (1993) 471.
- [54] M. Mazzotti, G. Storti, M. Morbidelli, *J. Chromatogr. A.* 769 (1997) 3.
- [55] M. Li, G. Zhang, Z. Su, *J. Chromatogr. A.* 959 (2002) 113.
- [56] P.K. Gupta, J.G. Nikelly, *J. High Resolut. Chromatogr.* 8 (1985) 177.
- [57] C.Y. Chin, N.-H.L. Wang, *Separation and Purification Reviews* 33 (2004) 77
- [58] Y. Ding, L. He, A.P.J. Middelberg, *Chem. Eng. Sci.* 63 (2008) 4333.
- [59] H.-C. Mahler, W. Friess, U. Grauschopf, S. Kiese, *J. Pharm. Sci.* 98 (2009) 2909.
- [60] A. Seidel-Morgenstern, L.C. Keßler, M. Kaspereit, *Chem. Eng. Technol.* 31 (2008) 826.
- [61] G. Ströhlein, L. Aumann, M. Mazzotti, M. Morbidelli, *J. Chromatogr. A.* 1126 (2006) 338.
- [62] L.C. Keßler, A. Seidel-Morgenstern, *J. Chromatogr. A.* 1126 (2006) 323.

# 6

## Techno-economic Evaluation of an Inclusion Body Solubilization and Recombinant Protein Refolding Process

---

### **Abstract**

The production of recombinant biopharmaceuticals using *Escherichia coli* (*E. coli*) is a well established production process. Expression of recombinant proteins in *E. coli* is normally accompanied by the formation of inclusion bodies. IBs are insoluble protein aggregates that contain the product of interest in an inactive, insoluble form. To obtain the protein product in an active (native), soluble form, the IBs must be first solubilized and thereafter the soluble, often denatured and reduced protein must be refolded. Several alternatives to conduct IBs solubilization and protein refolding, all technically feasible, have been proposed in recent years. However, rarely these alternatives have been evaluated from an economical point of view, which questions the feasibility of their implementation at a preparative scale. The presented study assesses the economic performance of four distinct processes alternatives. These include: pH induced IBs solubilization and protein refolding; IBs solubilization using urea, DTT and alkaline pH followed by batch size-exclusion protein refolding; IBs solubilization using urea, DTT and alkaline pH followed by simulated moving bed (SMB) size-exclusion protein refolding, and IBs solubilization using urea, DTT and alkaline pH followed by batch dilution protein refolding. The economic performance was judged on the basis of the direct fixed capital (DFC) and the contribution to the production cost per unit of product ( $P_c$ )

---

## 6.1 Introduction

The advent of DNA recombinant technology opened the possibility to produce large quantities of valuable proteins using a variety of expression systems, including prokaryotes, yeast, filamentous fungi, insect cells, and mammalian cells. Currently, some 30 gene recombinant pharmaceuticals, with an actual market volume of U.S.\$50-60 billion, have been introduced into therapy and about 300 compounds are estimated to be in development worldwide [1].

When it comes to the production of non-glycosylated proteins, the expression system of preference is the prokaryotic one. Within this system, *Escherichia Coli* (*E. coli*) is probably the preferred organisms because of reasons such as: (1) its molecular genetics are well understood, meaning that its genome can be modified with ease; (2) high fermentation yields are attainable using high cell density cultures (HCDC) [2]; (3) it can grown using relatively inexpensive culturing procedures and (4) it possesses a high capacity to accumulate recombinant proteins (up to 80% of its dry weight) [3]. *E. coli* has, however, some limitations including: (1) it is not suitable for the production of many large, complex proteins containing disulfide bonds, and (2) it is not suitable for the production of proteins requiring post-translational modifications [2-3]. Additionally, often the over-expression of either homologous or heterologous gene sequences in *E. coli*, result in the formation of inclusion bodies (IBs). Inclusion bodies are insoluble protein aggregates that contain the recombinant protein product in an inactive, insoluble form. To obtain the recombinant product in an active (native), soluble form, the IBs must be solubilized and thereafter, the soluble denatured and reduced product must be refolded. IBs solubilization is commonly done using a solution containing a chaotrope agent to disrupt non-covalent interactions, a reducing agent to break inter-chain disulfide bonds (covalent interactions) and alkaline pH. The chaotrope agents of choice are urea and guanidine hydrochloride (GuHCL), at concentrations in the range of 6-8 M [4-5]. As reducing agents, dithiothreitol (DTT) and  $\beta$ -mercaptoethanol are probably the most commonly used, at concentrations ranging from 5-100 mM [6]. The IBs solubilization step yields a soluble denatured and reduced (D&R) protein solution which is thereafter fed to the refolding unit. In the refolding unit, the concentration of the solubilization additives (i.e., denaturant and reducing agent) is decreased to a certain minimum, by a buffer exchange mechanism, inducing protein refolding and disulfide bond formation. The buffer exchange is achieved commonly by batch dilution, mixing the concentrated D&R protein solution with refolding buffer, or using liquid chromatography (chromatographic refolding). Finally, IBs solubilization may also be achieved concurrently with protein refolding using pH-induce solubilization and refolding systems [7].

Due to its simplicity, batch dilution refolding is certainly the most widely used approach. This approach basically consists of mixing a volume of the concentrated D&R protein solution with a volume of refolding buffer, often free of chaotropes and reducing agents. To attain relatively high refolding yields, generally the total protein concentration after dilution is set to a value in the range of 0.01-0.1mgml<sup>-1</sup>[8-9]. This relatively low protein concentration primarily slows down the rates of protein aggregation, which is competing against the first-order refolding, positively influencing the refolding yield. However, operating at such low protein concentration implies significant dilutions, resulting in large processing volumes, large vessels and extensive product concentration steps. Together, all these issues have a direct impact on the process economics driving the product cost upward, setting a limit to the process scale.

The ability to successfully refold proteins, fed to the refolding reactor at a concentration of denatured and reduced protein ( $C_{f,D\&R}$ ) of up to 80 mgml<sup>-1</sup> [10], is most likely among the reasons why the application of size-exclusion chromatography (SEC) to size-exclusion refolding has gain so much popularity in the scientific community. As for batch dilution, the SEC refolding reactor is fed by the IBs solubilization unit. Protein refolding in SEC occurs, as in the case of batch dilution refolding, as a result of a buffer exchange mechanism. However, in SEC the buffer exchange occurs as a consequence of the large difference in the partition coefficients between the solubilization additives (i.e., chaotrope, reducing agent) and the soluble D&R protein. This difference results in the separation of the concentration waves as the they migrate through the size-exclusion column, resulting in a decrease of the local concentration of the solubilization additives around the protein, inducing protein folding and disulfide bond formation. Compared to batch dilution refolding, SEC refolding offers a relatively high degree of process integration, represented by the possibility of performing protein refolding and product purification, in one unit operation. SEC refolding (SECR) suffers, however, from the same limitations of conventional SEC. By conventional it is meant, SEC operated in batch and solely for the purpose of purification. These limitations include: (a) small loading volumes and (b) long columns, leading to long processing times, product dilution, large eluent consumption, large requirement of gel filtration material, ultimately affecting the volumetric productivity and the overall performance of the unit. All these limitations are however, strictly related to the mode of operation and thus, they can be overcome by changing the mode of operation of the SEC refolding reactor from batch to continuous, as recently shown [11]. This change can be achieved with the aid of Simulated Moving Bed (SMB) technology. SMB technology is a well established technology, originally introduced for the sugar and petrochemical industries in the middle of the last century [12-14]. Most likely due to its well proven advantages over conventional batch chromatography,

SMB technology has been applied to and successfully used for the purification of biomolecules [15-17], and more recently for protein refolding [11,18-19].

Compared to the previous refolding approaches (batch dilution refolding, size-exclusion refolding), pH induced refolding is a relatively simple, one single step process. In a pH-induced refolding system, protein refolding occurs concomitantly with inclusion bodies solubilization. Commonly, the pH induced system operates at relatively high alkaline pH ( $pH \geq 12$ ) [7,20-21], in order to solubilize the inclusion bodies without or with a relatively low chaotrope concentration and in the absence of a reducing agent [20]. The fact that this system, from now on referred to as pH-induced solubilization and refolding (pH\_IndSR), does not require minimum or no chaotrope agent and no reducing agent, is most likely the most attractive feature of pH\_IndSR, since it makes it a relatively cheap process. The main limitations of the pH\_IndSR systems include (1) inefficient solubilization of the inclusion bodies, since pH alone may not be able to minimize the formation of soluble aggregates [22], and (2) is a highly protein dependent system, since not all IBs can be efficiently solubilized by pH alone and most proteins, once exposed for long periods to a  $pH > 12$ , will most likely undergo irreversible chemical modifications [7].

Summarizing, up to this point four distinct process configurations have been introduced, each of them well suitable to perform inclusion bodies solubilization and protein refolding. For three of the four process options, namely size-exclusion batch refolding, SMBSEC refolding and the dilution refolding, the inclusion bodies are solubilized using a chaotrope, a reducing agent and alkaline pH. The last alternative uses a pH\_IndSR system to achieve IBs solubilization concurrently with protein refolding. This paper discusses the contribution of each of these process alternatives, to the production cost ( $P_c$ ) of an industrially relevant protein. Additionally, this work also presents: (1) a technical discussion of the various alternatives considered; (2) a comparison of the alternatives based on their economic performance, judged on the basis of (a) the annual operational cost (AOC) and (b) the direct fixed capital (DFC); (3) the impact that the choice of either dithiothreitol (DTT) or  $\beta$ -Mercaptoethanol (BME) as reducing agent for the IBs solubilization, has on the overall process economics and (4) a sensitivity analysis evaluating the contribution to the product production cost ( $P_c$ ) from the operational variables of the selected ("best") process alternative.

## 6.2 Materials and Methods

### 6.2.1 Model protein and process alternatives

The model protein considered in this study has 117 amino acids, an approximate molecular weight of  $12\,710\text{ g mol}^{-1}$ , three disulfide bonds, no free cysteines, a theoretical isoelectric point (pI) of 7.64 (based on its primary sequence) and is expressed in *Escherichia coli* in the form of inclusion bodies.

Four process alternatives, all well suitable to conduct inclusion bodies (IBs) solubilization and protein refolding, were considered in this work. These alternatives are: (1) Alternative 1: pH induced IBs solubilization and protein refolding ( i.e., pH\_IndSR); (2) Alternative 2: IBs solubilization using urea, dithiothreitol (DTT) and alkaline pH, followed by protein refolding using batch size-exclusion chromatography (SEC); (3) Alternative 3: IBs solubilization using urea, dithiothreitol (DTT) and alkaline pH, followed by protein refolding using simulated moving bed size-exclusion chromatography (SMBSEC); and (4) Alternative 4: IBs solubilization using urea, dithiothreitol (DTT) and alkaline pH, succeeded by batch dilution protein refolding. These alternatives are described in detail in section 6.2.3

### 6.2.2 Basis of calculation

The four process alternatives discussed in this paper were analyzed assuming a batch throughput of 15 kg of product per batch. The analysis considered explicitly the solubilization and refolding steps, and captured the subsequent downstream processing steps assuming an overall downstream processing yield ( $Y_{\text{DSP,OV}}$ ). Table 6.1 presents the data used to estimate the plant throughput and the inflow of inclusion bodies ( $\dot{m}_{\text{IBs}}$ ) to each of the process alternatives considered.



Table 6.1 Data used to define the basis of calculation

Item	Symbol	Value	Units	Source
Annual operating time	AOT	7200	hrs	
Plant batch time	$t_{batch}$	180	hrs	
Effective plant batch time	$t_{effective}$	72	hrs	[23]
Annual campaign	AC	100	batches/yr	
Plant throughput	$\dot{m}_{plant}$	1500	kg/yr	
Batch throughput	$\dot{m}_{batch}$	15.0	kg/batch	
Overall DSP yield	$Y_{DSP,OV}$	0.90	kg/kg	Assumed
Total protein concentration <sup>a)</sup> S&R unit of alternative 1	$C_{P,total,Alt_1}$	5.0	kg/m <sup>3</sup>	Unpublished work
Total protein concentration in the refolding unit of alternative 4	$C_{P,total,Alt_4}$	0.10	kg/m <sup>3</sup>	Assumed
<sup>a)</sup> Total D&R protein concentration in the solubilization unit of alternative 2	$C_{D\&R,Alt_2}$	2.50	kg/m <sup>3</sup>	[11]
Total D&R protein concentration in the solubilization unit of alternative 3	$C_{D\&R,Alt_3}$	3.0	kg/m <sup>3</sup>	
Reducing agent (RA) concentration after dilution (alternative 4)	$C_{RA,dilution}$	0.50	mM	Assumed
Refolding yield alternative 1	$Y_{N,Alt_1}$	0.20	kg/kg	Unpublished work
Refolding yield alternative 2	$Y_{N,Alt_2}$	<sup>b)</sup> 0.50	kg/kg	[11]
Refolding yield alternative 3	$Y_{N,Alt_3}$	0.22	kg/kg	
Refolding yield alternative 4	$Y_{N,Alt_4}$	0.50	kg/kg	Assumed
Protein fraction on the inclusion bodies	$X_{Protein}$	1.0	kg/kg	Unpublished work
Life time of the gel filtration material	$\Theta$	300	cycles	Assumed

<sup>a)</sup>D&R: Denatured and reduced; S&R: Solubilization and refolding.

<sup>b)</sup>Best SEC refolding yield reported in [11]

### 6.2.3 Estimation of the inclusion bodies mass flow ( $\dot{m}_{IBs}$ )

As previously mentioned, the batch throughput for the four alternatives was fixed to 15 kg of product per batch, for the sake of the comparison. Having a fixed bath throughput permits the estimation of the mass of inclusion bodies ( $\dot{m}_{IBs}$ ) needed to be fed to each alternative, as described up next.

#### 6.2.3.1 $\dot{m}_{IBs}$ alternative 1

$$\dot{m}_{IBs,Alt_1} = \frac{\dot{m}_{Product}}{Y_{N,Alt_1} Y_{DSP,OV} x_{Protein}} \quad (6.1)$$

where  $\dot{m}_{IBs,Alt_1}$  represents the mass of inclusion bodies per batch (kg/batch),  $\dot{m}_{Product}$  the product batch throughput (kg/batch),  $Y_{N,Alt_1}$  the refolding yield of alternative 1 (kg/kg),  $Y_{DSP,OV}$  the overall downstream processing yield (kg/kg) and  $x_{Protein}$  represents the mass fraction of protein in the inclusion bodies (kg/kg). The volume of buffer required to dissolve the inclusion bodies is estimated with Eq.(6.2).

$$V_{SolBuff\_Alt_1} = \frac{\dot{m}_{Buffer}}{\rho_{Buffer}} = \frac{\dot{m}_{IBs,Alt_1}}{C_{P,total\_Alt_1}} x_{Protein} \quad (6.2)$$

where  $V_{SolBuff\_Alt_1}$  is the volume of buffer per batch (m<sup>3</sup>/batch),  $\dot{m}_{Buffer}$  is the mass flow of buffer per batch (kg/batch),  $\rho_{Buffer}$  is the buffer density (kg/m<sup>3</sup>), and  $C_{P,total\_Alt_1}$  is the total protein concentration in the solubilization and refolding vessel unit of alternative 1 (Fig.6.1) (kg/m<sup>3</sup>).

#### 6.2.3.2 $\dot{m}_{IBs}$ alternative 2

$$\dot{m}_{IBs,Alt_2} = \frac{\dot{m}_{Product}}{Y_{DSP,OV} Y_{N,Alt_2} x_{Protein}} \quad (6.3)$$

Where  $\dot{m}_{IBs,Alt_2}$  represents the mass of inclusion bodies per batch (kg/batch) and  $Y_{N,Alt_2}$  the refolding yield of the size-exclusion refolding unit (Fig. 6.2). The volume of solubilization buffer needed per batch was calculated using Eq. (6.4)

$$V_{SolBuff,Alt_2} = \frac{\dot{m}_{IBs,Alt_2}}{C_{PD\&R,Alt_2}} x_{Protein} \quad (6.4)$$

where  $V_{SolBuff,Alt_2}$  is the volume of solubilization buffer per batch (m<sup>3</sup>/batch) and  $C_{PD\&R,Alt_2}$  is the total, denatured and reduced, protein concentration in the solubilization unit.

### 6.2.3.3 $\dot{m}_{IBs}$ alternative 3

$$\dot{m}_{IBs,Alt_3} = \frac{\dot{m}_{Product}}{Y_{DSP,OV} Y_{N,Alt_3} x_{Protein}} \quad (6.5)$$

where  $\dot{m}_{IBs,Alt_3}$  represents the mass of inclusion bodies per batch (kg/batch) and  $Y_{N,Alt_3}$  the refolding yield of the simulated moving bed size-exclusion refolding unit (Fig. 6.3). The volume of solubilization buffer, to be processed by the SMBSEC unit, was calculated using Eq. (6.4), with  $C_{PD\&R,Alt_3}$ .

### 6.2.3.4 $\dot{m}_{IBs}$ alternative 4

$$\dot{m}_{IBs,Alt_4} = \frac{\dot{m}_{Product}}{Y_{DSP,OV} Y_{N,Alt_4} x_{Protein}} \quad (6.6a)$$

where  $\dot{m}_{IBs,Alt_4}$  represents the mass of inclusion bodies per batch (kg/batch) and  $Y_{N,Alt_4}$  the refolding yield of the batch dilution refolding unit (Fig. 6.4). The volume of solubilization buffer was determined with Eq. (6.4) and using  $C_{PD\&R,Alt_4}$ . The volume of refolding buffer was calculated using Eq. (6.6b)

$$V_{RefBuff} = V_{SolBuff,Alt_4} D_{F,Refold} \quad (6.6b)$$

Where  $V_{RefBuff}$  represents the volume of buffer ( $m^3 \text{ batch}^{-1}$ ) and  $D_{F,Refold}$  represents the refolding dilution factor. The dilution factor ( $D_{F,Refold}$ ) should minimally satisfy the following two requirements: (a) the total protein concentration after dilution, should be equal to the desired total protein concentration at which refolding should be conducted and (b) the residual concentration of reducing agent (RA) after dilution, should be low enough to permit disulfide bond formation and shuffling.

## 6.2.4 Description of the process alternatives

### 6.2.4.1 Alternative 1

Inclusion bodies solubilization occurring concomitantly with protein refolding is achieved by mixing the IBs slurry with a buffer at relatively high alkaline pH, typically a  $pH \geq 12$  [7]. This method has been successfully used with several proteins expressed as inclusion bodies, including growth hormones, pro-insulin, enolase and aldolase, to mention a few [7,20-21,24-25]. Protein refolding occurs concurrently with IBs solubilization, owing to the absence of a denaturant and/or a reducing agent, which are compounds that at certain concentrations, disrupt the secondary structure and reduce disulfide bonds, respectively. The  $pH_{IndSR}$

system offers advantages including: (a) IBs solubilization is achieved without a chaotrope or using a relatively low chaotrope concentration ( $\leq 2M$ ); (b) no reducing agent is used, and (c) IBs solubilization and protein refolding are attained in one step. The limitations of the pH\_IndSR include: (a) inefficient solubilization of the IBs, represented by the inability to prevent or minimize the formation of soluble aggregates [22]; (b) proteins exposed to relatively high alkaline pH, for relatively long periods of time, may undergo irreversible chemical modifications [7]; and (c) the success of the pH\_IndSR system is highly protein dependent, as not all inclusion bodies can be solubilized by pH alone, limiting its applicability. Fig. 6.1 presents a block flow diagram (BFD) depicting the pH\_IndSR system, showing the blocks solubilization & refolding and the overall downstream processing (DSP). The overall downstream processing block should include subsequent concentration, purification and polishing steps, required to obtain the product at the desired concentration and purity. These steps were not considered explicitly in this work, instead they are represented by an overall DSP yield ( $Y_{DSP,OV}$ ).

The pH\_IndSR system is included in this work because the IBs containing the model protein can be solubilized using this approach. Our experiments have shown that using the pH\_IndSR system, with a buffer composition of 10mM NaHCO<sub>3</sub>/0.1 mM EDTA, a refolding yield ( $Y_N$ ) of 0.20 can be obtained, at a total protein concentration of around 5 mg ml<sup>-1</sup> (unpublished work).

The volume of the solubilization/refolding buffer to be processed by alternative one is 16.67 m<sup>3</sup>. This volume was determined using the information in table 6.1, together with the experimental observations and Eqs. (6.1)-(6.2).

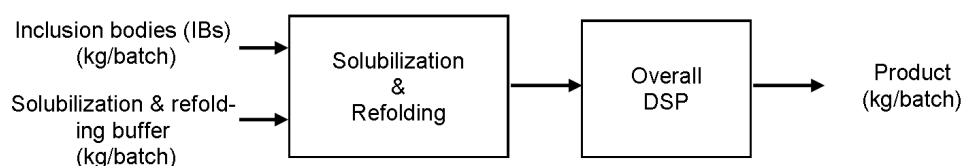


Figure 6.1: Block flow diagram of process alternative 1, pH induced inclusion bodies solubilization and protein refolding. DSP: Downstream processing

#### 6.2.4.2 Alternative 2

Inclusion bodies solubilization using urea, DTT and alkaline pH is an efficient method to solubilize inclusion bodies. The efficiency of this method, judged on the basis of the total amount of protein solubilized and the fraction of soluble aggregates formed, has been

shown to be dependent on the combined effect of the pH, the concentration of urea and the concentration of DTT [22]. As distinct from the pH\_indSR system, this solubilization method can be applied to solubilize practically any inclusion bodies, making it generally applicable. This solubilization method works by targeting the forces (non-covalent, covalent) holding the protein together in the form of an inclusion body, urea (chaotrope) disrupts the interactions based on non-covalent forces (e.g., hydrophobic, hydrogen bonding), while DTT (reducing agent) disrupts the interactions based on covalent forces (i.e., intra- and inter-chain disulfide bonds). As a result, the solubilization of the IBs using this method yields a protein solution containing the protein in a denatured and reduced (D&R) form. To refold the protein, the concentration of the solubilization agents (i.e., urea, DTT) must be decreased, using for instance size-exclusion chromatography (SEC). SEC refolding has been successfully applied to refold the model protein [11], as well as several other proteins [26].

SEC chromatographic refolding works, basically, based on a buffer exchange mechanism. This mechanism takes place as a result of the large difference in partition coefficients, between the small solubilization additives (i.e. urea, DTT) and the D&R protein. This difference results in the separation of the concentration waves, as they migrate through the column, decreasing the local concentration of the solubilization additives around the protein, inducing protein refolding and disulfide bond formation. Fig. 6.2 presents a schematic representation of alternative 2, in the form of a block flow diagram, showing the IBs solubilization unit preceding the batch SEC refolding unit(s).

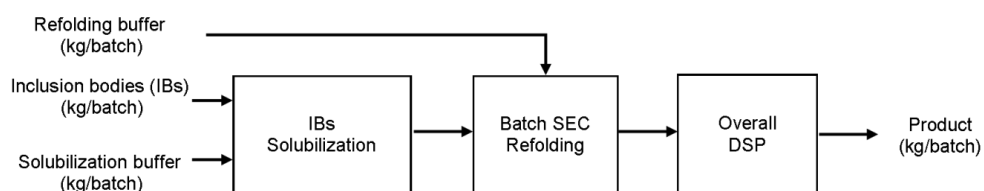


Figure 6.2 Block flow diagram of alternative 2: IBs solubilization using urea, DTT and alkaline pH, followed by batch size-exclusion chromatographic refolding

The size of the columns, the number of columns in parallel and the number of cycles per batch are determined as follows. The size of the column was fixed to a column diameter of 2 m and a bed height of 0.50 m, the biggest size commercially available to our knowledge. This dimensions result in a packed bed volume of 1.57 m<sup>3</sup>. The interstitial velocity was determined using Eq. (6.7).

$$u = \frac{L_c}{\tau_{exp}} \quad (6.7)$$

where  $u$  represents the interstitial liquid velocity (cm/h),  $L_c$  the packed bed height (cm), and  $\tau_{exp}$  the residence time (h). The residence time was estimated during the SEC batch refolding experiments as 36 min [11]. To calculate the superficial liquid velocity, and thereby the volumetric flow rate (Eq. (8)), the packed bed porosity ( $\varepsilon_b$ ) of the large scale column(s) is assumed to be the same as the laboratory scale column.

$$\dot{Q} = A_C u_0 = A_C u \varepsilon_b \quad (6.8)$$

Where  $\dot{Q}$  is the volumetric flow rate ( $m^3 h^{-1}$ ),  $A_C$  is the cross sectional area ( $m^2$ ),  $u_0$  is the superficial liquid velocity  $m h^{-1}$ , and  $\varepsilon_b$  is the bed porosity. The number of cycles ( $n_{Cycles}$ ), that a column should run to process the volume of the solubilization unit, was determined using Eq. (6.9).

$$n_{Cycles} = \frac{V_{SolBuff,Alt_2}}{n_{Col} V_{inj}} = \frac{V_{SolBuff,Alt_2}}{n_{Col} f_{inj} V_C} \quad (6.9)$$

where  $n_{Cycles}$  represents the number of cycles per batch (cycles/batch),  $n_{Col}$  the number of columns operating in parallel,  $V_{inj}$  the loading volume per column ( $m^3$ /column),  $f_{inj}$  the column loading fraction (loading volume/packed bed volume), and  $V_C$  represents the packed bed volume ( $m^3$ ).

The tradeoff between the number of columns in parallel and the number of cycles per batch, is given by the total time required to process the liquid volume in the solubilization unit ( $V_{SolBuff,Alt_2}$ ). As a guideline, the total time of this process should not be longer than the cycle time of the slowest unit operation. Otherwise, the SEC refolding unit(s) will become the new bottleneck, affecting the effective batch time and therefore the annual campaign. The total time to process  $V_{SolBuff,Alt_2}$  was estimated with Eq. (6.10).

$$t_{Op} = t_{Cycle} n_{Cycles} \quad (6.10)$$

where  $t_{Op}$  represents the time required to process the content of the solubilization unit (h/batch), and  $t_{Cycle}$  represents the cycle time of the size-exclusion refolding unit (h). The summary of the operational variables and performance indicators of alternative 2 are presented in table 6.2, including those calculated with Eqs. (6.7)-(6.10), (6.20), (6.21) and using the information in table 6.1. The columns of the SEC refolding unit(s) were packed

with Sephacryl S100, same material used during the SEC refolding experiments [11]. The Sephacryl S100 is commercialized by GE Healthcare life sciences at an average particle size of 47  $\mu\text{m}$ . The cycle time of the SEC refolding operation ( $t_{\text{Cycle}}$ ) was estimated assuming that the unit operates following a three step sequence, starting with an equilibration for 0.1 column volumes (CVs), followed by a loading for 0.02 CVs, finishing with an elution of 1CV.

Table 6.2: Summary of the operational variables and performance indicators pertaining to alternative 2

	Symbol	Units	Value
Volume of solubilization buffer	$V_{\text{SolBuff,Alt2}}$	$\text{m}^3/\text{batch}$	13.33
Packed bed volume	$V_{\text{C}}$	$\text{m}^3$	1.57
Loading fraction	$f_{\text{inj}}$	$^{\text{a)}}\text{CV}$	0.02
Volumetric flow rate	$\dot{Q}$	$\text{m}^3 \text{h}^{-1}$	2.36
Superficial velocity	$u_0$	$\text{cm hr}^{-1}$	75.00
$^{\text{b)}}\text{Maximum pressure drop}$	$\Delta P$	bar	1.58
Number of columns operating in parallel	$n_{\text{Col}}$	-	15
Number of cycles	$n_{\text{Cycles}}$	Cycles/batch	28
Cycle time	$t_{\text{Cycle}}$	h	0.75
Time required to process $V_{\text{SolBuff,Alt}_2}$	$t_{\text{Op}}$	h	21.13
$^{\text{c)}}\text{Productivity of the SEC reactor}$	$Pr_{\text{SECR}}$	$\text{kg m}^{-3} \text{h}^{-1}$	0.034
$^{\text{d)}}\text{Specific eluent consumption of the SEC reactor}$	$EC_{\text{SECR}}$	$\text{m}^3 \text{kg}^{-1}$	44.0

$^{\text{a)}}1\text{CV} = 1.57 \text{ m}^3$

$^{\text{b)}}\text{Calculated with Eq. (6.19), for } \varepsilon_b=0.3, d_p=47 \mu\text{m} \text{ and } \eta \text{ equal to the viscosity of a 4M urea solution [27]}$

$^{\text{c)}}\text{Estimated using Eq. (6.20a)}$

$^{\text{d)}}\text{Estimated using Eq. (6.21b)}$

### 6.2.4.3 Alternative 3

Batch size-exclusion refolding suffers from the same limitations as conventional size-exclusion chromatography (SEC). By conventional it is meant, SEC operated in batch and used solely for purification. These limitations include: (a) low loading volume to column volume ratio, contributing largely to product dilution and (b) long columns, resulting in long processing times, large consumption of eluent(s) and large requirement of column material. It is important to reiterate, that the aforementioned limitations are strictly related to the mode of operation. Thus, they can be overcome by changing the mode of operation of the SEC refolding reactor from batch to continuous, with the aid of for example simulated moving bed (SMB) technology. Simulated moving bed size-exclusion chromatographic

refolding (SMBSECR) has been successfully applied to refold the model protein, resulting in a refolding yield of 0.22 at a feed concentration of denatured and reduced protein of  $3 \text{ kg m}^{-3}$  [11]. Fig. 6.3 presents a block flow diagram representing alternative 3. Compared to alternative 2, alternative 3 employs SMB technology to operate the SEC refolding unit. It is important to mention that the IBs are solubilized, in this alternative, using urea, DTT and alkaline pH, just as in alternative 2.

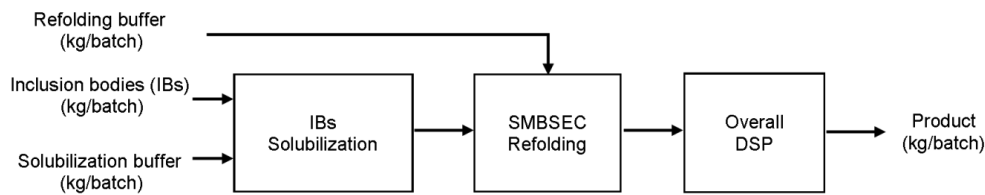


Figure 6.3: Block flow diagram of alternative 3: IBs solubilization using urea, DTT and alkaline pH, followed by protein refolding using SMBSEC refolding. SMB: Simulated Moving Bed; DSP: Downstream processing

The feed flow to the SMBSEC refolding unit ( $F$ ), the external flow rates ( $D$ ,  $E$ ,  $F$ ,  $R$ ,  $W$ ), the solid flow rate ( $Q_s$ ), the packed bed volume of a column on the SMB unit ( $V_{C,SMB}$ ), and the length of the packed bed column ( $L_{C,SMB}$ ), were determined using Eqs. (6.11), (6.13a)-(6.13e), (6.15), (6.16) and (6.17), respectively. The first operational variable determined was the feed flow to the SMBSEC refolding unit, which is given in Eq. (6.11). This flow rate is defined by the volume of the solubilization unit ( $V_{SolBuff,Alt_3}$ ), which is the volume that needs to be processed and the desired processing time ( $t_{Op}$ ). The latter, was set equal to the time needed by the batch SEC refolding unit to process the volume of the solubilization unit(s), for the sake of comparison.

$$F = \frac{V_{SolBuff,Alt_3}}{t_{Op}} \quad (6.11)$$

where  $F$  represents the feed flow rate to the SMB ( $\text{m}^3/\text{h}$ ),  $V_{SolBuff,Alt_3}$  the volume in the solubilization unit ( $\text{m}^3$ ), and  $t_{Op}$  is the desired operational time (h). The next operational variable estimated was the solid flow rate ( $Q_s$ ), which was estimated as a function of the feed flow ( $F$ ) and the flow rate ratios, also known as *m-flows*, which were used as the scale up parameters. The *m-flows* are a set of variables that interrelate the zone flow rates ( $Q_{j,SMB}$ ) and the solid flow rate, as described in Eq. (6.12a). The zone flow rates are connected with the external or operational flow rates on the SMB, via a set of continuity balances drawn around the external mixing nodes. These continuity balances are presented in Eqs. (6.13a)-



(6.13e), and they were derived for a four zones 2-2-2-2 SMB configuration, as presented in Fig. 6.3 of Freydehl et al [11]. Dividing both sides of Eqs (6.13a)-(6.13e) by  $Q_s$ , transforms the internal and external flow rates to relative flow rates, as presented in Eqs. (6.14a)-(6.14e). To obtain the expression for  $Q_s$  (Eq. (6.15)) the following steps were taken: (1) the internal zone flow rates and the external flow rates were transformed to relative flows, as presented in Eqs. (6.14a)-(6.14e); and (2) Eq. (6.15) was derived using Eqs. (6.11), (6.14c) and  $f = F/Q_s$ . The  $m$ -flows of zone 3 and 2 were taken from Freydehl et al [11]. These values were estimated using the triangle theory, constraint for linear isotherms [28], and using the distribution coefficients of urea and the D&R protein [11]. The SMBSEC unit was design with the aim of separating the D&R protein from urea and DTT, as this is the minimum criteria that must be satisfied to induce protein folding and disulfide bond formation.

$$m_{j,SMB} = \frac{Q_{j,SMB}}{Q_s} \quad (6.12a)$$

$$m_{j,SMB} = m_{j,TMB} + \frac{\varepsilon}{(1-\varepsilon)} \quad (6.12b)$$

where  $m_{j,SMB}$  represent the flow rate ratio in zone  $j$  of the SMB,  $m_{j,TMB}$  the flow rate ratio of the equivalent true moving bed (TMB) [29],  $Q_{j,SMB}$  the internal flow rate in zone  $j$  of the SMB ( $m^3/h$ ),  $Q_s$  the solid flow rate ( $m^3/h$ ), and  $\varepsilon$  represent the total bed porosity.

$$D = Q_{1,SMB} \quad (6.13a)$$

$$E = Q_{1,SMB} - Q_{2,SMB} \quad (6.13b)$$

$$F = Q_{3,SMB} - Q_{2,SMB} \quad (6.13c)$$

$$R = Q_{3,SMB} - Q_{4,SMB} \quad (6.13d)$$

$$W = Q_{4,SMB} \quad (6.13e)$$

where  $Q_{1...4,SMB}$  represent the zone flow rates of the SMB ( $m^3/h$ ),  $D$  the desorbent flow rate ( $m^3/h$ ),  $E$  the extract flow rate ( $m^3/h$ ),  $R$  the raffinate flow rate ( $m^3/h$ ), and  $W$  the waste flow rate ( $m^3/h$ ).

$$d = m_{1,SMB} \quad (6.14a)$$

$$e = m_{1,SMB} - m_{2,SMB} \quad (6.14b)$$

$$f = m_{3,SMB} - m_{2,SMB} \quad (6.14c)$$

$$r = m_{3,SMB} - m_{4,SMB} \quad (6.14d)$$

$$W = m_{4,\text{SMB}} \quad (6.14e)$$

where  $d$  represents the relative desorbent flow rate,  $e$  the relative extract flow rate,  $f$  the relative feed flow rate,  $r$  the relative raffinate flow rate,  $w$  the relative waste flow rate and  $m_{1...4,\text{SMB}}$  represent the  $m$ -flows of each zone.

$$Q_S = \frac{F}{f} = \frac{V_{\text{SolBuff,Alt}_3}}{t_{\text{op}}(m_{3,\text{TMB}} - m_{2,\text{TMB}})} \quad (6.15)$$

Eq. (6.15) makes evident that the  $m$ -flows, of sections two and three, can be used as the scale up parameters. Furthermore, Eq. (6.15) also shows that the amount of solid (i.e., gel filtration material) required, is inversely proportional to the operational time desired to process the feed and directly proportional to  $V_{\text{SolBuff,Alt}_3}$ . Having determined the expression to calculate  $Q_S$  allows, using Eq. (6.16), the calculation of the packed bed volume of a column in the SMB unit.

$$V_{\text{C,SMB}} = \frac{Q_S t_{\text{SW}}}{(1 - \varepsilon_b)} \quad (6.16)$$

where  $t_{\text{SW}}$  represents the SMB switching time (h), calculated during the laboratory scale experiment [11], and  $\varepsilon_b$  is the bed porosity. The length of each packed bed was calculated using Eq.(6.17).

$$L_{\text{C,SMB}} = \frac{V_{\text{C,SMB}}}{A_{\text{C,SMB}}} = \frac{4V_{\text{C,SMB}}}{\pi(D_{\text{C,SMB}})^2} \quad (6.17)$$

Where  $L_{\text{C,SMB}}$  represents the packed bed length (m), and  $D_{\text{C,SMB}}$  represents the column diameter (m). The diameter of the column ( $D_{\text{C,SMB}}$ ) was used as a degree of freedom, adjusted such that the resulting linear velocities in each section give a low pressure drop and a zone residence time close to the one used during the lab-scale experiment.

The summary of the operational variables and performance indicators of alternative 3 are presented in table 6.3, including those calculated with Eqs. (6.11), (6.13a)-(6.13e), (6.20), (6.21) and using the information in table 6.1.

Table 6.3: Summary of the operational variables and performance indicators pertaining to alternative 3

	Symbol	Units	Value
Volume of solubilization buffer	$V_{\text{SolBuff\_Alt}_3}$	$\text{m}^3/\text{batch}$	27.78
<sup>a)</sup> Packed bed volume of a column on the SMBSECR	$V_{\text{C,SMB}}$	$\text{m}^3$	0.31
Operational flows of the SMBSECR	D, E, F, R, W	$\text{m}^3 \text{h}^{-1}$	5.70, 3.51, 1.31, 2.45, 1.05
<sup>b)</sup> Maximum pressure drop per zone	$\Delta P_I, \dots, \Delta P_{IV}$	bar	2.52, 0.97, 1.55, 0.47
Switching time	$t_{\text{SW}}$	h	0.05
Time to process $V_{\text{SolBuff\_Alt}_3}$	$t_{\text{Op}}$	h	21.13
<sup>c)</sup> Productivity of the SMBSECR	$Pr_{\text{SMBSECR}}$	$\text{kg m}^{-3} \text{h}^{-1}$	0.32
<sup>d)</sup> Specific eluent consumption	$Ec_{\text{SMBSECR}}$	$\text{m}^3 \text{kg}^{-1}$	7.22

<sup>a)</sup> $D_{\text{C,SMB}}=1.6 \text{ m}; L_{\text{C}}=0.16 \text{ m}$

<sup>b)</sup>Calculated with Eq. (6.19), for  $\varepsilon_b=0.3$ ,  $d_p=47 \mu\text{m}$  and  $\eta$  equal to the viscosity of a 4M urea solution [27]

<sup>c)</sup>Estimated using Eq. (6.20b)

<sup>d)</sup>Estimated using Eq. (6.21a)

#### 6.2.4.4 Alternative 4

This alternative starts with the solubilization of the IBs using urea, DTT and alkaline pH, just as in alternatives two (Fig. 6.2) and three (Fig. 6.3). To refold the soluble, denatured and reduced (D&R) protein, the concentration of the solubilization additives (urea, DTT) is decreased by diluting the solution of D&R protein with refolding buffer. The refolding buffer is basically a solution with a low concentration of denaturant or no denaturant at all, no reducing agent and may contain a variety of additives, known to increase the refolding yield, at different concentrations [6,9,30]. Batch dilution refolding is the most widely used method for protein refolding and it is currently employed at an industrial scale. Typically, batch dilution refolding operates at relatively low total protein concentration ( $0.01\text{-}0.10 \text{ mgmL}^{-1}$ ) [8,31], to slow the protein aggregation rates down, favoring the refolding yield. Accordingly, a significantly high dilution is needed resulting in large processing volumes, requiring large vessels and subsequent concentration steps. Fig. 6.4 presents the BFD corresponding to alternative 4. This process alternative was not evaluated experimentally with the model protein. Thus, to include it in the comparison, the following process parameters were assumed: (a) the reducing agent concentration after dilution ( $C_{\text{RA,Refold}}$ ), (b) the total protein concentration in the refolding unit ( $C_{\text{P,total\_Alt}_4}$ ), and (c) the refolding yield ( $Y_{\text{N,Alt}_4}$ ). The values of the different assumptions are presented in table 1.

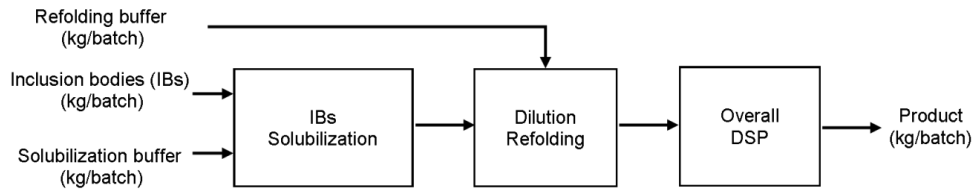


Figure 6.4 Block flow diagram of alternative 4, IBs solubilization using urea, DTT and alkaline pH, followed by protein refolding by batch dilution

The volumes of the solubilization and refolding units were calculated as follows: (1) the dilution factor ( $D_{F,Refold}$ ) was estimated using Eq. (6.18a) such that the final concentration of the reducing agent, after dilution, falls below the critical concentration of reducing agent ( $C_{RA,Critical}$ ). This critical value is important because a RA concentration close or above it may lead to a significant decrease of the refolding yield, as shown by Lanckriet et al [32]. Lanckriet et al show, using Lysozyme as model protein, that a DTT concentration close or higher than 1mM after dilution, decreased significantly the batch dilution refolding yield. For the simulations of alternative 4, a DTT concentration, after dilution, of 0.5mM was chosen. (2) The total D&R protein concentration in the solubilization unit ( $C_{P,D\&R,Solub}$ ) was estimated by fixing the desired total protein concentration in the refolding unit and using Eq. (6.18b); and (3) the volume of the solubilization unit was estimated using Eq. (6.18c) and the volume in the refolding unit was estimated with Eq. (6.18d).

$$D_{F,Refold} = C_{RA,Solub} \left( C_{RA,Refold} \right)^{-1} \quad (6.18a)$$

$$C_{P,D\&R,Solub} = C_{P,Refold} D_{F,Refold} \quad (6.18b)$$

$$V_{SolBuff,Alt_4} = \dot{m}_{IBs,Alt_4} \left( C_{P,D\&R,Solub} \right)^{-1} x_{Protein} \quad (6.18c)$$

$$V_{Refold,Alt_4} = D_{F,Refold} V_{SolBuff,Alt_4} \quad (6.18d)$$

where  $D_{F,Refold}$  represents the refolding dilution factor,  $C_{RA,Solub}$  the concentration of reducing agent used for the solubilization of the inclusion bodies (mM),  $C_{RA,Refold}$  is the concentration of reducing agent in the refolding unit (mM),  $C_{P,Refold}$  is the total protein concentration during refolding ( $\text{kg m}^{-3}$ ),  $C_{P,D\&R,Solub}$  is the total concentration of the denatured and reduced protein after the IBs solubilization ( $\text{kg m}^{-3}$ ),  $V_{SolBuff,Alt_4}$  is the volume of solubilization buffer ( $\text{m}^3 \text{ batch}^{-1}$ ) and  $V_{Refold,Alt_4}$  represent the refolding buffer volume ( $\text{m}^3 \text{ batch}^{-1}$ ).

The operational variables of alternative 4 are summarized in table 6.4, including those calculated with Eqs. (6.18a)-(6.18d) and using the information in table 6.1.

Table 6.4: Operational variables pertaining to alternative 4

	Symbol	Units	Value
Volume of solubilization buffer	$V_{\text{SolBuff,Alt}_4}$	m <sup>3</sup> /batch	6.67
Volume of refolding buffer	$V_{\text{Refold,Alt}_4}$	m <sup>3</sup> /batch	333.33
<sup>a)</sup> Number of tanks to accommodate the solubilization and refolding volume	$N_{\text{Tanks}}$	-	11
Total D&R protein concentration in the solubilization unit	$C_{f,\text{D\&R}}$	kg m <sup>-3</sup>	25.0
Refolding dilution factor	$D_{F,\text{Refold}}$	-	50

<sup>a)</sup>Estimated based on a maximum tank size of 40m<sup>3</sup> and a capacity utilization of 0.90.

### 6.2.5 Pressure drop over the packed bed

The pressure drop ( $\Delta P$ ) was estimated for both the SEC batch refolding unit(s) and for the SMBSEC refolding unit, used in alternatives two and three, respectively. The pressure drop was estimated using the Blake-Kozeny-Carman relation [33], as presented in Eq. (6.19)

$$\Delta P = \left( u \varepsilon_b \frac{L_c}{d_p^2} \eta \frac{(1 - \varepsilon_b)^2}{\varepsilon_b^3} - 180 \right) 1 \cdot 10^{-5} \quad (6.19)$$

where  $\Delta P$  represents the pressure drop over the packed bed (bar),  $u$  the interstitial velocity (cm s<sup>-1</sup>),  $\varepsilon_b$  the bed porosity,  $L_c$  the bed height,  $\eta$  the mobile's phase viscosity (Pa s) and  $d_p$  represents the average particle diameter (m). The pressure drop over the packed bed was estimated using the viscosity of the refolding buffer, which is practically the viscosity of water (*circa* 9.20·10<sup>-4</sup> Pa s), and the viscosity of a solution containing 4M urea (urea concentration used for the solubilization of the inclusion bodies) which is roughly 1.26 times the viscosity of water [27].

### 6.2.6 Performance indicators of the size-exclusion refolding units

The performance of the SEC batch and the SMBSEC refolding reactors was judged on the basis of the volumetric productivity (Pr) and the specific eluent consumption (Ec), estimated using Eqs (6.20a)-(6.20b) and (6.21a)-(6.21b), respectively.

$$Pr_{SECR} = \frac{C_{f,D\&R} V_{inj}}{t_{Cycle} V_C} Y_{N,AIt_2} \quad (6.20a)$$

$$Pr_{SMBSECR} = \frac{FC_{f,D\&R}}{N_{Col} V_C} Y_{N,AIt_3} \quad (6.20b)$$

where  $Pr_{SECR}$  and  $Pr_{SMBSECR}$  represent the volumetric productivity, of the SEC batch and the SMBSEC refolding reactors, respectively ( $\text{kg m}^{-3} \text{h}^{-1}$ ).  $C_{f,D\&R}$  represents the feed concentration of the denatured and reduced protein ( $\text{kg m}^{-3}$ ),  $t_{Cycle}$  is the cycle time of the SEC batch unit (h),  $V_C$  is the packed bed volume ( $\text{m}^3$ ),  $F$  is the feed flow to the SMBSEC reactor ( $\text{m}^3 \text{h}^{-1}$ ) and  $N_{Col}$  is the number of columns in the SMBSEC reactor.

$$Ec_{SECR} = \frac{\dot{Q} t_{Cycle}}{V_{inj} C_{f,D\&R} Y_{N,AIt_2}} \quad (6.21a)$$

$$Ec_{SMBSECR} = \frac{D}{FC_{f,D\&R} Y_{N,SMBSECR}} \quad (6.21b)$$

where  $Ec_{SECR}$  and  $Ec_{SMBSECR}$  represent the specific eluent consumption of the SEC batch and the SMBSEC refolding reactors, respectively ( $\text{m}^3 \text{kg}^{-1}$ ).  $D$  represents the desorbent flow ( $\text{m}^3 \text{h}^{-1}$ ) and  $F$  the feed flow ( $\text{m}^3 \text{h}^{-1}$ ), both pertaining to the SMBSECR reactor.

### 6.2.7 Economic indicators

The economic performance of the process alternatives considered was judged on the basis of the contribution to the product production cost ( $P_C$ ), from primarily the IBs solubilization and protein refolding operations. The contribution to the production cost was derived using (1) the annual operational cost due to raw material ( $AOC_{RM}$ ) and consumables ( $AOC_{Cons}$ ); (2) the equipment depreciation, calculated over a 5 year period and assuming a linear depreciation model; and (3) the plant annual throughput ( $P_t$ ). The contribution to the product production cost ( $P_C$ ) and the equipment depreciation, for each alternative, were estimated using Eqs. (6.22a) and (6.22b), respectively.

$$P_C = \frac{AOC_{Total}}{P_T} = \frac{AOC_{RM} + AOC_{Cons} + \Psi}{P_T} \quad (6.22a)$$

$$\Psi = \frac{DFC}{n} \quad (6.22b)$$

Where  $P_C$  represents the total contribution to the product production cost, due to the IBs solubilization and protein refolding operations ( $\text{€ kg}^{-1}$ ),  $AOC_{Total}$  is the total annual

operational cost (€ yr<sup>-1</sup>),  $AOC_{RM}$  is the annual operational cost due to raw materials (€ yr<sup>-1</sup>),  $AOC_{Cons}$  is the annual operational cost due to consumables (e.g., gel filtration material) (€ yr<sup>-1</sup>),  $n$  represents the depreciation period (yr) and  $\Psi$  represents the contribution coming from the equipment depreciation (€ yr<sup>-1</sup>). DFC represents the direct fixed capital, calculated following the approach presented by Petrides et al [23], which uses the equipment purchase cost as the basis of calculation.

### 6.2.7.1 AOC alternative 1

The annual operational cost of this alternative ( $AOC_{RM,Alt_1}$ ) is primarily derived from the consumption of the solubilization and refolding (S&R) buffer, as presented in Eq. (6.23)

$$AOC_{RM,Alt_1} = \dot{m}_{S\&RBuff} UC_{S\&RBuff} AC \quad (6.23)$$

Where  $AOC_{RM,Alt_1}$  represents the raw material costs from alternative 1, incurred during the IBs solubilization and protein refolding (USD yr<sup>-1</sup>),  $\dot{m}_{S\&RBuff}$  represents the mass flow of solubilization and refolding buffer (kg batch<sup>-1</sup>),  $UC_{S\&RBuff}$  the unit cost per kilogram of S&R buffer and AC is the annual campaign (batches yr<sup>-1</sup>).

### 6.2.7.2 AOC alternative 2

The annualized operational cost of this process alternative is composed of the operational cost of solubilization and the operational cost of the SEC batch refolding. The latter cost accounts for the consumption of refolding buffer and gel filtration material. The annual operational costs incurred due to raw materials and consumables were estimated using Eqs. (6.24a)-(6.24e).

$$AOC_{IBsSol,Alt_2} = (\dot{m}_{SolBuff} UC_{SolBuff}) AC = A \quad (6.24a)$$

$$AOC_{SECR} = (V_{GFM} R_F UC_{GFM}) AC + (\dot{m}_{Eluent} n_{Cycles} UC_{Eluent}) AC = B + C \quad (6.24b)$$

$$R_F = n_{Cycles} \Theta^{-1} \quad (6.24c)$$

$$AOC_{RM,Alt_2} = A + C \quad (6.24d)$$

$$AOC_{Cons, Alt_2} = B \quad (6.24e)$$

where  $AOC_{IBsSol,Alt_2}$  represents the annual operational costs due to the solubilization of the inclusion bodies (€ yr<sup>-1</sup>),  $AOC_{SECR}$  is the annual operational costs incurred during the batch size-exclusion refolding operation (€ yr<sup>-1</sup>),  $\dot{m}_{SolBuff}$  is the mass flow of solubilization buffer (kg batch<sup>-1</sup>),  $\dot{m}_{Eluent}$  is the mass flow of eluent (kg batch<sup>-1</sup>),  $V_{GFM}$  is the volume of gel

filtration material used ( $m^3$ ),  $R_F$  is the repacking frequency (batch<sup>-1</sup>),  $\Theta$  is the lifetime of the gel filtration material (cycles),  $n_{Cycles}$  is the number of cycles the SECR operates per batch (cycle batch<sup>-1</sup>),  $UC_{SolBuff}$  is the unit cost of the solubilization buffer ( $\text{€ kg}^{-1}$ ) and  $UC_{Eluent}$  is the unit cost of the eluent used in the SEC refolding reactor, which is the refolding buffer ( $\text{€ kg}^{-1}$ ).

### 6.2.7.3 AOC alternative 3

The various expressions used to estimate the annual operational cost of this process alternative are described in Eqs. (6.25a)-(6.25e).

$$AOC_{IBsSol,Alt_3} = (\dot{m}_{SolBuff} UC_{SolBuff}) AC = A \quad (6.25a)$$

$$AOC_{SMBSECR} = (Dt_{Op,SMB} UC_{Eluent}) AC + (V_{GFM} N_{Col} R_F UC_{GFM}) AC = B + C \quad (6.25b)$$

$$R_F = \frac{n_{Cycles,SMB}}{\Theta} = \frac{t_{Op,SMB}}{\Theta t_{Cycle,SMB}} \quad (6.25c)$$

$$AOC_{RM,Alt_3} = A + B \quad (6.25d)$$

$$AOC_{Cons,Alt_3} = C \quad (6.25e)$$

where  $D$  represents the desorbent flow to the SMBSECR unit ( $m^3 \text{ hr}^{-1}$ ),  $t_{Op,SMB}$  is the time the SMBSEC refolding unit operates in a batch (h batch<sup>-1</sup>),  $t_{Cycle,SMB}$  is the duration of one cycle of the SMBSEC refolding unit (h cycle<sup>-1</sup>),  $n_{Cycles,SMB}$  is the number of cycles the SMBSEC unit operates in a batch (cycles batch<sup>-1</sup>) and  $N_{Col}$  represents the number of packed bed columns mounted on the SMBSEC refolding unit.

### 6.2.7.4 AOC alternative 4

The annual operational cost of this alternative mainly consists of the costs incurred by the consumption of raw materials to carry the inclusion bodies solubilization and to perform the subsequent batch dilution refolding. These cost contributions are presented in Eq. (6.26).

$$AOC_{RM,Alt_4} = \underbrace{(\dot{m}_{SolBuff} UC_{SolBuff}) AC}_{IBs \text{ Solubilization}} + \underbrace{(\dot{m}_{RefBuff} UC_{RefBuff}) AC}_{Protein \text{ refolding}} \quad (6.26)$$



### 6.2.8 Raw materials and consumables prices

The prices of the chemicals and consumables are presented in Table 6.5. These prices were obtained from the on-line catalogues of Sigma-Aldrich and GE Healthcare life sciences, for the chemicals and the gel filtration material, respectively. It is important to point out though, that the prices presented on table 6.5 are catalogue prices, thus compared to bulk prices they are overpriced. A chemical that exemplifies this situation is urea, with a reported average free-on-board price (FOB), at mid February 2010, of 282 USD/ton (<http://www.icis.com/v2/chemicals/9076558/urea/pricing.html>). This price in Euros, applying an exchange rate of 1.27 USD/€, translates to 0.22 €/kg. Compared to the catalogue price, listed in table 6.5, this price is 543 fold cheaper. This difference, exemplifies the significant impact that buying chemicals in bulk amounts has on their price. It is common practice in techno-economic evaluation to use discount factors, in combination with the catalogue prices, to draw an estimation of the bulk price. Discount factors however, are in essence difficult to estimate because they depend in multiple factors including supplier, country where the purchase is made, amount purchased, type of chemical, purity of the chemical, etc. To eliminate this uncertainty, a single fixed discount factor of 90% was assumed for all the raw materials except for urea. For urea, the reported bulk price of 0.22 €/kg was used. The price of water was taken as 0.10 USD/kg [34]. The exchange rate of the euro against the US dollar (USD) was fixed at 1.27 USD/€. The discount factor of the gel filtration material Sephacryl S100 was fixed at 4 fold. This value was derived from a comparison of the catalogue price from Sephacryl S200, which sells approximately at the same price as Sephacryl S100, against a reported bulk price of 112 USD/l [33].

Table 6.5: List of raw materials and consumables prices (prices at 2010)

Item	Unit cost	Units	Purity
Ethylenediaminetetraacetic acid (EDTA)	99.00	€/kg	≥99.0%
Sodium hydrogen carbonate	44.80	€/kg	≥99.5%
Urea	119.60	€/kg	Meets <sup>a)</sup> USP testing specifications
Dithiothreitol (DTT)	16 880	€/kg	≥99.0
2-Mercaptoethanol (BME)	71.67	€/kg	≥99.0
Sodium Chloride	56.10	€/kg	Meets USP testing specifications
Tris	78.90	€/kg	Meets EP and USP specifications
Sephacryl S100	686.66	€/l	<sup>b)</sup> n.a.

<sup>a)</sup>USP: United States Pharmacopoeia; EP: European Pharmacopoeia

<sup>b)</sup>n.a.: Not applicable

### 6.2.9 Process simulation tool

All the calculations were done on Microsoft Excel 2007. The purchase equipment cost, including tanks, pumps and chromatography skids were estimated from the database of SuperPro Designer (SPD) version 7.50.

## 6.3 Results and Discussion

### 6.3.1 Inclusion bodies mass flow and solubilization volume

The mass flow of inclusion bodies ( $\dot{m}_{IBS}$ ) required by each process alternative to satisfy the required batch throughput ( $\dot{m}_{Product}$ ), is primarily determined by the refolding yield ( $Y_N$ ), the overall downstream processing yield ( $Y_{DSP,OV}$ ) and the mass fraction of protein in the inclusion bodies ( $x_{Protein}$ ), as shown in Eqs. (6.1), (6.3) and (6.6a). It is important to mention that  $Y_{DSP,OV}$  and  $x_{Protein}$  were kept constant for all process alternatives because they do not depend significantly on the solubilization and refolding operations.  $Y_{DSP,OV}$  depends mainly on the efficiency of the downstream purification and polishing steps, assumed to be equal for the process alternatives evaluated. And  $x_{Protein}$  depends primarily, on (1) the efficiency of the protein expression happening during the fermentation operation and (2) the efficiency of the inclusion bodies washing procedure.  $Y_N$  on the other hand, strongly depends on the refolding operation, thus it varies for each process alternative. The amount of inclusion bodies required by each process alternative is presented in table 6.6. Notice that  $\dot{m}_{IBS}$  increases as  $Y_N$  decreases, this is explained by Eqs. (6.1), (6.3) and (6.6a), which show that  $\dot{m}_{IBS}$  is inversely proportional to  $Y_N$ . A higher inclusion bodies input implies a higher fermentation and primary recovery capacity. Though not included in this analysis, the contribution of the upstream operations (fermentation and primary recovery) to the operational costs of a recombinant protein produced in *E. coli* have been shown to be relatively not significant, contributing to roughly 3.1% of the production costs [23].

Table 6.6, additionally presents the volume of solubilization buffer ( $V_{SolBuff}$ ) required to solubilize the inclusion bodies, for each process alternative. These values were calculated using Eqs. (6.2), (6.4) and (6.18c). It is important to point out, that for alternatives 1 to 3,  $V_{SolBuff}$  depends mainly on the mass flow of inclusion bodies ( $\dot{m}_{IBS}$ ) and the target total protein concentration in the solubilization unit ( $C_{P,total}$  or  $C_{PD\&R}$ ); whereas  $V_{SolBuff}$  for alternative 4 depends not only on  $\dot{m}_{IBS}$  but also on the concentration of reducing agent (RA) after dilution ( $C_{RA,Refold}$ ), the RA concentration used during IBs solubilization ( $C_{RA,Solub}$ ) and the target total protein concentration in the refolding unit ( $C_{P,Refold}$ ). In mathematical notation, the difference between the solubilization volume of alternative 4 (taken as a reference) and the solubilization volume of alternatives 1, 2 and 3 is explained in Eq. (6.27). Notice that in Eq. (6.27)  $C_{P,Solub\_Alt\_1...3} \equiv C_{P,total\_Alt\_1} \equiv C_{PD\&R\_Alt\_2,3}$ .

$$\frac{V_{\text{SolBuff,Alt}_{1..3}}}{V_{\text{SolBuff,Alt}_4}} = \underbrace{\left( \frac{Y_{N,\text{Alt}_4}}{Y_{N,\text{Alt}_{1..3}}} \right)}_{\text{Relative refolding Efficiency}} \underbrace{\left( \frac{C_{P,\text{Refold}} C_{RA,\text{Solub}}}{C_{RA,\text{Refold}}} \right)}_{\text{Parameters belonging to Alternative 4}} \underbrace{\left( \frac{1}{C_{P,\text{Solub\_Alt}_{1..3}}} \right)}_{\text{Parameter belonging to Alternatives 1,2 and 3}} \quad (6.27)$$

Table 6.6: Inclusion bodies inflow to satisfy the plant throughput

Process alternative	$Y_N$	$C_{P,D\&R,Solub}$ (kg m <sup>-3</sup> )	$V_{\text{SolBuff}}$ (m <sup>3</sup> batch <sup>-1</sup> )	$\dot{m}_{\text{IBs}}$ (kg batch <sup>-1</sup> )
1	0.20	5.0	16.67	83.33
2	0.50	2.50	13.33	33.33
3	0.22	3.0	25.25	75.76
4	0.50	5.0	6.67	33.33

### 6.3.2 Inclusion bodies solubilization

Inclusion bodies solubilization is an important step owing to reasons such as (a) it directly influences the performance of the protein refolding step, as the output of the solubilization unit is the input of the refolding unit and (b) the general solubilization approach is a relatively costly step, as it uses the chosen chaotrope (e.g., urea, GndHCL) in relatively large quantities (depending on the amount of  $V_{\text{SolBuff}}$  needed) and it uses relatively costly reducing agents (e.g., DTT, BME). By general solubilization it is meant an approach that can virtually be used to solubilize any IBs.

To recapitulate, in the presented work two alternatives to solubilize the inclusion bodies (IBs) are taken into consideration and these are (1) IBs solubilized solely using a relatively high alkaline pH ( $\geq 12$ ), used in alternative 1 and (2) IBs solubilized using urea and DTT or BME, used in alternatives 2, 3 and 4. 2-Mercaptoethanol (BME) was introduced in this study as an alternative to DTT due to its relatively low cost price (table 6.5). To compensate for the difference in reducing power between BME and DTT (BME less potent than DTT), BME was introduced in the solubilization buffer at a fourfold higher concentration. Additionally, it is assumed that having 100mM BME in the solubilization buffer results in the same solubilization efficiency, judged in terms of the fraction of denatured and reduced soluble monomer, as having 25mM DTT. The impact on the unit cost of the solubilization buffer, of having DTT at 25mM or BME at 100mM, is a decrease in the cost price of roughly 37 fold, when using BME instead of DTT. Additionally, the cost contribution of the reducing agent to the cost of the solubilization buffer, changed from 98% using DTT, to 32% using BME. The unit cost price of the solubilization buffer, as well as its composition is presented in table 6.7, for all the process alternatives considered.

Table 6.7 also presents the cost of the solubilization buffer for the pH induced solubilization approach. From the data it is clear that solubilization of IBs using pH alone is a relatively cheap process, own to the lower cost of the solubilization buffer. The relatively low cost of the buffer is explained by the absence of especially reducing agents. Despite the economic advantage, it is important to reiterate that (a) pH induced solubilization is not a generally applicable method, as not all IBs can be solubilized by pH alone, (b) exposure of proteins to a high alkaline pH, for long period of time, may lead to irreversible chemical modifications [7] and (c) IBs solubilization using alkaline pH is less efficient, in terms of the fraction of soluble D&R monomer formed after IBs solubilization, than IBs solubilization using a chaotrope, a reducing agent and a relatively lower pH [22].

Table 6.7 Solubilization buffer: Unit cost and buffer composition

Process Alternative	Buffer I.D.	Buffer Composition	<sup>a)</sup> UC (€ kg <sup>-1</sup> )
1 (no DTT nor BME)	Buffer A	10mM NaHCO <sub>3</sub> / 0.1mM EDTA	0.05
2_DTT, 3_DTT, 4_DTT	Buffer B	4M Urea/ 25mM DTT/ 10mM NaHCO <sub>3</sub> / 0.1mM EDTA	6.40
2_BME, 3_BME, 4_BME	Buffer C	4M Urea/ 100mM BME/ 10mM NaHCO <sub>3</sub> / 0.1mM EDTA	0.17

<sup>a)</sup>UC: Unit cost

### 6.3.3 Direct fixed capital (DFC)

The direct fixed capital (DFC), of all the process alternatives considered, was determined based on the equipment purchase cost (PE), using the approach presented by Petrides et al [23]. This approach consists of estimating the total plant direct cost (TPDC) using the equipment purchase cost and a set of multipliers. Then, the total plant indirect cost (TPIC) is estimated using TPDC. After that, the total plant cost (TPC) is estimated as  $TPC = TPDC + TPIC$ . Finally, DFC is estimated by adding to the TPC, the contractor's fee and the contingency. The latter two amounts were determined from the TPC. The purchase equipment cost was determined as follows. The purchase costs of the stirred tanks and the packed beds (used for SEC batch refolding) were estimated from the equipment data base of SuperPro Designer version 7.5. The equipment purchase cost concerning the SEC SMB unit, including the pumps, columns and the SMB valve, were obtained via personal communication with the company XENDO ([www.xendo.com](http://www.xendo.com)). This company is located in the Netherlands and has experience commercializing and installing SMB systems. Table 6.8 presents the major pieces of equipment and the corresponding purchase equipment cost, for each process alternative. Fig. 6.5 presents the direct fixed capital (DFC) for each process alternative. From the data it is evident that alternative 1 has the lowest DFC, followed, in increasing order, by alternative 3, alternative 4 and alternative 2. Alternative 1 has the

lowest DFC because it only requires a single tank to perform solubilization and refolding, resulting in a relatively low purchase and equipment costs (PE). The relatively high DFC of alternative 2 (SEC batch refolding), results from the high PE owing to the relatively high number of SEC columns that need to operate in parallel, to process the volume of the solubilization tank(s). However when an SMB unit is used for SEC on-column refolding, as in alternative 3, the DFC decreases roughly 5.8 fold compared to alternative 2.

Table 6.8: Major pieces of equipment and purchase equipment cost (PE)

Alternative 1				
Unit	Description	Quantity	Unit Cost (€ unit <sup>-1</sup> )	Cost (€)
Solubilization and refolding tank	$V_{\text{Sol\_Tank\_Alt}_1} 22 \text{ m}^3$	1	600 000	600 000
			$PE_{\text{Alt}_1}$	600 000
Alternative 2				
Unit	Description	Quantity	Unit Cost (€ unit <sup>-1</sup> )	Cost (€)
Solubilization tank	$V_{\text{Sol\_Tank\_Alt}_2} 18 \text{ m}^3$	1	571 000	571 000
Packed bed columns	$D_C 2\text{m}; L_C 0.50 \text{ m}; V_C 1.50 \text{ m}^3$	15	961 000	14 415 000
			$PE_{\text{Alt}_2}$	14 986 654
Alternative 3				
Unit	Description	Quantity	Unit Cost (€ unit <sup>-1</sup> )	Cost (€)
Solubilization tank	$V_{\text{Sol\_Tank\_Alt}_3} 33 \text{ m}^3$	1	667 000	667 000
SMBSEC unit pumps	-	4	50 000	200 000
SMB valve	-	1	500 000	500 000
SMB columns	$V_C 0.28 \text{ m}^3$	8	125 000	1 000 000
			$PE_{\text{Alt}_3}$	2 367 000
Alternative 4				
Unit	Description	Quantity	Unit Cost (€ unit <sup>-1</sup> )	Cost (€)
Solubilization tank	$V_{\text{Sol\_Tank\_Alt}_4} 9 \text{ m}^3$	1	528 000	528 000
Refolding tank	$V_{\text{Sol\_Tank\_Alt}_4} 40 \text{ m}^3$	10	706 000	7 060 000
			$PE_{\text{Alt}_4}$	7 528 000

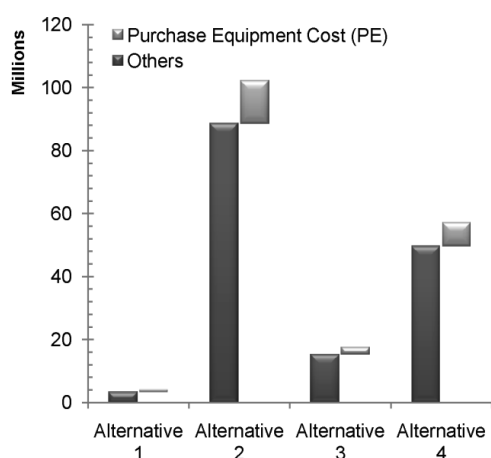


Figure 6.5 Direct fixed capital (DFC). The PE corresponds to the values reported in table 8. The data series Others includes: (a) the total plant direct cost (TPDC) accounting for installation (0.51xPE), process piping (0.10xPE), instrumentation (0.35xPE), insulation (0.03xPE), electrical (0.10xPE), buildings (0.85xPE), yard improvements (0.15xPE) and auxiliary facility costs (0.60xPE); (b) the total plant indirect costs (TPIC) accounting for engineering (0.25xTPDC) and construction (0.35xTPDC) costs and (c) contractor's fee (0.05xTPC) and contingency (0.10xTPC) costs. TPC: Total plant direct cost

### 6.3.4 Contribution to the production cost ( $P_C$ ) due to IBs solubilization and protein refolding operations

The contribution to the production cost ( $P_C$ ) accounts for the costs incurred due to raw materials, consumables (i.e., gel filtration material) and equipment depreciation. This economic indicator ( $P_C$ ) was calculated using Eq. (6.22) and the assumptions presented in table 6.1. The magnitude of  $P_C$  is interpreted as follows. A low  $P_C$  means that the process alternative is economically favorable and vice versa. This is because a low  $P_C$  increases the gross profit, providing that the cost contribution from the purification and polishing steps remains constant.

Fig 6.6 shows the cost contribution ( $P_C$ ) for the various alternatives (Fig. 6.6A) and the breakdown of the magnitude of  $P_C$  into its various constituents (Fig. 6.6B). From the data in figure 6.6A the following observations are evident: (1) alternative 1 has the lowest  $P_C$  compared to alternatives 2 to 4; (2) Alternative 3 has a significantly lower  $P_C$  compared to alternative 2; (3) substituting DTT at 25mM for BME at 100mM in the solubilization buffer, decreases the  $P_C$  of alternatives 2, 3 and 4, as these use a reducing agent for the solubilization of the IBs and (4) the  $P_C$  of alternative 3 is comparable to the  $P_C$  of alternative 4.

The lower  $P_C$  of alternative 1 is explained by the following reasons: (a) a relatively cheap IBs solubilization step, as a result of the low cost of the solubilization buffer (section 6.3.2) and (b) a low purchase equipment cost, as both solubilization and refolding can be done in the same unit. Figure 6.6B shows that 90% of the  $P_C$  of alternative 1 comes from equipment



depreciation and the remainder 10% comes from the raw materials. Overall, alternative 1 is economically very attractive but it is not applicable to all proteins, due to reasons discussed in previous sections, which is not the case for either alternatives 2, 3 and 4.

Figure 6.6A also shows that alternative 2 has a significantly higher  $P_C$  than alternative 3. Actually, when BME is used as reducing agent alternative 3 is roughly 5.6 fold more economical, based on the magnitude of  $P_C$ , than alternative 2. When DTT is used, alternative 3 is approximately 2.7 fold more economical. The fact that operating the SEC refolding reactor in batch mode is less economic than operating it in continuous mode (SMBSEC) is reasonably explained by the following reasons: (a) the batch system has a high purchase equipment cost, as multiple packed bed columns (15 in this case) operating in parallel are required to process the volume of the solubilization unit, in the desired time and (b) the batch system employs large columns, requiring a large quantity of gel filtration material. The positive impact of the SMB technology is mainly explained by the relatively low requirement of gel filtration material, owing to the relatively small packed bed units employed by the SMBSEC unit, driving the annual consumables cost down. The contribution of the gel filtration material to the magnitude of  $P_C$  is presented in figure 6.6B.

Concerning the solubilization of the inclusion bodies, the data presented in Figs. 6.6A-B show that the major contributor to the  $P_C$  is the reducing agent (DTT or BME), rather than urea (chaotropic agent). Moreover, the data in Fig. 6.6B indicates that substituting 25mM DTT for 100mM BME decreases the  $P_C$  of alternatives 3 and 4, as in these alternatives DTT contributes significantly to the magnitude of  $P_C$ .

Overall, from the data in Figs. 6.6A-B it can be concluded that the most economically attractive alternatives, organized in increasing value of  $P_C$ , are: alternative 1, followed by alternatives 3\_BME and 4\_BME. As mentioned before, alternative 1 is highly protein dependent (not all IBs can be solubilized solely by alkaline pH) and thus it is not considered further. The coming discussion focuses on alternatives 3\_BME and 4\_BME, generally applicable processes. These alternatives are generally applicable because all IBs can be solubilized using a reducing agent (i.e., BME) together with a chaotropic agent (i.e., urea) and alkaline pH. Since both alternatives (3 and 4) use a similar solubilization approach, the following discussion compares their corresponding protein refolding operations. In alternative 4, protein refolding is achieved by diluting the D&R protein solution with refolding buffer (section 6.2.4.4). The dilution factor is dictated by the final total protein concentration in the refolding vessel and the final reducing agent concentration, resulting in a large dilution factor (50 fold in this case). Accordingly, a significant number of large tanks

are required by the refolding operation, resulting in a PE roughly 3 fold higher than the PE of alternative 3.

In alternative 3, the protein refolding is achieved by the separation of the solubilization additives from the D&R soluble protein. This separation is attained by size-exclusion chromatography, operating in a continuous mode with the aid of SMB technology. This unit is fed by a relatively concentrated D&R protein solution (table 6.1) and it delivers in the raffinate stream (R) the protein, including the native protein and in the extract stream (E) the solubilization additives [11]. The separation of the protein from the solubilization additives gives alternative 3 a technical advantage over alternative 4. This is because on a typical process, using alternative 4 [23], the refolding reactor(s) are followed by a battery of diafiltration units, to perform a buffer exchange, separating the protein from the remainder amount of the solubilization additives. Although the diafiltration operation was not accounted for in the estimation of the PE and the  $P_C$  of alternative 4, should it was it will drive both PE and  $P_C$  upward. Accordingly, the current  $P_C$  difference of 1.11 fold, between alternatives 4\_BME and 3\_BME, will increase; making alternative 3\_BME a more economically attractive option.

Overall, it can be concluded that alternative 3\_BME is an attractive option both technically and economically. Thus, it is interesting to study how the economic performance of this alternative is influenced by changes in key process variables. To address this question a sensitivity analysis was performed and it is presented in section 6.3.5.

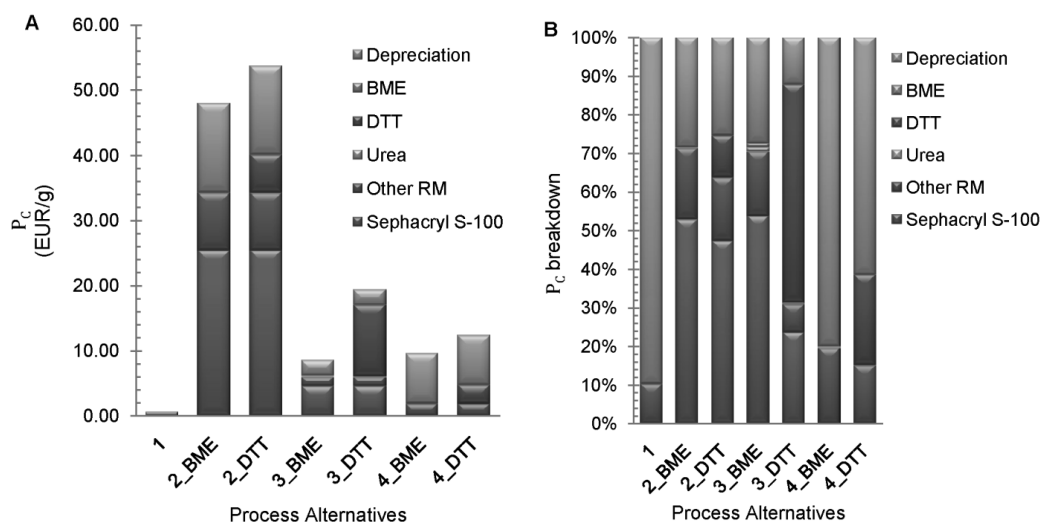


Figure 6.6 Contribution to the production cost due to IBs solubilization and protein refolding operations. (A)  $P_C$  for alternatives 1 to 4, including the option of substituting DTT for BME as reducing agent to achieve IBs solubilization. (B) Breakdown of the cost contribution, showing the relative contributions of the raw materials (i.e., DTT, BME, urea, other RMs), equipment depreciation and consumables to the magnitude of  $P_C$ . RMs: Raw materials

### 6.3.5 Sensitivity analysis

#### 6.3.5.1 Scenario 1: Impact of $C_{PD\&R}$ and $\Theta$ on $P_C$

This scenario fixes the purchase equipment cost (PE) and the refolding yield ( $Y_N$ ), whereupon  $P_C$  is calculated solely as a function of (a) the total D&R protein concentration in the solubilization unit ( $C_{PD\&R}$ ) and (b) the life time of the gel filtration material ( $\Theta$ ). Because PE is fixed, this scenario is actually considering the increase or decrease of  $P_C$  due to raw materials and consumables, as the contribution of the depreciation to  $P_C$  is fixed.  $C_{PD\&R}$  and  $\Theta$  were chosen as sensitivity parameters because both have a significant impact on the magnitude of  $P_C$ . Increasing  $C_{PD\&R}$  decreases the volume to be processed by the SMBSEC refolding unit, decreasing the number of cycles per batch. Accordingly, the consumption of eluent(s) and the repacking frequency of the SMBSEC unit decreases. On the other hand,  $\Theta$  has a direct influence in the costs because it determines the repacking frequency (Eq. (6.24c)). The repacking frequency ( $R_F$ ) is defined as the number of times the columns should be repacked and it dictates the annual costs of the gel filtration material (Eq. (6.24e)).

Figs. 6.7A-B present the contour plots of  $P_C$  as a function of  $C_{PD\&R}$  and  $\Theta$  for alternatives 3\_BME and 3\_DTT, respectively. As anticipated, the contribution of the production cost decreases as both variables increases. It is important to mention that  $\Theta$  for any given

chromatographic stationary phase is product specific and depends on a variety of factors including (a) resin type, (b) concentration and nature of the feed, (c) regeneration procedures, (d) storage conditions, among others [35]. And thus it is empirically determined from resin life time studies. Such studies could be guided using the information presented in Figs. 6.7A-B, as from these figures a target  $\Theta$  can be defined based on a desired  $P_C$ . Lastly, it is important to point out that changes in  $C_{PD\&R}$  will most likely have an impact on the refolding yield, as this is a function of the former process parameter. However, if the dependency of  $Y_N$  on  $C_{PD\&R}$  is known this could be readily incorporated in the economic analysis.

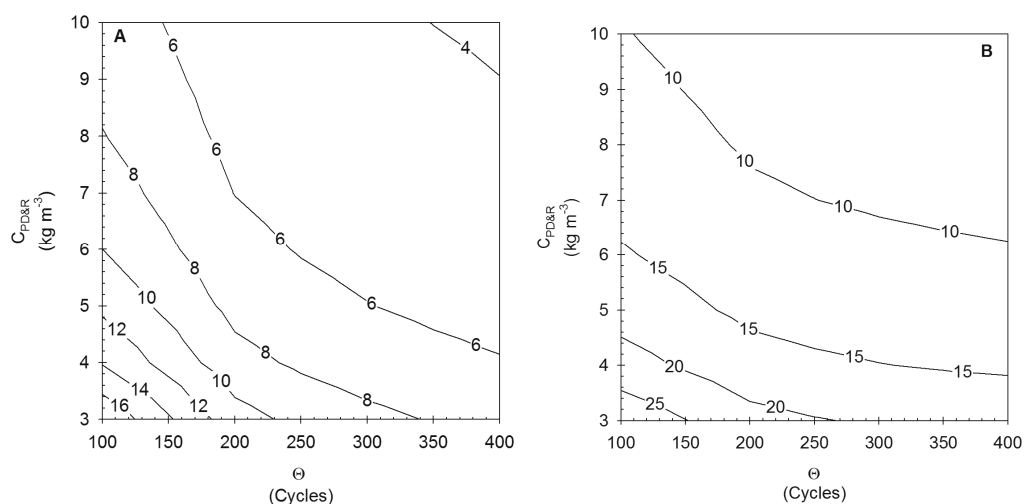


Figure 6.7  $P_C$  ( $\text{€ g}^{-1}$ ) contour plot as a function of the life time of the gel filtration material ( $\Theta$ ) and the concentration of denatured and reduced protein in the solubilization tank(s) ( $C_{PD\&R}$ ). (A) Contour plot for alternative 3\_BME. (B) Contour plot of alternative 3\_DTT. The contour lines represent  $P_C$

### 6.3.5.2 Scenario2: Impact of $\Theta$ and $f$ on $P_C$

This scenario keeps the total D&R protein concentration ( $C_{PD\&R}$ ) in the solubilization tank constant and assumes a constant refolding yield ( $Y_N$ ). These two assumptions permitted the evaluation of the effect that the relative feed flow ratio ( $f$ ) (Eqs. (6.14c), (6.15)) and the life time of the gel filtration material ( $\Theta$ ) have on  $P_C$ . In essence, this scenario presents a possibility to influence the magnitude of  $P_C$  while keeping the operational variables pertaining to the IBs solubilization constant.

The relative feed flow ratio  $f$  (Eq. (6.14c)) is an important variable because it determines the solid flow rate (Eq. (6.15)), which in turn determines the amount of gel filtration material

required by the SMBSEC (Eq.(6.16)). As  $f$  increases, the volumetric productivity of the SMBSEC refolding unit ( $P_{rSMBSEC}$ ) increases whereas the specific solvent consumption ( $EC_{SMBSEC}$ ) decreases. Accordingly, as  $f$  increases  $P_C$  decreases rendering alternative 3\_BME more economically attractive. Fig. 6.8B presents the contour plot of  $P_{CAIT\_3BME}$  as a function of  $f$  and  $\Theta$ . Notice how  $P_C$  decreases as  $f$  and  $\Theta$  increases. The dependency of  $P_C$  on  $\Theta$  was explained in section 6.3.5.1. It is important to point out that the value that  $f$  can take is restricted by the triangle of complete separation (Fig. 6.8A). This triangle was built using the distribution coefficients of urea and the D&R protein on the Sephacryl S100 gel filtration media. Values of  $f$  within the triangle lead to the complete separation of the D&R protein from urea (as well as the reducing agent), inducing disulfide bond formation and protein folding [11].

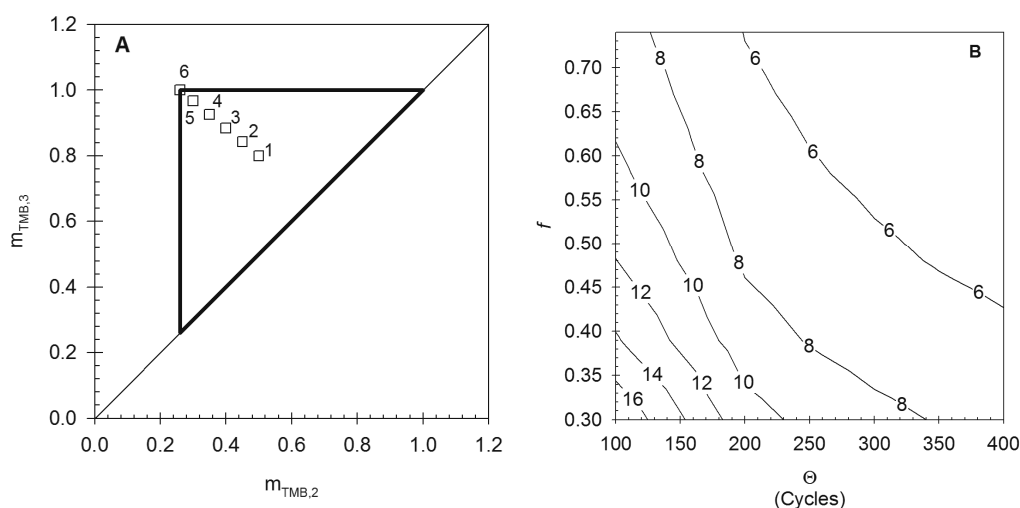


Figure 6.8 (A) SMB triangle of complete separation, constraint for linear isotherms. The triangle was built such that the SMBSEC refolding unit completely separates the solubilization additives from the D&R protein, inducing protein refolding. Points 1 to 6 are assumed operational points where the complete separation should be attainable. The theoretical operational points have the following coordinates: (1) (0.50, 0.80), (2) (0.45, 0.84), (3) (0.40, 0.88), (4) (0.35, 0.93), (5) (0.30, 0.97), (6) (0.26, 1.00). (B)  $P_C$  contour plot as a function of the relative SMBSEC feed flow ( $f$ ) and the life time of the gel filtration material ( $\Theta$ ). The contour lines represent  $P_C$

#### 6.4 Conclusions

This work presented a comparison, based primarily on their economic performance, of four distinct process alternatives, all suitable for the inclusion bodies solubilization and subsequent protein refolding of a recombinant protein, produced in *E. coli* and expressed in inclusion bodies. These four alternatives include: (a) a pH induced IBs solubilization and refolding system (pH\_IndSR); (b) IBs solubilization using urea and DTT or BME, followed by protein refolding by batch size-exclusion chromatography; (c) IBs solubilization using urea and DTT or BME, followed by simulated moving bed (SMB) SEC refolding and (d) IBs solubilization using urea and DTT, followed by batch dilution refolding. This work showed: (1) pH\_IndSR system is a relatively economical process, which is explained by the low cost of the solubilization process. It was also pointed out in this work that the major limitation of the pH is most likely its protein dependency, as not all IBs can be solubilized solely by alkaline pH; (2) Substituting BME for DTT is an attractive alternative as it decreases the product cost contribution from the IBs solubilization; (3) protein refolding by size-exclusion chromatography becomes economically attractive by changing the mode of operation of the chromatographic reactor from batch to continuous using SMB technology. Actually, changing the mode of operation led to a 5.8 fold and a 2.7 fold decrease in production cost ( $P_c$ ), when BME or DTT were used as reducing agent for the solubilization of the IBs, respectively; and (4) compared to batch dilution refolding, the novel SMBSEC refolding operation is economically attractive when BME is used as reducing agent during IBs solubilization. It is important to mention, that the economic performance of the SMBSEC operation compared to the batch dilution operation is expected to improve if the cost of the dialfiltration operation (operation required to obtain the refolded product free of solubilization additives) is added to the cost of batch dilution refolding. Lastly, this work showed using a sensitivity analysis how the economic performance of the SMBSEC operation can be further improved. This information can be used to guide future research, helping it to achieve concurrently economic and scientific targets.

## 6.5 References

- [1] F.R. Schmidt, *Appl. Microbiol. Biotechnol.* 65 (2004) 363.
- [2] J.H. Choi, K.C. Keum, S.Y. Lee, *Chem. Eng. Sci.* 61 (2006) 876.
- [3] A.L. Demain, P. Vaishnav, *Biotechnology Advances* 27 (2009) 297.
- [4] K. Tsumoto, D. Ejima, I. Kumagai, T. Arakawa, *Protein. Expres. Purif.* 28 (2003) 1.
- [5] A. Mukhopadhyay, in *Biotreatment, Downstream Processing and Modelling*, 1997, p. 61.
- [6] L.D. Cabrita, S.P. Bottomley, *Biotechnology annual review* 10 (2004) 31.
- [7] E.D.B. Clark, *Curr. Opin. Biotechnol.* 12 (2001) 202.
- [8] A. Jungbauer, W. Kaar, *J. Biotechnol.* (2007) 587.
- [9] E.D.B. Clark, *Curr. Opin. Biotechnol.* 9 (1998) 157.
- [10] B. Batas, J.B. Chaudhuri, *Biotechnol. Bioeng.* 50 (1996) 16.
- [11] E.J. Freydell, Y. Bultink, S.H. Van Hateren, L.A.M. van der Wielen, M. Eppink, M. Ottens, *Chem. Eng. Sci.* 65 (2010) 4701.
- [12] D.B. Broughton, *Chem. Eng. Prog.* 64 (1968) 60.
- [13] D.B. Broughton, C.G. Gerhold, in U.S. Patent 2985589, 1961.
- [14] D.B. Broughton, R. Neuzil, J. Pharis, C. Brearley, *Chem. Eng. Prog.* 66 (1970) 70.
- [15] S. Lucena, P. Rosa, L. Furlan, C. Santana, in *Engineering and Manufacturing for Biotechnology*, 2002, p. 325.
- [16] Y. Xie, S. Mun, J. Kim, N.-H.L. Wang, *Biotechnol. Prog.* 18 (2002) 1332.
- [17] D.A. Horneman, M. Ottens, L.A.M.v.d. Wielen, J.T.F. Keurentjes, *AIChE J.* 53 (2007) 1441.
- [18] B.-J. Park, C.-H. Lee, Y.-M. Koo, *Korean J. Chem. Eng.* 22 (2005) 425.
- [19] B.-J. Park, C.-H. Lee, S. Mun, Y.-M. Koo, *Process Biochem.* 41 (2006) 1072.
- [20] S.M. Singh, A.K. Upadhyay, A.K. Panda, *Journal of Chemical Technology & Biotechnology* 83 (2008) 1126.
- [21] A.K. Patra, R. Mukhopadhyay, R. Mukhija, A. Krishnan, L.C. Garg, A.K. Panda, *Protein. Expres. Purif.* 18 (2000) 182.
- [22] E.J. Freydell, M. Ottens, M. Eppink, G.v. Dedem, L.v.d. Wielen, *Biotechnol. J.* 2 (2007) 678.
- [23] D. Petrides, E. Sapidou, J. Calandranis, *Biotechnol. Bioeng.* 48 (1995) 529.
- [24] R.H. Khan, K.B.C.A. Rao, A.N.S. Eshwari, S.M. Totey, A.K. Panda, *Biotechnol. Prog.* 14 (1998) 722.
- [25] K. Raina, A.K. Panda, M.M. Ali, G.P. Talwar, *Protein. Expres. Purif.* 37 (2004) 8.
- [26] A. Jungbauer, W. Kaar, R. Schlegl, *Curr. Opin. Biotechnol.* 15 (2004) 487.
- [27] K. Kawahara, C. TANFORD, *The Journal of Biological Chemistry* 241 (1966) 3228.
- [28] G. Storti, M. Mazzotti, M. Morbidelli, S. Carrà, *AIChE J.* 39 (1993) 471.
- [29] F. Charton, R.-M. Nicoud, *J. Chromatogr. A.* 702 (1995) 97.
- [30] B. Fischer, I. Sumner, P. Goodenough, *Biotechnol. Bioeng.* 41 (1993) 3.
- [31] G. Lee, D. Cooney, A. Middelberg, W. Choe, *Bioprocess Biosyst. Eng.* 29 (2006) 73.
- [32] H. Lanckriet, A.P.J. Middelberg, *J. Chromatogr. A.* 1022 (2004) 103.
- [33] G. Sofer, L. Hagel, *Handbook of Process Chromatography: A Guide to Optimization, Scale-Up and Validation*, Academic Press, San Diego, 1997.
- [34] D.P. Petrides, A. Koulouris, P.T. Lagonikos, *Pharmaceutical Engineering* 22 (2002) 1.
- [35] C. Jiang, J. Liu, M. Rubacha, A.A. Shukla, *J. Chromatogr. A.* 1216 (2009) 5849.

# 7

## Outlook

---



The aim of the work presented in this thesis was, essentially, to look for and to evaluate solubilization and refolding strategies that could potentially be part of new IBs processing schemes. The areas covered by this work were: (1) Inclusion bodies solubilization; (2) Ion-exchange chromatographic refolding; (3) Size-exclusion chromatographic refolding; (4) Model development for size-exclusion protein refolding; (5) Simulated Moving Bed SEC protein refolding and (6) a techno-economic evaluation. The results of the various investigations are presented in Chapters 2-5, however while generating the presented data new questions were formulated. The following sections contain thoughts and ideas that were not included in the main chapters but that may be important to consider during future studies.

## **7.1 Inclusion bodies solubilization**

The inclusion bodies of our model protein can be readily dissolved using solely a relatively high alkaline pH ( $pH > 12$ ). The steps of this process could not be explicitly presented because they are part of the know-how package of Merck Sharp & Dohme (MSD). This approach has certain limitations, including: (a) it does not prevent the formation of soluble aggregates and (b) it is highly protein specific, as not all IBs can be solubilized by alkaline pH alone. Despite its technical limitations, as shown in chapter 6 this type of process is economically attractive. This finding certainly encourages future research efforts. These should aim at identifying the specific traits that allow the IBs of the model protein, to readily solubilize at alkaline pH. The identification of such trait(s) may then open the possibility to design new proteins whose IBs can also be solubilized by alkaline pH.

### **7.1.1 Measuring the content of soluble aggregates**

Initially the degree of protein aggregation was estimated using SDS-PAGE. However, the invasive nature of this technique may alter the composition of the samples, as prior to the analysis the samples are mixed with lithium dodecyl sulfate (LDS), and incubated at 75 °C for 10 min. A more suitable technique, employed in this work was size-exclusion chromatography. SEC is more suitable for this sort of analysis because the chemical composition of the sample remains unaltered during the measurement. The SEC data analysis, however, should be done bearing in mind that: the elution position of a solute does not only depend on its molecular mass, but also on its shape, and the elution position may be altered if the solute interacts with the gel matrix. To overcome the former limitation, future research would benefit from using an absolute size-exclusion chromatography (ASEC) set-up. Such set-up consists of a SEC column, followed by a dynamic light scattering (DLS)

detector, and it is capable of providing more accurate estimation of molecular weights, sizes and shape of the eluting solutes [1]. ASEC has been successfully used in studies on protein aggregation and degradation [2].

## 7.2 Ion-exchange chromatographic protein refolding

### 7.2.1 Effect of the adsorbent backbone

One of the on-going challenges on the area of ion-exchange chromatographic refolding (IExCR) is to comprehend the role that the backbone of the adsorbent plays and to quantify its contribution, to the fractional mass recovery ( $f_{Rec,Prot}$ ) and the refolding yield ( $Y_N$ ). During the work conducted the strategy on Fig. 7.1 was synthesized. This strategy suggests an experimental approach that may serve to define a suitable adsorbent for the optimal refolding of a given protein, and to provide relevant information to address the aforementioned challenges. The strategy takes into consideration the effect of the structural changes on the isotherm behavior, and the relation between fractional surface coverage and spatial isolation. It is important to mention that the strategy has not been validated using other model proteins (besides Lysozyme) and thus it is left for future research.

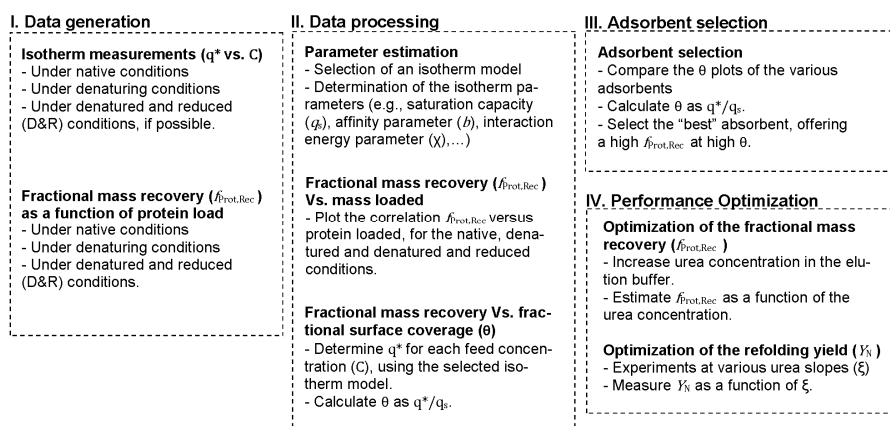


Figure 7.1: Strategy to investigate and quantitatively assess the effect that a given adsorbent on the refolding yield and the fractional mass recovery of ion-exchange protein refolding

### 7.2.2 Effect of the chaotrope gradient

Relatively fast changes in the denaturant concentration have been suggested to cause aggregate formation, and thus to decrease the refolding yield [3]. To overcome this limitation it has been suggested that a relatively slow change in the denaturant concentration should minimize this rapid structure collapse, considered to be the cause of the aggregate formation[4]. In chapter 3, it was shown that the denaturant gradient does

affect the refolding yield, and that this effect can be quantitatively assess using the slope of the gradient ( $\xi$ ). However, the following questions still remain: (a) Is the change in refolding yield caused by the change in residence time (occurring concurrently with the change in slope) or by the change in the speed of the chaotrope removal? And (b) what are the contributions of the residence time and the speed of the chaotrope removal to the magnitude of the change in refolding yield? To give an answer to these questions the following experiments are suggested. At a fixed protein load (mg/ml\_adsorbent) perform an IEx refolding experiment at a fixed gradient slope ( $\xi$ ). A second experiment should then be done were the chaotrope is removed using a step change, but the protein remains bound for a time equal to the time required by the denaturant gradient block (in the previous experiment). The fractions collected during the salt elution should be analyzed for activity and the proper protein mass balances should be performed. The comparison of the data from these two experiments should provide answers to the questions posed. These experiments, however, are left for future studies.

### **7.3 Size-exclusion chromatographic protein refolding**

Size-exclusion chromatography was the most exploited technique in this work. It was used as an analytical tool to quantify soluble aggregates (chapter 2 and 3), as an analytical tool to estimate suitable refolding mechanism(s) (chapter 4), and as a refolding reactor (chapter 5).

#### **7.3.1 Size-exclusion chromatography as a tool to estimate a suitable refolding mechanism**

Estimating a suitable refolding mechanism is important because without a mechanism, models to describe a chromatographic refolding reactor cannot be constructed. Conventionally, the competition between folding and aggregation has been represented as a first-order refolding reaction competing against a higher-order aggregation reaction. The kinetic constants of such mechanisms are usually determined using a least-squares fitting procedure and the measured transient concentration of the active protein [5-6]. In chapter 4, a novel method to estimate a suitable competitive reaction mechanism was presented. The approach made use of size-exclusion refolding data (the SEC chromatogram), a column model to describe the separation, mass balances, and a library of suitable mechanisms to determine the mass distribution of product and by-products. The main advantage of this approach over the conventional one is that it takes into consideration besides the native protein, the by-products formed after refolding. The experimental measurements conducted, however, had a limited resolution because of limitations of the SEC-UV set-up.

Such limitations can be overcome by coupling a DLS and a refractive (RI) detector to the SEC-UV system [1]. And thus, future research in this area should benefit if such a set-up is used during the experimental period. Using such set-up should give the ability to identify more species and accurately assign molecular masses and shapes, information that can be fed into the SEC reaction model, potentially improving its predictions.

### **7.3.2 Size-exclusion chromatography as a refolding reactor**

Size-exclusion chromatographic refolding has been shown to be a suitable refolding reactor that could potentially be employed for the refolding of recombinant proteins at a preparative scale. The following are the advantages of size-exclusion refolding observed during this work: negligible interactions between the gel matrix and the protein, relatively inexpensive column material, generally applicable for the refolding of most proteins, can be readily transferred to a continuous operation with the aid of Simulated Moving Bed technology, either urea or guanidine hydrochloride can be used for the IBs solubilization (this gives freedom to the design of the solubilization conditions), and on-line monitoring of the soluble aggregates formed.

During this thesis it was clearly shown based on economical (chapter 6) and technical evidence (chapter 5) that the foreseeable future of SEC refolding reactors lies in the application of SMB technology to SEC protein refolding. Additionally, it was shown that a conventional SMB configuration would be limited to the separation of the protein from the solubilization additives. Thus, future research should investigate the possibility of designing a three component SMBSEC for the purpose of refolding. In theory such unit would be capable of separating the solubilization additives from the protein and the native protein from the soluble aggregates (refolding by-products).

#### **7.4 References**

- [1] J. Wen, T. Arakawa, J.S. Philo, *Anal. Biochem.* 240 (1996) 155.
- [2] H. Ye, *Anal. Biochem.* 356 (2006) 76.
- [3] E.D.B. Clark, D. Hevehan, S. Szela, J. Maachupalli-Reddy, *Biotechnol. Prog.* 14 (1998) 47.
- [4] M. Li, G. Zhang, Z. Su, *J. Chromatogr. A.* 959 (2002) 113.
- [5] T. Kiefhaber, R. Rudolph, H.-H. Kohler, J. Buchner, *Nat. Biotechnol.* 9 (1991) 825.
- [6] D.L. Hevehan, E.D.B. Clark, *Biotechnol. Bioeng.* 54 (1997) 221.

## Curriculum Vitae

---

Gabriel Esteban Jaramillo Freydell was born the 24<sup>th</sup> of August of 1978 in Medellin (Colombia). In 1995 he completed his high school studies from Colegio Benedictinos. In 1996 he enrolled, as part of the first promotion, the program Process Engineering in the School of Administration Finances and Technology (EAFIT). In 2001 he graduated from EAFIT, after conducting his research project on photocatalysis of phenols under the supervision of Edison Gil Pavas. In 2002 he moved to Delft (The Netherlands) to start the program Designer in Bioprocess Engineering at the Delft University of Technology. In September 2003 he relocated to Oss (The Netherlands) and conducted his internship project at the former Organon N.V., nowadays Merk Sharp & Dohme (MSD). The focus of the internship project was to develop a process design to improve the refolding yield of proteins produced via inclusion bodies. This work was conducted under the supervision of Prof.dr.ir. Luuk van der Wielen, Dr.ir. Marcel Ottens and Dr. Michel Eppink. After successfully completed his designer project, he accepted a PhD position at the Delft University of Technology. The project was based on the results obtained during his internship at Organon N.V. In 2005 he conducted his first year as a PhD candidate in Organon (Oss) and in January 2006 he moved back to Delft to continue with his research. He completed his PhD education in 2009. In 2010 he relocated to Emeryville, California. Currently, he works as a Senior Scientist for the Process Chromatography Division at Bio-Rad laboratories (Hercules, CA).

# Publications and Conferences

---

## Publications

- E.J. Freydell, M. Ottens, M. Eppink, G.v. Dedem, L.v.d. Wielen. Efficient solubilization of inclusion bodies. *Biotechnol. J.* 2 (2007) 678.
- E.J. Freydell, Yvonne. Bulsink, Stef van Hateren, Luuk van der Wielen, Michel Eppink, Marcel Ottens. Size-Exclusion Simulated Moving Bed Chromatographic Protein Refolding. *Chem.Eng. Sci.* 65 (2010) 4701
- E.J. Freydell, Luuk van der Wielen, Michel Eppink, Marcel Ottens. Size-Exclusion Chromatographic Protein Refolding: Fundamentals, Modeling and Operation. *J. Chrom. A.* 2010 (under review)
- E.J. Freydell, Luuk van der Wielen, Michel Eppink, Marcel Ottens. Ion-Exchange Chromatographic Protein Refolding. *J. Chrom. A.* 2010 (under review)

## Conferences

- Freydell E., van der Wielen L.A.M, Eppink M.H.M, Ottens M. Protein refolding by gradient chromatography: Size-exclusion and its application to protein refolding. ACS-Biot division. Washington, USA. 2009. Oral presentation.
- Freydell E., van der Wielen L.A.M, Eppink M.H.M, Ottens M. Protein refolding by gradient ion-exchange chromatography. Biopartitioning and Biopurification conference (BPP). Uxbridge, UK. 2009. Oral presentation.
- Freydell E., Ottens M. Eppink M.H.M, L.A.M van der Wielen. Gradient chromatography protein refolding. Netherlands Biotechnology Congress (NBC). Ede, The Netherlands. 2008. Oral presentation.
- Freydell E., Ottens M., Eppink M.H.M, van Dedem G., van der Wielen L.A.M. Chromatographic oxidative refolding of proteins expressed as IBs. Netherlands Process Technology Symposium (NPS7). Veldhoven, The Netherlands. 2007. Oral presentation.
- Freydell E., Ottens M., Eppink M.H.M, van Dedem G., van der Wielen L.A.M. Chromatographic oxidative refolding of proteins expressed as IBs. AiCHE. Salt Lake city, USA. 2007. Oral presentation.
- Freydell E., Ottens M., Eppink M.H.M, van Dedem G., van der Wielen L.A.M. Chromatographic oxidative refolding of proteins expressed as IBs. Fundamentals of Adsorption Conference (FOA). Sicily, Italy. 2007. Oral presentation.
- Freydell E., van der Cammen M., Eppink M.H.M, Ottens M., van Dedem G., van der Wielen L.A.M. Dissolution and refolding without chaotropic agents. European Symposium on Biochemical Engineering Science (ESBES), Salzburg, Austria. 2006. Oral presentation.

# Acknowledgements/Dankwoord

---

I would like to start by extending my gratitude to the following persons: Marcel Ottens, Luuk van der Wielen and Michel Eppink, for providing me with the means to carry on this project, and for believing that at the end of the day, I could get the job done. Marcel once said to me in one of our meetings at the beginning of the project, "...you are now the Capitan of this ship..." This comment stayed with me through the years and I can even cite it today. It helped me a lot, thank you for sharing it. I had the opportunity to work with you these last years, practically on a daily basis. It was a wonderful experience. We had fun, we did science and I certainly learned a lot from you.

Luuk, I clearly remember the chat we had while driving around the refineries close to Rotterdam. You made a valuable comment about how important it is to find our own inspiration. I am happy to tell you that I have found my inspiration, thank you for pointing out its importance.

Michel, during the time we shared when I was in Oss, I remember the valuable lessons you taught me about courage and drive. I think that most of them started by you telling me about the 250 km that you cycled the weekend before. It was a pleasure to work with you, thank you.

I have been fortunate to meet many wonderful people throughout the duration of this project and I would like to thank all of you for the time you spend with me. To my friends in the former Diosynth (Yvonne, Diana, Kees, Joan, Cliff, Manon, Anke, Bart, Bianca, Peter) thank you for the contributions you made and I wish you lots of success. To my friends in the Department of Biotechnology (TU Delft), I wish I could mention you all, thank you for the different type of contributions you made for the culmination of this project. It was a wonderful, wonderful time. To Max Zomerdijk (-bon jour Monsieur-) it was great to have had the chance to work with you in the (blue) lab and I will never forget you saying "...there are no challenges, just opportunities". To Stef van Hateren, it was a pleasure to work with you. I specially remember the work on the SMB, taking it down and putting it back together, it was very cool. To both of you gentleman, thank you and good luck. To my friends (or should I say Bros) Beckley, Kedar, Sreekanth, Cagri (Effe), and Lodewijk (Bacan), thank you for the many discussions about science, Freakonomics (sorry guys but I had to mention it) and many other relevant topics. To Peter Verheijen, thank you for the early morning discussions during



coffee, they meant a lot to me. To my friends from the Soccer for Nerds, thank you for the many great memories; I wish you all good luck for the future.

Special thanks to Maarten van de Cammen for his help and assistance throughout this project. I must say that your lessons did not end in the laboratory (SDS-PAGE, RPHPLC, SELDI, etc) but you took them one step further, as they even included crash courses to learn Dutch with Jip and Janneke, cooking recipes, and how to play good level Badminton. To Emrah Nikerel, thank you for the time you spend teaching me MATLAB (the tool) and for helping me understand the potential that models have. To Beckley Nfor, thank you for the many critical discussions we had about experimental design, MATLAB in general (sometimes about the points that Emrah and Peter made), and specially the modeling of chromatography and the topic of chromatography in general. A Diego Mejía y Carol Roa, muchas gracias por haberme ayudado de la forma que lo hicieron, por siempre permaneceré profundamente agradecido. Igualmente tomo la oportunidad de extender mi gratitud a Edwin Habeych (Ed), Diego Romero y David Habeych, por haber estado presentes en momentos complejos que surgieron en este camino; Señores, gracias. Un saluto molto speciale a Mario Valvo (El loco), e stato un gran piacere; bouna fortuna. To Mark van Tongeren, thank you for being always there, in the tough times and in the fun times, landgenoot.

Lastly, I would like to thank my family. To Nestor and Beatriz Freydell, thank you for your help and support during difficult times. To my parents, Romy and Samuel, thank you for your valuable support and the constant encouragement. To my sister and brother, David and Sara, thank you for your love and for having such curious minds. A Lyda Gómez, muchas gracias por transmitirme tantas valiosas e innumerables lecciones, todas ellas hacen parte de mi esencia, gracias. Aan mij familie hier in Nederland (Toon, Beppie, Emma en Miguel), ik wil jullie bedanken voor jullie steun en energie, ze hebben en belangrijke rol gespeeld. To Lu, thank you for being there every day willing to listen and for helping me reflect on many things. You became my teammate on this adventure, thank you for the essential role you played. I just want to end by saying that I am looking so much forward to the next adventure.

