Development of High Throughput Screening Methods for the Automated Optimization of Inclusion Body Protein Refolding Processes

zur Erlangung des akademischen Grades eines

DOKTORS DER INGENIEURWISSENSCHAFTEN (Dr.-Ing.)

von der Fakultät für Chemieingenieurwesen und Verfahrenstechnik der

Universität Fridericiana Karlsruhe (TH)

angenommene

DISSERTATION

von

Dipl. Biotechnol. Annette Berg

aus Braunschweig

Tag der mündlichen Prüfung: 19.12.2009

Referent: Prof. Dr.-Ing. Jürgen Hubbuch

Korreferent: Prof. Dr. Christoph Syldatk

Acknowledgements

Abstract

Aim of the current thesis was the establishment and validation of automated high throughput screening (HTS) methods for development and optimization of downstream processing steps. Focus was laid on inclusion body protein refolding processes. Although a high demand exists for high throughput refolding screening several challenges could not be met up to now which are e.g. incompatibility of analytics with automated HTS platforms, the lack of universal methods to measure protein folding and an intelligent experimental design to economically search for optima in a multi parameter system.

Basic experimental methods for HTS screening were developed on a commercially available pipetting station Tecan Freedom Evo® 200 according to the HTS guidelines: automation, parallelization and miniaturization. Lysozyme was used as a model protein. Refolding buffer screening was performed in a 96 well plate format by dilution of denatured lysozyme into different refolding buffer systems with constant shaking and an additional aspirating and dispensing step of the refolding sample. Mixing in this dilution step was optimized to avoid high local protein concentration resulting in protein aggregation. A refolding time of 1 hour was observed to be sufficient to reach a stable equilibrium between protein aggregates and soluble protein species.

As a reference method for analytical development a standard lysozyme activity assay could be automated in a 96 well plate format. Decreasing turbidity of a *Micrococcus* suspension due to lysis of the bacterial cell walls can be easily detected and correlated to native lysozyme concentration.

An interesting finding during screening validation was that constant conditions during denaturation of lysozyme feedstock like unfolding time, temperature, concentration of reducing component and protein concentration are crucial for comparability of refolding results and thus for step wise optimization. Furthermore control of the redox environment by addition of oxidizing and reducing reagents was a prerequisite for high refolding yields and a systematic improvement of buffer composition. Air oxidation led to a completely non-systematic distribution of parameter values within the best refolding results.

An intense literature study revealed protein solubility and tryptophan fluorescence as potential non-specific analytical methods to estimate protein refolding success in an automated HTS approach. Measurement of protein solubility turned out to be complicated because of complex buffer matrices in refolding, low protein concentrations and protein structure variations interfering with commonly used techniques. Dye based assays and UV 280 nm absorption

1

revealed to be sensitive to either the presence of different protein folding states or to refolding buffer components like oxidizing or reducing reagents. Consequently automated methods for buffer exchange like ultrafiltration, size exclusion chromatography and dialysis had to be investigated. All approaches turned out to have some serious draw backs. Ultrafiltration led to a protein concentration and buffer dependent loss of protein whereas dialysis and size exclusion chromatography failed to separate protein from oxidizing reagents in a manageable time with robotic compatible equipment. As direct measurement of soluble protein turned out to be impossible in HTS refolding screening an indirect method was established determining aggregate concentration instead. In the resulting procedure lysozyme aggregates were separated by filtration in 96 well filter plates and subsequently resolubilized in denaturing buffer for absorption measurement at 280 nm. Soluble protein content could then be calculated on the basis of mass balances with the known initial and the assessed aggregated protein concentration. Validation with 40 highly divers refolding samples showed mass balances close to 100 % and an average standard deviation of 2 % thus qualifying this method for high throughput refold screening.

The correlation between solubility and folding was evaluated intensely. Highly soluble samples in refolding do not necessarily show high activity. Nevertheless optimization of solubility with a controlled redox environment led to optimum parameter sets overlapping with those favoured for protein folding. These observations point out the possibility to use fast and easy solubility measurements in a more excessive first screening round to reduce the parameter space of interest in later screenings with higher analytical effort.

As a second non-specific method intrinsic tryptophan fluorescence was validated as a tool to estimate refolding success. Emission spectra could be easily measured in a 96 well plate format. Due to the rather small red shift of spectra from completely active to completely denatured lysozyme an asymmetric Gauss fit was performed to calculate the exact emission maximum. In 40 lysozyme refolding systems with random buffer composition all measured tryptophan spectra laid in a close range showing no correlation with the specific yield of active lysozyme. This observation is likely due to a hydrophobic collapse of lysozyme being obligatory for protein solubility. Soluble protein will thus exhibit a rather uniform spectral behaviour. Interestingly a larger variation in tryptophan emission spectra is observed in the absence of redox components hinting at a higher flexibility of the protein due to the lack of disulfide bridges. Summarizing tryptophan fluorescence is a tool to follow extreme changes in protein structure like unfolding but unfortunately the resolution at least for lysozyme is too low to use it for refolding analytics in optimization experiments.

After the technical basis for refolding screening was provided two different methods for intelligent experimental design were investigated: a genetic algorithm and a classical factorial design. Optimization of up to five parameters (pH, concentration of NaCl, $MgSO_4$, DTT and GSSG) could be fully automated with the genetic algorithm including several rounds of experimentation, determination of refolding yield and calculation of the next starting parameter matrix. A high data density is provided in the parameter range of interest close to the optimum. With this stochastic method the average active lysozyme concentration could be improved with a factor of f = 8.3 from 0.019 mg/ml to 0.154 mg/ml within 5 generations consisting of 40 experiments each. The concentration of soluble lysozyme is improved with a factor of f = 2 from 0.195 mg/ml to 0.384 mg/ml including experiments with maximum solubility of 0.4 mg/ml within 4 generations.

A full factorial design in the reduced parameter space close to the optimum is applied to confirm or refine the optimization. In our case the genetic algorithm did not lead to a global optimum for two of the five parameters investigated which could be revealed by the full factorial experiments. The algorithm is caught on a diagonal line between two parameters representing the molar ratio of oxidizing to reducing component or in other words the redox potential of the solvent. The linear recombination method of the genetic algorithm in combination with the chosen mutation rate likely accounts for these undesired effects. This knowledge offers the possibility to adapt the GA parameters to the screening objective and points out the usefulness of an additional full factorial experiment.

During refolding screening development protein solubility was revealed to be of major importance for the qualified selection of suitable buffer systems. Consequently an available HTS tool for the determination of protein precipitation curves was further improved in terms of throughput to characterize the effect of pH, temperature and different additives like PEG, sorbitol, sucrose or Tween 20 on the solubility behaviour of lysozyme. Measurement of lysozyme solubility in additive buffer mixtures resulted in complex solubility surfaces demonstrating interdependent effects of both solvent components. As these interdependent effects are not simply additive, optimization of such systems should always be performed in the presence of both substances. An additional observation was that the pH dependency of lysozyme solubility correlated with the calculated net charge of the molecule. Decreasing net charge at pH values in proximity to the isoelectric point led to increasing protein interactions. Repulsive forces thus seem to play a major role in protein solubility.

Besides the development of screening methods for dilution refolding, HTS techniques for the optimization of solid phase refolding processes were established on the robotic workstation.

Batch binding of denatured lysozyme on different cation exchange adsorber resins was characterized by measurement of adsorption isotherms and kinetics revealing a significantly weaker binding of denatured lysozyme. The disintegration of charged patches by unfolding and the stretched molecular structure with solvent exposed hydrophobic residues presumably accounts for these observations.

Refolding was initiated by buffer exchange from denaturing to refolding buffer after transfer of loaded adsorber particles to filter plates. Effects of protein loading, incubation time in refolding buffer, refolding buffer pH and urea concentration on protein yield could be analyzed with the developed automated method. A strong correlation between folding and adsorption/desorption processes could be observed with only completely folded active protein being amenable to elute with high salt concentrations. Misfolded or aggregated protein could only be eluted in denaturing buffer conditions. These results hint at different adsorption mechanisms of the different protein species. In summary optimization of on-column refolding process parameters like resin type, protein loading and buffer compositions is possible with low protein and resin consumption using the developed HTS approach.

Zusammenfassung

Ziel der vorliegenden Arbeit war die Etablierung und Validierung von automatisierten Hochdurchsatz Screening (HTS) Methoden zur Entwicklung und Optimierung von Proteinaufarbeitungsprozessschritten. Hierbei wurde der Schwerpunkt auf Inclusion Body Protein Rückfaltungsprozesse gelegt. Obwohl der Bedarf für Hochdurchsatzscreening im Bereich der Proteinrückfaltung hoch ist, konnten zahlreiche Probleme bis heute noch nicht gelöst werden. Hierzu zählen z.B. die ungenügend auf Automationsstationen angepasste Analytik, das Fehlen universeller Methoden zur Bestimmung der Proteinfaltung und das Fehlen intelligenter Versuchsplanungsverfahren zur ökonomischen Optimierung von Multiparametersystemen.

Alle experimentellen Methoden für das Hochdurchstzscreening wurden auf einem kommerziell erhältlichen Pipettierroboter Tecan Freedom Evo® 200 unter Berücksichtigung der HTS Grundsätze: Automation, Parallelisierung und Miniaturisierung entwickelt. Rückfaltungspufferscreening wurde in 96 well Platten durch Verdünnung des denaturierten Lysozyms in verschiedenen Puffersystemen unter konstantem Schütteln durchgeführt. Das Mischen in diesem Verdünnungsschritt wurde optimiert um lokale Proteinkonzentrationsspitzen zu vermeiden, die zu Aggregation führen können. Eine Inkubationszeit von einer Stunde wurde als auseichend für die stabile Gleichgewichtseinstellung zwischen aggregiertem und gelöstem Protein ermittelt.

Als Referenzmethode für die Entwicklung der Analytik wurde ein Standardlysozymassay in einem 96 well Plattenformat automatisiert. Die abnehmende Trübung einer Micrococcus Suspension durch Auflösen der Zellwände mittels Lysozym kann leicht in einem Photometer gemessen und mit der nativen Lysozymkonzentration korreliert werden.

Bedingungen zur Herstellung des denaturierten Proteins wie Entfaltungszeit, Temperatur und die Konzentration von Protein und Reduktionsmittel wurden als wichtige Einflussgrößen für das spätere Rückfaltungsergebnis identifiziert und müssen zur Vergleichbarkeit der Ergebnisse und zur gezielten, schrittweisen Optimierung von Rückfaltungsbedingungen konstant gehalten werden. Weiterhin wurde festgestellt, dass die Kontrolle der Redoxumgebung durch Zugabe von Reduktions- und Oxidationsmittel für hohe Rückfaltungsausbeuten und eine systematische Verbesserung der Pufferzusammensetzung zwingend notwendig ist. Die reine Luftoxidation führte zu niedrigen Ausbeuten und einer unsystematischen Verteilung der Rückfaltungsparameter bei den Experimenten mit der höchsten Ausbeute.

Eine der wesentlichen Herausforderungen war die Entwicklung der Analytik. Die Korrelation von universell, schnell und automatisiert messbaren Größen wie der Tryptophanfluoreszenz und Proteinlöslichkeit mit der Lysozymaktivität wurde untersucht. Die nötige Analytik wurde in den Robotikprozess integriert, um manuelle Intervention und off-line Messungen zu umgehen. Lysozym wurde für alle Prozesse als Modellprotein verwendet.

Da übliche Methoden zur Bestimmung der Proteinlöslichkeit wie die Absorption bei 280 nm oder der Bradfordassay mit der Anwesenheit verschiedener Proteinstrukturvarianten und Lösungsmittelkomponenten nicht kompatibel sind, wurden Möglichkeiten zum automatisierten Pufferwechsel validiert. Ultrafiltration, Dialyse und Größenausschlusschromatorgaphie wurden auf ihr Potential zur Trennung von GSSG und Lysozym untersucht. Auf Grund von Proteinverlusten, langen Prozesszeiten und aufwändiger Implementierung in das Robotersystem wurden diese Ansätze zu Gunsten einer indirekten Bestimmung des gelösten Proteins aufgegeben. Proteinaggregate wurden über Filterplatten abgetrennt, in Denaturierungspuffer gelöst und die Konzentration übe Absorption bei 280 nm bestimmt. Gelöstes Protein kann auf Basis der Massenbilanz mit der bekannten Anfangskonzentration und der gemessenen Aggregatkonzentration berechnet werden. Die Validierung dieser Methode mit 40 Rückfaltungsansätzen bei hoher Puffervariabilität ergab Massenbilanzen nahe 100 % und eine mittlere Standardabweichung von 2 % und qualifizieren das Verfahren somit für den Einsatz in einem HTS Optimierungsprotokoll. Die Korrelation von Proteinlöslichkeit und Faltung zu aktivem Protein wurde intensiv untersucht. Es wurde eine Überlappung der Parameteroptima für diese beiden Messgrößen festgestellt, die es ermöglicht, Löslichkeit als Zielfunktion in einem exzessiven ersten Screening einzusetzen, um den interessanten Parameterbereich für folgende Screenings mit höherem experimentellen Aufwand zu reduzieren.

Als zweite Methode wurde intrinsische Tryptophanfluoreszenz zur universellen Messung von Proteinfaltung validiert. Die Wellenlänge im Emissionsmaximum bei Anregung von Tryptophan mit 280 nm zeigt bei Denaturierung von Lysozym eine signifikante Rotverschiebung, die auf einer höheren Zugänglichkeit des sonst innerhalb des Moleküls liegenden Tryptophan für die Lösungsmittelmoleküle beruht. In Lysozymrückfaltungsansätzen zeigte die lösliche Proteinfraktion keine Korrelation zwischen Wellenlänge im Emissionsspektrum und spezifischer Aktivität. Die Emissionsmaxima lagen in einem relativ engen Bereich über dem Wert für natives Lysozym. Ohne Zugabe von Redoxreagenzien streuten die Werte stärker. Die beobachteten Effekte beruhen sehr wahrscheinlich auf dem Bestreben des Proteins, alle hydrophoben Gruppen bei Verdünnung

in Rückfaltungspuffer schlagartig ins Innere zu verlagern, was zu einer Separierung des löslichen Proteins mit innen liegenden Tryptophanen und des aggregierenden Proteins mit Tryptophan an der Moleküloberfläche führt. Eine höhere Flexibilität des Lysozyms ohne Redoxsubstanzen durch die verlangsamte Bildung von Disulfidbrücken hat eine höhere Zugänglichkeit der Tryptophane zur Folge, die sich durch stärkere Abweichungen im Emissionsspektrum zeigt. Zusammenfassend lässt sich sagen, dass extreme Strukturveränderungen, wie Entfaltung, über Tryptophanfluoreszenz gemessen werden können. Dennoch ist das Auflösungsvermögen der Tryptophanfluoreszenz in Rückfaltungsansätzen zumindest für Lysozym für eine gezielte Optimierung zu niedrig.

Nach Etablierung der experimentellen Methoden wurden zwei verschiedene Versuchsplanungsverfahren untersucht: ein genetischer Algorithmus und ein klassischer faktorieller Versuchsplan. Um eine hohe Anzahl an Prozessparametern robust optimieren zu können, wurde der genetische Algorithmus in den Roboterprozess integriert. Die Optimierung durchläuft vollautomatisch die aufeinander folgenden Zyklen bestehend aus der Durchführung der Experimente, der Berechnung der Zielfunktion und der Erstellung des Versuchsplans für die nächste Runde. Bis zu fünf Parameter wie pH-Wert und die NaCl, $MgSO_4$, DTT and GSSG Konzentration wurden zeitgleich optimiert. Über den stochastischen Optimierungsalgorithmus konnte die Konzentration an aktivem Protein um den Faktor 8,3 von mittleren 0,019 mg/ml auf 0,154 mg/ml in fünf Generationen mit je 40 Experimenten erhöht werden. Die mittlere Konzentration an gelöstem Protein wurde in 4 Generationen um den Faktor 2 von 0,195 mg/ml auf 0,384 mg/ml verbessert.

Ein vollfaktorieller Versuchsplan im Optimum wurde genutzt, um Parametereffekte zu bestätigen und die Auflösung in diesem reduzierten Parameterraum zu erhöhen. Hierbei konnte festgestellt werden, dass das bestimmte Optimum für GSSG und DTT vom globalen Optimum abweicht, was auf eine nötige Anpassung des genetischen Algorithmus im Bezug auf die gewählte Rekombinationsmethode und die Mutationsrate hinweist. Zusammengefasst führt die Kombination eines genetischen Algorithmus zur Verkleinerung des interessanten Parameterbereiches und vollfaktorieller Versuchspläne zur Untersuchung der Parametereffekte im Optimum neben der Verbesserung der Zielfunktion auch zu einem erhöhten Verständnis des Prozess und der Arbeitsweise des genetischen Algorithmus.

Da die Proteinlöslichkeit bei der Optimierung von Rückfaltungsprozessen eine prominente Rolle spielt wurde der Durchsatz in einer bereits vorhandenen automatisierten Methode zur Bestimmung von Proteinlöslichkeit erhöht, um den Einfluss komplexerer Puffersysteme bestehend aus Additiv und Salz zu messen. Mit der adaptierten Methode konnte der Einfluss

der Temperatur und des pH Werts in Übereinstimmung mit der Literatur bestimmt werden. Weiterhin konnten komplexe Löslichkeitsflächendiagramme für Lysozym, Salz und Additive, wie PEG mit verschiedenen Molekulargewichten, Tween 20, Sorbitol und Sucrose, erstellt werden. Beobachtete Wechselwirkungen zwischen Additiv und Salz in verschiedenen Konzentrationen lassen sich in einem einfachen Screening mit einer Komponente nicht erfassen.

Neben der Etablierung von HTS Methoden zur Proteinrückfaltung über Verdünnung wurden auch Methoden zur Untersuchung von Rückfaltungsprozessen auf Chromatographieadsorbern entwickelt. Die Adsorption von denaturiertem Lysozym an Ionenaustauschmaterialien wurde im Batchmodus über Isothermen und Kinetiken charakterisiert und eine signifikant niedrigere Bindung von denaturiertem im Vergleich zu nativem Lysozym festgestellt. Dies kann durch die Zerstörung der geladenen Bindungsstellen auf der Proteinoberfläche bei Denaturierung erklärt werden. Die Rückfaltung wurde durch einen Pufferwechsel in Rückfaltungspuffer nach dem Transfer der beladenen Adsorberpartikel in Filterplatten durchgeführt. Die Effekte von Proteinbeladung, Verweilzeit in Rückfaltungspuffer, Rückfaltungspuffer pH und Harnstoffkonzentration auf die Rückfaltungsausbeute konnten mit de entwickelten Methode untersucht werden. Eine starke Korrelation zwischen Faltungs- und Adsorptions-, Desorptionsprozessen wurde beobachtet. So konnte im Elutionsschritt mit 1 M NaCl ausschließlich aktives Lysozym eluiert werden, während mißgefaltetes oder aggregiertes Protein nur unter denaturierenden Bedingungen desorbiert werden konnte. Weiterhin wurde die verfrühte Elution von Protein in Rückfaltungspuffer durch die Zugabe von Harnstoff und kurze Inkubationszeiten verstärkt, was vermutlich auf niedrigere Anteile an aktivem, unter diesen Bedingungen stark adsorbierendem Protein zurückzuführen ist. Zusammengefasst konnte gezeigt werden, dass Parameter für die Rückfaltung auf Chromatographieadsorber, wie Adsorbertyp, Proteinbeladung und Zusammensetzung des Rückfaltungs- und Elutionspuffers, mit niedrigem Bedarf an Protein und Adsorber optimiert und komplexe Abhängigkeiten zwischen Faltung und Adsorption untersucht werden können.

Abbreviations

ANS	1, anilino-naphthalene 8-sulfonate
BCA	bicinchoninic acid
BSA	bovine serum albumin
CD	circular dichroism
CPA	cis-parinaric acid
CV	column volume
DSC	differential scanning calorimetry
DLS	dynamic light scattering
DTT	reduced dithiothreitol
DWP	deepwell plate
ELISA	enzyme linked immunosorbent assay
FP	filter plate
GA	genetic algorithm
GSSG	oxidized glutathione
HIC	hydrophobic interaction chromatography
HPLC	high performance liquid chromatography
HTS	high throughput screening
MTP	microtiter plate
PEG	polyethylene glycol
PRODAN	6-propionyl-2-dimethylaminonaphthalene
pI	isoelectic point
RP-HPLC	reversed phase high performance liquid chromatography
RT	room temperature
SDS-PAGE	sodium dodecylsulfate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SIC	self-interaction chromatography
SLS	static light scattering
SPR	surface plasmon resonance
Tris	tris(hydroxymethyl)-aminomethan
UV MTP	UV microtiter plate (UV transparent bottom)

Symbols

c	concentration [M] [mg/ml]
L	number of parameter levels [-]
I	intensity of emitted light [-]
K	BET adsorption coefficient [ml/mg]
k	velocity constant [min^{-1}]
K_L	Langmuir adsorption coefficient [ml/mg]
K_S	salt specific solubility contant [M^{-1}]
m	mutation rate [-]
n	number of parameters [-]
q	binding capacity [mg/ml_{ads}]
S	protein concentration in the Cohn equation [g/l]
s	selection pressure [-]
T	temperature [°C]
t	time [min] [h]
V	volume [l]
Y	yield [mg/ml]

Greek Symbols

β	hypothetical maximum solubility without salt [g/l]
Δ	difference (absolute value) [-]
λ	wavelength lambda [nm]

Indices

0	initial value
ads	adsorber
prot aggregate	aggregate
e	end
sol filtrate	filtered
i	at time i
max	maximum value
min	minimum value
norm	normalized
resol	re-solubilized aggregate
sat	value at saturation
SN	supernatant
sol	soluble, solution
t	value at time t
total	total
wash	wash fraction

Contents

1 Introduction

1.1 Inclusion Body Protein

Reducing conditions in the cytosol and the absence of chaperones during the expression of eukaryotic proteins in *E.coli* lead to misfolded protein prone to degradation. High expression rates overload the cellular degradation machinery and protein aggregates, so called inclusion bodies are formed (Carrio and Villaverde 2002).

An advantage of the inclusion body formation is their high content of recombinant protein up to 98 % (Singh and Panda 2005) and the opportunity to easily harvest the protein by differential centrifugation of the dense aggregates (Middelberg 2002). On the other hand inclusion bodies have to be solubilised in denaturing reducing agents and subsequently refolded to gain active, soluble protein. Independent of these pros and cons *E.coli* is the most widely used expression host due to its cheap and easy cultivation in defined media, high growth rates, high content of recombinant protein (up to 50 % of the total protein) and a well established system for genetic manipulation (Choe, Nian et al. 2006). Biopharmaceutical products lacking posttranslational modifications are thus frequently produced in *E.coli*. In Table 1 products produced as inclusion bodies are summarized.

Table 1: Important biopharmaceutical products produced as inclusion bodies

Protein	Medical Indication
Human Insulin	Diabetes
Human Growth Hormone	Growth Failure
Interferon-β-1b	Multiple Sclerosis
Interferon-α-2a	Hepatitis B and C
Tissue Plasminogen Activator	Acute Myocardial Infarction
Granulocyte Colony-Stimulating Factor	Neutropenia
Interleukin 1	Thrombocytopenia
Interleukin 2	Renal Cell Carcinoma

Insulin is presumably the best-known product formed as inclusion bodies. Furthermore growth hormone and several growth factors, like G-CSF (Granulocyte Colony-Stimulating Factor), interferons and interleukins are produced in *E.coli* and have to be refolded (Misawa

and Kumagai 1999). At this point it should also be mentioned that in huge genomic projects a high number of proteins are produced as insoluble protein aggegates in *E.coli*. A tendency towards expression for example of antibody fragments or biosimilars in *E. coli* in combination with renaturation processes is observed and might partially substitute mammalian cell expression systems (Farid 2007; Patil, Rudolph et al. 2008; Vanz, Renard et al. 2008).

1.2 Refolding Processes

Refolding implicates the reduction of denaturant by simple dilution, dialysis or in a chromatographic column and, for cysteine containing proteins, supply with redox components for disulfide bridge formation.

1.2.1 Dilution

Dilution based refolding offers the possibility to investigate different buffer compositions in a flexible and simple manner. The main draw back is the complexity of the buffer matrix for subsequent analytics. Protein concentrations are typically low, in the range of 0.1 to 1 mg/ml to reduce aggregation of hydrophobic folding intermediates. This point will be further discussed below and is a general problem in refolding processes leading to low space time yields. Serial addition of denatured protein is used to increase the concentration of soluble protein and therefore to facilitate subsequent detection methods. Here waiting time between pulses can cost several hours of time depending on the folding kinetic of the protein and has to be evaluated carefully to reach a stable protein conformation in between addition steps.

In dilution refolding fast mixing of solubilised protein and refolding buffer is essential to reach a homogeneously low protein concentration at the starting point of renaturation. In addition with the injection rate of the denatured protein an optimum can be found for high process yields (Lee, Buswell et al. 2002; Mannall, Titchener-Hooker et al. 2006). Furthermore mixing is the most important process parameter during scale up of refolding by dilution (Jungbauer and Kaar 2007).

1.2.2 Dialysis

Dialysis is used for slow buffer exchange, but it is for both refolding and subsequent sample preparation mostly limited by slow diffusion processes. The protein concentration has to be kept low to prevent the protein from aggregating. Precipitation is especially observed if the protein folding rate is low and the folding intermediates tend to interact (Tsumoto, Ejima et al. 2003). Leon and Middelberg (2006) compared dilution refolding and dialysis refolding for human Alpha-fetoprotein. Dialysis as a slow process with high buffer consumption led to an overall lower refolding yield. One of the reasons was the loss of protein by non-specific adsorption at the membrane. A second reason mentioned was increased protein precipitation likely caused by a slow buffer exchange and consequently a different folding pathway of the protein.

Time for dialysis depends on diffusion properties of the molecules, the membrane area, thickness and composition, temperature, concentration gradient between buffer reservoir and sample and mixing on each side of the membrane (Bowen 1993). A step wise dialysis is time-consuming but offers the possibility for a controlled gradual exchange of buffer compositions.

1.2.3 Matrix-assisted Refolding

Refolding by means of chromatographic resins can be divided into refolding of adsorbed protein on ion exchange, hydrophobic interaction or affinity adsorbents and refolding of proteins in size exclusion matrices. Chromatographic renaturation methods can combine refolding by removal of denaturant and reducing agent and separation of native, misfolded and contaminating protein in one process step (Geng and Wang 2007). In size exclusion chromatography (SEC) proteins pass the gel matrix equilibrated with refolding buffer faster than the denaturing agent and thereby refolding is initiated. Applying a decreasing denaturant gradient leads to a slower transfer of the protein to conditions promoting folding. Protein aggregates formed during refolding are presumably re-solubilized by the delayed running front of the denaturant. Low chromatographic velocities and sample volumes are limiting factors in SEC refolding processes.

Refolding of adsorbed protein uses different binding mechanisms to reduce protein-protein interactions by immobilisation. In hydrophobic interaction chromatography (HIC) the hydrophobic domains of the protein bind under high salt conditions and intermolecular interactions based on hydrophobic patches are thus reduced (Geng, Quan et al. 2004).

Solubility problems in solutions with high ionic strength restrict operating conditions in HIC refolding. In addition binding might prevent hydrophobic residues to move inside the core of the molecule which is a crucial step during folding. Ion exchange chromatography can be used for refolding as the binding strength is rather low and therefore does not inhibit the formation of secondary and tertiary structures. Reduced mobility of the proteins reduces aggregation (Langenhof, Leong et al. 2005).

Affinity resins can consist of immobilized binding partners, chaperons or, and most frequently, of immobilized metal ions like nickel or copper for the binding of His-tagged proteins. Binding to affinity matrices is often highly specific and therefore leads to concomitant purification of the target protein. Anyway metal affinity resins are prone to ligand leakage and sensitive to low pH and frequently used additives like arginine (Cowieson, Wensley et al. 2006). Specific ligand coupling to a resin backbone is laborious and not feasible in industrial scale (Machold, Schlegl et al. 2005). All matrix assisted refolding methods show higher soluble protein concentrations in the refolded eluate in comparison to samples refolded by dilution or dialysis. A critical factor is the loading of the column, the gradient of denaturant and the flow rate (Langenhof, Leong et al. 2005).

1.3 Parameters in Protein Refolding

Besides technical parameters also buffer composition has to be optimized to provide an appropriate environment for folding and disulfide bond formation. For lack of a known functional relation between refolding yield and buffer components process design is still rather a matter of trail and error. The complexity of the optimization problem is due to a high number of parameters and parameter interactions. The development of refolding processes takes into account the solubility of the protein during refolding and the formation of native disulfide bonds in cysteine containing proteins. Parameters like ionic strength, salt type, pH value and several additives listed in Table 2 can contribute to an improved solubility of the protein.

Table 2: Common additives in refolding processes

Additive	Typical Concentration
Urea	2 M
Guanidinium hydrochloride	1 M
Arginine	0.5 M
Glycerol	20 - 40 % v/v
Sucrose	0.4 M
Lauryl maltoside	0.3 mM
Polyethylene glycol (3550 MW)	0.05 % w/v
Tris buffer	0.5 M
Triton X-100	10 mM
n-hexanol	5 mM
Sodium sulfate or potassium sulfate	0.5 M
Sodium chloride	0.5 M
Methylurea	1.5 - 2.5 M
Ethylurea	1.0 - 2.0 M
Formamide	2.5 - 4.0 M
Acetamide	1.5 - 2.5 M
Ethanol	up to 25 %
n-Pentanol	1 - 10 mM
Cyclohexanol	0.01 - 10 mM
Sorbitol	20 - 30 % v/v
α-Cyclodextrin	20 - 100 mM
CHAPS	10 - 60 mM
Mixed micelles	Depending on compounds used
Dodecyl maltoside	2 - 5 mM

Nevertheless common guidelines for protein solubility are not always true when refolding protein. Although e.g. native protein shows maximum solubility with increasing distance of the pH from the isoelectric point for refolding protein an alkaline pH is essential for the formation of the thiolate anion and in a second step native disulfide bonds. This example shows already the interdependencies of different buffer components. Furthermore native protein and protein undergoing refolding might differ strongly in their surface properties influencing solubility in different solvent systems.

Disulfide bond formation can be promoted by addition of divalent metal ions for improved air oxidation and by adaptation of the redox potential with reducing and oxidizing chemicals like dithiothreitol (DTT), glutathione (GSSG) and cysteine. As refolding always starts with reduced protein in a first step mixed disulfides are built with oxidizing agents. In a second step disulfide bonds are formed. Non-native disulfide bonds can be reshuffled by reducing agent present in the refolding buffer (Lilie, Schwarz et al. 1998). Consequently the ratio of oxidizing and reducing agent is besides the concentrations a significant parameter in refolding. In Figure 1 processes in disulfide bond formation are visualized. The ratio of oxidizing and reducing agent is also referred to as redox potential of a solution.
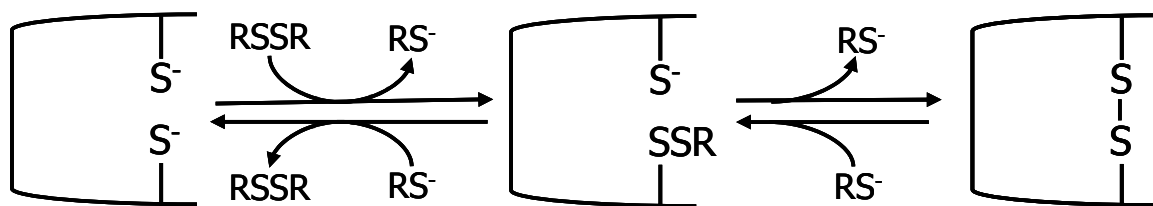
Figure 1: Mechanism of disulfide bond formation

1.4 Thermodynamic and Kinetic Aspects in Protein Refolding

The native structure of a protein is coded in its amino acid sequence (Anfinsen 1972). According to the current understanding of folding mechanisms the native structure is the most stable structure under physiological conditions (Dill and Chan 1997). Denatured protein has a high degree of entropy due to high flexibility of the polypeptide chain. During folding the entropy as well as the enthalpy, the driving force for this stochastic search process, are decreased frequently, visualized in energy landscape diagrams like in Figure 2. The first step in folding is always a hydrophobic collapse of the molecule towards a native like topology. Hydrophobic residues are buried inside the molecule and water molecules are excluded for a highly compact structure. The pathway a protein takes decides on the formation and stability of intermediates and the folding speed (Dobson 2004).
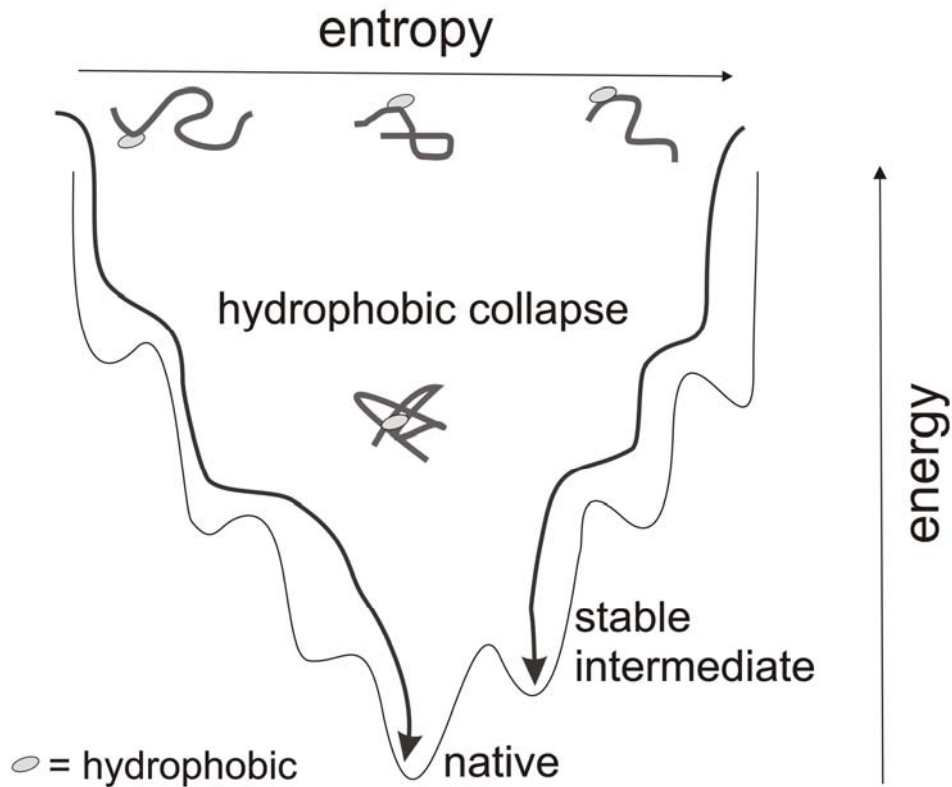
Figure 2: Folding funnel showing a reduction in entropy and increase in free energy during folding

Folding intermediates are often prone to aggregation for hydrophobic residues exposed on the protein surface. Furthermore aggregation shows a second or higher order kinetic being favoured at high protein concentrations whereas folding shows a kinetic of first order and is independent of the protein concentration (Kiefhaber, Rudolph et al. 1991). In refolding processes with slow folding pathways or high protein concentrations aggregation of intermediates is one of the main cause for protein loss. In Figure 3 folding and aggregation processes are visualized.
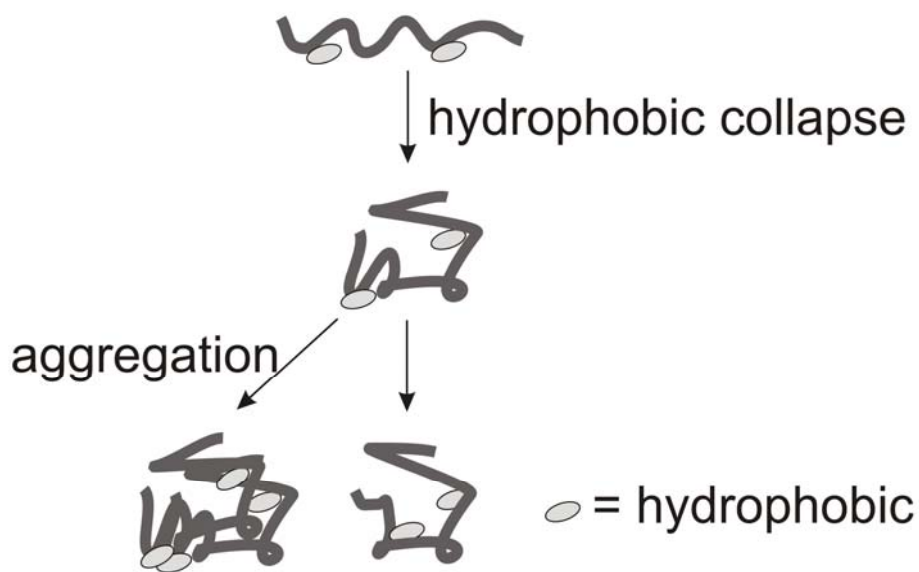
Figure 3: Aggregation of hydrophobic residues in refolding intermediates

1.5 Analytical Methods in Protein Refolding Process Development

Analytical methods can be split into measurement of soluble protein content and techniques to gain structural information. As this aspect will be discussed in detail in the current thesis the introduction only contains a short overview.

1.5.1 Solubility

Except for turbidity measurements which show no linear correlation with protein content during protein solubility screens the determination of soluble protein concentration in refolding buffer systems is limited by inhibitory matrix effects on common protein assays and low protein concentrations. Colorimetric assays as well as absorption at 280 nm might be incompatible with solvent components like detergents, amino acids and reducing and oxidizing agents. Furthermore the signals in dye binging assays are sensitive towards structural varieties in protein samples, which leads to problems especially during protein refolding and formulation studies taking into account protein stability and solubility. SDS-

PAGE analysis of the soluble and re-solubilized aggregated protein frequently used for quantification of solubility includes manual intervention for complex handling of gel systems (Armstrong, De Lencastre et al. 1999; Wang, John et al. 2004; Cowieson, Wensley et al. 2006; Ejima, Ono et al. 2006; Cowan, Davies et al. 2008). Analytical size exclusion chromatography to assess protein molecular weight and soluble aggregate content is time consuming for serial sample processing. Finally non-destructive methods like static and dynamic light scattering are used to qualitatively measure the monodispersity of a sample and an average particle size hinting at protein interactions (Ho, Middelberg et al. 2003; Zhang and Liu 2003; Jancarik, Pufan et al. 2004).

1.5.2 Folding

Methods for estimation of protein folding can be divided into specific binding or activity assays and measurement of conformational properties.

Structural based methods offer the advantage to be potentially applicable to a high number of proteins. Circular dichroism is a method frequently used to determine secondary structure elements qualitatively by comparison with signals of native protein. This method is restricted to pure protein samples free of components absorbing light at 280 nm. Besides these incompatibilities secondary structure can be observed already early in refolding and does not give any information on tertiary structure. Reversed Phase HPLC and Hydrophobic Interaction HPLC is used to measure surface hydrophobicity differences of native and misfolded proteins in correlation to retention behaviour on hydrophobic column materials. Sample preparation prior to HPLC analysis can alter protein conformations and serial analysis is generally time-consuming (Shire, Shahrokh et al. 2004). Surface hydrophobicity can also be detected with hydrophobic dyes like ANS (1, anilino-naphthalene 8-sulfonate), PRODAN (6-propionyl-2-dimethylaminonaphthalene) or CPA (cis-parinaric acid) (Evans and Engelhard 1996; Lakowicz 2000; Royer 2006). These dyes bind to hydrophobic patches in folding intermediates and upon interaction with protein emit light with high intensities. As dye binding also depends on solvent characteristics applicability to different buffer systems should be carefully evaluated. Characteristics of fluorescent probe and formed hydrophobic surface in folding intermediates have to fit to allow for significant binding (Cardamone and Puri 1992; Alizadeh-Pasdar and Li-Chan 2000). Intrinsic fluorescence spectroscopy is limited to proteins with buried tryptophane residues in native conformation to allow for a characteristic red shift of fluorescence emission during unfolding due to solvent interaction with the

fluorophore. Furthermore fluorescence intensity is sensitive to quenching buffer components like disulfides and to temperature shifts. The interpretation of spectra implies knowledge on structural properties or the fluorescent behaviour of the native protein (Royer 2006). Another non-specific method to investigate structural properties of proteins is limited proteolysis, harnessing higher stability of tightly folded native protein structures against proteolytic cleavage (Heiring and Muller 2001). A great draw back is the analysis of protein fragments performing SDS-PAGE after the digestion reaction and a sometimes low stability of the native protein against protease digestion.

Specific enzyme and cell based activity assays yield reliable information on protein structure integrity and are also automatable and easily parallelized. For this reason they are in many publication the chosen method for structure determination and quite often absolute enzyme activity is the exclusive objective function for optimization of refolding conditions (Ahn, Lee et al. 1997; Armstrong, De Lencastre et al. 1999; Sijwali, Brinen et al. 2001; Lange, Patil et al. 2005; Willis, Hogan et al. 2005; Rahimpour, Mamo et al. 2007).

Methods like ELISA and SPR harness specific binding of protein to binding partners like substrate molecules, antibodies or receptors and are already applied in high throughput refolding approaches although surface design and production and process set up are challenging tasks (Jones, Hutchinson et al. 2004; Lange, Patil et al. 2005; Cowan, Davies et al. 2008). For functional assays restricted to one protein time for assay development should not be underestimated in comparison to just adaptation of universal structure based techniques.

1.6 Refolding of Lysozyme

1.6.1 Lysozyme as Model Protein

Lysozyme is a frequently used model protein for refolding studies because of its well characterized structure. It is a small globular protein with 14.4 kDa and an isoelectric point (pI) at a pH of approximately 11. Furthermore lysozyme is well suited for studies concerning oxidative refolding due to its 4 disulfide bonds. Six tryptophan residues of which one is completely buried in the hydrophobic core of the molecule offer the possibility to study lysozyme folding by tryptophan fluorescence. Many publications describe refolding of lysozyme and reveal folding pathways of the protein as well as suitable process conditions.

1.6.2 Effects of Unfolding and Refolding Conditions

Published unfolding conditions for native lysozyme vary strongly in protein concentration, DTT concentration, unfolding time and temperature, whereas the buffer pH is always kept alkaline between pH 8 and 9 and either 8 M urea or 6 M guanidinium hydrochloride is used as chaotropic agent (Lin, Ruaan et al. 2007). Unfortunately the unfolded state of the protein is seldom characterized which makes comparison of refolding results difficult.

On the molecular basis two different folding pathways influenced by buffer conditions like the ionic strength are described (Bieri, Wildegger et al. 1999; Kulkarni, Ashcroft et al. 1999; Dobson 2004). According to these studies lysozyme undergoes a fast hydrophobic collapse and than the folding pathway branches into a fast track and a slow track. Both refolding tracks lead over a native-like structure to the native state.

In general a slower refolding kinetic is observed with higher protein concentrations in lysozyme refolding (Raman, Ramakrishna et al. 1996). Furthermore aggregation is increased with high lysozyme concentrations (Kiefhaber, Rudolph et al. 1991; Hevehan and Clark 1996; Clark, Hevehan et al. 1998; Buswell and Middelberg 2002) and can be suppressed by low concentrations of urea (~ 2 M) (Lanckriet and Middelberg 2004) and guanidinium hydrochloride (~ 1.75 M) (Hevehan and Clark 1996) without inhibiting native lysozyme formation. Formed aggregates in lysozyme renaturation with oxidative redox potential in the refolding buffer are disulfide linked and can thus only be solubilised in the presence of reducing agents (Ho, Middelberg et al. 2003). Lysozyme aggregation contributes to protein loss in refolding processes to a high extent but misfolding for example by choice of inappropriate redox conditions also plays a role in reduced yields (Roux, Ruoppolo et al. 1999).

In general the type of redox agents (Raman, Ramakrishna et al. 1996) and the ratio of oxidizing to reducing component (Ho, Middelberg et al. 2003; Lin, Ruaan et al. 2007) are significant parameters for refolding success. Air oxidation only leads to low renaturation yields due to mass transfer limitations (Clark, Hevehan et al. 1998; Ho, Middelberg et al. 2003).

Several publications deal with matrix assisted refolding of lysozyme on ion exchange chromatography columns (Li and Su 2002; Li, Zhang et al. 2002) or by size exclusion chromatography (Gu, Su et al. 2001; Lanckriet and Middelberg 2004). In ion exchange chromatography a dual gradient method with reducing urea concentration and concomitant increasing pH is described. With low elution flow rates and column loading refolding yield

could be improved. In addition a urea concentration gradient from 6 M and 1 M and a pH gradient from pH 6 to pH 10 were found to be suitable for high refolding yields (Li and Su 2002; Li, Zhang et al. 2002). In size exclusion chromatography the choice of an appropriate column material being able to fractionate the target protein from chaotrops and reducing agents and the composition of the equilibration buffer which is in this case the refolding buffer are mentioned to have a great influence on refolding success. After optimization of a batch size exclusion method a continuous chromatographic process is realized by annular chromatography. Despite intense optimization the yields in SEC and dilution refolding lie close together when comparing on the basis of equal final protein concentrations after refolding (Lanckriet and Middelberg 2004). An ascending urea gradient is used to promote re-solubilisation of aggregates formed during the refolding process in a SEC approach by a delayed running front of denaturing buffer. This additional process step increases the activity recovery compared to a SEC method without subsequent urea gradient between 20 and 40 % depending on the protein load (Gu, Su et al. 2001).

A comparison between chromatographic refolding methods and dilution refolding is difficult due to problems arising when trying to keep influencing parameters constant. Furthermore a significant objective function should be defined before making a decision on an appropriate refolding process. As a conclusion column based methods are an interesting alternative to dilution refolding because of higher space time yields although process development is still challenging due to the high complexity of the process.

2 Research Proposal

The introduction already points out the high complexity of protein refolding and its dependency on numerous interacting parameters. Nevertheless it becomes evident that a fast and efficient development and optimization of refolding processes is relevant for economic production of recombinant protein in industry as well as in research projects.

The empirical optimization of process parameters in manual screenings is the most wide spread method to design refolding processes. A major draw back of these empirical optimizations is the necessity to restrict the number of experiments by a restriction of the analyzed parameter space. As a consequence resulting processes are potentially operated far off the real optimum conditions. In the late 1990s scientists tried to gain a more systematic understanding of parameter effects on protein folding by the implementation of statistic methods for experimental design. For lack of automated screening platforms these studies were limited in number of experiments which had to be paid with a reduction of gained information. Only in the last few years automated high throughput experimentation was applied on the development of protein refolding processes still leaving a lot of challenging tasks open. One of the major and mainly unsolved problems in protein refolding screening is the automated, parallelized and quantitative measurement of protein folding to evaluate refolding success. On this basis one of the main aims of this work was the validation of analytical methods meeting the basic requirements of HTS: automation, parallelization and miniaturization. Refolding was performed by simple dilution of denatured protein in different refolding buffer systems. In this context the applicability of protein solubility as an indicator for protein folding was investigated. Due to the huge importance of solvent effects on protein solubility in refolding and the sparse data base on systematic characterization of parameters influencing protein aggregation this work was extended by an automated high throughput screening study on protein solubility behaviour in different buffer systems. Analytical development of the refolding screening platform also implied the validation of another more universal analytical method, the measurement of tryptophan fluorescence, to reduce time for development of specific assay technologies. The compatibility of analytical methods with high flexibility in buffer design and appearance of different protein folding states was validated.

Another important objective was the identification of process parameters which have to be controlled to guarantee for reproducible refolding results building a precondition for

parameter optimization. Besides, knowledge on the significance of process parameters is essential for the evaluation of published data.

The developed platform for high throughput experimentation should be combined with two methods for design of experiments: an evolutionary algorithm as a first step and in a second step a classic full factorial design. General advantages of an evolutionary algorithm like high robustness against experimental error and effective experimentation at a high number of parameters in a wide parameter space would thus be harnessed in refolding screening for the first time. The full factorial design covering a limited parameter space close to the optimum should be used as a tool to validate data from the evolutionary algorithm optimization. In summary the development of analytics and the investigation of methods for experimental design built the most challenging part for dilution refolding screening.

In matrix-assisted refolding the focus is laid on the technical transfer of the refolding process onto the robotic platform. Adsorption capacities and kinetics of denatured protein on different ion exchange resins should be investigated and compared to data gained with active protein. The development of an automated method for buffer exchange from denaturing to refolding buffer is a precondition for optimization of process parameters like pH, urea concentration and refolding time and one of the major challenges in matrix-assisted refolding screening.

3 Publications & Manuscripts

1. Automated High Throughput Technologies for Determination of Protein Solubility and Refolding Conditions

Annette Berg, Juergen Hubbuch

In this manuscript a review on published methods for protein solubility screening and protein refold screening is given and challenges for HTS development in these fields are pointed out. HTS compatibility of commonly used analytical methods is in the focus of this literature review.

2. Automated Measurement of Protein Solubility to Rapidly Assess Complex Parameter Interactions

Annette Berg, Maren Schuetz, Juergen Hubbuch

In this manuscript data from automated lysozyme solubility screenings are summarized and the effects of buffer concentration and pH as well as incubation temperature on lysozyme solubility are analyzed. Furthermore complex solubility surfaces are determined to investigate interdependencies between the concentration of buffer salt and of commonly used additives like polyethylene glycols of different molecular weight, Tween 20, sorbitol and sucrose. For an improved characterization of the solubility curves gained with the presented automated approach a comparison between thermodynamic solubility lines from literature and precipitation curves from HTS experiments is given and discussed.

3. Development and Characterization of an Automated High Throughput Screening Method for Optimization of Protein Refolding Processes

Annette Berg, Jörg Kittelmann, Jürgen Hubbuch

Development of methods for protein refold screening by dilution of denatured protein in refolding buffer is described intensely in this manuscript. Besides a miniaturization of the refolding process the focus was laid on validation of automatable and parallelized analytics to qantitatively measure refolding success. Furthermore crucial parameters during refolding are identified and the necessity of parameter control in refold screening is pointed out.

4. Automated Optimization of Protein Refolding Processes with an Evolutionary Algorithm

Annette Berg, Anna Siudak, Eric von Lieres, Jürgen Hubbuch

This manuscript deals with the implementation of a genetic algorithm for optimization of five refolding buffer parameters: buffer pH and the concentrations of NaCl, MgSO$_4$, GSSG and DTT to improve lysozyme solubility and activity. A reasonable combination of this evolutionary optimization method with full factorial experimental design in the optimum parameter range is shown. Furthermore a correlation between parameter optima for protein solubility and activity are evaluated.

5. Solid Phase Refolding

Annette Berg, Jörg Kittelmann, Jürgen Hubbuch

In this manuscript methods for automated high throughput screening of solid phase refolding parameters are presented. The adaptation of experimental procedures like liquid handling and mixing to highly viscous solutions of denatured protein are described. Adsorption of denatured lysozyme is analyzed by automated measurement of binding kinetics and isotherms. Furthermore the effect of protein loading, refolding time, refolding buffer pH and urea concentration are studied to confirm the applicability of the method as a screening tool for the automated optimization of solid phase refolding processes. A deeper insight into the interplay between protein adsorption/desorption and protein folding is gained with the performed batch experiments

Automated High Throughput Technologies for Determination of Protein Solubility and Refolding Conditions

Annette Berg, Juergen Hubbuch*

Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Science, University of Karlsruhe (TH), 76131 Karlsruhe, Germany

*Corresponding author. Tel.: +049 721 608-2557; fax: +049 721 608-6240. Email-adress: juergen.hubbuch@kit.edu

Abstract

High throughput screening (HTS) methods are a powerful tool for fast process development and optimization with low material consumption. Both parametes are critical issues determining economic success in biopharmaceutical industry due to shorter innovation cycles and the need to minimize production costs. Especially if a high number of process parameters needs to be considered automated high throughput technologies are the only means to cover the respective parameter space. When assessing protein solubility and refolding conditions no predictive physical model is known to facilitate optimization and consequently process development is based on empirical rules and experimental effort. In this paper high throughput screening methods for protein solubility and inclusion body protein refolding are reviewed. For all techniques described the focus was laid on a critical validation of published approaches according to the three basic guidelines for effective HTS: automation, parallelization and miniaturization including not only sample preparation but also analytics.

1 Introduction

The most valuable source of solubility data can be gathered from crystallization studies using nano-litre scale solution droplets. The needed concentrations of protein and solvent components are reached by controlled evaporation. Faster screening processes to evaluate precipitating conditions use evaporation in air atmosphere or ultrafiltration in microtiter plate formats. A bottleneck in solubility screening is the quantification of protein in solutions containing various substances being incompatible with commonly used protein assays. Furthermore up to now no method for fast, automatable and parallelizable buffer exchange with good and buffer independent protein recoveries is available. Refold and solubility screening are linked by high aggregation rates of folding intermediates occurring during refolding processes significantly decreasing active protein yields. A lack of universal, non-specific methods to measure structural integrity with low demands for time, sample volume and manual intervention is an obstacle for process optimization. Design of Experiments is used to decrease experimental effort. For systems with high number of parameters and parameter interdependencies evolutionary algorithms offer the possibility to find a global optimum in a wide parameter range.

2 High Throughput Screening

2.1 General Principles

High throughput screening can be characterized according to three basic principles: automation, parallelization and miniaturization. The aim lies in achieving a maximum amount of experimental data within a minimum time and material consumption.

While sample generation lies naturally in the heart of any screening process, a special focus during any establishment of HTS technologies should be laid on the choice of an appropriate analytical method. Analytics which are operated in sequential order, need manual intervention or special requirements for sample purity and amount should be circumvented. As described in Bensch et al. [1] with most techniques the extend of information gained with a specific analytical method can be correctly related to analytical complexity and time needed to perform the analysis. Therefore analytics used should be chosen according to the amount of

samples to be analysed. More sophisticated methods should be performed only in advanced screening procedures.

2.2 Choice of Measured Function

The famous rule of directed evolution "you get what you screen for" borrowed from You and Arnold [2] holds true for every other screening problem addressed. Thus a special interest lies in the choice of an appropriate objective function measured during the screening. With the restrictions of analytical methods adaptable in HTS formats a solution for this problem will often be a compromise between justifiable timeframes for the screening and the information gained. In general three different types of objective functions for the optimization of a process can be considered: quantitative measurement of the objective function of the process directly (for example solubility as mg protein per ml solvent for a solubility screening), qualitative classification of the objective function by measurement of a corresponding value (turbidity to classify protein samples as soluble or aggregated) and quantitative measurement of a function correlating with the objective function concerning favourable parameters (for example the concentration of solubilised aggregated protein in mg per ml solvent to calculate the amount of soluble protein). Obviously the first possibility is the most comfortable case as optimization can be performed on the basis of numerical data and screening towards the respective function of interest is performed directly. In the second case a lot of information is lost by using a qualitative method providing only a direction of parameter development when addressing system changes. The last case implies that the correlation of the true objective function of the process and the measured function is validated and thus optimization does not run into the wrong direction. The chosen example is trivial as it is based on the accepted rule of mass balances but for different functions careful evaluation experiments might be indispensable.

2.3 HTS for Solubility and Inclusion Body Refolding Parameter – Similarities between the two Screening Tasks

For protein solubility and refolding no predictive model is known to facilitate optimization and consequently process development is based on experimental procedures only. For both objectives, solubility and refolding, numerous affecting parameters are known which exhibit

interdependencies complicating systematic optimization approaches. Besides these similarities both processes are closely linked by solubility problems occurring in refolding processes. Though solubility screening approaches reviewed here are not related to protein during refolding the amount of soluble protein is in many cases used as an objective function in refolding screening or at least measured to asses the fraction of active protein gained. Problems in the quantitative determination of soluble protein have to be met for both, solubility and refolding screening, with even higher sample complexity in renaturation processes.

3 Solubility

3.1 Protein Solubility

Solubility of proteins is a major field of investigation during bioprocess development and still a challenge for fundamental research for its complexity. Aggregation does not only lead to lower yields of active protein but also creates technical problems in production and especially purification processes by clogging of filter membranes and chromatographic columns [3]. Furthermore solubility is a prerequisite in often highly concentrated biopharmaceutical formulations to guarantee for constant quality and accurate dosage [4]. During biopharmaceutical profiling in drug discovery the solubility of protein drugs plays a major role for bioavailability [5]. In protein structure analysis or in down stream processing by crystallization, solubility of proteins at high concentration levels is still one of the greatest obstacles [6].

It is important to discriminate between the two possible solid states of proteins, protein crystals and amorphous aggregates, also called precipitates. From the phase diagram in Figure 1 it becomes obvious that both processes, crystallization and precipitation, occur in supersaturated solutions with a protein and precipitant concentration above the thermodynamically stable solubility curve.
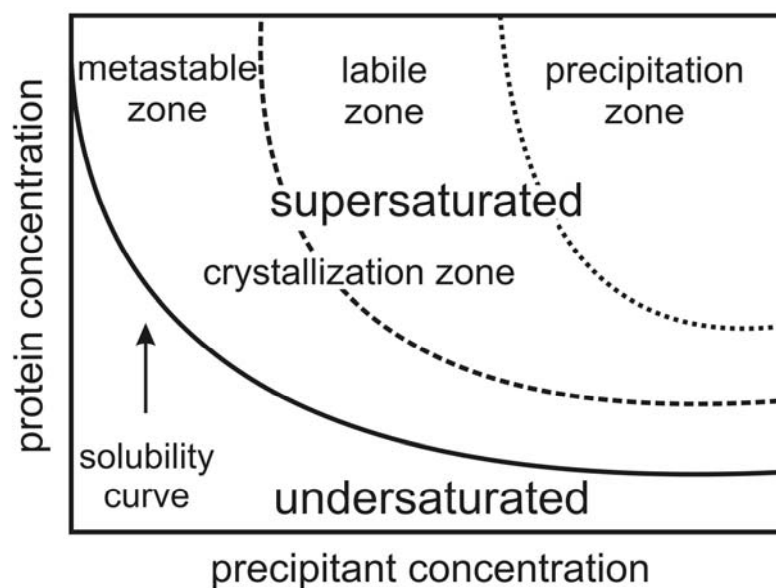
Figure 1: Schematic protein phase diagram (adapted from [7])

As solid phase formation is kinetically controlled, boundaries between supersaturated regions amenable to crystal nucleation, the metastable and labile phase, and the precipitation zone are dependent for example on the speed in which supersaturation is created or on particle content inside the solution [8, 9].

It should be taken into account that solubility is only one necessary criterion for buffer design despite several effects destabilizing proteins in solution. A review on protein solubility and stability is given in [10, 11].

Solubility of proteins depends on various parameters like solvent pH, ionic strength, added salt types and additives, temperature and of course on protein structure which leads to a high number of required experiments for optimization. High throughput screening methods offer the possibility to generate a high data density with low demands for material and time by automated, miniaturized and parallelized experiments [1]. On this basis a deeper understanding of factors affecting protein solubility and their interdependencies becomes possible.

3.2 Methods for Solubility Screening

In principle solubility screening parameters should be optimized to either prevent the formation of a solid protein state, for example in formulation screenings or in buffer design

studies for down stream processing steps, or to drive the protein into a solid and here usually crystallized state for structure analysis or preparative crystallization. As in every screening method development it first has to be decided what is the objective function and under which experimental conditions it can be analyzed. Screening for the formation of crystals will, due to the slow crystallization process, be much slower than screening for precipitation which is magnitudes faster.

3.2.1 Preparation of Screening Solutions

The first step of every method used is mixing of protein stock solutions or a solid protein preparation with the buffer system of interest. Solid protein states such as crystals are used when measuring the thermodynamic solubility line in equilibrium between crystal and soluble phase. If lyophilized or precipitated protein is used, a phase transition of these solid states into the solubilized species implies a stable protein structure in solid and soluble state. Moreover stable solid protein species such as crystals are however seldom available and an automated handling of solid protein is a quite complex task.

An important aim in the preparation of protein solutions for solubility screening is to reduce the overall material consumption. With a microfluidic formulator device for automated mixing of buffer and protein solutions volumes can be reduced to nanolitre scale [12-14]. A more wide spread method is the reduction of sample volumes to nanolitre drops in crystallization trails [15] or at least to volumes around 350 µl handled in microplates [15-17]. Commercially available robotic systems guarantee for accurate liquid handling and allow for the establishment of a fully automated screening process.

3.2.2 Screening Methods

It can be differentiated between incubation processes of protein solutions under static conditions with no change of protein or buffer concentration and incubation under dynamic conditions with increasing concentration of substances in solution. Methods for a concentration of protein in dynamic solubility screenings are summarized in Table 1.

Table 1: Concentration methods in solubility screening

Method	Limitations	HTS Compatible	Reference
Rate Controlled Evaporation (sitting/hanging drop/micro batch)	Slow due to vapour atmosphere or oil layer; all substances are concentrated, quantification of substances is difficult	Yes for crystallization (slow process imperative)	[15, 18]
Evaporation	Slow but faster than in vapour atmosphere; all substances are concentrated, quantification of substances is difficult	Yes	[16, 17, 19]
Ultrafiltration	Adsorption and aggregation of protein	Yes	[20]

A major draw back of the static approach is the restriction of protein and buffer concentrations in screening samples by the maximum solubility in stock solutions. This holds true for all approaches where a protein concentration shift from the undersaturated to the supersaturated phase has to be performed. This is the case for example in crystallization screenings, or if maximum protein solubility in the screening sample is higher than in the initial protein stock solution.

When considering dynamic situations the velocity of concentration should be adapted to the screening problem; i.e. in crystallization screening the supersaturated phase should be entered slowly to favour slow nucleation of crystals over fast precipitation. Dynamic methods either change the concentration of protein or concomitantly the concentration of all substances including salts. Quantification of protein in dynamic systems at maximum solubility is a problem discussed in the analytic section below.

All methods used for concentration of protein are compatible with high throughput principles as they can be automated and parallelized with the need for low sample volumes. Rate controlled evaporation frequently used for crystallization screening is a very slow process which is intended to meet long induction times for crystal nucleation. Evaporation in an air atmosphere is faster especially if the air-liquid surface is increased for example by shaking [16, 17, 21] or by generation of levitated sample drops in an acoustic field [19]. Shear forces inducing protein denaturation can be a problem if intense shaking is performed. An approach

using centrifugal concentrators to increase concentration velocity by centrifugal forces driving liquid through a membrane is presented in [20]. The last study however disregards possible protein adsorption at the filtration membrane falsifying the results obtained.

3.2.3 Measurement of Phase Diagrams

Measurements of different curves in phase diagrams differ mainly in the incubation time of protein solutions and in the rate of protein concentration. Thermodynamic solubility curve determination requires equilibrium between a solid crystal and a clear soluble protein phase which usually takes days to months. Generated data are independent from kinetic effects [9]. Nevertheless long-term storage to reach equilibrium conditions can only be justified in high throughput experimentation if thermodynamic solubility curves are indispensable.

Precipitation curves are preferentially determined for high speed and ease of detection [13, 14, 22] as precipitation is magnitudes faster than crystallization. Furthermore precipitation is one of the most important reasons for protein loss in downstream processing and formulation. Being dependent on the speed with which the supersaturated region of the phase diagram is reached kinetics of phase changes during screening should be adapted to the desired outcome. A compromise solution is published in [18], were the samples are incubated for 3 weeks to be close to equilibrium. A so called supersolubility line is constructed between conditions of crystallized or precipitated and clear samples separating the labile zone with spontaneous solid phase formation from the metastable or undersaturated zone.

3.3 HTS Analytics for Solubility Screening

3.3.1 Quantification of Soluble Protein

Direct quantification of the protein concentration of the samples obtained is most desirable. Measurement of soluble protein concentration however poses several problems in high throughput screenings as methods have to meet the needs for a compatibility to a wide range of buffers and additives, automation and parallelization or at least fast sample processing and low sample volumes. Commonly used methods are summarized in Table 2.

Table 2: Methods to quantify the concentration of soluble protein

Method	Principle	HTS Compatible	Reference
Colorimetric Assays	Protein in soluble supernatant after equilibration is measured.	Yes but with low flexibility in buffer composition	[23]
UV Absorption	UV absorption of soluble supernatant in equilibrium is measusred.	Yes but with low flexibility in buffer composition	[24]
SDS-PAGE/Native PAGE	Soluble and aggregated protein is stained and bands are quantified by densitometry or online detection.	Yes with capillary systems No with classical gel systems	[25-27]
Measurement of Sample Volume to Calculate Concentrations	Sample volume is measured by measurement of a corresponding value (liquid level, drop diameter).	Yes at least by conductive liquid level detection with robotic tips in 96 well plates.	[16, 17, 19, 21]
UV 28 0nm of Separated and Solubilized Aggregate	Aggregates are separated and solubilised with denaturant and reducing agent and measured at 280 nm.	Yes	[28]

Colorimetric assays as well as absorption at 280 nm might be incompatible with solvent components such as detergents, amino acids and reducing and oxidizing agents leading to low flexibility in screening conditions. Furthermore the signals in dye binging assays are sensitive towards structural varieties in protein samples, which especially during protein refolding and formulation studies has to be taken into account as inhomogeneous protein conformations have to be expected. A good overview of colorimetric assay techniques is given in [29]. SDS-PAGE analysis of soluble and re-solubilized aggregated protein commonly includes manual intervention for complex handling of gel or chip systems. Automated solutions for capillary electrophoreses are entering the market but are still not widespread.

An elegant method to assess protein and buffer concentrations is the measurement of sample volume at the time point of aggregation. In [19] for example an acoustically levitated sample

drop is mixed by a micro dispensing system and light scattering is detected online while water is slowly evaporating. Calculation of the sample volume and subsequently of protein and buffer concentration is possible by determination of the drop diameter with an imaging system in combination with the assumption of rotational symmetry of the droplet. A very similar approach is performed in [16, 21] generating protein buffer mixture with a liquid handling robot in microtiter plates. Sample volumes are detected over time by the conductive liquid level detection with the robotic tips. Turbidity is measured in parallel to identify the point of aggregation. The disadvantage of higher total protein and buffer consumption compared to the levitated drop system is compensated by low requirements concerning complex instrumental equipment and the opportunity for automation and parallelized sample preparation and measurement.

A method to estimate soluble protein content in refolding screening independent of structural variety and buffer incompatibilities is the separation of aggregate by filtration or centrifugation with subsequent solubilisation in denaturing buffer. The protein concentration in solubilized precipitates can thus be measured via absorption at 280 nm or a colorimetric assay and soluble protein content calculated from mass balances.

3.3.2 Buffer Exchange

Adaptation of effective methods for buffer exchange to a robotic workstation enlarges the applicability of established bioassays to different and even highly complex system compositions. Commonly used methods are summarized in Table 3.

Table 3: Methods for buffer exchange

Method	Limitations	HTS Compatible	Reference
SEC	Slow, high buffer consumption, protein dilution, long columns.	No	[30]
Ultrafiltration	Protein adsorption and aggregation.	Yes	[28, 31]
Dialysis	Slow, high buffer consumption.	No	[32-36]
Adsorption	Dependent on sample buffer.	Yes	[36, 37]

Methods for separation of proteins and low molecular weight compounds include size exclusion chromatography, ultrafiltration and dialysis. Also binding of protein or

contaminating substances to adsorber matrices leads to purified biological samples. As selective precipitation of active proteins is strongly dependent on buffer composition and optimization of the process is laborious, this method will not be discussed further. For size exclusion chromatography column length, particle size, particle pore size and ratio of sample volume to column volume are significant parameters for separation result [38]. Low sample loading, high sample dilution and serial buffer exchange are the major draw backs of size exclusion chromatography. Nevertheless gentle sample processing under diverse buffer conditions is possible using gelfiltration.

Ultrafiltration and dialysis harness the separation effect of membranes acting as molecular sieves. They differ in driving force for compounds able to pass the molecular barrier. In ultrafiltration liquid is forced through the membrane .High throughput systems like 96 well filter plates are evacuated by centrifugation or vacuum subsequently increasing protein concentration in the retentate. Protein solutions with various flow properties complicate flow control. Vacuum driven liquid transport strongly depends on equal liquid passing time for maintenance of the driving force, whereas the retentate volume in ultrafiltration by centrifugation is dependent on well position and applied speed profile [31]. Increased protein concentration in close proximity to the membrane often leads to loss of protein by adsorption or aggregation [39]. Varying buffer pH, salt composition and initial protein concentration are strongly related with loss of protein [28]. Required time for ultrafiltration depends on the molecular weight cut off of the membrane and on the parameters chosen for liquid transport like centrifugation speed or vacuum pressure.

In dialysis diffusion of molecules through the membrane leads to an equilibrium concentration of passing molecules between the sample compartment and a buffer reservoir of much higher volume dimension. Adsorption effects are related to dialysis time whereas aggregation of protein does not play a major role in sample loss. Time for dialysis depends on diffusion properties of the molecules, membrane area, thickness and composition, temperature, concentration gradient between buffer reservoir and sample and mixing on each side of the membrane [39].

Literature on buffer exchange as such but especially on buffer exchange in high throughput systems is limited to short notes in method descriptions. Up to now no automatable, parallelized system for size exclusion chromatography is described. Size exclusion by loading of microliter sample volumes into a low-pressure microfluidic device shows, that high sample throughput and good sample resolution are possible even with serial sample processing by miniaturisation [30].

Ultrafiltration in 96 well plates is evaluated and optimized for the separation of small molecule - protein complexes and small molecules free in solution [31].

Companies producing 96 well ultrafiltration plates at least provide the costumer with data on protein recovery, cross talk between wells and desalting efficiency for model proteins (Millipore, Pall). Nevertheless protein loss at the membrane and the depletion of disturbing buffer agents should be investigated for every protein. For example for lysozyme and BSA the protein yield after ultrafiltration is strongly dependent on buffer pH and protein concentration as described in [28].

Commercially available 96 well plate dialysis devices are used for complete sample desalting and removal of glycerol and arginine and show rather long dialysis times between 24 and 48 hours [35].

Binding of protein to metal chelate affinity resins offers the possibility for buffer exchange by washing the bound protein but is limited to constraints concerning binding and elution buffer properties [36, 37] which of course also holds true for all other adsorber materials.

Although being a prerequisite for many following process steps one can conclude, that development and validation of methods for buffer exchange seem to be underestimated or scientifically too trivial to be worth publishing. Consequently these problems have to be reinvestigated by every scientific group anew costing time and material resources. Specifications of commercially available devices are sometimes rare and should be checked for applicability to a set problem. Optimization might be necessary concerning protein recovery, contaminant depletion and process time. For quantitative buffer exchange gentle methods like size exclusion chromatography or binding of protein to affinity matrices or mixed mode adsorber resins should be further investigated and adapted to robotic work stations.

3.3.3 Classification of Samples According to Their Coordinates in Phase Diagrams

The coordinates of a sample in a phase diagram are defined either by measuring the point of protein phase transition or by measuring the existence of two phases. In most cases this means classification of a sample into soluble and aggregated or crystallized depending on the existence of the corresponding phase state of the protein. Most commonly used methods are based on changes in optical properties of a protein solution if crystals or aggregates are

present. A summary of techniques to classify samples according to their position in phase diagrams is given in Table 4.

Table 4: Methods to classify samples according to their position in phase diagrams

Method	Principle	HTS Compatible	Reference
Interferometry	Crystal dissolution is measured.	No	[24]
Turbidity (Absorbance at 350 to 600 nm)	Turbidity is detected by absorption of light.	Yes	[40-42]
Turbidity (Visual)	Turbidity is detected by visual inspection.	No	[40]
DLS	Particle size and distribution is measured.	Yes but slow	[40, 43]
SLS	Particle size and shape is measured.	Yes	[44]
Imaging Technology	Images of the solution are statistically analyzed for homogeneity.	Yes	[15, 19]
SEC-HPLC	Monomers and multimers are separated.	No	[45, 46]
DSC	Crystallization or aggregation enthalpies are measured.	No	[45, 47, 48]
Analytical Centrifugation	Sedimentation velocity of aggregates is measured.	No	[49, 50]

Interferometry can be used to assess if a crystal in solution is growing, dissolving or if equilibrium is reached. Complex instrumentation and the need for crystals are inconsistent with HTS applicability. Light scattering of solid particles in solution can also be measured by absorption at wavelengths between 350 and 600 nm for visual particles and by dynamic or static light scattering for even smaller particles like soluble protein aggregates. Imaging technologies substituting visual inspection of samples for turbidity can only distinguish between presence and absence of particles, whereas visual inspection is dependent on the experimenter. No quantitative information on the amount of aggregates or crystals can be gained with the described optical techniques being an obstacle to systematic optimization

approaches. SEC is also used to estimate the aggregate to monomer fraction separated in a gel matrix. SEC is restricted to soluble aggregates only. Being a serial analytical technique including slow flow rates and high buffer consumption SEC is not applicable in automated HTS approaches. DSC, a method to measure phase transition enthalpies, is also slow and automation is difficult due to complex instrumentation. The same holds true for analytical ultrafiltration measuring particle sizes and forms by their sedimentation velocity.

3.3.4 Thermodynamic Classification of Protein in Solution

A systematic approach to measure a proteins tendency to interact under the investigated solvent conditions is the determination of the second virial coefficient. Different analytical methods used to calculate the second virial coefficient are summarized in Table 5.

Table 5: Methods for thermodynamic classification of protein solutions

Method	Principle	HTS Compatible	Reference
DLS	Particle size and distribution is measured.	Yes but slow	[40, 43]
SLS	Particle size and shape is measured.	No (sample has to be separated on SEC first)	[44, 51]
SEC-HPLC	Monomers and multimers are separated.	No (not parallelizable with appropriate resolution)	[52]
SIC	Protein is bound to adsorber and retention time of the same protein is measured.	Yes but slow	[53, 54]

Analytical size exclusion chromatography [4], a standard method to analyse aggregation in biopharmaceutical products and formulations, is used in [52] to asses the second virial coefficient by a protein concentration dependency of the retention time.

In self-interaction chromatography (SIC) the protein of interest is covalently bound in a statistically random orientation on the adsorber surface. The same sort of protein is diluted in a mobile phase and the retention on the column gives qualitative information on the strength of the interaction between bound and soluble protein in the eluate buffer [54-56]. The serial processing of samples and difficulties in immobilisation of protein in a random orientation

keeping its structural integrity are major pitfalls of that approach. Nevertheless higher flow rates in comparison to size exclusion chromatography enhance sample throughput.

Static light scattering is used as a source to calculate the second virial coefficient in a study on aggregation behaviour of proteins during refolding [51].

In addition dynamic light scattering gives information on the monodispersity and an average particle size inside a solution hinting for example at invisible aggregates with a high potential to grow overtime [43, 57].

These spectroscopic techniques generally offer the possibility of parallelized analysis without sample loss in contrast to the demand for an appropriate sample volume and a serial and therefore time-consuming column based analytic.

The opportunity to automate the measurements and to integrate sample preparation and analytics in one processing station is strongly dependent on the construction of the analytical devices.

3.3.5 Design of Experiments for Solubility Screening

There are basically three approaches to set up solubility screening trials: a random distribution of parameter settings based on experience to directly find favourable conditions, a statistic and balanced design of experiments to gain insight into parameter effects and interdependencies and a design to measure partial or complete phase diagrams with maximum information on parameter effects for a reduced parameter number.

In the most frequently used and also commercially available method buffer systems are randomly generated with parameter values based on experience. This approach either yields the desired solvent properties or, if the parameters are not chosen correctly, the screening fails. Usually no quantitative information on parameter effects is gained or even aspired in such screening trails. In a second screening set up favourable parameters are combined or parameter concentrations optimized. One of the major draw backs of this method is the potential oversight of optimum conditions in unexpected parameter ranges.

This disadvantage should be met by a statistical design of experiment with a balanced fractional factorial design meaning that each parameter level is sampled an equal number of times including also binary combinations of parameters [58]. A qualitative score of this fractional factorial screening is used to select favourable parameter ranges for an advanced screening. This second screening then should yield data that can be fit in a response surface to calculate parameter values at the optimum. Success of such response surface optimization

protocols is strongly dependent on the reproducibility of the experimental method and on the quality of the surface fit which is especially a problem for unexpected or complex surfaces.

The third screening approach aims at a systematic knowledge of complete or at least partial phase diagrams which implies on one hand a higher number of experiments to get a reasonable resolution covering the whole parameter range but on the other hand also yields a good basis for optimization of buffer conditions. One of the major draw backs is the restriction to a low number of varying parameters due to high experimental effort. In [16] three parameters: protein, buffer and additive concentration were measured in small sample volumes in a single 96 well plate and phase diagrams were constructed as solubility surfaces. This approach is not useful for screening of high parameter numbers but the knowledge of phase diagrams for example for crystallisation trail design often leads to a huge increase of the success rate [13, 22].

Thermodynamic classification of proteins in solution by measurement of the second virial coefficient can help to reduce the parameter range interesting for formulation development [54] or crystallization screening (the so called crystallization slot) [56, 59] though being no guarantee for success. As methods to measure the virial coefficient are to a large extent experimentally complex and material and time consuming the screening usually is restricted to smaller parameter ranges and a full factorial experimental design for a small parameter number is performed to estimate parameter interactions and effects [56].

3.4 Conclusion and Outlook

High throughput screening techniques building the basis for high data densities should fulfil several requirements. Consumption of protein and buffer components should be minimized in combination with fast automated sample processing and analytics. None of the systems described fulfils all of these criteria. Microfluidic dispensers and mixers can provide automated sample processing with volumes in the nano litre range but measurement of maximum solubility of a protein is restricted to systems mixed from stock solutions. Dynamic methods with liquid evaporation solve this problem but complicate the estimation of protein and buffer concentrations. The measurement of sample volumes is an elegant way to calculate the concentrations of all components over time and to correlate these concentrations to solubility measured for example by turbidity. First steps into this direction are already published and offer a good starting point for further investigations.

In design of experiments there is a strong tendency towards more systematic screening approaches to gain a better understanding of solubility processes. In combination with improvements in molecular simulations, modelling for solubility prediction in different solvents should be considered a long-term objective.

Design of experiments helps to create more efficient screening protocols. Also elucidation of complex synergistic effects between solvent additives could be facilitated with statistical methods. Nevertheless evolutionary algorithms should be introduced in solubility screenings as they are superior to statistical experimental design if the investigated number of parameters exceeds three and parameter interdependencies are likely.

4 Inclusion Body Protein Refolding

As expression of eukaryotic proteins in *E.coli* often leads to formation of intracellular protein aggregates so called inclusion bodies subsequent solubilisation and refolding is indispensable to obtain active protein in its native conformation. Despite time consuming development of renaturation processes, *E. coli* is still a frequently used host for its simple and cheap cultivation and a well established system for genetic modification.

Refolding implicates the reduction of denaturant by simple dilution, dialysis or in a chromatographic column and, for cysteine containing proteins, supply with redox components for disulfide bridge formation. Choice of buffer composition is most important for preventing aggregation of hydrophobic intermediates and for keeping molecules flexible enough to be transformed into a stable native conformation.

In this field process development is rather dependent on experience and trial and error than on a known functional relation between active protein yield and numerous influencing parameters like buffer pH, salt concentration, salt type, redox components and various additives including poly ethylene glycols, detergents, amino acids, polyoles, sugars, organic solvent and ionic liquids in changing concentrations. Furthermore in structural biology a high number of different proteins have to be refolded showing great discrepancies in successful refolding conditions. Information on refolding process development is given in numerous reviews [60-64].

As aforementioned for solubility studies high throughput screening systems enable to perform empirical optimization in a time and material effective way.

4.1 Methods for Refold Screening

4.1.1 Preparation of Refold Screening Samples

Published screening approaches cover a wide range of refolding modes to reduce denaturant concentration. A summary of methods is given in Table 6.

Table 6: Summary of methods for refold screening sample preparation

Method	HTS Compatible	Reference
Dilution	Yes	[32-36, 41, 65-70]
Repeated Dilution	Yes	[32, 36]
Dialysis	No (slow, high buffer consumption)	[71]
Matrix-assisted Refolding	Yes (except for SEC)	[36, 37]
Dilution on a Chip	Yes	[69, 70]

Screening for appropriate buffer composition is most frequently performed by dilution [32-36, 41, 65-70]. Usually a 10 to 100 fold dilution of solubilised protein in refolding buffer is accompanied by fast and intense mixing.

Protein concentrations are typically low, in the range of 0.1 to 1 mg/ml to reduce aggregation of hydrophobic folding intermediates. Serial addition of denatured protein is used to increase the concentration of soluble protein and therefore to facilitate subsequent detection methods [32, 36]. Here waiting time between pulses can use up to several hours of experimental time depending on the folding kinetic of the protein and has to be evaluated carefully in order to reach a stable protein conformation in between addition steps.

In addition to rapid dilution dialysis is used for slow buffer exchange, but is for both refolding and subsequent sample preparation mostly restricted to manual sample processing and dependent on slow diffusion processes [32, 36]. A published step-wise buffer exchange for renaturation by dialysis in a 96 well plate format takes 6 to 8 hours equilibration time for each step [71].

Up to now, HTS refolding on adsorber matrices is mainly limited to His-tagged proteins bound to metal affinity resins [36, 37]. Adsorption and elution are performed in a batch mode. Intense washing of resin material after refolding leads to removal of buffer components disturbing protein assays. On the contrary imidazole for protein elution might also interfere

with analytical methods. Often used reducing agents in inclusion body solubilisation like DTT or low pH, ionic detergents and arginine in refolding buffers are not compatible with metal chelate affinity chromatography [37]. A first automated method to characterize effects of refolding buffer, resin type and residence time in refolding buffer on ion exchange materials in batch experiments is described for lysozyme in [72] using filter plates for easy buffer exchange.

The first attempts towards optimization on microchips by mixing solubilised protein with refolding buffer in nano-scale channels are published, but not applicable to protein refolding accompanied with high aggregation rates [73, 74].

4.1.2 Analytics of Refold Screening

Solubility as Objective Function

Analytical methods can be split into measurement of soluble protein content and techniques to gain structural information.

Concentration of soluble protein does not necessarily correlate with high yields of active protein but nevertheless minimization of protein aggregation is a necessary objective in a first screening [41, 69]. It could be observed in [75] that optimum conditions for lysozyme solubility and activity are correlating with closer parameter limits for activity.

In a variation of refolding by dilution, a reverse screening method investigates the effects of various additives on aggregation of native protein during partial unfolding in a basic buffer recipe. The suppression of protein aggregation is then correlated with a stabilisation of protein in refolding preparations [76]. With its high consumption of native protein the applicability of this reverse screening depends on the availability of active protein and produces additional costs. It has to be further evaluated if the behaviour of partially unfolded protein and refolding protein concerning aggregation can be universally correlated.

Measurement of Structural Integrity

Methods for structure estimation are often linked to sophisticated instrumentation and special requirements on sample quality and quantity. An overview of analytical techniques is given in Table 7.

Table 7: Summary of analytical methods to measure structural integrity of proteins

Method	HTS Compatible	Reference
CD	No	[33, 35, 37, 69]
NMR	No	[77]
RP-HPLC	No	[76]
SEC	No	[33, 35, 37, 69]
DLS	Yes but slow	[78]
ELISA	Yes	[79, 80]
SPR	Yes	[69, 81]
Intrinsic Fluorescence	Yes	[36, 73]
Fluorescent Dyes	Yes	[82-84]
Activity Assays	Yes	[28, 32, 65-68, 70, 76]
Limited Proteolysis	Yes with SDS-chip system	[34]

Serial analysis of protein solutions is time consuming and laborious. Consequently well known standard methods like CD (circular dichroism) spectroscopy, NMR (nuclear magnetic resonance) spectroscopy, reversed phase HPLC or analytical SEC (size exclusion chromatography) and DLS (dynamic light scattering) are not compatible with automated high throughput systems although frequently used for lack of more efficient strategies [32, 33, 35, 37, 69, 76].

Specific enzyme assays yield reliable information on protein structure integrity and are also automatable and easily parallelized. For this reason they are in most publication on high throughput screening development the chosen method. Quite often absolute enzyme activity is the exclusive objective function for optimization [32, 65-68, 70].

Methods like ELISA (enzyme linked immunosorbent assay) and SPR (surface plasmon resonance) harness specific binding of protein to binding partners like substrate molecules, antibodies or receptors and are already applied in high throughput refolding approaches [67, 69, 79-81]. A major disadvantage is that surface design and process set up are challenging tasks. In addition for proteins in genomic projects no function or binding partner is known and proteins like antibody fragments and growth factors industrially produced from inclusion bodies do not exhibit enzymatic activity. For functional assays restricted to one protein time for assay development should not be underestimated in comparison to just adaptation of universal structure based techniques.

For high throughput measurements intrinsic fluorescence of tryptophane [36, 73] or usage of fluorescent hydrophobic dyes like ANS are easy applied methods for structure estimation [82-84] though rarely used up to date in screening systems. Intrinsic fluorescence spectroscopy is limited to proteins with buried tryptophane residues in native conformation to allow for a characteristic red shift of fluorescence emission during unfolding due to solvent interaction with the fluorophore. Furthermore fluorescence intensity is sensitive to quenching buffer components like disulfides and to temperature shifts [82] and interpretation of spectra implies knowledge on structural properties or the fluorescent behaviour of native protein. Another problem is the hydrophobic collapse of the protein which is often a prerequisite for protein solubility when denatured protein is transferred to refolding buffer. As nearly all tryptophan residues are buried instantly, no structural variety can be observed in soluble protein populations [28].

Hydrophobic dyes bind to hydrophobic patches in folding intermediates and upon interaction with protein emit light with high intensities. As dye binding also depends on solvent characteristics applicability to different buffer systems should be carefully evaluated. Characteristics of fluorescent probe and formed hydrophobic surface in folding intermediates have to fit to allow for significant binding [82, 85].

Another non specific method to investigate structural properties of proteins is limited proteolysis, harnessing higher stability of tightly folded native protein structures against proteolytic cleavage [34]. A great draw back is the analysis of protein fragments performing SDS-PAGE after digestion reaction and a sometimes low stability of the native protein against protease digestion.

4.3 Design of Experiments for Refold Screening

Statistical design of experiments is up to now the method of choice to handle a high number of parameters in refold optimization screening [32, 33, 35, 41, 68, 70]. Sample number can be dramatically reduced by setting up a fractional factorial buffer matrix to estimate main effects and multifactor interactions [86]. In this connection the basic assumption, that the objective function is linearly dependent on the parameters and that the parameters do not interact is a simplification which is likely to lead to a suboptimal screening result. Moreover in a fractional or full factorial design with sparse data points in a rather wide parameter space the calculated optimum might be strongly affected by the experimental error. Only if boundary conditions for the considered parameter range are already known experimental design and

subsequent approximation of response values by surface response methodology can lead to optimized refolding conditions [65].

Evolutionary algorithms show a high potential to improve robustness of multi parameter optimizations despite experimental deviations and parameter interactions. A higher data density in the parameter region of interest with good results for the objective function leads to a good balance between gained information and experimental effort. In [75] it could be shown that a global optimum can be found despite for parameter interdependencies by use of a combination of the genetic algorithm with a full factorial in the optimum parameter range. The study also points out the need to validate the optimization algorithm itself, here by means of additional full factorial experiments, to adapt the algorithm´s parameters to the screening task. Frequently performed surface fits to quantify parameter effects showed to be difficult in the screened five parameter system due to optima on the parameter boundaries.

4.4 Conclusion and Outlook

Literature on HTS refolding deals in most cases with automated, parallelized refolding processes followed by manual or serial analytical techniques inconsistent with the idea of fully automated time efficient screening systems. Commercially available Kits (Pro-Matrix Protein Refolding Kit, Pierce; Quickfold™ Protein Refolding Kit, AthenaES™; Protein Refolding Kit, BioAssay™; FoldIt Screen, Hampton Research) consist of different buffer mixes whereas development of analytics is left to the customer. Specific activity or binding assays are up to now the only possibility for automated and parallelized estimation of structural integrity.

Investigations should be driven towards methods for high throughput parallel protein characterisation for example with structure sensitive probes or parallelized analytical chromatography. It should be focussed on non-invasive and universal techniques to circumvent protein specific assay development and sample alteration or loss during analytics.

Measurement of protein solubility in refolding samples offers a good starting point for buffer optimization and leads to a reduced number of interesting samples for more elaborate analytics. Solubilization and measurement of aggregated protein prevents disturbing effects of buffer and protein structure variability and can be fully automated.

Microfluidic devices for a serial screening of refolding buffer compositions with low sample volumes and fast sample processing and analytics are restricted to soluble samples to prevent the systems from clogging by protein aggregates.

Restrictions of statistic Design of Experiments like low efficiency with a high number of parameters and the need for a basic knowledge on the locality of optimum conditions should be bypassed. If the system is completely unknown or if optimization should integrate "unexpected" parameter values derivative free algorithms like the genetic algorithm are the most promising solution.

References

1. Bensch, M., P.S. Wierling, E. von Lieres, and J. Hubbuch, *High throughput screening of chromatographic phases for rapid process development.* Chemical Engineering & Technology, 2005. **28**(11): p. 1274-1284.

2. You, L. and F.H. Arnold, *Directed evolution of subtilisin E in Bacillus subtilis to enhance total activity in aqueous dimethylformamide.* Protein Engineering, 1996. **9**(1): p. 77-83.

3. Cromwell, M.E.M., E. Hilario, and F. Jacobson, *Protein aggregation and bioprocessing.* Aaps Journal, 2006. **8**(3): p. E572-E579.

4. Shire, S.J., Z. Shahrokh, and J. Liu, *Challenges in the development of high protein concentration formulations.* Journal of Pharmaceutical Sciences, 2004. **93**(6): p. 1390-1402.

5. Kerns, E.H. and L. Di, *Pharmaceutical profiling in drug discovery.* Drug Discovery Today, 2003. **8**(7): p. 316-323.

6. McPherson, A., *Introduction to protein crystallization.* Methods, 2004. **34**(3): p. 254-265.

7. Asherie, N., *Protein crystallization and phase diagrams.* Methods, 2004. **34**(3): p. 266-272.

8. Hu, H.Q., T. Hale, X.Y. Yang, and L.J. Wilson, *A spectrophotometer-based method for crystallization induction time period measurement.* Journal of Crystal Growth, 2001. **232**(1-4): p. 86-92.

9.   Garcia-Ruiz, J.M., *Nucleation of protein crystals.* Journal of Structural Biology, 2003. **142**(1): p. 22-31.

10.   Wei, W., *Instability, stabilization, and formulation of liquid protein pharmaceuticals.* International journal of pharmaceutics, 1999. **185**(2): p. 129-188.

11.   Pace, C., S. Trevino, E. Prabhakaran, and J. Scholtz, *Protein structure, stability and solubility in water and other solvents.* Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 2004. **359**(1448): p. 1225-1234.

12.   Lau, B.T.C., C.A. Baitz, X.P. Dong, and C.L. Hansen, *A complete microfluidic screening platform for rational protein crystallization.* Journal of the American Chemical Society, 2007. **129**(3): p. 454-455.

13.   Anderson, M.J., C.L. Hansen, and S.R. Quake, *Phase knowledge enables rational screens for protein crystallization.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(45): p. 16746-16751.

14.   Hansen, C.L., M.O.A. Sommer, and S.R. Quake, *Systematic investigation of protein phase behavior with a microfluidic formulator.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(40): p. 14431-14436.

15.   Bard, J., K. Ercolani, K. Svenson, A. Olland, and W. Somers, *Automated systems for protein crystallization.* Methods, 2004. **34**(3): p. 329-347.

16.   Berg, A., M. Schuetz, and J. Hubbuch, *Automated measurement of protein solubility to rapidly assess complex parameter interactions* in preparation, 2008.

17.   Wiendahl M., Voelker C., Husemann I., Krarup J., Staby A., Scholl S., and H. J., *A novel method to evaluate protein solubility using a high throughput screening approach.* 2008.

18.   Saridakis, E. and N.E. Chayen, *Systematic improvement of protein crystals by determining the supersolubility curves of phase diagrams.* Biophysical Journal, 2003. **84**(2): p. 1218-1222.

19.   Santesson, S., E.S. Cedergren-Zeppezauer, T. Johansson, T. Laurell, J. Nilsson, and S. Nilsson, *Screening of nucleation conditions using levitated drops for protein crystallization.* Analytical Chemistry, 2003. **75**(7): p. 1733-1740.

20. Gosavi, R.A., T.C. Mueser, and C.A. Schall, *Optimization of buffer solutions for protein crystallization.* Acta Crystallographica Section D-Biological Crystallography, 2008. **64**: p. 506-514.

21. Wiendahl M, Voelker C, Husemann I, and H. J, *High throughput screening of kinetic solubilities.* 2008.

22. Zhu, D.W., A. Garneau, M. Mazumdar, M. Zhou, G.J. Xu, and S.X. Lin, *Attempts to rationalize protein crystallization using relative crystallizability.* Journal of Structural Biology, 2006. **154**(3): p. 297-302.

23. Mikol, V. and R. Giege, *Phase diagram of a crystalline protein - determination of the solubility of concavalin a by a microquantitation assay* Journal of crystal growth, 1989. **97**(2): p. 324-332.

24. Nakazato, K., T. Homma, and T. Tomo, *Rapid solubility measurement of protein crystals as a function of precipitant concentration with micro-dialysis cell and two-beam interferometer.* Journal of Synchrotron Radiation, 2004. **11**: p. 34-37.

25. Bondos, S.E. and A. Bicknell, *Detection and prevention of protein aggregation before, during, and after purification.* Analytical Biochemistry, 2003. **316**(2): p. 223-231.

26. Won, C.M., T.E. Molnar, R.E. McKean, and G.A. Spenlehauer, *Stabilizers against heat-induced aggregation of RPR 114849, an acidic fibroblast growth factor (aFGF).* International Journal of Pharmaceutics, 1998. **167**(1-2): p. 25-36.

27. Curatolo, L., B. Valsasina, C. Caccia, G.L. Raimondi, G. Orsini, and A. Bianchetti, *Recombinant human IL-2 is cytotoxic to oligodendrocytes after in vitro self aggregation.* Cytokine, 1997. **9**(10): p. 734-739.

28. Berg, A., A. Kittelmann, and J. Hubbuch, *Development and characterization of an automated high throughput screening method for optimization of protein refolding processes* in preparation, 2008.

29. Sapan, C.V., R.L. Lundblad, and N.C. Price, *Colorimetric protein assay techniques.* Biotechnology and Applied Biochemistry, 1999. **29**: p. 99-108.

30.   Chirica, G., J. Lachmann, and J. Chan, *Size exclusion chromatography of microliter volumes for on-line use in low-pressure microfluidic systems.* Analytical Chemistry, 2006. **78**(15): p. 5362-5368.

31.   Zhang, J. and D.G. Musson, *Investigation of high-throughput ultrafiltration for the determination of an unbound compound in human plasma using liquid chromatography and tandem mass spectrometry with electrospray ionization.* Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences, 2006. **843**(1): p. 47-56.

32.   Armstrong, N., A. De Lencastre, and E. Gouaux, *A new protein folding screen: Application to the ligand binding domains of a glutamate and kainate receptor and to lysozyme and carbonic anhydrase.* Protein Science, 1999. **8**(7): p. 1475-1483.

33.   Chen, G.Q. and E. Gouaux, *Overexpression of a glutamate receptor (GluR2) ligand binding domain in Escherichia coli: Application of a novel protein folding screen.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(25): p. 13431-13436.

34.   Heiring, C. and Y.A. Muller, *Folding screening assayed by proteolysis: application to various cystine deletion mutants of vascular endothelial growth factor.* Protein Engineering, 2001. **14**(3): p. 183-188.

35.   Lin, L., J. Seehra, and M.L. Stahl, *High-throughput identification of refolding conditions for LXR beta without a functional assay.* Protein Expression and Purification, 2006. **47**(2): p. 355-366.

36.   Scheich, C., F.H. Niesen, R. Seckler, and K. Bussow, *An automated in vitro protein folding screen applied to a human dynactin subunit.* Protein Science, 2004. **13**(2): p. 370-380.

37.   Cowieson, N.P., B. Wensley, P. Listwan, D.A. Hume, B. Kobe, and J.L. Martin, *An automatable screen for the rapid identification of proteins amenable to refolding.* Proteomics, 2006. **6**(6): p. 1750-1757.

38.   Boschetti, E., *Advanced sorbents for preparative protein separation purposes.* Journal of chromatography, 1994. **658**(2): p. 207-236.

39.   Bowen, R., *Understanding flux patterns in membrane processing for protein solutions and suspensions.* Trends in biotechnology, 1993. **11**(11): p. 451-460.

40.   Mahler, H.C., R. Muller, W. Friess, A. Delille, and S. Matheus, *Induction and analysis of aggregates in a liquid IgG1-antibody formulation.* European Journal of Pharmaceutics and Biopharmaceutics, 2005. **59**(3): p. 407-417.

41.   Vincentelli, R., S. Canaan, V. Campanacci, C. Valencia, D. Maurin, F. Frassinetti, L. Scappucini-Calvo, Y. Bourne, C. Cambillau, and C. Bignon, *High-throughput automated refolding screening of inclusion bodies.* Protein Science, 2004. **13**(10): p. 2782-2792.

42.   Hu, H.Q., T. Hale, X.Y. Yang, and L.J. Wilson. *A spectrophotometer-based method for crystallization induction time period measurement.* in *8th International Conference on Crystallization of Biological Macromolecules.* 2000. Sandestin, Florida.

43.   Jancarik, J., R. Pufan, C. Hong, S.H. Kim, and R. Kim, *Optimum solubility (OS) screening: an efficient method to optimize buffer conditions for homogeneity and crystallization of proteins.* Acta Crystallographica Section D-Biological Crystallography, 2004. **60**: p. 1670-1673.

44.   Sluzky, V., J.A. Tamada, A.M. Klibanov, and R. Langer, *Kinetics of insulin aggregation in aqueous solutions upon agitation in the presence of hydrophobic surfaces* Proceedings of the National Academy of Sciences of the United States of America, 1991. **88**(21): p. 9377-9381.

45.   Fatouros, A., T. Osterberg, and M. Mikaelsson, *Recombinant factor VIII SQ - influence of oxygen, metal ions, pH and ionic strength on its stability in aqueous solution.* International Journal of Pharmaceutics, 1997. **155**(1): p. 121-131.

46.   Chang, L.Q., D. Shepherd, J. Sun, X.L. Tang, and M.J. Pikal, *Effect of sorbitol and residual moisture on the stability of lyophilized antibodies: Implications for the mechanism of protein stabilization in the solid state.* Journal of Pharmaceutical Sciences, 2005. **94**(7): p. 1445-1455.

47.   Darcy, P.A. and J.M. Wiencek, *Estimating lysozyme crystallization growth rates and solubility from isothermal microcalorimetry.* Acta Crystallographica Section D-Biological Crystallography, 1998. **54**: p. 1387-1394.

48.   Igarashi, K., M. Azuma, J. Kato, and H. Ooshima, *The initial stage of crystallization of lysozyme, a differential scanning calorimetric (DSC) study.* Journal of crystal growth, 1999. **204**(1-2): p. 191-200.

49.   Richards, J.P., M.P. Stickelmeyer, D.B. Flora, R.E. Chance, B.H. Frank, and M.R. DeFelippis, *Self-association properties of monomeric insulin analogs under formulation conditions.* Pharmaceutical Research, 1998. **15**(9): p. 1434-1441.

50.   Pekar, A. and M. Sukumar, *Quantitation of aggregates in therapeutic proteins using sedimentation velocity analytical ultracentrifugation: Practical considerations that affect precision and accuracy.* Analytical Biochemistry, 2007. **367**(2): p. 225-237.

51.   Ho, J.G.S., A.P.J. Middelberg, P. Ramage, and H.P. Kocher, *The likelihood of aggregation during protein renaturation can be assessed using the second virial coefficient.* Protein Sci, 2003. **12**(4): p. 708-716.

52.   Bloustine, J., V. Berejnov, and S. Fraden, *Measurements of protein-protein interactions by size exclusion chromatography.* Biophysical Journal, 2003. **85**(4): p. 2619-2623.

53.   Ahamed, T., B.N.A. Esteban, M. Ottens, G.W.K. van Dedem, L.A.M. van der Wielen, M.A.T. Bisschops, A. Lee, C. Pham, and J. Thommes, *Phase behavior of an intact monoclonal antibody.* Biophysical Journal, 2007. **93**(2): p. 610-619.

54.   Valente, J.J., B.G. Fryksdale, D.A. Dale, A.L. Gaertner, and C.S. Henry, *Screening for physical stability of a Pseudomonas amylase using self-interaction chromatography.* Analytical Biochemistry, 2006. **357**(1): p. 35-42.

55.   Patro, S. and T. Przybycien, *Self-interaction chromatography: A tool for the study of protein-protein interactions in bioprocessing environments.* Biotechnology and bioengineering, 1996. **52**(2): p. 193-203.

56.   Tessier, P.M., H.R. Johnson, R. Pazhianur, B.W. Berger, J.L. Prentice, B.J. Bahnson, S.I. Sandler, and A.M. Lenhoff, *Predictive crystallization of ribonuclease A via rapid screening of osmotic second virial coefficients.* Proteins-Structure Function and Genetics, 2003. **50**(2): p. 303-311.

57. Zhang, J. and X.Y. Liu, *Effect of protein-protein interactions on protein aggregation kinetics.* Journal of Chemical Physics, 2003. **119**(20): p. 10972-10976.

58. DeLucas, L.J., T.L. Bray, L. Nagy, D. McCombs, N. Chernov, D. Hamrick, L. Cosenza, A. Belgovskiy, B. Stoops, and A. Chait, *Efficient protein crystallization.* Journal of Structural Biology, 2003. **142**(1): p. 188-206.

59. Loll, P.J., M. Allaman, and J. Wiencek. *Assessing the role of detergent-detergent interactions in membrane protein crystallization.* in *8th International Conference on Crystallization of Biological Macromolecules.* 2000. Sandestin, Florida: Elsevier Science Bv.

60. Lilie, H., E. Schwarz, and R. Rudolph, *Advances in refolding of proteins produced in E-coli.* Current Opinion in Biotechnology, 1998. **9**(5): p. 497-501.

61. Clark, E., M, *Protein refolding for industrial processes.* Current opinion in biotechnology, 2001. **12**(2): p. 202-207.

62. Middelberg, A., *Preparative protein refolding.* Trends in biotechnology, 2002. **20**(10): p. 437-443.

63. Tsumoto, K., M, *Practical considerations in refolding proteins from inclusion bodies.* Protein expression and purification, 2003. **28**(1): p. 1-8.

64. Rinas, U., Ursula|Vallejo,L,Luis|,L,Luis, *Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins.* Microbial cell factories, 2004. **3**(1): p. 11.

65. Ahn, J.H., Y.P. Lee, and J.S. Rhee, *Investigation of refolding condition for Pseudomonas fluorescens lipase by response surface methodology.* Journal of Biotechnology, 1997. **54**(3): p. 151-160.

66. Sijwali, P.S., L.S. Brinen, and P.J. Rosenthal, *Systematic optimization of expression and refolding of the Plasmodium falciparum cysteine protease falcipain-2.* Protein Expression and Purification, 2001. **22**(1): p. 128-134.

67. Langenhof, M., S.S.J. Leong, L.K. Pattenden, and A.P.J. Middelberg, *Controlled oxidative protein refolding using an ion-exchange column.* Journal of Chromatography A, 2005. **1069**(2): p. 195-201.

68.   Willis, M.S., J.K. Hogan, P. Prabhakar, X. Liu, K. Tsai, Y.Y. Wei, and T. Fox, *Investigation of protein refolding using a fractional factorial screen: A study of reagent effects and interactions.* Protein Science, 2005. **14**(7): p. 1818-1826.

69.   Cowan, R.H., R.A. Davies, and T.T.J. Pinheiro, *A screening system for the identification of refolding conditions for a model protein kinase, p38 alpha.* Analytical Biochemistry, 2008. **376**(1): p. 25-38.

70.   Rahimpour, F., G. Mamo, F. Feyzi, S. Maghsoudi, and R. Hatti-Kaul, *Optimizing refolding and recovery of active recombinant Bacillus halodurans xylanase in polymer-salt aqueous two-phase system using surface response analysis.* Journal of Chromatography A, 2007. **1141**(1): p. 32-40.

71.   Wang, T., S. John, S. Archuleta, and C.B. Jonsson, *Rapid, high-throughput purification of HIV-1 integrase using microtiter plate technology.* Protein Expression and Purification, 2004. **33**(2): p. 232-237.

72.   Berg, A., A. Kittelmann, and J. Hubbuch, *Automated characterization of denatured protein adsorption and refolding on adsorber matixes.* in preparation, 2008.

73.   Kerby, M.B., J. Lee, J. Ziperstein, and A. Tripathi, *Kinetic measurements of protein conformation in a microchip.* Biotechnology Progress, 2006. **22**(5): p. 1416-1425.

74.   Zaccai, N.R., K. Yunus, S.M. Matthews, A.C. Fisher, and R.J. Falconer, *Refolding of a membrane protein in a microfluidics reactor.* European Biophysics Journal with Biophysics Letters, 2007. **36**(6): p. 581-588.

75.   Berg A., Siudak A., von Lieres E., and J. Hubbuch, *Automated optimization of protein refolding processes with an evolutionary algorithm.* in preparation, 2008.

76.   Ejima, D., K. Ono, K. Tsumoto, T. Arakawa, and Y. Eto, *A novel "reverse screening" to identify refolding additives for activin-A.* Protein Expression and Purification, 2006. **47**(1): p. 45-51.

77.   Chiku, H., A. Kawai, T. Ishibashi, M. Takehara, T. Yanai, F. Mizukami, and K. Sakaguchi, *A novel protein refolding method using a zeolite.* Analytical Biochemistry, 2006. **348**(2): p. 307-314.

78. Gast, K., D. Zirwer, and G. Damaschun. *Time-resolved dynamic light scattering as a method to monitor compaction during protein folding.* in *Workshop on Data Evaluation in Light Scattering of Polymers (LS '99).* 1999. Bad Schandau, Germany: Wiley-V C H Verlag Gmbh.

79. Fahey, M.S., D. Dawbarn, S.J. Allen, I.C. Paterson, and S.S. Prime, *Expression of recombinant extracellular domain of the type II transforming growth factor-ss receptor: Utilization in a modified enzyme-linked immunoabsorbent assay to screen TGF-ss agonists and antagonists.* Analytical Biochemistry, 2001. **290**(2): p. 272-276.

80. Leong, S.S.J. and A.P.J. Middelberg, *The refolding of different alpha-fetoprotein variants.* Protein Science, 2006. **15**(9): p. 2040-2050.

81. Jones, D.B., M.H. Hutchinson, and A.P.J. Middelberg, *Screening protein refolding using surface plasmon resonance.* Proteomics, 2004. **4**(4): p. 1007-1013.

82. Engelhard, M. and P.A. Evans, *Experimental investigation of sidechain interactions in early folding intermediates.* Folding & design, 1996. **1**(2): p. R31-R37.

83. Lakowicz, J., *On spectral relaxation in proteins.* Photochemistry and photobiology, 2000. **72**(4): p. 421-437.

84. Royer, C.A., *Probing protein folding and conformational transitions with fluorescence.* Chemical Reviews, 2006. **106**(5): p. 1769-1784.

85. Cardamone, M. and N.K. Puri, *Spectofluorimetric assessment of the surface hydrophobicity of proteins* Biochemical Journal, 1992. **282**: p. 589-593.

86. Qoronfleh, M.W., L.K. Hesterberg, and M.B. Seefeldt, *Confronting high-throughput protein refolding using high pressure and solution screens.* Protein Expression and Purification, 2007. **55**(2): p. 209-224.

Automated Measurement of Protein Solubility to Rapidly Assess Complex Parameter Interactions

Annette Berg, Maren Schuetz, Juergen Hubbuch*

Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Science, University of Karlsruhe (TH), 76131 Karlsruhe, Germany

*Corresponding author. Tel.: +049 721 608-2557; fax: +049 721 608-6240. Email-adress: juergen.hubbuch@kit.edu

Abstract

Protein solubility is one of the most important objective functions during optimization of biopharmaceutical production conditions. The characterization of protein solubility is important for either preventing protein aggregation which is the case in downstream processing steps like inclusion body refolding, hydrophobic interaction chromatography and liquid drug formulation, or to decrease solubility as is aimed for in protein purification by precipitation or crystallization.

In our study we used a high throughput screening method automated on a Tecan liquid handling robot to rapidly assess the solubility of lysozyme and its dependence on parameters like pH, ionic strength and additives. Combinatorial parameter effects could be measured in a reasonable time frame of approximately 1 day yielding a high data basis with low material consumption.

Temperature and distance of pH from the isoelectric point showed a positive coupled correlation with solubility. In addition we found an influence of ionic strength on the solubility changes induced by additives for all studied systems. Polyethylene glycol (PEG) 300 and Tween 20 were found to improve lysozyme solubility at higher salt concentrations. The addition of sorbitol and sucrose resulted in two distinct solubility maxima at low salt concentrations. While an explanation for single parameter effects on protein solubility was possible for example for pH by correlation of net charge and solubility at different pH values this becomes more and more difficult with an increasing number of parameters. By reducing the experimental effort needed it is possible to build a solid data basis to elucidate solvation mechanisms. Automated high throughput methods are thus a powerful tool not only for process optimization but also for a better understanding of precipitation processes.

1 Introduction

Protein solubility plays a major role in bioprocess development when considering aggregation of the target but also of the contaminating protein as it not only results in lower yields but also due to its possibly disturbing effect on process steps such as clogging of membranes and chromatographic columns. Especially in inclusion body refolding processes, aggregation due to strong interactions between folding intermediates with higher surface hydrophobicity is one of the major obstacles during process design and buffer optimization. In formulation studies solvent compounds are investigated for their effect on long term protein solubility to guarantee for constant product quality. An overview of protein aggregation and its consequences on bioprocess development is given in two recommendable reviews [1, 2].

Most of the published data deals with solubility in crystallization trails for protein structure determination. Although a high interest exists among fundamental researchers and industrial process designers, it is up to now not possible to describe parameter effects on protein solubility behaviour by a simple model. Therefore optimization has to be performed empirically with some guidelines based on experience. With an increasing number of parameters high throughput methods have to be established to reduce costs emerging from high material and time consumption during experimentation.

In Figure 1 a schematic protein phase diagrams is given. Screening of solubility is often performed in an equilibrium between aggregated and soluble protein in order to gain information on the thermodynamic solubility curve below which the protein shows no tendency to build a solid [3]. For relatively fast processes with rapid changes in buffer composition or protein concentration, the protein will be presumably far off the equilibrium. As the nucleation kinetics of crystals in the supersaturated region is dependent on the rate at which supersaturation is created, precipitation is more likely than crystal formation due to longer induction times for nucleation in comparison to usually fast precipitation processes [3, 4]. Therefore knowledge of the precipitation curve is generally sufficient for appropriate experimental design in bioprocess development. Even for studies on long-term storage stability the precipitation curve describes the upper border of a potentially suitable parameter space and should be assessed as a starting point due to simplicity and speed of its estimation. Knowledge of the phase diagram including the precipitation zone is a prerequisite for a rational screening design [5-7] to optimize crystallization of proteins for structure determination. In purification processes like chromatography, protein concentration and solvent composition change faster, and as a consequence no solubility equilibrium in solution

is reached. After a fast shift of the protein into the supersaturated phase, e.g. by addition of highly concentrated precipitants, formation of crystals or aggregates might not occur directly as it is dependent on the kinetics of the aggregation process. Nevertheless protein solutions in the supersaturated phase show a limited storage stability of the solution which has to be taken into account for example if the protein solution has to be stored between two process steps [3].
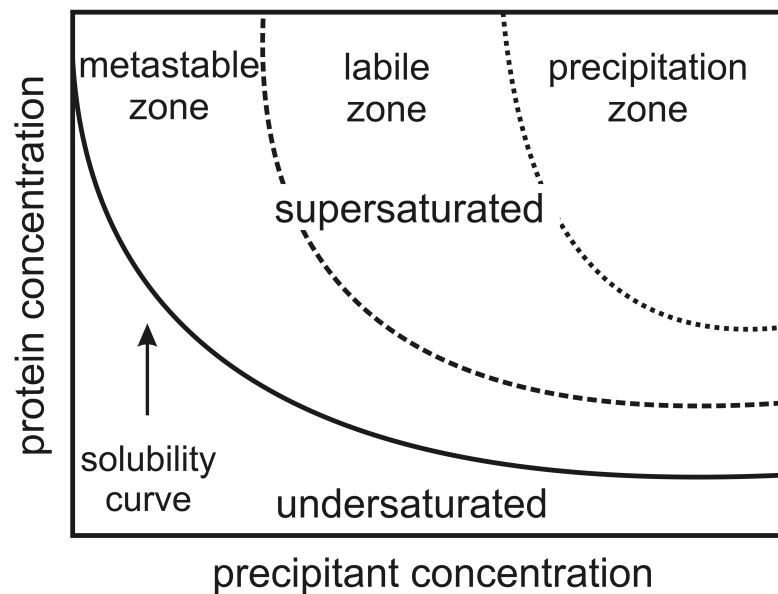

Figure 1: Schematic protein phase diagram (adapted from [3])

For the optimization of protein solubility numerous parameters have to be considered which can be divided into physical parameters and the chemical environment of the protein, *i.e.* the "solvent composition". Important physical parameters affecting protein solubility are mechanical stress induced for example by intense mixing, surface or interface adsorption and temperature. Important chemical parameters are solvent pH, ionic strength and salt type. As proteins display differences in net charge at different pH values according to their isoelectric point (pI) the optimum pH for high protein solubility is protein specific. Effects of ionic strength and salt type on solubility are complex and strongly dependent on the solution pH [8]. An empirical hierarchy of salt types being based on their effect on protein solubility was first published by Hofmeister et al. [9] and since then frequently verified by other researchers though pH was not adjusted in his studies, since the concept of pH values had not been established at that time yet. Ionic strength dependencies can be roughly divided into a region with rather low salt concentrations where salting-in of proteins dominates and a region with higher salt molarities where a higher tendency of the proteins to precipitate is observed, called

salting-out effect. Solubility curves describing the salting-out effect can be empirically fitted according the Cohn equation (equation 1) [10] with S as protein and c as salt concentration.

$$\ln S = \beta - K_S \cdot c \qquad \textbf{(1)}$$

A hypothetical maximum protein solubility β in the absence of salt is induced in which the salting-in effect of some salt protein systems is not considered. The slope of the straight line resulting from plotting the logarithmic protein concentration and the salt concentration can be interpreted as the salt specific constant K_s which is dependent on the protein salt system but independent of pH and temperature. In contrast to K_s, β is a function of both temperature and pH.

Another parameter to be mentioned are additives which frequently are added for example in biopharmaceutical formulations and inclusion body protein refolding processes to improve solubility, but also in protein precipitation approaches to reduce solubility. These additives belong to different groups of molecules and their effect on protein solubility is not well understood up to now. Additives can roughly be divided into at least the following types of substances: sugars and polyols (like sucrose or sorbitol), amino acids (like arginine), polymers (like PEG) and surfactants (like Tween 20). Sometimes they are also called osmolytes due to their high abundance in organisms viable under environmental stresses like high temperature or ionic strength [11]. The interdependencies of the effect on protein solubility between additive concentrations and other important parameters like pH and ionic strength have to our knowledge not been characterized in literature before.

Protein stability and solubility are strongly linked, as for example proteins during unfolding tend to aggregate due to a higher degree of hydrophobic residues accessible at the molecular surface. Thus protein structure can also be used as another important parameter for the optimization of solubility, restricting the parameter space to a range preserving comparable protein structure properties. Furthermore, one has to deal with a mixture of folding intermediates and the active structure when optimizing protein solubility for inclusion body protein refolding processes, which potentially complicates the search for favourable buffer systems.

In this study we show data resulting from an automated high throughput screening method using evaporation of water to concentrate protein and buffer over time. Since time and material can be reduced, the measurement of precipitation curves can be performed in one day. We compared the location of our precipitation curve with that of published

thermodynamic solubility curves to show the difference between both approaches. It is demonstrated that even complex parameter interdependencies for example between buffer salts and detergents can be assessed within short timeframes.

2 Material and Methods

2.1 Material

2.1.1 Robotic Work Station

A Freedom Evo 200 (Tecan Crailsheim, Germany) equipped with one liquid handling arm and one gripper was used as an automated pipetting station. Pipetting was performed with 8 fixed standard tips. A H+P Thermoshake with one position and a shaking diameter of 2 mm (Inheco Munich, Germany) and a spectrophotometer InfiniTe 200 (Tecan Crailsheim, Germany) were integrated into the robotic platform. A temperature-controlled microplate carrier with three positions was connected to a F12-ED refrigerated/heating circulator (Julabo, Seelbach, Germany).

2.1.2 Disposables

Microtiter plates with a well volume of 360 µl were purchased from Greiner (Frickenhausen, Germany).

2.1.3 Chemicals and Proteins

All chemicals had analytical grade and were purchased from Sigma Aldrich (St. Louis, USA) as well as hen egg white lysozyme (L-6876) with $\geq 90 \%$ protein content.

2.2 Methods

2.2.1 Preparation of Stock Solutions

Buffer pH was adjusted by titration of acidic and basic components. Sodium chloride solutions were titrated with HCl. Lysozyme solution was prepared by solubilization of lysozyme in Milli-Q water by stirring at 300 rpm on a magnetic MR 2000 stirrer (Heidolph, Kehlheim, Germany). The composition of additive stock solutions is given in Table 1 and the composition of salt stock solutions is given in Table 2.

Table 1: Composition of additive stock solutions

Additive	Concentration Stock Solution
PEG 300	0.067 M
PEG 3000	0.067 M
PEG 8000	0.067 M
Sorbitol	4 M
Sucrose	2 M
Tween 20	0.6 %

Table 2: Composition of buffer salt solutions

Buffer Salt	pH	Composition	Concentration Stock Solution
Potassium phosphate Buffer (KPi)	5.0 / 6.0 / 7.0 / 8.0 / 9.0 / 10.0	KH_2PO_4 / K_2HPO_4	1.5 M
Sodium chloride	4.3, 6.0, 6.5, 8.4	NaCl / HCl	0.5 M

2.2.2 Automated Determination of Precipitation Curves

A modification of the method described in Wiendahl et al. [12] was used. A flow scheme of the robotic process is given in Figure 2 showing steps for sample preparation, concentration of protein and solvent components by liquid evaporation and analytical steps like measurement of turbidity and sample volume.
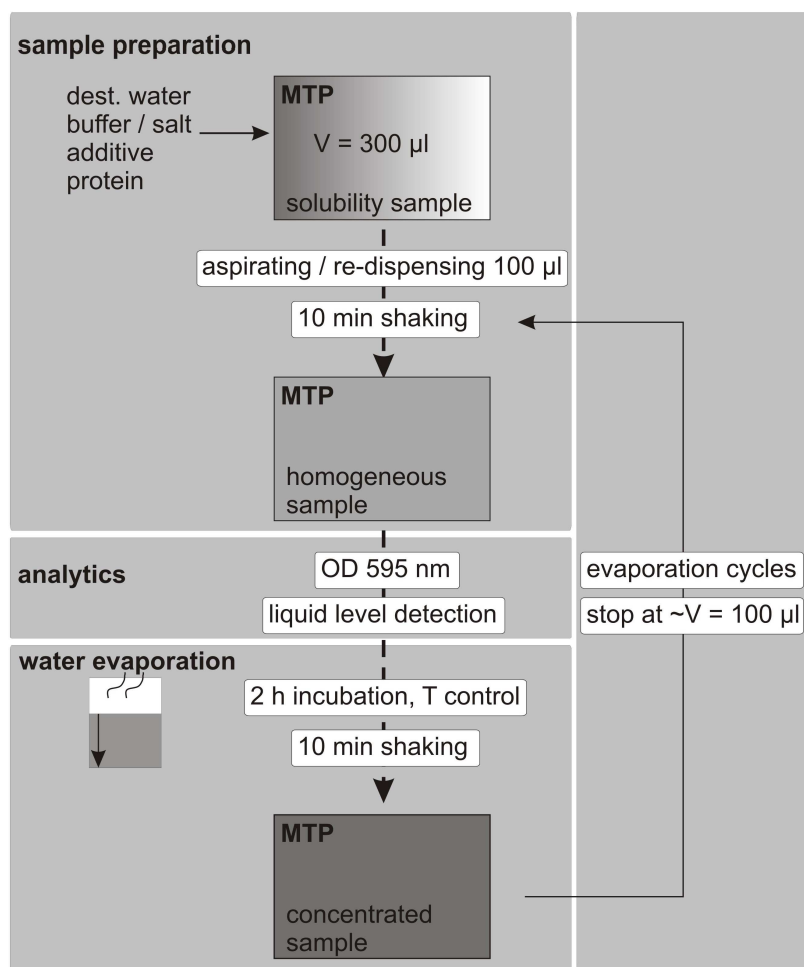
Figure 2: Flow scheme of the automated solubility screening process

Initial solution conditions were generated by preparing a mixture of Milli-Q water, buffer or salt, additive and lysozyme stock solutions in the given order in a microtiter plate. The sample volume was 300 µl. Mixing of samples was performed by aspirating and re-dispensing of 100 µl and shaking on the thermoshaker at 1020 rpm for 10 minutes. The absorption at 595 nm was measured and the volume of the samples was determined by conductive liquid level detection with the robotic tips. After incubation of the microtiter plate for 2 h on a

temperature controlled carrier the samples were transferred onto a thermoshaker for homogenization by 10 minutes of shaking. In the following, absorption measurements at 595 nm and liquid level detection were performed. The cycle is repeated 8 times to reach a sample volume of around 100 µl and resulting in an increase in protein and buffer concentration by a factor of approximately 3. Three 96 well microtiter plates can be processed in parallel. A general scheme for solubility determination is shown in Figure 3 A. For the measurement of additive effects 16 buffer protein systems close to the maximum solubility without additives are generated as given in Figure 3 B. Six of these 16 identical buffer and protein solutions are generated and supplemented with different additive starting concentrations including one set without additive addition.

With measurement of eight evaporation cycles a total of 768 or, for experiments with additives, 128 different conditions are analyzed using turbidity measurements.
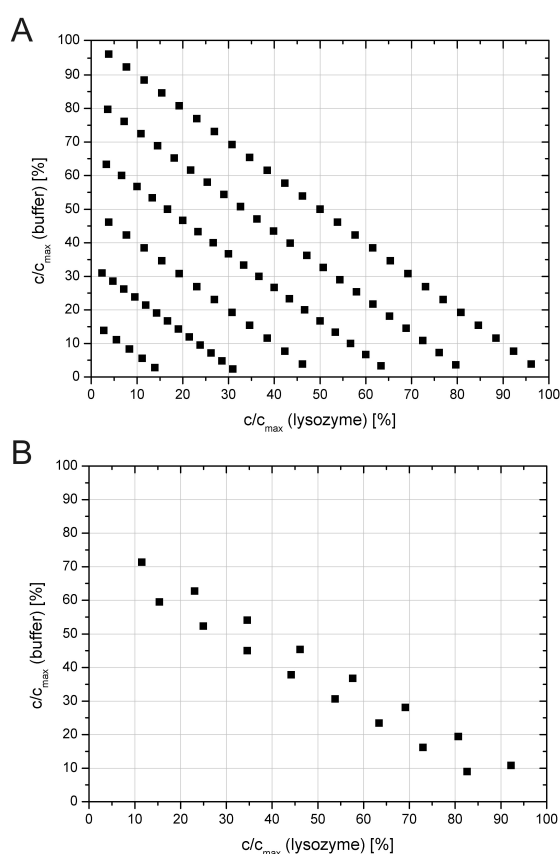

Figure 3: A Experimental starting conditions for the determination of solubility curves with different buffer or salt systems and lysozyme. **B** Experimental buffer and lysozyme starting concentrations for the determination of solubility curves with addition of additives.

2.2.3 Data Evaluation

For every measuring cycle the concentration of buffer, salt, additive and protein is calculated from the starting concentrations and the determined volume according to equation 2 with $c(t)$ as concentration at time t, c_0 as initial concentration, $V(t)$ as the liquid volume at time t and V_0 as initial liquid volume.

$$c(t) = \frac{c_0 \cdot V_0}{V(t)} \tag{2}$$

Samples are classified as insoluble if their absorption at 595 nm exceeds a static threshold value of 0.08 or if the absorption increment between two cycles exceeds 0.006. The center of the last soluble and the first precipitated condition is used as a data point. A curve fit according to the empirical Cohn equation [10] was calculated from the collection of data points for each condition. Equation 3 depicts the Cohn equation with c as salt concentration, S as protein concentration, β as hypothetical maximum solubility without salt and K_S as a salt specific constant.

$$S = \beta \cdot \exp(-K_S \cdot c) \tag{3}$$

Solubility data with addition of additives are visualized in surface plots respectively.

3 Results and Discussion

3.1 Comparison of Precipitation and Thermodynamic Solubility

To compare thermodynamic solubility curves for lysozyme and results gained with the robotic approach we used data published by Rettailleau et al. [13]. In [13], samples had to be stored for weeks to months to reach thermodynamic equilibrium between protein in solution and protein crystals. In the contrary, the time frame needed to gain the solubility data applying the robotic approach was approximately 22 hours which with a high probability is too short to reach thermodynamic equilibrium. Furthermore the experiments to estimate the

thermodynamic solubility curves started from a solution containing crystals in contrast to the robotic approach where undersaturated solutions were used as starting points.

The data for lysozyme solubility curves from literature and from our robotic experiments in NaCl at pH 4.3, 6.5 and 8.4 at 18 °C are plotted in Figure 4 A to C. All precipitation curves show a declining asymptotic behavior of lysozyme solubility with increasing salt concentration. This observation is also referred to as salting out effect which was for the first time published by Hofmeister in [9] and up to now no simple theory is able to account for it [18]. No curve crossing of the solubility curves from literature and from the robotic solubility measurements is observed. At all pH values the solubility curve from literature has a stronger curvature and exhibits a steeper decline in the supernatant with increasing salt concentration in comparison to the data obtained with the robotic system.
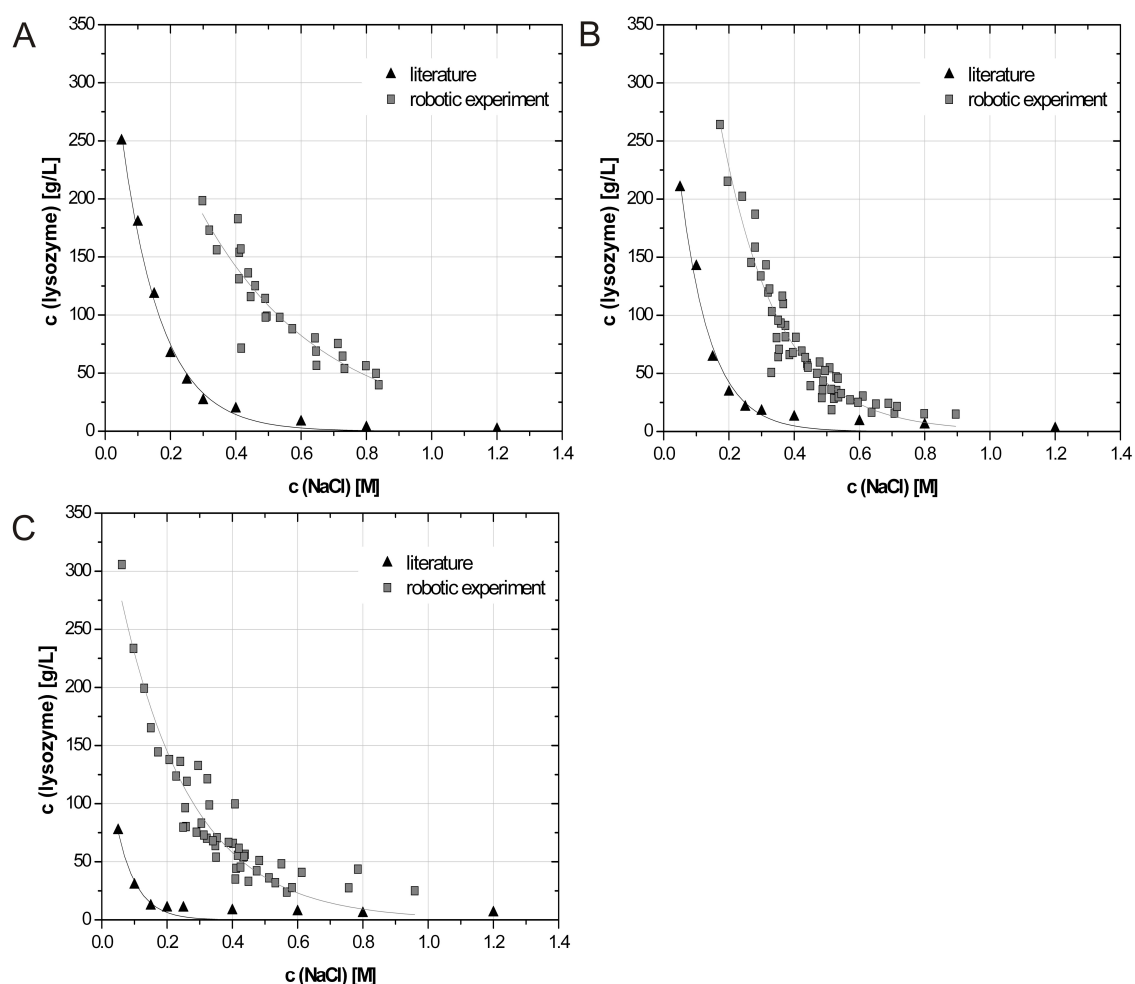

Figure 4: A Comparison of the thermodynamic solubility curve from literature **[13]** and the solubility curve determined with the robotic experiments incubated at 18 °C for NaCl at pH 4.3; **B** for NaCl at pH 6.5; **C** for NaCl at pH 8.4.

These differences are also represented by the fit parameter β, a measure for the curvature and the K_s value, a measure for the slope of the solubility curve. The fit parameters are summarized in Table 3 according to the empirical Cohn equation. The standard error of the parameter fit lies between 4 and 17 % for values from [13] and between 4 and 13 % for our results pointing out that a curve fit with the Cohn equation can be applied to data from both experimental approaches.

An average standard deviation was determined for fitted parameters from the automated measurements discussed. The experimental error was calculated for Cohn parameters from three different microtiter plates in a single experimental run and from three experiments performed completely independent from each other. In the case of fit parameters from one experimental run with three different microtiter plates for the determination of β an experimental error between 6 and 16 % was obtained and for the determination of K_s an experimental error between 6 and 17 %. For independent experiments an experimental error of 20 % is calculated for β and an error of 14 % for K_s hinting at a huge impact of slight differences in prepared stock solutions on β. Based on these observations we only compared data from experiments which were performed in a single experiment. Despite the lack of information on the experimental error of the published results used in this section, the Cohn parameters are compared in the following section.

The K_s value from [13] are 203 % (pH 6), 92 % (pH 7) and 262 % (pH 8) higher than the K_s value from our experiments. A higher value of the salt specific constant K_s is physically related to a stronger decrease of the protein concentration in the supernatant with increasing salt concentration. Thus the published solubility curves show a higher dependency of protein solubility on salt content. This discrepancy between the K_s values indicates differences between the measured phase transitions in the compared approaches.

Furthermore β values from [13] are a magnitude of 8 % (pH 6), 47 % (pH 7) and 52 % (pH 8) smaller in comparison to our results. This can also be observed in Figure 4 as a stronger curvature of the curves from literature data being strongest at pH 8.

No correlation between pH and the K_s values can observed in equations derived from HTS data whereas equations in [13] show an increase of K_s values with increasing pH. A dependency of K_s on pH is in contrast to the definition of K_s as a salt-protein-system specific constant independent from temperature and pH.

The parameter β only shows a dependency from pH for the literature data set which is in agreement with the definition in the Cohn equation. With increasing pH the β value decreases. As β represents the theoretical maximum solubility in the absence of salt an increase in β at

lower pH can be explained by a higher net charge of lysozyme with higher distance of the pH from the isoelectric point lying approximately at pH 11. Data gained by means of automation are not capable of resolving this effect.

Table 3: Cohn curve fit parameters for the solubility curves of lysozyme from [13] and from robotic experiments in NaCl at pH 4.3, pH 6.5 and pH 8.4

System	Parameters		Error	
	β [g/L]	K_S [M^{-1}]	β [g/L]	-K_S [M^{-1}]
lysozyme in NaCl, pH 4.3 [13]	386.07	8.23	± 14.275	± 0.36
lysozyme in NaCl, pH 4.3	420.23	2.72	± 54.72	± 0.29
lysozyme in NaCl, pH 6.5 [13]	370.44	10.88	± 25.394	± 0.80
lysozyme in NaCl, pH 6.5	705.26	5.66	± 53.36	± 0.24
lysozyme in NaCl, pH 8.4 [13]	173.98	16.61	± 23.137	± 2.75
lysozyme in NaCl, pH 8.4	363.17	4.59	± 20.08	± 0.23

The observed significant discrepancies between calculated solubility curves from the different experimental systems led to the assumption that the solubility line, or more generally spoken the phase transition measured with the robotic system does not correspond to the solubility line obtained with the experiments described in [13]. The solubility measured with the robotic approach is significantly higher and less dependent on pH and salt concentration. Protein and salt concentration increased much faster in the automated experiments than in [13] taking hours instead of days or weeks. In this case the protein solution undergoes a fast transition from the undersaturated to the supersaturated region of the phase diagram without reaching equilibrium. The formation of crystals or aggregates in this region is kinetically controlled. As a consequence of the short holding time no precipitate which can be detected at 595 nm is formed although the solution is supersaturated [3]. Another important difference between the solubility curves measured by Retailleau et al. and those determined with our experiments is the quality of the solid phase Retailleau analyzes the formation of crystals in the equilibrium experiments whereas in the fast robotic experiments the formation of amorphous aggregates is observed. Here again the kinetic control of solid phase transition plays an important role as crystallization is significantly slower than precipitation. In the light of these findings it would be more adequate to call the solubility curves from the automated measurements precipitation curves.

In cases where short time stability of protein solutions is sufficient, for example in protein purification processes, precipitation curves can offer the required information. This should be confirmed by a storage stability study of systems prepared closely below the measured precipitation curve. Eight solutions at pH 4.3, 6.5 and 8.4 were mixed manually from NaCl and protein stock solutions and stored with shaking for 23 hours at 18 °C. The turbidity at 595 nm was measured directly after sample preparation, after 6.5 hours and after 23 hours storage. The absorption of all samples at all measured time points was below 0.1 AU showing no visual turbidity. Consequently the determined precipitation curves give a good indication for the preparation of soluble protein solutions stable for at least 1 day.

3.2 Effect of Temperature on Lysozyme Solubility in Potassium Phosphate Buffer

For the investigation of temperature effects on lysozyme solubility and potential interdependencies with the solution pH lysozyme precipitation curves were determined for potassium phosphate buffer (KPi) at pH 6, pH 7 and pH 8 at 18 °C, 25 °C and 30 °C. In Figure 5 A to C precipitation curves are depicted separately for all three buffer pH values. As observed for the sodium chloride systems the obtained curves show an exponential decrease of lysozyme solubility with increasing buffer concentration. For all pH values a ranking from the highest to the lowest temperature is possible exhibiting increasing soluble protein content with increasing temperature at a defined salt concentration.
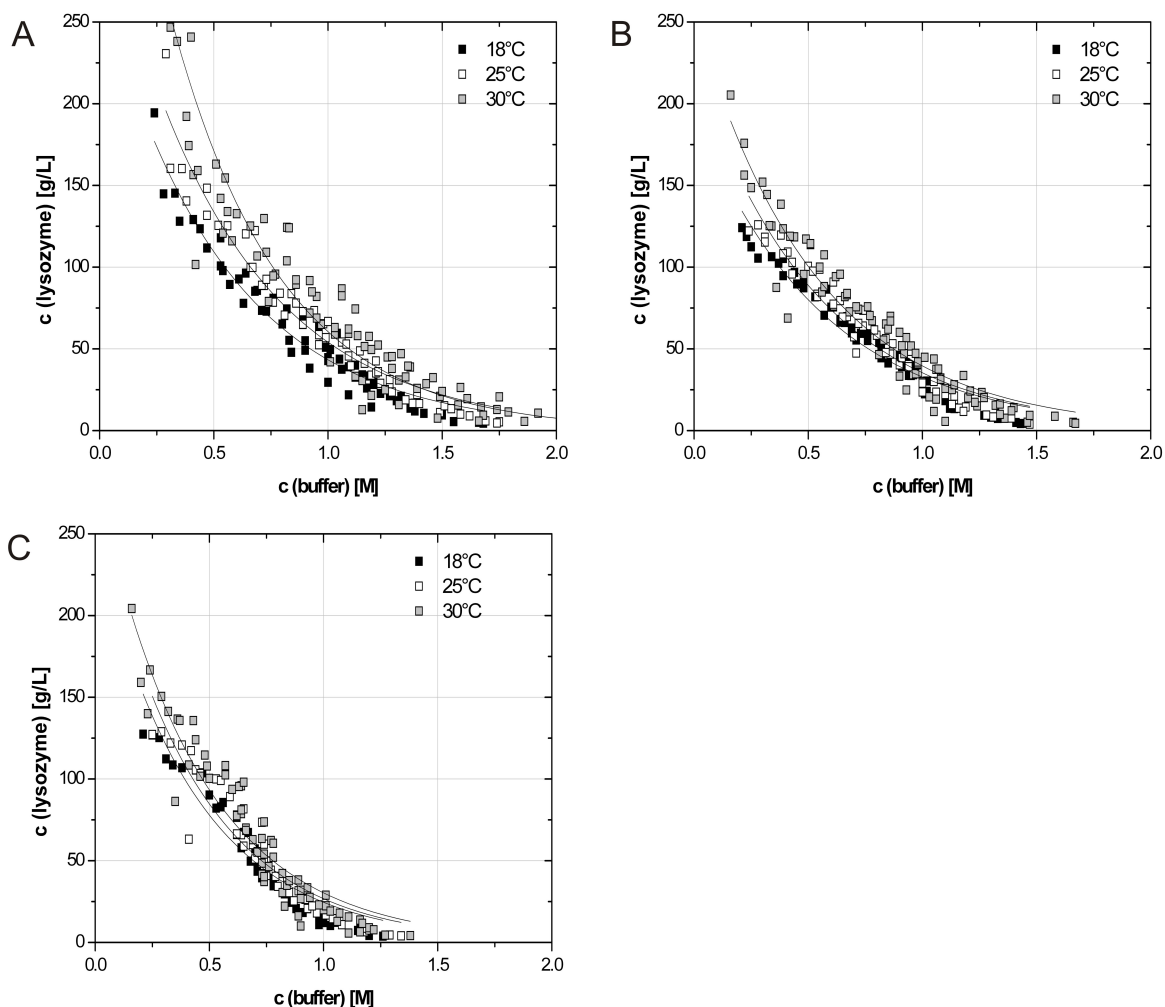
Figure 5: Temperature effect on lysozyme solubility in potassium phosphate buffer (KPi) at **A** pH 6, **B** at pH 7, **C** at pH 8

The fit parameters in Table 4 to Table 6 confirm this observation. A significant correlation of β with the incubation temperature indicates an improved maximum lysozyme solubility at higher temperatures. At pH 6 the fitted β value at 18 °C is only 57 % in proportion to β at 30 °C. At 25 °C β is 68 % high compared to the parameter value at 30 °C. At pH 7 a β of 77 % and 88 % is reached at 18 °C and 25 °C in comparison to the 30 °C parameter value. For pH 8 the temperature effect on β is smallest with 86 % at 18 °C and 93 % at 25 °C. These findings are consistent with effects described for lysozyme crystal growth in [14-16]. A negative enthalpy for the formation of the solid phase which is dependent on the buffer conditions is responsible for an increase of solubility with higher temperature, also called normal solubility [17].

In general the highest maximum solubility over the entire range is obtained at pH 6 whereas differences between values at pH 7 and pH 8 lie within the experimental error. The influence

of higher repulsive forces between lysozyme molecules at lower pH values due to an isoelectric point at pH 11 hints at an important role of charged residues in intermolecular interactions. Moreover the effect of temperature is higher at low salt concentrations. This interdependency between pH value, salt concentration and temperature is also mentioned in [15] for formation of orthorhombic and tetragonal lysozyme crystals with a temperature and solvent dependent change in crystal structure [16].

According to the empirical finding of Cohn the incubation temperature has no influence on the K_s values.

Table 4: Cohn curve fit parameters for the solubility curves of lysozyme in potassium phosphate buffer (KPi) at pH 6, pH 7 and pH 8 at 18 °C

System	Parameters		Error	
	β [g/L]	$-K_S$ [M^{-1}]	β [g/L]	$-K_S$ [M^{-1}]
lysozyme in KPi, pH 6	276.07	1.85	± 10.43	± 0.06
lysozyme in KPi, pH 7	195.36	1.78	± 7.403	± 0.068
lysozyme in KPi, pH 8	246.62	2.30	± 14.473	± 0.11

Table 5: Cohn curve fit parameters for the solubility curves of lysozyme in potassium phosphate buffer (KPi) at pH 6, pH 7 and pH 8 at 25 °C

System	Parameters		Error	
	β [g/L]	$-K_S$ [M^{-1}]	β [g/L]	$-K_S$ [M^{-1}]
lysozyme in KPi, pH 6	330.99	1.81	± 13.634	± 0.06
lysozyme in KPi, pH 7	223.28	1.84	± 10.02	± 0.08
lysozyme in KPi, pH 8	267.89	2.30	± 19.754	± 0.13

Table 6: Cohn curve fit parameters for the solubility curves of lysozyme in potassium phosphate buffer (KPi) at pH 6, pH 7 and pH 8 at 30 °C

System	Parameters		Error	
	β [g/L]	$-K_S$ [M^{-1}]	β [g/L]	$-K_S$ [M^{-1}]
lysozyme in KPi, pH 6	483.9	2.08	± 36.221	± 0.13
lysozyme in KPi, pH 7	255.23	1.86	± 9.818	± 0.07
lysozyme in KPi, pH 8	286.82	2.25	± 14.516	± 0.10

3.3 Effect of pH on Lysozyme Solubility

The influence of pH on lysozyme solubility was determined in potassium phosphate buffer systems. Precipitation was examined at pH 5 to 10 in steps of one unit at 25 °C. In Figure 6 the precipitation curves of lysozyme in potassium phosphate buffer at different pH values are depicted. Precipitation curves obtained from experiments performed at pH 7 to pH 9 lie close together and cross each other at low buffer concentrations. The precipitation curve determined at pH 5 shows by far the highest lysozyme solubility at all buffer concentrations. Lysozyme solubility decreases with higher pH values and the precipitation curves show a higher curvature. This is also mirrored by the Cohn fit parameters summarized in Table 7. The β values decrease from pH 5 to pH 7 representing an increase in curvature and then stay constant for all used higher pH values except for pH 9 which seems to be an outlier. An interesting point is a 100 % increase in K_s values for systems measured above pH 8 though this parameter should stay constant with changing pH. An explanation might be the absence of buffer capacity of the potassium phosphate system in this pH range. As a consequence lysozyme might have a stronger influence on the solvent properties as it also changes pH to a value of 3.9 being dissolved in Milli-Q water.
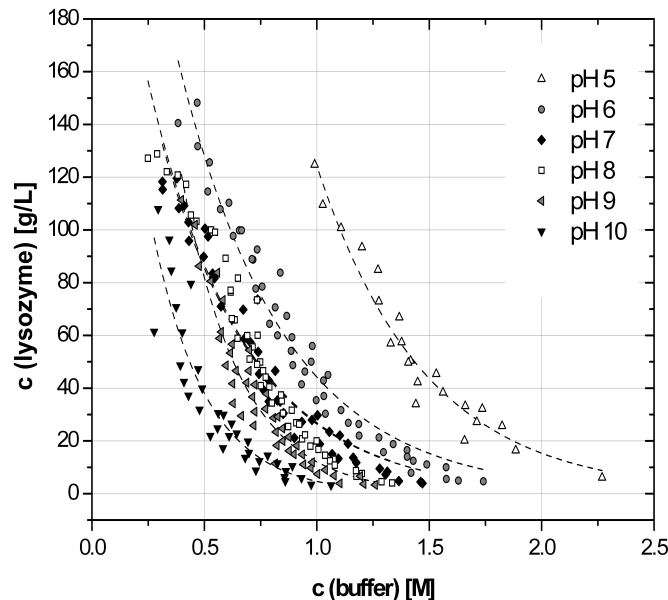

Figure 6: Dependence of lysozyme solubility in potassium phosphate buffer (KPi) on pH at 25 °C

Table 7: Cohn curve fit parameters for the solubility curves of lysozyme in potassium phosphate buffer (KPi) at pH 5 to pH 10 in steps of 1 at 25 °C

System	Parameters		Error	
	β [g/L]	$-K_S$ [M^{-1}]	β [g/L]	$-K_S$ [M^{-1}]
lysozyme in KPi, pH 5	1005.29	2.09	± 148.6	± 0.12
lysozyme in KPi, pH 6	370.58	2.13	± 20.09	± 0.08
lysozyme in KPi, pH 7	275.88	2.35	± 14.90	± 0.10
lysozyme in KPi, pH 8	280.52	2.35	± 18.16	± 0.12
lysozyme in KPi, pH 9	554.1	3.83	± 59.45	± 0.18
lysozyme in KPi, pH 10	309.08	4.19	± 53.66	± 0.43

As a conclusion, lysozyme maximum solubility is higher at pH values far from its isoelectric point (pH 11). Figure 7 gives a deeper insight into the correlation of net charge and solubility at a constant ionic strength of 1 M potassium phosphate. It can be observed that a saddle point at pH 7 and 8 is found for the net charge (calculated according to [19]) and for lysozyme solubility respectively. Our data are confirmed by published results for example for rhGCSF (recombinant human Granulocyte-Colony-Stimulating-Factor) and hemoglobin showing higher solubility at a pH far from the isoelectric point and at lower salt concentrations [20, 21]. These results show a positive correlation of net charge and solubility. Repulsive forces at a higher net charge seem to prevent protein interactions.
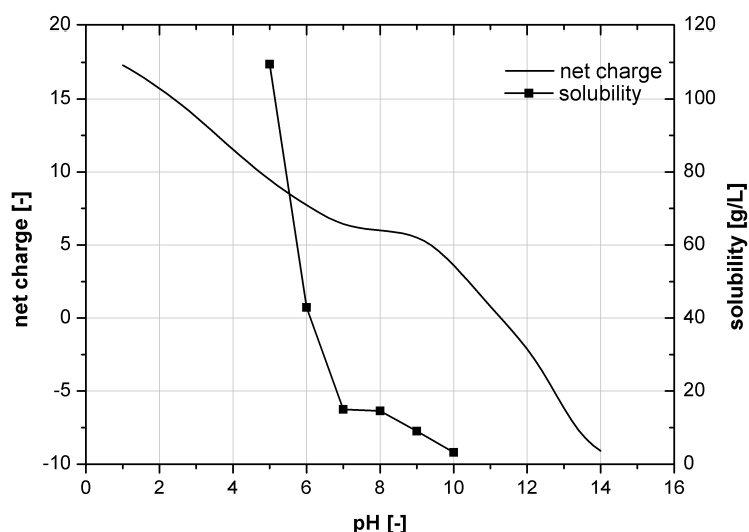

Figure 7: Correlation of lysozyme solubility and calculated net charge at different pH values in 1 M potassium phosphate buffer (KPi) at 25 °C

3.4 Effect of Additives on Lysozyme Solubility

3.4.1 Effect of Polyethylene Glycol (PEG)

Already in 1981 [22] a correlation between precipitating effect and molecular weight of PEGs was observed. The ability of PEG to promote protein aggregation is due to a steric exclusion of proteins from the liquid phase by PEG molecules and thus a local increase in protein concentration. On the other hand PEG is also used as additive in protein refolding processes to prevent the protein from precipitating [23]. Moreover the combination of salt and PEG does not necessarily lead to a simple addition of effects and is still not well understood [18].

Thus we investigated the concentration of a potassium phosphate buffer and the concentrations of PEGs of different molecular weight in respect to possible combinatorial effects. For the evaluation of PEG 300, PEG 3000 and PEG 8000 the PEG starting concentrations in the 96 well plates were set to values between 5 and 15 mM in steps of 2.5 mM. In Figure 8 three solubility surfaces of lysozyme in potassium phosphate buffer at pH 8 with addition of polyethylene glycols of different molecular weights are plotted.
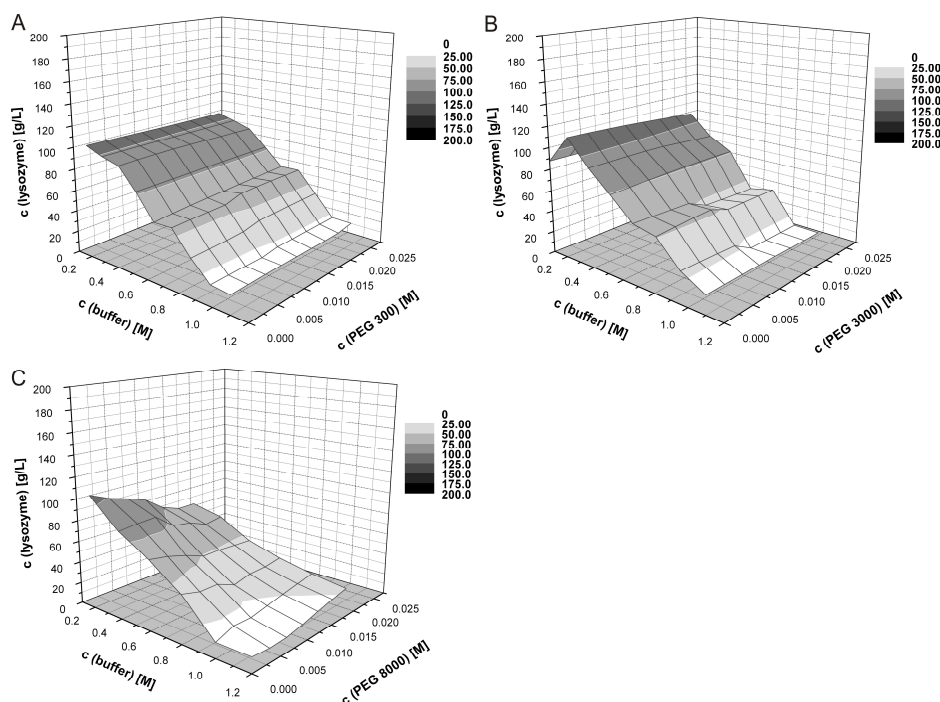
Figure 8: Solubility surface of lysozyme in potassium phosphate buffer at pH 8 and 25 °C with **A** PEG 300, **B** PEG 3000, **C** PEG 8000.

All surfaces show a decrease in lysozyme solubility with increasing ionic strength. This was already observed without additives. For PEG 300 and PEG 3000 the surfaces show a plateau at a buffer concentration of 0.6 M and PEG concentrations of 0.015 M for PEG 300 and 0.01 M for PEG 3000. On the other hand PEG 300 improves solubility at salt and PEG concentrations above these values and has no effect in the lower concentration range. PEG 3000 decreases lysozyme solubility especially in concentrations above 0.01 M at high salt contents. PEG 8000 acts as a strong precipitant for all investigated buffer concentrations. This precipitating effect is increased at higher PEG and salt concentrations as was already observed for PEG 3000.

The dependency of the precipitating effect on the PEGs` molecular weight and concentration was also observed in Atha et al. [22]. A decrease of human serum albumin solubility with PEG molecular weights above 400 described confirms our data for PEG 3000 and 8000. An improvement of protein solubility with addition of a lower molecular PEG, PEG 200, was observed during refolding of an interferon [24]. Though parameter effects on protein solubility in refolding processes might differ from effects on active protein solubility the

stabilizing effect of low molecular weight PEG is also observed in our experiments with the addition of PEG 300 at salt concentrations above 0.6 M.

3.4.2 Effect of Tween 20

As Tween 20 is one of the most frequently used non-ionic surfactants to prevent protein aggregation induced by shaking or on surfaces [8], we investigated the influence of Tween 20 on lysozyme solubility in combination with different concentrations of potassium phosphate buffer. Tween 20 was added in starting concentrations between 0.12 mM and 0.6 mM in steps of 0.12 mM.

Tween 20 increases solubility of lysozyme especially at high salt concentrations as can be seen in the solubility surface of lysozyme in a potassium phosphate buffer at pH 8 and Tween 20 in Figure 9.
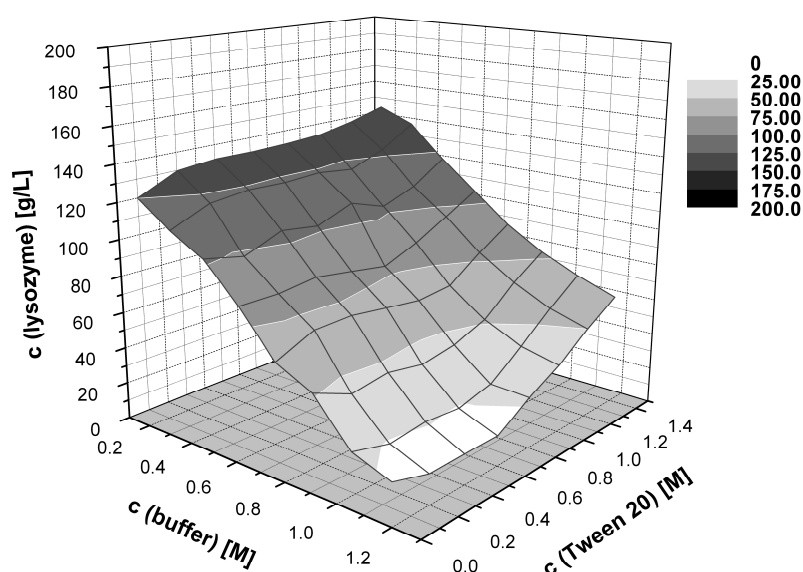

Figure 9: Lysozyme solubility in potassium phosphate pH 8, Tween 20, 25 °C

Tween 20 at concentrations above 0.8 M counteracts the decrease in lysozyme solubility induced at increasing potassium phosphate concentrations. The positive impact on lysozyme solubility becomes stronger with increasing salt concentrations. In respect to the interdependency between salt concentration and Tween 20 concentration, Tween 20 behaves similar to PEG 300 but with an even stronger positive influence on solubility over the whole salt concentration range. In literature [25] Tween 20 is also described to completely suppress

aggregation of 10 mg/ml human factor VII in 10 mM Tris buffer induced by shaking at Tween 20 concentrations above 120 µM. The proposed mechanism of the surfactant Tween 20 on protein aggregation is completely different from that described for PEG 300. A competition between Tween 20 and the protein for the air-water interface leads to improved solubility as stress-induced protein aggregation at the interphase is reduced [25]. To our knowledge no description of salt and Tween 20 interdependencies can be found in literature until today.

3.4.3 Effect of Sorbitol

As sorbitol is an additive frequently used to stabilize the native protein structure and to suppress aggregation in protein formulation [26] as well as in refolding processes [27] we were interested in its effect on lysozyme solubility in combination with varying salt concentrations.

Sorbitol was used in starting concentrations between 0.15 M and 0.75 M in steps of 0.15 M. Lysozyme solubility with addition of sorbitol can be estimated from Figure 10 where the solubility surface for lysozyme in potassium phosphate buffer at pH 8 and sorbitol is given.
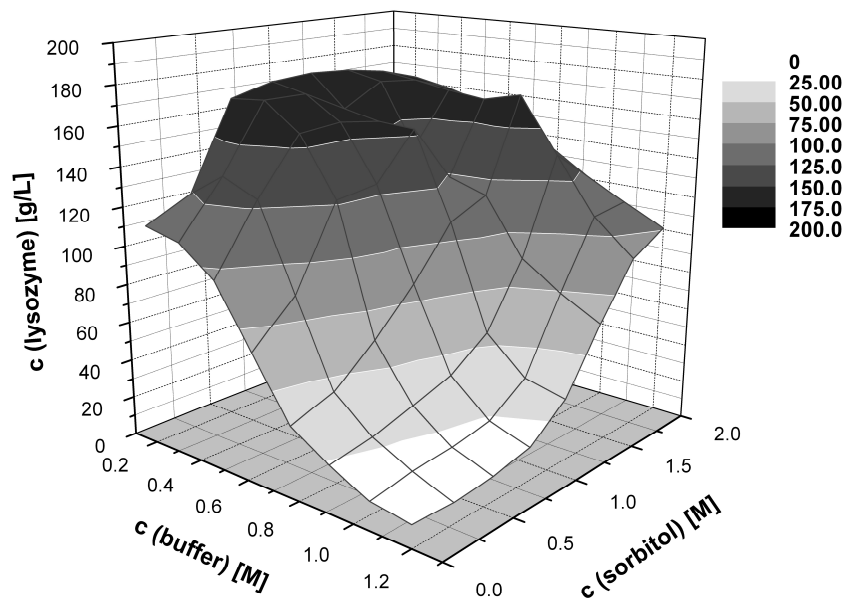

Figure 10: Solubility surface of lysozyme in potassium phosphate buffer at pH 8 and sorbitol at 25 °C

A plateau with maximum solubility is reached for low potassium phosphate concentrations below 0.7 M and sorbitol concentration above 0.3 M. At potassium phosphate concentrations above 0.7 M, lysozyme solubility is improved by sorbitol at concentrations above 1 M. A

further increase in sorbitol concentration does not result in a further improved lysozyme solubility. Between these two regions a slightly lower solubility is observed. Sorbitol is said to lead to a preferential hydration of protein molecules. Furthermore a specific binding of sorbitol to lysozyme is known to stabilize the protein structure [28]. In [29] these findings are discussed controversially pointing out that solubility is still a field leaving a lot of questions open. The observed solubility behaviour can only be explained by two distinct effects, one at lower buffer concentrations where sorbitol does not influence lysozyme solubility and one at high buffer concentrations where lysozyme solubility is dependent on sorbitol concentration.

3.4.4 Effect of Sucrose

Investigations on sucrose and salt concentration interdependencies on lysozyme solubility were carried out to gain a deeper inside into the impact of sucrose which is frequently used to prevent aggregation in formulation and refolding processes [8].

Sucrose was used in starting concentrations between 0.15 and 0.55 M in steps of 0.1 M. A second experiment using sucrose was performed for starting values between 0.02 M and 0.06 M with an increment of 0.01 in oder to obtain a better data resolution. For sucrose the most complex solubility diagram of the three examined additives resulted (Figure 11).
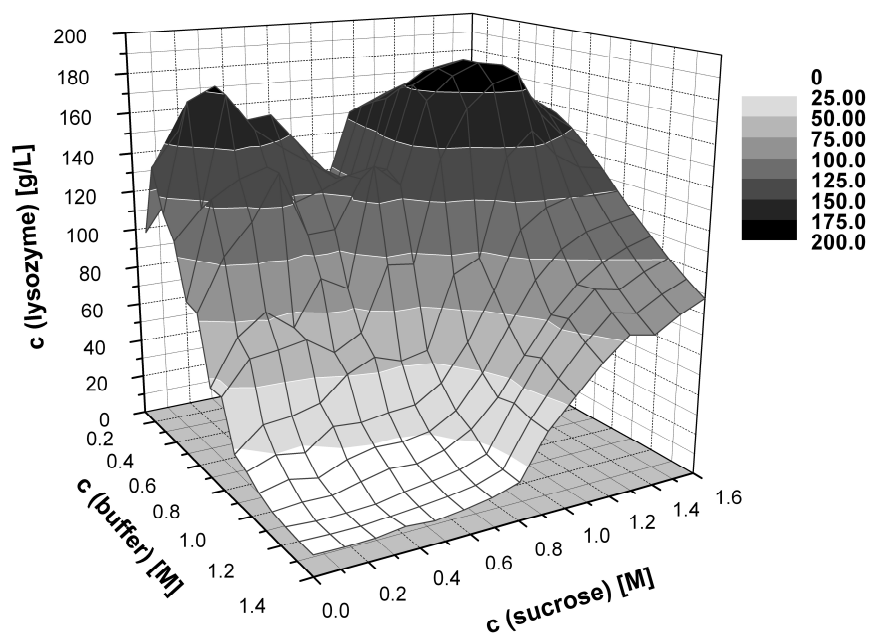

Figure 11: Solubility surface of lysozyme in potassium phosphate buffer at pH 8 and sucrose at 25 °C

Sucrose shows two peaks at buffer concentrations > 0.6 M with highly improved lysozyme solubility for sucrose concentrations below 0.5 M and above 1.0 M. The lower concentration optimum of sucrose shows a strong dependence on the salt concentration with maximum solubility at up to 0.5 M. The two peaks could be found for both experiments carried out with sucrose. Binding of sucrose to lysozyme is discussed in literature as preferential hydration caused by exclusion of sucrose from the protein surface [26, 29]. Buffer and additive concentration dependencies of lysozyme solubility are very similar for sucrose and sorbitol. These observations are also confirmed by similar proposed mechanisms for the additive effect on protein solubility. A major difference between both additives is a solubility minimum observed for sucrose which is not present in sorbitol data.

4 Conclusions

The combination of many potential interactions between different solvent and protein molecules makes an interpretation of solubility data a difficult task. Changing properties of all involved molecules including concentration dependencies further complicate modelling of solubility processes. All molecules in solution show interdependent properties which makes modelling of protein solubility a difficult task. Screening parameters cannot be investigated separately but have to be examined simultaneously, which increases the experimental effort. Our method opens the possibility to generate a high data density within a short time as a basis for a deeper understanding of salvation processes. Combinatorial effects between the concentration of buffer and additive can be investigated. Furthermore the influence of temperature, buffer concentration and pH on lysozyme solubility can be easily studied. We could show that pH, salt concentration and the concentration of different additives have interdependent effects on lysozyme solubility which to a high degree cannot be explained by simple precipitation mechanisms reported in literature. The results are in good agreement with published data from manual laboratory investigations showing validity of the automated high throughput method used. A high data density combined with thermodynamic and kinetic solubility principles and especially molecular modelling might be the only way to a better understanding of the protein precipitation process [30].

References

1. Cromwell, M.E.M., E. Hilario, and F. Jacobson, *Protein aggregation and bioprocessing.* Aaps Journal, 2006. **8**(3): p. E572-E579.

2. Wang, W., *Protein aggregation and its inhibition in biopharmaceutics.* International Journal of Pharmaceutics, 2005. **289**(1-2): p. 1-30.

3. Asherie, N., *Protein crystallization and phase diagrams.* Methods, 2004. **34**(3): p. 266-272.

4. Garcia-Ruiz, J.M., *Nucleation of protein crystals.* Journal of Structural Biology, 2003. **142**(1): p. 22-31.

5. Anderson, M.J., C.L. Hansen, and S.R. Quake, *Phase knowledge enables rational screens for protein crystallization.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(45): p. 16746-16751.

6. Zhu, D.W., et al., *Attempts to rationalize protein crystallization using relative crystallizability.* Journal of Structural Biology, 2006. **154**(3): p. 297-302.

7. Santesson, S., et al., *Screening of nucleation conditions using levitated drops for protein crystallization.* Analytical Chemistry, 2003. **75**(7): p. 1733-1740.

8. Wei, W., *Instability, stabilization, and formulation of liquid protein pharmaceuticals.* International journal of pharmaceutics, 1999. **185**(2): p. 129-188.

9. Hofmeister, F., *Zur Lehre von der Wirkung der Salze.* Arch. Exp.Pathol.Pharmakol., 1888. **24**: p. 1 - 16.

10. Cohn E.J. and E. J.T., *Proteins, amino Acids and peptides as ions and dipolar ions.* Reinhold Publishing, New York, 1943.

11. Bolen, D.W., *Effects of naturally occurring osmolytes on protein stability and solubility: issues important in protein crystallization.* Methods, 2004. **34**(3): p. 312-322.

12. Wiendahl M., et al., *A novel method to evaluate protein solubility using a high throughput screening approach.* Chemical Engineering Science, 2009.

13. Retailleau, P., M. RiesKautt, and A. Ducruix, *No salting-in of lysozyme chloride observed at how ionic strength over a large range of pH.* Biophysical Journal, 1997. **73**(4): p. 2156-2163.

14. Forsythe, E.L., R.A. Judge, and M.L. Pusey, *Tetragonal chicken egg white lysozyme solubility in sodium chloride solutions.* Journal of Chemical and Engineering Data, 1999. **44**(3): p. 637-640.

15. Ewing, F., E. Forsythe, and M. Pusey, *Orthorombic lysozyme solubility.* Acta Crystallographica Section D-Biological Crystallography, 1994. **50**: p. 424-428.

16. Cacioppo, E., S. Munson, and M.L. Pusey, *Protein solubilities determined by a rapid technique and modification of that technique to a micromethod.* Journal of Crystal Growth, 1991. **110**(1-2): p. 66-71.

17. Lu, J., X.J. Wang, and C.B. Ching, *Batch crystallization of soluble proteins: effect of precipitant, temperature and additive.* Progress in Crystal Growth and Characterization of Materials, 2002. **45**(3): p. 195-205.

18. Tardieu, A., et al., *Understanding salt or PEG induced attractive interactions to crystallize biological macromolecules.* Acta Crystallographica Section D-Biological Crystallography, 2002. **58**: p. 1549-1553.

19. Bashford, D., *Macroscopic electrostatic models for protonation states in proteins.* Front Biosci, 2004. **9**: p. 1082-99.

20. Chi, E.Y., et al., *Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation.* Pharmaceutical Research, 2003. **20**(9): p. 1325-1336.

21. Green, A.A., *Studies in the physical chemistry of the protein VIII. The solubility of hemoglobin in concentrated salt solutions. A study of salting out of proteins.* Journal of Biological Chemistry, 1931. **93**(2): p. 495-516.

22.   Atha, D.H. and K.C. Ingham, *Mechanism of precipitation of proteins by polyethylen glycols - analysis in terms of excluded volume.* Journal of Biological Chemistry, 1981. **256**(23): p. 2108-2117.

23.   Middelberg, A.P.J., *Preparative protein refolding.* Trends in Biotechnology, 2002. **20**(10): p. 437-443.

24.   Wang, F.W., et al., *On-colunm refolding of consensus interferon at high concentration with guanidine-hydrochloride and polyethylene glycol gradients.* Journal of Chromatography A, 2006. **1115**(1-2): p. 72-80.

25.   Kreilgaard, L., et al., *Effect of Tween 20 on freeze-thawing- and agitation-induced aggregation of recombinant, human factor XIII.* Journal of Pharmaceutical Sciences, 1998. **87**(12): p. 1597-1603.

26.   Timasheff, S.N., *The control of protein stability and association by weak-interactions with water-how do solvents affect these processes* Annual Review of Biophysics and Biomolecular Structure, 1993. **22**: p. 67-97.

27.   Yu, Z.H. and B. Li, *The effect of polyols on the reactivation of guanidium chloride-denatured arginine kinase from shrimp Feneropenaeus chinensis muscle.* Protein and Peptide Letters, 2003. **10**(2): p. 199-211.

28.   Wimmer, R., et al., *Towards a molecular level understanding of protein stabilization: The interaction between lysozyme and sorbitol.* Journal of Biotechnology, 1997. **55**(2): p. 85-100.

29.   Datta, S., B.K. Biswal, and M. Vijayan, *The effect of stabilizing additives on the structure and hydration of proteins: a study involving tetragonal lysozyme.* Acta Crystallographica Section D-Biological Crystallography, 2001. **57**: p. 1614-1620.

30.   Ulrich, J. and M.J. Jones, *Industrial crystallization - Developments in research and technology.* Chemical Engineering Research & Design, 2004. **82**(A12): p. 1567-1570.

Development and Characterization of an Automated High Throughput Screening Method for Optimization of Protein Refolding Processes

Annette Berg, Joerg Kittelmann, Juergen Hubbuch*

Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Science, University of Karlsruhe (TH), 76131 Karlsruhe, Germany

*Corresponding author. Tel.: +049 721 608-2557; fax: +049 721 608-6240. Email-adress: juergen.hubbuch@kit.edu

Abstract

Optimization of protein refolding parameters by automated, miniaturized and parallelized high throughput screening (HTS) is a powerful approach to meet the demand for fast process development with low material consumption. In this study we validated methods applicable on a standard liquid handling robot for screening of refolding process parameters by dilution of denatured lysozyme in refolding buffer systems. One of the major challenges was the implementation of fast and automatable analytics. Different approaches for the estimation of protein solubility and folding were validated concerning resolution and compatibility with the robotic system and with the complex buffer and protein structure composition. We established an indirect method to assess soluble lysozyme concentration independent of matrix effects and protein structure varieties by automated separation of aggregated protein, resolubilization and measurement of absorption at 280 nm. Soluble lysozyme content was calculated on the basis of mass balances. Furthermore resolution of tryptophan fluorescence for estimation of protein structure was validated with a lysozyme activity assay implemented in the automated process. The resolution in refolding samples was too low to serve as function for optimization. As solubility can be measured with non-specific assays the correlation between favourable parameters for high active and soluble lysozyme yields were evaluated. An overlap of good refolding buffer compositions was found provided that the redox environment is controlled with redox reagents. In addition the need to control unfolding conditions like time, temperature, lysozyme and DTT concentration is pointed out as different feed stocks resulted in different refolding yields.

1 Introduction

E. coli is still one of the most frequently used expression hosts for the production of biopharmaceutical products. High expression titers in fully synthetic media and established systems for genetic modification do not only decrease time for development of expression strains and fermentation strategies but also drive down production costs. Unfortunately high expression levels and reducing conditions in the bacterial cytosol often lead to formation of aggregated recombinant proteins, so called inclusion bodies, which have to be solubilized with denaturing reagents and subsequently refolded to gain active protein. An overview of basic principles in inclusion body protein refolding is given in numerous reviews [1-5].

As development and optimization of refolding processes is mainly based on experience and trial and error, a high number of experiments have to be conducted to find favorable conditions for each protein. Furthermore protein renaturation can be improved with control of various process parameters like buffer salt and pH, ionic strength, reducing and oxidizing substances for disulfide bond formation and several additives promoting folding and solubility. There are some approaches published to solve this complex problem of refolding process development on the basis of high throughput screening technologies with refolding of denatured protein by dilution in different buffer systems [6-12]. The major hurdle when introducing HTS (high throughput screening) technologies lies in the use of adequate analytical methods. This has, however, so far not been addressed. Sample analysis performed manually or the usage of time consuming serial sample handling spoils positive effects on time for development decreased by parallelized and frequently automated sample generation in most published studies. For instance SDS-PAGE analysis, sample preparation by dialysis, CD spectroscopy, RP-HPLC or analytical SEC and DLS are time-consuming and need human intervention for handling of sophisticated instrumentation or complex operations. Enzymatic and specific binding assays are often the methods of choice, but have to be set up individually for every protein [13-17]. Commercially available refolding Kits (Pro-Matrix Protein Refolding Kit, Pierce; Quickfold™ Protein Refolding Kit, AthenaES™; Protein Refolding Kit, BioAssay™; FoldIt Screen, Hampton Research) provide potential refolding buffer systems whereas the development of analytics remains to the customer. Especially refolding buffers containing a diverse set of additives are incompatible with commonly used protein assays. Moreover structural diversity of the samples often leads to significant varieties in

signal height. All these points massively complicate development of automated high throughput screening methods for protein refolding.

In this study we developed a protein refolding screening system that fulfils the requirements of parallelization and automation for sample generation and subsequent analytics. A method for measurement of soluble protein concentration is presented that lacks the mentioned disadvantages of manual and time-consuming buffer exchange, incompatibility with disturbing reagents or influence of structural diversities.

As protein solubility represents one of the critical objective functions in renaturation process optimization and an important parameter in the course of process development this paper elucidates different methods of measuring this parameter in the light of requirements of HTS based applications and investigates process parameter correlations of solubility and folding. Moreover tryptophan fluorescence is validated as a marker for structural integrity. Lysozyme is taken as an ideal model protein with 4 disulfide bonds and 1 of 6 tryptophan residues fully buried in the interior of the molecule. An enzymatic assay is adapted to the robotic platform for fast and automated analysis of activity. Finally our results provide a deeper insight into the necessity of parameter control during refolding process development concerning quality of starting protein material and redox potential.

To guarantee for high sample variability during validation of the presented HTS method, buffer composition follows a random distribution of process parameters within a wide parameter space. A fractional factorial design was not used in this study because in this case a restriction of the parameter space or of the number of parameters would have been obligatory to reduce experimental effort. Furthermore the presented model system should show the validity of measured functions and their relevance for refolding buffer design rather than the final optimization of the refolding buffer system.

2 Materials and Methods

2.1 Materials

2.1.1 Robotic Workstation

For this study, the automated pipetting station Freedom Evo 200® (Tecan Crailsheim, Germany) equipped with one liquid handling arm and two grippers was used. Liquid handling on this platform was performed with 8 fixed standard tips. A centrifuge Rotanta RSC46 (Hettich Kirchenlengern, Germany), a magnetic orbital shaker with four positions and a shaking diameter of 2 mm (Inheco Munich, Germany) and a spectrophotometer InfiniTe 200 (Tecan Crailsheim, Germany) were integrated into the robotic platform.

2.1.2 Technical Laboratory Equipment

The used water bath was the MC-4 heating circulator from Julabo (Seelbach, Germany) and the thermomixer compact from Eppendorf (Hamburg, Germany). Size exclusion chromatography validation was performed on an AKTA HPLC system with Tricorn 5/50 columns purchased from GE Healthcare (Uppsala, Sweden).

2.1.3 Disposables

96 well microtiter plates and UV microtiter plates with a well volume of 360 µl were purchased from Greiner (Frickenhausen, Germany). Deep well plates with a well volume of 2.2 ml were obtained from TreffLab (Degersheim, Switzerland). 350 µl AcroPrep 96 filter plates with a pore size of 0.2 µm and a Bio-Inert membrane, AcroPrep 96 well ultrafiltration plates with a 3 kDa cut off and an Omega membrane were purchased from Pall (Dreieich, Germany). Greiner tubes were purchased from Greiner bio-one GmbH (Frickenhausen, Germany) and Eppendorf cups from Eppendorf (Hamburg, Germany). 96-Well DispoDIALYZER plates with a molecular weight cut off of 2000 Da from Havard Apparatus (Kent, Great Britain) were used for high throughput dialysis.

2.1.4 Chemicals and Proteins

All chemicals (analytical grade), Bradford reagent and hen egg white lysozyme (L-6876) with $\geq$ 90 % protein content were purchased from Sigma Aldrich (St. Louis, USA). Sephadex G25 superfine gelfiltration resin was purchased from GR Healthcare (Uppsala, Sweden).

2.2 Methods

2.2.1 Preparation of Stock Solutions

Buffer Solution, Salt and Redox Components

All buffer stock solutions had a concentration of 0.5 M. The pH of the respective buffer solutions was adjusted by titration of the acidic and basic components. Glycine buffer was used for pH 3, sodium acetate buffer for pH 4 and pH 5, potassium phosphate buffer for pH 6, pH 7 and pH 8 and bicine buffer for pH 9.

Salt containing stock solutions such as $MgSO_4$ and NaCl were prepared with a concentration of 1.5 M.

Denaturation buffer used to unfold lysozyme contained 8 M urea and 50 mM potassium phosphate buffer at pH 8. The unfolding or denaturation buffer was prepared without DTT and stored for at most two weeks. Immediately before use DTT was added from a frozen 1 M stock solution in a concentration of 5 mM.

DTT stock solution for refolding was solubilized in a concentration of 50 mM in water. GSSG was prepared in a concentration of 35 mM in water. The pH of the GSSG solution had to be adjusted to pH 7 by addition of 1 M NaOH. Both redox component solutions were prepared directly before start of the experiment.

Native and Denatured Lysozyme

Solutions of native lysozyme were prepared directly before use by dissolving lysozyme in Milli-Q water in the following concentrations: 1 mg/ml, 0.5 mg/ml, 0.25 mg/ml, 0.125 mg/ml, 0.063 mg/ml, 0.031 mg/ml and 0.016 mg/ml. Mixing was performed by short vortexing of the solution.

Lysozyme in a concentration of 2 mg/ml, 4 mg/ml or 8 mg/ml was dissolved in denaturation buffer by either gently stirring in Greiner tubes on a magnetic stirrer or by vortexing in small scale validation experiments in Eppendorf cups. In large scale experiments with several milliliters the solution is incubated at the specified temperature either in a water bath (37 °C), at room temperature or in the fridge (4 °C). For small scale validation experiments of one to two milliliters the Eppendorf cups were incubated in a thermomixer with a set temperature of 4 °C, 25 °C or 37 °C without shaking. Validation samples were measured at different time points from the start up to 46 h. In larger scale denaturation at room temperature was performed for 3 h and denaturation at 37 °C for 1.5 h.

Micrococcus lysodeikticus Suspension

Micrococcus lysodeikticus was suspended in 100 mM potassium phosphate buffer at pH 6 in a concentration of 0.65 mg/ml. Mixing was performed by intense vortexing. The stability of the suspension was investigated over time by measuring 12 calibration curves of the same native lysozyme preparation (1 mg/ml, 0.5 mg/ml, 0.25 mg/ml, 0.125 mg/ml, 0.063 mg/ml, 0.031 mg/ml and 0.016 mg/ml) at different time points from 0 min to 125 min. When assaying refolded lysozyme the suspension was prepared 30 minutes before use.

2.2.2 Automated Sample Preparation for High Throughput Refolding Screening

Volumes of stock solutions for generation of refolding buffers were imported into the robotic control software as variables from Excel files. Pipetting with the robotic system was calibrated for the different solutions used by adapting liquid handling parameters like aspirating and dispensing speed and location of the tips in relation to the liquid level or the labware bottom.

Automated Preparation of Refolding Systems

The automated generation of refolding systems is depicted in Figure 1. Prior to the procedure developed for screening of refolding conditions, unfolded protein was prepared as described above. The actual screening procedure is than initiated by mixing of a set of refolding systems. The refold systems contained several additives summing up to a volume of 900 µl and were prepared by using stock solutions of 0.5 M buffer, 1.5 M NaCl, 1.5 M MgSO$_4$, 35 mM glutathione (GSSG) at pH 7 as oxidizing and 50 mM dithiothreitol (DTT) as reducing agent and Milli-Q water as fill to reach the final volume. Each system was initially mixed by applying one aspiration-dispensing step with a volume of 900 µl. After the preparation of all systems continuous mixing was ensured by orbital shaking of the deep well plate at 1500 rpm. Without interruption of shaking, 100 µl of denatured protein were added to the refold systems resulting in a protein concentration of 0.4 mg/ml. This procedure led to a fast reduction of denaturant concentration with a dilution factor of 10 of denatured, reduced lysozyme in refolding buffer. Concomitant shaking on the orbital shaker and one aspirating-dispensing step of 900 µl were performed for fast homogeneous distribution of the protein. Effects of local protein concentration maxima on protein aggregation should thus be minimized. Incubation for refolding was performed at room temperature with constant shaking at 1500 rpm for 1 h, to gain a stable protein solution with aggregated and soluble protein in equilibrium.
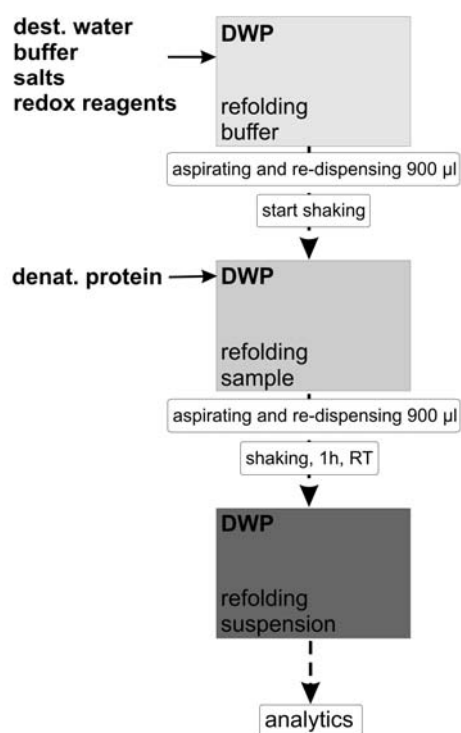
Figure 1: Flow scheme of the automated refolding process

Evaluation of Parameter Relevance and Experimental Error

When evaluating the relevance of the various additives and the experimental error accompanying the screening approach 40 random experiments were performed prior to the actual screening procedures. Buffer concentration, protein concentration and dilution factor of denatured protein into refolding buffer were kept constant. The pH is varied in steps of one between pH 3 and pH 9 whereas the salts (NaCl, $MgSO_4$) and the redox components were handled as indiscrete parameters varying between 0 and 150 mM for salts, between 0 and 20 mM for the oxidizing component GSSG and between 0 and 10 mM for the reducing component DTT.

2.2.3 Sample Preparation Prior to Analysis

After a 1 h incubation period, the refolding suspension is directly transferred into the analytical pathway vizualized in Figure 2. The flow scheme also shows the samples measured during method validation.

In the standard procedure 300 µl of the refold suspension were transferred to a 0.2 µm Bio-Inert membrane plate. This aliquot of the refold suspension was further processed starting with a solid-liquid separation step. The filter plates set on top of a UV microtiter plate were centrifuged at 4754 RCF for 10 min at room temperature. This procedure led to a separation of liquid containing soluble protein and aggregates. For validation purposes the remaining suspension in the deep well plate was centrifuged at 1500 rpm for 10 min at room temperature and 300 µl of the aggregate free supernatant were then transferred to a UV plate and protein content measured using UV 280 nm.

The filtrate was than used for structural analysis using tryptophane fluorescence, biological activity using the described enzymatic assay and protein content using UV 280 nm.

The aggregate containing retentate in the filter plate was washed once with 300 µl of Milli-Q water to remove residual buffer components using the above described centrifugal procedure. The filtrate resulting from the washing step was analyzed for protein content using UV 280 nm. Remaining aggregates in the filtration plate are then resolubilized using 300 µl of denaturation buffer. DTT is added to reduce intermolecular disulfide bonds stabilizing aggregates and thus preventing solubilization. To ensure quantitative solubilization of aggregates - including potentially adhering aggregates to the filter membrane - the solution was finally passed through the filter using centrifugation at 1500 rpm for 10 min at room temperature. The filtrated solution is than assayed for protein content using UV 280 nm.
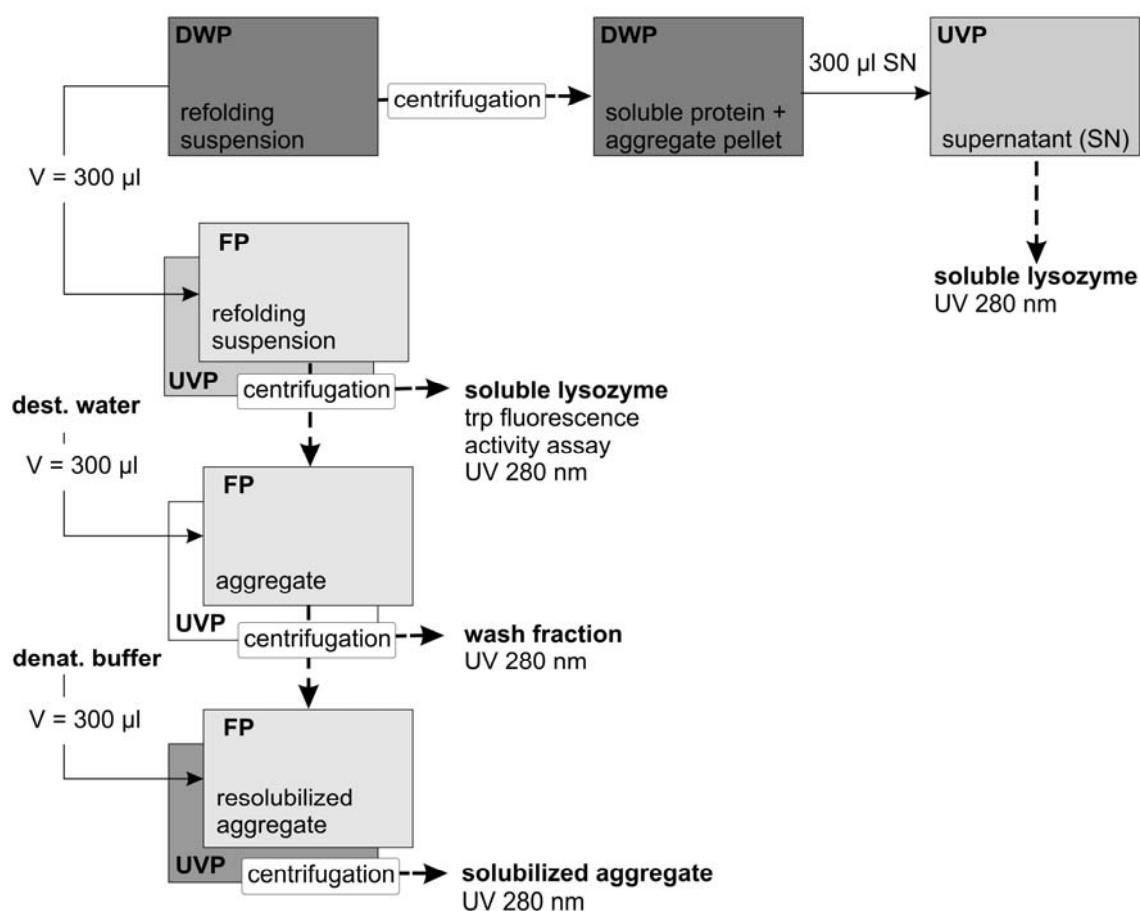
Figure 2: Flow scheme of the automated analytics including steps for method validation

2.2.4 Analytical Assays

Protein Content

a) Direct Analysis of Soluble Protein Using UV 280 nm

The absorption at 280 nm of 300 µl protein solution was measured in microtiter plates with a UV transparent bottom. Measurement was conducted three times in each well of the microtiter plate in the middle of the well and an average value was calculated. For validation purposes native lysozyme was solubilized in 50 mM potassium phosphate buffer at pH 8 by vortexing. Different protein concentrations up to 0.5 mg/ml and a buffer blank were processed in parallel. The calibration is fitted with a linear regression curve. To measure denatured and

reduced lysozyme during validation experiments lysozyme concentrations between 0 and 0.5 mg/ml are incubated for 3 h at 25 °C in denaturing buffer.

b) Direct Analysis of Soluble Protein Using a Bradford Assay

The Bradford assay used was performed according to [18]. In short, 200 µl of Bradford reagent were pipetted into each well of a standard microtiter plate. After addition of 40 µl of the protein containing sample mixing was performed by two repeated aspiration-dispensing steps with a volume of 200 µl. Following this the absorption was measured at 595 nm.

c) Determination of Aggregates

The determination of protein aggregates in refolding samples is performed by separation of aggregates from 300 µl refolding sample in 0.2 µm Bioinert filter plates. Aggregates were separated by centrifugation for 10 minutes at 1500 rpm, washed with Milli-Q water and resolubilized in 300 µl denaturing buffer.. The absorption at 280 nm of the solubilized aggregates is measured in UV microtiter plates. A buffer blank is processed in parallel. A native lysozyme calibration was used to calculate the protein concentration of resolubilized aggregates.

d) Turbidity

Turbidity was measured to validate the quantification of aggregated protein by re-solubilisation and UV 280 nm. Refolding samples with 1 mL total volume were re-suspended by one aspirating and re-dispensing step of 900 µl before a 300 µl sample volume is transferred to a microtiter plate for a threefold measurement of absorption at 410 nm in the middle of each well. As blank an average absorption value of 8 different native, and completely soluble lysozyme concentrations between 0 and 1 mg/ml in water is used.

e) Analysis of Lysozyme Activity

Lysozyme activity was measured by a decrease in absorption of a *Micrococcus lysodeikticus* suspension at 595 nm when subjected to lysozyme enzymatic activity. 200 µl of a 0.65 mg/ml *Micrococcus* suspension are distributed in each well of a 96 well microtiter plate. 30 µl

aliquots from the sample preparation scheme outlined above and a calibration standard containing 8 different concentrations of native lysozyme from 0 to 1 mg/ml were diluted in 20 mM potassium phosphate buffer at pH 6 with a dilution factor of 30 to bring samples into the linear range of the assay. The samples were mixed by orbital shaking at 1500 rpm and by one aspirating and re-dispensing step of 900 µl. 40 µl of each diluted sample were transferred into the bacterial suspension. The resulting assay solution was mixed by two repeated aspiration-dispensing step with a volume of 170 µl. Prior to any pipetting step the liquid handling tips were washed with 800 µl of 1 M NaOH and subsequently 20 mL Milli-Q water to avoid false positive results due to liquid carry-over. Directly after the sample solution was mixed into the Micrococcus suspension the absorption was measured five times in 31 s intervals to be able to measure the starting velocity of the reaction and a linear decrease of turbidity. Because liquid handling was performed with an eight tip liquid handling arm only eight samples could be processed in parallel which leads to a longer holding time for samples measured at later time points.

f) Measurement of Tryptophan Fluorescence

After removal of aggregated protein 300 µl of the soluble lysozyme filtrate in the UV microtiter plate were used for tryptophan fluorescence measurements. Tryptophan emission spectra were measured by excitation at 280 nm between 300 and 400 nm in intervals of 2 nm. An asymmetric Gauss fit of the obtained spectra was performed. The value assayed for was the emission wavelength at peak maximum λ (I_{max}).

2.2.5 Buffer Exchange on a HTS Platform

Ultrafiltration

Lysozyme and BSA were used at concentrations between 60 and 200 µg/ml. The initial buffer for lysozyme was 50 mM sodium acetate buffer at pH 4 and 50 mM bicine buffer at pH 9 and for BSA 50 mM sodium acetate buffer at pH 4 and 50 mM potassium phosphate buffer at pH 8. A change to 50 mM potassium phosphate buffer at pH 8 is performed in three steps consisting of sample volume reduction to approximately 50 µl by centrifugation of the ultrafiltration plate for 45 min at 4754 RCF and addition of 300 µl potassium phosphate buffer. A liquid level detection with the robotic tips was performed to assess the sample

volume inside the ultrafiltration plate after addition of 200 µl potassium phosphate elution buffer. The retentate was then homogenously mixed by aspirating and re-dispensing of 200 µl for three times inside the ultrafiltration plate prior to transfer in a UV microtiter plate. The protein concentration was measured by absorption at 280 nm.

Dialysis

The removal of 20 mM GSSG from BSA and lysozyme solutions should be analyzed and the protein recovery evaluated. BSA and lysozyme were used in 8 different concentrations between 0 and 1 mg/ml. Both proteins were dissolved in 50 mM potassium phosphate buffer at pH 8 with and without 20 mM GSSG. Furthermore a BSA solution in 50 mM sodium acetate buffer at pH 4 was prepared. Buffer exchange was performed against 50 mM potassium phosphate buffer at pH 8. A volume of 300 µl of each sample was distributed into the wells of a DispoDIALYZER plate with a cut off of 2 kDa pre-equillibrated with 200 µl dialysis buffer. The dialysis plate was directly set on the liquid surface of the dialysis reservoir filled with 3 L of potassium phosphate buffer. Incubation was performed at ambient temperature with stirring of the dialysis buffer using a magnetic stirrer. After 5 h of incubation the dialysis buffer was changed and the dialysis proceeded overnight. A liquid level detection is performed with the sensor system of the robotic tips to calculate the sample volume after dialysis. Afterwards 200 µl sample was transferred to a UV microtiter plate and the absorption at 280 nm is measured. Unprocessed control samples without GSSG were used as 100 % recovery standard.

Size Exclusion Chromatography (SEC)

The separation of 0.1 mg/mL lysozyme (14000 Da) from 20 mM GSSG (612 Da) was investigated on different gelfiltrations resins (Superose 12, Sephadex G25, Sephadex G50, Sephacryl S100) starting with a column length of 20 cm and a sample load of 8 % CV. As only Sephadex G25 showed potential baseline separation with optimization of column load, this approach was used to perform experiments with 5 cm column length applicable on the robotic platform. A sample load of 2 % CV was chosen. Lysozyme and GSSG were also loaded separately onto the column to see possible interactions between protein and oxidizing agent.

3 Results and Discussion

3.1 Analytical Procedures

A necessary prerequisite for HTS compatible analytical procedure in the field of inclusion body refold screening lies in a maximum flexibility of the operator when considering buffer design including redox components for disulfide bond formation, additives such as detergents and amino acids as well as added salt components. Moreover diversities in protein structure developing during protein folding lead to an additional increase in sample complexity and directly influence all assays exploiting interactions with specific amino acids. Low concentrations of soluble protein typically occurring during refold screens applying a dilution based method complicate the requirements by the need for high sensitivity. In general when evaluating the effect of different process parameters quantitative data are preferred against qualitative data as a basis for rational experimental design and optimization. In addition all these challenges posed by complex buffer systems, low protein concentration and the need for quantification of objective functions should be met in a fully automated, miniaturized and parallelized screening set up. Up to our knowledge these criteria are not met by any published or commercially available high throughput screening approach for the optimization of protein refolding.

3.1.1 Quantification of Soluble Protein Content

UV 280 Based Measurements

Protein concentration measurements based on the absorption at UV 280 rely on the absorption of radiation in the near UV mainly by the amino acids tyrosine and tryptophan and to a small extend phenylalanine and disulfide bonds. In theory higher protein structures also may absorb UV light or modify the molar absorptivities of tyrosine and tryptophan. Finally any additive present in the sample which is a chromophore will interfere with the measurements.

Figure 3 compares UV 280 measurements for native and denatured BSA and lysozyme respectively. The good congruence of the respective calibration curves indicates that there is no significant sensitivity against changes in protein structure. Though the extinction

coefficients being represented by the slope of the calibration curves are six times lower for BSA compared to lysozyme due to different contents of absorbing amino acids like tryptophan and tyrosine no differences can be observed for completely denatured and active protein.
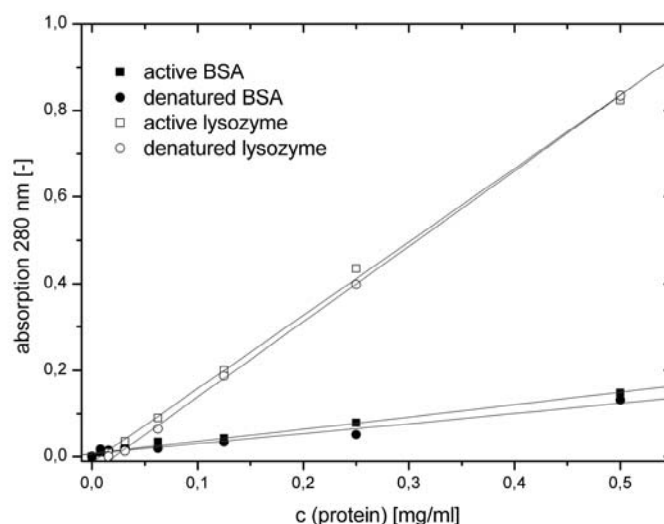

Figure 3: Calibration curves measured at 280 nm for BSA and lysozyme. Active lysozyme and BSA are solubilized in 50 mM potassium phosphate buffer at pH 8 and denatured lysozyme and BSA are solubilized for 3 hours in 8 M urea, 50 mM potassium phosphate buffer and 5 mM DTT at 25°C

On the other hand oxidizing components in refolding buffer systems show high absorption at 280 nm being an obstacle to UV 280 nm based protein determination. Furthermore changes in the UV absorption of redox components are observed due to pH dependent redox reactions between redox couples and also air. This has to be addressed additionally for every type of sample. The absorption at 280 nm was measured during incubation of 8 M urea buffer at pH 8 with 5 and 10 mM DTT in an uncovered microtiter plate with shaking. In Figure 4 an exponential increase in absorption at 280 nm over time is shown with an equal initial slopes of 0.002 min^{-1} for both DTT concentrations and a saturation absorption value being proportional with DTT concentration.
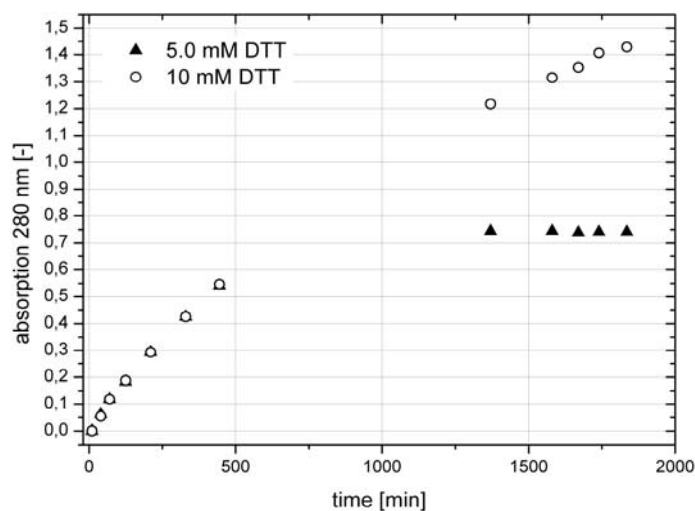
Figure 4: Measurement of absorption at 280 nm of a 5 mM and a 10 mM DTT solution in 8 M urea and 50 mM potassium phosphate buffer at pH 8 over time. The solution was incubated with shaking in a non-covered microtiter plate

This increase in absorption can be explained by fast oxidation of DTT with air at pH 8. As the initial slope of the absorption over time is equal for both DTT concentrations the mass transfer of oxygen seems to be the limiting factor. Direct protein quantification by absorption measurement at 280 nm is therefore difficult as the buffer absorption changes overtime and is dependent on all factors influencing the redox environment.

Dye Based Methods

The most commonly used methods for direct quantification of soluble protein content are colorimetric assays such as Bradford or BCA (bicinchoninic acid). However due to a number of reasons pinpointed below both methods can not be applied during a refold screening.
The Bradford assay is based on the non-covalent bond between the hydrophobic patches of an unfolded protein and the non-polar region of the dye via van-der-Waals forces, positioning the positive amine groups in proximity with the negative charge of the dye. The electrostatic interaction arising from this coordination strengthens this bond. This protein dye complex leads to a stabilization of the blue form of Coomassie dye, and thus the amount of complex formation is a direct measure of protein concentration.

The BCA assay is based on the reduction of Cu2+ to Cu1+ by protein peptide bonds followed by a chelating complex of two molecules of bicinchoninic acid and the reduced copper. This complex expresses a purple color and is measured at a wavelength of 562 nm.

It becomes immediately clear that reducing components being indispensable in disulfide bond reshuffling and also being carried over into the refold with the denatured protein strongly interfere with the BCA assay and lead to a high background signal [19]. Next to a number of additives not compatible with both, the BCA and the Bradford assay, Triton or Tween, which are used in refolding to enhance protein solubility, cannot be used in combination with these assays. Arginine, also one of the most frequently used additives for stabilization of protein during folding leads to high signals with Bradford reagent [19].

Besides these substance incompatibilities a major problem we observed in our studies is the dependence on protein structure when fully denatured and active lysozyme and BSA were measured in a calibration experiment with the Bradford assay. The result can be seen in Figure 5.
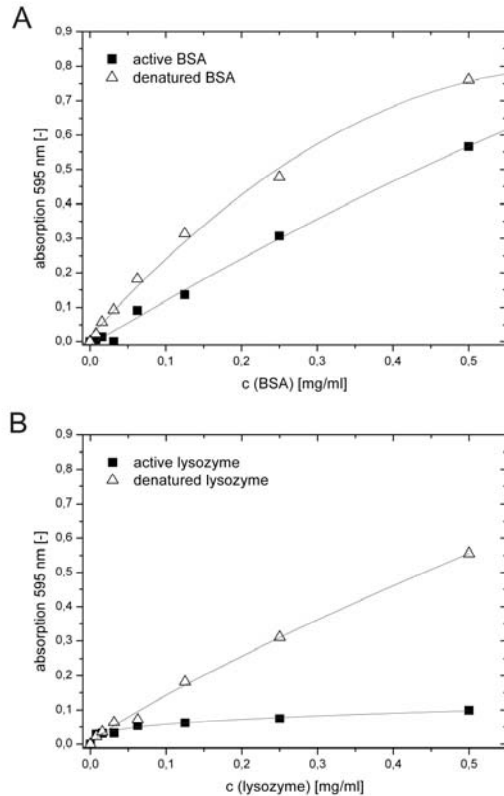
Figure 5: Calibration curves measured at 595 nm for the Bradford assay of 40 µl protein sample with 200 µl assay solution for A) BSA and B) lysozyme. Active lysozyme and BSA are solubilized in 50 mM potassium phosphate buffer at pH 8 and denatured lysozyme and BSA are solubilized for 3 hours in 8 M urea, 50 mM potassium phosphate buffer and 5 mM DTT at 25°C

A clear difference in the concentration measurement between native and denatured protein can be seen in both cases. In general BSA shows higher signals than lysozyme and a broader linear range pinpointing the signal dependency on the actual protein present. Furthermore the signals of the denatured species were consistently higher when compared to those of the native species. However, as the Bradford assay is based on Coomassie Blue binding to protonated arginine and lysine residues, it is not surprising that a difference in accessibility of these residues by denaturation leads to signal changes. This seems to be the case although an at least partial denaturation of proteins with Bradford reagent should be reached.

Buffer Exchange

As direct quantification by absorption at 280 nm is disturbed by absorbing buffer components like glutathione or oxidized DTT, methods for high throughput buffer exchange were validated to remove these substances prior to analysis. The most obvious candidates investigated in this study were ultrafiltration, dialysis and size exclusion chromatography.

a) Ultrafiltration

Ultrafiltration harnesses the separation effect of membranes acting as molecular sieves. Liquids and compounds able to pass the pores are forced through the membrane by centrifugation or vacuum subsequently increasing protein concentration in the retentate.

For high throughput applications 96 well ultrafiltration plates are available which can be operated either by vacuum or centrifugation. Vacuum driven liquid transport strongly depends on a homogenous flow of liquid throughout the 96 wells. As this is not always given, especially when using buffer systems with high viscosities, centrifugation was used throughout the study. Nevertheless, in both methods increased protein concentration in close proximity to the membrane might lead to loss of protein by aggregation often combined with non-specific adsorption at the membrane.

Figure 6 shows the recovery of lysozyme and BSA at different initial concentrations after automated 96 well plate ultrafiltration. Lysozyme and BSA were used at concentrations between 60 and 200 µg/ml. The initial buffer systems for lysozyme was sodium acetate buffer at pH 4 and bicine buffer at pH 9 while for BSA sodium acetate buffer at pH 4 and potassium phosphate buffer at pH 8. The final buffer system was potassium phosphate buffer at pH 8. The three consecutive ultrafiltration steps consisted each of a sample volume reduction to approximately 50 µl followed by an addition of 300 µl potassium phosphate buffer. This dilution strategy enabled a near quantitative removal of 20 mM GSSG present in the initial buffer systems.
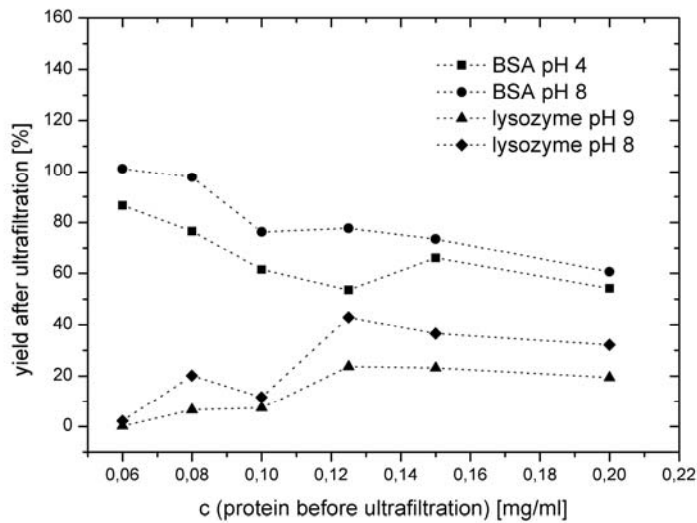
Figure 6: Yield after ultrafiltration for buffer exchange into 50 mM potassium phosphate buffer (pH 8). Lysozyme is initially in 50 mM potassium phosphate buffer (pH 8) or in 50 mM bicine buffer (pH 9). BSA is initially in 50 mM potassium phosphate buffer (pH 8) or in 50 mM sodium acetate buffer (pH 4)

For lysozyme protein recovery for all samples was below 50 % with decreasing yields at lower protein concentration and higher initial pH. The effect of pH can be explained by lower distance to the isoelectric point (pI) which in the case of lysozyme lies around (pI) pH 11. The overall lower net charge leads to a higher tendency to aggregate or to adsorb to hydrophobic surfaces. As aggregation would increase with increasing protein concentration presumably adsorption accounts for loss of lysozyme. For BSA with a pI around pH 5.5 higher recoveries were observed for pH 8 when compared to pH 4. The latter is clearly influenced by the fact that the samples at an initial pH of 4 have to cross the pI of the protein during buffer exchange procedure. In contrast to lysozyme recovery of BSA led for all samples between 50 % and 100 %. The yield reached was decreasing with increasing initial protein concentration proposing strong aggregation effects. The strong relation between pH, protein type and concentration with loss of protein is confirmed by [20]. Thus a quantitative buffer exchange to subsequently measure protein concentration in refolding screening is not possible with automated ultrafiltration.

b) Dialysis

For investigation of HTS compatible dialysis the performance of 96 well dialysis plates with a membrane bottom was examined. The accumulated sample volume inside all wells of the dialysis plate was 24 mL. With a total dialysis buffer volume of 6 L a theoretical dilution factor of 250 is calculated which would lead to an appropriate removal of 20 mM GSSG with absorption at 280 nm on the level of buffer. We dialyzed eight different concentrations of BSA and lysozyme between 0 and 0.5 mg/mL in potassium phosphate buffer at pH 8 with and without addition of 20 mM GSSG. In addition BSA was also dissolved in sodium acetate buffer at pH 4. As dialysis buffer potassium phosphate buffer at pH 8 was chosen. The first dialysis step was performed with 3 L dialysis buffer for 5 h. In a second step the dialysis buffer was exchanged and dialysis proceeded overnight. The absorption at 280 nm was measured after the sample volume in the dialysis plate was measured by liquid level detection with the robotic tips. The blank samples without protein showed clear differences in UV signals for buffer systems with and without GSSG. An absorption of 0.303 ± 0.02 was gained for buffer systems with GSSG after dialysis. The residual concentration of GSSG was calculated with a GSSG calibration curve and laid at approximately 7.24 mM which refers to a dilution factor of 2.76 instead of the theoretical value of 250. Overnight incubation with stirring seems to be insufficient for complete removal of 20 mM GSSG. A crucial factor when considering HTS applications is processing time. Dialysis process time depends on the diffusion properties of the molecules, the membrane area, thickness and composition, temperature, concentration gradient between buffer reservoir and sample and mixing on each side of the membrane [21]. Commercially available 96 well plate dialysis devices like the ones we used in our study are marketed towards complete sample desalting and removal of glycerol and arginine and recommended dialysis time lies between 24 and 48 hours [7]. Vink et al. [22] optimized a 96 well dialysis block to remove detergent and achieved 100 % removal after approximately 250 h with frequent buffer exchange. This is not unusual for dialysis and correlates with non-automated processes using membrane tubes or dialysis chambers [10, 12, 23, 24]. In comparison to ultrafiltration adsorption effects are related to dialysis time whereas aggregation of protein does not play a major role in sample loss. We investigated protein loss in our dialysis experiments after overnight incubation. Protein concentration was measured by UV 280 nm. The yields after dialysis are depicted in Figure 7 for lysozyme and BSA in potassium phosphate buffer at pH 8 and for BSA in sodium acetate buffer at pH 4. High protein recovery between 95 and 160 % is gained for BSA and lysozyme

110

in 50 mM potassium phosphate buffer at pH 8. Lysozyme shows nearly no recovery at protein concentrations below 0.016 mg/mL. A good explanation for this would be membrane adsorption of lysozyme as already observed in ultrafiltration and thus a complete loss of lysozyme in low concentrated solutions. Yields above 100 % result from experimental error during liquid handling steps and measurement of liquid level and UV absorption. BSA initially dissolved at pH 4 has to be shifted across its isoelectric point at pH 5.5 leading to high losses of protein with a recovery between 18 and 41 % for protein concentrations between 0.063 and 0.5 mg/mL and a complete loss of BSA for protein concentrations below 0.0625 mg/mL. As increasing BSA concentration improves yield the observed effect is rather due to membrane adsorption than to aggregation corresponding to theory.
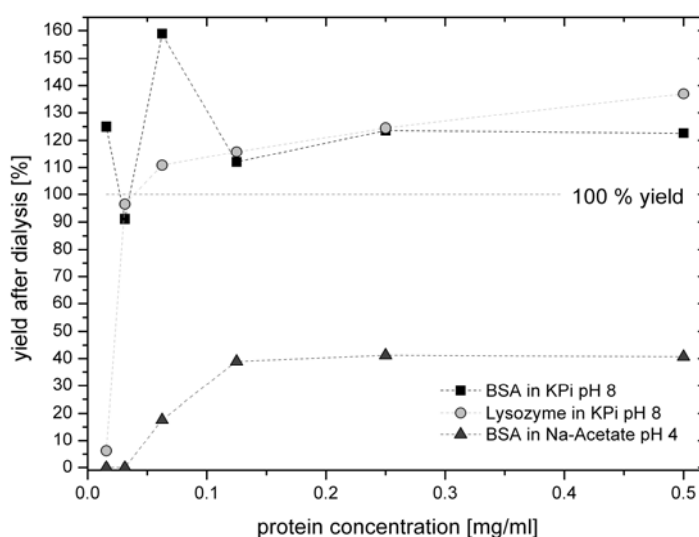

Figure 7: Protein yield after dialysis plotted versus initial protein concentration in the sample for BSA and lysozyme initially dissolved in 50 mM potassium phosphate buffer (KPi) at pH 8 and for BSA initially dissolved in 50 mM Na-acetate buffer at pH 4. Dialysis was performed versus 50 mM potassium phosphate buffer (KPi) at pH 8 overnight at room temperature

c) SEC

After investigation of different gelfiltration resins Sephadex G25 was chosen for further experiments being the most promising candidate. Nevertheless a column length of 5 cm applicable on the robotic platform was not sufficient for baseline separation of lysozyme and GSSG though column load was low with 2 % CV. The chromatogram for a solution containing 0.1 mg/ml lysozyme and 20 mM GSSG is shown in Figure 8. Both peaks are overlapping. To exclude an interaction between GSSG and lysozyme preventing separation

both molecules were injected separately and show the same retention behaviour as in the mixture (Figure 8). Consequently this approach was not pursued.
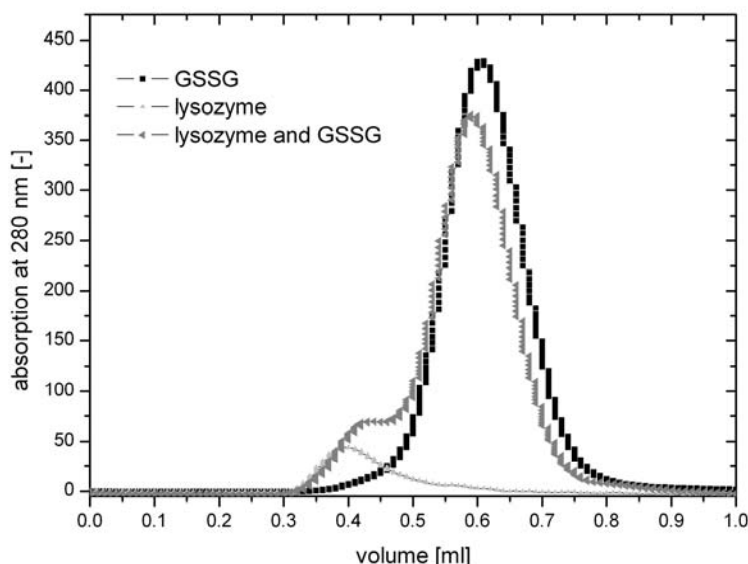

Figure 8: Chromatogram of size exclusion chromatography on a Sephadex G25 column with 0.5 cm diameter and 5 cm length. The injection volume was 2 % CV and the concentration of lysozyme 0.1 mg/mL and of GSSG 20 mM.

When operating size exclusion chromatography column length, particle size, pore size and the ratio of sample volume to column volume are significant parameters for separation results [25]. Low sample loading, high sample dilution and serial processing are major draw backs of SEC. The applicability of SEC on robotic workstations is thus restricted even though solutions for parallelized robotic chromatography are published for ion exchange chromatography in [26].

Quantification of Aggregated Protein

As a conclusion of the above studies no analytical method for a direct protein concentration could be found. We thus developed an indirect quantification method to elucidate soluble protein content. The method is based on the recovery of aggregates, solubilization with the above described denaturation buffer followed by a simple UV 280 nm analysis. Mass balancing enables the determination of soluble protein as depicted in equation 1 where m_{total} describes the total mass of protein introduced initially, $m_{prot\ aggregate}$ aggregated protein and m_{sol} the amount of soluble protein.

$$m_{sol} = m_{total} - m_{\text{prot aggregate}} \qquad \textbf{(1)}$$

The advantage of this method lies in constant buffer conditions, comparable protein structure throughout all samples and most important the absence of all interfering substances. For method validation 40 refolding samples were generated using random buffer systems with a buffer concentration of 50 mM, varying pH values between 3 and 9 and varying NaCl and $MgSO_4$ concentrations between 0 and 150 mM. No redox components were added to allow for UV 280 nm measurement of soluble protein and the wash fraction. After a ten fold dilution of denatured protein into the refolding buffer systems the total protein concentration was 0.4 mg/ml. Following a one hour incubation with constant shaking the aggregates of 300 µl refolding sample were separated as described above.

In order to close all mass balances, the recovery of soluble protein in the filtrate had to be assessed. The recovery of soluble protein in the filtrate $Y_{\text{sol filtrate}}$ was calculated according to equation 2 with the mass of soluble protein in the filtrate $m_{\text{sol filtrate}}$ and the mass of soluble protein m_{sol} measured after aggregate removal by centrifugation of the initial sample plate.

$$Y_{sol,filtrate} = \left(\frac{m_{solfiltrate}}{m_{sol}} \right) \times 100 \qquad \textbf{(2)}$$

The total recovery of protein Y_{total} was calculated according to equation 3 with the mass of protein m_{total} initially introduced, soluble protein in the filtrate $m_{\text{sol filtrate}}$, the mass of protein in the wash fraction m_{wash}, which was 0 mg for all measured fractions, and the mass of protein determined by re-solubilizing aggregates $m_{\text{prot aggregate}}$.

$$Y_{total} = \frac{(m_{sol.filtrate} + m_{wash} + m_{protaggregate})}{m_{total}} * 100 \qquad \textbf{(3)}$$

The recovery of soluble protein in the filtration step and the total recovery of protein using this approach are depicted in Figure 9 for all 40 experiments with error bars for dublicates. Mass balances closed for all samples reaching close to 100 % (for Y_{total} 106 % ± 4 % and for $Y_{\text{sol.filtrate}}$ 102 % ± 6 %) with a mean error of 2 %.
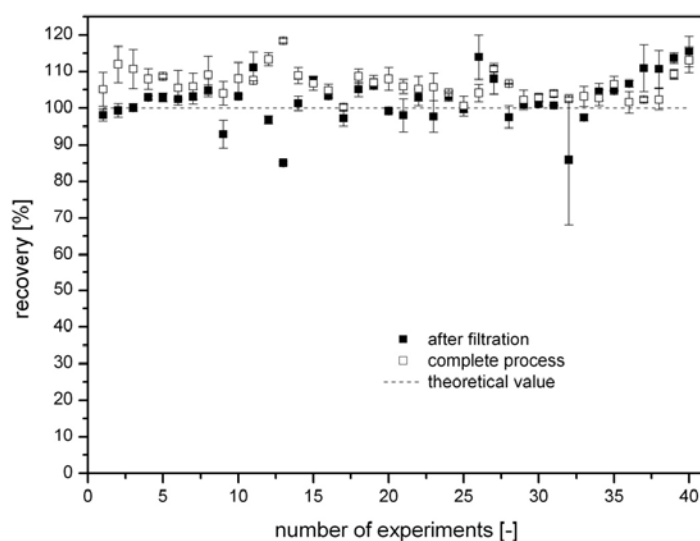
Figure 9: Recovery after filtration and total recovery calculated from protein in filtrate and resolubilized aggregates

Mass balances above 100 % are to a certain extent caused by a carry-over of 0.5 mM DTT with the denatured protein stock solution increasing 280 nm absorption values of the soluble protein and the wash fraction. In the case of complete DTT oxidation the absorption at 280 nm would be 0.075 which corresponds to a lysozyme concentration of 0.017 mg/ml or a yield of 4.35 %.

Hevehan et al. [27] used an approach similar to the one presented in our study to quantify lysozyme aggregates in refolding samples with added glutathione as oxidizing component by absorption at 280 nm. Lysozyme aggregates were separated by centrifugation, washed with TE buffer (0.05 M Tris-HCl, 1 mM EDTA, pH 8) and resolubilized in denaturing and reducing buffer containing 8 M guanidinium chloride and 32 mM DTT. Their findings regarding the potential of aggregate quantification after resolubilization support our results.

The method developed – based on aggregate separation by filtration – is to our knowledge the first published fully automated method for indirect measurement of protein solubility compatible with protein structure and buffer variabilities.

In [28] aggregates formed during refolding are analyzed on a reducing and non-reducing SDS-PAGE . Lysozyme dimers and trimers stabilized by disulfide bonds were detected when oxidizing substances including air were present during refolding. This observation is consistent with data obtained by the above described approach. Protein aggregates could only be solubilized in the presence of reducing components.

114

Turbidity measurements prior to aggregate separation used to assess aggregation might be applied but cannot be classified as a reliable methodology. In Figure 10 A and B turbidity measured at 410 nm and the concentration of resolubilized aggregates is plottes versus the concentration of soluble protein. Data were collected from 40 refolding experiments witht 0.4 mg/ml lysozyme in random buffer compositions. For both functions, turbidity and concentration of aggregates, a linear correlation with soluble lysozyme concentration is observed. The calculated correlation coefficient is better for the data set including resolubilized aggregates with $R^2 = 0.96$ in comparison to the one with 410 nm absorption ($R^2 = 0.91$). This reveals that the described resolubilization method is clearly superior to turbidity measurement for the quantification of solubility. Furthermore it should be mentioned that the linear correlation of turbidity with soluble protein content is worse at higher levels of precipitation.

The scatter observed in turbidity measurements is confirmed by Wang [9] who describes a great influence of aggregate dissolution on the measured absorption values and only limited linearity of turbidity with protein solubility. Therefore turbidimetric measurements are not applicable in a screening approach where quantitative rather than qualitative data are needed for optimization of process parameters. Nevertheless turbidity was used as an objective function in [29]. A threshold value for sample absorption at 410 nm was introduced and samples separated into two groups - soluble samples exhibiting absorption below the threshold value and precipitated samples showing absorption above the threshold value. As samples from these two classes are not further characterized, solubility differences within these classes are not resolved which leads to a great loss of information on parameter effects. In contrast the quantitative measurement of aggregated resolubilized protein allows for a differentiation of precipitated samples and thus increases the information on parameter effects for further optimization.
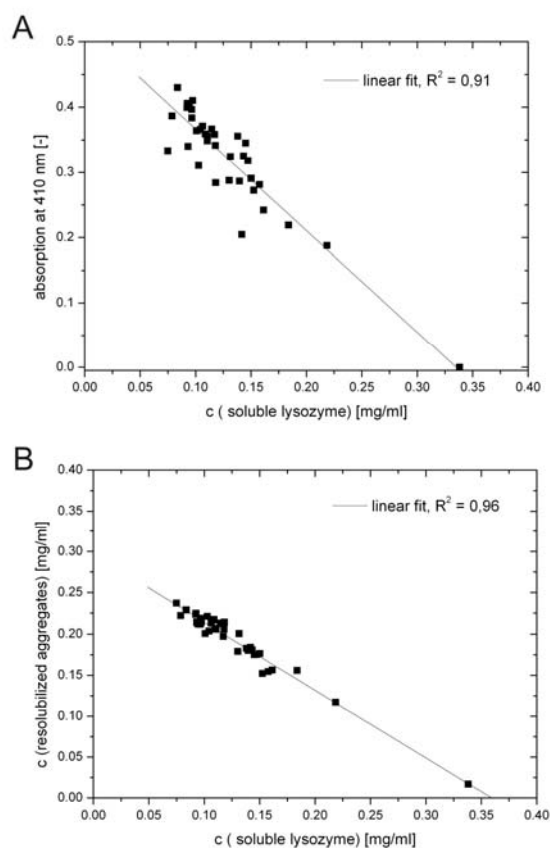
Figure 10: Correlation of soluble lysozyme concentration with **A** absorption at 410 nm of refolding suspension or **B** concentration of resolubilized aggregates (40 randomly mixed buffer systems with 0.4 mg/ml lysozyme)

3.1.2 Structure and Function

Well known methods for protein structure estimation including circular dichroism, RP-HPLC or analytical size exclusion chromatography are not applicable on robotic workstations due to the need for manual intervention and sophisticated instrumentation. Special requirements on sample quality and quantity of homogeneous material cannot be met in a high throughput approach with miniaturized process volumes and the problems mentioned above concerning automated sample preparation. Additionally sample analysis in a serial mode is not consistent with the principle of parallelization used in HTS and consequently leads to analytical bottle necks.

Protein specific techniques to assess structural integrity and activity are binding assays like ELISAs or enzymatic assays which are automatable and easily parallelized. Non-specific methods harnessing changes in intrinsic tryptophan fluorescence or the binding of

hydrophobic fluorescent dyes show the advantage of a more universal applicability and thus reduce time needed for method development if a new protein has to be analyzed.

Enzymatic Activity

The enzymatic assay used to determine lysozyme activity is based on the lysis of *Micrococcus* cells leading to a linear decrease in turbidity of a *Micrococcus* suspension. This approach is used as a standard lysozyme assay since the early 60s and was first described by Jolles [30]. However, when used on the HTS platform settle down of the bacteria before distribution from a storage trough into the assay plates and partial lysis of the suspension during storage was observed. As assay times for the individual samples differ these effects lead to a general decrease in turbidity and thus the applicability of the assay and the stability of the *Micrococcus* suspension had to be investigated to guarantee for comparable assay results during the robotic run time. The absorption at 595 nm of different *Micrococcus* concentrations was measured at different time points of storage in a 96 well plate. In Figure 11 A the absorption values for all measured *Micrococcus* concentrations are depicted for storage times between 0 and 125 min. A decrease of absorption at 595 nm is measured for all concentrations leading to a shift of the linear correlation between absorption at 595 nm and the *Micrococcus* concentration mainly in the y-axis intercept with a rather constant slope. This effect is due to a lysis of the Micrococcus cells over time, as the suspension was distributed into the microtiter plate wells at the beginning of the experiment and as usual mixed before absorption measurement. A settle down of particles was observed to be a problem if the Micrococcus suspension is stored longer than 30 minutes in a storage trough before distribution.
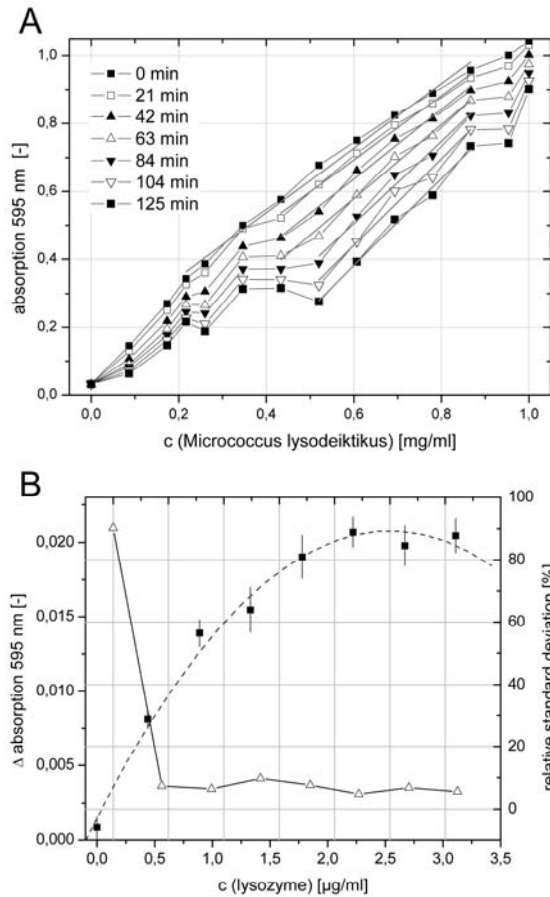
Figure 11: A Absorption 595 nm of *Micrococcus* assay suspensions over time. The linear range is marked by fitted lines. **B** The linear decrease of absorption at 595 nm of a Micrococcus assay suspension (Δ absorption 595 nm) after addition of different lysozyme concentrations (■). The relative standard deviation of Δ absorption 595 nm for 12 replicates (Δ)

All linear fit parameters are summarized in Table 1. A linear absorption range after storage of the *Micrococcus* suspension for more than 21 min requires a minimum *Micrococcus* concentration of 0.42 mg/ml. At 84 min incubation time the minimum *Micrococcus* concentration increases to 0.52 mg/ml. As a consequence of the decreasing linear range of the absorption of the Micrococcus suspension the suspension was always prepared 30 min before start of the distribution in the 96 well plate for the subsequent enzyme assay. A calibration curve of eight different native lysozyme concentrations including a buffer blank is always run in parallel.

Table 1: Overview of the linear curve fit parameters and the upper and lower Micrococcus concentration limit of the linear range (c_{min}, c_{max}) for the decrease of absorption at 595 nm of different Micrococcus concentrations at different time points of storage

Time	0 min	21 min	42 min	63 min	84 min	104 min	125 min
Slope	0.950	0.942	1.019	1.085	1.209	1.275	1.281
Y0	0.158	0.128	0.026	-0.071	-0.220	-0.324	-0.387
R^2	0.992	0.992	0.989	0.989	0.977	0.971	0.990
c_{min}	0.22	0.43	0.43	0.43	0.52	0.52	0.52
c_{max}	0.87	0.87	0.87	0.87	0.87	0.87	0.87

In our setting eight samples were measured in parallel within a time span of 4 minutes. The standard deviation of the applied assay was determined using 12 experimental cycles including 96 samples of native lysozyme. In Figure 11 B the decrease in 595 nm absorption and the relative standard deviation of of the 12 cycles are summarized. The data reveal a relative standard deviation between 7 and 8 % for all lysozyme concentrations above 0.5 µg/ml. In the actual refold screenings one standard calibration curve per 40 refolding experiments was introduced in the experimental set-up to be on the safe side.

Despite for the rather poor storage stability of a *Micrococcus* assay suspension automated measurement of lysozyme activity is possible during HTS applications. In Lee et al. [31] the *Micrococcus* based lysozyme assay is presented in a microplate format with faster sample processsing and measurement of 96 samples in 7 min resulting in a standard deviation of only 4.8 %. As details on the devices used for sample processing are not available in this publication an evaluation of the reasons for this reduced standard deviation could not be undertaken.

Tryptophan Fluorescence

The measurement of intrinsic tryptophan fluorescence emission spectra by excitation at 280 nm provides information on the direct environment of hydrophobic tryptophan residues. In correctly folded, active proteins most tryptophans are buried inside the hydrophobic core of the molecule. During unfolding or in misfolded protein intermediates tryptophan can be exposed to polar solvent molecules leading to a shift in the emission spectrum towards higher wavelengths due to an energy transfer from tryptophan to solvent. The method is restricted to proteins with at least one buried tryptophan. Typical emission spectra are shown in Figure 12

for 4 mg/ml active and correctly folded lysozyme with a maximum emission intensity at 330 nm and lysozyme denatured and reduced with 8 M urea and 5 mM DTT at 25 °C for 24 h and maximum emission intensity at 341 nm. It is important to note that a native tryptophan spectrum does not necessarily correlate with 100 % activity but gives only a hint at the integrity of the tertiary structure and the position of hydrophobic tryptophan.
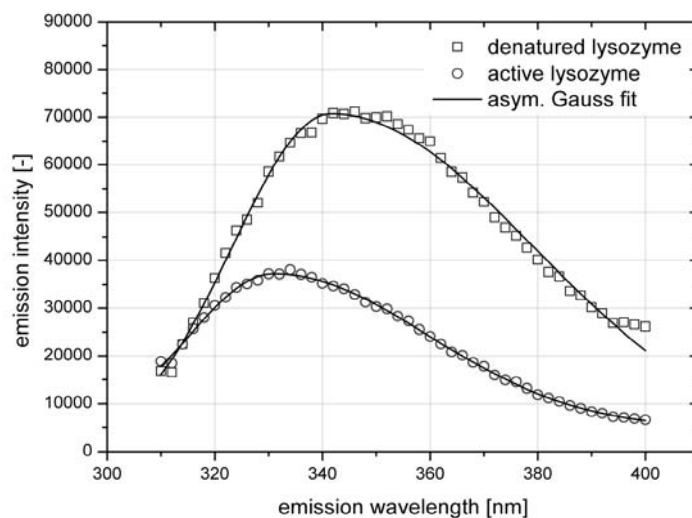

Figure 12: Tryptophan emission spectra of active lysozyme in 50 mM potassium phosphate buffer (pH 8) and of denatured lysozyme incubated for 24 h at 25 °C in 8 M urea, 50 mM potassium phosphate buffer (pH 8) and 5 mM DTT. The samples are excited at 280 nm

To quantify the rather small spectral shifts all spectra were fitted with an asymmetric Gaussian curve. The wavelength obtained at the maximum intensity of the fitted spectra – λ (I_{max}) – was used as a basis for structure estimation and as a measure for the ratio of active to denatured protein. The comparison of tryptophan spectry of lysozyme after refolding and of a native lysozyme standard is one of the frequently used methods to measure structural integrity of refolded protein [32]. Kinetic measurements to assess the progress of refolding [33] or unfolding [34] are performed by measurement of tryptophan fluorescence changes over time. A shift of 9 to 12 nm is confirmed for unfolding of lysozyme by Lin et al. [34] however with differing wavelengths for the maximum intensities of active (341 nm) and denatured lysozyme (352 nm) due to different spectroscopic equipment.

Measurement of single emission intensities in general also yields information on protein concentration but is strongly dependent on the folding state of the protein and the buffer composition [35]. The addition of GSSG for example quenches tryptophan emission by an energy transfer to the disulfide bond leading to an exponential decrease of emission intensity

with increasing GSSG concentration as depicted in Figure 13 for tryptophan and active lysozyme.
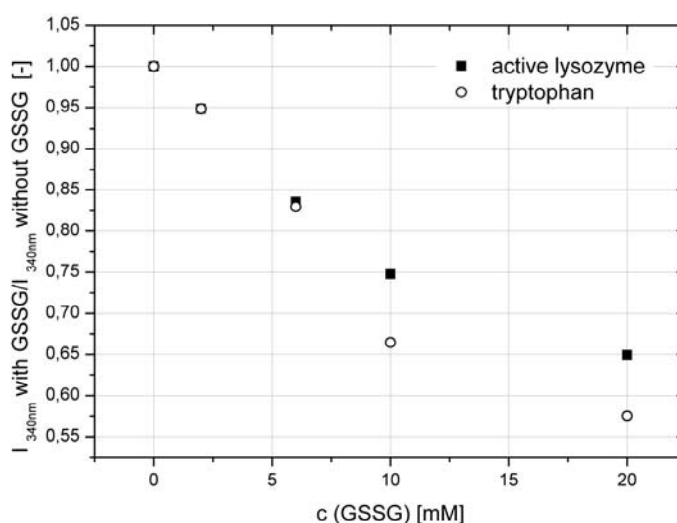

Figure 13: Ratio of emission intensity at 340 nm of lysozyme and tryptophan in 50 mM potassium phosphate buffer (pH 8) with different concentrations of GSSG to the emission intensity of samples in the same solution without added GSSG. Tryptophan was excited at 280 nm

In order to assess the potential of tryptophan spectra to estimate protein structure integrity in refolding samples refolding experiments with and without added redox system components were performed. Lysozyme was denatured for 4 h in 8 M urea, 50 mM potassium phosphate buffer at pH 8 and 5 mM DTT. Refolding was performed using a randomly distributed set of 160 different refolding buffer systems with and without addition of redox components. The parameter space for the random sample preparation is summarized in Table 2.

Table 2: Overview of the used parameter design space

Parameter	Design Space
buffer pH	3 - 9, discrete, step size of 1
NaCl concentration	0 - 150 mM
$MgSO_4$ concentration	0 - 150 mM
GSSG concentration	0 - 20 mM
DTT concentration	0 - 10 mM

Tryptophan emission spectra are measured between 300 and 400 nm in the fraction of soluble protein after removal of aggregates with excitation at 280 nm. The wavelength at maximum emission intensity is calculated from curve fits. Concentration of soluble and active lysozyme

was measured as described above to calculate the ratio of active to total soluble lysozyme. In Figure 14 the wavelength values in the emission maximum, λ (I_{max}), and the fraction of active to soluble lysozyme for all experiments are summarized to assess, if the position of fluorescence spectra can be used as a measure of specific activity.
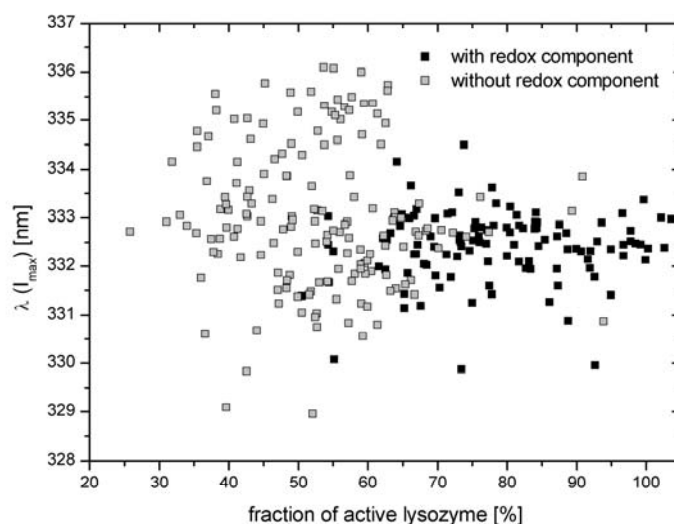

Figure 14: Correlation of active lysozyme fraction with the wavelength in the emission maximum ($\lambda(I_{max})$) for refolding samples in different buffer systems containing either GSSG and DTT as redox system or no redox components and 0,4 mg/ml lysozyme. Lysozyme was denatured for 4 h at 25°C. Tryptophan was excited in the soluble protein fraction at 280 nm after 1 h refolding

It becomes obvious that measured soluble lysozyme samples show a narrow range of emission maxima between 330 and 336 nm. As native lysozyme has its emission maximum at 330.9 ± 0.4, spectra of refold samples are red shifted, indicating a higher tryptophan accessibility for polar buffer molecules. In the investigated design space no correlation of high specific activity to native spectral behavior is observed. It is likely that hydrophobic tryptophan residues have to be buried to a certain extent to allow for lysozyme solubility. This is in high correspondence with a described hydrophobic collapse of the protein as a first step in refolding for higher thermodynamic stability [36]. Interestingly samples without added redox reagents show a wider distribution of spectral maxima, which might be due to a higher flexibility of the protein without stabilizing disulfide bonds. As described in literature the specific activity after refolding without addition of substances promoting disulfide bond formation is rather low although the protein is soluble and shows similar secondary structure and amount of buried tryptophan residues than the native lysozyme [37]. The importance of redox system control will be discussed in detail below. Summarizing from the data presented

above it is not possible to resolve lysozyme structure with the needed resolution by measurement of tryptophan fluorescence spectra in refolding screening.

3.2 Control of Refolding Parameters

3.2.1 Solubilization of Protein

The characterization of the unfolded state of a protein is essential for a comparison of data concerning optimum refolding conditions as well as refolding yields of active protein. Various unfolding conditions are used in literature for lysozyme, e.g. unfolding with different protein and DTT concentrations, different incubation time and buffer pH as summarized in Lin et al. [34]. Examples for different unfolding conditions are given in Table 3. In addition to varieties in solvent conditions the reaction temperature represents another variable factor in lysozyme refolding publications. While in Middelberg et al. [38] 37 °C is used during denaturation, Lin et al. performed denaturation at 25 °C [34]. In general a complete specification of the unfolding state of the protein is often missing and therefore comparability of published results is complicated. Additionally optimization of refolding parameters is only possible if the starting material including the unfolded protein feed stock is kept at constant quality.

Table 3: Different unfolding conditions as summarized by Lin et al. [34]

Lysozyme [mg/ml]	Denaturant	DTT [mM]	Buffer	Unfolding Time [h]
5	8 M urea	10	0.1 M Tris, pH 8.6	24
5 – 35	8 M urea	10	0.1 M Tris, pH 8.5	2
30 – 80	8 M urea	30	0.1 M Tris, pH 8.5	1,33
50	8 M urea	30	0.1 M Tris, pH 8.6	3
3 – 25	8 M urea	32	0.05 M Tris, pH 8.0	1
10	6 M Gnd/HCl	150	0.1 M Tris, pH 8.6	3
20	6 M Gnd/HCl	150	0.1 M Tris, pH 8.6	15
10 – 80	8 M urea	150	0.1 M Tris, pH 8.6	2
2.6 – 30	8 M urea	200	0.1 M Tris, pH 8.7	4 - 5

In order to obtain a high reproducibility of feedstocks we initially evaluated lysozyme unfolding kinetics for three different protein concentrations at three different temperatures in

a constant denaturing buffer (8 M urea, 50 mM potassium phosphate and 5 mM DTT) by measurement of tryptophan fluorescence. This method was earlier described in studies on lysozyme unfolding behavior [34].

Temperature

In Figure 15 the kinetic of lysozyme unfolding using a 4 mg/ml lysozyme solution at 4 °C, 25 °C and 37 °C is depicted as an increase in the wavelength at the maximum emission intensity of tryptophan excited at 280 nm.
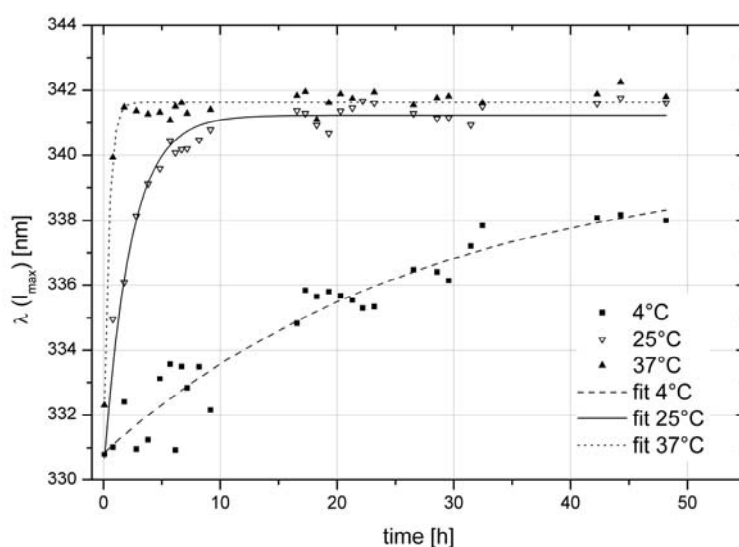

Figure 15: Effect of temperature on the unfolding kinetics of 4 mg/ml lysozyme in 8 M urea, 50 mM potassium phosphate buffer at pH 8 and 5 mM DTT measured by a shift of wavelength at maximum emission intensity (λ (I_{max})) with tryptophan excitation at 280 nm. The data points are fitted exponentially

Data was fitted using the following exponential function

$$y = A_1 * \exp(\frac{-x}{t_1}) + y_0 \tag{4}$$

where y_0 represents the wavelength at the maximum emission shift. The initial slope of the shift over time A_1/t_1 is a measure for the reaction velocity.

With increasing temperature an increase in the initial reaction velocity represented by the initial slope of the fit curve occurs. For all three temperatures the final λ (I_{max}) was found to lie in the range of 340 nm to 342 nm. The wavelength characterizing native lysozyme was determined to be 331 nm with a relative standard deviation of 0.15 % calculated from

triplicate measurement of active lysozyme. Measured differences at maximum shift of the spectrum represented by y_0 are significant with a deviation of 0.25 % hinting at an influence of temperature on unfolded protein quality also at steady state unfolding. In Table 4 the data calculated for unfolding at the three different temperatures are summarized. This finding is consistent with published literature where the temperature effect on lysozyme unfolding is explained by a decrease in stability of the folded lysozyme with increasing temperature [39].

Table 4: Parameters of the exponential curve fit of the unfolding of 4 mg/ml lysozyme in 8 M urea, 50 mM potassium phosphate and 5 mM DTT measured as a shift in tryptophan fluorescence spectra for three different temperatures

Unfolding Temperatur	4 °C	25 °C	37 °C
Initial slope m [nm/h]	0.33	4.75	27.00
y0 [nm]	339.95	341.22	341.60
R^2	0.92	0.97	0.98

Protein Concentration

In Figure 16 kinetic data for lysozyme unfolding at 25 °C with 2 mg/ml, 4 mg/ml and 8 mg/ml lysozyme are depicted.
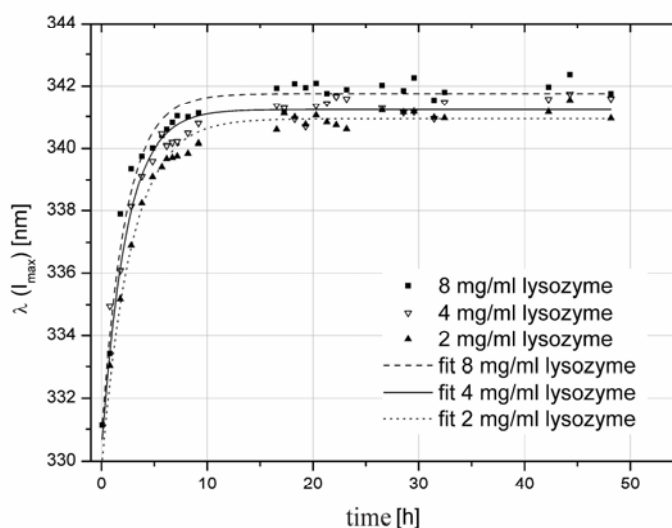

Figure 16: Effect of lysozyme concentration on unfolding kinetics at 25 °C in 8 M urea, 50 mM potassium phosphate buffer at pH8 and 5 mM DTT measured by a shift of wavelength at maximum emission intensity (λ (I_{max})) with tryptophan excitation at 280 nm.

An exponential increase of λ (I_{max}) converging towards a maximum value is measured for all protein concentrations. The initial reaction velocity – m – decreases with increasing protein concentration by a factor of 1.2 from 2 mg/ml to 8 mg/ml. The maximum shift of the emission spectrum y0 lies for all three protein concentrations between 341 nm and 342 nm with only a small deviation of 0.13 %. Thus an effect of protein concentrations on unfolding is not resolved by tryptophan spectroscopy. In Table 5 data from the corresponding curve fits are summarized.

Table 5: Parameters of the exponential curve fit of the unfolding of lysozyme in 8 M urea, 50 mM potassium phosphate and 5 mM DTT at 25 °C measured as a shift in tryptophan fluorescence spectra for three different protein concentrations

Lysozyme Concentration	2 mg/ml	4 mg/ml	8 mg/ml
initial slope m [nm/h]	4.90	4.75	4.03
y0 [nm]	341.80	341.22	340.90
R^2	0.97	0.97	0.99

The small differences observed for the unfolding of different protein concentrations might be due to a different ratio of reducing component DTT to lysozyme. The molar ratio of DTT to lysozyme changes from 36 at 2 mg/ml protein to 18 for 4 mg/ml and finally to 9 for 8 mg/ml lysozyme. For all lysozyme concentrations DTT is theoretically present in excess even if all four disulfide bonds have to be reduced. In practice lack of DTT for complete reduction of lysozyme at higher concentrations can be due to competitive oxidation by air.

DTT Concentration

To address possible limitations of the reducing reagent DTT during denaturation we added 8 mM, 20 mM and 100 mM DTT to three samples already exhibiting steady state unfolding after addition of 5 mM DTT. This experiment is performed with each of the above analyzed lysozyme concentrations. The molar ratios of total DTT to lysozyme for all combinations were 57 (8 mM DTT), 143 (20 mM DTT) and 716 (100 mM DTT) for 2 mg/ml; 29 (8 mM DTT), 72 (20 mM DTT) and 358 (100 mM DTT) for 4 mg/ml and 14 (8 mM DTT), 36 (20 mM DTT) and 179 (100 mM DTT) for 8 mg/ml. The λ (I_{max}) value was measured after unfolding at 25 °C for 40 hours. In Figure 17 λ (I_{max}) at steady state is plotted versus the molar DTT to protein ratio for all performed unfolding experiments. With an increasing ratio

of DTT to lysozyme an exponential increase of λ (I_{max}) is observed. As no saturation is reached within the experimental range DTT seems to be still limiting for lysozyme unfolding at least at 25 °C.
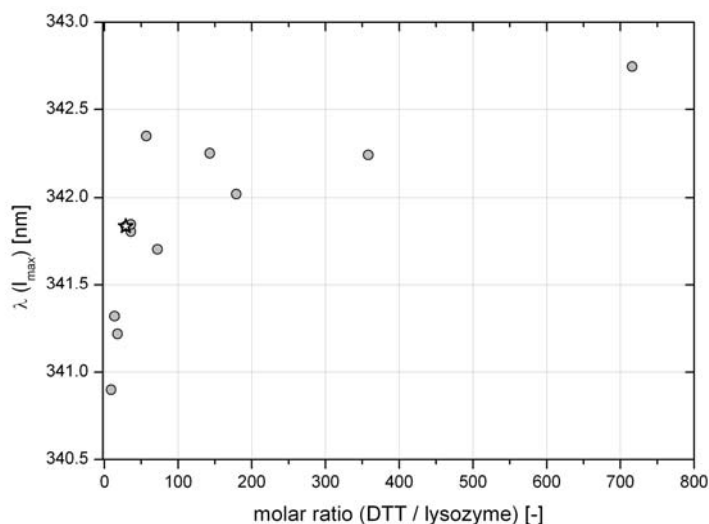

Figure 17: Wavelength at maximum emission intensity (λ (I_{max})) with tryptophan excitation at 280 nm for different molar ratios of DTT to lysozyme. Protein was unfolded for 40 h in 8 M urea, 50 mM potassium phosphate buffer (pH 8). A molar ratio of DTT to lysozyme of 29 is marked as a star representing the transition point from one- to two-phase unfolding in literature.

In Lin et al. [34] a second phase during unfolding was observed for molar DTT to lysozyme ratios above 29 with a maximum emission shift λ (I_{max}) of 12 nm whereas samples with lower DTT to lysozyme ratios show only the plateau of the first unfolding phase with an emission shift of 10 nm. Data points in Figure 17 exhibit a steeper slope of λ (I_{max}) with increasing ratios of DTT to lysozyme below 29 (marked with a star) and an asymptotic convergence of λ (I_{max}) values at DTT to lysozyme ratios above 29. The first curve segment thus represents a stronger limitation of DTT during unfolding whereas systems in the second curve segment are close to DTT saturation. In comparison to the data in [34] we see a shift of λ (I_{max}) with an increment between 9.9 and 10.8 nm for DTT to lysozyme ratios below 29 and a λ (I_{max}) shift with an increment between 10.8 and 11.8 for DTT to lysozyme ratios above 29. As a conclusion in the analyzed range a transition between one-phase and two-phase unfolding is observed. For our refolding experiments denaturation was performed at a DTT to lysozyme ratio of 18 which, with an emission shift of 10.2 nm, represents only partial unfolding of lysozyme.

3.2.2 Combined Parameter Effect on Refold Data

To further elucidate the effect of denaturation conditions on refold screening data we used 4 mg/ml lysozyme solution denatured for 4 hours at 25 °C and lysozyme denatured for 1.5 hours at 37 °C. The respective values for λ (I_{max}) were measured to be 339.3 nm and 341.3 nm respectively. The difference in the λ (I_{max}) value of 0.42 % between these two unfolding states lies above the experimental error and is thus significant.

Both solutions were subsequently used for 40 different refolding experiments randomly covering the parameter space summarized in Table 2 above.

In Figure 18 the ratios of active and soluble lysozyme obtained in the 40 refold systems for both denaturing conditions are summarized.
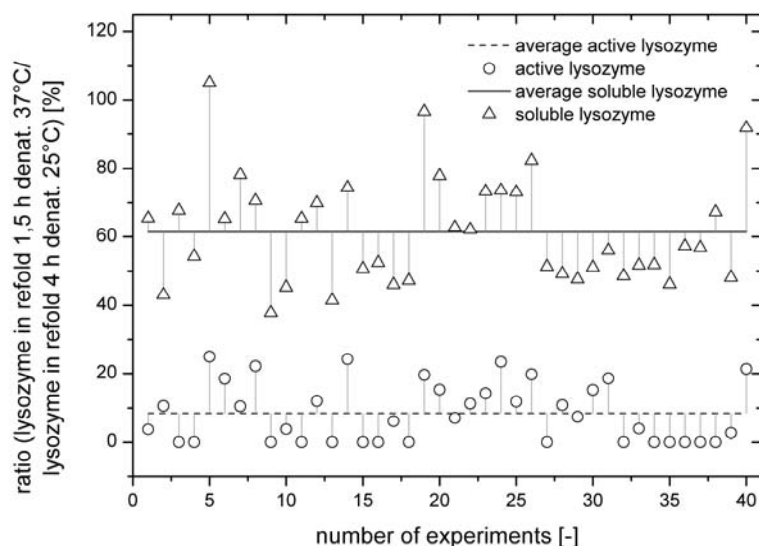

Figure 18: Ratio of active and soluble lysozyme yields for lysozyme denatured for 1.5 h at 37 °C compared to lysozyme denatured for 4 h at 25 °C. Refolding is performed in 40 different buffer systems with additions of GSSG and DTT. The average of all 40 experiments is depicted as solid line for the yield ratio of active lysozyme and as broken line for the yield ratio of soluble lysozyme

Data points describing the amount of soluble lysozyme obtained for the systems denatured at 37 °C scatter around a value of 61 % when compared to data obtained for systems denatured at 25 °C. A similar result was achieved when assaying for active lysozyme. Here data scattered around an average value of 8 % when comparing the systems denatured at 37 °C and 25°C. This equates to concentrations of active lysozyme between 0 mg/ml and 0.08 mg/ml for the systems denatured at 37 °C and 0.12 mg/ml and 0.3 mg/ml for systems denatured at 25 °C.

The concentration of active lysozyme was thus significantly lower than the soluble amount for both systems. Furthermore the ratio of active to soluble lysozyme obtained during refolding showed no consistent pattern between and within both systems. As data variation can not be explained by a consistent pattern both denaturation protocols seem to lead to significantly different initial feedstock characteristics. This also corresponds to the detected variance of the fluorescence spectra described above. These results point out that defined and constant denaturation conditions are indispensable when comparing performance and yield of refolding processes.

Furthermore tryptophan fluorescence has shown to be a good methodology to measure unfolding kinetics. The obtained spectra have to be evaluated carefully in order not to misinterpret information on structural properties of a protein due to low resolution of the method.

When comparing active and soluble amounts of lysozyme resulting from the two different feedstock qualities, it becomes clear that solubility is by far less sensitive towards this parameter than activity. Lin et al. [34] investigated the refolding of different preparations of denatured lysozyme which differ in tryptophan fluorescence emission maxima. A clear dependency of refolding kinetics for differently denatured lysozyme was observed. Relatively mild unfolding conditions with lower ratios of DTT to lysozyme and shorter unfolding times showed faster refolding kinetics leading to higher refolding yields in a given timeframe.

In Figure 19 average parameter values with corresponding standard deviations for the 5 experiments with the highest gained yield of active lysozyme (A) and soluble lysozyme (B) are visualized for both denaturation conditions.
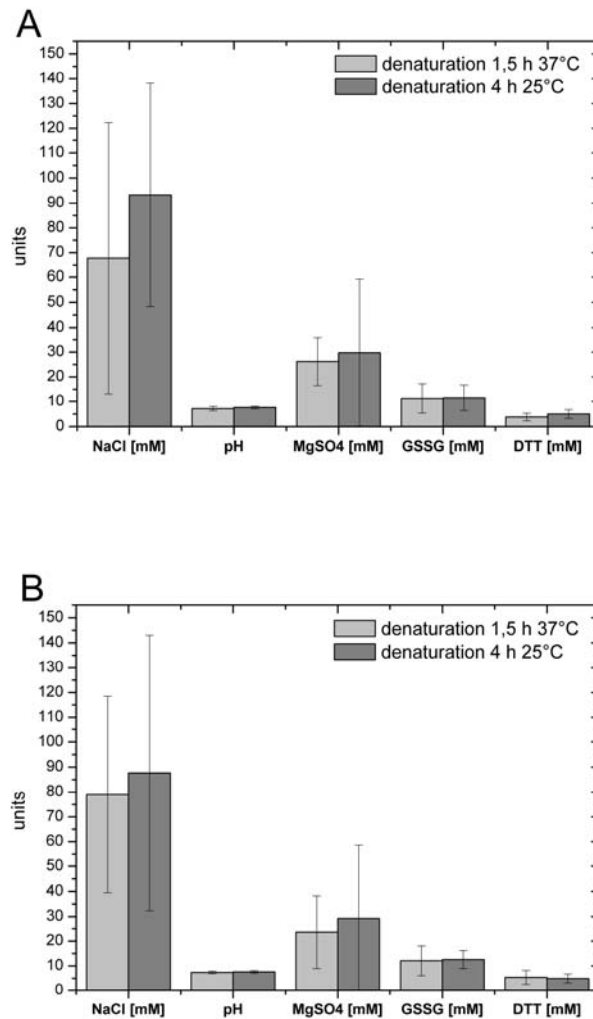
Figure 19: Average parameter values of the 5 best refolding experiments of 40 in respect to **A** active lysozyme yield and **B** soluble lysozyme yield with differently denatured lysozyme feed stocks (8 M urea, 50 mM potassium phosphate buffer (pH 8), 5 mM DTT) and the standard deviation

The most intriguing finding is that the best parameter sets regarding optimal solubility and lysozyme activity yield cover a similar parameter space regarding all five parameters investigated. Besides, this observation is independent of the differences encountered due to the variations in the unfolded feedstocks. The only and moderate differences were found in changes concerning the concentration of NaCl and $MgSO_4$ present in the refolding buffer. Following the above results, differences in feedstock quality might not be prone to change favorable refolding conditions significantly.

As a conclusion suitable conditions identified during two individual screenings each performed with constant conditions for protein denaturation might correspond to each other. Nevertheless gained yields will likely differ due to a strong influence of protein starting material on yield of active protein. As a consequence automated optimization including

multiple screening runs requires a constant feedstock preparation in order to perform parameter optimization on the basis of the highest reproducibility possible.

Furthermore the results gained demonstrate that the conditions for storage such as storage time and temperature have to be considered in addition to the composition of the denaturing feedstock preparation. This is especially true for automated optimization processes, as all solutions need to be placed on the robotic platform prior to the start of the screening process. When using unstable solutions these preparations have to be performed prior to every experimental run with a detailed protocol in terms of preparation time, storage and usage of the solutions.

3.3 Redox Systems during Refolding

Despite the control of unfolding conditions to guarantee for initial feedstock reproducibility the control of the redox potential in the refolding buffer plays a major role during optimization of refolding buffer conditions. Besides different combinations of reducing and oxidizing reagents like DTT, GSSG, cysteine, cystine and air oxidation often assisted by metal ions is frequently used during refolding of disulfide bridged proteins. The main advantage of air oxidation is the saving of expensive redox components which have to be removed afterwards. These advantages are compromised by the low mass transfer rate of oxygen in buffer solutions resulting in significantly lower refolding yields or even no refolding at all as observed for lysozyme. This is set in comparison to the samples including GSH/GSSG as a low molecular weight redox couple with refolding yields above 70 % [40, 41]. Ho et al. [28] published a value of 10 % active lysozyme yield by refolding without added disulfide components with a high loss of protein due to precipitation in contrast to 50 % yield with added GSSG. Other disadvantages of refolding with sole oxidation by air are slow refolding rates and a low reproducibility of refolding results due to insufficient control of aeration [42]. An influence of different types of redox couples is observed by Raman et al. [40]. In refolding of lysozyme between 20 and 50 % higher yields are gained for a cysteine/cystine couple compared to a GSH/GSSG couple. Interestingly this effect seems to be dependent on protein concentration because it was only observed for protein concentrations above 50 μg/ml.

In order to shed some light into these issues in regard to HTS strategies two times 40 refold systems with a concentration of 0.4 mg/ml lysozyme denatured for 4 h at 25 °C were generated. The two groups represented systems missing any redox component besides

introduced air and a combination of DTT/GSSG. Variations were further introduced by randomly distributed changes in buffer pH and concentrations of NaCl and MgSO$_4$. The data was initially used to assess whether differences can be observed between air oxidation and a controlled redox environment consisting of low molecular weight redox substances. Furthermore trends in buffer parameter effects and yields in a typical first screening were evaluated. The experimental space is summarized in Table 2. In Table 6 yields of active and soluble lysozyme are shown for the best 5 experiments in respect to active lysozyme yield with and without the addition of redox reagents and the corresponding parameter values. Corresponding data for the 5 best experiments in respect to soluble lysozyme yield are shown in Table 7.

Table 6: Parameters and yields of active and soluble protein for the best 5 of 40 experiments in respect to active lysozyme yield with and without DTT and GSSG addition. The refolding samples contained 0.4 mg/ml lysozyme and were incubated for 1 hour at room temperature

Best 5 Experiments of 40 with Added Redox Components						
c (NaCl) [mM]	pH	c (MgSO$_4$) [mM]	c (GSSG) [mM]	c (DTT) [mM]	yield active lysozyme [%]	yield soluble lysozyme [%]
62	8	10	9	5	74	96
30	7	34	4	3	73	90
122	8	77	17	7	64	96
119	7	2	13	5	63	94
134	8	26	15	6	63	94
Best 5 Experiments of 40 without Added Redox Components						
c (NaCl) [mM]	pH	c (MgSO$_4$) [mM]	c (GSSG) [mM]	c (DTT) [mM]	yield active lysozyme [%]	yield soluble lysozyme [%]
114	6	147	-	-	50	66
35	3	144	-	-	50	55
67	4	111	-	-	48	53
73	5	132	-	-	48	62
3	6	38	-	-	46	66

Table 7: Parameters and yields of active and soluble protein for the best 5 of 40 experiments in respect to soluble lysozyme yield with and without DTT and GSSG addition. The refolding samples contained 0.4 mg/ml lysozyme and were incubated for 1 hour at room temperature

Best 5 Experiments of 40 with Added Redox Components						
c (NaCl) [mM]	pH	c (MgSO$_4$) [mM]	c (GSSG) [mM]	c (DTT) [mM]	yield active lysozyme [%]	yield soluble lysozyme [%]
122	8	77	17	7	64	96
62	8	10	9	5	74	96
2	7	31	9	2	61	94
134	8	26	15	6	63	94
119	7	2	13	5	63	94
Best 5 Experiments of 40 without Added Redox Components						
c (NaCl) [mM]	pH	c (MgSO$_4$) [mM]	c (GSSG) [mM]	c (DTT) [mM]	yield active lysozyme [%]	yield soluble lysozyme [%]
92	4	20	-	-	39	73
53	8	64	-	-	36	73
134	8	26	-	-	38	73
101	7	8	-	-	40	72
30	7	34	-	-	38	71

Clearly the addition of redox components showed a positive effect on the yield of active lysozyme with an average of 67 % for the best five experiments in contrast to 48 % with air oxidation (Table 6) and on the yield of soluble lysozyme with an average of 95 % compared to 73 % (Table 7).

In Figure 20 average parameter values with corresponding standard deviations for the best 5 experiments are visualized for both oxidation with addition of GSSG and DTT and sole air oxidation in respect to the yield of active lysozyme in A and soluble lysozyme in B. Despite the overall lower yield in soluble protein content in the absence of controlled redox conditions, parameter values leading to high solubility achievable in the two groups are rather similar (Figure 20 B). This underlines the above findings that conditions to achieve soluble protein are less critical to sample history or process parameter variations. In contrast to this parameter settings for high yields of active lysozyme differed significantly for both groups (Figure 20 A). Again clear changes in the concentration of NaCl and MgSO$_4$ but also in pH were observed. .
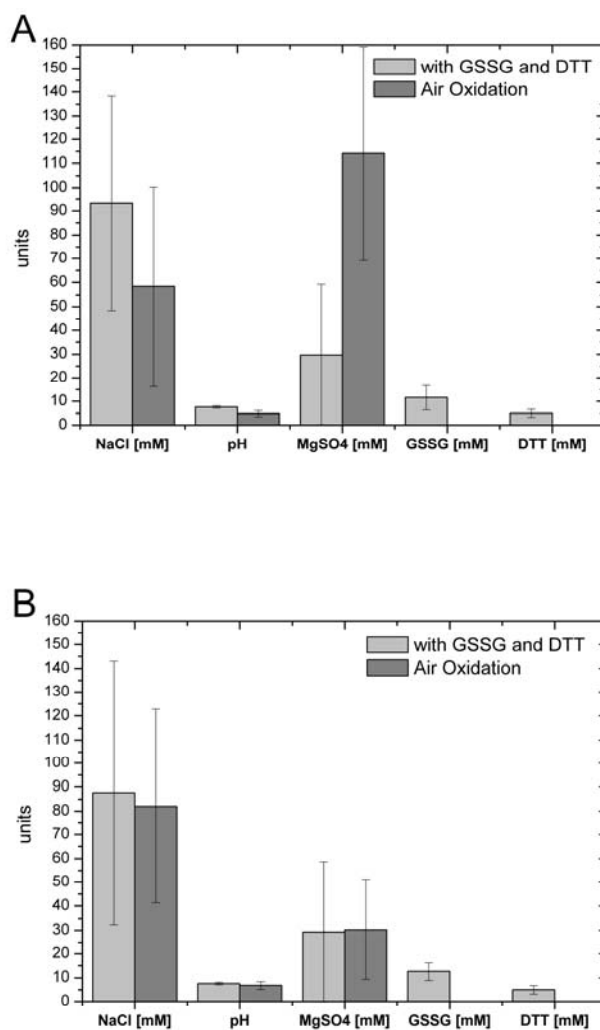
Figure 20: Average parameter values of 5 best refolding experiments of 40 in respect to **A** active lysozyme and **B** soluble lysozyme yield with addition of GSSG and DTT and without (air oxidation) and the standard deviation. Lysozyme was denatured 4 h at 25°C in 8 M urea, 50 mM potassium phosphate buffer (pH 8), 5 mM DTT

Furthermore, while for the redox controlled systems four out of the best five systems corresponded for high solubility and activity, the systems with only air as redox system leading to high solubility did not lead to high activity and vice versa. As a consequence optimization of lysozyme solubility will only lead to an optimized lysozyme folding with high yields of active protein with a controlled redox environment. Screening approaches with air as oxidizing environment will most likely not lead to any meaningful result when considering the post screening step where lab or pilot scale is investigated under controlled redox environment.

Data analysis revealed that in order to achieve high active lysozyme yields the best systems – with controlled redox environment – were characterized by a pH of 7 and 8 whereas in samples without redox control pH showed no effect on the refolding yield. In literature alkaline pH is reported to be crucial for disulfide bond formation, as the reaction is driven by formation of the thiolate anions [5]. Thus the pH plays a major role in samples with redox reagents due to the potentially strong influence of the redox components on the folding process whereas air oxidation with its restricted refolding rate does not lead to pH dependent refolding.

The favorable salt concentrations of NaCl and $MgSO_4$ showed a slight tendency towards 3 to 4 times higher concentrations of NaCl when compared to $MgSO_4$ for samples with added redox system. However a broad distribution of salt concentrations between 30 and 134 mM was observed. There is no systematic investigation of salt effects on refolding yields published despite their key function in protein solubility. The well-known Hofmeister series [43] is not only used to assess the influence of salts on the solubility of native but also on that of refolding protein [44]. The tendency to lower salt concentrations for $MgSO_4$ in comparison to NaCl fits with a higher salting-out effect of $MgSO_4$. Additionally the ionic strength of $MgSO_4$ is four times higher at the same concentration than for NaCl. In general solubility is improved with low ionic strength as is observed in our results as well.

In contrast without redox components high concentrations of $MgSO_4$ above 100 mM seem to be favorable for high yields of active lysozyme. This might arise from the salting-out effect of $MgSO_4$ which promotes folding just by reducing the amount of soluble protein competing for air oxidation. This assumption is further supported by the results for the best experiments concerning soluble lysozyme yield showing lower $MgSO_4$ concentrations between 8 and 64 mM. The best parameter sets for lysozyme solubility in return lead to decreased yields of active lysozyme from 50 to 40 % compared to the best conditions for active lysozyme yields with lower solubility. Furthermore a slowdown of folding is described in Raman et al. [40] for samples with higher protein concentration in lysozyme refolding with at least 1 mM GSSG. It might thus well be that we see a kinetic effect where low refolding rates due to higher protein concentrations lead to low yields of active lysozyme after a given refolding time of 1 h. As no refolding kinetics were measured in this study it might be that high concentrations of soluble lysozyme lead to high yields of active lysozyme at steady state.

The optimal ratio of DTT and GSSG lies between 1.8 and 2.6 are observed regarding the best active lysozyme yields. A ratio of oxidizing to reducing agent around 2.5 with excessive

oxidizing component is reported for good refolding results [27], [34], [38] and thereby confirms our data.

In summary we could observe in a random parameter screening of up to five parameters with the presented high throughput screening method good refolding results at buffer pH, salt concentration and redox systems confirmed by literature. The control of the redox environment has shown to be a prerequisite for high refolding yields and rates. Furthermore we investigated if an optimization of solubility leads to a concomitant optimization of folding towards active protein. This could only be confirmed for systems with added redox components and was not possible with air oxidation.

4 Conclusions

In this study a completely automated strategy for a screening procedure towards optimal refolding conditions was established. It became clear that conventional methods for the determination of protein concentration fail with the task given. In this context a new method was developed which bases on the recovery and resolubilization of aggregated protein followed by a subsequent concentration analysis of the now unfolded protein using UV 280. This method was successfully validated by closure of the overall mass balance. When assaying for structural integrity, two methods were investigated, emission of tryptophan fluorescence and enzyme activity. While tryptophan fluorescence showed to be useful to characterize process parameters it did not help to elucidate structural integrity. Unfortunately soluble lysozyme shows no spectral diversity correlating with specific activity. This effect is likely originating from the absolute need for buried tryptophan residues to prevent protein from aggregation and consequently a uniform distribution of tryptophans in all soluble lysozyme species. Nevertheless this observation might not be true for other protein species and should be evaluated as a potential marker for protein folding by measurement of spectral variety in refolding samples. Thus structural integrity was assayed by an automated enzymatic assay.

In respect to the actual screening strategy it has been shown that one of the most crucial aspects lies in the controlled and reproducible preparation of the unfolded sample solutions. In addition storage conditions and holding times on the HTS platform have to be kept constant. The denatured state of the starting protein material was investigated by tryptophan emission

measurements and was identified as a parameter tremendously influencing screening results. For a better comparability of refolding yields and parameter optima, unfolding conditions should therefore be kept constant and protein concentration, unfolding time, temperature and buffer compositions should be mentioned.

As suitable refolding parameter values were similar for different feedstock qualities results can be transferred even if denaturing conditions change.

Finally it became clear that the approach using air as redox controlling system does not lead to meaningful results. For systems with no controlled redox environment system parameters leading to improved yields of soluble lysozyme seemed to be unfavorable for the folding to active lysozyme.

In contrast it was observed in screening with addition of a redox system that protein solubility is increased in the same parameter range than protein folding assessed by activity measurements. Consequently this overlap of good parameter values with controlled redox environment qualifies protein solubility as a good objective function during refolding screening. A universal analytical method can be used in a first screening for improved solubility and subsequently a much closer parameter space can be screened for folding using specific assays or sophisticated analytics.

This work thus acts as a fundament for further studies on refold condition screening as it tackles critical system parameters both in respect to the actual screening process but also to potential parameter ranges for lysozyme refolding.

5 References

1.   Lilie, H., E. Schwarz, and R. Rudolph, *Advances in refolding of proteins produced in E-coli.* Current Opinion in Biotechnology, 1998. **9**(5): p. 497-501.

2.   Tsumoto, K., M, *Practical considerations in refolding proteins from inclusion bodies.* Protein expression and purification, 2003. **28**(1): p. 1-8.

3.   Clark, E., M, *Protein refolding for industrial processes.* Current opinion in biotechnology, 2001. **12**(2): p. 202-207.

4.   Rinas, U. and L. Vallejo, *Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins.* Microbial cell factories, 2004. **3**(1): p. 11.

5.   Middelberg, A.P.J., *Preparative protein refolding.* Trends in Biotechnology, 2002. **20**(10): p. 437-443.

6.   Cowieson, N.P., B. Wensley, P. Listwan, D.A. Hume, B. Kobe, and J.L. Martin, *An automatable screen for the rapid identification of proteins amenable to refolding.* Proteomics, 2006. **6**(6): p. 1750-1757.

7.   Lin, L., J. Seehra, and M.L. Stahl, *High-throughput identification of refolding conditions for LXR beta without a functional assay.* Protein Expression and Purification, 2006. **47**(2): p. 355-366.

8.   Ejima, D., K. Ono, K. Tsumoto, T. Arakawa, and Y. Eto, *A novel "reverse screening" to identify refolding additives for activin-A.* Protein Expression and Purification, 2006. **47**(1): p. 45-51.

9.   Wang, W., *Protein aggregation and its inhibition in biopharmaceutics.* International journal of pharmaceutics, 2005. **289**(1-2): p. 1-30.

10.   Armstrong, N., A. De Lencastre, and E. Gouaux, *A new protein folding screen: Application to the ligand binding domains of a glutamate and kainate receptor and to lysozyme and carbonic anhydrase.* Protein Science, 1999. **8**(7): p. 1475-1483.

11.   Chen, T.M., H. Shen, and C.Y. Zhu, *Evaluation of a method for high throughput solubility determination using a multi-wavelength UV plate reader.* Combinatorial Chemistry & High Throughput Screening, 2002. **5**(7): p. 575-581.

12.   Scheich, C., F.H. Niesen, R. Seckler, and K. Bussow, *An automated in vitro protein folding screen applied to a human dynactin subunit.* Protein Science, 2004. **13**(2): p. 370-380.

13.   Ahn, J.H., Y.P. Lee, and J.S. Rhee, *Investigation of refolding condition for Pseudomonas fluorescens lipase by response surface methodology.* Journal of Biotechnology, 1997. **54**(3): p. 151-160.

14.   Sijwali, P.S., L.S. Brinen, and P.J. Rosenthal, *Systematic optimization of expression and refolding of the Plasmodium falciparum cysteine protease falcipain-2.* Protein Expression and Purification, 2001. **22**(1): p. 128-134.

15. Willis, M.S., J.K. Hogan, P. Prabhakar, X. Liu, K. Tsai, Y.Y. Wei, and T. Fox, *Investigation of protein refolding using a fractional factorial screen: A study of reagent effects and interactions.* Protein Science, 2005. **14**(7): p. 1818-1826.

16. Lange, C., G. Patil, and R. Rudolph, *Ionic liquids as refolding additives: N '-alkyl and N '-(omega-hydroxyalkyl) N-methylimidazolium chlorides.* Protein Science, 2005. **14**(10): p. 2693-2701.

17. Jones, D.B., M.H. Hutchinson, and A.P.J. Middelberg, *Screening protein refolding using surface plasmon resonance.* Proteomics, 2004. **4**(4): p. 1007-1013.

18. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Analytical Biochemistry, 1976. **7**(72): p. 248-254.

19. Sapan, C.V., R.L. Lundblad, and N.C. Price, *Colorimetric protein assay techniques.* Biotechnology and Applied Biochemistry, 1999. **29**: p. 99-108.

20. Schein, C.H., *Solubility as a function of protein-structure and solvent components* Bio-Technology, 1990. **8**(4): p. 308-315.

21. Bowen, R., *Understanding flux patterns in membrane processing for protein solutions and suspensions.* Trends in biotechnology, 1993. **11**(11): p. 451-460.

22. Vink, M., K. Derr, J. Love, D.L. Stokes, and T. Ubarretxena-Belandia, *A high-throughput strategy to screen 2D crystallization trials of membrane proteins.* Journal of Structural Biology, 2007. **160**(3): p. 295-304.

23. Chen, G.Q. and E. Gouaux, *Overexpression of a glutamate receptor (GluR2) ligand binding domain in Escherichia coli: Application of a novel protein folding screen.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(25): p. 13431-13436.

24. Heiring, C. and Y.A. Muller, *Folding screening assayed by proteolysis: application to various cystine deletion mutants of vascular endothelial growth factor.* Protein Engineering, 2001. **14**(3): p. 183-188.

25. Boschetti, E., *Advanced sorbents for preparative protein separation purposes.* Journal of chromatography, 1994. **658**(2): p. 207-236.

26. Wiendahl, M., P.S. Wierling, J. Nielsen, D.F. Christensen, J. Krarup, A. Staby, and J. Hubbuch, *High throughput screening for the design and optimization of chromatographic processes - Miniaturization, automation and parallelization of breakthrough and elution studies.* Chemical Engineering & Technology, 2008. **31**(6): p. 893-903.

27. Hevehan, D.L. and E.D. Clark, *Oxidative renaturation of lysozyme at high concentrations.* Biotechnology and Bioengineering, 1997. **54**(3): p. 221-230.

28. Ho, G.S.J., A.P.J. Middelberg, P. Ramage, and H.P. Kocher, *The likelihood of aggregation during protein renaturation can be assessed using the second virial coefficient.* Protein Science, 2003. **12**: p. 708-716.

29. Vincentelli, R., S. Canaan, V. Campanacci, C. Valencia, D. Maurin, F. Frassinetti, L. Scappucini-Calvo, Y. Bourne, C. Cambillau, and C. Bignon, *High-throughput automated refolding screening of inclusion bodies.* Protein Science, 2004. **13**(10): p. 2782-2792.

30. Jolles, P., *Preparation and assay of enzymes.* Methods in Enzymology, 1962. **5**: p. 12-13.

31. Lee, Y.C. and D. Yang, *Determination of lysozyme activities in a microplate format.* Analytical Biochemistry, 2002. **310**(2): p. 223-224.

32. Desai, A., C. Lee, L. Sharma, and A. Sharma, *Lysozyme refolding with cyclodextrins: structure-activity relationship.* Biochimie, 2006. **88**(10): p. 1435-1445.

33. Itzhaki, L.S., P.A. Evans, C.M. Dobson, and S.E. Radford, *Tertiary interactions in the folding pathway of hen lysozyme - kinetic studies using fluorescent probes.* Biochemistry, 1994. **33**(17): p. 5212-5220.

34. Lin, J.L., R.C. Ruaan, and H.J. Hsieh, *Refolding of partially and fully denatured lysozymes.* Biotechnology Letters, 2007. **29**(5): p. 723-729.

35. Cowgill, R.W., *Fluorescence and Protein Structure .11. Fluorescence Quenching by Disulfide and Sulfhydryl Groups.* Biochimica Et Biophysica Acta, 1967. **140**(1): p. 37-45.

36.   Dill, K.A., *Dominant Forces in Protein Folding.* Biochemistry, 1990. **29**(31): p. 7133-7155.

37.   Yutani, K., A. Yutani, A. Imanishi, and T. Isemura, *Mechanism of Refolding of Reduced Random Coil Form of Lysozyme.* Journal of Biochemistry, 1968. **64**(4): p. 449-455.

38.   Buswell, A.M. and A.P.J. Middelberg, *Critical analysis of lysozyme refolding kinetics.* Biotechnology Progress, 2002. **18**(3): p. 470-475.

39.   Creighton, T.E., *Protein folding.* Biochemical Journal, 1990. **270**(1): p. 1-16.

40.   Raman, B., T. Ramakrishna, and C.M. Rao, *Refolding of denatured and denatured/reduced lysozyme at high concentrations.* Journal of Biological Chemistry, 1996. **271**(29): p. 17067-17072.

41.   Clark, E.D., D. Hevehan, S. Szela, and J. Maachupalli-Reddy, *Oxidative renaturation of hen egg-white lysozyme. Folding vs aggregation.* Biotechnology Progress, 1998. **14**(1): p. 47-54.

42.   Fischer, B.E., *Renaturation of recombinant proteins produced as inclusion bodies* Biotechnology Advances, 1994. **12**(1): p. 89-101.

43.   Hofmeister, F., *Zur Lehre von der Wirkung der Salze.* Arch. Exp.Pathol.Pharmakol., 1888. **24**: p. 1 - 16.

44.   Qoronfleh, M.W., L.K. Hesterberg, and M.B. Seefeldt, *Confronting high-throughput protein refolding using high pressure and solution screens.* Protein Expression and Purification, 2007. **55**(2): p. 209-224.

Automated Optimization of Protein Refolding Processes with an Evolutionary Algorithm

Annette Berg, Anna Siudak, Eric von Lieres, Juergen Hubbuch*

Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Science, University of Karlsruhe (TH), 76131 Karlsruhe, Germany

*Corresponding author. Tel.: +049 721 608-2557; fax: +049 721 608-6240. Email-adress: juergen.hubbuch@kit.edu

Abstract

Inclusion body protein refolding is an optimization task including a high number of potentially interdependent variables like buffer pH, salt types and ionic strength, type of redox components and their concentrations and ratio and numerous additives used to improve solubility and folding. As no model is known to describe parameter effects on the refolding yield a high number of experiments has to be performed during process development driving up costs. We show in our study that an evolutionary algorithm can successfully be used to optimize five refolding buffer parameters: pH, concentration of NaCl and $MgSO_4$ and of oxidizing and reducing components at the same time. Concentration of soluble and active lysozyme could thus be improved on a more robust data basis than provided by fractional factorial screening designs. We were furthermore able to define the correlation of parameter optima for lysozyme solubility and activity by analysing the development of both functions during convergence of parameters with one function as "driving force" for the genetic algorithm. In the investigated parameter matrix the conditions for high yields of active lysozyme are restricted to a much closer parameter range than for high yields of soluble lysozyme but show a complete overlap.

Restriction of the parameter range of interest by the evolutionary algorithm leads to the possibility to perform full factorial experiments within the optimum with reasonable experimental effort. With results from full factorial experiments the predicted optima on basis of data from the genetic algorithm optimization could be confirmed and refined. Additionally the performance of five replicates of the full factorial experiments was used to estimate the experimental error and its distribution in the design space. We found an average relative standard deviation of 15 % with an equal distribution in the investigated parameter range. As the concentrations of active lysozyme with different parameter values even in the case of 15 % error show differences of 28 % of the highest compared to the lowest concentration of active lysozyme parameter effects are identified to be significant for the lysozyme refolding process.

1 Introduction

The use of so called platform processes to express and especially purify biologics is the current aim when developing novel production processes. The most prominent example is the expression and purification of monoclonal antibodies where the overwhelming majority of purification processes is based on a sequence of two to three chromatographic steps – Protein A affinity chromatography followed by ion exchange and hydrophobic interaction chromatography or hydroxy apatite chromatography [1-3]. For a long time the production of biologics via inclusion body formation in Escherichia coli has also been considered to act as a novel platform process. The major drawback, however, lies in finding the optimal refolding conditions to reach high yields of correctly refolded protein in the first step. A number of systems to aid this has been developed. Their realization, however, is mostly connected with considerably higher costs [4-6].

While most refolding processes base on the experience of the respective biochemical engineer being responsible for the development, screening tools to aid the development are only considered within the last couple of years [7-16]. The major drawback of these systems is, however, the lack of appropriate analytics to cope with the number of samples and the failure to adequately control all relevant screening process parameters [17]. In addition to this the number of relevant process parameters is considerably high. When designing refold buffer systems the decision for appropriate buffer pH, types of salts, ionic strength, types and concentration of reducing and oxidizing agents is imperative. Complexity is further increased by combinatorial effects of parameters on refolding success. A material and time effective strategy for process development is thus one of the main requirements in the biopharmaceutical industry. Optimization of a multi parameter system such as protein refold systems can only be performed when combining adequate HTS approaches with experimental design based on mathematical and statistical methodologies in order to gather maximum information with minimum experimental effort.

In order to tackle the experimental challenge we have previously developed a fully automated high throughput method for screening of refolding buffer composition which is described in [18]. Even with automated HTS strategies the number of experiments in a brute force screening approach with a high number of parameters will increase to such an extent that the experimental effort is not manageable when considering time and costs.

Nevertheless a non-systematic brute force screening is performed for example in [19] to screen different additives at one level. No information on the effects of additive concentration

or synergistic effects of additives is gained. A similar approach is presented in [9] where in a first screening aggregation suppression of different detergents in refolding was estimated and in a second screening one favourable detergent was combined with a second additive. This step by step optimization leads to an oversight of potential positive or negative parameter interactions leading to suboptimal screening results.

In contrast to this, full factorial screening approaches offer the possibility to gain systematic insight into parameter effects and interactions but require a high number of experiments as described by equation 1 with e as number of experiments, p the number of parameters and L the .number of parameter levels.

$$e = L^p \tag{1}$$

In Figure 1 this functional relation between the number of experiments and the number of parameters is visualized for two parameter levels.
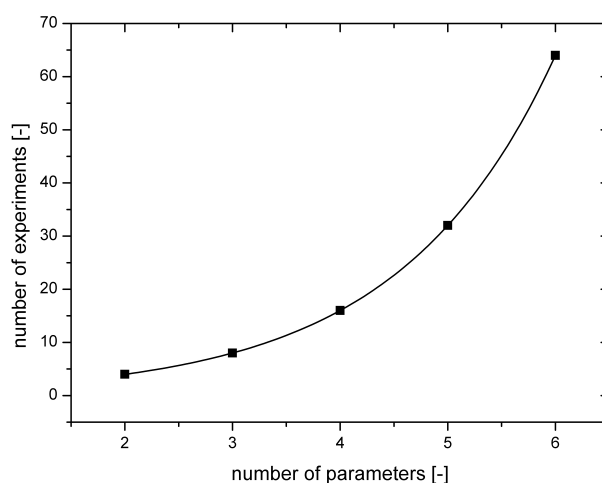

Figure 1: Number of experiments vs. number of parameters at two levels

This said, the number of experiments can be reduced when performing fractional factorial screenings. The use of such approaches which in the first instance provide an experimental plan is however not always fully exploited reducing its informational content to a minimum. In [11, 14, 20] for example data from fractional factorial experiments were not analyzed by statistical means. Insight into favourable refolding conditions is gained by a rather intuitive comparison of refolding results and a qualitative interpretation of positive and negative parameter effects. The best experiments from the screening were then directly chosen for refolding. Consequently the statistical approach is harnessed for the generation of a reduced experimental set up but not for data validation. In [8, 15] data from fractional factorial refolding screening were statistically analyzed to identify main parameter effects increasing

process understanding. Following the identification of important refolding parameters a second full factorial screening is performed for only two parameters to identify parameter interactions. An optimization on refolding buffer systems by a systematic fit of the screening results is not performed and thus the best experiment from the screening might still be far of the global optimum.

Furthermore all of these statistical evaluations are based on the underlying assumption that the objective function is linearly dependent on the parameters. This, however, is a simplification which is likely to lead to a suboptimal screening result. The latter especially holds true when considering a large parameter space. This is clearly shown by a study of [7, 13] where a response surface is used to predict optimum refolding parameters. The optimization of a matrix comprising five factors was shown to be possible leading to a good refolding performance verified by additional experiments. This, however, is only the front page of the complete effort. Parameter effecting refolding yield and an appropriate parameter space for improvement of surface fit with the chosen model were already investigated experimentally before starting experimental design for the response surface fit. Thus the presented results only represent part of the experimental effort required for optimization with response surface analysis. In [7] it was further mentioned that parameter space was reduced in previous experiments to a range close to an expected optimum. This approach was chosen to increase the likelihood of a realistic response surface fit covering the main optimum avoiding a complex surface with multiple optima. Furthermore data density can be chosen high enough to be less dependent on experimental error which is usually a problem if sparse data points are used for response surface fit.

We apply an evolutionary algorithm to improve the robustness of the optimization process and to make simultaneous optimization of five parameters in a wide parameter space possible despite potential parameter interactions or non-linear dependencies of the objective function. The genetic algorithm chosen for this study is a population based algorithm thus harnessing the advantage of the parallelized experimental screening method. Its high robustness is based on an iterative search mechanism with repetitive measurements in close proximity. Nevertheless the genetic algorithm is highly efficient because data density is increasing close to the optimum while experiments in less interesting regions with low refolding yield are reduced to a minimum.

During improvement of an objective function with the genetic algorithm aiming at lysozyme activity, a second function, in our case lysozyme solubility, was measured and analysed. Thus a correlation of parameter effects on both measured function values can be observed. A

correlation of parameter optima offers the potential to use the function measured with higher accuracy or lower experimental effort for future optimization experiments.

After optimization with a genetic algorithm a factorial screening was performed close to the optimum. In our study factorial experiments are used for estimation of experimental errors and to control the result of the evolutionary optimization in the optimum parameter range. With this combination of the genetic algorithm to screen a wide parameter range followed by a classical full factorial approach in a reduced design space close to the optimum we reach a good balance between experimental input and process optimization and understanding.

2 Materials and Methods

2.1 Automated Liquid Handling Platform

For all studies the automated liquid handling platform Freedom Evo 200 (Tecan Crailsheim, Germany) equipped with one liquid handling arm and two grippers was used. Pipetting was performed with 8 fixed standard tips. A centrifuge Rotanta RSC46 (Hettich Kirchenlengern, Germany), a magnetic orbital shaker with four positions and a shaking diameter of 2 mm (Inheco Munich, Germany) and a spectrophotometer InfiniTe 200 (Tecan Crailsheim, Germany) were integrated into the robotic platform. The software Evoware (Tecan Crailsheim, Germany) was used to control the robotic work station.

96 well microtiter plates and UV microtiter plates with a well volume of 360 µl were purchased from Greiner (Frickenhausen, Germany). Deep well plates with a well volume of 2.2 ml were from TreffLab (Degersheim, Switzerland). 350 µl AcroPrep 96 filter plates with a pore size of 0.2 µm and a Bio-Inert membrane were produced by Pall (Dreieich, Deutschland).

All chemicals had analytical grade and were purchased from Sigma Aldrich (St. Louis, USA) as well as hen egg white lysozyme (L-6876) with ≥ 90 % protein content.

2.2 Sample Preparation and Analytics

Methods for the preparation of refolding samples and measurement of soluble protein content and lysozyme activity on the robotic platform are described in detail in [18].

Protein denaturation was performed according to the following protocol. 4 mg/ml lysoyme was incubated for 1.5 h in 50 mM potassium phosphate buffer containing 8 M urea and 5 mM DTT at pH 8 at 37 °C.

Buffer systems for protein refolding were prepared by mixing Milli-Q water with buffer stock solutions (glycine at pH 3, sodium acetate at pH 4 and pH 5, potassium phosphate at pH 6, pH 7 and pH 8, bicine at pH 9) to a buffer concentration of 50 mM. In the given order NaCl, $MgSO_4$, GSSG and DTT were added. A constant protein dilution factor of 10 in a total volume of 1 ml was maintained. Protein refolding was initiated by adding 0.4 mg/ml denatured protein into the respective refolding buffer systems followed by mixing the solution by aspirating and re-dispensing of 900 µl sample volume. The refolding process was terminated after an incubation period of 1 h at 25 °C. Mixing conditions were maintained throughout the incubation period by constant orbital shaking at 1500 rpm. Aggregated and soluble protein content were analyzed by the separation of protein aggregates using filtration, their re-solubilization and subsequent protein determination using UV 280 measurements. Concentration of active lysozyme was measured via an automated lysozyme assay.

2.3 Experimental Parameter Space

The experimental matrix investigated in this study included five parameters: concentration of NaCl, $MgSO_4$, GSSG and DTT and buffer pH (see Table 1).

Table 1: Overview of the parameter design space for the genetic algorithm

Parameter	Design Space
Buffer pH	3 - 9, discrete, step size of 1
NaCl concentration	0 - 150 mM
$MgSO_4$ concentration	0 - 150 mM
GSSG concentration	0 - 20 mM
DTT concentration	0 - 10 mM

2.3 Genetic Algorithm

The process optimization method applied in this study had been described in detail by Susanto et al. [21] for the optimization of chromatographic protein separation.

An overview of the framework behind the optimization procedure using the genetic algorithm as a control tool for high throughput experimentation and the respective software tools is given in Figure 2. For the automated optimization circle three software tools were combined: Evoware (Tecan, Crailsheim, Germany) for control of the liquid handling station, MATLAB (MathWorks, Natick, MA, USA) for data management and the included GEATbx v1.95 [22] for performance of the genetic optimization algorithm and Excel (Microsoft, Redmond, WA, USA) as an easy user interface connecting Evoware and MATLAB. Excel offered the possibility to observe the produced data coming from automated experiments and also from the genetic algorithm optimization tool online and allowed to take corrective actions during the process. Furthermore the Excel interface was a prerequisite for the communication between MATLAB and Evoware and acted as a user interface where the parameters affecting the performance of the genetic algorithm such as the number of experiments, the selection method and pressure, the method of recombination and the rate of mutation could be set.
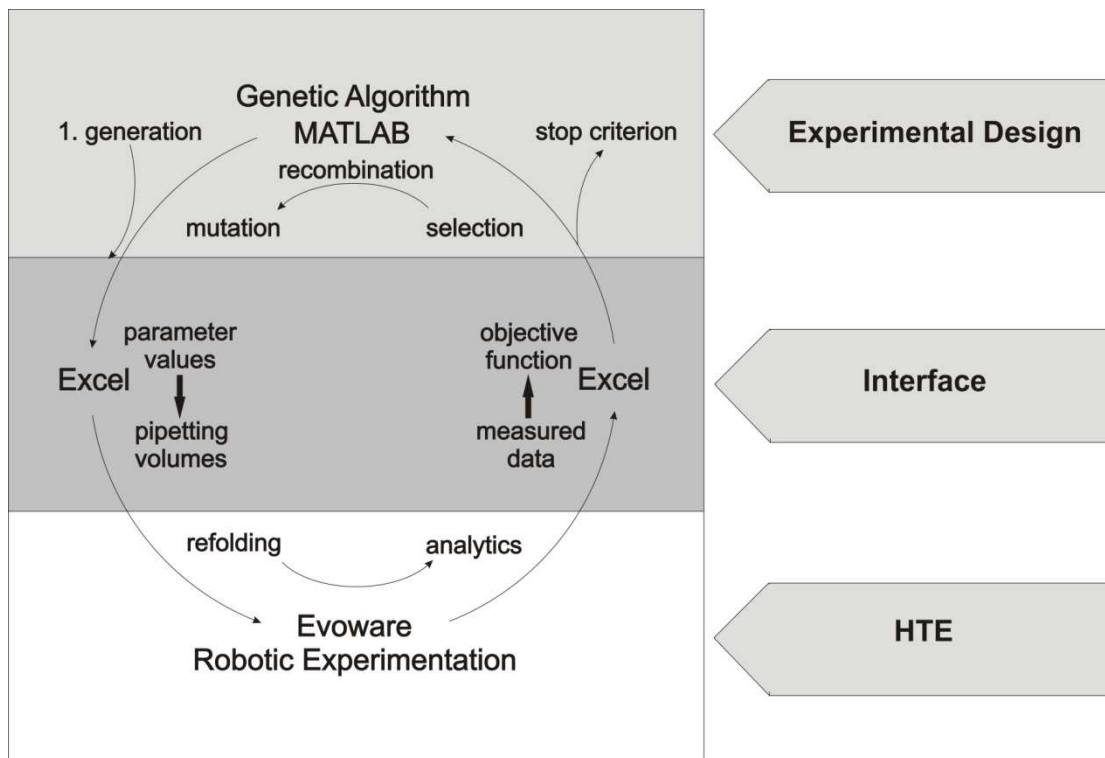
Figure 2: Overview of the evolutionary optimization cycle

The genetic algorithm was initialized by a first generation of experiments with randomly distributed experimental parameter values.

Two different objective functions, the concentration of soluble lysozyme and the concentration of active lysozyme in the refold samples were applied.

The performance of the genetic algorithm is dependent on algorithm parameters which include the number of experiments in a population, the selection pressure and the mutation rate. For our studies we chose a population size of 40 experiments, a selection pressure of 11.4 and a mutation rate of 0.2 [22, 23]. The optimization cycle was stopped when no further convergence of experimental parameter values could be observed represented by a constant or increasing mean standard deviation of the parameter values for the next generation.

2.4 Full Factorials

A full factorial comprising 88 experiments close to the optimum determined by the genetic algorithm was performed. For all experiments we kept the pH value constant at pH 9 as this pH showed to be invariably most favourable for refolding. The chosen parameter matrix is shown in Table 2. The full factorial experiments were performed five times on different days

with all solutions freshly prepared. Generation of the experimental matrix as well as statistic validations were performed with the Software Modde (Umetrics, Malmö, Sweden).

Table 2: Overview of the parameter design matrix for the full factorial experiments

Parameter	Design Space
Buffer pH	9
NaCl concentration	115 mM, 135 mM, 155 mM
MgSO$_4$ concentration	20 mM, 25 mM, 30 mM
GSSG concentration	15 mM, 16.5 mM, 18 mM
DTT concentration	4 mM, 5 mM, 6 mM

3 Results and Discussion

In a first screening with randomly distributed parameter settings in the applied experimental matrix we observed an overlap of parameter values between samples with a high content of active lysozyme and a high content of soluble lysozyme given that the redox environment was controlled by addition of reducing and oxidizing components [18]. Even though such a result was expected the observation gave rise to a more structured investigation into the interdependency of both objective functions.

3.1 Genetic Algorithm

The genetic algorithm was applied twice for the above described system with two independent objective functions aiming at the maximization of solubility and activity respectively. The aim of this investigation was centred around the question on data analysis as well as size, ease of reach and interdependency of the global optimum regions of solubility and activity.

3.1.1 Objective Function in Protein Refolding Optimization

When considering protein refolding processes one immediately highlights the necessity of correctly refolded protein, soluble protein and the specific activity of the obtained protein. While the maximum concentration of correctly refolded protein clearly defines the yield of the refolding process and should thus have highest priority, maximizing the specific activity

might gain importance if incorrectly refolded protein is difficult to separate from correctly refolded protein and thus leads to a high loss of correctly folded protein in subsequent steps and an overall lower yield. When put into an order with the other two functions screening for solubility might mislead the optimization procedure, especially when the region for maximum solubility and activity do not coincide (see below). However, solubility screens – if applicable for the respective system – offer a rapid tool for a first screen as the accompanied assay is rapid and HTS compatible. This absolute necessary prerequisite for a screen with a high number of samples is not given for most structural or activity related assays.

3.1.2 Potential Interdependency of Solubility and Activity

In the light of the necessity of an adequate analytical scheme for a first screen one needs to be aware of the potential settings when using an indirect screening objective such as solubility. Figure 3 depicts a scheme of potential settings of both regions – activity and solubility – within the investigated parameter space. At first sight one might expect these two values to coincide with the region for activity smaller than solubility accounting for soluble but incorrectly folded protein (Figure 3 A). However, in protein refolding it is often seen that incorrectly folded and thus inactive species contribute to soluble protein content to a much higher degree than correctly and thus active protein. The most unlikely event in this context might be that the amount of incorrectly folded protein leads to a solubility peak within the optimal region for active protein (Figure 3 B). More likely scenarios are that these regions overlap (Figure 3 C) or lead to a complete separation of the global optima (Figure 3 D).
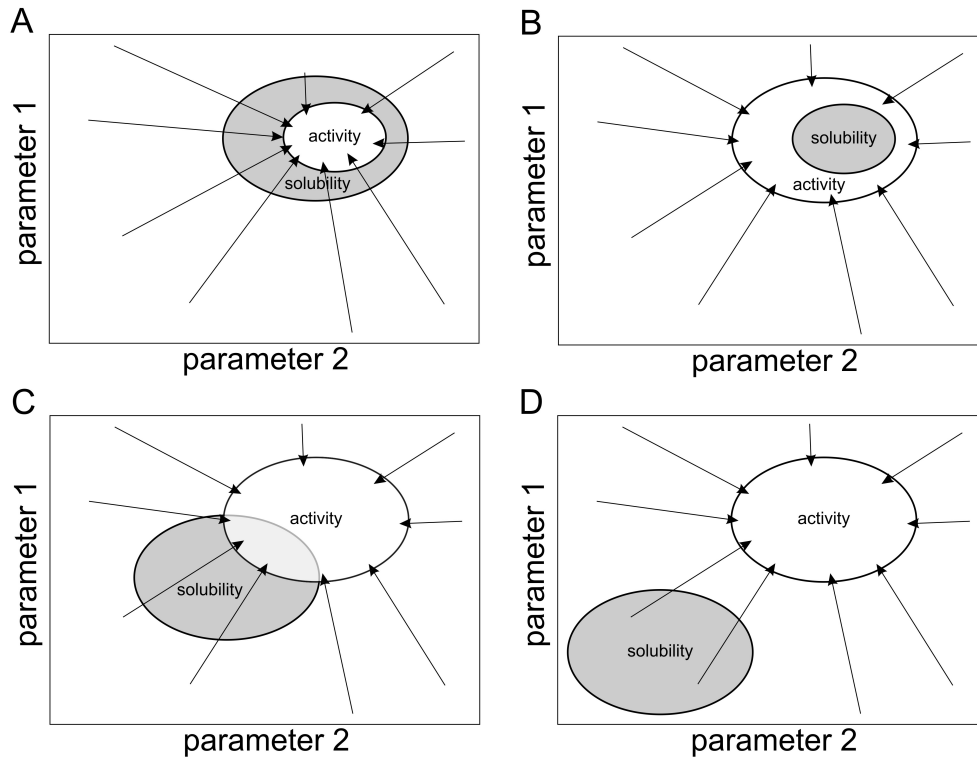Data obtained by the genetic algorithm can now be analysed by various approaches.

Figure 3:.Four alternatives for the positioning of parameter optima in a two parameter space for two potentially interdependent objective functions the concentration of active protein (activity) and the concentration of soluble protein (solubility) in refold systems. **A** both optima coincide with a smaller optimum region for activity **B** both optima coincide with a smaller optimum region for solubility **C** both optima partially overlap **D** optima are completely separated

3.1.1 Development of the Objective Function in Time

Given that the event shown in Figure 3 A is by far more likely than that shown in Figure 3 B most information will be gained when analysing a screening procedure aiming at maximum activity. The development of the optimization performance over the respective generations and the interdependency of activity and solubility are shown in Figure 4. Figure 4 A summarizes the average values of active lysozyme concentration and the corresponding values for soluble lysozyme concentration over five generations of an optimization procedure with the objective function aiming at maximum activity. The average specific activity or the fraction of active to soluble lysozyme are calculated and depicted in Figure 4 B for all generations. Concentrations of active and soluble lysozyme both increase in subsequent generations. The average active lysozyme concentration obtained changed with a factor of 8.3 from 0.019 mg/ml to 0.154 mg/ml. The concentration of soluble lysozyme improved with a

factor of 2 from 0.195 mg/ml to 0.384 mg/ml including samples with maximum solubility of 0.4 mg/ml. From generation 2 to generation 3 the concentration of active lysozyme stayed constant whereas the concentration of soluble lysozyme showed a slight increase. Subsequently a slight decrease in specific activity was observed. In this stage of the optimization procedure the experimental points used are probably situated on a plateau regarding protein activity while values for solubility experience a slope upwards. Subsequently this scheme is reversed showing a steady increase of protein activity while protein solubility reaches a plateau expressing maximum solubility in generation 4 and 5. The most plausible explanation for this behaviour lies in a scheme depicted in Figure 3 A where the protein activity peak is situated in a plateau describing maximum protein solubility.
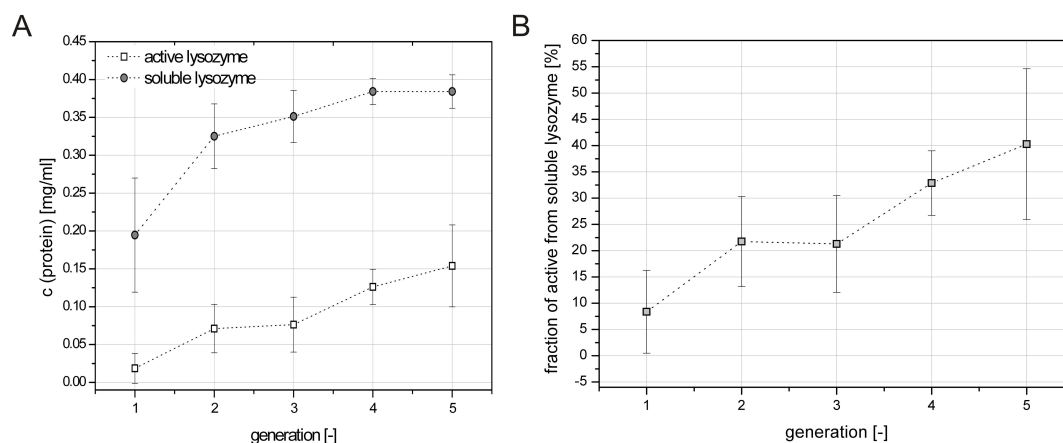

Figure 4: Average results of all genetic algorithm generations from an optimization aiming at maximum lysozyme activity showing **A** the development of active and soluble lysozyme concentration and **B** the fraction of active to soluble lysozyme concentration

A more detailed picture is obtained when comparing the outcome of both optimization procedures next to each other. In Figure 5 the concentration of soluble protein is plotted over the concentration of active protein. Figure 5 A, C represent data from the systems optimized towards maximum activity and Figure 5 B, D represent the corresponding data from systems obtained when optimizing towards maximum solubility.
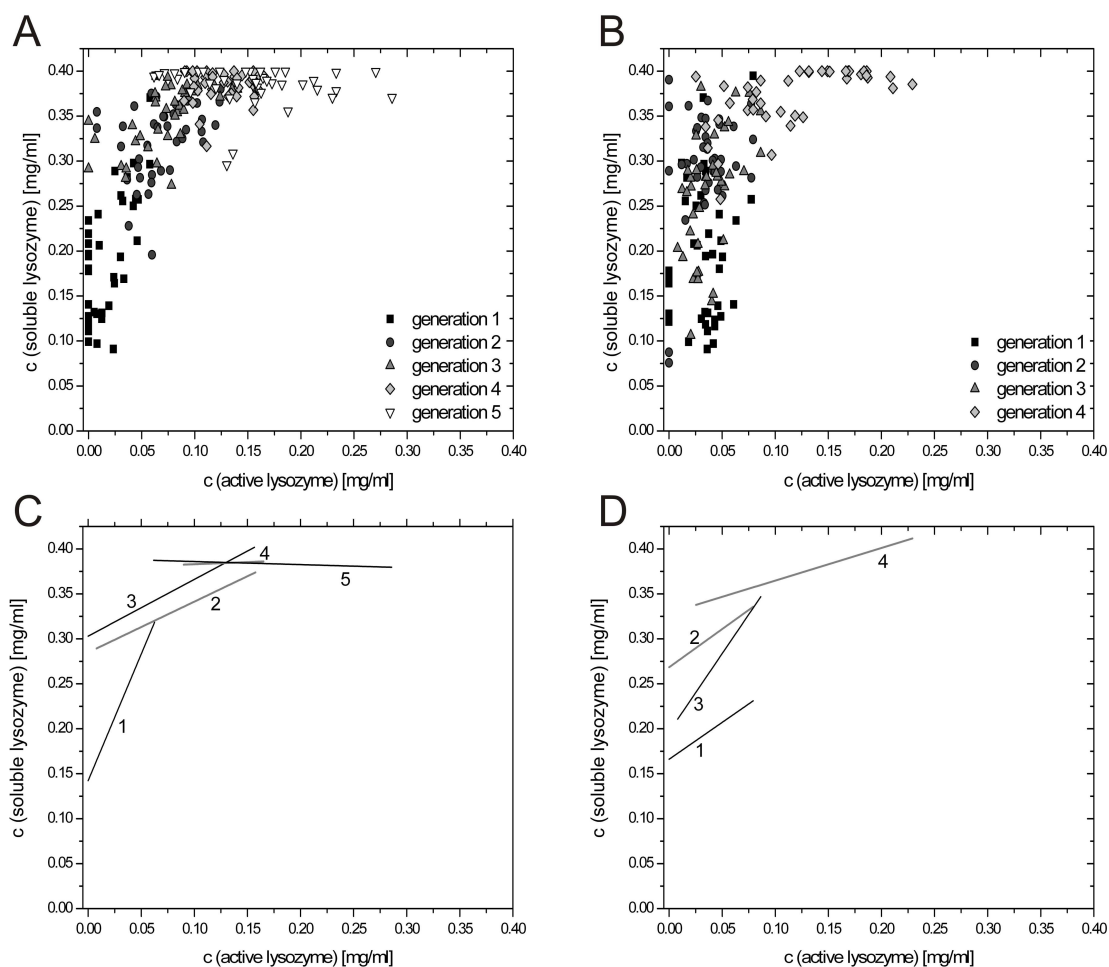
Figure 5: Comparison of experimental results from all generations of a genetic algorithm optimizations aiming at maximum active lysozyme concentration versus aiming at maximum soluble lysozyme concentration. In A concentrations of active and soluble lysozyme are depicted with active lysozyme concentration as objective function. C shows the corresponding linear fit for all generations for a better orientation. In B concentrations of active and soluble lysozyme are depicted with soluble lysozyme concentration as objective function. D shows the corresponding linear fit for all generations for a better orientation.

Based on two different initial populations the first generations show a slight variation expressing a broader solubility range for system 5 A while system 5 B shows a broader distribution for protein activity. The 'better' distribution of experimental points in the first generation concerning protein solubility might be responsible for a more rapid approach of the plateau region expressing maximum solubility, even though the objective function aimed at maximum activity. The rapid optimization towards maximum solubility in both systems and further improvement of protein activity in generation 5 indicates a setting according to Figure 3 A. Both systems describe a stagnant (system 5 A) and decreasing (system 5 B) phase in the

156

third generation indicating a plateau or slight downhill region regarding the investigated parameter. In conclusion it can be stated that for the given system, an initial rough screen towards maximum solubility followed by a more detailed investigation into the obtained solubility plateau region towards high protein activity might be justified.

3.1.2 Objective Function Values over the Experimental Space

Figure 6 shows surface plots describing the dependency of a parameter set onto the objective function. Column one represents data obtained when screening for optimal activity while column two presents the respective data for the solubility screen. For comparability of the two sets, concentration of soluble and active lysozyme, eight contour levels divide the range between the maximum and the minimum z-value achieved in the fit. The final optimum is marked with a grey arrow.

Focussing on the salt levels it becomes obvious that the high yield value regions at low concentrations of $MgSO_4$ and high concentrations of NaCl (Figure 6 E and F). The latter finding indicates that the parameter ranges for NaCl might be too narrow for the optimization of the given system and especially when screening for optimal solubility higher concentrations of NaCl should be investigated. However, both regions are found at approximately the same coordinates. This also holds true for the relationship of GSSG and DTT (Figure 6 A and B). Here GSSG offered a wider spectrum where optimal values are obtained (~ 8mM – 18 mM) while DTT restricted the combinations to a narrow band of ~ 4mM to 6 mM. The optimum pH region (Figure 6 C and D) lies for solubility and activity at the upper border of the analyzed parameter space at pH 9 indicting that the screened parameter space for this parameter should be enlarged.
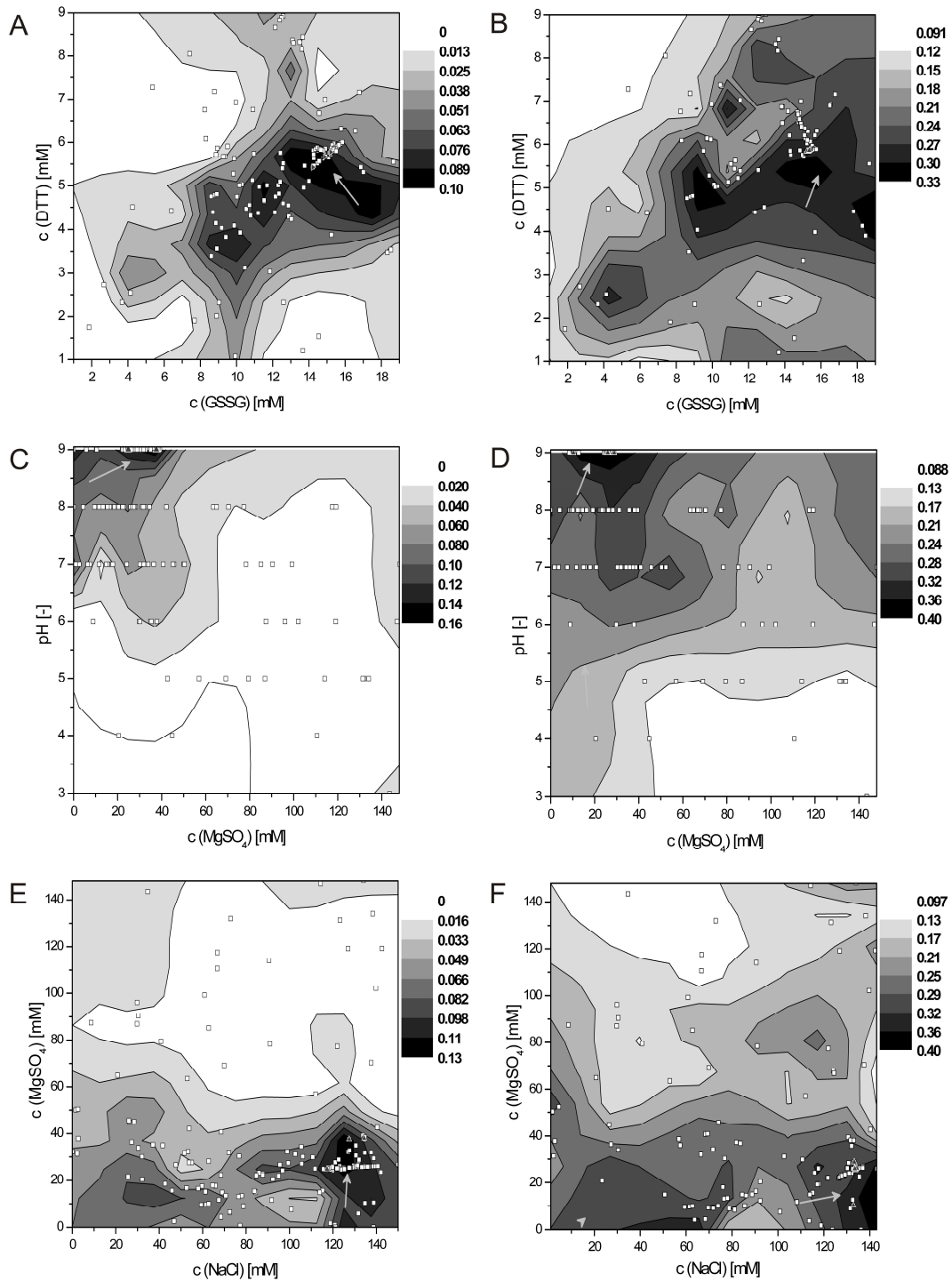
Figure 6: Contour plots for the optimization of lysozyme activity (column on the right: **A, C, E**) and the optimization of lysozyme solubility (column on the left: **B, D, F**) for the parameter combinations concentration of DTT and GSSG, pH and concentration of MgSO$_4$ and concentration of NaCl and MgSO$_4$. The objective function value in mg/ml is depicted as 8 contour levels between the maximum value in black and the minimum value of the fit surface in white. An arrow marks the position of the optimum.

In Figure 7 we compared the frequency of parameter settings in the final generations. The parameters can be sorted into three sets. In the first set describing salt dependency the optimal salt concentration for protein activity is less stringently defined than that for protein solubility where a single concentration was highlighted. This coincides with the finding that the optimal NaCl concentration for solubility lies on the edge of the experimental parameter space suggesting higher values for NaCl when screening for optimum solubility. When considering the redox environment both attributes, optimum solubility and activity are described by a single value namely 15 mM GSSG and 5.5 mM DTT. This is a quite surprising result when compared to the overall findings presented in Figure 6 and highlights the need to plot the overall performance rather than focussing on the final generation only. In the third set describing pH dependence we see a narrow concentration for protein activity (pH 9) and a wider range for solubility (pH 8 – pH 9).This coincides again with data from the overall run suggesting that the optimal pH value found is situated on the edge of the experimental parameter space.
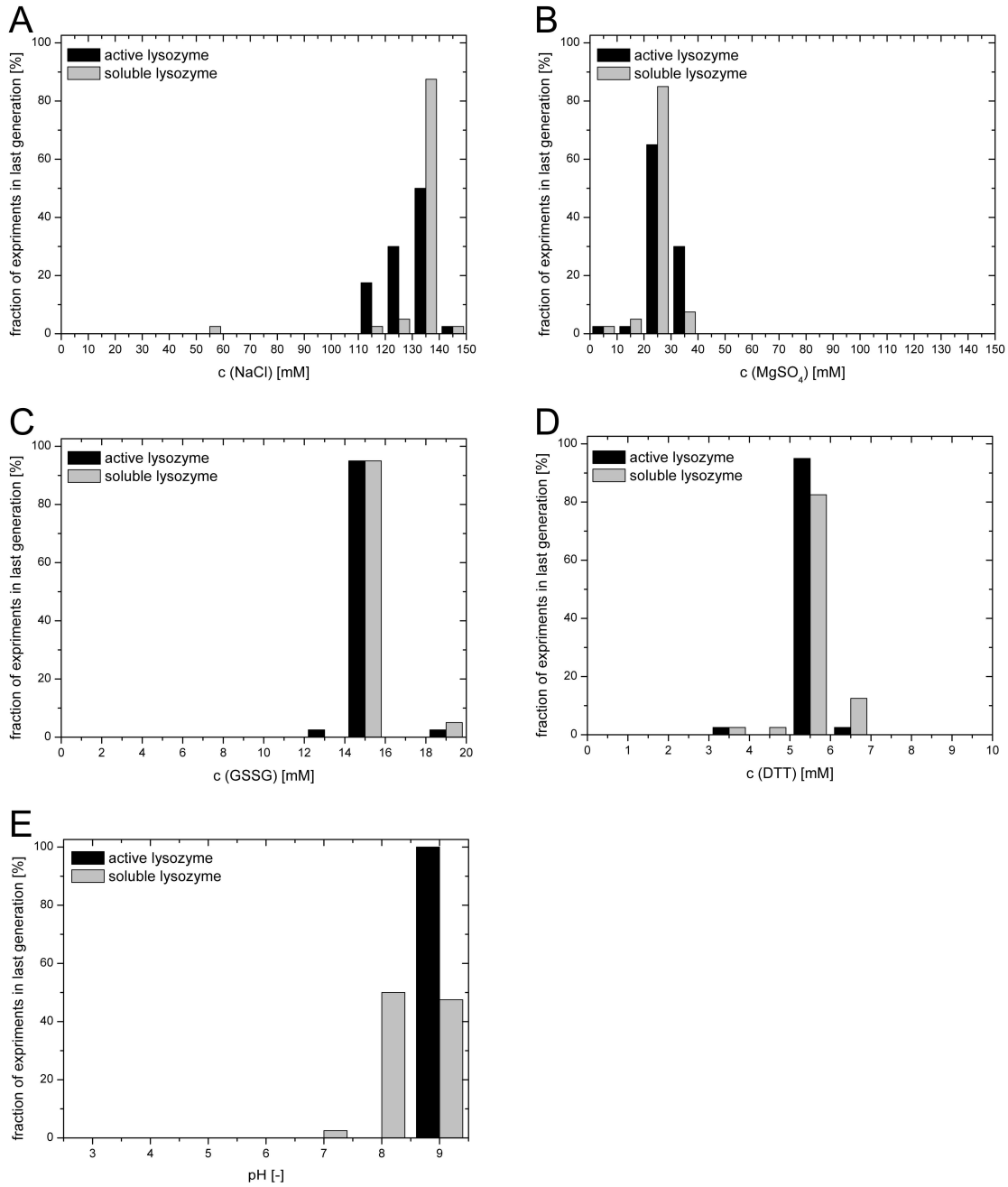
Figure 7: Fraction of experiments performed in the final generation of the genetic algorithm for active lysozyme optimization (black bars) and soluble lysozyme optimization (grey bars) for the parameters: **A** NaCl concentration, **B** MgSO₄ concentration, **C** GSSG concentration, **D** DTT concentration and **E** pH.

3.2 Full Factorial Analysis of the Optimum Region

In a second approach we analysed the optimal parameter regions using a full factorial screen. Five independent full factorial sets with 88 parameter combinations are performed close to the potential optimum for active lysozyme found with the genetic algorithm. Variable parameters were the concentration of $MgSO_4$, NaCl, GSSG and DTT. As the optimal pH was found to be situated on the edge of the experimental space pH 9 was set constant in this procedure. The optimum region of the surface plots shown in Figure 6 are depicted in a zoomed version in Figure 8 A and 9 A. The grey arrows mark the three optimal systems obtained during the GA screen.

For GSSG and DTT the best three results lie on a diagonal line with GSSG to DTT ratios of 2.6 ± 0.006 and between GSSG concentrations of 14 and 16 mM and DTT concentrations of 5 to 6 mM with better yields of active lysozyme at lower GSSG and DTT concentrations (Figure 8 A).

Only few data points are measured in the parameter range above 16 mM GSSG and below 5 mM DTT also resulting in a maximum after the empirical fit of the data. The full factorial covers this parameter range with reliable experimental data from arithmetic means of five replicates. The results of a surface fit including not only the data from the genetic algorithm but also an average of the five measurements of the full factorial experiments are shown in a second contour plot (Figure 8 B). Interestingly the optimum parameter range underdeterimed in the genetic algorithm optimization is confirmed to yield the highest concentrations of active lysozyme. The genetic algorithm seems to miss this part of the parameter matrix due to an optimization along the diagonal line with initially favourable GSSG to DTT ratios due to the extended line recombination method chosen. Higher mutation rates would in this case prevent the genetic algorithm from being caught at parameter values in proximity to but not within the optimum.
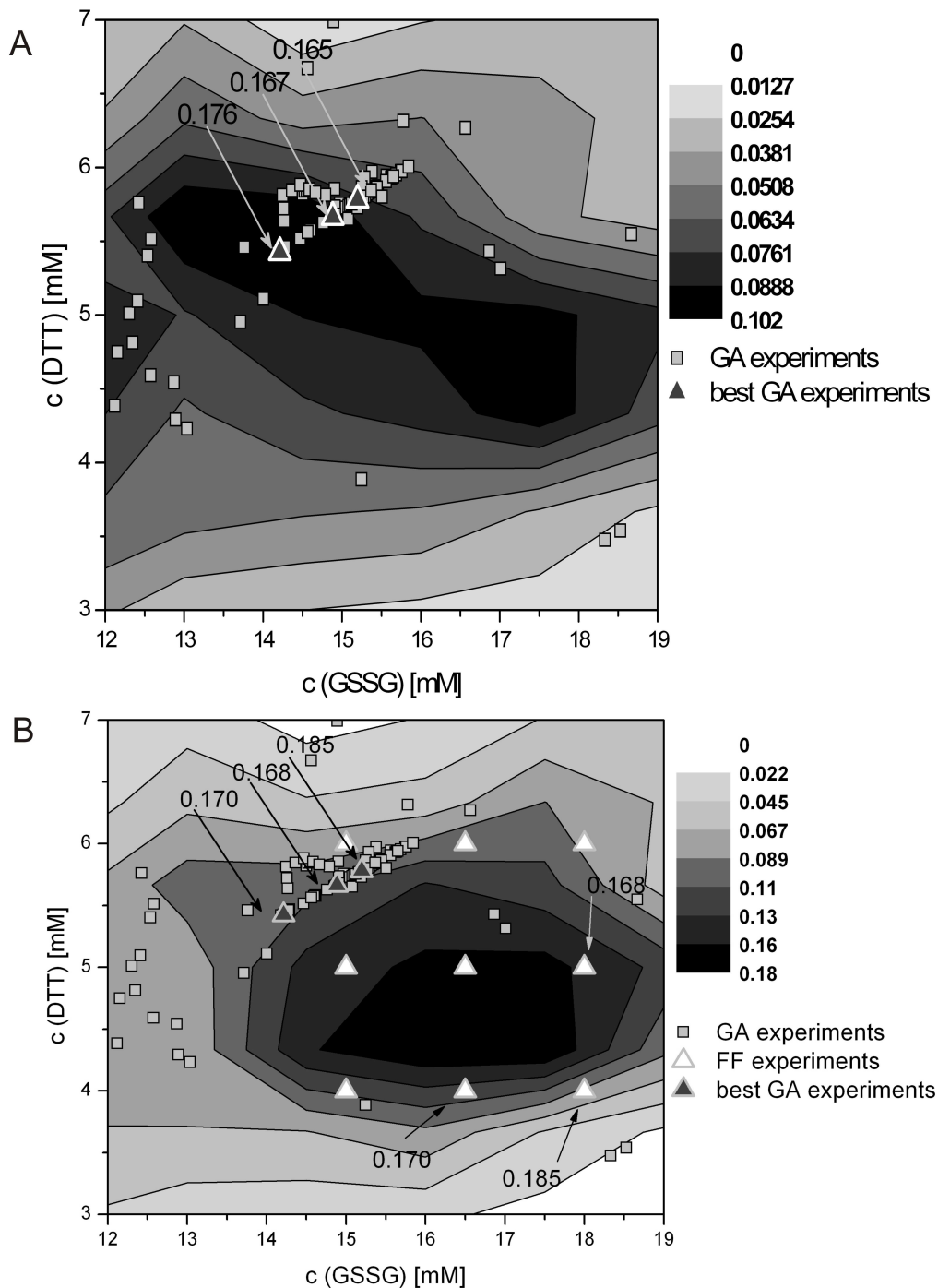
Figure 8: A Zoom plot of the optimum parameter range of GSSG and DTT concentration for the GA optimization of active lysozyme concentration. Contour levels show concentration of active lysozyme in mg/ml. Grey triangles mark the best three concentrations of active lysozyme measured including corresponding concentration values. **B** Zoom plot of optimum parameter range including full factorial experiments marked as white triangles. Best full factorial experiments and best experiments gained with the GA are marked including concentration values.

For the salt concentrations the zoomed surface plot (Figure 9 A) from the genetic algorithm experiments shows the best yields of active lysozyme between 115 and 140 mM NaCl and between 25 and 40 mM $MgSO_4$ with improved yields at the lowest concentrations of both salts.

The parameter range below 25 mM $MgSO_4$ and above 140 mM NaCl is not well validated and therefore included in the full factorial experiments (Figure 9 B). For these two parameters no change in the resulting surface plot can be observed when average full factorial data are included as can be seen in Figure 9 B. An even higher maximum objective function is reached by combination of all parameters close to the optimum, indicating parameter interdependencies.
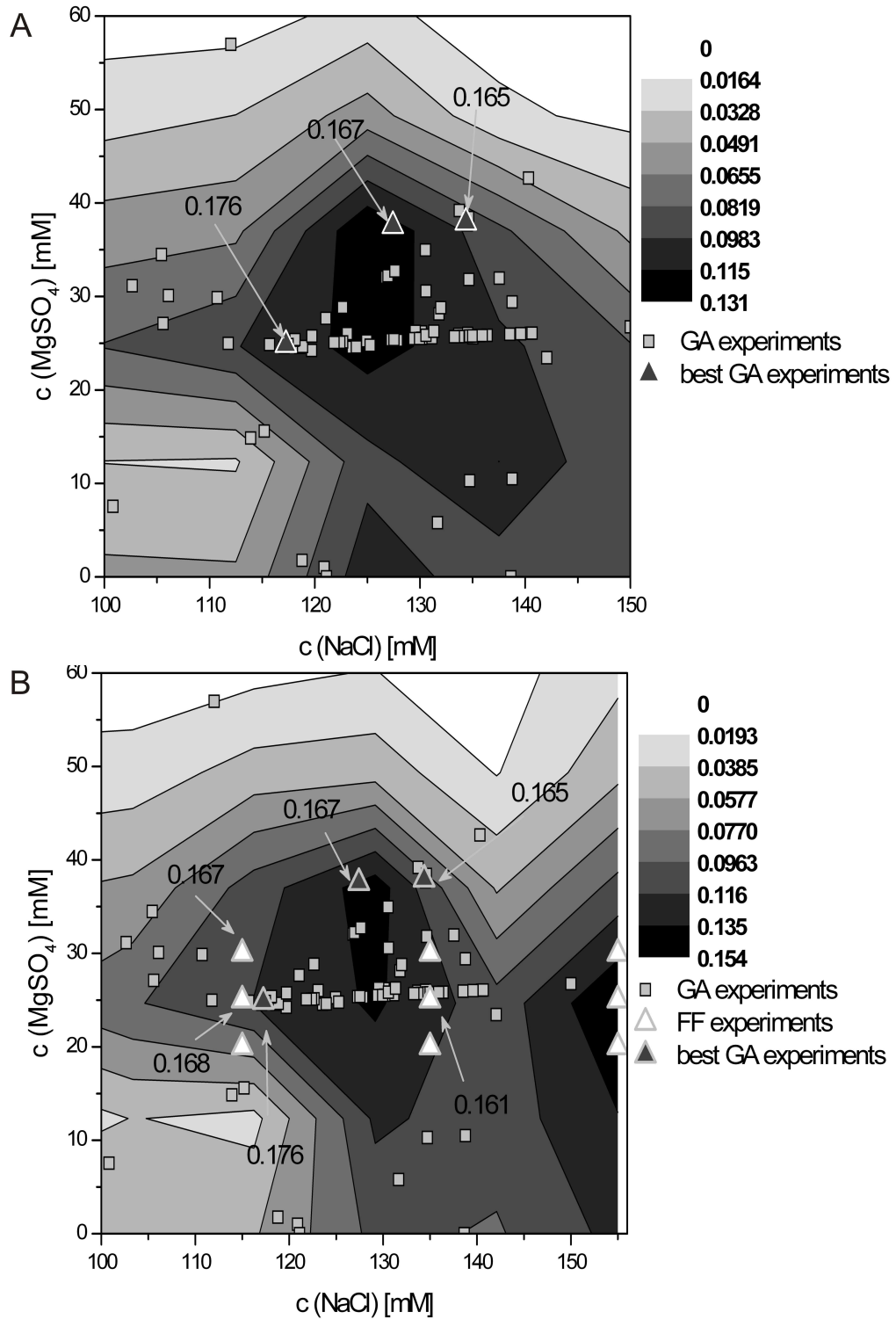
Figure 9: A Zoom plot of the optimum parameter range of NaCl and MgSO₄ concentration for GA optimization of active lysozyme concentration. Contour levels show the concentration of active lysozyme in mg/ml. Grey triangles mark the best three concentrations of active lysozyme measured including corresponding concentration values. **B** Zoom plot of optimum parameter range including full factorial experiments marked as white triangles. Best full factorial experiments and best GA experiments are marked including concentration values.

3.3 Experimental Error Estimation by Full Factorial Experiments

A restriction of the parameter range to the optimum increases the ability to describe the parameter effects with a lower amount of experiments. Nevertheless the experimental error is a crucial factor which has to be determined to estimate the significance of the gained correlations. If three levels of a parameter are investigated for its effect on an objective function an error in the centre point might make the difference in the correlation curve between a hill, a plateau or even a valley. Another possibility which has to be taken into consideration is the correlation of the experimental error with experimental parameter levels. In this case one can expect to find differences in the measured objective function values with higher errors in a range where the analytical method produces higher standard deviations for example at very low absorption values or in the non-linear range of a calibration curve. Thus knowledge on the distribution of the experimental error can be used to assess the quality of the process in a specified experimental matrix. Therefore the three level full factorial experiments were performed five times with a completely new set of solutions at different days to get information on the worst case experimental error of independent experiments.

The relative standard deviation was found to be 15 % with a normal distribution of errors and thus lies in a typical range for complex high throughput experiments on a TECAN workstation [24]. A dependency of the experimental error on the combination of different parameter levels was validated by calculation of the correlation factor of all experimental parameter settings with the gained standard deviation. An R^2 of 0.1901 (calculated with Modde) indicates the absence of any influence of the set parameters on the experimental error meaning that observed parameter effects are significant.

3.4 Response Surface Fit

A more systematic method to calculate the global maximum value of the objective function and to numerically specify parameter effects and interactions is the fit of the data in the optimum to a response surface represented by a polynomial of degree 2.

This polynomial has 21 unknown coefficients, which have to be estimated by using our experimental data. The coefficients belonging to the quadratic terms provide information

about the importance or in other words the effect of the considered parameters, the coefficients belonging to the mixed terms describe combined effects of two parameters.

The performance of the polynomial fit in our case however was quite poor, not only because of the parameter uncertainty caused by the lack of enough data to estimate 21 coefficients but also because of two special data structure problems, explained in the following.

The complexity of the "multidimensional" data space determined by our refolding data could only be examined by looking at 3-dimensional subspaces (Figure 8, 9). In our case the 3 dimensional case of GSSG and DTT concentrations and the concentration of active lysozyme we get a stretched non symmetrical surface building a diagonal line with a fixed ratio of GSSG to DTT representing the redox potential created in the buffer system (Figure 6 A). This can hardly be fitted to a symmetrical response surface.

The second point is concerned with the fact that some of the analyses optima are not in the middle of our space but at or on the border of it. Naturally you cannot fit these optima to a polynomial of degree 2 properly. The approximation of a hill will not represent the shape of an ascending surface. In our case this holds true for nearly all parameters: pH (Figure 6 C), NaCl and $MgSO_4$ concentration (Figure 6 E) and also GSSG concentration (figure 6 A) showed such behaviour at the borders.

4 Conclusions

With the genetic algorithm for a robuste detection of the parameter space of intrest with a low data density in less important regions and a subsequent full factorial in the now reduced parameter space for further refinement of results, the optimum parameters for a five parameter system are verified with high resolution and low experimental effort compared to a full factorial approach covering the whole design space. The optimum conditions for lysozyme refolding qualitatively agree with published results generally favouring alkaline pH for disulfide bond formation and oxidizing agent in excess [17, 25, 26]. NaCl is mentioned to stabilize lysozyme folding intermediates which correlates with our observations [27]. Low concentrations of $MgSO_4$ can be explained by a 4 times higher ionic strength compared with the same NaCl concentration and a stronger salting-out effect according to the Hofmeister series decreasing solubility [28]. One might say that our results are supported by generally accepted principles for lysozyme refolding but are for the first time determined in a robust multi parameter optimization approach excluding imprecise assumptions on parameter dependencies or oversight of the optimum by parameter space reduction. Furthermore

solubility is identified as a good objective function for first optimization cycles according to an optimum parameter space overlapping with the optimum found for lysozyme activity. The experimental error detected was equally distributed over the parameter range. As no correlation of experimental parameter sets with the standard deviation of protein activity was found the influence of the investigated parameters can be considered significant. In addition we found out that the chosen parameter matrix can have a great influence on the ability to fit a response surface with a polynomial of degree two.

5. References

1.   Shukla, A.A., B. Hubbard, T. Tressel, S. Guhan, and L. D., *Downstream processing of monoclonal antibodies - Application of platform approaches.* Journal of Chromatography B, 2007. **848**(1): p. 28-39.

2.   Gottschalk, U., *Bioseparation in antibody manufacturing: The good, the bad and the ugly.* Biotechnology Progress, 2008. **24**: p. 496-503.

3.   Kelley, B., *Very large scale monoclonal antibody purification : The case for conventional unit operations.* Biotechnology Progress, 2007. **23**: p. 995-1008.

4.   Goto, M., Y. Hashimoto, T. Fujita, T. Ono, and S. Furusaki, *Important parameters affecting efficiency of protein refolding by reversed micelles.* Biotechnology Progress, 2000. **16**: p. 1079-1085.

5.   Mishra, R., R. Seckler, and R. Bhat, *Efficient refolding of aggregation-prone citrate synthase by polyol osmolytes.* The Journal of Biological Chemistry, 2005. **280**(16): p. 15553-155560.

6.   van den Berg, B., E.W. Chung, C.V. Robinson, P.L. Mateo, and C.M. Dobson, *The oxidative refolding of hen lysozyme and its catalysis by protein disulfide isomerase.* The EMBO Journal, 1999. **18**(17): p. 4794-4803.

7.   Ahn, J.H., Y.P. Lee, and J.S. Rhee, *Investigation of refolding condition for Pseudomonas fluorescens lipase by response surface methodology.* Journal of Biotechnology, 1997. **54**(3): p. 151-160.

8.   Armstrong, N., A. De Lencastre, and E. Gouaux, *A new protein folding screen: Application to the ligand binding domains of a glutamate and kainate receptor and to lysozyme and carbonic anhydrase.* Protein Science, 1999. **8**(7): p. 1475-1483.

9.   Ejima, D., K. Ono, K. Tsumoto, T. Arakawa, and Y. Eto, *A novel "reverse screening" to identify refolding additives for activin-A.* Protein Expression and Purification, 2006. **47**(1): p. 45-51.

10.   Jones, D.B., M.H. Hutchinson, and A.P.J. Middelberg, *Screening protein refolding using surface plasmon resonance.* Proteomics, 2004. **4**(4): p. 1007-1013.

11.   Lin, L., J. Seehra, and M.L. Stahl, *High-throughput identification of refolding conditions for LXR beta without a functional assay.* Protein Expression and Purification, 2006. **47**(2): p. 355-366.

12.   Qoronfleh, M.W., L.K. Hesterberg, and M.B. Seefeldt, *Confronting high-throughput protein refolding using high pressure and solution screens.* Protein Expression and Purification, 2007. **55**(2): p. 209-224.

13.   Rahimpour, F., G. Mamo, F. Feyzi, S. Maghsoudi, and R. Hatti-Kaul, *Optimizing refolding and recovery of active recombinant Bacillus halodurans xylanase in polymer-salt aqueous two-phase system using surface response analysis.* Journal of Chromatography A, 2007. **1141**(1): p. 32-40.

14.   Vincentelli, R., S. Canaan, V. Campanacci, C. Valencia, D. Maurin, F. Frassinetti, L. Scappucini-Calvo, Y. Bourne, C. Cambillau, and C. Bignon, *High-throughput automated refolding screening of inclusion bodies.* Protein Science, 2004. **13**(10): p. 2782-2792.

15.   Willis, M.S., J.K. Hogan, P. Prabhakar, X. Liu, K. Tsai, Y.Y. Wei, and T. Fox, *Investigation of protein refolding using a fractional factorial screen: A study of reagent effects and interactions.* Protein Science, 2005. **14**(7): p. 1818-1826.

16.   Tobbell , D.A., B.J. Middletona , S. Rainesb, M.R.C. Needhamb, I.W.F. Taylora, J.Y. Beveridgea, and W.M. Abbott, *Identification of in vitro folding conditions for Procathepsin S and Cathepsin S using fractional factorial screens.* Protein Expression and Purification, 2002. **24**(2): p. 242-254.

17.   Lin, J.L., R.C. Ruaan, and H.J. Hsieh, *Refolding of partially and fully denatured lysozymes.* Biotechnology Letters, 2007. **29**(5): p. 723-729.

18.   Berg, A., J. Kittelmann, and J. Hubbuch, *Development and characterization of an automated high throughput screening method for optimization of protein refolding processes* in preparation, 2008.

19.   Scheich, C., F.H. Niesen, R. Seckler, and K. Bussow, *An automated in vitro protein folding screen applied to a human dynactin subunit.* Protein Science, 2004. **13**(2): p. 370-380.

20.   Chen, G.Q. and E. Gouaux, *Overexpression of a glutamate receptor (GluR2) ligand binding domain in Escherichia coli: Application of a novel protein folding screen.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(25): p. 13431-13436.

21.   Susanto, A., K. Treier, E. Knieps-Gruenhagen, and J. Hubbuch, *High throughput screening for the design and optimization of chromatographic processes: Automated optimization of chromatographic phase systems.* Chemical Engineering and Technology, 2007. **32**(1): p. 140-154.

22.   Pohlheim, H., *Evolutionäre Algorithmen*. 2000, Berlin, Heidelberg, New York: Springer Verlag.

23.   Hansen, N., *On self-adaption in evolutionary strategies.* Studies in computational intelligence. Vol. 136. 2006, Berlin / Heidelberg: Springer Verlag. 31-57.

24.   Bensch, M., B. Selbach, and J. Hubbuch, *High throughput screening techniques in downstream processing: Preparation, characterization and optimization of aqueous two-phase systems.* Chemical Engineering Science, 2007. **62**(7): p. 2011-2021.

25.   Buswell, A.M., M. Ebtinger, A.A. Vertes, and A.P.J. Middelberg, *Effect of operating variables on the yield of recombinant trypsinogen for a pulse-fed dilution-refolding reactor.* Biotechnology and Bioengineering, 2002. **77**(4): p. 435-444.

26.   Hevehan, D.L. and E.D. Clark, *Oxidative renaturation of lysozyme at high concentrations.* Biotechnology and Bioengineering, 1997. **54**(3): p. 221-230.

27.   Bieri, O., G. Wildegger, A. Bachmann, C. Wagner, and T. Kiefhaber, *A salt-induced kinetic intermediate is on a new parallel pathway of lysozyme folding.* Biochemistry, 1999. **38**(38): p. 12460-12470.

28.   Hofmeister, F., *Zur Lehre von der Wirkung der Salze.* Arch. Exp.Pathol.Pharmakol., 1888. **24**: p. 1 - 16.

Solid Phase Refolding

Annette Berg, Joerg Kittelmann, Juergen Hubbuch*

Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Science, University of Karlsruhe (TH), 76131 Karlsruhe, Germany

*Corresponding author. Tel.: +049 721 608-2557; fax: +049 721 608-6240. Email-adress: juergen.hubbuch@kit.edu

Abstract

Inclusion body protein refolding on chromatographic columns is described to be superior to refolding by simple dilution in refolding buffer concerning yields of soluble and active protein. Nevertheless these processes are rarely used for their high complexity being an obstacle to process optimization. Besides an appropriate buffer design for the folding as for the chromatographic process, parameters like protein loading, resin type and velocity of buffer exchange from denaturing to renaturing conditions have to be evaluated. In this publication automated methods for the investigation and optimization of process parameters during protein refolding on chromatographic resins are described. Adsorption and refolding was performed in a batch chromatography approach resulting in low resin and protein consumption. We focused on ion-exchange resins and lysozyme as model protein. Binding of denatured lysozyme was investigated by automated measurement of isotherms and kinetics. Significantly lower binding capacities of the denatured compared to native lysozyme were calculated from isotherm data with similar binding kinetics. Lower binding strength of the denatured species on ion-exchange materials could be explained by the disintegration of charged patches on the protein surface during denaturation. After resin screening a material with high binding capacity was chosen for refold screening. Refolding buffer pH between 6 and 9 and urea concentrations between 0 and 2 M were analyzed for their effect on soluble protein yield. Lysozyme was desorbed from the ion-exchange resin under two different conditions: at low ionic strength during refolding and during the high salt elution step which hints at two different lysozyme species formed during the process. The yield of both eluted species was not affected by refolding pH or urea concentration. In contrast elongation of the refolding step reduced preliminary eluted protein at low ionic strength hinting at a higher content of completely folded lysozyme with improved binding at low salt concentrations. These results show that the automated method is not only a useful tool to screen for on-column refolding parameters but in addition can be used to qualitatively elucidate interdependencies between protein folding and adsorption/desorption processes.

172

1 Introduction

The development of matrix assisted refolding processes is complex due to a high number of process parameters such as sample loading, composition of mobile phase including gradient characteristics during denaturant reduction, elution flow rate and of course properties of the solid phase [1]. In general, adsorption of proteins should inhibit intermolecular interactions leading to protein loss via aggregation in dilution refolding processes. Thus a higher concentration of refolded and concomitantly purified target protein should be reached [2]. Refolding of adsorbed protein on chromatographic resins is described for affinity resins with immobilized folding catalysts or artificial chaperons and of course also with immobilized metal ions like nickel or copper. Furthermore refolding on hydrophobic interaction and ion exchange resins is proved to be a potential approach [3].

There are numerous reports on the use of different chromatographic techniques each having its pros and cons [4-7]. In size exclusion chromatography (SEC) protein is not adsorbed to the column material, thus this technology does not fulfil the prerequisite of preventing intermolecular interactions. Affinity interaction chromatography is generally restricted to modified proteins or dependent on costly specific chromatographic material [2]. In hydrophobic interaction chromatography (HIC) refolding high salt concentrations during sample loading might lead to solubility problems [8] and an observed tendency of denaturation with binding to HIC materials might also prevent protein from folding [9]. Ion exchange chromatography is one of the most frequently used methods in protein purification due to high binding capacities, low salt concentrations during sample loading and a lower tendency of proteins to denature as interacting molecular residues are on the protein surface.

Fundamental characterization of adsorption processes is done in batch mode by measuring adsorption isotherms to asses the protein binding affinity and the static binding capacity of a protein resin combination under certain buffer conditions in equilibrium. Measurement of uptake kinetics yields insight into the velocity of the adsorption process. In dynamic experiments column capacity is assessed for a chosen residence time by means of protein break through studies. The resulting data provide information on dynamic and total binding capacity. The total binding capacity correlates with data from batch binding experimentation. During normal protein separation studies using chromatographic processes gradient elution is used to exploit different binding strengths of different protein species present. In protein refolding studies gradient elution can be used to separate differently folded proteins. In this

integrated approach refolding and separation of refolded species are carried out within a single process step.

Refolding on ion exchange resins and its optimization concerning protein loading, pH and urea concentration of the refolding buffer and residence time of adsorbed protein in refolding buffer is extensively described in literature. Common to all reports is that buffer composition has to be adapted to every protein individually with the general rule to use low residual urea concentrations and alkaline pH. Low sample loading is also described to be favourable for higher refolding yields. Sample loss on the column by aggregation or strong non specific binding is still a hurdle to be taken [2, 8, 10, 11].

In this study an automated high throughput method is established to characterize these effects on the yield of active lysozyme during matrix assisted refolding on ion exchange adsorber materials. As experimental tool batch adsorption studies were used due to their simplicity, speed and low material consumption of native and denatured protein. To do so liquid handling of viscous high molar urea solutions had to be optimized prior to the actual studies to assure good reproducibility. The overall aim was to define a strategy allowing the optimization of multiple parameters with low amounts of protein, buffer and adsorber resin in a rational time frame.

2 Materials and Methods

2.1 Materials

2.1.1 Robotic Work Station

For all studies we used an automated pipetting station Freedom Evo 200 (Tecan Crailsheim, Germany) equipped with one liquid handling arm and two grippers. Pipetting was performed with 8 fixed standard tips. A centrifuge Rotanta RSC46 (Hettich Kirchenlengern, Germany), a magnetic orbital shaker with four positions and a shaking diameter of 2 mm (Inheco Munich, Germany) and a spectrophotometer Genios Pro (Tecan Crailsheim, Germany) were integrated into the robotic platform. The software Evoware (Tecan Crailsheim, Germany) was used to control the robotic workstation.

2.1.2 Disposables

96 well microtiter plates and UV microtiter plates with a well volume of 360 µl were purchased from Greiner (Frickenhausen, Germany). Deep well plates with a well volume of 1200 µl were from ABgene (Surrey, UK). 1000 µl AcroPrep 96 well filter plates with a pore size of 0.45 µm and a GHP membrane were produced by Pall (Dreieich, Deutschland).

2.1.3 Chemicals and Proteins

All chemicals had analytical grade and were purchased from Sigma Aldrich (St. Louis, USA) as well as hen egg white lysozyme (L-6876) with ≥ 90% protein content.
SP Sepharose FF and CM Sepharose FF were from GE Healthcare (Freiburg), Toyopearl 650 SP from Tosoh (Tokyo, Japan) and EMD Fractogel SO_3 (M) from Merck (Darmstadt).

2.2 Methods

2.2.1 Preparation of Stock Solutions and Adsorbent Material

Buffer and Salt Solutions

Buffer pH was adjusted by titration of acidic and basic components. Buffer for denaturation of lysozyme (8 M urea, 50 mM potassium phosphate buffer, pH 8) was prepared without DTT. DTT was added from a frozen 1 M stock solution in a concentration of 5 mM directly before use.

Lysozyme Containing Solutions

A stock solution of denatured lysozyme was prepared by solubilization of 10 mg/mL lysozyme in denaturing buffer (8 M urea, 50 mM potassium phosphate buffer, pH 8, 5 mM DTT). Miximg was performed by short vortexing. The solution was incubated in an overhead shaker at room temperature for 16 hours.

Solutions of native lysozyme were prepared by solubilization of lysozyme in concentrations between 1 and 0.016 mg/ml for calibration purposes and in a concentration of 10 mg/ml as stock solution in Milli-Q water by short vortexing directly before use.

Micrococcus lysodeiktikus Suspension

Micrococcus lysodeitikus was suspended by intense vortexing in 100 mM potassium phosphate buffer at pH 6 using a concentration of 0.65 mg/ml. The preparation of the *Micrococcus* suspension was carried out 30 minutes before use.

Aliquotation of Adsorber Material

Adsorber material was aliquoted in a device presented in [12]. A commercially available solution, the MediaScout® ResiQuot (Atoll, Weinheim) was used to fill adsorber particles into small cylindrical cavities with a defined volume of 7.8 ± 0.3 µl by an applied vacuum. The particles were washed twice with Milli-Q water and twice with the loading buffer of the subsequent experiment for equilibration. Chromatographic resin aliquots were stored in 100 µl of equilibration buffer.

2.2.3 Automated Measurement of Adsorption Isotherms

In Figure 1 three different approaches for the determination of adsorption isotherms are summarized. Because method development performed on the measurement of isotherms also builds the basis for measurement of binding kinetics as well as for refolding experiments with adsorbed protein the validation of liquid handling is described in this section. Automated measurement of adsorption isotherms for native protein was already published by Bensch et al. [13] and corresponds to the method A in Figure 1. Stock solutions of protein, buffer and Milli-Q water are directly transferred into a deepwell plate with a total transfer volume of 500 µl. The deepwell plate was pre-filled with 15 µl adsorber wetted with 100 µl of sample buffer. Mixing of solvent components as well as re-suspension of adsorber particles is performed by orbital shaking of the plate for two hours at 1500 rpm. Subsequently adsorber material was sedimented by centrifugation at 760 RCF for 4 minutes and 300 µl of the supernatant were analysed for total protein content both undiluted and in a 1/5 dilution in UV microtiter plates at 280 nm. Samples without added adsorber resin were used for calibration curves. In this

approach buffer exchange to initiate refolding can only be performed quantitatively after a transfer of adsorber particles into a filter plate. In a second method shown in Figure 1 B water, protein and buffer stock solutions were pre-mixed to a total volume of 800 µl in a separate deepwell plate by a three times aspirating and re-dispensing step with 700 µl liquid volume. Following 500 µl of the pre-mixed protein solution was added to the resin particles in a second deepwell plate for a 2 hours incubation step with orbital shaking at 1500 rpm. To facilitate this procedure adsorption isotherms were directly prepared in filter plates as visualized in Figure 1 C. During shaking for resuspension of adsorber particles liquid loss from the membrane bottom of the plate could only be prevented by a sealing of the plate with an adhesive foil. All isotherms were measured in triplicates.
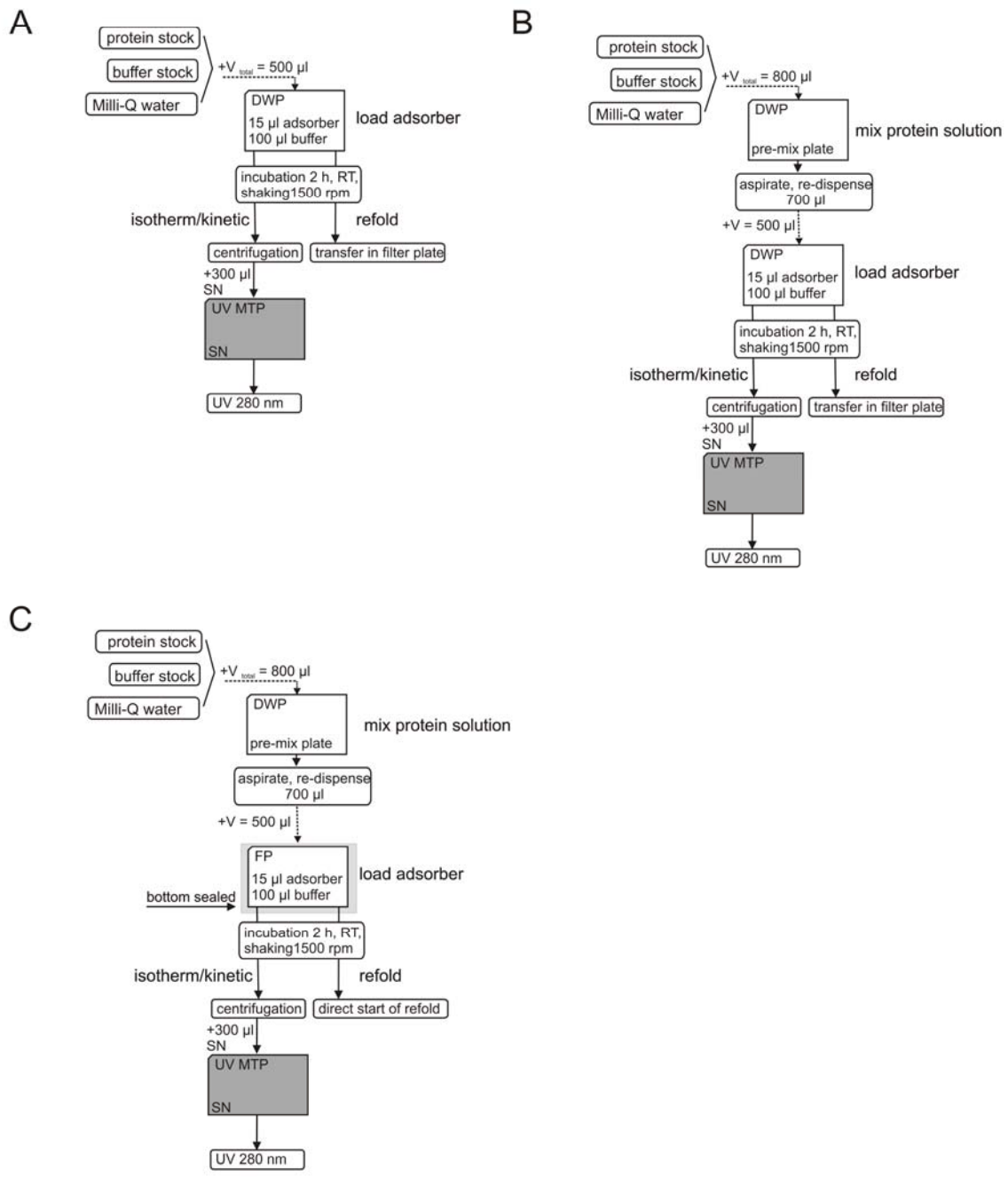
Figure 1: Flow scheme of protein loading on resin particles for isotherms/kinetics and for refolding processes. **A** Method transfer from [13] with direct mixing of all solvent components together with resin particles in a single deep well plate. **B** Pre-mixing of all solvent components in a separate plate by aspirating and re-dispensing of solvent and transfer of the solution into the resin containing deep well plate. **C** Pre-mixing of all solvent components in a separate plate by aspirating and re-dispensing of solvent and transfer of the solution into a sealed filter plate containing resin particles to facilitate buffer exchange in the refolding process.

178

The protein capacity in equilibrium q* was calculated from the known protein concentration c_0 at the beginning, the measured protein concentration in equilibrium c* after adsorption and the volume of the protein solution V_{sol} and the adsorber V_{ads} aliquots according to equation 1.

$$q* = \frac{V_{sol} + (c_0 - c*)}{V_{ads}}$$

(1)

The data were fitted either to a Langmuir shaped isotherm for native protein (equation 2) with the Langmuir adsorption coefficient K_L and the maximum capacity q_{max}

$$q* = \frac{K_L * q_{max} * c^*}{1 + K_L * c^*}$$

(2)

or to a BET shaped isotherm (equation 3) with the BET adsorption coefficient K and the maximum solubility of the protein c_{sat} for denatured protein.

$$q* = \frac{K * q_{max} * c^*}{(c_{sat} - c^*) * (1 + (K - 1) * c^* / c_{sat})}$$

(3)

2.2.4 Automated Measurement of Adsorption Kinetics

The experimental set up for the measurement of binding kinetics is very similar to that used to determine adsorption isotherms (see Figure 1). In the optimized procedure 500 µl of a pre-mixed protein solution is added to 15 µl adsorber at defined time points (Figure 1 C). The protein concentration used were 5.25 mg/ml and 2.19 mg/ml at the beginning of the experiment. All samples were incubated on an orbital shaker with a shaking frequency of 1500 rpm. The adsorption process is stopped by sedimentation of the adsorber particles using centrifugation at 3040 RCF for one minute. Evaluation of protein content is performed by measurements of the absorption at 280 nm in UV microtiter plates as described above. For comparability of kinetic curves the protein concentrations c_I were normalized with the end concentration c_e and the initial concentration c_0 according to equation 4.

$$c_{norm} = \frac{c_i - c_e}{c_0 - c_e}$$

(4)

The curve is empirically fitted with equation 5 with t as time and k as velocity constant.

$$c_{norm} = e^{-kt}$$

(5)

2.2.5 Automated Screening of Solid Phase Refolding Conditions

A process flow scheme is given in Figure 2. As described for the isotherms and kinetics deep well plates were pre-filled with 15 µl resin. 500 µl of denatured lysozyme solution previously mixed from stock solutions were added to the adsorber material and incubated for 2 hours with shaking at 1500 rpm.

To facilitate buffer exchange steps for washing and refolding the loaded adsorber resin was transferred to a filter plate with the fixed tips of the liquid handling system. The resin particles were re-suspended by three times repeated aspirating and dispensing steps with 450 µl volume. 600 µl adsorber suspension was aspirated after the tips were filled with 100 µl denaturing buffer without DTT for complete removal of resin particles after pipetting. A total volume of 700 µl was transferred to a filter plate and the deep well plate was washed in a second pipetting step with 700 µl denaturing buffer without DTT. This wash fraction was also transferred to the filter plate after evacuation of the filter by centrifugation at 3040 RCF for 1 minute. DTT was thus removed from the adsorber in a second centrifugation step.

Renaturation buffer systems (50 mM potassium phosphate buffer with the indicated pH and urea concentration) were mixed from stock solutions in a deep well plate. Refolding is initiated by addition of 300 µl refolding buffer with different compositions. The filter plate was incubated for 20 minutes with orbital shaking at 1500 rpm if not indicated differently. Subsequently the refolding buffer is removed by centrifugation at 3040 RCF for 1 minute and the concentration of lysozyme in the solvent is measured by UV absorption in a UV microtiter plate. This fraction of lysozyme is later called "preliminary eluted protein" or "eluate during refolding", as lysozyme measured here does not bind to the ion-exchange resin under refolding conditions.

Refolded protein was eluted with 300 µl refolding buffer containing 1 M NaCl. This fraction is called "eluted protein" as desorption occurs at high salt concentrations. The adsorber was incubated in elution buffer for 20 minutes with shaking at 1500 rpm before the eluted fraction is transferred to a UV microtiter plate by centrifugation at 3040 RCF for 1 minute. Protein concentration is measured by absorption at 280 nm. To remove aggregated or non-specifically bound protein from the adsorbent a second elution step with denaturation urea buffer is performed and the concentration measured according to the first elution step.
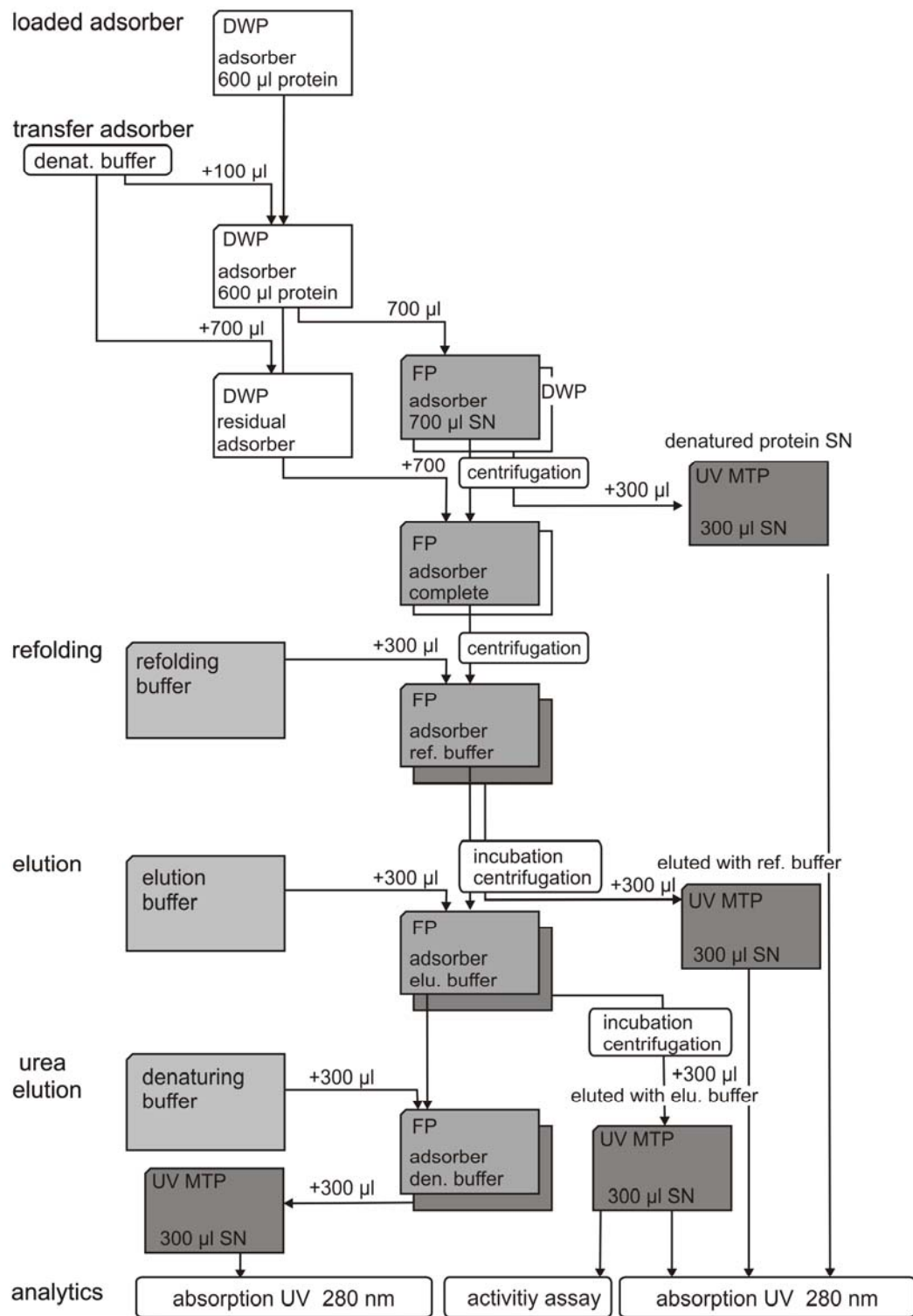
Figure 2: Flow scheme of the automated refolding process starting with loaded adsorber

2.2.6 Analytics

Measurement of Protein Concentration

Absorption measurements at 280 nm were performed in UV microtiter plates with a sample volume of 300 µl. In parallel to the process samples a seven point calibration curve consisting of different native lysozyme concentrations is prepared to calculate the slope of the calibration curve. In denatured protein samples including DTT a time and process dependent oxidation of DTT leads to absorption at 280 nm which corresponds to a shift of the y-axis intercept of the calibration curve. Therefore a parallel processing of a buffer blank is crucial. Samples from protein loading, refolding and elution steps can thus be measured.

Measurement of protein loss in the washing step after adsorption of denatured protein is impossible due to an undefined residual volume of buffer containing DTT.

Measurement of Lysozyme Activity

Lysozyme activity was measured by a decrease in absorption of a *Micrococcus lysodeikticus* suspension at 595 nm. 200 µl of a 0.65 mg/ml *Micrococcus* suspension were distributed in each well of a 96 well microtiter plate. Refold samples and a calibration standard with native lysozyme were diluted in 20 mM potassium phosphate buffer at pH 6 to bring the samples into the linear range of the assay. 40 µl of each diluted sample was transferred into the bacterial suspension. The assay solution was mixed by aspirating and re-dispensing of 170 µl. Afterwards the absorption at 595 nm was measured five times in 31 s intervals. The liquid handling tips have to be washed with 800 µl 1 M NaOH each before pipetting *Micrococcus* suspension or protein samples to avoid false positive results by a carry-over of lysozyme from inside the tubes.

3 Results and Discussion

3.1 Automated Measurement of Adsorption Isotherms under Denaturing Conditions

Data on the adsorption of denatured protein on the adsorbent surface can be used to assess the suitability of different adsorbents for the given process and helps to elucidate process economics. An automated method to measure adsorption isotherms for native protein was published by Bensch et al. [13]. In the course of this work this approach was adapted to the distinct parameters of a denaturing buffer system, i.e. high viscosity. A scheme of the different automated processes validated in this study to measure isotherms is given in Figure 1. Details are described in the following sections on method development.

3.1.1 Preparation of Protein Solutions

Due to the high viscosity of solutions used for denaturation like 8 M urea, sufficient mixing represents a critical process parameter. For the determination of different points on the adsorption isotherm protein solutions with different concentration should be mixed from a stock solution of 10 mg/ml denatured protein and denaturing buffer (8 M urea, 50 mM potassium phosphate pH 8, 5 mM DTT). It should be evaluated if a dilution of denatured protein stock solution with denaturing buffer can be mixed to homogeneity just by orbital shaking during the incubation period in the deepwell plate containing the resin (Figure 1 A). This approach was known to be possible for native protein if a deepwell plate with a round well geometry is used for the isotherm. In an alternative approach it should be validated if a pre-mixing of different protein solutions in a separate plate by aspirating and re-dispensing has to be performed prior to addition into the resin containing deepwell plate (Figure 1 B). Isotherms measured by both methods are shown exemplarily for denatured lysozyme on SP Sepharose FF in Figure 3. For both experiments a BET shaped isotherm was fitted.
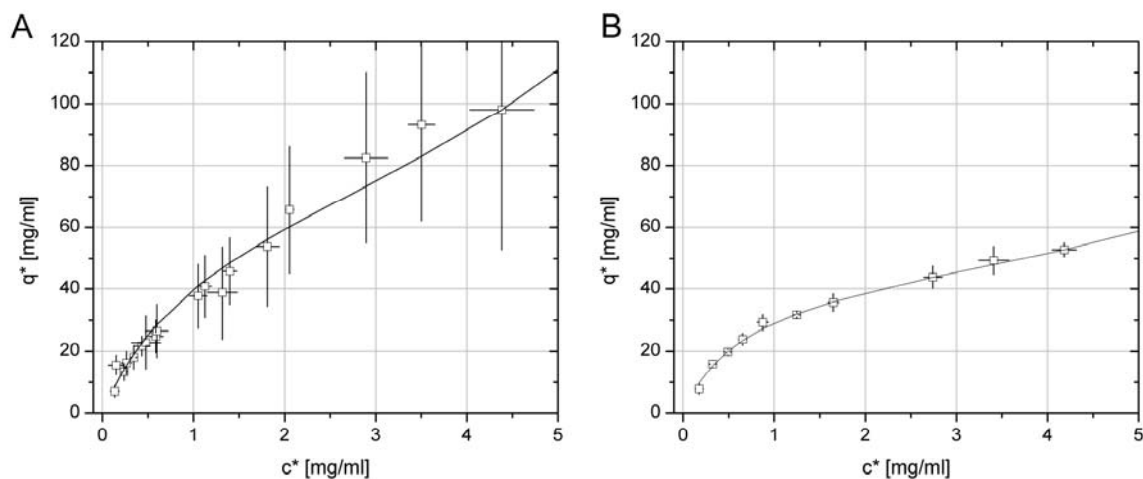
Figure 3: Isotherm on 15 µl SP Sepharose FF with denatured protein. Protein solutions were mixed from 10 mg/ml denatured lysozyme and denaturing buffer (8M urea, 50 mM potassium phosphate pH 8, 5 mM DTT) **A** by orbital shaking during incubation with the adsorber particles in a deepwell plate with round well geometry and **B** by aspirating and re-dispensing before addition to the adsorber particles in a separate deepwell plate

The most striking difference using both methods lies in the significantly different standard deviations as a result of the differences in solution handling. The average standard deviation for the process performing mixing of protein and buffer stock solutions by shaking together with the resin was 12 % for the protein concentration in equilibrium and 42 % for the capacity compared to 5 % and 6 % for the same values measured with intense pre-mixing by aspirating and re-dispensing steps in a separate plate. These results show that orbital shaking only leads to limited or slow mixing of protein and buffer stock solutions because 8 M urea buffer has a higher viscosity than water or buffer solutions with lower concentration. A critical rpm for mixing of resin particles in solution is defined in [13] but seems to be not sufficient for mixing with high urea content. This problem with liquid handling in highly concentrated urea solutions is also described in [3]. In contrary the preparation and mixing of stock solutions by aspiration and dispensing steps clearly leads to more precise data. A similar approach is also described for mixing of aqueous two phase systems with a highly viscous PEG phase in [14]. For isotherms under native buffer conditions with low concentrations of salt pre-mixing is not mandatory.

3.1.2 Influence of Plate Type on Resuspension of Resin

The screening process is highly dependent on a simple handling of different liquids and the respective buffer exchange and incubation of the adsorbent particles. An obvious solution to facilitate this would lie in the use of filter plates. For the measurement of adsorption isotherms two plate types were compared concerning their impact on resuspension of resin particles, deep well plates with a round bottom profile which were commonly used for native protein and filter plates with a flat filter bottom. In Figure 4 the results for both plate types and EMD Fractogel SO_3 are given.
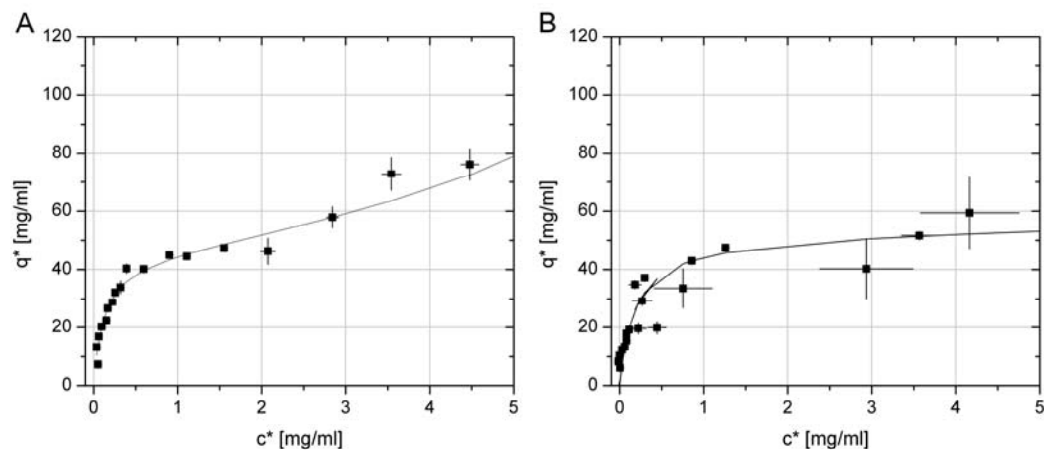

Figure 4: Isotherm on EMD Fractogel SO_3 with denatured lysozyme in 8 M urea, 50 mM potassium phosphate pH 8 and 5 mM DTT after **A** incubation and shaking in 96 well deep well plates with round bottom profile and **B** after incubation and shaking in 96 well filter plates with flat bottom profile

In general mixing in filter plates with a flat bottom leads to higher standard deviations of 23 % for the protein concentration in equilibrium and 7 % for the capacity compared to 5 % and 2 % for the same values measured in deep well plates. In deep well plates with a round bottom the development of a vortex is promoted and increases mixing and particle dispersion. This is unfortunately not given for the filter plates used. Besides filter plates have to be sealed manually to prevent fluid loss throughout the membrane bottom during shaking which is a great limitation in automated processes.

As a conclusion performing mixing of stock solutions and resuspension of particles by shaking in a deepwell plate yields the worst reproducibility. This is likely due to local differences in protein concentration by insufficient homogenization of protein in urea buffer.

3.1.3 Comparison of Adsorption Isotherms of Denatured and Native Protein

With the optimized conditions described above adsorption studies of native and denatured lysozyme on four different resins were performed. The selected adsorber materials included one weak cation exchange resin, CM Sepharose FF, and three strong cation exchange resins, SP Sepharose FF with the same backbone as CM Sepharose FF, Toyopearl SP 650 and the grafted adsorber EMD Fractogel SO_3.

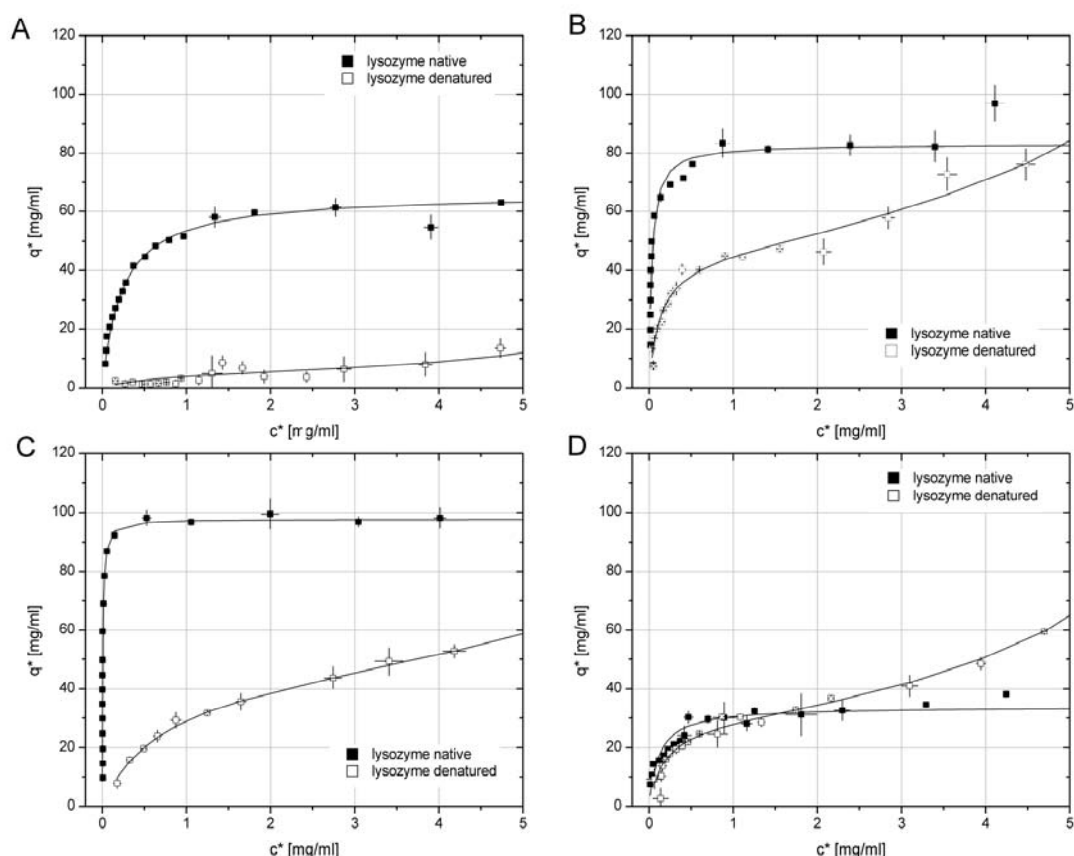In Figure 5 adsorption isotherms for the four different adsorber matrices with denatured and native lysozyme are depicted.

Figure 5: Isotherms of native (■) and denatured (□) lysozyme on **A** CM Sepharose FF; **B** EMD Fractogel SO_3 (M); **C** SP Sepharose FF; **D** Toyopearl SP 650. Lysozyme was denatured in 8 M urea, 50 mM potassium phosphate pH 8 and 5 mM DTT. Native lysozyme was solubilized in 50 mM potassium phosphate buffer pH 8.

Native lysozyme shows a nearly rectangular isotherm which could be fitted according to Langmuir (equation 2). Whereas denatured lysozyme reaches no saturation due to a curve

progression according to a BET isotherm (equation 3) hinting at a multilayer binding mechanism. This observation is also described in [15]. The parameters of the fit curves are given in Table 1.

Table 1: Fit parameters for isotherms of denatured and native lysozyme

Fit Parameters for Native Lysozyme (Langmuir)			
Adsorber	**K_L [ml*mg^{-1}]**	**q_{max} [mg/ml$_{ads}$]**	
CM Sepharose FF	4.23	66.5	
EMD Fractogel SO$_3$	39.95	83.2	
SP Sepharose FF	162.2	97.3	
Toyopearl SP 650	5.5	35.6	
Fit Parameters for Denatured Lysozyme (BET)			
Adsorber	**K [ml*mg^{-1}]**	**q_{max} [mg/ml$_{ads}$]**	**c_{sat} [mg/ml]**
CM Sepharose FF	15.84	5	8.4
EMD Fractogel SO$_3$	89.04	44.4	10.51
SP Sepharose FF	25.35	42.9	15.7
Toyopearl SP 650	52.58	28.3	8.8

For native lysozyme higher capacities and affinities on nearly all adsorber materials are observed except for Toyopearl SP650 which shows low binding for both folding states of lysozyme. The main reason for lower binding capacity and affinity of denatured lysozyme is a lack of distinct charged patches on the proteins surface. Hydrophobic residues previously buried inside the molecule and oppositely charged residues might modulate binding on the adsorber surface. These observations are supported by results from [7] showing shorter retention times of unfolded lysozyme compared to native lysozyme on ion exchange materials.

For native lysozyme capacity increases in the order Toyopearl SP 650 with 35.6 mg/ml, CM Sepharose FF with 66.5 mg/ml, EMD Fractogel SO$_3$ with 83.2 mg/ml and SP Sepharose FF with 97.3 mg/ml. Low capacity on Toyopearl SP 650 compared to SP Sepharose FF can be explained by a lower ligand density of Toyopearl [13]. CM Sepharose FF is superiour to Toyopearl SP 650 though being a weak cation exchange resin. The grafted adsorber Fractogel does not lead to improved capacities compared to Sepharose FF. As tentacle gels are usually favourable due to higher accessibility of protein binding sites, this factor seems to be less

important in lysozyme binding than other differences like backbone chemistry or pore size. In [16] lower capacities for lysozyme on EMD Fractogel SO₃ are also measured compared to SP Sepharose FF which confirms our results.

For denatured lysozyme Fractogel shows the highest affinity and a slightly higher capacity (44.4 mg/ml) than Sepharose FF (42.9 mg/ml) supporting the hypothesis of a wider distribution of charged binding sites due to random protein structure. Higher ligand densities of SP Sepharose FF lead to higher capacities compared to Toyopearl (28.3 mg/ml).

In the following SP Sepharose FF is used for refolding experiments due to its high binding capacity and broad industrial usage.

3.1.4 Comparison of Binding Kinetics of Denatured and Native Protein on SP Sepharose FF

The observed discrepancies between the adsorption behaviour of denatured and native lysozyme could also be explained by diffusional limitation and slower equilibrium adsorption kinetics of denatured lysozym within the 2 hours incubation period chosen.

In Figure 6 data on binding kinetics for 5.25 mg/ml and 2.19 mg/ml lysozyme in denatured and native conformation are shown by plotting capacity versus time.
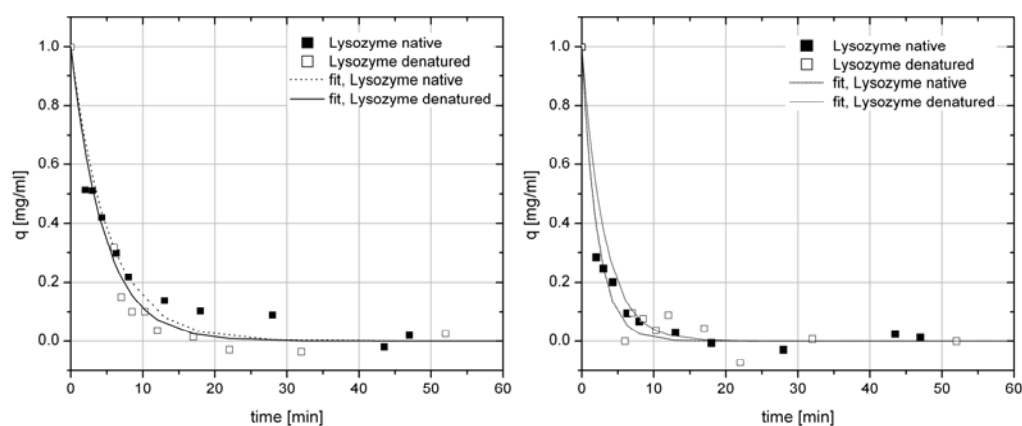

Figure 6: Adsorption kinetics of native (■) and denatured (□) lysozyme with **A** 2.19 mg/ml protein in added solution; **B** 5.25 mg/ml protein in added solution. Lysozyme was denatured in 8 M urea, 50 mM potassium phosphate pH 8 and 5 mM DTT. Native lysozyme was solubilized in 50 mM potassium phosphate buffer pH 8.

For both protein concentrations no difference in binding kinetics can be observed between denatured and native protein. After approximately 20 minutes near complete adsorption could be observed.

3.2 Parameter Effects during Solid Phase Refolding

In the following batch chromatography was validated towards its applicability as a tool to investigate important parameters such as protein loading, refolding buffer pH, urea concentration and residence time in refolding buffer. To facilitate protein concentration measurement no redox components or other substances absorbing light at 280 nm were used despite the high probability of low refolding yield already observed in refolding by dilution [17]. In contrast to common on-column strategies applying a gradient of refolding buffer a one step strategy was chosen with a 100 % shift from denaturing to renaturing buffer conditions. Protein elution was performed in a first elution step using high ionic strength followed by a second elution step applying denaturing buffer to elute aggregated or non-specifically bound protein. A detailed description and flow scheme (Figure 2) of the complete refolding process starting with the transfer of the loaded particles into the filter plates is given in the Materials and Method section.

3.2.1 Specific Activity Obtained during Solid Phase Refolding

In order to assess refold performance eluted protein needs to be analysed towards its activity. As refolding intermediates and aggregates tend to bind tightly to the resin material [18] it might well be possible, that only native and thus active protein is eluted in the high salt elution step, namely in the absence of high urea concentrations. After refolding of bound lysozyme with 50 mM potassium phosphate buffer at pH 8 and three different urea concentrations (0, 1 and 2 M) the first elution step containing 1 M NaCl was monitored towards lysozyme content and activity. Figure 7 shows a linear correlation of total soluble lysozyme and active lysozyme observed during this study. This supports the assumption that eluted protein is completely folded.
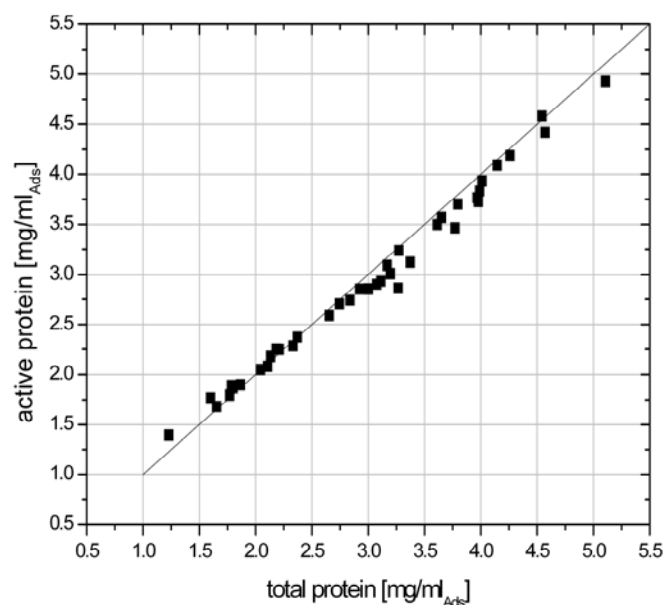
Figure 7: Active lysozyme concentration versus total lysozyme recovered. Refolding of bound lysozyme on SP Sepharose FF was performed with 50 mM potassium phosphate buffer pH 8 and different urea concentrations (0, 1, 2 M). 1 M NaCl was added for elution.

3.2.2 Protein Aggregation during Solid Phase Refolding and the Effect of Urea

Simple dilution of high denatured protein concentrations into refolding buffer might lead to low yields due to aggregation of folding intermediates [19]. The corresponding parameter in solid phase refolding describing the proximity of two proteins during refolding is the protein load realized on the adsorbent. We thus compared the influence of different adsorbent loadings on protein recovery. Next to a reduction in protein concentration to avoid aggregation, one might add various additives having a positive influence on protein solubility. A common method is the addition of low amounts of urea to the refolding buffer. It was shown in [20] that the addition of urea improves the solubility of the protein to be refolded.

Next to varying adsorbent load (approximately 10 to 105 mg/ml$_{ads}$) the refolding buffer used was varied by adding 0 M, 1 M and 2 M urea.

In Figure 8 the effect of protein loading on lysozyme yield after the two elution steps with 1 M NaCl and with denaturing buffer is shown.
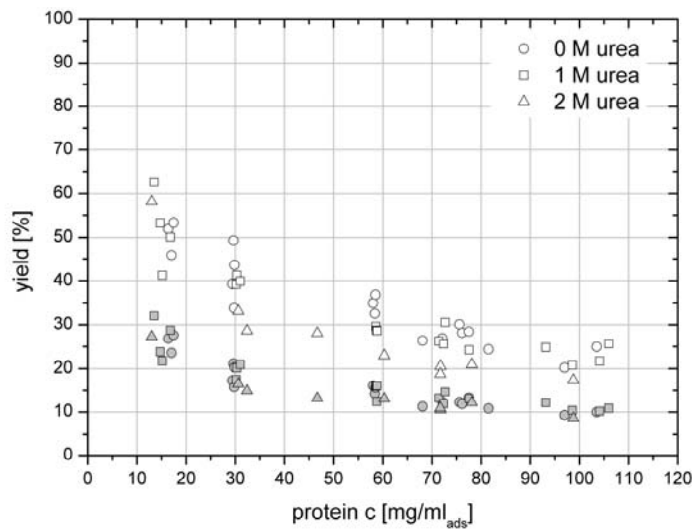
Figure 8: Effect of protein loading and urea concentration in refolding buffer on summed yields after elution with refolding buffer (50 mM potassium phosphate buffer, pH 8) and 1 M NaCl (grey symbols) and on yield after additional elution with 8 M urea, 50 mM potassium phosphate buffer, pH 8 and 5 mM DTT (white symbols). Refolding was performed on SP Sepharose FF.

An increase of protein loading from 10 to 105 mg/ml$_{ads}$ leads to an exponential decay of eluted lysozyme yield to approximately 71 % eluted without elution of non-specifically bound protein by urea. Additional elution with denaturing buffer leads to a twofold higher total yield but shows a higher dependency on adsorber loading with an exponential decrease to approximately 24 % in the investigated range. The observations are independent from the used urea concentrations in refolding buffer. First of all the fact that further elution is possible with denaturing buffer corroborates the assumption of protein aggregation and non-specific binding on the column material. This corresponds to findings by [2] for the refolding of bovine serum albumin. Furthermore a stronger impact of protein loading on the yield gained with additional urea elution can be explained by an increase in aggregation. At very high protein loading a single elution step with urea buffer seems to be not suffiecient to re-solubilize protein on the adsorber surface. It is likely that pores are blocked by aggregates and thus aggregated protein inside the pores is not accessible.

The observations concerning the impact of protein loading on recovery in matrix assisted refolding are consistent with effects described in literature [2, 11]. The maximum yield of eluted lysozyme ranging from 10 to 30 % is relatively low which is typical for on column refolding processes [2, 8, 10, 11].

An effect of varying urea concentration in the refolding buffer on lysozyme yield could not be observed in our study even though an improvement of protein solubility during dilution

refolding was observed earlier with addition of low concentrations of urea or guanidinium hydrochloride [21, 22]. This might be due to non-specific interactions of refolding intermediates with the chromatographic resin preventing solubilisation effects normally occurring in free solution. Furthermore higher concentrations of urea lead to a slower refolding kinetic [8] which compensates for potential positive effects as the refolding process has to be elongated for improved yields. Analysis of lysozyme amount eluting preliminary in the refolding step at low ionic strength showed significantly increased elution with higher urea concentrations in the refolding buffer as depicted in Figure 9. Eluted lysozyme mass plotted in relation to the adsorber volume is approximately three fold higher at 2 M urea concentration than with 1 M urea. This effect is independent of the protein loading on the resin particles. The difference between 1 M urea and the absence of urea lies within the experimental error.
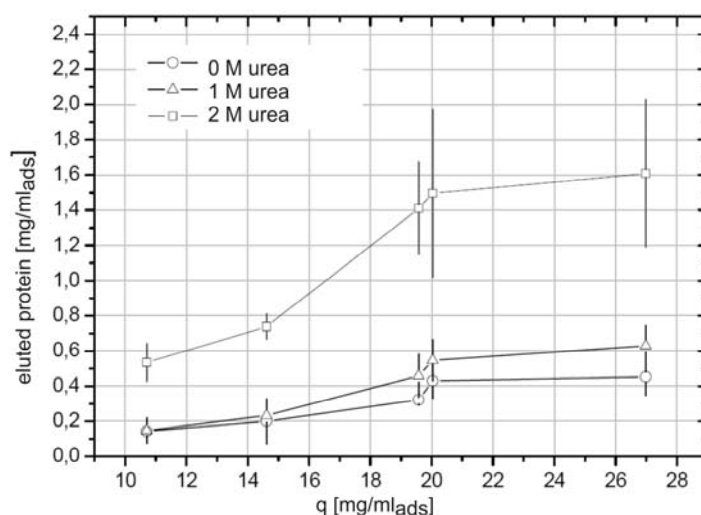

Figure 9: Effect of urea concentration on the amount of preliminary eluted lysozyme in refolding buffer without NaCl addition. Refolding was performed on SP Sepharose FF with 50 mM potassium phosphate buffer, pH 8.

Lower affinity of lysozyme and as a consequence desorption at low ionic strength can be explained by incomplete folding of the protein as already observed for isotherms from denatured protein. On the other hand we saw a correlation of active and soluble protein eluted with high salt conditions without urea which was explained by aggregation and non-specific binding of incompletely folded or misfolded protein. The results in this section show a combination of lower binding affinity of folding intermediates and an improvement of their solubility by addition of low urea concentrations. Protein elutes during refolding without high

ionic strength due to lower binding affinity of folding intermediates. These intermediates are not soluble without urea and as a consequence can not be eluted. As a conclusion also the analysis of protein folding effects on adsorption/desorption properties is possible in the developed batch matrix-assisted refolding process.

3.2.3 Effect of Incubation Time during Solid Phase Refolding

A critical parameter when designing refolding processes is the incubation time of denatured protein in refolding buffer systems. In order to elucidate this parameter dependency different incubation times of denatured protein in refolding buffer were assessed and protein eluted during the refolding step and total protein eluted in the refolding and elution step analysed.
In figure 10 A the fraction of preliminary eluted lysozyme during the refolding step is plotted versus the incubation time in refolding buffer for two load concentrations 4 mg/ml and 2 mg/ml.
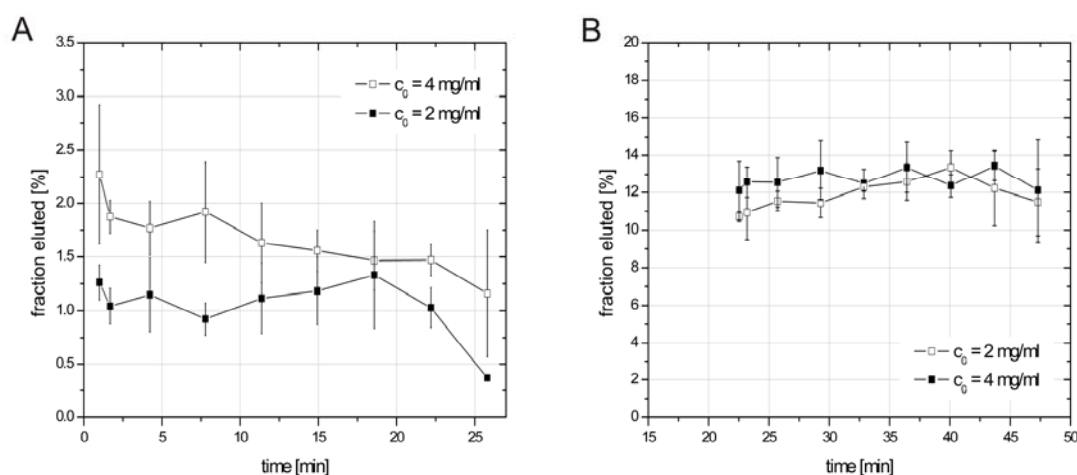

Figure 10: Effect of incubation time in refolding buffer for 2 mg/ml (■) and 4 mg/ml (□) load on **A** the protein fraction eluted during refolding without high ionic strength and on **B** the total eluted fraction during refolding at low and elution at high ionic strength . Refolding was performed on SP Sepharose FF with 50 mM potassium phosphate buffer, pH 8. Elution was performed with 50 mM potassium phosphate buffer, pH 8 and 1 M NaCl.

The incubation time had an influence on the loss of protein in the refolding step decreasing with a factor of around 2 from 1 minute to 26 minutes for 4 mg/ml load and with a factor of 3.4 for 2 mg/ml load. For the higher protein loading around 40 % higher fractions of lysozyme eluted during refolding in comparison to the lower protein load analyzed. In Figure 10 B the overall fraction of eluted protein including the refolding and the elution step are plotted versus the time needed for the complete process until elution is completed. The first

time point in this figure corresponds to the first time point in Figure 10 A. As the loss during refolding for all incubation times was below 2.5 % no effect on the overall yield of refolded protein can be observed.

An improved binding with longer incubation time in refolding buffer can be explained by a stronger binding affinity of completely refolded lysozyme in comparison to denatured lysozyme and an increasing fraction of the refolded species with time. As folding kinetics is described to be slower with higher protein concentrations the observed effect is stronger for higher protein loading. For BSA an increase in protein recovery and folding yield is observed with longer incubation times on an ion-exchange column supporting our results [2].

3.2.4 Effect of Buffer pH during Solid Phase Refolding

The pH set during a refolding reaction mainly influences refolding performance. When the refolding process is performed in solid phase refolding mode, the pH also plays a major role determining the binding strength. To elucidate potential effects binding of denatured protein and refolding was performed in a pH range from pH 6 to pH 9 using 50 mM potassium phosphate.

In Figure 11 average yields of eluted protein are depicted for every pH value for loadings between 2.5 and 22.5 mg/ml$_{ads}$.
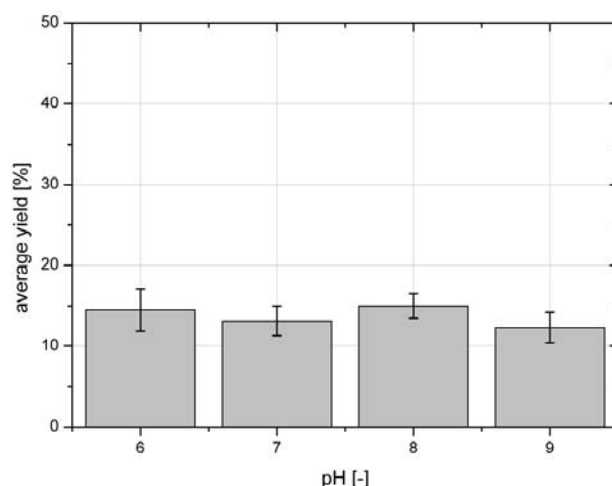

Figure 11: Effect of pH on the average yield of lysozyme eluted after refolding on SP Sepharose FF at loadings between 2.5 and 22.5 mg/ml$_{ads}$). The standard deviation for different loadings is given as error bars.

No significant difference could be observed for the yield of refolded protein at the four pH values. An effect of the net charge of lysozyme is possibly to small as lysozymes net charge only changes from 7.7 for pH 6 to 5.5 for pH 9 [23]. This finding agrees with our investigation on dilution refolding without addition of redox components where no dependency of yield on buffer pH could be found [17].

4 Conclusions

Protein binding of denatured protein on different ion exchange resins was characterized by automated measurement of adsorption isotherms and kinetics. Lower maximum capacities were observed for denatured lysozyme in comparison to native lysozyme and a BET shaped isotherm describing the obtained data points was used hinting at multilayer binding. Adsorption kinetics showed no significant difference for both species excluding a kinetic effect during isotherm assessment.

The resin with the highest binding capacity, SP Sepharose FF, was used to establish an automated refolding process including a single step shift to refolding buffer and a subsequent elution with high ionic strength. Protein amenable to elute with high salt concentrations was identified to be completely folded by an integrated enzyme assay.

It could be shown that lower protein loading leads to an increase in refolding yields. This corresponds to a correlation of high soluble protein yield and low initial protein concentrations in dilution refolding.

Refolding buffer pH had no influence on the refolding yield in the investigated pH range between 6 and 9 which is consistent with refolding results in the dilution mode in the absence of redox components. Besides no significant shift in the net charge of lysozyme occurs in the analyzed pH range and consequently adsorption properties should stay constant.

Interestingly even complex interactions between folding and adsorption/desorption processes became obvious during the screening of different incubation times in refolding buffer and different urea concentrations during refolding.

Incubation time in refolding buffer was analyzed for its effect on overall yield of soluble lysozyme and lysozyme eluted early in the refolding step. Longer incubation decreased elution in refolding buffer assumedly due to a re-adsorption of native lysozyme with higher binding affinity after complete folding.

Low urea concentrations of 1 and 2 M in refolding buffer did not lead to an improvement of overall soluble protein yield which is in contrast to described positive effects in dilution

refolding. Desorption of protein in refolding buffer was increased by increasing urea concentration hinting at slower folding kinetics in the presence of urea already described in literature.

References

1.	Geng, X.D. and C.Z. Wang, *Protein folding liquid chromatography and its recent developments.* Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences, 2007. **849**(1-2): p. 69-80.

2.	Langenhof, M., S.S.J. Leong, L.K. Pattenden, and A.P.J. Middelberg, *Controlled oxidative protein refolding using an ion-exchange column.* Journal of Chromatography A, 2005. **1069**(2): p. 195-201.

3.	Jungbauer, A., W. Kaar, and R. Schlegl, *Folding and refolding of proteins in chromatographic beds.* Current Opinion in Biotechnology, 2004. **15**(5): p. 487-494.

4.	Boschetti, E., *Advanced sorbents for preparative protein separation purposes.* Journal of chromatography, 1994. **658**(2): p. 207-236.

5.	Ueda, E.K.M., P.W. Gout, and L. Morganti, *Current and prospective applications of metal ion-protein binding.* Journal of Chromatography A, 2003. **988**: p. 1-23.

6.	Xiao, Y., T.T. Jones, A.H. Laurent, J.P. O´Connell, T.M. Przybyien, and E.J. Fernandez, *Protein instability during HIC: hydrogen exchange labeling analysis and a framework for describing mobile and stanionary phase effects.* Biotechnology and Bioengineering, 2007. **96**(1): p. 80-93.

7.	Yamamoto, S., S. Fujiia, N. Yoshimotoa, and P. Akbarzadehlaleh, *Effects of protein conformational changes on separation performance in electrostatic interaction chromatography: Unfolded proteins and PEGylated proteins.* Journal of Biotechnology, 2007. **132**(2): p. 196-201.

8.	Machold, C., R. Schlegl, W. Buchinger, and A. Jungbauer, *Matrix assisted refolding of proteins by ion exchange chromatography.* Journal of Biotechnology, 2005. **117**(1): p. 83-97.

9.   Jungbauer, A., C. Machold, and R. Hahn, *Hydrophobic interaction chromatography of proteins - III. Unfolding of proteins upon adsorption.* Journal of Chromatography A, 2005. **1079**(1-2): p. 221-228.

10.   Li, M. and Z. Su, *Refolding human lysozyme produced as an inclusion body by urea concentration and pH gradient ion exchange chromatography.* Chromatographia, 2002. **56**(1-2): p. 33-38.

11.   Li, M., G.F. Zhang, and Z.G. Su, *Dual gradient ion-exchange chromatography improved refolding yield of lysozyme.* Journal of Chromatography A, 2002. **959**(1-2): p. 113-120.

12.   Herrmann, T., M. Schroder, and J. Hubbuch, *Generation of equally sized particle plaques using solid-liquid suspensions.* Biotechnology Progress, 2006. **22**(3): p. 914-918.

13.   Bensch, M., P.S. Wierling, E. von Lieres, and J. Hubbuch, *High throughput screening of chromatographic phases for rapid process development.* Chemical Engineering & Technology, 2005. **28**(11): p. 1274-1284.

14.   Bensch, M., B. Selbach, and J. Hubbuch, *High throughput screening techniques in downstream processing: Preparation, characterization and optimization of aqueous two-phase systems.* Chemical Engineering Science, 2007. **62**(7): p. 2011-2021.

15.   Hunter, A.K., E.J. Suda, T.K. Das, R.E. Shell, J.T. Herberg, N. Ramasubramanyan, M.E. Gustafson, and S.V. Ho, *Impact of denaturation with urea on recombinant apolipoprotein A-IMilano ion-exchange adsorption: equilibrium uptake behavior and protein mass transfer kinetics.* Biotechnol J, 2007. **2**(1): p. 110-20.

16.   Staby, A., J.H. Jacobsen, R.G. Hansen, U.K. Bruus, and I.H. Jensen, *Comparison of chromatographic ion-exchange resins V. Strong and weak caion-exchange resins* Journal of Chromatography A, 2006. **1118**: p. 168-179.

17.   Berg, A., J. Kittelmann, and J. Hubbuch, *Development and characterization of an automated high throughput screening method for optimization of protein refolding processes* in preparation, 2008.

18.   Li, M., Z.G. Su, and J.C. Janson, *In vitro protein refolding by chromatographic procedures.* Protein Expression and Purification, 2004. **33**(1): p. 1-10.

19.   Kiefhaber, T., R. Rudolph, H.H. Kohler, and J. Buchner, *Protein aggregation in vitro and in vivo - A quantitative model of the kinetic competition between folding and aggregation* Bio-Technology, 1991. **9**(9): p. 825-829.

20.   Lanckriet, H. and A.P.J. Middelberg, *Continuous chromatographic protein refolding.* Journal of Chromatography A, 2004. **1022**(1-2): p. 103-113.

21.   Clark, E., M, *Protein refolding for industrial processes.* Current opinion in biotechnology, 2001. **12**(2): p. 202-207.

22.   Hevehan, D.L. and E.D. Clark, *Oxidative renaturation of lysozyme at high concentrations.* Biotechnology and Bioengineering, 1997. **54**(3): p. 221-230.

23.   Berg, A., M. Schuetz, and J. Hubbuch, *Automated measurement of protein solubility to rapidly assess complex parameter interactions.* in preparation, 2008.

4 Conclusions & Outlook

Although a high demand exists for the development of automated high throughput methods for protein refold screening several challenges could not be met up to now which are in few words insufficient HTS compatibility of analytics and the use of protein specific methods to measure folding. As a result the most prominent objective was the validation of non-specific methods to quantitatively estimate refolding success and the implementation of these methods in the robotic workstation. Two approaches were identified to potentially meet the above criteria which were protein solubility and tryphtophan fluorescence.

The development of analytical methods to measure protein solubility turned out to be complicated because of complex buffer matrices, low protein concentrations and protein structure variations interfering with commonly used techniques. Dye based assays and UV 280 nm absorption revealed to be sensitive to either the presence of different protein folding states or to refolding buffer components like oxidizing or reducing reagents. Consequently automation of methods for buffer exchange like ultrafiltration, size exclusion chromatography and dialysis had to be validated. All approaches turned out to have some serious draw backs. Ultrafiltration led to a protein concentration and buffer dependent loss of protein whereas dialysis and size exclusion chromatography failed to separate protein from oxidizing reagents in a manageable time with robotic compatible equipment. As a conclusion soluble protein content in refolding screening samples could not be measured directly. Protein aggregates were measured instead to calculate the soluble protein concentration on the basis of mass balances. Aggregates were separated by filtration in 96 well filter plates and were afterwards resolubilized in denaturing buffer for absorption measurements at 280 nm. Measured samples exhibit no variations in protein structure and buffer composition which minimizes the number of blank and calibration samples. Besides the achieved automation and parallelization of the process it is completely independent of the refolding buffer systems screened which offers maximum flexibility in buffer design.

Tryptophan fluorescence was chosen as second alternative to measure protein refolding success. Spectra of lysozyme could be easily measured in a 96 well plate photometer. As the red shift in fluorescence emission maximum from completely active to completely denatured protein is small an asymmetric Gauss fit had to be performed to calculate the exact emission intensity maximum. Single point measurements of tryptophan fluorescent showed a high sensitivity towards oxidizing reagents due to quenching by disulfide bonds and are thus not recommended for refolding screening.

A validation of protein solubility and tryptophan fluorescence to evaluate protein refolding systems is only possible if a reference method is available. In this study structural integrity was determined by means of a lysozyme activity assay. This standard assay detects a decrease in 595 nm absorption of a Micrococcus suspension up on lysis of the cell walls by lysozyme. Automation and parallelization could be performed in a 96 well plate format to provide the data basis for tryptophan fluorescence and solubility validation in a HTS approach.

A process for protein refolding by dilution in different buffer systems was completely automated on a commercially available robotic pipetting station including preparation of refolding buffer from stock solutions. Instantaneous intense mixing of denatured protein in renaturation buffer was reached by shaking during addition and aspirating and re-dispensing of the sample. Incubation time was set at 1 h to obtain equilibrium of aggregates and soluble protein for analytics.

In a validation screening 40 random refolding buffer systems were used consisting of buffer salts between pH 3 and pH 9, NaCl and $MgSO_4$ between 0 and 150 mM, DTT as reducing agent between 0 and 10 mM and GSSG as oxidizing agent between 0 and 20 mM. Tryptophan fluorescence spectra, protein solubility and enzymatic activity were determined according to the above described procedures. The screening revealed a good correlation between buffer parameters resulting in high soluble and high active lysozyme yields. On the other hand tryptophan emission spectra in lysozyme refolding samples only covered a close range. This observation is likely due to a hydrophobic collapse of lysozyme being obligatory for protein solubility. Soluble protein will thus exhibit a rather uniform spectral behaviour corresponding to this kind of conformation. As a conclusion tryptophan fluorescence does not offer the needed resolution to estimate refolding success at least in the case of lysozyme.

To guarantee for maximum reproducibility of refolding screening results significant parameters showing an effect on refolding yields have to be controlled. During validation runs the impact of differently denatured protein feedstocks on refolding yield and on favourable buffer composition for refolding was analyzed. Tryptophan fluorescence revealed to be a useful tool to measure unfolding kinetics during lysozyme feedstock preparation. The shift in emission maximum gives a quantitative measure of the denatured feedstock quality. Unfolding temperature, protein concentration and the molar ratio of reducing reagent and protein showed to have a detectable impact on denaturation kinetics.

Two different denaturation protocols were used to prepared feedstock exhibiting differences in tryptophan fluorescence spectra. Both denatured protein species were applied in an equally designed refolding screening including 40 experiments. The denatured protein stock solution

showing a higher red shift in fluorescence emission resulted in lower yields of active and of soluble lysozyme. This discrepancy between the two different feedstocks was not due to residual active lysozyme after unfolding as it was not represented by a constant factor. Interestingly solubility is less affected from differences in protein denaturation than activity and the best experiments of both screening runs show a very similar parameter range. Consequently the formation of soluble lysozyme is less sensitive towards starting protein quality than is the formation of active lysozyme. Optimized parameters found by a screening with constant feedstock quality seem to be most suitable also if the feedstock preparation changes.

The impact of a controlled redox environment was a second parameter investigated during screening validation. Screening runs were performed with and without addition of DTT and GSSG consisting each of 40 experiments. The addition of redox components was observed to be superior to air oxidation as higher refolding yields could be achieved in all measured buffer systems. Additionally the absence of a controlled redox environment led to a completely non-systematic distribution of parameter values within the best refolding results. Without a correlation of screened parameter values and refolding yield a systematic optimization is impossible. The main reason for these unsatisfying results lies in slow oxygen transfer rates and the lack of HTS methods to control or at least measure dissolved oxygen in microtiter plates. Interestingly refolding samples without added redox components exhibited a higher variation in tryptophan fluorescence spectra which already hints at a higher molecular flexibility due to the lack of native disulfide bonds.

After the technical basis for refolding screening was provided two different methods for intelligent experimental design were investigated: a genetic algorithm and a classical factorial design. Optimization of up to five parameters (pH, concentration of NaCl, MgSO$_4$, DTT and GSSG) could be fully automated with the genetic algorithm including several rounds of experimentation, calculation of refolding yield and calculation of the next starting parameter matrix. A high data density is provided in the parameter range of interest close to the optimum. A full factorial design in the reduced parameter space close to the optimum is applied to confirm or refine the optimization. In our case the genetic algorithm did not lead to a global optimum for all parameters which could be revealed by the full factorial experiments. The algorithm is caught on a diagonal line between two parameters representing the molar ratio of oxidizing to reducing component or simply spoken the redox potential of the solvent. The linear recombination method of the genetic algorithm in combination with the chosen mutation rate likely accounts for these undesired effects. This knowledge offers the possibility

to adapt the GA parameters to the screening objective and points out the usefulness of an additional full factorial experiment.

During the development of refolding screening methods the importance of protein solubility in refolding processes became prominent. Consequently an available HTS tool for the determination of protein precipitation curves was further improved in terms of throughput to characterize the effect of pH, temperature and different additives like PEG, sorbitol, sucrose or Tween 20 on the solubility behaviour of lysozyme. Measurement of lysozyme solubility in additive buffer mixtures resulted in complex solubility surfaces demonstrating interdependent effects of both solvent components. As these interdependent effects are not simply additive, optimization of such systems should always be performed in the presence of both substances. An additional observation was that the pH dependency of lysozyme solubility correlated with the calculated net charge of the molecule. Decreasing net charge at pH values in proximity to the isoelectic point led to increasing protein interactions. Repulsive forces thus seem to play a major role in protein solubility.

A last part of this work dealt with the development of methods for optimization of solid phase refolding processes. An automated method for characterization of denatured lysozyme adsorption by isotherms and kinetics was established for resin screening in matrix-assisted refolding process development. A major problem was the liquid handling and mixing of viscous urea solutions, the resuspension of resin particles and the performance of a buffer exchange from denaturing to renaturing conditions. Different concentrations of denatured protein representing different points on the isotherm had to be mixed by pipetting the solution up and down several time. Mixing to homogeneity was not possible by orbital shaking only. Another problem is the resuspension of resin particles during adsorber loading which is only possible by orbital shaking in deepweel plates with a round bottom profile whereas flat bottom filter plates facilitate buffer exchange. As a consequence loaded adsorber particles had to be transferred to a filter plate with the tips of the liquid handling robot. Refolding is then initiated by a buffer exchange in the filter plates. The developed system allows for screening of refolding parameters like protein loading, residence time in refolding buffer and refolding and elution buffer composition.

Binding of denatured lysozyme on ion exchange resins turned out to be weaker than binding of native lysozyme and exhibited a BET shaped isotherm hinting at multilayer binding. The disintegration of charged patches by unfolding and the stretched molecular structure with solvent exposed hydrophobic residues presumably accounts for these observations.

An additional discovery was that refolded lysozyme amenable to elute from a strong cation exchange adsorber at high ionic strength is completely active whereas misfolded or aggregated protein can only be eluted in denaturing buffer conditions. These results hint at different adsorption mechanisms of the different protein species.

Future work should focus on the validation of the established high throughput screening method with a real inclusion body system. Inclusion body solubilisation parameters could be additionally performed and optimized on the robotic workstation.

Scale-up studies for the optimization of dilution refolding process parameters like the injection rate of denatured protein and mixing could be performed in a second development step.

An additional screening task for evolutionary optimization would include the addition of frequently used additives to improve solubility and folding in protein refolding. Combinatorial effects between additive concentration and other solvent components could be revealed and optimized up to our knowledge for the first time in parallel.

As redox components are expensive, relatively instable at room temperature and have to be removed, air oxidation is still an interesting alternative for refolding of disulfide bonded proteins though control of the redox potential is difficult. The concentration of divalent ions is as important for air oxidation as mass transfer from the gas to the liquid phase and the temperature. Process optimization on this field would give a deeper insight into the effects of these parameters and might lead to faster process design of economically superior processes without added disulfide components.

On behalf of on column refolding optimization on the robotic platform robotic column systems could be used to screen for parameters like urea gradient slope and elution flow velocity.

The automated determination of precipitation curves could be improved by acceleration and control of liquid evaporation and temperature using a vacuum evaporator. The detection of soluble aggregates could be performed by dynamic light scattering measurements already hinting at potential aggregation at early time points. This would improve the stability of solutions subsequently prepared below the solubility curve.

5 References

Ahamed, T., B.N.A. Esteban, M. Ottens, G.W.K. van Dedem, L.A.M. van der Wielen, M.A.T. Bisschops, A. Lee, C. Pham, and J. Thommes, *Phase behavior of an intact monoclonal antibody.* Biophysical Journal, 2007. **93**(2): p. 610-619.

Ahn, J.H., Y.P. Lee, and J.S. Rhee, *Investigation of refolding condition for Pseudomonas fluorescens lipase by response surface methodology.* Journal of Biotechnology, 1997. **54**(3): p. 151-160.

Alizadeh-Pasdar, N. and E.C.Y. Li-Chan, *Comparison of protein surface hydrophobicity measured at various pH values using three different fluorescent probes.* Journal of agricultural and food chemistry, 2000. **48**(2): p. 328-34.

Anderson, M.J., C.L. Hansen, and S.R. Quake, *Phase knowledge enables rational screens for protein crystallization.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(45): p. 16746-16751.

Anfinsen, C. (1972). "The formation and stabalization of protein structure." Biochemical Journal **128**(4): 737-749.

Armstrong, N., A. De Lencastre, and E. Gouaux, *A new protein folding screen: Application to the ligand binding domains of a glutamate and kainate receptor and to lysozyme and carbonic anhydrase.* Protein Science, 1999. **8**(7): p. 1475-1483.

Asherie, N., *Protein crystallization and phase diagrams.* Methods, 2004. **34**(3): p. 266-272.

Atha, D.H. and K.C. Ingham, *Mechanism of precipitation of proteins by polyethylen glycols - analysis in terms of excluded volume.* Journal of Biological Chemistry, 1981. **256**(23): p. 2108-2117.

Bard, J., K. Ercolani, K. Svenson, A. Olland, and W. Somers, *Automated systems for protein crystallization.* Methods, 2004. **34**(3): p. 329-347.

Bashford, D., *Macroscopic electrostatic models for protonation states in proteins.* Front Biosci, 2004. **9**: p. 1082-99.

Bensch, M., B. Selbach and J. Hubbuch, *High throughput screening techniques in downstream processing: Preparation, characterization and optimization of aqueous two-phase systems.* Chemical Engineering Science, 2007. **62**(7): p. 2011-2021.

Bensch, M., P.S. Wierling, E. von Lieres and J. Hubbuch, *High throughput screening of chromatographic phases for rapid process development.* Chemical Engineering & Technology, 2005. **28**(11): p. 1274-1284.

Berg A., Siudak A., von Lieres E. and J. Hubbuch, *Automated optimization of protein refolding processes with an evolutionary algorithm.* in preparation, 2008.

Berg, A., A. Kittelmann, and J. Hubbuch, *Solid phase refolding.* in preparation, 2008.

Berg, A., A. Kittelmann, and J. Hubbuch, *Development and characterization of an automated high throughput screening method for optimization of protein refolding processes* in preparation, 2008.

Berg, A., M. Schuetz, and J. Hubbuch, *Automated measurement of protein solubility to rapidly assess complex parameter interactions.* in preparation, 2008.

Bieri, O., G. Wildegger, A. Bachmann, C. Wagner, and T. Kiefhaber, *A salt-induced kinetic intermediate is on a new parallel pathway of lysozyme folding.* Biochemistry, 1999. **38**(38): p. 12460-12470.

Bloustine, J., V. Berejnov, and S. Fraden, *Measurements of protein-protein interactions by size exclusion chromatography.* Biophysical Journal, 2003. **85**(4): p. 2619-2623.

Bolen, D.W., *Effects of naturally occurring osmolytes on protein stability and solubility: issues important in protein crystallization.* Methods, 2004. **34**(3): p. 312-322.

Bondos, S.E. and A. Bicknell, *Detection and prevention of protein aggregation before, during, and after purification.* Analytical Biochemistry, 2003. **316**(2): p. 223-231.

Boschetti, E., *Advanced sorbents for preparative protein separation purposes.* Journal of chromatography, 1994. **658**(2): p. 207-236.

Bowen, R., *Understanding flux patterns in membrane processing for protein solutions and suspensions.* Trends in biotechnology, 1993. **11**(11): p. 451-460.

Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Analytical Biochemistry, 1976. **7**(72): p. 248-254.

Buswell, A.M. and A.P.J. Middelberg, *Critical analysis of lysozyme refolding kinetics.* Biotechnology Progress, 2002. **18**(3): p. 470-475.

Buswell, A.M., M. Ebtinger, A.A. Vertes, and A.P.J. Middelberg, *Effect of operating variables on the yield of recombinant trypsinogen for a pulse-fed dilution-refolding reactor.* Biotechnology and Bioengineering, 2002. **77**(4): p. 435-444.

Cacioppo, E., S. Munson, and M.L. Pusey, *Protein solubilities determined by a rapid technique and modification of that technique to a micromethod.* Journal of Crystal Growth, 1991. **110**(1-2): p. 66-71.

Cardamone, M. and N.K. Puri, *Spectrofluorometric assessment of the surface hydrophobicity of proteins.* The Biochemical journal, 1992. **282**: p. 589-593.

Carrio, M.M. and A. Villaverde, *Construction and deconstruction of bacterial inclusion bodies.* Journal of Biotechnology, 2002. **96**(1): p. 3-12.

Chang, L.Q., D. Shepherd, J. Sun, X.L. Tang, and M.J. Pikal, *Effect of sorbitol and residual moisture on the stability of lyophilized antibodies: Implications for the mechanism of protein stabilization in the solid state.* Journal of Pharmaceutical Sciences, 2005. **94**(7): p. 1445-1455.

Chen, G.Q. and E. Gouaux, *Overexpression of a glutamate receptor (GluR2) ligand binding domain in Escherichia coli: Application of a novel protein folding screen.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(25): p. 13431-13436.

Chen, T.M., H. Shen, and C.Y. Zhu, *Evaluation of a method for high throughput solubility determination using a multi-wavelength UV plate reader.* Combinatorial Chemistry & High Throughput Screening, 2002. **5**(7): p. 575-581.

Chi, E.Y., S. Krishnan, T.W. Randolph, and J.F. Carpenter, *Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation.* Pharmaceutical Research, 2003. **20**(9): p. 1325-1336.

Chiku, H., A. Kawai, T. Ishibashi, M. Takehara, T. Yanai, F. Mizukami, and K. Sakaguchi, *A novel protein refolding method using a zeolite.* Analytical Biochemistry, 2006. **348**(2): p. 307-314.

Chirica, G., J. Lachmann, and J. Chan, *Size exclusion chromatography of microliter volumes for on-line use in low-pressure microfluidic systems.* Analytical Chemistry, 2006. **78**(15): p. 5362-5368.

Choe, W.S., R. Nian, and W.B. Lai, *Recent advances in biomolecular process intensification.* Chemical Engineering Science, 2006. **61**(3): p. 886-906.

Clark, E., M, *Protein refolding for industrial processes.* Current opinion in biotechnology, 2001. **12**(2): p. 202-207.

Clark, E.D., D. Hevehan, S. Szela, and J. Maachupalli-Reddy, *Oxidative renaturation of hen egg-white lysozyme. Folding vs aggregation.* Biotechnology Progress, 1998. **14**(1): p. 47-54.

Cohn E.J. and E. J.T., *Proteins, amino Acids and peptides as ions and dipolar ions.* Reinhold Publishing, New York, 1943.

Cowan, R.H., R.A. Davies, and T.T.J. Pinheiro, *A screening system for the identification of refolding conditions for a model protein kinase, p38 alpha.* Analytical Biochemistry, 2008. **376**(1): p. 25-38.

Cowgill, R.W., *Fluorescence and Protein Structure .11. Fluorescence Quenching by Disulfide and Sulfhydryl Groups.* Biochimica Et Biophysica Acta, 1967. **140**(1): p. 37-47.

Cowieson, N.P., B. Wensley, P. Listwan, D.A. Hume, B. Kobe, and J.L. Martin, *An automatable screen for the rapid identification of proteins amenable to refolding.* Proteomics, 2006. **6**(6): p. 1750-1757.

Creighton, T.E., *Protein folding.* Biochemical Journal, 1990. **270**(1): p. 1-16.

Cromwell, M.E.M., E. Hilario, and F. Jacobson, *Protein aggregation and bioprocessing.* Aaps Journal, 2006. **8**(3): p. E572-E579.

Curatolo, L., B. Valsasina, C. Caccia, G.L. Raimondi, G. Orsini, and A. Bianchetti, *Recombinant human IL-2 is cytotoxic to oligodendrocytes after in vitro self aggregation.* Cytokine, 1997. **9**(10): p. 734-739.

Darcy, P.A. and J.M. Wiencek, *Estimating lysozyme crystallization growth rates and solubility from isothermal microcalorimetry.* Acta Crystallographica Section D-Biological Crystallography, 1998. **54**: p. 1387-1394.

Datta, S., B.K. Biswal, and M. Vijayan, *The effect of stabilizing additives on the structure and hydration of proteins: a study involving tetragonal lysozyme.* Acta Crystallographica Section D-Biological Crystallography, 2001. **57**: p. 1614-1620.

DeLucas, L.J., T.L. Bray, L. Nagy, D. McCombs, N. Chernov, D. Hamrick, L. Cosenza, A. Belgovskiy, B. Stoops, and A. Chait, *Efficient protein crystallization.* Journal of Structural Biology, 2003. **142**(1): p. 188-206.

Desai, A., C. Lee, L. Sharma, and A. Sharma, *Lysozyme refolding with cyclodextrins: structure-activity relationship.* Biochimie, 2006. **88**(10): p. 1435-1445.

Dill, K.A. and H.S. Chan, *From Levinthal to pathways to funnels.* Nature Structural Biology, 1997. **4**(1): p. 10-19.

Dill, K.A., *Dominant Forces in Protein Folding*. Biochemistry, 1990. **29**(31): p. 7133-7155.

Dobson, C.M., *Principles of protein folding, misfolding and aggregation*. Seminars in Cell & Developmental Biology, 2004. **15**(1): p. 3-16.

Ejima, D., K. Ono, K. Tsumoto, T. Arakawa, and Y. Eto, *A novel "reverse screening" to identify refolding additives for activin-A*. Protein Expression and Purification, 2006. **47**(1): p. 45-51.

Engelhard, M. and P.A. Evans, *Experimental investigation of sidechain interactions in early folding intermediates*. Folding & design, 1996. **1**(2): p. R31-R37.

Evans, P.A. and M. Engelhard, *Experimental investigation of sidechain interactions in early folding intermediates*. Folding & design, 1996. **1**(2): p. R31-7.

Ewing, F., E. Forsythe, and M. Pusey, *Orthorombic lysozyme solubility*. Acta Crystallographica Section D-Biological Crystallography, 1994. **50**: p. 424-428.

Fahey, M.S., D. Dawbarn, S.J. Allen, I.C. Paterson, and S.S. Prime, *Expression of recombinant extracellular domain of the type II transforming growth factor-ss receptor: Utilization in a modified enzyme-linked immunoabsorbent assay to screen TGF-ss agonists and antagonists*. Analytical Biochemistry, 2001. **290**(2): p. 272-276.

Farid, S.S., *Process economics of industrial monoclonal antibody manufacture*. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences, 2007. **848**(1): p. 8-18.

Fatouros, A., T. Osterberg, and M. Mikaelsson, *Recombinant factor VIII SQ - influence of oxygen, metal ions, pH and ionic strength on its stability in aqueous solution*. International Journal of Pharmaceutics, 1997. **155**(1): p. 121-131.

Fischer, B.E., *Renaturation of recombinant proteins produced as inclusion bodies* Biotechnology Advances, 1994. **12**(1): p. 89-101.

Forsythe, E.L., R.A. Judge, and M.L. Pusey, *Tetragonal chicken egg white lysozyme solubility in sodium chloride solutions.* Journal of Chemical and Engineering Data, 1999. **44**(3): p. 637-640.

Garcia-Ruiz, J.M., *Nucleation of protein crystals.* Journal of Structural Biology, 2003. **142**(1): p. 22-31.

Gast, K., D. Zirwer, and G. Damaschun. *Time-resolved dynamic light scattering as a method to monitor compaction during protein folding.* in *Workshop on Data Evaluation in Light Scattering of Polymers (LS '99).* 1999. Bad Schandau, Germany: Wiley-V C H Verlag Gmbh.

Geng, X.D. and C.Z. Wang, *Protein folding liquid chromatography and its recent developments.* Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences, 2007. **849**(1-2): p. 69-80.

Geng, X.D., B. Quan, Y.J. Zhang, L. Xiang, and D. Wu, *Refolding and purification of interferon-gamma in industry by hydrophobic interaction chromatography.* Journal of Biotechnology, 2004. **113**(1-3): p. 137-149.

Gosavi, R.A., T.C. Mueser, and C.A. Schall, *Optimization of buffer solutions for protein crystallization.* Acta Crystallographica Section D-Biological Crystallography, 2008. **64**: p. 506-514.

Goto, M., Y. Hashimoto, T. Fujita, T. Ono, and S. Furusaki, *Important parameters affecting efficiency of protein refolding by reversed micelles.* Biotechnology Progress, 2000. **16**: p. 1079-1085.

Gottschalk, U., *Bioseparation in antibody manufacturing: The good, the bad and the ugly.* Biotechnology Progress, 2008. **24**: p. 496-503.

Green, A.A., *Studies in the physical chemistry of the protein VIII. The solubility of hemoglobin in concentrated salt solutions. A study of salting out of proteins.* Journal of Biological Chemistry, 1931. **93**(2): p. 495-516.

Hansen, C.L., M.O.A. Sommer, and S.R. Quake, *Systematic investigation of protein phase behavior with a microfluidic formulator.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(40): p. 14431-14436.

Hansen, N., *On self-adaption in evolutionary strategies.* Studies in computational intelligence. Vol. 136. 2006, Berlin / Heidelberg: Springer Verlag. 31-57.

Heiring, C. and Y.A. Muller, *Folding screening assayed by proteolysis: application to various cystine deletion mutants of vascular endothelial growth factor.* Protein Engineering, 2001. **14**(3): p. 183-188.

Hevehan, D.L. and E.D. Clark, *Oxidative renaturation of lysozyme at high concentrations.* Biotechnology and Bioengineering, 1997. **54**(3): p. 221-230.

Ho, J.G.S., A.P.J. Middelberg, P. Ramage, and H.P. Kocher, *The likelihood of aggregation during protein renaturation can be assessed using the second virial coefficient.* Protein Sci, 2003. **12**(4): p. 708-716.

Hofmeister, F., *Zur Lehre von der Wirkung der Salze.* Arch. Exp.Pathol.Pharmakol., 1888. **24**: p. 1 - 16.

Hu, H.Q., T. Hale, X.Y. Yang, and L.J. Wilson, *A spectrophotometer-based method for crystallization induction time period measurement.* Journal of Crystal Growth, 2001. **232**(1-4): p. 86-92.

Igarashi, K., M. Azuma, J. Kato, and H. Ooshima, *The initial stage of crystallization of lysozyme, a differential scanning calorimetric (DSC) study.* Journal of crystal growth, 1999. **204**(1-2): p. 191-200.

Itzhaki, L.S., P.A. Evans, C.M. Dobson, and S.E. Radford, *Tertiary interactions in the folding pathway of hen lysozyme - kinetic studies using fluorescent probes.* Biochemistry, 1994. **33**(17): p. 5212-5220.

Jancarik, J., R. Pufan, C. Hong, S.H. Kim, and R. Kim, *Optimum solubility (OS) screening: an efficient method to optimize buffer conditions for homogeneity and crystallization of proteins.* Acta Crystallographica Section D-Biological Crystallography, 2004. **60**: p. 1670-1673.

Jolles, P., *Preparation and assay of enzymes.* Methods in Enzymology, 1962. **5**: p. 12-13.

Jones, D.B., M.H. Hutchinson, and A.P.J. Middelberg, *Screening protein refolding using surface plasmon resonance.* Proteomics, 2004. **4**(4): p. 1007-1013.

Kelley, B., *Very large scale monoclonal antibody purification : The case for conventional unit operations.* Biotechnology Progress, 2007. **23**: p. 995-1008.

Kerby, M.B., J. Lee, J. Ziperstein, and A. Tripathi, *Kinetic measurements of protein conformation in a microchip.* Biotechnology Progress, 2006. **22**(5): p. 1416-1425.

Kerns, E.H. and L. Di, *Pharmaceutical profiling in drug discovery.* Drug Discovery Today, 2003. **8**(7): p. 316-323.

Kiefhaber, T., R. Rudolph, H.H. Kohler, and J. Buchner, *Protein aggregation in vitro and in vivo - A quantitative model of the kinetic competition between folding and aggregation* Bio-Technology, 1991. **9**(9): p. 825-829.

Kreilgaard, L., L.S. Jones, T.W. Randolph, S. Frokjaer, J.M. Flink, M.C. Manning, and J.F. Carpenter, *Effect of Tween 20 on freeze-thawing- and agitation-induced aggregation of recombinant, human factor XIII.* Journal of Pharmaceutical Sciences, 1998. **87**(12): p. 1597-1603.

Kulkarni, S.K., A.E. Ashcroft, M. Carey, D. Masselos, C.V. Robinson, and S.E. Radford, *A near-native state on the slow refolding pathway of hen lysozyme.* Protein Science, 1999. **8**(1): p. 35-44.

Lakowicz, J., *On spectral relaxation in proteins.* Photochemistry and photobiology, 2000. **72**(4): p. 421-437.

Lanckriet, H. and A.P.J. Middelberg, *Continuous chromatographic protein refolding.* Journal of Chromatography A, 2004. **1022**(1-2): p. 103-113.

Lange, C., G. Patil, and R. Rudolph, *Ionic liquids as refolding additives: N '-alkyl and N '-(omega-hydroxyalkyl) N-methylimidazolium chlorides.* Protein Science, 2005. **14**(10): p. 2693-2701.

Langenhof, M., S.S.J. Leong, L.K. Pattenden, and A.P.J. Middelberg, *Controlled oxidative protein refolding using an ion-exchange column.* Journal of Chromatography A, 2005. **1069**(2): p. 195-201.

Lau, B.T.C., C.A. Baitz, X.P. Dong, and C.L. Hansen, *A complete microfluidic screening platform for rational protein crystallization.* Journal of the American Chemical Society, 2007. **129**(3): p. 454-455.

Lee, C.T., A.M. Buswell, and A.P.J. Middelberg, *The influence of mixing on lysozyme renaturation during refolding in an oscillatory flow and a stirred-tank reactor.* Chemical Engineering Science, 2002. **57**(10): p. 1679-1684.

Lee, Y.C. and D. Yang, *Determination of lysozyme activities in a microplate format.* Analytical Biochemistry, 2002. **310**(2): p. 223-224.

Leong, S.S.J. and A.P.J. Middelberg, *The refolding of different alpha-fetoprotein variants.* Protein Science, 2006. **15**(9): p. 2040-2050.

Li, M. and Z. Su, *Refolding human lysozyme produced as an inclusion body by urea concentration and pH gradient ion exchange chromatography.* Chromatographia, 2002. **56**(1-2): p. 33-38.

Li, M., G.F. Zhang, and Z.G. Su, *Dual gradient ion-exchange chromatography improved refolding yield of lysozyme.* Journal of Chromatography A, 2002. **959**(1-2): p. 113-120.

Lilie, H., E. Schwarz, and R. Rudolph, *Advances in refolding of proteins produced in E-coli.* Current Opinion in Biotechnology, 1998. **9**(5): p. 497-501.

213

Lin, J.L., R.C. Ruaan, and H.J. Hsieh, *Refolding of partially and fully denatured lysozymes.* Biotechnology Letters, 2007. **29**(5): p. 723-729.

Lin, L., J. Seehra, and M.L. Stahl, *High-throughput identification of refolding conditions for LXR beta without a functional assay.* Protein Expression and Purification, 2006. **47**(2): p. 355-366.

Loll, P.J., M. Allaman, and J. Wiencek. *Assessing the role of detergent-detergent interactions in membrane protein crystallization.* in *8th International Conference on Crystallization of Biological Macromolecules.* 2000. Sandestin, Florida: Elsevier Science

Lu, J., X.J. Wang, and C.B. Ching, *Batch crystallization of soluble proteins: effect of precipitant, temperature and additive.* Progress in Crystal Growth and Characterization of Materials, 2002. **45**(3): p. 195-205.

Machold, C., R. Schlegl, W. Buchinger, and A. Jungbauer, *Matrix assisted refolding of proteins by ion exchange chromatography.* Journal of Biotechnology, 2005. **117**(1): p. 83-97.

Mahler, H.C., R. Muller, W. Friess, A. Delille, and S. Matheus, *Induction and analysis of aggregates in a liquid IgG1-antibody formulation.* European Journal of Pharmaceutics and Biopharmaceutics, 2005. **59**(3): p. 407-417.

Mannall, G.J., N.J. Titchener-Hooker, H.A. Chase, and P.A. Dalby, *A critical assessment of the impact of mixing on dilution refolding.* Biotechnology and Bioengineering, 2006. **93**(5): p. 955-963.

McPherson, A., *Introduction to protein crystallization.* Methods, 2004. **34**(3): p. 254-265.

Middelberg, A., *Preparative protein refolding.* Trends in biotechnology, 2002. **20**(10): p. 437-443.

Middelberg, A.P.J., *Preparative protein refolding.* Trends in Biotechnology, 2002. **20**(10): p. 437-443.

Mikol, V. and R. Giege, *Phase diagram of a crystalline protein - determination of the solubility of concavalin a by a microquantitation assay* Journal of crystal growth, 1989. **97**(2): p. 324-332.

Misawa, S. and I. Kumagai, *Refolding of therapeutic proteins produced in Escherichia coli as inclusion bodies.* Biopolymers, 1999. **51**(4): p. 297-307.

Mishra, R., R. Seckler, and R. Bhat, *Efficient refolding of aggregation-prone citrate synthase by polyol osmolytes.* The Journal of Biological Chemistry, 2005. **280**(16): p. 15553-155560.

Nakazato, K., T. Homma, and T. Tomo, *Rapid solubility measurement of protein crystals as a function of precipitant concentration with micro-dialysis cell and two-beam interferometer.* Journal of Synchrotron Radiation, 2004. **11**: p. 34-37.

Pace, C., S. Trevino, E. Prabhakaran, and J. Scholtz, *Protein structure, stability and solubility in water and other solvents.* Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 2004. **359**(1448): p. 1225-1234.

Patil, G., R. Rudolph, and C. Lange, *In vitro-refolding of a single-chain Fv fragment in the presence of heteroaromatic thiols.* Journal of Biotechnology, 2008. **134**(3-4): p. 218-221.

Patro, S. and T. Przybycien, *Self-interaction chromatography: A tool for the study of protein-protein interactions in bioprocessing environments.* Biotechnology and bioengineering, 1996. **52**(2): p. 193-203.

Pekar, A. and M. Sukumar, *Quantitation of aggregates in therapeutic proteins using sedimentation velocity analytical ultracentrifugation: Practical considerations that affect precision and accuracy.* Analytical Biochemistry, 2007. **367**(2): p. 225-237.

Pohlheim, H., *Evolutionäre Algorithmen.* 2000, Berlin, Heidelberg, New York: Springer Verlag.

Qoronfleh, M.W., L.K. Hesterberg, and M.B. Seefeldt, *Confronting high-throughput protein refolding using high pressure and solution screens.* Protein Expression and Purification, 2007. **55**(2): p. 209-224.

Rahimpour, F., G. Mamo, F. Feyzi, S. Maghsoudi, and R. Hatti-Kaul, *Optimizing refolding and recovery of active recombinant Bacillus halodurans xylanase in polymer-salt aqueous two-*

phase system using surface response analysis. Journal of Chromatography A, 2007. **1141**(1): p. 32-40.

Raman, B., T. Ramakrishna, and C.M. Rao, *Refolding of denatured and denatured/reduced lysozyme at high concentrations.* Journal of Biological Chemistry, 1996. **271**(29): p. 17067-17072.

Retailleau, P., M. RiesKautt, and A. Ducruix, *No salting-in of lysozyme chloride observed at how ionic strength over a large range of pH.* Biophysical Journal, 1997. **73**(4): p. 2156-2163.

Richards, J.P., M.P. Stickelmeyer, D.B. Flora, R.E. Chance, B.H. Frank, and M.R. DeFelippis, *Self-association properties of monomeric insulin analogs under formulation conditions.* Pharmaceutical Research, 1998. **15**(9): p. 1434-1441.

Rinas, U. and L. Vallejo, *Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins.* Microbial cell factories, 2004. **3**(1): p. 11.

Roux, P., M. Ruoppolo, A.F. Chaffotte, and M.E. Goldberg, *Comparison of the kinetics of S-S bond, secondary structure, and active site formation during refolding of reduced denatured hen egg white lysozyme.* Protein Science, 1999. **8**(12): p. 2751-2760.

Royer, C.A., *Probing protein folding and conformational transitions with fluorescence.* Chemical Reviews, 2006. **106**(5): p. 1769-1784.

Santesson, S., E.S. Cedergren-Zeppezauer, T. Johansson, T. Laurell, J. Nilsson, and S. Nilsson, *Screening of nucleation conditions using levitated drops for protein crystallization.* Analytical Chemistry, 2003. **75**(7): p. 1733-1740.

Sapan, C.V., R.L. Lundblad, and N.C. Price, *Colorimetric protein assay techniques.* Biotechnology and Applied Biochemistry, 1999. **29**: p. 99-108.

Saridakis, E. and N.E. Chayen, *Systematic improvement of protein crystals by determining the supersolubility curves of phase diagrams.* Biophysical Journal, 2003. **84**(2): p. 1218-1222.

Scheich, C., F.H. Niesen, R. Seckler, and K. Bussow, *An automated in vitro protein folding screen applied to a human dynactin subunit.* Protein Science, 2004. **13**(2): p. 370-380.

Shire, S.J., Z. Shahrokh, and J. Liu, *Challenges in the development of high protein concentration formulations.* Journal of Pharmaceutical Sciences, 2004. **93**(6): p. 1390-1402.

Shukla, A.A., B. Hubbard, T. Tressel, S. Guhan, and L. D., *Downstream processing of monoclonal antibodies - Application of platform approaches.* Journal of Chromatography B, 2007. **848**(1): p. 28-39.

Sijwali, P.S., L.S. Brinen, and P.J. Rosenthal, *Systematic optimization of expression and refolding of the Plasmodium falciparum cysteine protease falcipain-2.* Protein Expression and Purification, 2001. **22**(1): p. 128-134.

Singh, S.M. and A.K. Panda, *Solubilization and refolding of bacterial inclusion body proteins.* Journal of Bioscience and Bioengineering, 2005. **99**(4): p. 303-310.

Sluzky, V., J.A. Tamada, A.M. Klibanov, and R. Langer, *Kinetics of insulin aggregation in aqueous solutions upon agitation in the presence of hydrophobic surfaces* Proceedings of the National Academy of Sciences of the United States of America, 1991. **88**(21): p. 9377-9381.

Susanto, A., K. Treier, E. Knieps-Gruenhagen, and J. Hubbuch, *High throughput screening for the design and optimization of chromatographic processes: Automated optimization of chromatographic phase systems.* Chemical Engineering and Technology, 2007. **32**(1): p. 140-154.

Tardieu, A., F. Bonnete, D.S. Finet, and D. Vivares, *Understanding salt or PEG induced attractive interactions to crystallize biological macromolecules.* Acta Crystallographica Section D-Biological Crystallography, 2002. **58**: p. 1549-1553.

Tessier, P.M., H.R. Johnson, R. Pazhianur, B.W. Berger, J.L. Prentice, B.J. Bahnson, S.I. Sandler, and A.M. Lenhoff, *Predictive crystallization of ribonuclease A via rapid screening of osmotic second virial coefficients.* Proteins-Structure Function and Genetics, 2003. **50**(2): p. 303-311.

Timasheff, S.N., *The control of protein stability and association by weak-interactions with water-how do solvents affect these processes* Annual Review of Biophysics and Biomolecular Structure, 1993. **22**: p. 67-97.

Tobbell , D.A., B.J. Middletona , S. Rainesb, M.R.C. Needhamb, I.W.F. Taylora, J.Y. Beveridgea, and W.M. Abbott, *Identification of in vitro folding conditions for Procathepsin S and Cathepsin S using fractional factorial screens.* Protein Expression and Purification, 2002. **24**(2): p. 242-254.

Tsumoto, K., D. Ejima, I. Kumagai, and T. Arakawa, *Practical considerations in refolding proteins from inclusion bodies.* Protein Expression and Purification, 2003. **28**(1): p. 1-8.

Tsumoto, K., M, *Practical considerations in refolding proteins from inclusion bodies.* Protein expression and purification, 2003. **28**(1): p. 1-8.

Ulrich, J. and M.J. Jones, *Industrial crystallization - Developments in research and technology.* Chemical Engineering Research & Design, 2004. **82**(A12): p. 1567-1570.

Valente, J.J., B.G. Fryksdale, D.A. Dale, A.L. Gaertner, and C.S. Henry, *Screening for physical stability of a Pseudomonas amylase using self-interaction chromatography.* Analytical Biochemistry, 2006. **357**(1): p. 35-42.

van den Berg, B., E.W. Chung, C.V. Robinson, P.L. Mateo, and C.M. Dobson, *The oxidative refolding of hen lysozyme and its catalysis by protein disulfide isomerase.* The EMBO Journal, 1999. **18**(17): p. 4794-4803.

Vanz, A.L.S., G. Renard, M.S. Palma, J.M. Chies, S.L. Dalmora, L.A. Basso, and D.S. Santos, *Human granulocyte colony stimulating factor (hG-CSF): cloning, overexpression, purification and characterization.* Microbial Cell Factories, 2008. **7**.

Vincentelli, R., S. Canaan, V. Campanacci, C. Valencia, D. Maurin, F. Frassinetti, L. Scappucini-Calvo, Y. Bourne, C. Cambillau, and C. Bignon, *High-throughput automated refolding screening of inclusion bodies.* Protein Science, 2004. **13**(10): p. 2782-2792.

Vink, M., K. Derr, J. Love, D.L. Stokes, and T. Ubarretxena-Belandia, *A high-throughput strategy to screen 2D crystallization trials of membrane proteins.* Journal of Structural Biology, 2007. **160**(3): p. 295-304.

Wang, F.W., Y.D. Liu, J.J. Li, G.H. Ma, and Z.G. Su, *On-colunm refolding of consensus interferon at high concentration with guanidine-hydrochloride and polyethylene glycol gradients.* Journal of Chromatography A, 2006. **1115**(1-2): p. 72-80.

Wang, T., S. John, S. Archuleta, and C.B. Jonsson, *Rapid, high-throughput purification of HIV-1 integrase using microtiter plate technology.* Protein Expression and Purification, 2004. **33**(2): p. 232-237.

Wang, W., *Protein aggregation and its inhibition in biopharmaceutics.* International Journal of Pharmaceutics, 2005. **289**(1-2): p. 1-30.

Wei, W., *Instability, stabilization, and formulation of liquid protein pharmaceuticals.* International journal of pharmaceutics, 1999. **185**(2): p. 129-188.

Wiendahl M., Voelker C., Husemann I., Krarup J., Staby A., Scholl S., and H. J., *A novel method to evaluate protein solubility using a high throughput screening approach.* Chemical Engineering Science, 2009.

Wiendahl, M., P.S. Wierling, J. Nielsen, D.F. Christensen, J. Krarup, A. Staby, and J. Hubbuch, *High throughput screening for the design and optimization of chromatographic processes - Miniaturization, automation and parallelization of breakthrough and elution studies.* Chemical Engineering & Technology, 2008. **31**(6): p. 893-903.

Willis, M.S., J.K. Hogan, P. Prabhakar, X. Liu, K. Tsai, Y.Y. Wei, and T. Fox, *Investigation of protein refolding using a fractional factorial screen: A study of reagent effects and interactions.* Protein Science, 2005. **14**(7): p. 1818-1826.

Wimmer, R., M. Olsson, M.T.N. Petersen, R. HattiKaul, S.B. Petersen, and N. Muller, *Towards a molecular level understanding of protein stabilization: The interaction between lysozyme and sorbitol.* Journal of Biotechnology, 1997. **55**(2): p. 85-100.

Wu, C.M., T.E. Molnar, R.E. McKean, and A. Spenlehauer, *Stabilizers against heat-induced aggregation of RPR 114849, an acidic fibroblast growth factor (aFGF).* International Journal of Pharmaceutics, 1998. **167**(1-2): p. 25-36.

You, L. and F.H. Arnold, *Directed evolution of subtilisin E in Bacillus subtilis to enhance total activity in aqueous dimethylformamide.* Protein Engineering, 1996. **9**(1): p. 77-83.

Yu, Z.H. and B. Li, *The effect of polyols on the reactivation of guanidium chloride-denatured arginine kinase from shrimp Feneropenaeus chinensis muscle.* Protein and Peptide Letters, 2003. **10**(2): p. 199-211.

Yutani, K., A. Yutani, A. Imanishi, and T. Isemura, *Mechanism of Refolding of Reduced Random Coil Form of Lysozyme.* Journal of Biochem istry, 1968. **64**(4): p. 449-455.

Zaccai, N.R., K. Yunus, S.M. Matth ews, A.C. Fisher, and R.J Falconer, *Refolding of a membrane protein in a microfluidics reactor.* European Biophysics Journa l with Biophysics Letters, 2007. **36**(6): p. 581-588.

Zhang, J and D.G Mus son, *Investigation of high-throughput ultrafiltration for the determination of an unbound compound in human plasma using liquid chromatography and tandem mass spectrometry with electrospray ionization.* Journal of Chrom atography B-Analytical Technologies in the Biomedical and Life Sciences, 2006. **843**(1): p. 47-56.

Zhang, J and X Liu, *Effect of protein-protein interactions on protein aggregation kinetics.* Journal of Chemical Physics, 2003. **119**(20): p. 10972-10976.

Zhu, D.W, A. Garneau, M. Mazumdar, M. Zhou, G X, and S.X Lin, *Attempts to rationalize protein crystallization using relative crystallizability.* Journal of Structural Biology, 2006. **154**(3): p. 297-302.

CURRICULUM VITAE

Zeppelinallee 57
60487 Frankfurt Main
Tel: +49 69 74730330
email : Annette_Berg@gmx.de

Annette Berg

EDUCATION

08/2005 - 11/2008	**Graduate student** at the Institute of Life Sciences, TU Karlsruhe (relocation of the group)
08/2005 - 02/2008	**Graduate student** at the Institute of Biotechnology 2, Research Centre Juelich
06/2004 - 06/2005	**Graduate student** at the Institute of Physiological Chemistry, LMU, Munich (PhD not proceeded)
04/2004	**Diploma** in Biotechnology at the TU Braunschweig, **overall mark:** excellent
05/2003 - 02/2004	**Diploma thesis** at the Institute of Microbiology and Genetics and at the Institute of Biophysics, TU Berlin and Charité Berlin
	"Expression, purification and characterization of Rhodopsinkinase from Saccharomyces cerevisiae"
	mark: 1.3 (excellent)
10/2002 - 12/2002	**Student research project** at the German Research Centre for Biotechnology (GBF), Braunschweig
	"Purification and refolding of BMP-2 from inclusion bodies"
	mark: 1.3 (excellent)
09/2000	**Intermediate Diploma** **mark:** 1.7 (good)
10/1998 - 04/2004	**Studies of Biotechnology at the TU Braunschweig**

WORK EXPERIENCE

11/2008 - now	**Laboratory Head** in Downstream Processing Development at Sanofi-Aventis, Frankfurt Main

Scientific Publications

Posters

Annette Berg, Jürgen Hubbuch
Development of an Automated High Throughput Screening Method for Optimization of Inclusion Body Protein Refolding
ISPPP 2008 (International Symposium on the
Separation of Proteins, Peptides and Polynucleotides), 22. - 24.09.2008, Baden Baden

Annette Berg, Jürgen Hubbuch
Critical Evaluation of High Throughput Screening Techniques for the Optimization of Proein Refolding Processes
International Symposium on Biothermodynamics, 21. - 22.02.2008, Frankfurt Main

Annette Berg, Jürgen Hubbuch
Critical Evaluation of High Throughput Screening Techniques for the Optimization of Proein Refolding Processes
GVC / DECHEMA Congress "Aufarbeitung biotechnologischer Produkte",
13. - 16.05.2007, Osnabrück

Annette Berg, Jürgen Hubbuch
Critical Evaluation of High Throughput Screening Techniques for the Optimization of Proein Refolding Processes
ISPPP 2006 (International Symposium on the
Separation of Proteins, Peptides and Polynucleotides), 17. - 20.10.2006, Innsbruck

Papers

Dejana Mokranjac, Annette Berg, Alexander Adam, Kai Hell
Association of the Tim14.Tim16 subcomplex with the TIM23 translocase is crucial for function of the mitochondrial protein import motor
Journal of Biological Chemistry, 282 (25) (2007) 18037 - 18045

Dejana Mokranjac, Martin Sichting, Dušan Popov-Čeleketić, Annette Berg, Kai Hell, Walter Neupert
The import motor of the yeast mitochondrial TIM23 preprotein translocase contains two different J proteins, Tim14 and Mdj2
Journal of Biological Chemistry, 2820 (36) (2005) 31608 - 31614

Annette Berg