Novel approaches in chromatographic purification process development for low concentrated proteins in complex mixtures

zur Erlangung des akademischen Grades eines
DOKTORS DER INGENIEURWISSENSCHAFTEN (Dr.-Ing.)

der Fakultät für Chemieingenieurwesen und Verfahrenstechnik des
Karlsruher Instituts für Technologie (KIT)

genehmigte
DISSERTATION

von
(M.Sc.) Frieder Kröner
aus Karlsruhe

Referent: Prof. Dr. Jürgen Hubbuch
Korreferent: Prof. Dr. Matthias Franzreb
Tag der mündlichen Prüfung: 20.09.2013

Acknowledgements

It is on my heart to pass awards to the following people,

Prof. Dr. Dr. Fritz Frimmel for giving me the opportunity to work on my PhD thesis at the Department for Biomolecular Separation Engineering at the Karlsruhe Institute of Technology.

Prof. Dr. Jürgen Hubbuch for the scientifical and personal support as supervisor of my PhD thesis. I also want to thank him for the nice and friendly working atmosphere at his department as well as giving me the freedom to independently develop my project.

Prof. Dr. Matthias Franzreb for taking the time to be the second referee of my thesis.

Dr. Richard Kneusel for organizing the collaboration with the Diarect AG (Freiburg) in a very friendly manner, but also for all the helpful hints and the scientific support.

Dr. Marcel Ottens for the nice and friendly way to welcome and supervise me during my research visit at the TU Delft.

Bianca Stolzenberger, Sarah Nagel and Pascal Baumann for sharing the office with a constantly grumbling, sometimes dishevelled looking PhD student, with a distinct interest in silly jokes and conversations.

Dr. Florian Dismer, Dr. Stefan Oelmaier for the scientific support and the fruitful discussions but also for the nice stays at conferences as well as the countless times of watching crappy “sneak preview” movies.

Dr. Anna Osberghaus for helping me to get started with the enormously complex tool called “MATLAB”, something that i would probably have never touched.

Dr. Sigrid Hansen for proof reading this thesis and sometimes joining in for an occasional cigarette in the late evenings.

Benjamin Maiser, Patrick Diederich, Natalie Schnabel, Katharina Lang and Jörg Kittelmann for getting friends and not only staying colleagues in the last few years.

Alex Hanke for being a very good student but also for all the nice evenings spent together in Delft, barbecuing, cooking, watching series or simply drinking belgium beer, while being philosophical.

Dennis Elsäßer for spending countless hours tinkering around with the UHPLC, while working on his diploma thesis.

All the members of the Biomolecular Separation Engineering department, which are not especially mentioned here for the nice working atmosphere, the scientific discussions, the coffee breaks and the nice evenings spent together.

Thanks to all my friends not mentioned here, especially to those not living in Karlsruhe for staying in contact even though seeing each other became a rare but still a very welcome event.

Thanks to all former an present members of the "Hirsch-WG" and even more to those of the "Kaiser-WG" for being such a nice social circle in the last few years.

And of course thank you to my family and especially my parents for raising me in the way they did. Without you i would have never developed enough endurance and idealism for being a scientist, while never loosing fascination with science.

Additionally, i want to acknowledge the following institutions,

The Federal Ministry of Education and Research (BMBF) for the financial support, making this project possible.

The Diarect AG in Freiburg for the collaboration, the financial and the scientific support.

The Karlsruhe House of Young Scientists (KHYS) for granting me a research travel scholarship for my research stay at the TU Delft.

Table of Contents

TABLE OF CONTENTS

TABLE OF CONTENTS

Zusammenfassung

Eine Vielzahl der Proteine des menschlichen Körpers werden mit verschiedensten Krank- heiten in Verbindung gebracht. Dies führt dazu, dass viele dieser Proteine als Biomarker für die Diagnose eines bestimmten Krankheitsbildes genutzt werden können. So z.B. Autoantikörper, welche im Falle einer Autoimmunerkrankung auftauchen und somit Biomarker für die jeweilige Autoimmunerkrankung sind. Durch die Detektion dieser, mittels der Verwendung des entsprechenden Autoantigens in diagnostischen Assays kann die jeweilige Autoimmunerkrankung diagnostiziert, sowie deren Verlauf überwacht werden. Proteinautoantigene werden biotechnologisch, mittels rekombinanter Expression, Zellkultur- und Extraktionstechniken hergestellt. Die Reinigung dieser Proteine gestaltet sich oft besonders schwierig, da die Proteine meist sehr niedrig konzentriert in äußerst komplexen Gemischen (z.B. Zelllysaten) vorliegen. In der vorliegenden Doktorarbeit wurden Lösungsansätze für die Schwierigkeiten, welche mit der Prozessentwicklung für die Reinigung dieser Proteine assoziiert sind, erarbeitet. Das Ziel der Arbeit war es durch die Entwicklung systematischer Screeningmethoden die Prozessentwicklung zur Reinigung von Autoantigenen deutlich zu vereinfachen und zu beschleunigen. Die Arbeit ist in drei Hauptteile und einen zusätzlichen Teil gegliedert:

I. Entwicklung einer systematischen Vorgehensweise zur Generierung von Puffersystemen für die pH Gradienten Ionenaustauschchromatographie (IAC).

II. Multi-dimensionale, chromatographische Analytikmethoden basierend auf pH- und Salzgradienten IAC zur Ermittlung physikochemischer Parameter für die Auslegung von Reinigungsprozessen für komplexe, biotechnologische Gemische mit niedrigen Produkttitern.

III. Entwicklung eines neuartigen Ansatzes zur Ermittlung mechanistischer Parameter in einem weiten pH Bereich, für die *in silico* Optimierung (SMA Modell [1]) chromatographischer Prozesse (IAC).

IV. Zusätzlich: Entwicklung und Anwendung einer hochauflösenden Methode basierend auf pH Gradienten IAC, zur Charakterisierung von PEGylierungs - Positions - Varianten von mono-PEGyliertem Lysozym.

Im ersten Teil der Arbeit ist die systematische Generierung von Puffersystemen, bestehend aus mehreren Puffersubstanzen, für die pH Gradienten IAC dargestellt. Hierbei wurde ein Ansatz entworfen die Zusammensetzung der verwendeten Puffersysteme *in silico* für deren Anwendung

zu optimieren. Um lineare, kontrollierbare pH Gradienten in der IAC zu erhalten, muss die Pufferkapazität eines Puffersystems so konstant wie möglich gehalten werden. Durch die Variation der Puffersubstanzen, sowie deren Konzentrationen kann die Abweichung der Pufferkapazität eines Puffersystems, von einem konstanten Wert, minimiert werden. Hierzu wurde eine MATLAB-Prozedur (Mathworks, Natick, MA, USA) zur Berechnung aller notwendigen Gleichungen, der Pufferkapazität, der Titrationskurve und der Ionenstärke eines Puffersystems bestehend aus mono-, di- und tri- basischen/protischen Substanzen, aufgesetzt. Um Puffersysteme mit möglichst konstanter Pufferkapazität zu erhalten wurde die Prozedur durch einen Non-linear Least Squares Algorithmus zur Minimierung der Abweichung der Pufferkapazität von einem konstanten Wert, erweitert. Im Anschluss wurden Puffersysteme mit konstanter Pufferkapazität (10 mmol*l^{-1}) über einen Bereich von 7.5 pH Einheiten für die pH Gradienten An- und Kathionenaustauschchromatographie (AAC / KAC) erstellt. Bei der Verwendung der generierten Puffersysteme (Mono Q / S 4.6/100 Säule, GE Healthcare, Uppsala, Sweden) wurden hochlineare pH Gradienten mit einer Linearität $R^2 > 0.99$ erzielt. Die Anwendbarkeit der Puffersysteme für die Bestimmung der Elutions-pH Werte von Proteinen mittels pH Gradienten IAC wurde im Anschluss gemäß den ICH Richtlinien [2], durch den Einsatz vier verschiedener Modellproteine, validiert. Die ermittelten Abweichungen der Wiederholbarkeits- und Präzisionsmessung für die Bestimmung der Elutions-pH Werte von Proteinen betrugen ΔpH < 0.1 pH Einheiten. Die validierte Methodik wurde im Anschluss genutzt um das Elutionsverhalten von 22 verschiedenen Modellproteinen in der pH Gradienten IAC zu untersuchen. Die Resultate zeigen klare Unterschiede zwischen den verschiedenen Methoden, der pH Gradienten AAC / KAC und der isoelektrischen Fokussierung (IEF). Diese resultieren aus der Wechselwirkung der Proteine in der IAC mit einer geladenen Oberfläche. Darüber hinaus wurde eine unterschiedliche Selektivität der beiden Ionenaustauschermodi beobachtet, welche durch die unterschiedliche Ladung und somit abgeänderte Wechselwirkung der Proteine mit dem Adsorber erklärt werden kann. Die gewonnenen Erkenntnisse bezüglich der Selektivität der unterschiedlichen Methoden für verschiedene Trennprobleme erleichtern deren Auswahl für spezifische, analytische Fragestellungen. Darüber hinaus wurde eine Möglichkeit geschaffen, durch ein einfaches und schnelles Verfahren Puffersysteme für die pH Gradienten IAC als Hilfsmittel zur Entwicklung chromatographischer Reinigungsprozesse [3], zu generieren.

Im zweiten Teil der Arbeit ist die Entwicklung und Validierung zweier multidimensionaler Trennansätze zur Charakterisierung biotechnologischer Rohextrakte mit niedrigen Produkttitern, mit dem Hintergrund der Prozessentwicklung für die Reinigung enthaltener Zielproteine, dargestellt. Dieser Teil kann in zwei Ansätze unterteilt werden: 1. Die multidimensionale Charakterisierung eines Zelllysates unter Einsatz von pH Gradienten IAC zur Auslegung eines einfachen Reinigungsprozesses; 2. Die multidimensionale Charakterisierung eines Zelllysates mithilfe von

Salzgradienten IAC zur Ermittlung mechanistischer Parameter für die *in silico* Simulation / Optimierung chromatographischer Prozesse.
Zunächst wird eine generelle Methodik vorgestellt, mittels pH Gradienten IAC, SDS PAGE und LC-MS basierter Proteinidentifikation ein Zelllysat zu charakterisieren, um im Anschluss die gewonnenen Informationen zur systematischen Auslegung eines Reinigungsprozesses zu nutzen. Das biotechnologische Rohextrakt wird hierzu zunächst durch pH Gradienten IAC (AAC pH 10.5 - 3 / KAC pH 4 - 11.5) aufgetrennt um spezifisch den Elutions-pH Wert des Zielproteins zu bestimmen. Danach wird der verwendete pH Bereich auf den Elutionsbereich des Zielproteins verringert (Elutions-pH +/- 1). Die resultierenden Fraktionen der Trennung des Rohextraktes durch die fokussierte pH Gradienten IAC werden zunächst aufkonzentriert (Lyophilisation / Resolubilisation) und dann mittels Gelelektrophorese weiter aufgetrennt. Die enthaltenen Proteine werden im Anschluss extrahiert und durch LC-MS basierte Proteinidentifikation identifiziert. Die identifizierten Proteine stellen die kritischen Verunreinigungen für die Reinigung des Zielproteins mittels IAC dar, da sie aufgrund ihrer elektrostatischen Eigenschaften dem Zielprotein ähnlich sind. Die erhaltenen Daten (Protein-ID, Molekulargewicht, Elutions-pH AAC / KAC) bezüglich des Zielproteins, sowie der kritischen Kontaminanten können anschließend zur systematischen Auslegung der Reinigung des Zielproteins mittels IAC verwendet werden. Die Anwendbarkeit des beschriebenen Ansatz wurde durch dessen Verwendung zur Charakterisierung eines Insektenzell / Baculovirus Lysat, mit einem enthaltenen Zielprotein (Nucleolin, GI 55956788, Fragment RRM1-4) demonstriert. Bei dem Protein Nucleolin handelt es sich um ein RNA-bindendes Protein, welches als Antigen für die Diagnose von *systemischem Lupus Erythematosus* eingesetzt wird. Die Elutions-pH Wert Bestimmung mittels zielproteinspezifischer Dot-Blot Analytik der Fraktionen ergab Elutions-pH Werte von 4.85 (AAC) und 10.44 (KAC, bei 200 mM NaCl). Der enorme Unterschied der beiden Elutions-pH Werte kann durch die positiv geladenen RNA-bindenden Domänen des Proteins erklärt werden, da diese eine stark anisotropische Ladungsverteilung auf der Proteinoberfläche und somit ein stark unterschiedliches Elutionsverhalten auf den beiden unterschiedlichen Adsorbertypen verursachen. Um die kritischen Verunreinigungen zu identifizieren wurde das Zelllysat mittels pH Gradienten IAC im Elutions-pH Bereich des Zielproteins (AAC pH 5.5 - 3.5 / KAC pH 9.5 - 11.5, 200 mM NaCl) fraktioniert, durch SDS PAGE weiter aufgetrennt und die enthaltenen Proteine abschließend mithilfe LC-MS basierter Proteinidentifikation identifiziert. Die hierbei gewonnenen Informationen (Protein-ID, Molekulargewicht, Elutions-pH Werte AEC/CEC) konnten erfolgreich zur Auslegung einer Zweischrittreinigung des Zielproteins mittels IAC genutzt werden. Das Zelllysat wurde dazu zunächst auf einen Kathionenaustauscher gebunden (SP Sepharose FF, pH 8.75, 200 mM NaCl) und im Anschluss mit einem Elutionsschritt eluiert (pH 8.75, 500 mM). Die Ionenstärke der Elutionsfraktion wurde nachfolgend mittels Ultrafiltration halbiert. Die Reinigung wurde dann durch Salzgradienten

AAC (Q Sepharose FF, pH 5.2, 0-500 mM NaCl) fortgesetzt. Durch die Zweischrittreinigung konnte bereits eine Reinheit des Zielproteins von 80% erreicht werden. Die erreichte Reinheit verdeutlicht klar den Nutzen des beschriebenen, analytischen Ansatzes um möglichst effizient Reinigungsprozesse für Proteine aus komplexen, biotechnologischen Extrakten zu entwerfen.
Im zweiten Ansatz ist eine multidimensionale Analytikmethode zur Bestimmung chromatographischer Retentionsvolumina von Einzelproteinen im komplexen Gemisch dargestellt. Die ermittelten Retentionsvolumina werden im Anschluss dazu benutzt mechanistische Parameter (SMA Modell [1]) für die Simulation / Optimierung chromatographischer Trennverfahren zu errechnen. Die Analytikmethode besteht aus einer Salzgradienten IAC, einem Entsalzungsschritt mittels Umkehrphasenchromatographie, einem Aufkonzentrierungsschritt und abschließender Analytik durch Hochdurchsatz-Kapillar-Gelelektrophorese (HT-CGE). Die geschilderte Methodik wurde dazu benutzt ein Insektenzell / Baculovirus Lysat mit einem enthaltenen Zielprotein zu analysieren. Hierbei wurden die Konzentrationen, die Molekulargewichte und die Retentionsvolumina der enthaltenen Einzelproteine ermittelt. Die Retentionsvolumina konnten durch die Rekonstruktion der Peaks (Gauss-Peak-Fitting) aus den erhaltenen Daten der multidimensionalen Trennung für vier unterschiedliche IAC Salzgradientenlängen bestimmt werden (5, 10, 20, 80 Säulenvolumina, MonoQ 4.6/100, pH 5.2, 0-500 mM). Anhand der vier ermittelten Retentionsvolumina jedes Einzelproteins, wurden die linearen SMA Parameter, die charakteristische Ladung ν und die Gleichgewichtskonstante K_{SMA} mittels des Formalismus von Shukla et al. [4] berechnet. Die erhaltenen Parameter wurden im Anschluss erfolgreich, durch den Vergleich errechneter mit experimentel ermittelten Retentionsvolumina, validiert. Der Vergleich resultierte in einer mittleren Abweichung von 5 % und einer maximalen Abweichung von 13.6 %. Darüber hinaus wurden die Chromatogramme dreier Proteine exemplarisch über das Transport-Dispersive-Modell berechnet. Der hierfür notwendige, nicht-lineare, sterische Faktor σ wurde für jedes Protein festgelegt ($\sigma = 30$), da dieser sich experimentell nicht aus einem Gemisch ermitteln lässt. Der Einfluss dieses Faktors ist in beschriebenem Fall jedoch vernachlässigbar, da die Konzentrationen der enthaltenen Proteine im linearen Bereich der Adsorptionsisothermen liegen und σ somit keinen Einfluss auf die resultierenden Chromatogramme hat. Die vorhergesagten Chromatogramme stimmen gut mit den experimentell ermittelten überein, dies bestätigt klar die Qualität der bestimmten Parameter. Die Resultate verdeutlichen, dass mittels einer multidimensionalen Analytikmethode schnell und im analytischen Maßstab die notwendigen Parameter für die *in silico* Optimierung einer chromatographischen Trennung aus einem komplexen Gemisch ermittelt werden können. Darüber hinaus kann eine solche Methodik auch zur kontinuierlichen Überwachung der Zusammensetzung biotechnologischer Rohextrakte verwendet werden.
Im dritten Teil der Arbeit wird ein neuartiger Ansatz zur Ermittlung von SMA Parametern für weite pH Bereiche vorgestellt. Die experimentelle Bestimmung solcher Parameter, für deren

Verwendung zur *in silico* Optimierung / Simulation von chromatographischen Trennprozessen, ist meist sehr zeit- und arbeitsaufwändig. Aus diesem Grund werden diese meist nur für einzelne, oder einige wenige verschiedene pH Werte bestimmt. Trotz alledem spielt der pH Wert eine essentielle, nicht zu vernachlässigende Rolle für die Entwicklung von Reinigungsprozessen über IAC. Daher wurde ein Ansatz entwickelt der es ermöglicht durch pH Gradienten IAC und einigen wenige isokratische Elutionsexperimenten die SMA Parameter mehrerer Proteine im Gemisch über einen gesamten pH Bereich (hier: pH 4 - 11.5) zu bestimmen. Die SMA Parameter werden dabei durch die Anpassung des existierenden SMA Modells [5] an die pH Gradienten Elution in der IAC, ermittelt. Da sich ν und K_{SMA} während des pH Gradienten kontinuierlich ändern, wird der Verlauf dieser durch die Verwendung geeigneter Funktionen limitiert. Der Verlauf von K_{SMA} wurde über eine inverse, sigmoidale Boltzmann Funktion beschrieben, während der Verlauf von ν am besten über eine inverse bi-sigmoidale Boltzmann Funktion beschrieben werden konnte. Die Grenzen der einzelnen Kurvenparameter wurden durch die Anpassung von 100.000 beispielhaften Proteintitrationskurven festgelegt. Für die drei Modelproteine, Cytochrom C, Lysozym und Ribonuklease A wurden die SMA Parameter durch acht pH Gradienten Elutionen bei Salzkonzentrationen von 10 - 350 mM NaCl und je vier isokratischen Elutionen (pH 5, 7, 9) auf einer Mono S 4.6/100 Säule ermittelt. Die erhaltenen Parameter konnten durch den Vergleich vorhergesagter, mit experimentel bestimmten Retentionsvolumina der Proteine in verschiedenen Elutionsmodi validiert werden. Die mittlere Abweichung betrug hierbei 8%. Anschließend wurde eine Monte-Carlo Simulation mit der Annahme einer 5%-igen Abweichung des pH Wertes und der Salzkonzentration durchgeführt, um zu überprüfen ob die Ursache der Abweichung der Retentionsvolumina durch einen falschen Ansatz zur Parameterbestimmung, oder durch kleine experimentelle Variationen hervorgerufen wurde. Nahezu alle Abweichungen der Retentionsvolumina befanden sich im Bereich der Monte-Carlo Simulation und können daher durch experimentelle Variationen erklärt werden. Die Parameter wurden im Anschluß dafür genutzt die chromatographische Auflösung, sowie die Robustheit aller Elutionsmodi der IAC zu untersuchen. Die Resultate bestätigten die Erwartungen, dass über pH Gradienten und isokratische Elution höhere chromatographische Auflösungen als bei der Salzgradienten Elution erreichbar sind, während die Robustheit deutlich geringer ausfällt. Die Auflösung der Salzgradienten Elution kann durch einen zusätzlichen pH Gradienten, ohne einen negativen Effekt auf die Prozessrobustheit, erhöht werden. Abschließend kann gesagt werden, dass mittels geringem, experimentellem und zeitlichen Aufwand die SMA Parameter mehrerer im Gemisch enthaltener Proteine für einen ganzen pH Bereich bestimmt werden konnten. Die gewonnenen Daten ermöglichen groß angelegte Optimierungs- / Robustheitsstudien, welche das Erlangen eines detaillierten Prozessverständnis durch *in silico* Studien ermöglichen.

Im letzten und zusätzlichen Teil der Doktorabeit ist eine Methodik zur analytischen Charakterisierung eines mono-PEGylierten Proteins durch pH Gradienten IAC dar-gestellt. Die PEGylierung von Proteinen ist eine häufig eingesetzte Technik zur Ver-besserung der pharmakokinetischen Eigenschaften von Proteintherapeutika. Die Charakterisierung PEGylierter Proteine gestaltet sich jedoch oft sehr schwierig, da hierfür analytische Methoden mit sehr hoher Auflösung notwendig sind. Aufgrund der hohen chromatographischen Auflösung für die Trennung von Ladungsvarianten eignet sich die pH Gradienten IAC. Um dies zu demonstrieren wurde 5- und 10 kDa mono-PEGyliertes Lysozym mit einem hochauflösenden pH Gradienten (pH 10.5 - 11.5 Mono S 4.6/100) aufgetrennt. Lysozym kann an sechs verschiedenen Lysinen PEGyliert werden, was zu ebenso vielen positionellen Isomeren bei mono-PEGyliertem Lysozym führt. Mithilfe der pH Gradienten IAC konnten fünf der Isoformen getrennt werden. Um die PEGylierungsstelle zu bestimmen wurden die pI Werte der einzelnen Isoformen anhand 3d-Struktur-basierter (PDB-ID: 132L) Berechnung [6] errechnet und im Anschluss mit den Elutions-pH Werten korreliert. Die Werte korrelierten mit einem Koeffizienten von $R^2 > 0.99$. Die hierdurch vorhergesagte Elutionsreihenfolge wurde durch die Identifikation der PEGylierungsstelle mittels tryptischem Verdau und MALDI-MS bestätigt. Die Resultate bestätigten eindrücklich die erhöhte chromatographische Auflösung der pH Gradienten IAC für die Separation von Proteinladungsvarianten. Die erhöhte Auflösung der pH Gradienten IAC könnte in Zukunft auch für die Analytik anderer Protein-Ladungsvarianten therapeutischer Proteine (Glykosylierung, Phosphorylierung ...) von Nutzen sein.

Es kann zusammengefasst werden, dass die Hauptproblematiken bezüglich der Prozessentwicklung für die Reinigung von Autoantigenen, die niedrigen Produkttiter, die Komlexität der Rohextrakte, sowie das Fehlen von Affinitätstechniken, durch den Entwurf verschiedener, systematischer Screeningexperimente zur Prozessentwicklung, erfolgreich verringert werden konnten. Durch die Anwendung dargestellter Screeningexperimente kann die Prozessentwicklung für die Reinigung dieser Proteine deutlich beschleunigt und vereinfacht werden. Darüber hinaus wurde eine Methodik zur systematischen Erstellung von Puffersystemen für die pH Gradienten IAC entwickelt. Die pH Gradienten IAC bewährte sich dabei nicht nur als Screeningmethodik für die Prozessentwicklung, sondern auch als Werkzeug zur Bestimmung von SMA Parametern und als Methode der Wahl für die Charakterisierung PEGylierter Proteine.

Abstract

Many proteins in the human body are associated with diseases, thus a large number of proteins has a potential of being used as a disease biomarker. Autoantibodies, which appear in case of an autoimmune disease are such disease biomarkers. By applying the corresponding antigens in diagnostic assays, these diseases can be diagnosed and monitored. Protein autoantigens, are usually produced biotechnologically, using recombinant expression, cell culturing and purification techniques. Purification of these proteins is often very challenging, as they are mostly contained at low concentrations in very complex matrices (e.g. cell lysates). The challenges associated with the process development for the purification of these proteins are treated in this PhD thesis. By introducing systematic screening approaches to rationally design chromatographic purification processes, downstream process development is simplified and accelerated. The thesis can be divided in three main parts and one additional:

I. Development of a systematic methodology, to generate buffer systems for pH gradient ion exchange chromatography (IEC).

II. Multi-dimensional, analytical, chromatographic screenings using pH and salt gradient IEC to obtain physicochemical parameters of single proteins contained in complex mixtures for purification process development.

III. A novel approach to obtain mechanistic parameters covering a wide pH range for *in silico* optimization (SMA model [1]) of protein purification, using IEC.

IV. Additional: High resolution pH gradient IEC for the characterization of positional PEGylation variants of mono-PEGylated lysozyme.

In the first part an *in silico* approach for the systematic generation of buffer systems consisting of multiple buffer substances for pH gradient IEC was successfully developed and experimentally validated. To generate buffer compositions for the formation of linear, controllable pH gradients in IEC, their buffer capacity has to be kept constant in the chosen pH range. By varying buffer substances used in the buffer composition as well as the single substance concentrations, the deviation of a composition's buffer capacity from a constant value can be minimized. A MATLAB (Mathworks, Natick, MA, USA) procedure was set up using all necessary equations for the calculation of the buffer capacity, the titration curve and the ionic strength of a buffer composition consisting of mono-, di- and tri- basic/protic substances. A non-linear least square algorithm was implemented to minimize the deviation of the composition's buffer capacity from

a constant value. Hereby, buffer compositions with constant buffer capacity (10 mmol*l^{-1}) for pH gradient anion or cation exchange (AEC / CEC) chromatography spanning 7.5 pH units, were generated. Applying these buffer systems to form pH gradients in IEC (Mono Q / S 4.6/100 column, GE Healthcare, Uppsala, Sweden) resulted in highly linear pH gradients with $R^2 > 0.99$. The applicability of the buffer systems for the analysis of proteins using pH gradient IEC was validated according to the ICH Guidelines [2] by analysing four model proteins. The determined intermediate precision and the repeatability for the proteins' elution-pH values both showed a ΔpH < 0.1 pH units. Validated methods were subsequently used to investigate the elution behaviour of 22 model proteins in pH gradient AEC / CEC. The results clearly showed that there are fundamental differences between IEC and isoelectric focusing (IEF) resulting from the proteins' interaction with a charged surface in IEC. Also differences in selectivity between the two ion exchange modes (AEC / CEC) orginating from the changed interaction of the proteins with the oppositely charged surface, were observed. These gained insights in the mechanistic fundamentals of the three separations, pH gradient AEC / CEC and IEF clearly motivate for a selective use of the methods for specific separation problems. Furthermore, a methodology was created to easily develop buffer systems for pH gradient IEC, which can be used as a very useful tool in chromatographic process development for the purification of proteins [3].

The second part of this thesis describes the development and validation of multidimensional, analytical methods to characterize crude biotechnological feedstocks with low product titers for accelerated purification process development. It can be further subdivided into two different approaches, one applying pH gradient IEC based fractionation for the rational layout of an early purification process and one applying salt gradient IEC to acquire mechanistic parameters for *in silico* prediction / optimization of chromatographic separations.

First, a general approach applying pH gradient IEC for the characterization of complex biotechnological feedstocks to rationally design IEC purification steps is demonstrated. The feedstock is analysed using long-range pH gradient IEC (AEC pH 10.5 - 3 / CEC pH 4 - 11.5) to specifically determine the elution-pH of the target protein. Afterwards, the range of pH gradient IEC is narrowed around the elution-pH (+/- 1 pH) of the target protein. The fractions resulting from the separation of the cell lysate using focused pH gradient IEC are further separated after a preconcentration step (Lyophilization / Resolubilization) using SDS-PAGE. Contained proteins are identified by tryptic digestion and LC-MS. Due to eluting close to the target protein, the identified proteins represent the major contaminants, difficult to remove by IEC because of having comparable electrostatic properties. Evaluation of the suggested approach was carried out by characterizing an insect cell / Baculovirus lysate, containing a target protein (Nucleolin, GI 55956788, protein fragment containing RRM1-4). Nucleolin is a RNA binding protein, used as a protein diagnostic for the autoimmune disease *systemic lupus erythematosus* (SLE). The

elution-pH values of the target protein, determined by identifying the pH gradient IEC fractions containing Nucleolin (Dot-Blot analytics), were 4.85 in AEC and 10.44 in CEC (at 200 mM NaCl). This surprisingly large difference is most likely reasoned by the fact, that the four positively charged RRM1-4 RNA-binding domains of Nucleolin cause a strong anisotropic charge distribution on the protein surface and therefore a strongly differing elution behaviour between both chromatographic modes. The cell lysate was then analysed by focused pH gradient IEC (AEC pH 5.5 - 3.5 / CEC pH 9.5 - 11.5, 200 mM NaCl), SDS PAGE and LC-MS. Hereby, all critical impurities were identified. The obtained information was used to rationally lay out a simple two-step purification applying IEC. The cell lysate was captured on a CEC adsorber (SP Sepharose FF, pH 8.75, 200 mM NaCl) and eluted by an elution step (500 mM NaCl, pH 8.75). Ionic strength of the eluted sample was afterwards reduced by one half, using ultrafiltration. The sample was further purified using salt gradient AEC (Q Sepharose FF, pH 5.2, 0-500 mM NaCl) with salt gradient elution. A purity of 80% was achieved for Nucleolin, a very good value for such an early developmental stage. The achieved purity clearly showed how well the described analytical approach to characterize a complex biotechnological feedstock could be applied for purification process development.

The second multidimensional fractionation technique was established to track retention behaviours of single proteins in complex mixtures with low product titers for the estimation of mechanistic parameters (SMA model [1]) to simulate / optimize chromatographic separation *in silico*. A fast, multi-dimensional fractionation approach using salt gradient IEC, reversed phase desalting, sample concentrating and high-throughput capillary gel electrophoresis (HT-CGE) was established to analyse cell lysates containing target proteins. Using the multidimensional separation method an insect cell/Baculovirus lysate containing a target protein was analyzed. The retention volumes of contained, single proteins were determined by reconstructing single protein chromatograms (Gaussian peak fitting) from the data obtained by the multidimensional analysis. The analysis was performed for salt gradient separations with four different gradient lengths (5, 10, 20, 80 CV, Mono Q 4.6/100, pH 5.2, 0-500 mM). Also single protein concentrations and sizes were acquired from the analysis. The obtained retention volumes were further used to estimate the proteins' linear SMA parameters, the characteristic charge ν and the equilibrium constant K_{SMA} using the formalism developed by Shukla et al. [4]. The obtained values were successfully validated by comparing predicted retention volumes with experimentally derived ones. Comparison resulted in a mean deviation of 5 % and a maximal deviation of 13.6 %. Also chromatograms for three contained proteins were exemplarily predicted using the transport-dispersive model. The necessary non-linear SMA parameter σ, the steric factor, was defined ($\sigma = 30$), because its determination is not possible from mixtures. However, its influence is neglectable as the concentrations of all contained proteins were in the linear range. Predicted chromatograms were

highly comparable to experimentally determined chromatograms, underlining the quality of the obtained parameters. The results clearly demonstrate that the described multi-dimensional fractionation approach can be used to rapidly obtain parameters usable for *in silico* optimization of chromatographic procedures, which can dramatically decrease the amount of time needed for purification process development. Furthermore, it can also be used as an analytical tool to monitor the composition of biotechnological feedstocks also enabling the prediction of possible consequences on the purification process in case of a changed composition.

In the third part of this work an experimental approach to obtain SMA parameters covering a broad pH range, is demonstrated. Parameter estimation is generally a very time consuming procedure and mostly parameters are only determined for several pH values and not for a complete pH range. Nevertheless, pH is a crucial parameter in protein separations by ion exchange chromatography. Therefore, an approach applying a set of pH gradient and isocratic elution operations was developed to estimate SMA parameters in a wide pH range (4 - 11.5). SMA parameters are obtained from the acquired data by extending the applicability of the existing SMA model [5] to protein elution in pH gradients and by adding certain constraints to the parameter estimation. In pH gradient elution the two linear SMA parameters K_{SMA} and ν are constantly changing in a rational way. K_{SMA} was best described using an inverse sigmoidal Boltzmann function, while ν is following an inverse bi-sigmoidal trend. Boundaries for the function parameters were defined by fitting 100.000 randomly generated protein titration curves. The estimation of the SMA parameters was performed for three model proteins, cytochrome C, lysozyme and ribonuclease A. Eight pH gradient elutions (pH 4 - 11.5) at salt concentrations from 10 - 350 mM and four isocratic elution experiments at three different pH values (5, 7, 9) were performed using a Mono S 4.6/100 column. From the resulting data a set of SMA parameters was determined. Validation of these parameters was carried out by comparing predicted retention volumes with experimentally determined ones, of the single proteins in different elution modes, isocratic, salt gradient, pH + salt gradient. The mean deviation of the predicted retention volumes from the experimentally determined retention volumes is 8 %. A Monte-Carlo simulation, assuming a 5 % normally distributed error on pH and salt concentration for the parameter estimation was performed to access if this error was caused by a general problem with the parameter estimation or by simple experimental variations. Nearly all variations of the retention volumes were in the range of the Monte-Carlo simulation, thus it can be concluded that the estimated parameters are accurate within the margin of expectable deviations. The parameter set was subsequently used to calculate chromatographic resolutions and robustness for all IEC elution modes. The results confirmed that isocratic and pH gradient elution reach higher resolution values than salt gradient IEC, while robustness was much lower. The combination of salt and pH gradient elution increased chromatographic resolution without a negative effect on process robustness. In con-

clusion the results showed impressively, that applying a set of pH gradient and isocratic elution experiments as well as related computations for the determination of SMA parameters resulted in an impressive and very useful amount of data applicable for *in silico* process optimization and robustness studies.

PEGylation is a frequently used technique to improve the pharmacokinetics of protein therapeuticals. However, characterization of PEGylated proteins is often very challenging as separation of variants needs highly resolving analytical methods. The additional part of the thesis demonstrates an application of pH gradient IEC for the analysis of positional isoforms of PEGylated proteins. mono-PEGylated lysozyme with 5 and 10 kDa PEG was analysed by pH gradient IEC (pH 10.5 - 11.5) using a Mono S 4.6/100 cation exchange column. Lysozyme has six possible PEGylation sites (lysine residues) resulting in six positional isoforms, of which five isoforms were successfully separated. To identify lysozyme isoforms the pI of each positional isoform was calculated using 3d-structure (PDB-ID: 132L) dependent pI calculation [6]. For each positional isoform the corresponding, PEGylated lysine residue was neutralized for pI calculation. Resulting pI values were correlated with the experimentally determined elution-pH values, showing a correlation coefficient of $R^2 > 0.99$. Predicted elution order of the positional isoforms was successfully validated by tryptical digestion of collected fractions and subsequent determination of the PEGylated peptide by MALDI-MS. With the demonstrated chromatographic method the evaluation of PEGylation reaction kinetics for lysozyme is possible. Furthermore, pH gradient IEC showed its superior resolution for the separation of protein charge variants. Therefore, pH gradient IEC might also be interesting for the chromatographic analysis of other charge variants e.g. differently glycosylated, phosphorylated ..., therapeutic proteins.

It can be concluded that the main challenges in purifying autoantigens from biotechnological feedstocks, the low expression levels, the complexity of the biotechnological feedstock and the lack of effective affinity techniques for purification are successfully addressed by designing suitable screening experiments to obtain the necessary information to rationally lay out the purification steps. Applying these screening procedures will lead to a considerable acceleration in downstream process development for proteins contained in very complex mixtures at a low titre. Furthermore, pH gradient IEC proved to be useful not only as a screening tool for purification process development but also for the determination of SMA parameters covering a wide pH range as well as the analysis of protein charge variants.

Introduction

Proteins are the functionary components of organisms carrying out the various functions encoded in their related genes. Therefore proteins bear a multitude of different functions e.g. enzymatic-, regulatory-, structural-, transport functions and many more. The number of protein coding genes in humans ranges from 20.000 to 40.000 [7–9]. As this is only the number of protein coding genes, the number is supposed to be much higher, due to alternative splicing, post-translational modification and complex formation. Many of those proteins are associated with specific diseases due to e.g. mutations, false expression levels, defective regulation and many more. Therefore, a large number of proteins has a potential of being used as a protein therapeutic or as a potential disease biomarker. Reasoned by the large amounts of possible applications the market for biopharmaceuticals was constantly growing in the last few years and is expected to grow to 167 billion US Dollars by 2015 [10].

Biopharmaceuticals can be classified in four different classes [11]: **Group I:** protein therapeutics with enzymatic and regulatory activity; **Group II:** protein therapeutics with special targeting activity, e.g. monoclonal antibodies (mAbs); **Group III:** protein vaccines; **Group IV:** protein diagnostics. Group I-III, especially group II play the most important role in the market for biopharmaceuticals, but also the market for protein diagnostics (Group IV) is expected to reach a financial volume of 19.8 billion US Dollars by 2015 [12]. Biopharmaceutical proteins are mostly, recombinantly produced using various expression systems, e.g. bacteria, yeast, insect cells, mammalian cells or even transgenic plants (Upstream processing). After recombinant production of the target protein, purification is carried out using several protein separation technologies (Downstream processing). The requirements for purity and quality of the final product, which have to be met, are mostly very strictly defined. In the last few years, due to advances in recombinant technologies and fermentation techniques the major costs of the whole production process for biopharmaceuticals shifted to the downstream processing [13–15]. Also the increasing interest in more complex proteins, as well as the market entry of biosimilars are posing further challenges in protein downstream process development. This results in a growing demand for faster, more cost efficient methodologies to develop highly understood downstream processes.

A specifically challenging case is the process development for the purification of protein diagnostics. Mostly only low amounts of a target protein are required but the variety of diseases, which can be diagnosed using protein diagnostics, demands for a very diverse product portfolio. Therefore, the scientific focus in the production of protein diagnostics is on accelerating the development of time- and cost-efficient downstream processes for such molecules. The facts that many of those molecules are hardly produced recombinantly and only with low titres, as well as

the lack of affinity purification steps, are further complicating downstream process development. This results in an especially strong demand for new, systematic approaches to rapidly develop purification processes for protein diagnostics.

1 Purification of proteins

1.1 Overview

Each protein has individual physicochemical properties like e.g. surface charges, hydrophobicity, molecular size. To purify proteins from mixtures the differences between the single molecules' properties are utilzed to achieve separation. Several separation technologies, making use of different physicochemical properties, are applicable to purify proteins from mixtures. An overview of the separation technologies and the related properties is shown in Tab. 1

Table 1: Purification of proteins: Technologies, methods and properties

Technology	**Method**	**Physicochemical parameter**
Chromatographic methods	IEC	Electrostatic properties
	HIC	Hydrophobicity
	SEC	Molecular size
	AC	Affinity
	RPC	Hydrophobicity
Non-chromatographic methods	ATPS	Solubility
	Filtration	Molecular size
	Precipitation	Solubility
	Crystallization	Solubility

IEC Ion Exchange Chromatography; **HIC** Hydrophobic Interaction Chromatography; **SEC** Size Exclusion Chromatography; **AC** Affinity Chromatography; **RPC** Reversed Phase Chromatography; **ATPS** Aqueous Two-Phase Systems;

Purification processes can mostly be divided in four stages [16]: **A**: Preparation, Extraction, Clarification; **B**: Capture; **C**: Intermediate purification; **D**: Polishing. Depending on the characteristics of the target protein, the contaminants and the properties of the biotechnological feedstock, different methodologies are sequentially combined to achieve the required purity of the target protein.

To develop each applied single unit operation many process parameters have to be chosen well, thus a large design space has to be screened. Purification process development for proteins is often very challenging as biotechnological feedstocks are mostly very complex mixtures containing many different impurities, some of which have very similar physicochemical properties to the target protein. Nevertheless, the regulatories for purity and quality of pharmaceutical proteins are very strictly defined. A further complicating fact is the sensitivity of proteins to denaturation, aggregation, proteolytic degradation, etc.. Because of the risk of denaturation as well as the complexity of the feedstock, chromatographic methods are still the most applied separation technique for the purification of proteins, due to their high selectivity and the mild separation conditions [17], even though they are mostly very cost-intensive. Reasoned by all the difficulties associated with the purification of proteins downstream process development is very often economically and technically challenging.

1.2 Purification process development

The main challenge in purification process development for proteins is, that mostly little or nothing is known about the composition of the biotechnological feedstock as well as the physicochemical properties of the target and the impurities. The approaches facing that challenge can be classified in five classes [17]: Heuristic (knowledge-based) approaches, experimental approaches, platform approaches, model-based methods and hybrid methods.

Heuristic approaches: Purification process development is carried out by making use of expert knowledge, experience derived from past processes and several trial-and-error experiments. Even though such an approach mostly ends up in non-optimal purification processes, it is still a common way to develop purification processes for proteins.

Experimental approaches: Systematic, experimental approaches to obtain relevant parameters for purification process development. The experiments are mostly HTS (**H**igh-**T**hroughput-**S**creening) or HTE (**H**igh-**T**hroughput-**E**xperimentation) based. Often robotic platforms are used to perform a large amount of miniaturized experiments in a short time-scale. To reduce the amount of experiments needed, these approaches are mostly done applying DoE (**D**esign **o**f **E**xperiments) strategies. Time and costs are strongly reduced by applying these approaches for protein purification process development [18].

Platform approaches: A special case of process development for the purification of highly uniform proteins, e.g. mAbs. An established process for the purification of a protein is used as a template for the generation of purification processes for different proteins from the same molecular class. A typical purification platform for the purification of mAbs consists of two to three sequential chromatographic steps, Protein A affinity chromatography, followed by polishing ba-

sed on AEC/CEC and sometimes HIC [17, 19]. Platform process based purification development does have the significant advantages of being straightforward, simple, cost efficient and less time consuming.

Model-based methods: Physicochemical models are applied to predict / optimize separation behaviours of the different components *in silico*. The major challenge in model-based process development is the acquisition of the necessary physicochemical parameters of the proteins. Process development using mechanistic models is further complicated by the fact that the accuracy of the prediction directly depends on the quality of these parameters. The parameters are mostly estimated experimentally using defined screening experiments, but there are also approaches to predict them using protein structures in QSAR (**Q**uantitative **S**tructure **A**ctivity **R**elationship) or MD (**M**olecular **D**ynamics) simulations [20, 21]. However, process development based on the usage of mechanistic models for process simulation has significant advantages, like low developmental time, costs and increased process understanding due to the possibility to carry out a large amount of experiments *in silico*. Nevertheless, model-based process development is often not possible as the determination of the parameters is neglected by the unavailability of rational methodologies to obtain these from complex mixtures.

Hybrid methods: A combination of experimental and model-based methods, e.g. the determination of mechanistic parameters for *in silico* process development using HTS or HTE. Protein mixtures are systematically characterized using experimental screenings to obtain the relevant physicochemical parameters for process simulation and optimization as demonstrated by Nfor et al. [22]. Hybrid methods combine the advantages of the large amounts of obtained information from the biotechnological feedstock using HTS or HTE but also the advantages of model-based process development, low developmental time and costs as well as increased process understanding. Anyhow, clearly defined experimental procedures to systematically acquire high quality parameters for process simulation of the single components in complex mixtures are very rare.

Industrial process development for the purification of proteins needs to lead rapidly to optimal, but also safe and robust processes, fulfilling the requirements on protein purity and quality. Especially for very complex biotechnological feedstocks, e.g. cell lysates, containing low titer expressed target proteins, process development can be very time consuming and economically challenging.

2 The production of protein diagnostics

2.1 Protein diagnostics

Besides genetic biomarkers, antigens play an essential role in clinical diagnostics of diseases. One big group of applied protein diagnostics are antigens, being used as binding partners for the detection of present serum antibodies, associated with the respective disease. Mostly, the proteins are applied in ELISA (Enzyme-Linked-Immunosorbent-Assay) or microarrays for disease diagnosis. Examples are the diagnosis of *Lyme borreliosis*, HIV, Malaria and many more, by ELISA. A major field of application is the diagnosis of autoimmune diseases, which emerge in case of a malfunctioning immune system attacking endogenous structures due to the false recognition of own proteins as pathogens. Autoimmune diseases are related with genetic, environmental and epigenetic factors [23–25]. By determining the concentration of related autoimmune antibodies, e.g. using ELISA, autoimmune diseases can be prognosed, diagnosed and in some cases monitored. The availability of well established diagnostic assays makes it possible to identify the form and the progression of the disease, based on the results decisions on possible therapeutic intervention can be made. To establish such diagnostic assays the proteins containing the relevant epitopes have to be identified and produced in a pure form. Recombinant expression of those proteins is very often challenging, as the right conformational structure, post-translational modification [26] and complex formation with DNA, RNA or other proteins (e.g. small ribonuclear particles, snRNP, [27]) are necessary to properly bind the related autoimmune antibodies. Some proteins / protein complexes like the U1-snRNP complex for the diagnosis of *Systemic lupus erythematosus* (SLE) have not been recombinantly expressed so far. Also the purification of these proteins is consequently very challenging as the feedstocks are very complex mixtures containing only low amounts of the target protein.

2.2 Recombinant expression of diagnostic autoantigens

After isolating the gene-of-interest it is cloned into the appropriate vector for the applied expression organism. Most autoantigens are produced using, *Escherichia coli*, insect–cell / Baculovirus or mammalian expression systems, e.g. CHO-cells (**C**hinese **H**amster **O**vary cells). The expression system is chosen due to the fact if a functional protein at preferably high expression level can be obtained. The molecules activity in immune diagnostics is directly influenced by its structural authenticity, which depends on the correct folding as well as post-translational modifications. Reasoned by these structural requirements an expression system able to correctly fold and modify the target protein has to be chosen. Therefore, the expression of protein diagnostics in *E.coli* is often not possible and an eukaryotic system, with the disadvantage of a lower

expression rate, has to be chosen. Currently, the most frequently used expression system is the insect–cell / Baculovirus expression system [28], due to the facts that it allows production of post-translationally modified proteins at a relatively high protein expression rate. Mammalian cell expression systems are only used, if proteins are to be produced, which can not be expressed stable in insect cell expression systems e.g. due to the need of specific cofactors for correct folding. The major drawback of mammalian cell expression systems is that they require huge efforts until being fully established, yielding high protein production rates.
Further challenges in the recombinant production of autoantigens are insolubility of the overexpressed proteins, aggregation, instability as well as host cell toxicity. The latter is mostly caused by the function of the expressed protein. Also most proteins are expressed intracellularly resulting in the need of cell lysis for protein isolation. The low expression rates of the proteins, the need of cell lysis result in very complex mixtures and thus in challenging purification process development.

2.3 Purification of diagnostic autoantigens

To use diagnostic autoantigens for disease diagnosis in immunoassays, all contaminants, which might result in false positive results have to be removed. Recombinantly expressed autoantigens are mostly purified using attached hexa-histidine tags by applying IMAC (**I**mmobilized-**M**etal-**A**ffinity-**C**hromatography). IMAC is making use of the high affinity and binding strength of hexa-histidine tags to divalent cations immobilized on the chromatographic resin. Using IMAC has the advantages of very mild chromatographic conditions, while offering the chromatographic selectivity of an affinity purification step. After using IMAC for capturing and purifying the protein, mostly one additional polishing step, involving IEC or HIC, is sufficient to remove the leftover contaminants. Purity rates of a percentage larger than 90% are mostly sufficient for the protein's intended use.
However, sometimes no hexa-histidine tag can be effectively attached to the expressed protein, due to the loss of the proteins diagnostic activity because of inauthentic protein folding or due to inaccessible, e.g. intramolecular localisation of the tag. In the case that a protein can not be produced recombinantly or only without having a hexa-histidine tag attached, purification is getting very challenging as the biotechnological feedstocks are highly complex and the target proteins are mostly only expressed with low titres. Also purification process development is getting very time-consuming in such cases and sometimes no purification process leading to acceptable results can be reached. The following table (Tab. 2) is showing data on developmental time scales for purification process development of several antigens from the industrial partner (Diarect AG, Freiburg, Germany).

Table 2: Data on purification process development for autoimmune antigens

Autoimmune disease	Autoantigen	Time and yield
Celiac disease	tissue Transglutaminase (tTG)	24 months, 15 mg/l cell culture
SLE	Ro/SS-A 60 kDa	33 months, 5 mg/l cell culture
Autoimmune thyroiditis	Thyroid peroxidase (TPO)	15 months, 2 mg/l cell culture
SLE	U1-snRNP 68 kDa	35 months, 4 mg/l cell culture
Autoimmune thyroiditis	Thyreoglobuline	not possible

Tab. 2 shows the enormous increase in developmental time for purification process development for specific proteins, which are expressed at low titres and without an attached hexa-histidine tag. For such proteins, purification process development is accompanied by a very long time-to-market and enormous costs, especially in case of proteins, which have to be derived from native sources.

2.4 Research objectives

The previously described difficulties concerning downstream process development for the purification of autoantigens, the low product titers, the high feedstock complexity, high product diversity and the lack of affinity techniques, clearly demand for a faster, more rational methodology to accelerate purification process development for these proteins. One approach to address the described challenges is the concept of a developmental platform, a series of screening experiments to characterize the biotechnological feedstock with the aim to obtain the relevant physicochemical parameters of the contained components to rationally lay out a purification process based on that information. Such approaches demand for clearly defined, systematical, analytical screening experiments enabling the described characterization of biotechnological feedstocks at a very early stage in purification process development. Approaches adressing comparable challenges by employing chromatographic screening experiments were previously described in publications by Ahamed and Nfor [3, 22]: A multidimensional fractionation approach applying salt, pH gradient IEC and SDS PAGE as a 2nd analytical dimension to systematically obtain physicochemical properties of the contained proteins, by tracking their elution behaviour in the mixture.
In these previous publications pH gradient IEC was successfully applied to determine the optimal pH value under given salt concentration for the layout of a purification process for a monoclonal antibody from a cell culture supernatant using IEC. However, the formation of con-

trollable, linear pH gradients in IEC with preferably low ionic strength is very challenging and so far there was no approach delivering a fast, simple and systematic way to generate buffer compositions for controllable pH gradients in a chosen pH range with minimized ionic strength. Therefore, an *in silico* approach was developed for the optimization of multi-component buffer systems to achieve buffer compositions with linear titration curves applicable for the formation of controllable pH gradients in IEC. After validating the buffers applicability to form linear pH gradients in IEC for the separation of proteins, the elution behaviour or 22 model proteins was investigated. Thereby, differences between chromatographic separations and isoelectric focusing were examined to acquire deeper understanding of the underlying mechanisms of the separation behaviour. The work and the results are described in detail in chapter I: *"Systematic generation of buffer systems for pH gradient ion exchange chromatography and their application"*.

In the second chapter of this work, the previously developed method to generate buffer systems for pH gradient IEC was applied to analytically characterize a very complex biotechnological feedstock containing a low titer target protein to obtain parameters useful for purification process development. The approach implementing four pH gradient IEC operations was applied to determine the elution behaviour of the target protein and the major impurities, which were further identified using LC-MS based techniques. The gained information on the elution behaviour of relevant contaminants and the target protein was subsequently used to lay out purification of the target protein using IEC. The results of the developed screening approach and the resulting purification procedure are discussed in detail in chapter II: *"Analytical characterization of complex, biotechnological feedstocks by pH gradient ion exchange chromatography for purification process development"*.

Parameters obtained by applying the described screening experiments (chapter II) are so called "heuristic" parameters, which are not applicable for process simulation using mechanistic models. Process development based on applying mechanistic models for the *in silico* optimization of chromatographic procedures has several advantages as low developmental time, costs and increased gain of process understanding due to the possibility to carry out a large amount of experiments *in silico*. However the determination of mechanistic parameters of single proteins contained in complex mixtures, allowing for simulation of chromatographic separations, is very challenging. A rapid, multidimensional fractionation approach, employing salt gradient IEC, reversed phase desalting, sample concentrating and high throughput capillary gel electrophoresis was developed to obtain the parameters allowing for simulation of chromatographic processes applying the steric-mass action model [1]. The approach was subsequently applied to characterize a cell lysate containing a low titer expressed target protein. Validation of the parameters was carried out by comparing predicted retention volumes and chromatograms to experimentally determined ones. All results as well as the experimental procedure are described in chapter III:

"A high-throughput 2D-analytical technique to obtain single protein parameters from cell lysates for in silico process development of ion exchange chromatography".
Mostly SMA parameters of proteins are only determined for single pH values, completely neglecting optimization of IEC separations at different pH values. Furthermore slight changes in pH and salt concentration values influence chromatographic separations and it is of great interest for process development to investigate these effects. In the forth chapter a newly developed methodology to obtain a full set of SMA parameters of single proteins in a pH range of 4 to 11.5 is described. The method is based on a fully experimental approach, applying pH gradient IEC at different salt concentrations and a set of isocratic elution experiments. SMA parameters were obtained by extending the applicability of the existing SMA model [5] to protein elution in pH gradients and by adding certain constraints to the parameter estimation. Determined model parameters were used to predict retention times for isocratic, mono- and bi-linear salt gradient elution as well as for combined pH and salt gradient elution. A Monte Carlo study was made to investigate the effects of experimental errors on all critical parameters. Furthermore, robustness, resolution and efficiency of all elution modes in IEC were investigated and compared. The results and the procedure to determine the parameters are shown in chapter IV: *"Parameter estimation for a wide pH range to model chromatographic separations on ion exchange materials for both salt and pH gradient elution".*
Analytical technologies, providing sufficient resolution to characterize protein isoforms are very rare and therefore of great interest for the quality control of therapeutic proteins. PEGylation of proteins is a common methodology to enhance the therapeutic properties of biopharmaceuticals as it improves their pharmacokinetic behaviour. The analytical characterization of PEGylated proteins is a challenging task but necessary for the therapeutics approval. In chapter V a method, applying pH gradient IEC for the analytical separation of mono-PEGylated lysozyme isomers is demonstrated, proving its exceptionally high chromatographic resolution for the separation of protein variants. PEGylation sites of the separated variants were identified, using tryptic digestion and MALDI-MS. Obtained results, as well as the experimental procedure are shown in chapter V: *"Isoform separation and binding site determination of mono-PEGylated lysozyme with pH gradient chromatography".*
Chapter I-IV are focusing on the analytical acquisition of parameters for purification process development of proteins contained in complex mixtures with a low titer. All methods are designed as experimental tools in purification process development especially for the described, challenging biotechnological feedstocks containing diagnostic antigens. In chapter V an alternative application for pH gradient IEC as an analytical tool for the characterization of protein variants is demonstrated, as it provides high chromatographic resolution.

Systematic generation of buffer systems for pH gradient ion exchange chromatography and their application

Frieder Kröner, Jürgen Hubbuch

Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Abstract

pH gradient protein separations are widely used techniques in the field of protein analytics, of which isoelectric focusing is the most well known application. The chromatografic variant, based on the formation of pH gradients in ion exchange columns is only rarely applied due to the difficulties to form controllable, linear pH gradients over a broad pH range. This work describes a method for the systematic generation of buffer compositions with linear titration curves, resulting in well controllable pH gradients. To generate buffer compositions with linear titration curves an *in silico* method was successfully developed. With this tool, buffer compositions for pH gradient ion exchange chromatography with pH ranges spanning up to 7.5 pH units were established and successfully validated. Subsequently, the buffer systems were used to characterize the elution behavior of 22 different model proteins in cation and anion exchange pH gradient chromatography. The results of both chromatographic modes as well as isoelectric focusing were compared to describe differences in between the methods.

Journal of Chromatography A (Volume 1285, 12 April 2013, Pages 78-87)

1 Introduction

pH gradient based separations of proteins for analytical purposes are among the most used techniques in protein analytics, e.g. isoelectric focusing or 2D-gelelectrophoresis. The chromatographic variant, pH gradient ion exchange chromatography (IEC), has been valued as an analytical separation technique in the field of proteomics [29] and as a screening tool for the selection of pH-related parameters in industrial bioseparation process development [30]. In addition, pH gradient ion exchange chromatography showed its potential as an excellent analytical tool in the routine analysis of monoclonal antibody charge variants [31] as well as PEGylation variants [32]. Another study on the differences between pH gradient and salt gradient IEC showed that pH gradient IEC might provide a higher chromatographic resolution [33].
Analogous to conventional IEC, protein binding is charge mediated in pH gradient IEC. In pH gradient IEC proteins are eluted by changing the pH of the mobile phase gradually and thus titrating the interaction of the protein with the resin. The fact that an interaction with a charged surface is involved suggests a fundamental difference between classical isoelectric focusing and pH gradient IEC, which could cause a deviation between the pI and the elution-pH of proteins. There are two main ways of generating pH gradients in IEC: an internal and an external method. For the internal method, a weak ion exchange resin with an intrinsic buffer capacity is used. The column is first equilibrated with the application buffer. The elution is done by titrating the intrinsic buffer capacity with the running buffer to generate the outlet pH gradient. The internal method is mostly referred to as chromatofocusing [34]. For the external method, the pH gradient is formed before entering the column by gradually mixing the running buffer with its titrant. The focus of this work is on the externally generated pH gradients, which are regularly run on strong ion exchange resins.
The major challenge in running pH gradients in IEC is the controllability of the pH gradient. Additionally, it is preferable to keep the ionic strength of the buffer system low, in order to reduce its influence on the proteins retention behavior. One approach is to use ampholytes, providing a high buffer capacity covering a broad pH range. However the usage of ampholytes has major disadvantages, being their lot-to-lot variability, their interaction with the proteins and chromatographic adsorbers, as well as the fact that they are difficult to remove from solution. Another approach is to use multiple, equally concentrated buffer substances with equally spaced pKa values in the chosen pH range [3]. The latter leads to reproducible, linear pH gradients covering a broad pH range on the corresponding chromatographic resin, but still with relatively high ionic strength and an inconstant buffer capacity. To lower the ionic strength of the buffer composition, a further approach uses a very similar buffer chemistry with lower concentrated buffer substances, compensating gradient non-linearities with a software-enabled, algorithmic

control of the gradient-mixing [35], resulting in linear pH gradients with uneven buffer capacity, but reduced ionic strength.
This means that so far there is no approach delivering a simple, fast, systematic way to generate buffer compositions for controllable pH gradients in every chosen pH range with minimized and known ionic strength. To generate controllable pH gradients, the buffer capacity of the buffer has to be kept constant, resulting in a buffer composition with a linear titration curve, which can then be controlled by the liquid chromatography system to generate the pH gradient. Other publications focusing on a similar problematic, the *in silico* optimization of buffer compositions for isoelectric focusing, as well as the simulation of the resulting pH gradient, have proven to be successful for the specific case presented [36–39].
In this work, the idea of optimizing a given buffer composition's capacity, predicting its titration curve and the course of the ionic strength throughout the gradient was adapted for the purpose to generate buffer systems for pH gradient IEC. The established *in silico* buffer optimization tool was used to generate buffer compositions for long range pH gradient IEC. To demonstrate their applicability, the optimized buffer compositions were applied for the formation of linear pH gradients and validated for the characterization of protein elution behavior in pH gradient IEC. Subsequently, the methods were used for an extensive study of the elution behavior of 22 model proteins on differently charged chromatographic resins. A comparison of the proteins elution-pH values of the two different chromatographic modes with literature values on the pIs of the proteins was made to gain better understanding of the differences between the methods, as well as the electrostatics of the single proteins.

2 Theory

For the generation of controllable pH gradients, the titration curve of the applied buffer system has to be linear, which results from an even buffer capacity throughout the pH range. The buffer capacity, the titration curve and the ionic strength from single buffer substances as well as mixtures can be calculated. By varying the buffer substances and their single concentrations, buffer mixtures can be optimized to achieve linear titration curves. The theoretical background, the derivation of the needed equations and the optimization procedure are described in the following section.

2.1 Calculations

For chromatographic operations, most used buffer substances are mono-, di- or tribasic/-protic, while higher degrees are only rarely used. Therefore, this work is only focusing on these substances. In the following part, the derivation of the necessary equations is shown for a weak

monoprotic substance.
The dissociation of a weak monoprotic acid

$HA \overset{k_a}{\rightleftharpoons} H^+ + A^-$

concludes to the following mass balance, whereby C_A is the total acid concentration.

$C_A = [HA] + [A^-]$, while $[HA] = \frac{[A^-][H^+]}{k_a}$

Combining both results in the following term for $[A^-]$:

$[A^-] = C_A \frac{k_a}{[H^+]+k_a}$

The electroneutrality condition defines the following charge balance for the dissolved acid

$[H^+] = [A^-] + [OH^-]$

Adding a strong, monobasic substance changes the charge balance to

$[B^+] + [H^+] = [A^-] + [OH^-]$

Including the ionic product of water $[H^+][OH^-] = k_w$ the charge balance can be written as

$[B^+] + [H^+] = [A^-] + \frac{[k_w]}{[H^+]}$

combining the mass balance and the charge balance leads to

$$[B^+] = C_A \frac{k_a}{[H^+]+k_a} + \frac{k_w}{[H^+]} - [H^+] \tag{1}$$

from which the titration curve of the monoprotic acid can be calculated iteratively. To calculate the buffer capacity B, the equation has to be differentiated with respect to $[H^+]$, while remembering that

$pH = -log_{10}[H^+] = -\frac{1}{ln(10)} ln([H^+])$

the buffer capacity is

$$B = \frac{dB^+}{dpH} = ln(10)(C_A \frac{k_a[H^+]}{([H^+]+k_a)^2} + \underbrace{\frac{k_w}{[H^+]} + [H^+]}_{B(w)}) \tag{2}$$

which includes the intrinsic buffer capacity of water ($B(w)$). To calculate the ionic strength of a buffering solution at different pH values, the single ionic species have to be summed up, independent of their charge. The general formula for the ionic strength of a solution including n different ionic species is

$I = \frac{1}{2} \sum_{i=1}^{n} C_i z_i^2$;

C= mol. conc. of ions; z= charge number

and in case of a monoprotic acid the equation is

$$I = \frac{1}{2}(\underbrace{\frac{C_A k_a}{[H^+]+k_a}}_{}\,[\mathrm{A}^-] + \underbrace{\frac{k_w}{[H^+]}}_{}\,[\mathrm{OH}^-] + [H^+] + [B^+]) \tag{3}$$

The derived equations (1-3) can be used to compute the course of the titration curve, the buffer capacity and the resulting ionic strength of a single substance. To calculate the values for di- and tribasic/-protic buffer substances the corresponding equations can be derived in the same way, the results are:

Diprotic substances

The titration of a diprotic substance

$$[B^+] = C_A(\frac{k_{a1}[H^+]+2k_{a1}k_{a2}}{[H^+]^2+k_{a1}[H^+]+k_{a1}k_{a2}}) + \frac{k_w}{[H^+]} - [H^+] \tag{4}$$

The buffer capacity of a diprotic substance

$$\begin{aligned} B =& ln(10) \\ & * (C_A(\frac{k_{a1}[H^+]^3+4k_{a1}k_{a2}[H^+]^2+k_{a1}^2k_{a2}[H^+]}{([H^+]^2+k_{a1}[H^+]+k_{a1}k_{a2})^2}) \\ & + \frac{k_w}{[H^+]} + [H^+]) \end{aligned} \tag{5}$$

The ionic strength of a diprotic substance

$$\begin{aligned} I =& \frac{1}{2}(C_A\frac{k_{a1}[H^+]+4k_{a1}k_{a2}}{[H^+]^2+k_{a1}[H^+]+k_{a1}k_{a2}} + \frac{k_w}{[H^+]} \\ & + [H^+] + [B^+]) \end{aligned} \tag{6}$$

Triprotic substances

The titration of a triprotic substance

$$\begin{aligned} [B^+] =& C_A(\frac{k_{a1}[H^+]^2+2k_{a1}k_{a2}[H^+]+3k_{a1}k_{a2}k_{a3}}{[H^+]^3+k_{a1}[H^+]^2+k_{a1}k_{a2}[H^+]+k_{a1}k_{a2}k_{a3}}) \\ & + \frac{k_w}{[H^+]} - [H^+] \end{aligned} \tag{7}$$

The buffer capacity of a triprotic substance

$$\begin{aligned}\mathrm{B} =& ln(10)(C_A * \\ & (\frac{k_{a1}[H^+]^5+4k_{a1}k_{a2}[H^+]^4+k_{a1}^2k_{a2}[H^+]^3+9k_{a1}k_{a2}k_{a3}[H^+]^3}{([H^+]^3+k_{a1}[H^+]^2+k_{a1}k_{a2}[H^+]+k_{a1}k_{a2}k_{a3})^2} \\ & + \frac{4k_{a1}^2k_{a2}k_{a3}[H^+]^2+k_{a1}^2k_{a2}^2k_{a3}[H^+]}{([H^+]^3+k_{a1}[H^+]^2+k_{a1}k_{a2}[H^+]+k_{a1}k_{a2}k_{a3})^2}) \\ & + \frac{k_w}{[H^+]} + [H^+])\end{aligned} \tag{8}$$

The ionic strength of a triprotic substance

$$\begin{aligned}I =& \frac{1}{2}(C_A \frac{k_{a1}[H^+]\hat{A}^2+4k_{a1}k_{a2}[H^+]+9k_{a1}k_{a2}k_{a3}}{[H^+]^3+k_{a1}[H^+]^2+k_{a1}k_{a2}[H^+]+k_{a1}k_{a2}k_{a3}} \\ & + \frac{k_w}{[H^+]} + [H^+] + [B^+])\end{aligned} \tag{9}$$

With the derived equations 1-9, the titration curve, the buffer capacity and the ionic strength of mono-, di-, and triprotic substances can be calculated. To calculate the values for basic substances k_a has to be substituted by $k_b = k_w/k_a$ and $[H^+]$ by $[OH^-]$, with $[OH^-] = 10^{-14}/[H^+]$. For the computation of the buffer capacity, the titration curve, or the ionic strength of buffer compositions consisting of multiple single substances, the values for the single substances are calculated and summed up. For example, the buffer capacity of m monoprotic, d diprotic and t triprotic acids is a sum of the single substances buffer capacity and the intrinsic buffer capacity of water.

$$\begin{aligned}B_{sum} =& \sum_{i=1}^{m} B_i + \sum_{j=1}^{d} B_j + \sum_{k=1}^{t} B_k \\ & + ln(10)(\frac{k_w}{[H^+]} + [H^+])\end{aligned}$$

The corresponding equations can also be formulated for the computation of the titration curve and the ionic strength. Combining all the equations in the computing environment MATLAB (MathWorks, Natick, MA, USA), made it possible to calculate the values for any mixture of mono-, di- and triprotic/-basic substances.

2.2 Optimization

The key parameter for buffer compositions with a linear titration curve is an even buffer capacity throughout the chosen pH range. To cover broad pH ranges, multiple buffer substances have to be chosen with pKa values covering the chosen pH range. The buffer capacity can be kept constant throughout the pH range by choosing buffer substances with equally distributed pKa values [3, 36]. Another approach is to vary the concentrations of the single buffer substances in the mixture. Due to the lack of compatible buffer substances with a broad variety of pKa values, a buffer composition with evenly distributed pKa values for a linear pH gradient with a broad pH range is difficult to achieve. A combination of the right choice of commercially

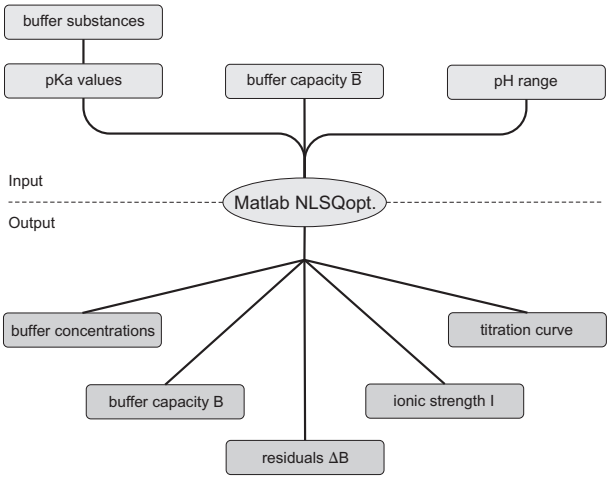

Figure 1: The optimization method
Schematic overview of the *in silico* optimization process for buffersystems applicable for the formation of controllable, linear pH gradients in IEC.

available buffer substances with optimized concentrations leads to buffer compositions with constant buffer capacities resulting in linear titration curves.
The deviation of a buffer composition's buffer capacity from its mean buffer capacity $\overline{B}$, throughout a chosen pH range is formulated in the following equation:

$$\Delta B_{pH_{min-max}} = |B_{pH_{min-max}} - \overline{B}|$$

$B_{pH_{min-max}}$ is the buffer compositions summed buffer capacity B_{sum} in the chosen pH range. To keep the buffer capacity as constant as possible, the deviation $\Delta B_{pH_{min-max}}$ has to be as small as possible, therefore the target function for the optimization of buffer compositions is: $\Delta B_{pH_{min-max}} \rightarrow min$. To optimize buffer compositions the MATLAB-based calculation of a mixture's buffer capacity was enabled to calculate the target function $\Delta B_{pH_{min-max}}$. Further on, the concentrations of the single buffer substances were defined as variables. By varying the single buffer concentrations the minimum of the target function $\Delta B_{pH_{min-max}}$ can be determined. The non-linear least square algorithm was used to minimize the target function. Fig. 1 describes the MATLAB-based optimization procedure graphically. For the optimization of a buffer composition, the single pKa values of the chosen buffer substances, the chosen mean buffer capacity $\overline{B}$, as well as the chosen pH range have to be entered. After minimizing the target function by varying the single buffer concentrations, the titration curve, the buffer capacity and the course of the ionic strength are calculated. This tool enables the generation of buffer compositions with a linear titration curve in every pH range with a chosen mean buffer capacity $\overline{B}$. The only major requirement that has to be fulfilled is the availability of suitable buffer substances for the chosen

pH range.

2.3 Buffer substances

The buffer substances for the pH gradient IEC need to fulfill different criteria. The substances should not interact with the ion exchange resin and thus they have to be oppositely charged. Therefore, acidic substances are used for cation exchange chromatography (CEC), while basic substances are used for anion exchange chromatography (AEC). Another important criterion is that the substances do not interact with proteins, which is the case for ampholytes. We also decided to limit the substances to mono-, di-, and triprotic/-basic acids, because substances with a higher possible dissociation degree are rarely used in chromatography. Additionally, the substances have to be commercially available at an acceptable price and with analytical grade purity. A collection of useful buffer substances with their corresponding pKa values is shown in Tab. 1.

Table 1: Buffer substances

Basic buffer substances				**Acidic buffer substances**			
Substance	pKa_1	pKa_2	pKa_3	Substance	pKa_1	pKa_2	pKa_3
Piperidine	11.12[a]	–	–	CABS	10.70[c]	–	–
Methylamine	10.75[b]	–	–	CAPS	10.50[a]	–	–
1-Ethylpiperidine	10.45[a]	–	–	CAPSO	9.83[a]	–	–
1,2-Ethanediamine	9.93[b]	6.99[b]	–	CHES	9.39[a]	–	–
1,2-Propanediamine	9.82[a]	6.61[a]	–	AMPSO	9.14[a]	–	–
Piperazine	9.78[b]	5.52[b]	–	TABS	8.90[c]	–	–
2-Methylpiperazine	9.54[b]	5.24[b]	–	TAPS	8.44[a]	–	–
1-Methylpiperazine	9.16[b]	4.78[b]	–	EPPS	8.00[a]	–	–
Bis-tris propane	8.93[b]	6.59[b]	–	POPSO	7.80[a]	–	–
Morpholine	8.34[b]	–	–	TAPSO	7.64[a]	–	–
Tris	8.16[b]	–	–	HEPES	7.56[a]	–	–
1,4-Dimethylpiperazine	8.15[b]	4.04[b]	–	MOPS	7.18[a]	–	–
Triethanolamine	7.52[b]	–	–	MOPSO	6.90[a]	–	–
4-Methylmorpholine	7.34[b]	–	–	MES	6.10[a]	–	–
Bis-tris	6.22[b]	–	–	Acetate	4.76[a]	–	–
Hydroxylamine	5.67[b]	–	–	Succinate	4.21[a]	5.64[a]	–
Pyridine	5.23[a]	–	–	Formate	3.75[a]	–	–
–	–	–	–	Malate	3.40[a]	5.11[a]	–
–	–	–	–	Citrate	3.13[a]	4.76[a]	6.40[a]
–	–	–	–	Phosphate	2.15[a]	7.20[a]	12.35[a]

a Handbook of Chemistry and Physics 89th Edition, 2008, CRC Press

b Measured pKa value, at 25°C

c Thiel et al. [40]

3 Materials and methods

3.1 Chemicals & Buffers

For the pH gradient CEC the chosen substances were MES, formic and acetic acid (Merck, Darmstadt, Germany), HEPPSO (Molekula, Dorset, UK), and MOPSO, TAPS, CHES, CAPS (Applichem, Darmstadt, Germany). The buffer for pH gradient AEC consisted of hydroxylamine, methylamine, 1,2-ethanediamine, 1,4-dimethylpiperazine from Merck (Darmstadt, Germany), 1-methylpiperazine from Sigma-Aldrich (St. Louis, MO, USA) and Bis-Tris from Molekula (Dorset, UK). Sodium chloride, hydrochloric acid and sodium hydroxide were obtained from Merck (Darmstadt, Germany). All substances were purchased in analytical grade. The used model proteins were purchased from Sigma-Aldrich (St. Louis, MO, USA), except glucose isomerase, which was purchased from Hampton Research (Hampton Research, Aliso Viejo, CA, USA). To set up the buffers for the pH gradient IEC, all substances were weighed in, dissolved in ultrapure water and split in two equal volumes. One part was titrated to the low pH extreme, the other to the high pH extreme, with the appropriate strong titrant. The pH adjustment was carefully performed with a freshly, five-point calibrated pH meter (HI-3220, Hanna Instruments, Woonsocket, RI, USA). The pH calibration buffers, pH 3, 5, 7, 9, 11, were high precision standards from Hanna Instruments. After adjusting the pH, the buffers were brought to their final volume. All buffers have been filtered through 0.2 μm cellulose acetate filters (Sartorius, Goettingen, Germany). All buffer solutions were prepared with ultrapure water from a Arium water purification system (Sartorius, Goettingen, Germany). After preparation, the buffers were used for a maximum time period of two weeks.

3.2 Buffer optimization for pH gradient IEC

For the optimization of the buffer composition a mean buffer capacity of 10 mM was chosen. The pH range for the optimization was pH 10.5-3.5 for pH gradient AEC and 4.0-11.0 for pH gradient CEC. After optimizing the compositions, the buffers were prepared as described previously.
The applicability of the buffer systems was validated by examining the linearity of the resulting pH gradients. For the pH gradient CEC, a Mono S 4.6/100 column (GE Healthcare, Uppsala, Sweden) was used. The pH gradient AEC was run on a Mono Q 4.6/100 column (GE Healthcare, Uppsala, Sweden).
The experiments were done on an Aekta Purifier (GE Healthcare, Uppsala, Sweden) equipped with a pH electrode for online pH measurement. The pH electrode was calibrated with pH 3 and pH 12 high precision calibration standards. The column was equilibrated offline with ten column volumes (CV = 1.662 ml) of the application buffer. The gradient from 0-100% was performed in

15 column volumes. After running the gradient, a post-gradient, at 100% of the elution buffer was maintained for another five column volumes. The chromatographic run was executed with a flow rate of 1.5 ml/min.
To validate the reproducibility, three different experimental setups were run by three different persons. This means, for all three chromatographic runs the buffers were prepared fresh and the online-pH measurement was calibrated again. Each experiment was performed by a different person.

3.3 Elution-pH measurement of model proteins

The model proteins were prepared by dissolving 2 mg protein in 2 ml of the application buffer. Particulates in the protein solution were removed by centrifugation. The protein elution-pH values were determined by using the same chromatographic procedure as described in 3.2, but with an injection of 250 μl of the protein solution. The pH was monitored online, as well as the UV-absorption at 280 nm. To analyse the chromatograms of the pH gradient AEC the UV280 absorption of a blank run had to be subtracted first as the buffer showed a reproducible ghost peak at the beginning of each run probably resulting from minor impurities of the buffer substances. Afterwards, the proteins' elution-pH values could be determined from the peak maxima in the chromatograms. The Unicorn 5.2 software (GE Healthcare, Uppsala, Sweden) was used to analyse the chromatograms.
To validate the applicability of the method for the characterization of proteins in terms of repeatability as well as intermediate precision [2], the elution-pH values of four different model proteins were determined in three different experimental setups by a six time repeat measurement run by three different people.
After validation of the method's applicability, 22 model proteins were characterized by pH gradient IEC. Every model protein was analyzed by a repeat measurement on the cation and the anion exchange resin. The average protein's elution-pH value was then calculated for each chromatographic mode.

4 Results and discussions

4.1 Buffer optimization for pH gradient IEC

The aim was to generate buffer compositions for pH gradient AEC and CEC providing a broad pH range with low ionic strength. At first the buffer substances were chosen. With the known pKa values and a mean buffer capacity of 10 mM, the MATLAB optimization procedure was applied for the minimization of the previously mentioned target function ΔB, varying the single

Table 2: Buffer systems for pH gradient IEC

AEC buffer pH 10.5 - 3.5

Substance	pKa_1	pKa_2	Conc.[mM]
Methylamine	10.75	–	9.8
1,2-Ethanediamine	9.93	6.99	9.1
1-Methylpiperazine	9.16	4.78	6.4
1,4-Dimethylpiperazine	8.15	4.04	13.7
Bis-tris	6.22	–	5.8
Hydroxylamine	5.67	–	7.7
min. ionic str. = 8.5 mM	/	max. ionic str. = 104.3 mM	

CEC buffer pH 4.0 - 11.0

Substance	pKa_1	pKa_2	Conc.[mM]
CAPS	10.50	–	15.6
CHES	9.39	–	9.4
TAPS	8.44	–	4.6
HEPPSO	8.04	–	9.9
MOPSO	6.90	–	8.7
MES	6.10	–	11.0
Acetate	4.76	–	13.0
Formate	3.75	–	9.9
min. ionic str. = 8.3 mM	/	max. ionic str. = 89 mM	

buffer concentrations. The optimization resulted in the buffer compositions listed in Tab. 2. The course of the buffer capacity and the calculated titration curve of each buffer composition are visualized in Fig. 2. The high linearity of the calculated titration curve clearly proves that this approach was successful. The ionic strength of the buffer systems at the extreme pH values (Tab. 2) is acceptably low, leading to a minimal effect of the ionic strength on the chromatographic elution behavior of proteins.

4.2 Validation

To validate the applicability of the buffer systems, they have been applied on their correlating chromatographic resin. The resulting pH gradients were determined by measuring the pH online at the column outlet. The resulting pH gradients are shown in Fig. 3. Both gradients reached a linearity with $R^2 > 0.99$.

To reach high linearity at the extreme pH values 10 mM NaCl was added to suppress the exchange of the resins counterion: This means the exchange of Cl^- by OH^- at very basic pH values on the anion exchange resin or of Na^+ by H^+ at very acidic pH values on the cation exchange resin. Suppressing this ion exchange procedure by adding a small amount of salt, improved linearity

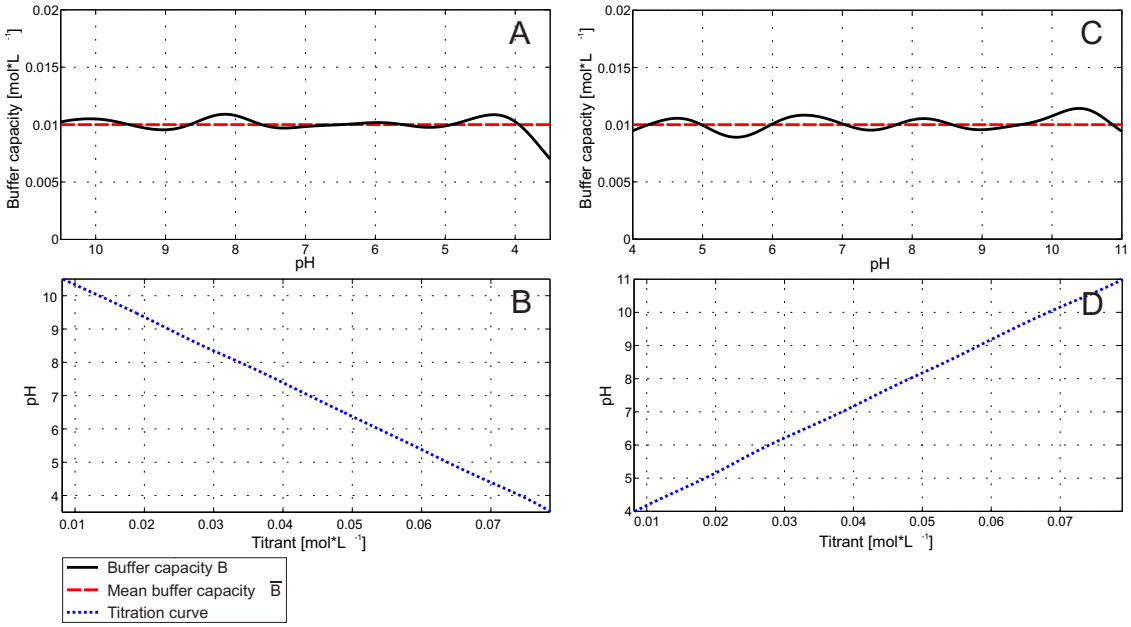

Figure 2: Buffer calculations
Illustration of the course of the buffer capacity and the titration curve of the optimized buffer compositions described in Tab. 2. Optimization of the buffer systems and the calculation of the related plots were done in MATLAB. **A & B**: AEC buffer pH 10.5 - pH 3.5: A: calculated course of the buffer capacity, B: calculated titration curve; **C & D**: CEC buffer pH 4.0 - pH 11.0: C: calculated course of the buffer capacity, D: calculated titration curve;

greatly. The concentration of NaCl depends on the ion exchange rate of the resins counterion with $H^+{}_7OH^-$ and would therefore probably need adaption for other resins. Obviously, the ionic strength of 10 mM NaCl has to be added to the calculated buffers ionic strength.
However, there was still a small effect of the columns on the resulting pH gradients. It can be seen from the chromatograms in Fig. 3, that both gradients slightly deviate from linearity in the beginning and their starting points are delayed by 4–5 ml. This is little more than the gradient delay volume, which was 3.5 ml from the pump to the pH probe for the chosen setup. It is likely to be that this is still an effect caused by the described not completely suppressed ion exchange process. It seems that a small amount of H^+/OH^- ions still bound to the column and were then titrated off the adsorbent by the gradual increase of the elution buffer. This consequently would have leaded to delayed gradient starting points and an increased steepness in the beginning of the gradient, which correlates well with the presented results in Fig. 3. Therefore, a further increase of the amount of NaCl in the buffer would eliminate this effect. However, increasing the salt concentration further would effect the proteins' elution behaviour, which was not wanted. As a gradient linearity with $R^2 > 0.99$ was already reached, we decided that a further increase

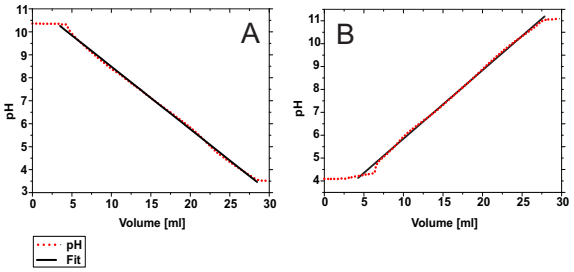

Figure 3: pH gradients - linearity
Experimental determination of pH gradient linearity. pH gradients were formed by the application of each optimized buffer (Tab. 2) on the corresponding MonoQ / MonoS column. pH was monitored online at the column outlet with a delay volume from pump to pH probe of 3.5 ml. The gradient length was 24,9 ml with the starting point at 0 ml. The columns were equilibrated offline before the run. The linearity of the gradient was determined by linear regression. **A**: AEC buffer applied on MonoQ column, linearity $R^2 > 0.99$; **B**: CEC buffer applied on MonoS column, linearity $R^2 > 0.99$;

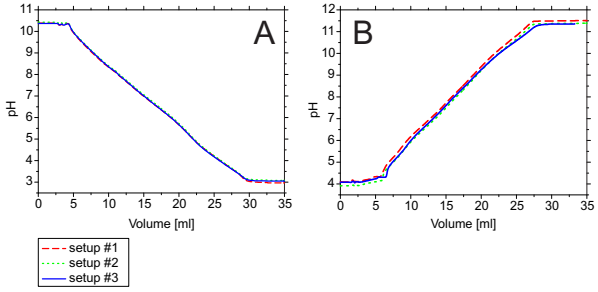

Figure 4: pH gradients - reproducibility
Illustration of the overlay of three pH gradients (length = 24.9 ml, start at 0 ml) on a MonoS / MonoQ column formed by using the optimized buffer systems (Tab. 2). pH was monitored at the column outlet with a delay volume of 3.5 ml. Experiments were run as three independent setups by three different users. **A**: AEC buffer pH 10.5 - 3.0, 3x experimental setups; **B**: CEC buffer pH 4.0 - 11.5, 3x experimental setups;

of the NaCl concentration would not be of any advantage.

The results of the validation, done by running the experiments in three different experimental setups by three persons, are shown in Fig. 4. The nearly identical course of the three pH gradients indicates the reproducibility of the method under common lab circumstances. Additionally we recognized that the gradients linearity could be maintained even when the range was extended by 0.5 pH units (AEC 10.5 - 3.0; CEC 4.0 - 11.5). This implies the minor disadvantage that the buffer capacity is lower, out of the optimized buffer systems pH range. To validate the method for its actual purpose, being the separation and characterization of proteins with pH gradient IEC, the elution-pH values of four model proteins were determined in a six time repeat measurement in three different experimental setups by three different people. The experiments were done for both chromatographic modes. The model proteins for pH gradient AEC were cytochrome C, myoglobin, glucose oxidase and amyloglucosidase. For the pH gradient CEC, glucose oxidase, bovine serum albumine, alpha-chymotrypsinogen A and cytochrome C were chosen. The decision for the model proteins was made due to their clearly defined chromatografic peak in the applied chromatografic mode. The results are shown in Tab. 3.

The results clearly indicate a high repeatability and intermediate precision with a standard deviation for both 0.1 pH units, in both chromatografic modes. Thus the experimental setup using

Table 3: Experimental validation of pH gradient IEC

1. pH gradient AEC

Protein	pH	Repeatability	Int. prec.
RNAse A	9.38	$^{+}/_{-}$ 0.03	$^{+}/_{-}$ 0.04
Myoglobin	8.65	$^{+}/_{-}$ 0.01	$^{+}/_{-}$ 0.02
Glucose Oxidase	4.68	$^{+}/_{-}$ 0.01	$^{+}/_{-}$ 0.06
Amyloglucosidase	3.93	$^{+}/_{-}$ 0.01	$^{+}/_{-}$ 0.04

2. pH gradient CEC

Protein	pH	Repeatability	Int. prec.
Cytochrome C	10.25	$^{+}/_{-}$ 0.01	$^{+}/_{-}$ 0.09
α-Chymo-trypsinogen A	10.07	$^{+}/_{-}$ 0.01	$^{+}/_{-}$ 0.05
BSA	5.84	$^{+}/_{-}$ 0.00	$^{+}/_{-}$ 0.06
Glucose Oxidase	4.34	$^{+}/_{-}$ 0.01	$^{+}/_{-}$ 0.09

the optimized buffer system clearly qualified for the specified purpose, meaning the separation and characterization of proteins with pH gradient IEC.

4.3 Elution-pH determination of model proteins

To characterize the elution behavior of proteins in pH gradient IEC, the validated method was used for an extensive characterization of 22 model proteins, with anion and cation exchange chromatography, comparing the proteins' elution behavior in the two chromatographic modes and literature values of experimentally determined pI values of the proteins.

In previous studies by Ahamed et al. [3], different model proteins were analyzed in terms of their elution behavior in pH gradient AEC, correlating the results with the pI values as well as the proteins titration curves. Their results showed that acidic (pI $<$ 6) and basic proteins (pI $>$ 8) elute roughly at their pI value while neutral proteins (pI $\sim$ 7) eluted at higher pH values in pH gradient AEC. Ahamed et al. [3] explained this phenomenon with the flat nature of the titration curve of neutral proteins. To our knowledge no study has been made, comparing the elution-pH value in pH gradient AEC and CEC to discover possible differences between the two chromatographic modes. Additionally the ionic strength of the buffer system, proposed by Ahamed et al. [3] was relatively high. The buffer's (pH 10.5 - 4.0 with 5 mM NaCl) calculated ionic strength was 10 - 198 mM charge equivalents. This means that the ionic stength of our optimized buffer systems is reduced roughly by one half, while providing a constant buffer capacity and also a linear, controllable pH gradient. This results in a reduced influence on the proteins retention behavior by ionic strength, meaning that these effects are minimized for the characterization of proteins.

The determined elution pH values as well as literature values of experimentally determined pIs

are shown in Tab. 4. For better interpretation of the results, they were sorted in two groups, proteins with and without isoforms.

4.3.1 Isoform-free proteins

The identity of the detected chromatographic peak is clearly defined in both chromatographic modes as isoform free proteins are analyzed. This allows the direct comparison of the two elution-pH values and the isoelectric point. Interpretation of the results was done with regard to the differences between the three values, to gain insight into mechanistical differences between the three methods. The differences are shown in numbers for a better overview in Tab. 5.

Table 4: Elution-pH of model proteins

Protein	Organism	Ref.	pI	AEC Elution-pH	CEC Elution-pH
Cytochrome C	*Bos taurus*	[41]	10.01	9.95	10.25
Cytochrome C	*Equus caballus*	[41]	10.03	9.93	10.23
Amyloglucosidase	*Aspergillus niger*	S.I.	3.60	3.93	< 4.00
Glucose oxidase	*Aspergillus niger*	S.I.	4.2	4.68	4.34
Lysozyme	*Homo sapiens*	[42]	10.0	≈ 10.5	10.19
α-Lactalbumin	*Bos taurus*	S.I.	4.53	5.29	5.34
α-Chymotrypsinogen A	*Bos taurus*	[42]	8.97	8.88	10.07
Glucose isomerase	*Streptomyces rubignosus*	calc.	5.00	4.34	n.a.
Lysozyme	*Gallus gallus*	S.I.	11.35	> 10.5	10.72
Hemoglobin A_0	*Homo sapiens*	[42]	6.95	7.70	n.a.
RNAse A	*Bos taurus*	[43]	9.7	9.38	9.16
RNAse B	*Bos taurus*	[43]	9.7	9.38	8.83
beta-Lactoglobulin	*Bos taurus*	[42]	5.26, 5.34	4.25	5.97
Carbonic anhydrase	*Bos taurus*	[42]**	5.89*	8.07, 8.23, 8.64*	6.48, 7.41, 10.66*
Myoglobin	*Equus caballus*	S.I.	6.8, 7.2	8.65, 8.96	n.a.
Thaumatin	*Thaumatococcus*	[44]	$11.7 - 12$	9.82*, 10.02*	9.17*, 9.49*
Serum albumin	*Homo sapiens*	[45]	4.8, 5.6	5.25, 5.35	6.06
Serum albumin	*Bos taurus*	[42]	4.98, 5.07, 5.18	5.18, 5.28, 5.32	5.79
Conalbumin	*Gallus*	[42]	5.62, 5.78*, 6.05, 6.25*, 6.50, 6.73*	5.89*, 6.21*, 6.72*	7.47
holo-Transferrin	*Bos taurus*	[46]	5.2, 5.4, 5.6, 5.7	5.64, 5.84, 5.98	6.54, 6.95
Ovalbumin	*Gallus gallus*	[47]	4.8, 4.9, 5.0	5.14*, 5.23*, 5.33*, 5.49	4.89
Trypsin inhibitor	*Gallus gallus*	[48]	3.83, 4.01, 4.17, 4.28, 4.41	4.20, 5.17, 5.35*, 5.70, 5.89	4.72
Catalase	*Bos taurus*	S.I.	5.40	5.96, 6.09*	5.44, 5.99*, 6.19

- S.I.: Supplier information, Sigma-Aldrich (St. Louis, MO, USA)
- Calc.: Calculated pI (Compute pI/Mw, $http://web.expasy.org/compute_pi/$)
- n.a.: Not available, protein denatured in application buffer
- * : Major isoform
- ** : Literature pI values incomplete

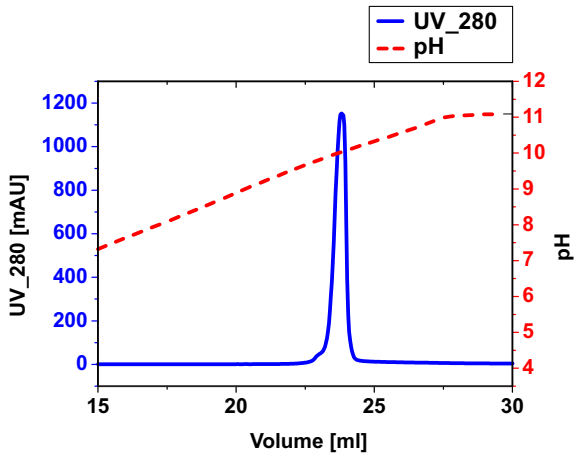

Figure 5: pH gradient CEC of Cytochrome C
CEC elution-pH determination of the isoform-free protein Cytochrome. 250 μl of 1 mg/ml protein solution were injected on a MonoS column and analysed by a 24.9 ml pH gradient from pH 4.0 - 11.5, while monitoring the pH at the column outlet and the UV 280 nm signal. Elution-pH was determined by reading out the pH value at the determined peak maximum. Chromatograms were analysed with the Unicorn 5.2 software.

The determination of a protein's elution-pH value is exemplarily illustrated in Fig. 5. The illustration shows the analysis of the isoform-free protein Cytochrome C by pH gradient CEC. To obtain the elution-pH value of the protein the pH at the peak maximum was determined.
The hemeproteins, **cytochrome C** from bovine or horse heart eluted merely at the same pH values, close to their pI values. Comparing the two IEC modes, the elution-pH values in pH gradient CEC are slightly shifted to higher pH values.
The enzyme **amyloglucosidase** is very acidic, therefore the protein was not retained in pH gradient CEC at pH 4.0 and eluted in the flow-through. The AEC elution-pH value is in the range of the proteins pI.
The acidic protein **glucose oxidase** eluted close to its pI values, which is lower than both

Table 5: Comparison of protein elution-pH values and their pI values

Protein	ΔpH pI-AEC	ΔpH pI-CEC	ΔpH AEC-CEC
Cytochrome C (bovine)	+ 0.06	− 0.24	− 0.30
Cytochrome C (equine)	+ 0.01	− 0.20	− 0.30
Amyloglucosidase	− 0.33	-	-
Glucose oxidase	− 0.48	− 0.14	+ 0.34
Lysozyme (human)	− 0.50	− 0.19	+ 0.31
α-Lactalbumin	− 0.76	− 0.81	− 0.05
α-Chymotrypsinogen A	− 0.09	− 1.10	− 1.19
Lysozyme (chicken)	-	+ 0.63	-
Glucose isomerase	+ 0.66	-	-
Hemoglobin A_0	− 0.75	-	-

elution-pH values. The AEC elution-pH value is a little higher than the corresponding CEC elution-pH value.
Human milk **lysozyme** eluted at very basic pH values in both chromatographic modes. The protein's elution-pH in CEC is close to its pI but lower than in AEC.
While having two nearly identical elution-pH values, the pI of **α-lactalbumin** differs strongly from the proteins elution-pH values. Both elution-pH values are shifted about $\sim$ 0.8 pH units to higher pH values.
The AEC elution-pH value of **α-chymotrypsinogen A**, is roughly in the range of its pI value. Surprisingly the elution-pH in CEC is more than one unit higher than in AEC.
Lysozyme from chicken egg is a very basic protein, which was therefore not retained in pH gradient AEC. The elution-pH value in pH gradient CEC is lower than the protein's pI and than the corresponding value in AEC.
The pI value of **glucose isomerase** is $\sim$ 0.8 pH units higher than the corresponding AEC elution-pH. The CEC elution-pH could not be determined due to the known instability of the protein below pH 5.0 (Supplier information).
The hemeprotein **hemoglobin** A_0 is the major isoform of human hemoglobins. The protein's AEC elution-pH differs strongly from its pI value, both of them are in the neutral pH range . The proteins CEC elution-pH could not be determined due to denaturation of the protein in the application buffer at pH 4.0.
Summarizing the results leads to the following conclusions. The difference of the pI value and the elution-pH value varies from protein to protein, e.g. the elution-pH values of cytochrome c are very close to the proteins pI, while the elution-pH values of α-lactalbumin vary very strong from the proteins pI. A plausible explanation for this lies in the fundamental difference between isoelectric focusing and pH gradient IEC, the interaction with the stationary phase. Due to the threedimensionaltiy of the protein only a patch of the protein might be directly involved into the binding of the stationary phase. Depending on the strength of the attractive interaction between the resin and the interacting patch of the protein, this might result in a difference between the pI and the elution-pH value. The more anisotropic the proteins charge distribution is, the bigger the difference between the values would get. By means of this explanation it is very likely that the surface charge distribution of cytochrome c is more isotropic as the one of α-lactalbumin.
The same explanation is plausible for the elucidation of the differences between the elution-pH values of anion- and cation exchange chromatography. Due to the oppositely charged adsorbers, different patches of the protein surface might interact with the different adsorber type. The resulting difference in the attractive interaction with the stationary phase concludes to a variation of the two elution-pH value.
One major drawback of using long range pH gradients is the denaturation of the protein at the

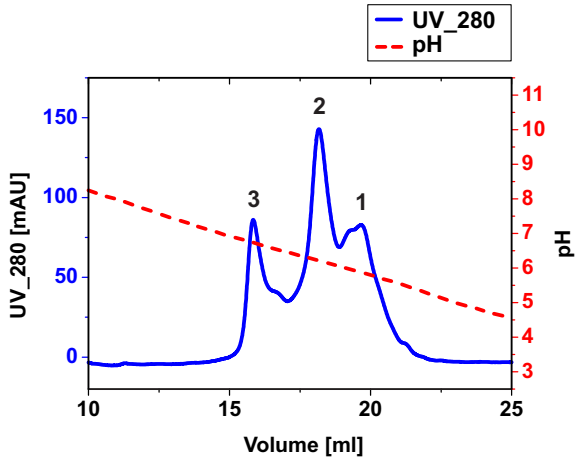

Figure 6: pH gradient AEC of Conalbumin
AEC elution-pH determination of the isoforms of Conalbumin. 250 μl of 1 mg/ml protein solution were injected on a MonoQ column and analysed by a 24.9 ml pH gradient from pH 10.5 - 3.0, while monitoring the pH at the column outlet and the UV 280 nm signal. Elution-pH values were determined by reading out the pH values at the determined peak maxima. Chromatograms were analysed with the Unicorn 5.2 software.

extreme pH value of the application buffer. This problem might be easily fixed in most cases by using protein specific pH gradients with a narrower pH range and hence a less disruptive application buffer.

4.3.2 Proteins with isoforms

Many proteins have isoforms, resulting in charge variants of the single protein. Possible reasons are differences in the amino acid sequence, differences in the oxidative state, post-translational modifications, introducing a charge or masking charged amino acids and many more. Interpretation of the obtained results was done with regard to the separation behavior of the isoforms in the different methods.
The characterization of a protein with isoforms is exemplarily shown in Fig. 6. The illustration shows the analysis of the protein Conalbumin by pH gradient AEC. To obtain the elution-pH values of the single isoforms the pH values at the detected peak maxima were determined.
β-Lactoglobuline shows two known isoforms (A and B) in isoelectric focusing [42]. The proteins differ in their primary structure by two amino acids (A: Asp64 ; B: Gly64). The isoforms were not resolved by either of the two modes of long range pH gradient IEC. The resolution of the long range pH gradient seems to be too low, most probably using a pH gradient with a shorter pH range would resolve the two isoforms.
It is well known that **carbonic anhydrase** has many different isoforms. Up to 16 known isoforms exist in mammals [49] isoforms differing in their primary structure. The results of the pH gradient

IEC of carbonic anhydrase from bovine erythrocytes show three well separated major isoforms. Due to the inhomogeneity of the protein it is not a simple task to assign the peaks to isoforms. Also there is a big difference of the elution-pH between the two IEC modes, which makes it difficult to correlate the isoforms in between the chromatographic separations. Most probably the three major isoforms are related to carbonic anhydrase I, II, III, due to their higher abundancy in erythrocytes [50], whereby isoform II has the highest abundancy. Isoform II would therefore relate to the major isoform with the highest elution-pH value in both chromatographic modes.
Myoglobin showed one major and one minor isoform in pH gradient AEC. The two isoforms are natural variants, varying in their primary structure. Comparing the proteins AEC elution-pH with its pI shows the same result like reported previously. The elution-pH of the neutral protein is significantly higher than its pI [3]. Ahamed et al. explain the difference with the flat shape of the titration curve of neutral proteins, leading to a small change of charges as the pH is changed strongly. This means the attractive interaction of the protein with the resin at pH $\sim$ 9 is already too low to retain the protein. The determination of the elution-pH on the cation exchange resin was not possible due to the denaturation of the hemeprotein at pH 4, which was also reported by others [51].
Thaumatins are extremely basic proteins with two known isoforms, **thaumatin I** and **thaumatin II** [52]. The two isoforms differ in their primary structure. By pH gradient IEC we were able to separate the two major isoforms. The measured elution-pH values differed greatly from the proteins pI, but literature values on the proteins pI are inconsistent. In another reference a pI range from 10.2 - 10.5 [53] is reported. This indicates that the well accepted pI value of 11.7-12.0 is likely to be wrong, which was already proposed by others [54]. Thus, the elution-pH values lie in the range of the proteins pIs.
Human serum albumin is known to form two isoforms due to different amounts of bound lipids, related to a conformational change [45]. The two isoforms are separated by isoelectric focusing and also by pH gradient AEC, but not by pH gradient CEC. A possible explanation is the bound fatty acids. The negatively charged molecules have a much stronger influence on the protein's interaction with the anion exchange resin, than with the cation exchange resin.
Bovine serum albumin has three known major isoforms due to different oxidative states of the intermolecular disulfide bonds resulting in the formation of multimeres [55]. Isoelectric focusing is able to separate three of them, which were also closely separated in pH gradient AEC, there was only one peak resolved in pH gradient CEC.
The chicken egg protein **conalbumin**, also known as ovotransferrin, has three major isoforms. Those are a result of the amount of bound iron ions, as the protein's pI drops with the amount of bound iron [56]. The elution-pH values decrease in the following order, iron-free, mono-ferric and the di-ferric version. pH gradient AEC of conalbumin shows the expected three major isoforms

with almost exactly the same pH values than the pIs. pH gradient CEC was not able to separate the isoforms, also the determined elution-pH is significantly higher. The reason for this might be the fact, that at acidic pH values the iron ions are released from the protein [57]. Therefore the ferric-isoforms are removed in pH gradient CEC and only the iron-free protein is detected, which elutes at a higher pH value in pH gradient CEC. Comparing the AEC elution-pH values with the elution-pH values, determined by Ahamed et al. [3] (pH 6.7, 7.0, 7.5) one will notice, that these values are slightly higher than the values presented here. The reason for this is most probably the lowered ionic strength of the used, optimized buffer, which leads to a reduced effect on the elution behavior of the proteins and consequently lower AEC elution-pH values.

Holo-transferrin, also known as siderophilin has four major isoforms, resulting from post-translational modification with sialic acid from mono- up to hexasialotransferrin. Di- to pentasialoferrine especially tetrasialotransferrin are the most abundant [46, 58]. The references for the pI values [46] were determined from human holo-transferrin and are used here for comparison due to the unavailability of the same data from the bovine protein. Three major isoforms could be separated in pH gradient AEC, while only two were separated in pH gradient CEC. This difference in the selectivity of both chromatographic modes is likely to be a result of the post-translational modification with the negatively charged sialic acid. This modification has a stronger influence on the interaction with the anion exchange, than with the cation exchange resin. Another effect that might also be involved in the lower selectivity of pH gradient CEC is the partial denaturation of the protein at low pH values as it is known that the bound iron ions might be released at low pH values [59].

The hen-egg protein, **ovalbumin** has three known isoforms differing in their electrostatic properties a di-, a mono- and a non-phosphorylated version [47]. Three major isoforms are also detected in pH gradient AEC. A further, minor isoform is detected at a slightly higher elution-pH, probably this chromatographic peak is a result of the glycoforms of ovalbumin [60]. pH gradient CEC was not able to separate the different phosphoforms. This means the interaction with the cation exchange resin is barely affected by the negative charges introduced due to the phosphorylation.

Trypsin inhibitor, also known as ovomucoid, is a heavily glycosylated protein. The five different isoelectric isoforms are related to different glycoforms of the protein [47]. Only one elution-pH could be determined by cation exchange chromatography, while five were determined by pH gradient AEC. This indicates that the glycosylation does not have much influence on the proteins interaction with a negatively charged resin. Ovomucoid is glycosylated at Asn residues by various, negatively charged glycosylations with sialic acid or sulfated carbohydrates [61]. The different glycosylation degrees lead to a varying strength of attractive interaction of the different glycoforms with the anion, but not with the cation exchange resin. This concludes to

the differing selectivity between the two chromatographic modes.
Ribonuclease A and **B** are two isoforms differing in the glycosylation pattern, whereby the B form is glycosylated at Asn34. Both isoforms have an identical pI and identical AEC elution-pH values, but interestingly they differ in the CEC elution-pH. The reason might be the positively charged amino acid, which is masked by the glycosylation and therefore not accessible by the cation exchange resin, thus the glycoform RNAse B elutes at a lower pH than the non-glycosylated RNAse A. This results are in accordance with previously reported results, stating that the two isoforms could be separated by CEC but not by AEC [62].
Catalase from bovine liver is uniform in isoelectric focusing, but in both pH gradient IEC modes, two to three isoforms were detected. The cause for this is likely to be the partial dissociation of the tetrameric protein at low [63] or high pH values [64] resulting in induced isoforms. To determine the detailed composition of the resulting protein fragments, further analysis would be needed.
Summarizing the results from the characterization of the different proteins via pH gradient ion exchange chromatography leads to the following conclusion. It became obvious that different modifications resulting in isoforms have a different impact on the separation efficiency of the single methods. This issue manifests in the differing selectivity of cation and anion exchange chromatography, e.g. the different phosphoforms of ovalbumin could be separated in pH gradient AEC but not CEC. This means that depending on the nature of the modification pH gradient AEC, CEC or isoelectric focusing might be the method of choice. Regarding the results it became clear that pH gradient IEC proved to be a potent tool to separate isoforms based on their different electrostatic properties.

5 Conclusions

In this work, we successfully demonstrated, that the *in silico* optimization of buffer compositions for pH gradient IEC is a fast and simple way to generate buffer compositions for well controllable pH gradients with low ionic strength. With this method we could successfully generate buffer compositions for long range pH gradients, spanning up to 7.5 pH units. The applicability of the buffer compositions was successfully validated by applying them on their corresponding resins. The resulting pH gradients showed a high linearity and reproducibility. Additionally we collected a list of possible buffer substances, enabling us to generate optimized buffer compositions for nearly every pH range.
In the second part we successfully validated, the buffer compositions applicability for the characterization of proteins via pH gradient IEC. As a consequence, we applied pH gradient AEC and CEC for the characterization of 22 proteins, comparing the results with literature values of the

proteins experimental pIs. The results clearly showed that there are major differences between the three methods, due to the proteins interaction with a charged adsorber surface. These differences often resulted in different selectivities of the three methods, motivating a selective use of the methods for specific separation problems. Additionally, pH gradient IEC clearly showed that it can be used as a real alternative for the characterization of protein charge variants.
Still there are many questions on the elution behaviour of proteins in pH gradient IEC left open for future studies. One interesting point would be to study the differences between salt and pH gradient IEC or a combination of both e.g. in terms of chromatographic resolution or robustness. The comparison of protein separation behaviour in pH gradient IEC using different columns, with different lengths and resins is also not treated by this article. The described methodology for the development of buffer systems for pH gradient IEC makes it possible to easily generate applicable buffer systems to work on those or further scientific questions.

Analytical characterization of complex, biotechnological feedstocks by pH gradient ion exchange chromatography for purification process development

Frieder Kröner[a], Alexander T. Hanke[b], Beckley K. Nfor[b], Martijn W. H. Pinkse[b], Marcel Ottens[b], Jürgen Hubbuch[a]

[a] *Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany*

[b] *Department of Biotechnology, Delft University of Technology, Delft, The Netherlands*

Abstract

The accelerating growth of the market for proteins and the growing interest in new, more complex molecules are bringing new challenges to the downstream process development of these proteins. This results in a demand for faster, more cost efficient, and highly understood downstream processes. Screening procedures based on high-throughput methods are widely applied nowadays to develop purification processes for proteins. However, screening highly complex biotechnological feedstocks, such as complete cell lysates containing target proteins often expressed with a low titre, is still very challenging. In this work we demonstrate a multidimensional, analytical screening approach based on pH gradient ion exchange chromatography (IEC), gel electrophoresis and protein identification via mass spectrometry to rationally characterize a biotechnological feedstock for the purpose of purification process development. With this very simple characterization strategy a two-step purification based on consecutive IEC operations was rapidly laid out for the purification of a diagnostic protein from a cell lysate reaching a purity of $\sim 80\%$. The target protein was recombinantly produced using an insect cell expression system.

Journal of Chromatography A (UNDER REVIEW), submitted March 2013

1 Introduction

In the past decade research and development in biopharmaceutical process engineering was strongly focused on the optimization of routine production platforms for monoclonal antibodies (mAbs). In terms of upstream process development, focus was on improving the expression organism towards ever increasing titres. To our knowledge, the current largest mAb titre is 27 g/l of cell culture volume (XD® process by DSM with PER.C6® cell line from Crucell). To eliminate time consuming process development on the downstream processing side so called platform processes were developed centering around the initial capture step using Protein A affinity chromatography. A typical purification platform consists today of two to three sequential chromatographic steps, Protein A affinity chromatography, followed by polishing based on anion and/or cation exchange chromatography (AEC/CEC) [19]. The high titres of the antibodies, the low concentration of contaminants due to the extracellular expression of the molecule, as well as the platform process with an affinity-based purification step, has made the production of mAbs a rather simple, routinely performed task.

However, this achievement overshadows challenges exhibiting more complex tasks and diverse feedstocks for the production of enzymes, protein hormones, cytokines, protein vaccines and diagnostic proteins [11]. Using protein tags for simplified purification of those targets is very often not an option due to potential risks for the therapeutic's safety [65, 66]. A more general use of expression organisms, potential intracellular products as well as more complex and diverse products will result in lower product titres, and higher contaminant concentrations. Furthermore, a general platform process for the purification of all different proteins is unlikely to be realized in the near future.

In the light of the described situation, various approaches for purification process development, such as heuristic (knowledge-based), experimental (high-throughput screening (HTS)), model-based approaches and combinations of these, are extensively discussed in literature [17]. One trend in downstream processing of biopharmaceuticals that could clearly be observed in the last few years is the progress towards more rational synthesis and design methodologies [18, 22, 67–69]. This means that the development of screening approaches that provide a rational methodology to acquire the relevant information needed to rapidly set up modular purification processes for different targets. In a recent publication, Nfor et al. [22] proposed the use of different chromatographic screenings and LC-MS based protein identification to acquire the purification relevant physicochemical parameters of the target protein as well as the identified contaminants from a cell culture supernatant containing a monoclonal antibody. The acquired physicochemical parameters can be classified into so called "heuristic parameters" and "mechanistic parameters". Heuristic parameters can be used to empirically synthesize purification steps knowledge-based,

while mechanistic parameters allow *in silico* prediction of the mixture's chromatographic behaviour. In the previous work by Nfor et al. [22] two different methods are used to acquire the said parameters for IEC: 1. pH gradient IEC to determine the mAb's elution-pH to rationally choose the pH for salt gradient IEC [3]; 2. IEC, using multiple, linear salt gradients to estimate the parameters for the simulation of IEC operations [70, 71]. However, using a comparable characterization approach for the previously described feedstock would pose several, mainly analytical challenges due to the significantly increased complexity of the mixture and the strongly decreased target protein concentration.
In this work we show the use of an adapted screening procedure based on optimized pH gradient IEC [72] and LC-MS based protein identification to characterize a highly complex, complete cell lysate containing a diagnostic protein with a very low expression level. A detailed description of the screening procedure is given in the next section. The obtained data was subsequently used to lay out a simple two step purification of the target protein based on IEC to evaluate the applicability of the developed method.
The diagnostic protein is a 45 kDa fragment of the human protein Nucleolin (Ncl, GI number 55956788), containing all the relevant epitopes for the diagnosis of the autoimmune disease systemic lupus erythematosus (SLE). The target protein is recombinantly expressed in a SF9/Baculovirus expression system.

1.1 The screening procedure

Characterizing highly complex biotechnological feedstocks with low product titres poses several challenges, mainly caused by the low target protein concentration and the high number of contained proteins. To address these challenges, a general strategy consisting of three experimental steps is suggested.
First, the elution-pH values of the target protein in AEC / CEC are determined by fractionating the cell lysate using long range pH gradient AEC (pH 10.5-3.5) / CEC (pH 4.0-11.0). A target specific analytical technique is subsequently applied to determine the fractions containing the target protein. Possible analytical techniques are immunoassays or LC-MS based methods.
The determined elution-pH values are used to select the pH range (elution-pH +/- 1 pH unit) for short range pH gradient AEC / CEC. By submitting the cell lysate to short range pH gradient IEC, proteins with electrostatic properties comparable to those of the target protein, are specifically fractionated. Significantly concentrated proteins contained in the fractions of the short range pH gradient fractionation represent the relevant impurities for the purification of the target protein, which will be difficult to remove by IEC. To define those impurities, the fractions are further analysed by SDS PAGE after a concentrating step. The concentrating step

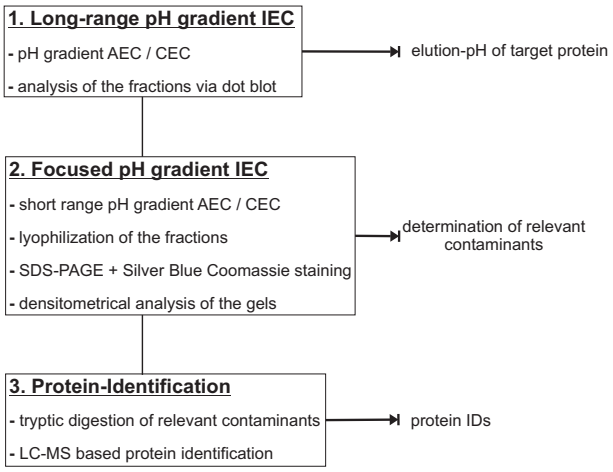

Figure 1: Screening procedure for the SF9/Baculovirus cell lysate containing Ncl

(e.g. lyophilization) is mostly mandatory as only small amounts of the biotechnological feedstock are analysed to keep the procedure in analytical scale.
Finally, the selected, relevant contaminants are identified via tryptic digestion and LC-MS. The obtained informations are, the protein IDs, the elution-pH values in AEC / CEC, the electrophoretic molecular weight, the molecular weight derived from the used protein database as well as the calculated pI of each relevant protein. These informations provide a basis for rationally designing early purification processes based on IEC. The whole process is illustrated in Fig. 1 for the presented case. The screening procedure is applied on the insect cell lysate containing the recombinantly expressed protein Nucleolin. On the basis of the obtained information an early purification process of Nucleolin by IEC is laid out.

2 Materials and methods

2.1 Chemicals

The chemicals, acetonitrile, Tris, NaCl, DTT, SDS, phosphoric acid, ammoniumbicarbonate, DMSO, 1-methylpiperazine and the enzyme benzonase, were purchased from Sigma-Aldrich (St. Louis, MO, USA). The substances, ammoniumsulfate, EDTA, urea, methylamine, hydroxylamine, 1,2-ethanediamine, 1,4-dimethylpiperazine, hydrochloric acid, sodium hydroxide, formic acid and acetic acid were purchased from Merck (Darmstadt, Germany). The buffer substances for pH gradient CEC (section 3.2), Triton X-100 and Tween-20 were purchased from Applichem (St. Louis, MO, USA). Coomassie G250 was purchased from Bio-Rad (Hercules, CA, USA). Bis-tris was purchased from Molekula (Molekula, Dorset, UK). All used buffer substances were

of analytical grade purity.

2.2 Buffers

SF9 lysis buffer: The buffer for the SF9 cell lysis consisted of 20 mM Tris-Cl at pH 8.0, 150 mM NaCl, 0.5% (v/v) Triton X-100, 2 mM EDTA and 1x SigmaFast protease inhibitor (Sigma-Aldrich St. Louis, MO, USA).
Solubilization buffer: The solubilization buffer for the freeze-dried samples consisted of 6 M urea, 100 mM Tris, 5 mM EDTA and 0.5% (w/v) SDS, the buffer was adjusted to pH 8.0.
Long range pH gradient AEC: The composition of the buffer for pH gradient AEC pH 10.5 – 3.5 is shown in the following table.

Substance	Conc.[mM]
Methylamine	9.8
1,2-Ethanediamine	9.1
1-Methylpiperazine	6.4
1,4-Dimethylpiperazine	13.7
Bis-tris	5.8
Hydroxylamine	7.7
+ 10 mM sodium chloride	

Long range pH gradient CEC: The composition of the buffer for pH gradient CEC pH 4.0 – 11.0 is shown in the following table.

Substance	Conc.[mM]
CAPS	15.6
CHES	9.4
TAPS	4.6
HEPPSO	9.9
MOPSO	8.7
MES	11.0
Acetate	13.0
Formate	9.9
+ 200 mM sodium chloride	

Short range pH gradient AEC: The composition of the buffer for pH gradient AEC pH 5.5 – 3.5 is shown in the following table.

Substance	Conc.[mM]
Piperazine	15.2
1,4-Dimethylpiperazine	16.7
+ 10 mM sodium chloride	

Short range pH gradient CEC: The composition of the buffer for pH gradient CEC pH 9.5 – 11.5 is shown in the following table.

Substance	Conc.[mM]
CHES	15.2
CAPS	16.7
+ 200 mM sodium chloride	

The buffer systems for pH gradient IEC were systematically developed as it is reported in a previous publication [72]. To set up the buffers for the pH gradient IEC, all substances were weighed in, dissolved in ultrapure water and split in two equal volumes. One part was titrated to the low pH extreme, the other to the high pH extreme, with the appropriate strong titrant.
Buffers for dot-blot development: For the development of the dot-blots, TBS (20 mM Tris, pH 7.5, 150 mM NaCl) and TBS-T (20 mM Tris, pH 7.5, 500 mM NaCl, 0.5%(w/v) Tween-20) were used.
All buffer solutions were prepared with ultrapure water from a MilliQ water purification system from Millipore (Billerica, MA, USA). The pH adjustment was carefully performed with a freshly, five-point calibrated pH meter (HI-3220, Hanna Instruments, Woonsocket, RI, USA). The pH calibration buffers, pH 3, 5, 7, 9, 11, were high precision standards from Hanna Instruments. After adjusting the pH, the buffers were brought to their final volume. Buffers used for chromatographic separations were degassed and filtered using vacuum filtration with a 0.2 μm cellulose acetate filters from Sartorius Stedim (Goettingen, Germany).

2.3 Cell lysis

Cultivation and infection of the SF9 insect cells for the recombinant production of the nucleolin fragment was done by the industrial partner, Diarect AG (Freiburg, Germany). For the preparation of the cell lysate, cell pellets of 50 ml cultivation broth were dissolved in lysis buffer

with a ratio of 1 : 5 (w/w), by shaking the suspension for 10 min on an overhead shaker. After dissolving the pellet, the cells were lysed by sonication: 6x 15 s sonication at 40% amplitude, with 30 s cooling steps in between. Sonication was carried out on ice with a Branson Sonicator 450D equipped with a 1/8 inch microtip horn (Branson Ultrasonics, Danbury, CT, USA). After lysis, 10 mM $MgCl_2$ and 1 μl of benzonase ($\triangleq$ 250 units) were added to digest the DNA in the lysate. The cell lysate was then incubated for 30 min on an overhead shaker at room temperature. Particulates were removed by centrifugation for 30 min at 40.000 g. The prepared cell lysates were then aliquoted and stored at -32°C. The cell lysates contained ∼10 g/l of protein and ∼10–100 mg/l of the target protein.

2.4 Sample preparation for LC

The frozen cell lysates were thawed in a sonication bath (Branson Ultrasonics, Danbury, CT, USA) to prevent strong formation of precipitates. Afterwards, particulates were removed by centrifugation. To prepare the cell lysates for subsequent chromatographic separations, a buffer exchange via PD-10 columns (GE Healthcare, Uppsala) into the loading buffer of the following chromatographic step, was performed. The lysates were then filtered with 0.2 μm cellulose acetate filters (Sartorius Goettingen Germany).

2.5 Long range pH gradient fractionation / elution-pH determination

The long range pH gradient IEC was performed to fractionate the crude lysate for the determination of the target protein's elution-pH. The chromatographic runs were done on an Äkta Purifier (GE Healthcare, Uppsala, Sweden) and monitored with online UV and pH measurement. The used chromatographic columns were a MonoQ 4.6/100 and a MonoS 4.6/100 column (GE Healthcare, Uppsala, Sweden). Long pH gradient AEC was run from pH 10.5 – 3.5 and pH gradient CEC was run from pH 4.0 – 11.0. The chromatographic sequence consisted of the following steps: 1. Equilibration for 5 CV; 2. Injection of 1 ml cell lysate; 3. Wash-step for 10 CV in AEC and 20 CV in CEC; 4. Gradient 0 – 100% in 15 CV; 5. Post-gradient at 100% for 5 CV 6. Cleaning-step with 1N NaCl for 5 CV. The flow rate was kept constant at 1.5 ml/min. For CEC the volume of the wash-step was increased to 20 CV to wash out all unspecifically bound contaminants until a stable UV 280 nm baseline was reached. A fraction collection device (GE Healthcare, Uppsala, Sweden) was used to collect 1 ml fractions throughout the gradient. The flow through step, the post-gradient step and the wash step were collected in 50 ml tubes. The collected fractions and 1 ml of each 50 ml tube were stored at -32°C in a 96-deepwell MTP (Thermo Scientific, Waltham, USA) for further analysis by dot-blot.
To identify the nucleolin-containing fraction a dot-blot was performed. 1 μl of each fraction

was spotted on a nitrocellulose membrane (Bio-Rad, Hercules, CA, USA). After washing the membrane twice with TBS (5 min) the membrane was blocked for 1 h with 3% (w/v) BSA in TBS. Subsequently the membrane was washed twice with TBS-T and once with TBS. The primary antibody (detecting the 6-His-tag of Nucleolin), a mouse-anti-pentaHIS IgG (Qiagen GmbH, Hilden, Germany), was diluted 1:1000 in blocking solution and then applied for 1 h. The second antibody, an anti-mouse-IgG alkaline phosphatase conjugate (Thermo Scientific, Waltham, USA) was applied for 1 h at 1:10000 dilution ratio in blocking solution after a washing step. To develop the dot-blot it was washed first and then stained for 10 min with an alkaline phosphatase color development kit (Bio-Rad, Hercules, USA). All steps were carried out on a rocking shaker.

2.6 Short range pH gradient fractionation

After determining the elution-pH of the target protein, a pH gradient IEC method with a narrower pH range around the determined elution pH (+/–1 pH unit of the target protein's elution-pH) was designed. The chosen pH ranges were 5.5 – 3.5 for AEC and 9.5 – 11.5 for CEC. The buffer compositions are described in section 3.1. The chromatographic runs were performed according to the description in section 2.5, with two minor changes: the volume of the injected lysate was raised to 2 ml per run and the flow-through step to 20 CV. The collected fractions and 1 ml of each 50 ml tube were frozen at -82°C for minimum 2 h. The pre-frozen samples were subsequently submitted to a freeze dryer (Labconco, Fort Scott, USA) and concentrated to complete dryness. The samples were then stored at -32°C until further analysis.

2.7 Identification of the target protein and the relevant contaminants

The chromatographic fractions of the short range pH gradient were analysed to define and identify the relevant contaminants and the target protein. The fractions were dissolved under heavy shaking in 100 μl solubilization buffer (see 3.2) for 30 min on a thermo shaker (Bio-Rad, Hercules, CA, USA) at 37 °C. This resulted in a 10x concentration compared to the initial fractions. The resolubilized samples were then prepared for the SDS PAGE according to the supplier information for 4-12% Bis-Tris NuPAGE gels under non-reducing conditions (Invitrogen, Carlsbad, CA, USA). A protein standard (Mark12, Invitrogen, Carlsbad, CA, USA) was used to estimate the molecular weight of the proteins. The gels were stained for 24 h with the highly sensitive "Blue silver staining", a colloidal coomassie staining method [73]. The gels were scanned in a Gel Doc XII station (Bio-Rad, Hercules, USA) after destaining with ultrapure water. The gels were analyzed densitometrically via ImageJ (NIH, Bethesda, USA). All proteins with an intensity $\geq$ 20% of the intensity of the most prominent protein (Nucleolin) were defined as relevant

contaminants, due to their concentration and their comparable electrostatic behaviour. The corresponding bands were excised and identified via tryptic digestion and LC-MS, as described in the sections 2.8 & 2.9.

2.8 LC-MS sample preparation

The excised protein bands were destained by shrinking them for 10 min at room temperature in 100 μl 25 mM ammoniumbicarbonate, 50% (v/v) acetonitrile. The solution was removed and the gels were reswollen in 25 mM ammoniumbicarbonate. This procedure was repeated until the gel was completely destained. The destained gel slices were shrunk for 10 min in acetonitrile. Afterwards, they were immersed in 25 mM ammoniumbicarbonate with 10 mM dithiothreitol (DTT) for 30 min to reduce the disulphide bonds of the proteins. To inhibit the reformation of the disulphide bonds by alkylation, the gel slices were shrunk again and then incubated in the absence of light with 25 mM ammoniumbicarbonate buffer containing 20 mM iodoacetamide for 45 min. The gel slices were shrunk in ice cold acetonitrile (30 min, on ice) and reswollen (30 min, on ice) in 25 μl trypsin solution, which consisted of 20 μg/ml trypsin in 40 mM ammoniumbicarbonate with 9% acetonitrile. After reswelling the gel pieces (30 min) with trypsin solution, another 50 μl 25 mM ammoniumbicarbonate buffer was pipetted on the gel. The samples were then digested for 24 h at 37C on a thermoshaker (Eppendorf, Hamburg, Germany). The liquid was separated from the gel pieces after digestion and stored at -32°C until LC-MS analysis. All digestion steps were performed very carefully in a clean environment to avoid possible contaminations.

2.9 Identification of proteins by LC-MS

After digestion, formic acid and DMSO were added (both 5% v/v) to increase peptide recovery. Protein digests were analysed by reversed phase nano-LC coupled to a LTQ Orbitrap Velos (Thermo Fisher Scientific, Bremen, Germany). An Agilent 1200 series HPLC system was equipped with an in-house packed trapping column (100 μm ID and 20 mm length) and analytical column (50 μm ID and 250 mm length) filled with Reprosil Pur 120 C18-AQ (Dr. Maisch, Ammerbuch-Entringen, Germany) essentially as described by Meiring et al. [74]. Trapping was performed at 5 μl/min for 10 minutes in solvent A (0.1 M acetic acid), and elution was achieved with a gradient from 0 to 40% solvent B (0.1 M acetic acid in 8:2 v/v acetonitrile:water) for 40 minutes. The LTQ Orbitrap Velos was operated in data dependent mode automatically switching between MS and MS/MS. Survey full scan MS spectra were acquired from m/z 400 to 1500 in the Orbitrap with a resolution of 30,000 at m/z 400 after accumulation to a target value of 1e6 in the linear ion trap. The ten most intense, multiply charged ions at a threshold of

above 1000 were fragmented in the linear ion trap using collision-induced dissociation (CID) at a target value of 1e4. All raw data files were processed into peaklists using Proteome Discoverer 1.1. Mascot 2.2 was used to search the spectra against the NCBInr database using taxonomy restrain from human, moth and viral proteins. To improve the certainty of the results, the Mascot data files were additionally re-analysed in Scaffold 3.6 (Proteome Software, Portland, OR, USA) by using the X!Tandem algorithm with the same database. pH gradient IEC fractionation, SDS-PAGE and LC-MS were performed two to three times for protein samples with uncertain identification to decrease the possibility of falsely identified proteins.

2.10 Two-step purification of Nucleolin

The collected data on the elution behaviour of the target protein and the contaminants was finally used to generate a crude purification process consisting of a CEC capture and an AEC purification step.

CEC capture step: Proteins were captured at pH 8.75 (20 mM TAPS), 200 mM NaCl on a SP Sepharose FF column (GE Healthcare, Uppsala, Sweden). The used column was a custom packed Kronlab class column (YMC, Wilmington, NC, USA) with an inner diameter of 20 mm and a bed volume of 7.5 ml. Elution was carried out with a salt step from 200 – 500 mM NaCl. The chromatographic operation was performed on an Äkta purifier, while monitoring the UV 254 nm and 280 nm signals as well as the conductivity. The chromatographic sequence was: 1. Equilibration step for 2 CV; 2. Injection of 10 ml lysate; 3. Wash-step for 4.5 CV; 4. Step-elution 0 – 100% hold for 2.75 CV; 5. Cleaning-step with 1N NaCl for 2.5 CV. All steps were performed at 1.25 ml/min except the injection step which was performed at 0.5 ml/min. The elution fraction (7.5 ml) and the flow-through fraction were collected. The collected elution fraction was first 1:2 diluted using CEC running buffer at pH 8.75 without salt and further concentrated to a volume of 2 ml using centrifugal concentrators, VivaSpin 20 (Sartorius) with a molecular weight cutoff of 3 kDa according to the manufacturers instructions.

AEC purification step: A salt gradient elution at pH 5.2 (20 mM piperazine) from 0 – 1000 mM NaCl was used to purify the sample further after the capture step. The used chromatographic column was a 1 ml (5 mm ID x 50 mm L) Q Sepharose FF (GE Healthcare, Uppsala, Sweden) column from Atoll (Weingarten, Germany). The chromatographic operation was also performed on an Äkta purfier, while monitoring the same signals as for the CEC capture step. The chromatographic sequence was: 1. Equilibration step for 5 CV; 2. Sample load (conc. CEC elution fraction); 3. Wash-step for 10 CV; 4. Elution with a salt gradient of 0 – 5% in 0 CV, 5 – 50% in 25 CV and 50 – 100% in 0 CV, hold for another 5 CV. The proteins eluted by the salt gradient were collected in 1 ml fractions, while the flow-through and the cleaning step were

collected in 50 ml tubes.

Dot-Blot analysis: All collected samples had to be brought to identical volumes for better semi-quantitative interpretation of the results. The CEC flow-through fraction was concentrated to a volume of 10 ml using VivaSpin 20 centrifugal concentrators (Sartorius) with a molecular weight cutoff of 3 kDa. Samples with a volume of 5 μl were taken of the concentrated CEC elution fraction and diluted 1:5, while 5 μl samples of each AEC fraction were diluted 1:10 with ultrapure water. After bringing all samples to a comparable concentration in relation to the injected cell lysate (10 ml) 1 μl of each was dotted on a nitrocellulose membrane (BioRad) and developed as previously described in section 2.5.

SDS PAGE analysis: The samples were analysed by SDS PAGE and Coomassie staining to control the purification qualitatively. The concentrations of all collected AEC fractions were increased ten times by freeze-drying and resolubilization prior to SDS PAGE analysis as described in section 3.6. Preconcentration of samples was carried out to improve result interpretation by the following SDS PAGE analysis. Samples of the cell lysate, the CEC elution fraction and the concentrated CEC elution fraction were untreated for analysis by SDS PAGE. The SDS PAGE & the Coomassie staining were prepared as described previously (section 2.7). The gels were densitometrically analyzed using ImageJ.

3 Results and discussions

3.1 Long range pH gradient fractionation / elution-pH determination

To determine the elution-pH of Nucleolin from the lysate, pH gradient AEC / CEC screenings were performed as described in section 2.5. Dot-blot analysis was used for the identification of the Nucleolin containing fractions. The results are shown in Fig. 2. The AEC elution range of Nucleolin was pH 4.9 – 4.3, while the CEC elution was in the range of pH 9.7 – 10.6 at a sodium chloride concentration of 200 mM. The protein was only eluted from the CEC column by the 1M NaCl wash step when pH gradient CEC was performed without the addition of NaCl. This reveals a surprisingly big difference between the protein's elution behaviour on the different IEC columns. According to a previous publication, shifts between AEC and CEC elution-pH values are likely to be caused by an anisotropic charge distribution on the protein surface [72]. This might be reasoned by the fact, that the surface charge distribution of the protein dictates the binding orientation of the protein to the charged surface [75–79]. It is known that the diagnostic nucleolin fragment consists of the four RNA-binding domains RRM1-4. When taking the above findings into account, it is highly probable that the very locally, positively charged RNA-binding sites have an extreme effect on the protein's surface charge distribution and therefore its elution

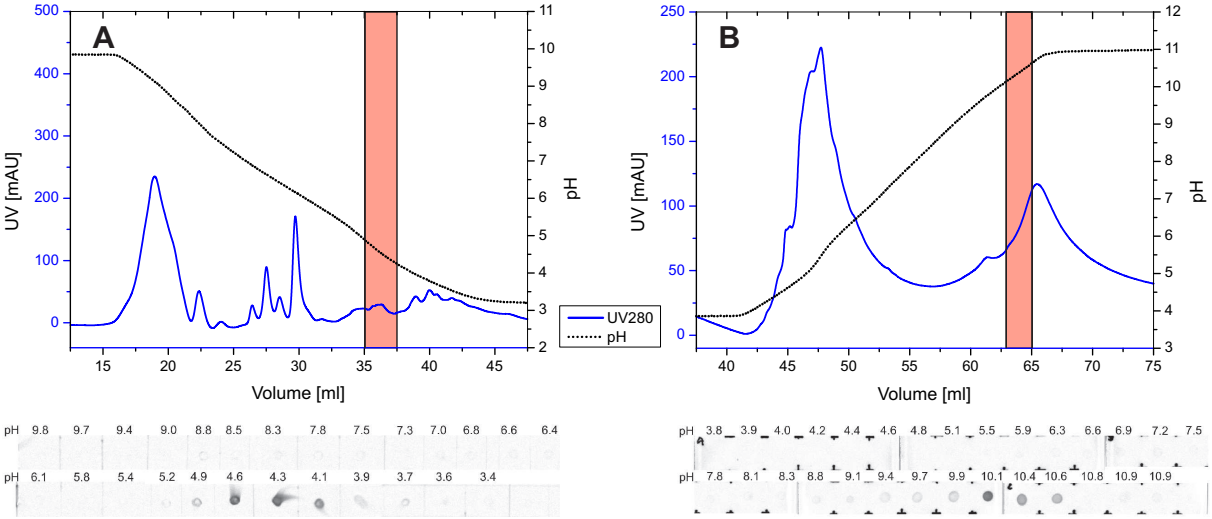

Figure 2: pH gradient AEC/CEC screening for elution-pH determination of nucleolin from the cell lysate: **A**: pH gradient AEC pH 10.5 – 3.5, 10 mM NaCl and dot-blot analysis to determine fractions containing the target protein. The elution range is shaded in the chromatogram; **B**: pH gradient CEC pH 4.0 – 11.0, 200 mM NaCl and dot-blot analysis to determine fractions containing the target protein. The elution range is shaded in the chromatogram;

behaviour in IEC. This fact is likely to be the explanation for the exceptionally strong binding of the protein to the negatively charged CEC resin by the positively charged RNA-binding sites as well as the large variation between the elution-pH values in AEC and CEC. The extreme difference found for Nucleolin with respect to calculated pI (calc. pI: 8.56) and actual elution behaviour clearly demonstrates that using the elution-pH for IEC process development has a significant advantage over using the protein's pI. Utilizing the pI of the protein for IEC process development in the presented case would have led to a serious misjudgement of process parameters. This is reasoned by the fact that the electrostatic interaction of the protein with an oppositely charged solid phase can not be explained only by the protein's net charge. Also the surface charge distribution of the protein influences binding of the protein to the charged surface which was already reported by others [75–79].
The two determined elution-pH values were further used in the presented case to choose the pH range for the short pH gradient IEC as well as the pH and salt concentration of the subsequently developed IEC purification steps.

3.2 Identification & characterization of the target protein and the relevant contaminants

The most problematic contaminants for the purification of a target protein are the proteins with closely similar physicochemical properties, such as the electrostatic properties, in ion exchange

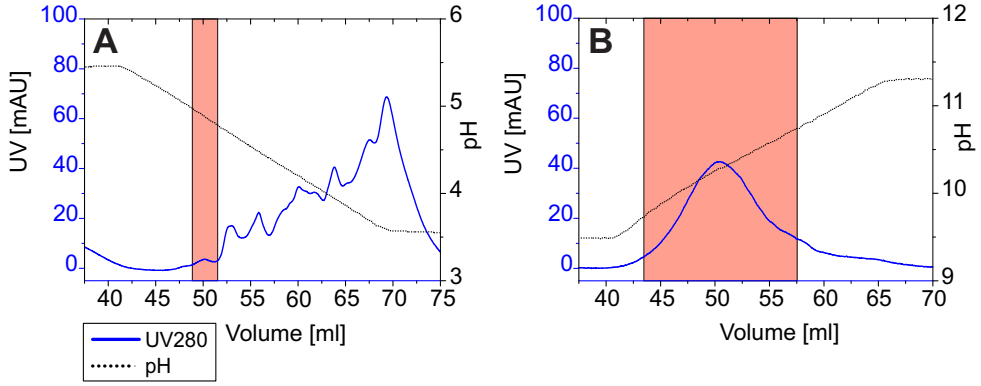

Figure 3: focused pH gradient AEC/CEC: **A**: pH gradient AEC pH 5.5 – 3.5, 10 mM NaCl, Nucleolin elution shaded; **B**: pH gradient CEC pH 9.5 – 11.5, 200 mM NaCl, Nucleolin elution shaded;

chromatography. By running a focused pH gradient IEC with a shorter pH range it was possible to fractionate the lysate with increased chromatographic resolution. The proteins, eluting closely to the target protein are the proteins with very similar electrostatic properties.

The applied pH ranges were 5.5 – 3.5 for pH gradient AEC and 9.5 – 11.5 for pH gradient CEC. The pH gradient CEC was run at an ionic strength of 200 mM sodium chloride. The pH was monitored to determine the pH of the collected fractions and therefore the elution-pH values of the eluted proteins. The chromatograms of the focused pH gradients are shown in Fig. 3. To define the relevant contaminants the fractions were further separated by SDS-PAGE. The gels were stained and analysed densitometrically. The criteria for the definition of the relevant contaminants was an intensity $\geq$ 20% of the protein with the strongest intensity, in this case Nucleolin. The gels with the defined relevant contaminants of both pH gradient fractionations are shown in Fig. 4.

To identify the relevant contaminants, the corresponding protein bands were excised, tryptically digested and the resulting peptides were identified by LC-MS and database searching. The identified proteins are listed with the determined values for each chromatographic mode in Tab. 1. The theoretical protein pIs were calculated from the proteins' amino acid sequence using Compute pI/Mw (www.expasy.org). The proteins' sizes were also determined via gel electrophoresis allowing the comparison with the theoretical masses which were determined using the sequence of each identified protein.

As expected, proteins from the host organism as well as proteins from the viral expression vector were identified. Regarding the two pH gradients, it was expected that less proteins bind in pH gradient CEC, as the applied conditions (pH 9.5 – 11.5, 200 mM NaCl) probably prevent many proteins from binding to the column. For this reason, more relevant contaminants were found in the AEC fractions compared to the CEC fractions. Some identified proteins, especially the

proteins involved in protein synthesis, like HSP70/HSC70 are already known to be overexpressed after infection with the baculovirus [80].

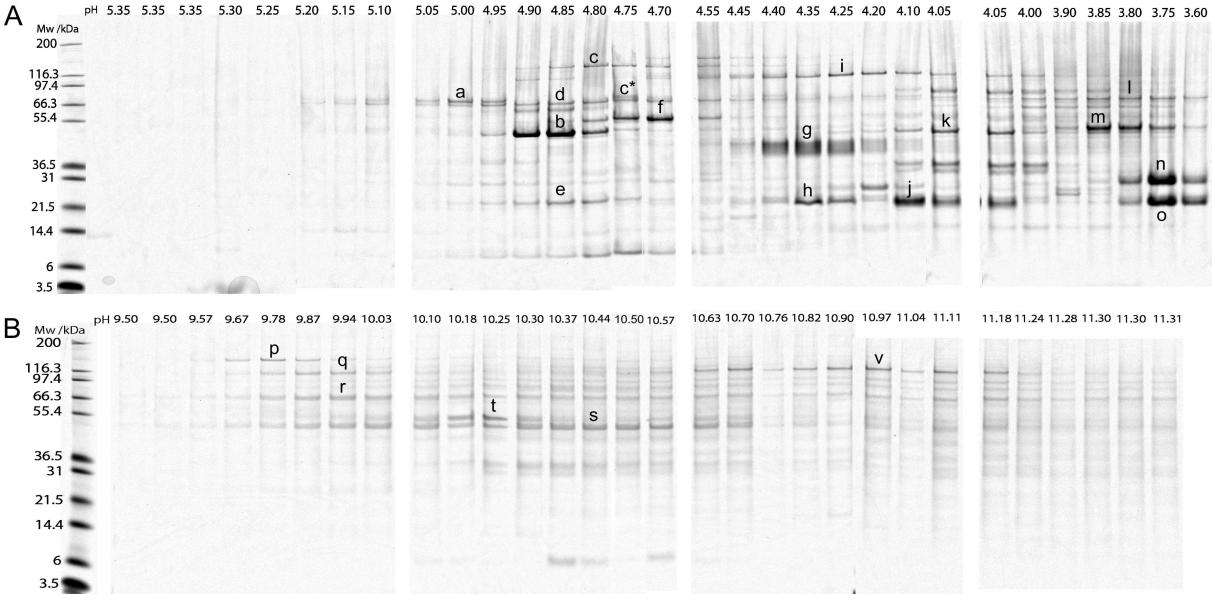

Figure 4: SDS PAGE analysis of the pH gradient IEC fractions: **A**: pH gradient AEC pH 5.5 – 3.5, 10 mM NaCl, ProteinIDs of the relevant contaminants a-o are shown in Tab. 1
B: pH gradient CEC pH 9.5 – 11.5, 200 mM NaCl, ProteinIDs of the relevant contaminants p-v are shown in Tab. 1

Several proteins (HSP70, Moesin, LEF-3) were identified multiple times on different positions in the gel, with slightly different electrophoretic, molecular size and elution-pH values. HSP70 is likely to exist in highly homologous variants, as it does in other species [81]. Those variants are very difficult to distinguish via LC-MS based identification, as the unique peptides have to be detected. The identification of highly homologous variants is further complicated by the fact that the proteomic database of a higher taxonomical class had to be used due to the unavailability of a complete SF9-cell proteome. Therefore, the identified HSP70 proteins at position a and d are likely to be different HSP70 variants, which could not be distinguished by the LC-MS based identification.
The viral protein LEF-3 (Late-expression factor 3) is detected with different molecular weights. The explanation for this is possibly the differing oxidation states of the protein. LEF-3 is known to exist as a dimer due to intermolecular disulphide bonds as well as monomer with reduced or non-reduced disulphide bonds [82]. This might result in variants with slightly different chromatographic elution behaviour and different molecular sizes, as the sample preparation is done under non-reductive conditions. The monomer with intramolecular disulphide bonds was reported to migrate slightly faster than the reduced monomer as a diffuse band in gelelectrophoresis

Table 1: Protein IDs

focused pH gradient AEC pH 5.5 - 3.5, 10 mM NaCl

ID	Protein	Database	GI number	MWgel	MWms	AEC-elpH	AEC-ΔpH	pI
a	HSP70	moth	305693941	81 kDa	72 kDa	5.00	5.10-4.95	5.32
b	Nucleolin	human	55956788	52 kDa	45 kDa	4.85	4.95-4.75	8.56
c, c*	Moesin	moth	122920502	70, 146 kDa	70 kDa	4.80	4.90-4.40	5.90
d	HSP70	moth	305693941	72 kDa	72 kDa	4.85	5.10-4.40	5.32
e	Thioredoxin peroxidase	moth	159459926	22 kDa	22 kDa	4.85	4.95-4.75	5.93
f	Protein disulfide-isomerase like protein	moth	62241290	58 kDa	55 kDa	4.70	4.80-4.55	4.60
g	LEF-3	AcNPV	9627810	43 kDa	45 kDa	4.35	4.45-4.20	5.14
h	TCTP	moth	112982880	20 kDa	20 kDa	4.35	4.40-4.25	4.66
i	LEF-3	AcNPV	9627810	120 kDa	45 kDa	4.25	4.55-4.00	5.14
j	eIF-5A	moth	51702278	22 kDa	18 kDa	4.10	4.10-4.05	5.16
k	LEF-3	AcNPV	9627810	51 kDa	45 kDa	4.05	4.10-4.05	5.14
l	HSC70	moth	27260894	72 kDa	73 kDa	3.80	4.00-3.60	5.20
m	Calreticulin precursor	moth	28804517	48 kDa	46 kDa	3.85	3.90-3.75	4.49
n	Occlusion-derived virus envelope protein	AcNPV	9627837	28 kDa	26 kDa	3.75	3.80-3.60	5.94
o	Telokin-like protein-20	AcNPV	9627825	20 kDa	20 kDa	3.75	3.80-3.60	4.46

focused pH gradient CEC pH 9.5 – 11.5, 200 mM NaCl

ID	Protein	Database	GI number	MWgel	MWms	CEC-elpH	CEC-ΔpH	pI
p	Hypothetical protein KGM_03313	moth	357626559	142 kDa	132 kDa	9.78	9.67-9.87	6.29
q	26S proteasome regulatory subunit RPN1	moth	357612280	115 kDa	110 kDa	9.94	9.78-10.03	5.34
r	Moesin	moth	122920502	67 kDa	70 kDa	9.94	9.78-11.31	5.90
s	Nucleolin	human	55956788	52 kDa	45 kDa	10.44	9.78-10.70	8.56
t	ATP synthase	moth	114052278	54 kDa	60 kDa	10.25	10.10-10.30	9.21
v	Kinesin heavy chain	moth	182511222	115 kDa	108 kDa	10.97	10.57-11.18	5.82

[82]. According to this we can conclude that most probably, g = LEF-3 with intramolecular disulphide bonds, i = LEF-3 as a dimer with intermolecular disulphide bonds and k = LEF-3 with free disulphide bonds.

The protein Moesin is found with an electrophoretical size of 70 and 146 kDa. This issue might be caused by the fact that Moesin is forming homodimers in the anion exchange chromatography step which are also stable under denaturing conditions. Those dimers were not detected in pH gradient CEC, probably caused by dissociation of the dimers due to the extreme pH. The stable formation of Moesin-Moesin homodimers has been reported previously [83].

Comparing the results of both focused pH gradients it becomes obvious that there is only one contaminant, Moesin, detected in the fractions of both chromatographic modes. Thus, the protein will be especially difficult to remove via ion exchange chromatography due to its highly similar chromatographic behaviour. Regarding the quality of the two chromatographic separations the CEC mode might be a highly effective step to capture the target protein. The conditions at which the target protein still binds to the cation exchange column exclude the binding of the majority of the proteins contained in the feedstock. Therefore a CEC capture step would have high binding capacity and could be performed with a high volume of the cell lysate. The results show that a combination of both chromatographic modes should already lead to a very pure product as only one relevant protein contaminant shows comparable chromatographic behaviour

in both IEC modes.

3.3 Purification of the target protein

As a proof-of-principle, we used the results from the analytical screening to set up a simple, small scale, two-step purification process based on salt gradient ion exchange chromatography. Salt gradient IEC was preferred over pH gradient IEC for several reasons. An important fact is the much wider acceptance of salt gradient IEC in industrial, preparative bioseparation engineering compared to pH gradient IEC. Also precipitation effects might occur in pH gradient IEC as the pH is changing throughout the gradient, resulting in strongly reduced column lifetimes.
The suggested two-step purification consisted of a CEC capture step with salt step elution and a subsequent AEC purification step with salt gradient elution.
The CEC capture step was conducted at pH 8.75, 200 mM NaCl, bound proteins were eluted by a salt elution step from 200 – 500 mM sodium chloride at pH 8.75 after washing out unbound sample. No buffer exchange was necessary to bind the protein to the column (Lysate: pH 8.0, 150 mM NaCl). Due to the pH gradient screening, the chosen conditions were known to allow binding of the target protein to the column, while binding of the majority of proteins contained in the lysate is prevented. The latter ensures a high binding capacity for the target protein. A safety margin of one pH unit between the pH of the mobile phase and the previously screened elution-pH range of the target protein (pH 9.78-10.70) was included to prevent its loss due to isocratic elution, while washing out unbound proteins by the washing step. A possible loss of the target protein during the wash-step was indicated by its broad elution in the pH gradient CEC screening (Fig. 4, Tab. 1). The collected elution fraction was first diluted 1:2 with salt free CEC running buffer pH 8.75 to halve the ionic strength and then concentrated 7.5 times by ultrafiltration to reduce sample volume for further purification by salt gradient AEC. Salt gradient AEC was chosen to purify the target further, using a very long, shallow gradient for elution in order to achieve separation of the critical contaminant Moesin from the target protein. The difference between the elution-pH values of both proteins is slightly bigger in AEC than in CEC (ref. Tab. 1, Fig. 4) increasing the possibility to achieve separation. Also the chromatographic elution profile of both proteins was much sharper in pH gradient AEC than CEC, indicating a higher chromatographic selectivity of the AEC resin. The conditions for salt gradient AEC were pH 5.2 and a salt gradient from 50 – 500 mM NaCl after washing out unbound proteins at pH 5.2 with 0 mM sodium chloride. Due to the sharp elution profile of Nucleolin (pH 4.95-4.75) in pH gradient AEC a safety margin of 0.3 pH units of the mobile phase was chosen to be sufficient to inhibit elution of Nucleolin, while washing out weakly or unbound proteins. Sample conditions were pH 8.75 with a salt concentration of 250 mM sodium chloride.

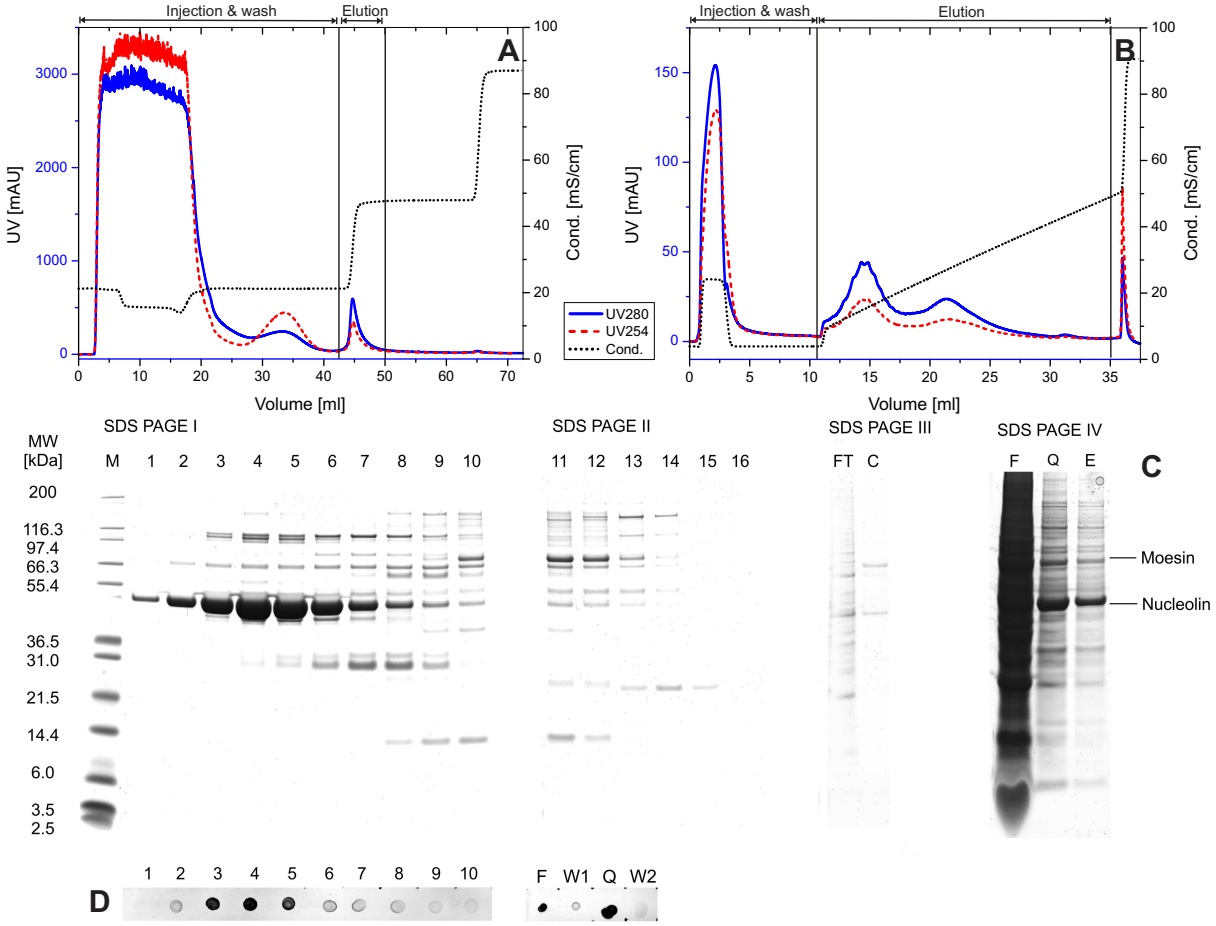

Figure 5: Two-step purification: **A**: CEC capture step at pH 8.75, 200 mM NaCl, step-elution 200-500 mM NaCl, 10 ml injection of cell lysate, 43.75 ml flow-through fraction, 7.5 ml elution fraction, column cleaning at 1000 mM NaCl; **B**: AEC purification at pH 5.2, 0 mM NaCl, gradient-elution 0-500 mM NaCl, 2 ml injection of CEC-pool, 11 ml flow-through fraction, 1 ml elution fractions, column cleaning at 1000 mM NaCl; **C**: SDS PAGE I and II: AEC fractions 1-16 ; SDS PAGE III: FT = AEC flow-through, C = 1N NaCl AEC cleaning step ; SDS PAGE IV: F = SF9 feedstock, Q = conditioned CEC capture, E = CEC capture; **D**: Dot-blot analysis of AEC fractions 1-10, F = SF9 feedstock, W1 = CEC flow-through, Q = AEC sample, W2 = AEC flow-through, all samples were brought to identical volumes before dot-blot analysis for better interpretation of the semi-quantitative result.

The chromatograms of the two-step purification process and the analysis of the chromatographic fractions by gel electrophoresis and dot-blot are shown in Fig. 5. Regarding the analysis via gel electrophoresis (Fig. 5: F: cell lysate ; Q: conc. CEC elution fraction; E: CEC elution fraction;) as well as the chromatogram it becomes obvious that the major fraction of contaminating components are already removed by the highly efficient CEC capture step. The dot-blot of the flow-through fractions clearly shows that binding conditions were chosen well as no substantial amount of the target protein was detected in either of them. The critical contaminant Moesin

Table 2: Purities

I. Fractions

number	purity [%]
1	100
2	96
3	86
4	80
5	80
6	62
7	44
8	31
9	21

II. Samples

sample	purity [%]
Feed	~ 0.1
CEX	38
AEX 1-9	66
AEX 1-7	79

is seen in the electrophoresis of the eluted fraction after CEC capture (Fig. 5: lane E).
The chromatogram of the salt gradient AEC shows two chromatographic peaks during elution. All collected AEC fractions were concentrated ten times before gel electrophoresis increasing band intensity to improve interpretation of the results. The analysis of the fractions collected during elution reveals that the first peak mainly contains the target protein and the second peak contains the critical contaminant moesin. The purity of each fraction was estimated via densitometrical analysis of the gels. The determined purities are shown in Tab. 2. A real quantitative analysis of the purification step to describe the process yield was not possible as no analytical technique with the necessary specifity to analyse the cell lysate was available. This is a frequently occuring issue when working with highly complex biotechnological mixtures. Quantitative analytical methods with the necessary specificty, e.g. well established ELISAs, are very rarely available.
Nevertheless, the results clearly show the potential of the chosen approach. By applying the suggested two-step purification a purity of the target protein of min. 66% was achieved. The suggested CEC capture step is a very advantageous capture step due to the specific chromatographic elution behaviour of the target protein. Using the capture step with the suggested conditions already resulted in a purity of 38%. After submitting the captured material to the AEC purification step, we ended up with a purity of 66% due to the successful removal of the critical contaminant, Moesin. Leaving out fractions with a lower purity could increase purity further, e.g. pooling fractions 1-7 would already result in 79% purity. However there are still some lower concentrated impurities left which were previously not defined as relevant contaminants. Those impurities might be removed by further polishing steps and / or optimization of the two chromatographic operations.

4 Conclusions

We successfully demonstrated that analytical scale characterization of a cell lysate was sufficient to generate highly useful information for protein purification process development. The used approach implemented two long and two short range pH gradient IEC operations. By using long range pH gradients and dot-blot analytics it was possible to determine its elution-pH values in IEC. Based on this information, the range of the short pH gradient was defined. Applying short pH gradient IEC together with the SDS PAGE analysis, made it possible to define the relevant contaminations, significantly concentrated proteins with comparable electrostatic properties. These were subsequently identified via tryptic digestion and LC-MS. The gained information on the relevant contaminants, the target protein and their elution behaviour was successfully used to establish a simple two-step purification for the target protein, resulting in a purity of the target of $\sim 80\%$. As the target protein is used for diagnostic purposes in a current case study, the achieved purity might be sufficient. The presented proof of concept showed impressively how well the chromatographic conditions could be chosen based on the information gained from the pH gradient screening. The strongly shifted elution-pH values of the target protein in AEC and CEC show once more that the pI value can be strongly differing from the protein's elution-pH values. This fact clearly indicates the usefulness of pH gradient IEC as a tool for protein downstream process development.

There are additional benefits to be gained from such a screening approach. As the whole characterization was done in analytical scale, it only needs a low amount of sample. This enables the usage of the screening for the determination of essential downstream process parameters already at a very early stage of process development. Another advantage of using such a screening approach is the acquired knowledge on the detailed composition of the biological feedstock. This leads to simplified validation and monitoring of the subsequently designed downstream process, in terms of feed, intermediate & product quality. Also having the information in hand which proteins are the relevant contaminants for the purification of one target protein makes it possible to influence their expression by recombinant engineering or by simply adapting fermentation / harvest conditions.

A future challenge will be to find methodologies to acquire parameters usable for chromatographic modelling and to implement other separation techniques, like other chromatographic interaction modes. As proteomic technologies have become more and more affordable and technically advanced in the last few years, it might be a future perspective that standard expression systems get completely characterized allowing *in silico* development of complete downstream processes.

A high-throughput 2D-analytical technique to obtain single protein parameters from complex cell lysates for in silico process development of ion exchange chromatography

Frieder Kröner, Dennis Elsäßer, Jürgen Hubbuch

Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Abstract

The accelerating growth of the market for biopharmaceutical proteins, the market entry of biosimilars and the growing interest in new, more complex molecules constantly pose new challenges for bioseparation process development. In the presented work we demonstrate the application of a multidimensional, analytical separation approach to obtain the relevant physicochemical parameters of single proteins in a complex mixture for *in silico* chromatographic process development. A complete cell lysate containing a low titre target protein was first fractionated by multiple linear salt gradient anion exchange chromatography (AEC) with varying gradient length. The collected fractions were subsequently analysed by high-throughput capillary gel electrophoresis (HT-CGE) after being desalted and concentrated. From the obtained data of the 2D-separation the retention-volumes and the concentration of the contained single proteins were determined. The retention-volumes of the single proteins were used to calculate the related steric-mass action model parameters. In a final evaluation experiment the received parameters were successfully applied to predict the retention behaviour of the single proteins in salt gradient AEC.

Journal of Chromatography A (UNDER REVIEW), submitted May 2013

1 Introduction

The number of protein coding human genes ranges from 20.000 to 40.000 [7–9]. As this is only the number of protein coding genes, the number of proteins is supposed to be much higher, due to alternative splicing, post-translational modifications and complex formation. Many of those are associated with specific diseases due to e.g. mutations, faulty expression levels and many more. Therefore, a large number of proteins has a potential of being used as a protein therapeutic or as a potential disease biomarker. Also the market for biopharmaceuticals was constantly growing in the last few years and is expected to grow to 167 billion US Dollars by 2015 [10]. The most strongly represented class of protein therapeutics are the monoclonal antibodies (mAbs). Expression of mAbs is mostly carried out in mammalian cell cultures, secreting the proteins into the cell culture supernatant. High titre expression processes currently reach around 10 mg/l [84]. Downstream processing of mAbs is performed using a platform process, consisting of a protein A affinity step and subsequent purification by ion exchange chromatography [17]. The production of other protein therapeutics like enzymes, protein hormones, cytokines, protein vaccines, autoantigens and many more is often more challenging than for mAbs. Reasons for that are: 1. Intracellular expression of the proteins; 2. Lower expression rates; 3. The lack of a purification platform process; 4. The lack of an affinity purification step. Nevertheless the demands for product quality and purity are still the same, resulting in a growing demand for new, more cost- and time-efficient ways of rational process development.

Because of the high requirements the most important separation techniques for proteins is still chromatography as it provides high selectivity. Furthermore, it can be performed under relatively mild conditions [17]. Classical chromatographic process development based on heuristic, knowledge-based approaches, is mostly resulting in non-optimal purification processes. A significant progress in chromatographic process development was made by the introduction of HTS-approaches for the screening of chromatographic conditions [85–88]. Also progresses in understanding the mechanistics of chromatographic separations and advances in computing technologies are currently changing classical chromatographic process development [89–93]. Process development based on the usage of mechanistic models for process simulation has significant advantages, like low developmental time, costs and increased process understanding due to the possibility to carry out a large amount of experiments *in silico*. However, to use mechanistic models for the prediction of chromatographic processes, the relevant physicochemical parameters of the proteins have to be experimentally determined. Determining those parameters can be very challenging for highly complex mixtures, e.g. complete cell lysates containing a target protein. The most well regarded mechanistic model for the simulation of ion exchange chromatography (IEC) is the steric-mass action (SMA) model developed by Brooks and Cramer [1]. Determina-

tion of the SMA parameters in mixtures is mostly done by tracking single proteins in multiple linear salt gradient IEC with different gradient lengths [4], which is still very challenging for highly complex mixtures.
A recent publication by Nfor et al. [22] showed a multidimensional fractionation approach, applying IEC, SDS-PAGE and LC-MS based protein identification, to determine the SMA parameters for a mAb and the contaminating proteins in a cell culture supernatant. After analysing the chromatographic fractions of the linear salt gradient IEC with different gradient lengths by SDS-PAGE they were able to define pseudo-components, groups of proteins showing similar retention / elution behaviour [94], for the *in silico* reconstruction of the chromatograms. By tracking the retention-volumes of these pseudo components the parameters needed for the mechanistic model were determined. Even though the method illustrates a big progress in chromatographic process development it is still very time consuming and the physicochemical parameters are only determined for pseudo-components and not the single proteins.
In this publication we show an alternative multidimensional fractionation approach, avoiding protein identification by LC-MS, with significantly increased speed and the possibility to determine SMA parameters, the concentrations and the molecular weights of contained, single proteins.
The experimental technique is based on the fractionation of multiple linear salt gradient IEC with four different gradient lengths, a rapid desalting step, sample concentrating and fraction analysis by high-throughput capillary electrophoresis. To demonstrate the potential of the experimental setup, a cell lysate containing a low titre target protein is exemplarily analysed to determine the physicochemical parameters as well as the concentrations of the contained, single proteins. The quality of the obtained parameters is finally validated by comparing predicted chromatographic behaviour of the single proteins with experimentally derived chromatograms.

1.1 Theory

1.1.1 The SMA model

The steric-mass action (SMA) model introduced by Brooks and Cramer [1] is a highly regarded model, describing the sorption processes in IEC. The model incorporates three protein specific parameters, in case of a rapid equilibrium. The equilibrium constant k_{SMA}, the protein's characteristic charge ν and its steric factor σ. In the case of a rapid equilibrium the equation of the SMA isotherm can be written as:

$$k_{SMA} = \frac{q_p}{c_p}\left(\frac{c_s}{\Lambda - (\nu_p + \sigma_i)q_p}\right)^{\nu_p} \quad (1)$$

c_s is the salt concentration in the liquid phase, c_p the protein concentration in the liquid phase, q_p the adsorbed protein concentration and Λ the columns ionic capacity.

1.1.2 Parameter determination from protein mixtures

The parameters ν and k_{SMA} are mostly determined for mixtures containing multiple proteins by running linear salt gradient IEC experiments with differing gradient lengths. The obtained retention-volumes of the proteins are then related to the parameters, using the following equation by Shukla et al. [4], which was derived of a retention model developed by Parente and Wetlaufer 1984 [95]:

$$V_R = \left(\left(c_{i,s}^{\nu+1} + \frac{V_d k_{SMA} \epsilon_c \Lambda^{\nu} (\nu+1)(c_{e,s} - c_{i,s})}{V_G}\right)^{1/(\nu+1)} - c_{i,s}\right) * \frac{V_G}{c_{e,s} - c_{i,s}} \quad (2)$$

$c_{i,s}$ and $c_{e,s}$ are the gradient salt concentrations at the beginning and the end of the salt gradient, V_G is the gradient volume, V_d the column's dead volume, V_R the retention-volume and ϵ_c the total column porosity. Usually 3-4 gradient experiments with different gradient lengths are performed to estimate k_{SMA} and ν using equation (2). The non-linear parameter σ is regularly determined by performing frontal chromatographic experiments with pure components, therefore it can not be obtained properly from mixtures. Furthermore, the steric factor is only influencing chromatographic separation if chromatography is performed out of the linear range of the columns binding capacity, which is not the case in the presented work. For these reasons, the steric factor was neglected in this work.

1.1.3 The transport-dispersive model

The employed transport-dispersive model can be used to describe mass-transfer and adsorption processes if a rapid equilibrium is assumed, e.g. for the previously described SMA model in IEC. The time- and position-dependent change of concentration for component i, $\partial c_i / \partial t$, on column level, is described by:

$$\frac{\partial c_p}{\partial t} = - u_{int}\frac{\partial c_p}{\partial x} + D_{ax}\frac{\partial^2 c_p}{\partial x^2} - \frac{1-\varepsilon_c}{\varepsilon_c}\cdot\frac{3}{r_b}k_{eff,p}[c_p - c_{b,p}] \tag{3}$$

The first term in the equation describes the convective transport, the second term the dispersive transport and the third term the mass transfer of the protein to the particle surface. The contained parameters are: $k_{eff,p}$ the lumped film diffusion coefficient, r_b the particle radius, u_{int} the interstitial velocity and D_{ax} the axial dispersion.
The time-dependent change of concentration of the component p, $\partial c_{b,p}/\partial t$, on particle level, is described by:

$$\frac{\partial c_{b,p}}{\partial t} = \frac{3}{\varepsilon_b r_b}k_{eff,p}\,[c_p - c_{b,p}] - \frac{1-\varepsilon_b}{\varepsilon_b}\frac{\partial q_p}{\partial t} \tag{4}$$

with q_p the conc. of the particle-bound protein p and ε_b the particle porosity. Thus, the expression $\partial q_p/\partial t$ is defined by equation (1) as a rapid equilibrium is assumed.
Holding in hand all the necessary parameters the model can be solved numerically to predict the chromatographic elution behaviour of proteins *in silico*.

2 Experimental procedure

To give a better overview of the performed experiments the procedure is briefly described in the following:

I. Multiple, linear salt gradient IEC: The cell lysate is injected on an IEC column and fractionated by salt gradient IEC. By performing the gradient elution with different gradient lengths the retention-volumes of the proteins can be related to their characteristic charge ν and the equilibrium constant k_{SMA}. A protein was added as an internal standard to all collected fractions, for later protein quantification.

II. Desalting: Each collected fraction is desalted by RP-trapping to remove all salt ions, which would interfere with the later following HT-CGE. Bound proteins are eluted in organic solvent.

III. Sample concentrating: The organic solvent is evaporated of all fractions and the proteins are concentrated to complete dryness. Protein concentration is increased 125x, which was

necessary for the presented case due to the low protein concentrations contained in the sample.

IV. HT-CGE: The dried samples are resolved in the sample buffer used for the HT-CGE. All fractions are subsequently analysed. Protein concentration and molecular weight are adjusted by using the internal standard.

V. Parameter estimation: The single protein concentrations are extracted from the HT-CGE results and assigned to their retention-volume. Data is then fitted by gaussian peak fitting to determine the retention-volumes of the proteins, which are finally used to determine the SMA parameters by applying equation (2).

3 Materials and methods

3.1 Chemicals

The chemicals, DTT, dextran 2000 kDa, trifluoracetic acid, lysozyme, benzonase and Sigma Fast Protease Inhibitor tablets were purchased from Sigma-Aldrich (St. Louis, MO, USA). The substances, EDTA, piperazine, isopropanol, acetonitrile, tris, sodiumchloride, magnesiumchloride, sodium hydroxide and hydrochloric acid were purchased from Merck (Darmstadt, Germany). Triton X-100 was purchased from Applichem (Darmstadt, Germany). All used substances were purchased with analytical grade purity.

3.2 Buffers

The buffer for the SF9 cell lysis consisted of 20 mM Tris-Cl at pH 8.0, 150 mM NaCl, 0.5% (v/v) Triton X-100, 2 mM EDTA and 1x SigmaFast protease inhibitor. A 20 mM piperazine buffers at pH 5.2 with concentrations of 0 and 1000 mM NaCl were used for salt gradient AEC (A: 0 mM NaCl; B: 1000 mM NaCl). The solvents for RP-trapping of proteins consisted of H_2O/Acetonitrile/Isopropanol with a ratio of 95/5/0 for the binding of the proteins and a ratio of 5/75/20 for the elution. Trifluoracetic acid was added to both solvents with a final concentration of 0.1% (v/v). All buffer solutions were prepared with ultrapure water drawn from a water purification system (Sartorius Goettingen, Germany). The pH adjustment was carefully performed with a freshly, calibrated pH meter (HI-3220, Hanna Instruments, Woonsocket, RI, USA). The used pH calibration buffers were high precision standards from Hanna Instruments. After adjusting the pH using the appropriate titrant, the buffers were brought to their final volume. Buffers used for chromatographic separations were degassed and filtered using vacuum filtration with 0.2 μm cellulose acetate filters from Sartorius (Goettingen, Germany).

3.3 Cell lysis

Cultivation and infection of the SF9 insect cells for the recombinant production of the target protein was done by the industrial partner, Diarect AG (Freiburg, Germany). For the preparation of the cell lysate, cell pellets of 50 ml cultivation broth were dissolved in lysis buffer with a ratio of 1 : 5 (w/w), by shaking the suspension for 10 min on an overhead shaker. After dissolving the pellet, the cells were lysed by sonication: 6x 15 s sonication at 40% amplitude, with 30 s cooling steps in between. Sonication was carried out on ice with a Branson Sonicator 450D equipped with a 1/8 inch microtip horn (Branson Ultrasonics, Danbury, CT, USA). After lysis, 10 mM $MgCl_2$ and 1 μl of benzonase ($\triangleq$ 250 units) were added to digest the DNA in the lysate. The cell lysate was then incubated for 30 min on an overhead shaker at room temperature. After that, particulates were removed by centrifugation for 30 min at 40.000 g. The prepared cell lysates were then aliquoted and stored frozen at -32°C.

3.4 Sample preparation for LC

The frozen cell lysates were completely thawed in a sonication bath (Branson Ultrasonics, Danbury, CT, USA) to prevent strong formation of precipitates. Afterwards, particulates were removed by centrifugation. To prepare the cell lysates for subsequent chromatographic separations, a buffer exchange via PD-10 columns (GE Healthcare, Uppsala) into the application buffer of the following chromatographic step, was performed. The lysates were filtered before injection with 0.2 μm cellulose acetate filters.

3.5 Salt gradient AEC

Multiple linear salt gradient AEC with different gradient lengths was performed to acquire the proteins' SMA parameters. The chromatographic runs were carried out on an Äkta Purifier (GE Healthcare, Uppsala, Sweden) with online monitoring of UV and conductivity. The used chromatographic column was a MonoQ 4.6/100 anion exchange column (GE Healthcare, Uppsala, Sweden) with a column volume of 1.662 ml. The chromatographic sequence consisted of the following steps: 1. Equilibration for 5 CV with 0% B; 2. Injection of 2 ml prepared sample; 3. Flow-through step for 20 CV; 4. Gradient 0-50% B in 5, 10, 20, 80 CV; 5. Wash step at 100% B for 5 CV. The flow rate was kept constant at 1.5 ml/min. To keep the number of fractions equal for the four different gradient lengths, fractionation volumes 100, 200, 400 and 1600 μl were chosen to fractionate the salt gradient. All fractions were collected in 96-deep-well MTPs. The collected fractions were subsequently prepared for RP-trapping.

3.6 RP-trapping

RP-trapping of the AEC fractions was performed to remove salt ions from the solution which would interfere with the subsequent analysis by HT-CGE. Before submitting the samples to the RP-trapping procedure an internal standard, Lysozyme, was added. Lysozyme was dissolved in water at an appropriate concentration and added to each fraction with a final concentration of 2 ng/μl for the fractions of the 5, 10, 20 CV gradient and 0.66 ng/μl for the fractions of the 80 CV gradient. The amount was chosen to reach 500 ng in the injection volume (250 μl for 5, 10, 20 CV and 750 μl for 80 CV). After adding the appropriately concentrated lysozyme solution, the final volumes of the fractions were 650 μl for the 5, 10 and 20 CV gradient and 1950 μl for the 80 CV gradient. RP-trapping was performed on an Ultimate 3000 X2 RSLC (Thermo Scientific, Waltham, MA, USA) using a RSpak RP-18G column (Showa Denko America Inc., New York, NY, USA). The trapping procedure was carried out with a flow rate of 1.0 ml/min. The procedure was conducted in the following way: 1. Equilibration for 0.95 min at 0% B; 2. Injection of 250 μl of the sample; 3. Flow-through for 1.10 min at 0% B; 4. Step-elution for 2.15 min at 100% B. To collect the eluted proteins, 300 μl of the elution step were fractionated in a V-shape 96-well MTP or V-shape 96-deep-well MTP for the trapping of the 80 CV gradient fractions. The same trapping procedure was performed for the fractions of the 80 CV gradient but repeated three times per fraction, as an injection of 750 μl, to inject the same amount of protein, is necessary. Both MTP, the AEC sample plate and the collection plate were properly sealed with silicon plate mats to avoid evaporation and contamination. To overcome the dead-volumes of the injection system an amount of 290 μl was drawn of each sample. To clean the sampler needle after injection, flow-through was performed with the collection valve switched to collect, shortly switching it to drain before fractionation to move the needle to the collection position on the MTP. The sample loop was washed after fractionation with 1000 μl of 5% Acetonitrile in water. The autosampler was set to 8°C to inhibit protein degradation. To maximally concentrate the samples in the MTP afterwards, the organic solvent was evaporated in a rotation vacuum concentrator (Martin Christ GmbH, Osterode, Germany) until complete dryness. By using V-shape MTP plates small spots of dried proteins were achieved at the bottom of each well.

3.7 High-throughput capillary gelelectrophoresis

High throughput capillary gel electrophoresis (HT-CGE) was used as the 2nd analytical dimension to trace the single proteins elution behaviour in the salt gradient AEC fractions. The HT-CGE was performed on a LabChip GX-II device, using HT Protein Express LabChip Kits (PerkinElmer, Waltham, MA, USA). As the dried protein samples had to be dissolved first, the manufacturers standard protocol was adapted. 7 μl denaturing solution with 35 mM DTT as

reducing agent and 5 μl ultrapure water were added to each sample. To dissolve the dried pellet, the MTP was properly sealed and incubated on a MTP thermomixer with lid heating (Ditabis AG, Pforzheim, Germany) at 37°C for 30 min at 300 rpm. To fully denature the dissolved proteins the thermomixer was then heated to 95°C for 5 min. The dissolved and denatured protein samples were than cooled to room temperature, shortly centrifuged to remove condensed liquid from the sealing and mixed with 32 μl of ultrapure water. In the meantime the Protein Express chip was prepared as described in the manufacturer's protocol (PerkinElmer). Protein samples were analysed by using the HT Protein Express 200 assay. The LabChip GX 3.1 software (PerkinElmer) was used for sample analysis and data interpretation. The resulting electropherograms were aligned to the internal standard Lysozyme (14.3 kDa). Quantification of the proteins in each sample was done by using the internal standard (250 ng/μl) with peak-baseline integration.

3.8 Determination of column parameters

The column parameters of the used 4.6/100 MonoQ column were experimentally determined on an Äkta purifier. To determine the interstitial porosity 25 μl of a 10 mg/ml Dextran 2000 Da solution were injected. The determination of the total porosity was carried out by injecting 25 μl of 1 M NaCl. The porosities were determined with a three time repeat measurement at a flow rate of 1.5 ml/min, while monitoring UV 225 nm and the conductivity. The particle porosity was calculated from the two experimentally determined porosities. The ionic capacity of the anion exchange column was determined via titration. The column was equilibrated with 0.01 mM sodium hydroxide and titrated with 0.01 mM hydrochloric acid. The breakthrough of hydrochloric acid, replacing bound hydroxide ions from the column was monitored by online conductivity measurement. The columns ionic capacity was determined from the amount of used hydrochloric acid and the columns total porosity. The axial dispersion was estimated by using the 1st and 2nd moment of the chromatographic peak of the non-penetrating tracer molecule, Dextran 2000.

3.9 Determination of protein concentrations, retention-volumes and molecular weights

Most abundant proteins, selected by interpretation of the HT-CGE results, were chosen to be monitored in the fractions of the 5, 10, 20 and 80 CV long gradients. The concentrations of those proteins were manually extracted from the electropherogram of each fraction. The determined single protein concentrations were then assigned to their retention-volume by the fraction number. Afterwards, peak fitting was performed to determine the retention-volume for each single protein. Peak fitting was performed using OriginLab (OriginLab Corporation,

Northampton, MA, USA) applying the standard settings for gaussian peak fitting. The retention-volumes of the fitted peaks were corrected by the dead volumes and further used for SMA parameter estimation. The concentrations of the single proteins in different fractions were added up to determine the complete concentration of each single proteins. The molecular weight was determined for each protein from the electropherograms.

3.10 Determination of the SMA parameters

The parameters ν and k_{SMA} were determined from the retention volumes by using the previously described formalism (section 1.1). Therefore, retention-volumes of the single proteins were assigned to their gradient length. The data points were subsequently fitted using the described equation (eq. 2). Fitting procedure was carried out in MATLAB (MathWorks, Natick, MA, USA) by using the "fminsearch" algorithm. The parameters ν and k_{SMA} were obtained from the fitting procedure.

4 Results and discussion

The described approach to fractionate a biotechnological feedstock and analyse the fractions further by HT-CGE was carried out to acquire the SMA parameters of the single proteins for *in silico* process prediction and optimization. The used biotechnological feedstock was a SF9/Baculovirus insect cell lysate containing the target protein (protein No. 6, Fig. 1). The protein is a 45 kDa fragment of the human protein Nucleolin (Ncl, GI number 55956788), containing all the relevant epitopes for the diagnosis of the autoimmune disease *systemic lupus erythematosus*.

The cell lysate was first fractionated by multiple linear salt gradients in AEC, with gradient lengths of 5, 10, 20 and 80 CV. All four gradients (pH 5.2, 0 – 500 mM) were fractionated in 83 chromatographic fractions. After fractionation the internal standard, lysozyme was added to increase accuracy of the later following protein concentration and size determination by HT-CGE. All 83 chromatographic fractions were subsequently desalted by RP-LC, and maximally concentrated by evaporation as described in section 3.6. Dried, desalted fractions were then dissolved in sample buffer and analysed by HT-CGE.

By analysing each chromatographic fraction with the HT-CGE, single proteins contained in the fractions could be resolved. From these results concentrations and molecular weights were determined for each protein present at a significant concentration level. Protein concentrations and the molecular weight were normalized / aligned by using the internal standard. The whole procedure is illustrated in Fig. 1 showing the results in detail for the 5 CV salt gradient IEC. The gel pictures show the electropherograms of each fraction listed in the elution / fractionation

Table 1: Proteins & values

	Size	Concentration		Retention-volumes [ml]				Calc. values	
ID	[kDa]	Ø[ng]	mean dev. [%]	5CV	10CV	20CV	80CV	ν	k_{SMA}
1	120	48	50.9	3.05	5.43	9.45	32.70	7.74	2.19e-08
2	89	554	40.1	1.87	3.29	5.65	15.05	3.15	2.12e-04
3	80	1088	44.1	2.30	3.94	7.31	18.20	2.77	1.44e-03
4	68	997	26.6	2.38	3.93	7.08	20.70	4.12	3.57e-05
5	63	2899	20.9	3.09	5.55	9.41	30.40	5.41	6.16e-06
6	51	9498	43.0	3.11	5.41	8.64	28.17	5.01	1.18e-05
7	53	308	33.9	1.77	2.90	4.16	7.09	1.12	6.50e-02
8	36	240	33.6	1.96	3.13	5.18	16.63	5.09	5.36e-07
9	26	196	21.0	1.63	2.93	5.49	18.30	7.94	1.29e-10
10	22	478	29.3	1.73	2.94	4.04	13.00	4.10	4.90e-06
11	19	119	34.3	1.53	2.56	3.99	13.11	5.35	6.41e-06
12	24	564	15.5	2.49	4.37	7.47	25.80	7.50	6.70e-09
13	11	150	24.3	2.61	4.44	6.93	21.45	4.07	4.54e-05
14	43	1944	46.3	5.53	10.32	19.66	72.90	16.60	2.47e-12
15	25	373	58.8	4.15	7.83	14.23	49.75	9.19	2.96e-08

order. From the band intensities it can be easily recognized that the detected proteins show typical gaussian elution behaviour, as expected. Single protein concentrations of the selected proteins were extracted and assigned to their retention-volume related with the fraction number. The retention-volumes of the proteins were finally obtained by standard gaussian peak fitting procedure in OriginLab. Same analysis was performed with the fractions of the 10, 20 and 80 CV long salt gradient. The determination of the proteins' retention-volumes is exemplarily illustrated for the selected target protein in Fig. 2. The whole procedure was carried out in a two times repeat measurement. The obtained, mean retention-volumes for the selected proteins are shown in Tab. 1.

Holding in hand the retention-volumes of each selected single protein for all gradient lengths, consequently enabled the determination of the linear SMA parameters (k_{SMA} and ν) of the proteins. The necessary column parameters are shown in Tab. 3. These parameters were obtained as described in section 3.8. The linear SMA parameters of the proteins were determined by relating the retention-volumes of the proteins with the gradient length and fitting the described equation (2) to it. This is exemplarily illustrated in Fig. 3 for the target protein. By repeating this procedure for all selected proteins, k_{SMA} and ν values of each protein were determined. The determination of the non-linear SMA parameter σ, accounting for the shielding effect was omitted, reasoned by the previously described facts (section 1.1.2).

Single protein concentrations were determined for the selected proteins from the protein concentrations contained in each fraction. As all four salt gradient IEC runs and the analysis were performed in a repeat measurement single protein concentrations were determined eight times.

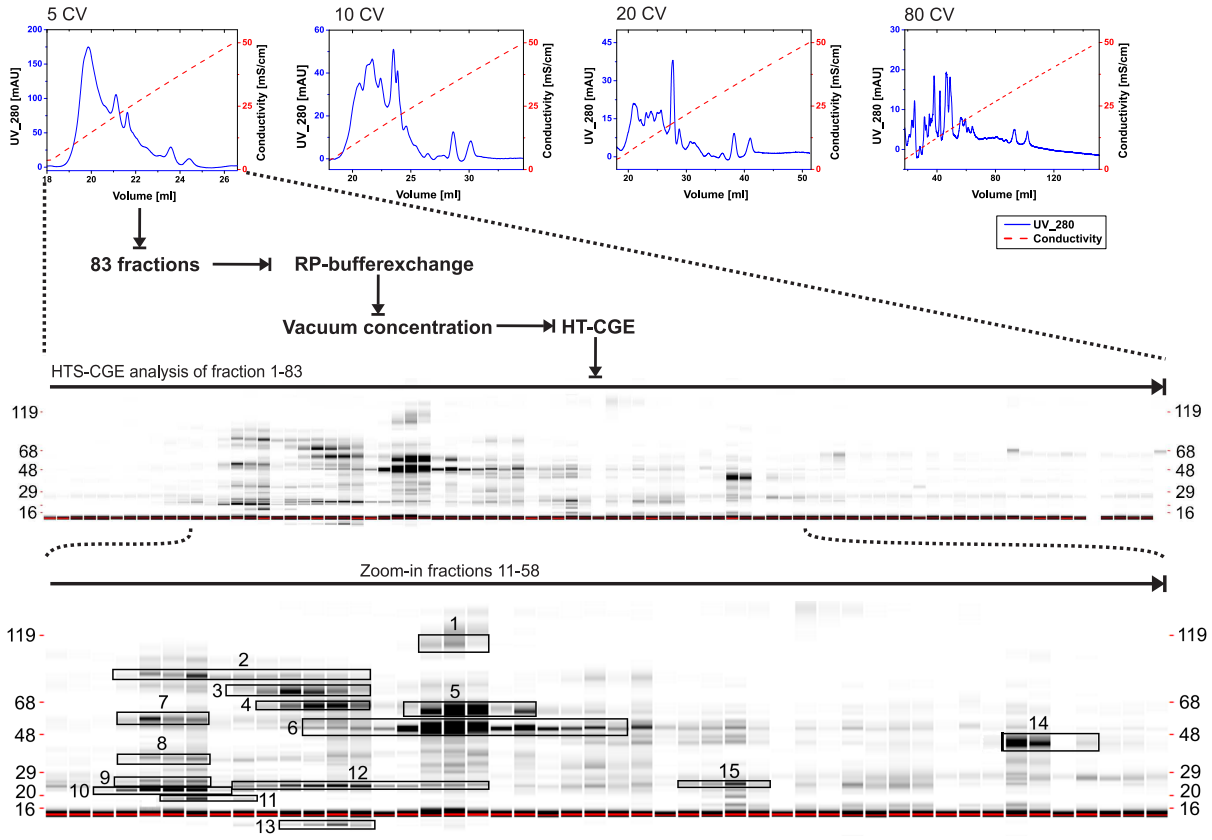

Figure 1: Multidimensional analysis

The figure illustrates the multidimensional analysis of the cell lysate. Cell lysate is first fractionated into 83 fractions by 5, 10, 20 and 80 CV long salt gradients (pH 5.2 0 – 500 mM NaCl). A desalting buffer exchange of each fraction is then carried out by using RPC. All resulting RP-fractions are evaporated to complete dryness and dissolved in sample buffer for HT-CGE. The 83 fractions are then analysed by HT-CGE. The upper gel picture shows the analysis of the chromatographic fractions of the 5 CV gradient with a rising number from left to right. The lower gel picture is focused on the protein containing fractions, selected proteins are marked by black squares and numbered. Protein no. 6 is the target protein. The lower protein band at 14.3 kDa, apparent in all samples is the internal standard at 250 ng/μl. Reconstructed gel lanes were aligned to the internal standard.

Mean concentration values as well as the deviation were calculated. The resulting data is shown in Tab. 1. Determined protein concentrations were in the range of $\sim$ 50 – 10000 ng/μl. The single protein concentrations show a mean deviation of $\sim$ 34%, while varying between 15 – 59%. Several reasons might have caused these variations. Dissolving the proteins in HT-CGE sample buffer after concentrating the samples to complete dryness might suffer from incomplete solubilization. This effect was already reduced by using an internal standard for normalization of the concentration. But also with adding the internal standard, varying solubilization of different

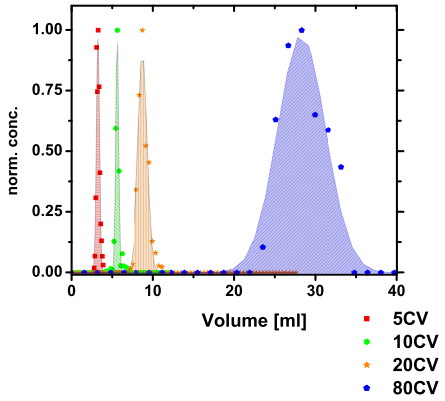

Figure 2: Peak fitting

The determination of the retention-volumes of protein no. 6 by fitting the single concentration values in the different gradient elutions 5, 10, 20 and 80 CV. Peak retention-volumes were determined from the resulting gaussian peaks.

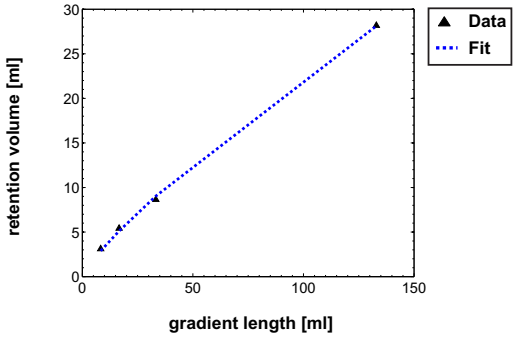

Figure 3: Parameter estimation

The determination of the SMA parameters of protein no. 6 was carried out by fitting the assigned retention-volumes (triangulars) by a "fminsearch" based MATLAB procedure using eq. 2.

proteins can not be ruled out. Also, variation of protein concentrations might be reasoned by the changed sample preparation, probably increasing variation of the analytical procedure. Another fact that surely influences the determined protein concentrations is the overlap of proteins with comparable molecular size. These overlaps are likely to be influenced by the gradient length of the salt gradient IEC, as the increased chromatographic resolution of the longer IEC salt gradients probably separates overlapping proteins. Furthermore single protein concentrations vary due to regular variations caused by the device, the protein chips and the analytical procedure. Nevertheless, single protein concentrations were determined eight times and are, due to that issue most probably close to the real value of the protein concentration.

To evaluate the applicability of the determined SMA parameters the values were used to predict the retention-volumes of the single proteins in salt gradient IEC with a 15 CV long salt gradient, by applying the obtained SMA parameters. The resulting retention-volumes were subsequently compared to experimentally determined retention-volumes. The obtained results are shown in Tab. 2.

Comparing the predicted retention-volumes with the experimentally obtained retention-volumes for the 15 CV long salt gradient results in a mean deviation of 5%. The deviation varies between 0.1 and 13.6% for the different proteins. This clearly shows that the SMA parameters obtained by using this multidimensional separation / analysis procedure can be used for predicting the retention-volumes of the single proteins. The applicability of the parameters for *in silico* optimization of the chromatographic separation was further evaluated by exemplarily predicting the

chromatograms of three proteins. Prediction of the chromatographic elution is done by applying the transport-dispersive model as described previously. The model was solved numerically, by applying an in-house software tool written in `C++`. A density of 400 knots was applied over the whole column length. The necessary film diffusion coefficient k_{film} was estimated by prediction of the proteins diffusion constant [96], which was then related with k_{film} as described elsewhere [97]. The steric factor σ was chosen to be 30 for all proteins because it was not possible to determine σ experimentally from such a crude mixture. However, it did not effect the prediction of the chromatographic behaviour due to the fact that concentrations of the contained single proteins were in the linear range of the column's capacity. The resulting chromatographic peaks are shown in Fig. 4. Predicted chromatographic peaks match the experimentally determined peaks very well. Only peak width shows variation from the experimentally determined results.
Variations between the experimental and the predicted elution behaviour of the single proteins are most likely to be reasoned by slight deviations of the parameters from the real values. The determined single protein concentrations in each fraction underly the previously described variations. As the peak fitting procedure is depending on the protein concentrations determined in each fraction, variations in the obtained retention-volumes of the single proteins are expected. The occurrence of protein charge variants, which might get partly separated in longer salt gradients as chromatographic resolution is increased, would also affect determination of the retention-volumes. Slightly deviating retention-volumes are consequently influencing the estimation of SMA parameters and therefore the predicted retention-volumes. Also, the precision of the retention-volume is limited by the amount of fractions collected, a higher amount of fractions would lead to an increased precision but decreased protein concentration in the fractions as well as increased investment in time and experimental effort.
Still a mean deviation of the predicted retention-volumes of 5% is a very good result, definitely allowing the *in silico* prediction of the single proteins' retention-volumes. Also the *in silico* prediction of single protein chromatograms using the obtained parameters matched the experimentally determined chromatograms very well. The whole procedure, four multiple salt gradients, the buffer exchange by RPC, solvent evaporation and the HT-CGE analysis were performed at a time-scale of two days. Still the data analysis was very time consuming as no automated data processing exists so far.

5 Conclusions

The developed multidimensional fractionation technique was successfully applied to characterize a complex biotechnological feedstock for the acquisition of relevant parameters for *in silico* prediction of the chromatographic behaviour of the contained proteins. Multidimensional frac-

Table 2: Comparison of predicted with experimentally determined V_R for 15 CV gradient length

ID	Retention-volumes [ml] Calc.	Exp.	Δ [%]
1	7.51	7.20	4.3
2	4.47	4.93	9.4
3	5.61	5.87	4.4
4	5.65	5.08	11.3
5	7.56	7.55	0.2
6	7.13	7.29	2.1
7	3.41	3.41	0.0
8	4.25	4.24	0.5
9	4.00	4.24	0.3
10	3.57	3.78	13.2
11	3.33	3.86	13.6
12	6.00	6.35	5.5
13	5.78	5.89	1.8
14	15.08	15.24	1.0
15	11.06	11.15	0.8

Table 3: Column parameters, MonoQ 4.6/100

Parameter	Value
V_{Dead}	0.366 ml
ε_{total}	0.77
ε_{Bead}	0.648
Λ	1522 mM
D_{ax}	2.5 mm^2/s
u_{int}	1.5 mm/s
r_p	5 μm
$C_{salt,i}$	10 mM
$C_{salt,e}$	510 mM

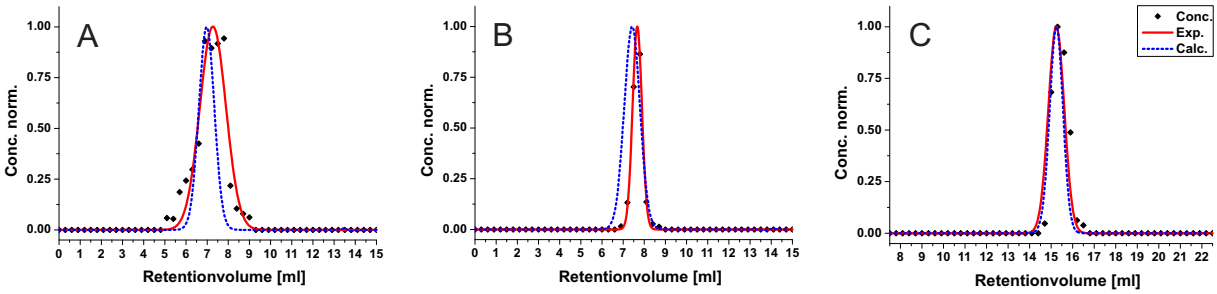

Figure 4: Comparison of predicted with experimental results

The experimentally determined chromatograms as well as the predicted chromatograms for a gradient length of 15 CV are illustrated for the three most concentrated proteins, no. 5, 6 and 14. The data for protein no. 6 is shown in A, no. 5 in B and no. 14 in C. Experimentally determined protein concentrations in the collected fraction are illustrated by black diamond symbols. The fitted gaussian curve is shown by a red solid line and the predicted chromatographic peak is shown by the dashed, blue line. Peak maxima were normalized for better illustration.

tionation was based on salt gradient IEC, fast RPC-based sample desalting and concentrating and a the second analytical dimension (HT-CGE). By running multiple linear salt gradients with the described experimental set-up the concentration, the molecular weight, the characteristic charge ν as well as the equilibrium constant k_{SMA}, were determined for each protein. Validation was successfully carried out by comparing predicted, with experimentally determined retention-volumes. The comparison of predicted chromatograms based on the obtained parameters with experimentally obtained chromatograms successfully evaluated their usability for *in*

silico process prediction or further process optimization. The major qualities of this approach are, its velocity, the analytical amount of sample necessary and the possibility to track the single protein's retention behaviour in complex lysates allowing the determination of parameters usable for chromatographic process modelling. These attributes enable the application of the method e.g. as a routinely used tool to analyse biotechnological feedstocks at a fully analytical scale to obtain the described parameters, but also to monitor the composition of the feedstock and probably predict possible influences of changed compositions on the chromatographic process. Nonetheless, the approach also has its weaknesses, determined protein concentrations are still significantly varying even though the eight times repetition of the single protein concentrations seems to compensate for that. As this is probably reasoned by the changed sample preparation for the HT-CGE a future revision of the experimental procedure addressing that fact might bring improvement. Another flaw is the time consuming analysis of the resulting data. Therefore, the generation of automatized data analysis tools is essential for the routine application of the technique. Concluding all that it can be clearly stated that the application of rapid multi-dimensional fractionation techniques, like the presented one, might be a very useful future tool for chromatographic but possibly also non-chromatographic bioseparation process development and monitoring.

Parameter estimation for a wide pH range to model chromatographic separations on ion exchange materials for both salt and pH gradient elution

Florian Dismer[a], Frieder Kröner[a], Jürgen Hubbuch

Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[a] *These authors contributed equally to this manuscript*

Abstract

For the development and optimization of chromatographic purification steps for proteins, predictive modelling has long proven to be a valuable tool, although modelling is most often applied in the academic world for several reasons. One of them is that protein parameter estimation is a tedious work. Parameters are pH specific and their determination often requires relatively pure protein. Nevertheless, the pH is a crucial parameter of great interest in most optimizations it would thus be helpful if it could be included in modelling without the need of running a lot of calibration runs at different pH. In this article an approach is presented to determine model parameters for the steric mass action model for a wide pH range (pH 4 to 11.5) with a limited set of calibration experiments (linear pH gradient elution at different salt concentrations). Parameter estimation was achieved by extending the applicability of the existing SMA model to protein elution in pH gradients and by adding certain constraints on the estimated model parameters. These constraints ensure a systematic progression of the characteristic charge and the equilibrium constant K_{SMA} over a certain pH range. Model parameters were accurate enough to predict retention times for isocratic, mono- and bi-linear salt gradient elution and for combined pH and salt gradient elution. A Monte Carlo study was made to investigate the effects of a 5% experimental error on all critical parameters. It confirmed that the higher resolutions that can

Journal of Chromatography A (UNDER REVIEW), submitted March 2013

be achieved in isocratic elution mode come at the cost of a lower robustness compared to both salt and pH gradient elution, meaning that retention times can shift significantly already for small experimental errors.

1 Introduction

The use of chromatographic techniques for the purification of biopharmaceuticals on both lab and industrial scale is well established and a lot of efforts were made to speed up process development. Especially over the last years both industry and academia were trying to establish a platform strategy, or in other words a "plug and play" approach for finding suitable process parameters, independent of the target molecule to be purified [98, 99]. For structurally similar molecules, that can be purified by affinity chromatography (i.e. monoclonal antibodies) platform strategies can be a defined as a series of two to three consecutive chromatographic steps, often being Protein A affinity followed by an anion exchange and a cation exchange step.
Nevertheless, other recombinant proteins are to be expected to enter the market in the future and finding a platform process for these structurally and functionally diverse molecules will most certainly fail. Alternatively, one could think of a platform approach rather than a platform process, meaning that all new molecules would have to pass a series of characterization steps that will then define the layout of the process. One such approach for the identification of suitable chromatographic separation steps and their parameters could be predictive modelling. Having a new molecule to purify, one could characterize the molecule by determining all parameters necessary for the chosen model to perform an in silico process optimization for choosing suitable process conditions. In order to be sufficiently efficient to replace traditional (meaning experimentally driven) process development, parameter estimation needs to be limited to a minimum number of experiments giving a maximum amount of information on the behaviour of the target molecule in the unit operation of interest. One additional benefit of establishing a model-based platform approach is its value in QbD, as it implies a detailed understanding of the single unit operations and additionally allows for robustness analysis with only little additional work.
In the field of chromatography a lot of effort has been made over the last 30 years to develop both empirical and mechanistic models to fill this gap of a missing platform process for future molecules to come. One step stone was the development of the "Linear-Solvent Strength" (LSS) model [100, 101] in 1979 that was initially designed to predict retention behaviour in reversed-phase chromatography. The basic concept was then extended to ion-exchange chromatography in 1989 by Sasagawa et al. [102], finally leading to the "Steric-Mass Action" (SMA) model by Brooks & Cramer published in 1992 [1], which for the first time translated the mechanistical understanding that only part of the protein surface interacts with the adsorber material into

mathematical equations allowing for gradient shape optimizations.
The SMA model is capable of predicting retention times on ion-exchange resins after determination of a set of two protein specific parameters: the affinity constant (K_{SMA}) and the characteristic charge of the protein (ν). When working outside the linear range of the adsorption isotherm, a third parameter (the steric factor σ) is needed for accurate predictions. As these parameters are independent of salt concentration, the model can be used to optimize isocratic and gradient elution, which has been shown in numerous publications [5, 103–106]. Although these parameters are independent of the salt concentration, they strongly depend on the pH in the mobile phase, limiting the use of the SMA model to serve as a platform approach for developing new processes. For that reason, a few academic groups have tried to extend the SMA model to different mobile phase pH.
In 1998, Bosma and Wesselingh [107] were the first that extended the SMA model to account for changes in mobile phase pH by adding a rather complex thermodynamic framework. One of the main assumptions of their approach was though, that the characteristic charge of the protein, which is, based on the SMA formalism, the charge of the binding site of the protein, is independent of pH and only the affinity of binding (K_{SMA}) changes. From more recent studies (including the one discussed in the next paragraph) we know today, that this assumption does not hold.
In 2007 Yang et al. successfully predicted SMA parameters for different pH (pH 4 to 8) by applying a quantitative structure-property relationship (QSPR) approach [108]. They used a set of 17 different proteins and determined a set of pH specific descriptors based on the 3D structure of these proteins. Although being capable of yielding accurate characteristic charge predictions, the QSPR approach showed relatively high deviations for predicted K_{SMA} values. Additionally it was experimentally challenging (45 gradient elution runs to generate data for the training dataset), required crystal structures for all proteins and expert knowledge on how to determine structure-related parameters.
One of the most recent publications on incorporating pH changes into modelling was published in 2010 by Guélat et al. [109]. They established a new thermodynamic model based on the DLVO theory, reducing the protein-adsorber interaction to a sphere-plane interaction. In contrast to the SMA model, their approach is based on the assumption that a protein has a homogeneous charge distribution. By calculating pH dependent protonation states of titratable amino acids, the surface charge density of a given protein could be calculated. Similar to the previous approach, this one requires detailed information about the protein and the adsorber surface.
In the following we will present a purely experimental approach that is capable of yielding SMA parameters for a wide pH range with a limited number of calibration experiments. This approach uses a linear pH gradient (pH 4.0 to 11) ran at eight different salt concentrations (10

to 350 mM) combined with isocratic elution experiments at three different pH values (pH 5, 7 and 9). We also show that a pH gradient at constant or changing salt concentration can be well described mathematically by a series of isocratic elution steps with systematically changing SMA parameters.
With this approach, SMA parameters could be determined for the whole pH range allowing for the optimization of isocratic or gradient elutions at different pH values, which is demonstrated for three model proteins (Ribonuclease A, Cytochrome C and Lysozyme). The parameters can then be used to determine retention times at different pH for isocratic and gradient elution.
Additionally, the model was used to study process robustness and maximum resolution of different modes of operating a cation exchange column: isocratic elution, salt gradient elution and pH gradient elution combined with an additional salt gradient.

2 Material and methods

2.1 Chemicals & buffers

For pH gradient CEC the chosen substances were MES, formic and acetic acid (Merck, Darmstadt, Germany), HEPPSO (Molekula, Dorset, UK), and MOPSO, TAPS, CHES, CAPS (Applichem, Darmstadt, Germany). The detailed composition of the buffer system for pH gradient CEC is shown in Tab. 1. The derivation of the buffer systems composition is described in a

Table 1: Buffer system for pH gradient CEC

Substance	Conc.[mM]
CAPS	15.6
CHES	9.4
TAPS	4.6
HEPPSO	9.9
MOPSO	8.7
MES	11.0
Acetate	13.0
Formate	9.9

previous work [72]. Sodium hydroxide, hydrochloric acid and sodium chloride were purchased from Merck.

2.2 Determination of column parameters

The column parameters of the used 4.6/100 MonoS column (GE Healthcare, Uppsala, Sweden) were experimentally determined on an Aekta purifier (GE Healthcare, Uppsala, Sweden).
To determine the interstitial porosity 25 μl of a 10 mg/ml Dextran 2000 Da solution was injected. The determination of the total porosity was carried out by injecting 25 μl of 1 M NaCl. Both

porosities were determined with a three time repeat measurement under a flow rate of 1.5 ml/min, while monitoring UV 225 nm and the conductivity. The particle porosity was calculated from the two experimentally determined porosities.
The ionic capacity of the cation exchange column was determined via titration. The column was equilibrated with 0.01 mM hydrochloric acid and titrated with 0.01 mM sodium hydroxide. The breakthrough of sodium hydroxide was monitored by online conductivity measurement. The columns ionic capacity was determined from the amount of used hydrochloric acid and the columns total porosity.

2.3 pH gradient elution experiments

The pH gradient elution experiments were performed on an Aekta purifier, while UV 280 nm, the conductivity and the pH were online monitored. The used column was the previously mentioned MonoS 4.6/100. The chromatographic run consisted of a four column volumes (1.662 ml) equilibration step, a flow through step for one column volume, a linear gradient from 0-100% over 15 column volumes and a post-gradient step at 100% for five column volumes. The flow rate was 1.5 ml/min. The pH gradient reached from pH 4.0 to 11.0. pH gradients were run at a salt concentration of 10, 50, 100, 150, 200, 250, 300 and 350 mM NaCl. The injection volume was 100 μl. The protein solution consisted of 5 mg/ml of each protein in the application buffer. The retention times of the single proteins were determined afterwards from the chromatograms.

2.4 Isocratic elution experiments

The isocratic elution experiments were also performed on the Aekta purifier with the same column as in 2.3, while monitoring the UV 280 nm signal. The chromatographic run was performed at a flow rate of 1.5 ml/min with an injection of 100 μl of 5 mg/ml single protein solution.

2.5 Calculating retention volumes in isocratic, salt and pH gradients

To determine the SMA parameters an equation published by Pedersen et al. [5] was used for the isocratic elution of a protein:

$$V_r = V_0 + V_{column} * (1 - \epsilon_i) * \epsilon_p * K_{SEC} * K_{SMA} * (\Lambda/(z_{salt} * c_{salt}))^{z_{protein}} \quad (1)$$

V_r is the retention volume, V_0 the volume of an unattained solute, ϵ_i is the interstitial porosity, ϵ_p in the particle porosity, K_{SEC} is a protein specific coefficient accounting for the size exclusion effect of the column, Λ is the total ionic capacity of the resin, z_{salt} is the charge of the salt ions, c_{salt} the salt concentration and K_{SMA} and $z_{protein}$ are the SMA parameters. The third SMA parameter (steric hindrance factor σ) is only needed when not working in the linear region

of the isotherm and was not included in this study as its determination requires additional experiments. We have modified eq. 1 by substituting the volume of an unattained solute by the following term:

$$V_0 = K_{SEC} * V_{column} \tag{2}$$

By doing that, the retention volume of a protein under non-binding conditions does no longer equal the retention volume of a small tracer such as acetone. Eq. 2 accounts for the size exclusion effect of the column as the molecular size of a protein is considerably larger than that of acetone. With this change calculated retention volumes for weakly retained proteins becomes more accurate.
Eq. 1 was used for all four elution modes: isocratic, salt-gradient, pH-gradient at constant salt and pH gradient combined with a salt gradient based on the following concept:

1. Isocratic retention volumes can directly be calculated from eq. 1.
2. Salt gradient elution is a series of consecutive isocratic elutions steps of infinitesimal small size with increasing salt concentrations but constant SMA parameters.
3. pH gradient elution at constant salt concentration is a series of consecutive isocratic elution steps of infinitesimal small size with constant salt concentrations but with systematically changing SMA parameters.
4. Simultaneous pH and salt gradient elution is a series of isocratic elution steps with changing salt concentrations and systematically changing SMA parameters at the same time.

A step width of 30 to 60 μl was used throughout this work. In each step, the protein peak moves towards the outlet of the column, the relative movement can be calculated as:

$$V_i = (V_s/V_{iso,step}) * K_{SEC} * V_{column} \tag{3}$$

V_i is the retention volume for step i, V_s is the step volume, $V_{iso,step}$ is the retention volume given by eq. 1 for this step at this particular salt concentration. A simple example helps to understand: For a step volume of $V_s = 0.1$ ml, a calculated retention volume of $V_{iso,step} = 10$ ml, a K_{SEC} of 0.9 and a column volume of 1 ml, the protein would pass through 0.009 ml of the available volume per step. When the sum of all V_i values equals $K_{SEC} * V_{column}$, the protein has passed through the whole column. If $V_{iso,step}$ is constant for all steps (as for an isocratic elution), the protein would elute after 100 steps = 10 ml. In this simple example the retention volume could also be directly calculated from eq. 1.

2.6 Constraints for SMA parameter estimation from pH gradients

As discussed in the previous section, retention volumes in pH gradients were calculated based on the assumption that a pH gradient is a series of isocratic steps with changing SMA parameters (K_{SMA} and $z_{protein}$). The SMA parameters would not change randomly with increasing pH but would rather follow some rational:

1. The charge of the protein or binding site ($z_{protein}$) remains constant or decreases with increasing pH, similar to a titration curve of a protein.
2. The affinity of the protein (K_{SMA}) remains constant or decreases.

To add these constraints two different mathematical functions were chosen that capture these trends. For K_{SMA} a Boltzmann function was used describing an inverse sigmoidal trend:

$$K_{SMA} = (A_1 + A_2)/(1 + e^{((x_1 - x_2)/k)} + A_2) \tag{4}$$

For the characteristic charge a double Boltzmann function was used describing an inverse bi-sigmoidal trend (some examples can be seen in Fig. 1):

$$Z_{protein} = y_0 + A * (p/(1 + e^{((x-x_1)/k_1)}) + ((1-p)/(1 + e^{((x-x_2)/k_2)}) \tag{5}$$

Using these two functions serves two purposes: 1) both parameters automatically fulfill the constraints mentioned above and 2) Parameter estimation is easier since only 11 parameters need to be determined rather than a few hundred (since K_{SMA} and $z_{protein}$ is needed for every pH of every small isocratic elution step).
Since eq. 4 does not only capture inverse bi-sigmoidal trends but other trends that would violate the prerequisites defined above (points 1 and 2), boundary conditions were determined for all 7 parameters. To account for the inherent characteristics of proteins this was done by generating 100,000 artificial titration curves of proteins with a random number of titratable amino acids (lysine, arginine, histidine, glutamic acid and aspartic acid) with their respective pKa. Additionally pKa values were randomly altered to increase variability and to reflect 3D structure effects on the pKa of an individual amino acid.
Titration curves showing a pI between 4 and 12 were selected others discarded. These titration curves where then fitted by eq. 5. The boundary conditions summarized in Tab. 2 resulted from this study.

2.7 SMA parameter estimation

For SMA parameter estimation based on experimental data, the following experiments were made:

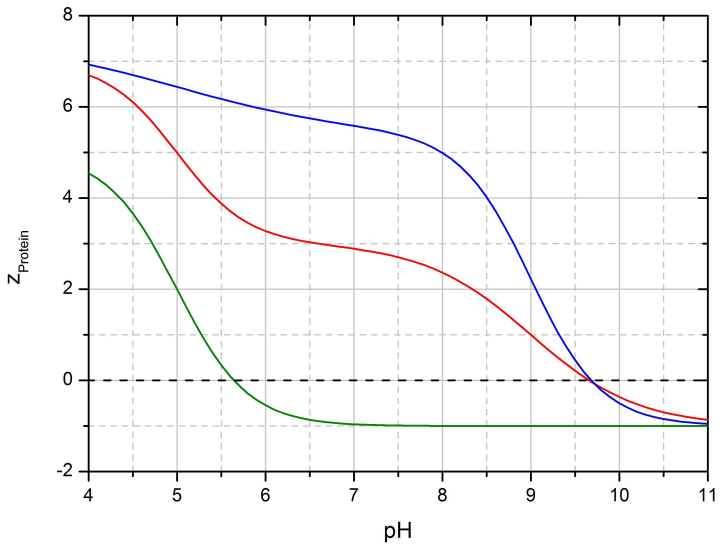

Figure 1: Double Boltzmann functions according to eq. 4. Parameters are within the boundaries defined in Tab. 2.

1. pH gradients (4.0 to 11) at eight different salt concentrations (10-350 mM)
2. four isocratic elution conditions per pH at three different pH (pH 5, 7 and 9)

Then a random set of parameters for eq. 4 and eq. 5 within the boundaries was generated as starting set, SMA parameters were calculated at different pH values using these parameters, retention volumes were estimated and compared to experimental data. Parameters for eq. 4 & 5 were then iteratively refined to minimize the error.
Column and protein parameters used for all simulations are summarized in Tab. 3. The set of parameters that best captured the experimental data was then chosen for that protein.

Table 2: Boundaries for fit parameters for eq. 5

Parameter	**Boundaries**
A	-5 to -1
k_1	6 to 12
k_2	0.25 to 1
p	3.5 to 6
x_1	0.1 to 1
x_2	6 to 12
y_0	0.2 to 0.95

Table 3: Column and protein parameters used for modelling

Parameter	Value
V_{column}	1.662 ml
$\epsilon_{interstitial}$	0.332
$\epsilon_{particle}$	0.639
Λ	0.584 mol/l
$K_{SEC,Lysozyme}$	0.73
$K_{SEC,CytochromeC}$	0.71
$K_{SEC,RibonucleaseA}$	0.71

2.8 Isocratic and gradient elution runs and optimization of process conditions

The model was used to identify experimental conditions that show a separation of all three proteins as these are suitable for model validation. This was done for all four different elution modes. Experiments were run at these conditions.
Finding suitable process conditions was done by running 250.000 in silico experiments with random values for the parameters to optimize:

1. **Isocratic elution**: salt concentrations of two consecutive elution steps, both step volumes and the pH.
2. **Salt gradient elution**: start and end salt concentrations of two consecutive linear salt gradients, both gradient volumes and the pH.
3. **pH + salt gradient elution**: start and end salt concentrations of salt gradient, start and end pH and gradient volume.

Peak resolution was calculated for ranking as this is the critical parameter for judging the quality of separation. For these three model proteins the resolution was defined as retention volume difference between cytochrome C and the next closest component as cytochrome C always eluted in the middle. The optimal system was chosen according to certain criteria (a combination of resolution and efficiency) that are discussed later in more detail. These parameters were then refined by random variations within close boundaries of +/- 10 % of each parameter value until no further improvement could be detected.

2.9 Robustness analysis

Robustness analysis was done by running 20.000 simulations with a random, normal distributed error on some parameters that are discussed later in detail.

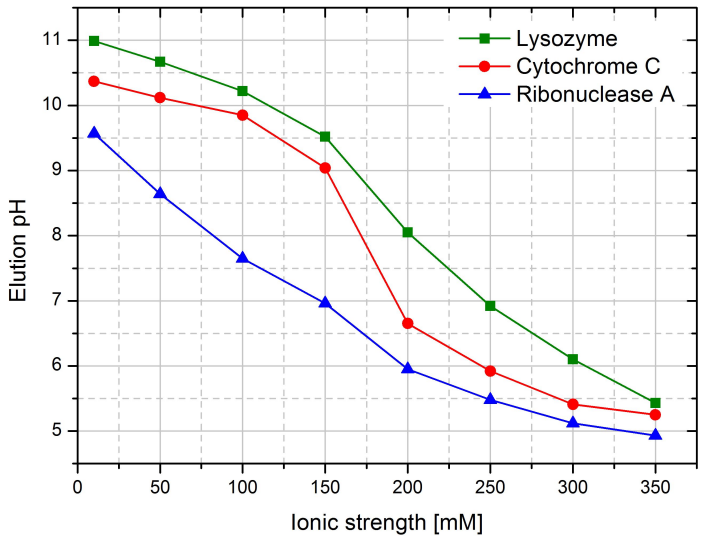

Figure 2: Salt concentration dependent elution pH of lysozyme (square), cytochrome C (circle) and ribonuclease A (triangular).

3 Results and discussion

3.1 pH gradient elution experiments

Fig. 2 shows the elution behaviour of lysozyme, cytochrome C and ribonuclease A in linear pH gradients (pH 4.0 to 11, salt concentrations between 10 and 350 mM). All three proteins used in this study have a pI in the basic region: 11.3 for lysozyme, 9.9 for ribonuclease A and 9.7 for cytochrome C (calculated using the online tool MEAD [6], which was also reflected in their elution behaviour, especially at low salt concentrations, where the elution pH should be close the pI of the proteins as binding strength is not weakened by the presence of salt.
With increasing salt concentration, the elution pH was shifted towards lower pH as salt effects dominated over protein charge effects. As to be expected, the three proteins showed a different sensitivity towards changes in salt concentration. While the retention pH of ribonuclease A and lysozyme changed more or less consistently, cytochrome c showed a more pronounced drop between 150 and 250 mM ionic strength. Without any modelling it is obvious that ribonuclease could be best purified at low ionic strength, cytochrome C at about 180 mM ionic strength and lysozyme at ionic strengths above 200 mM.

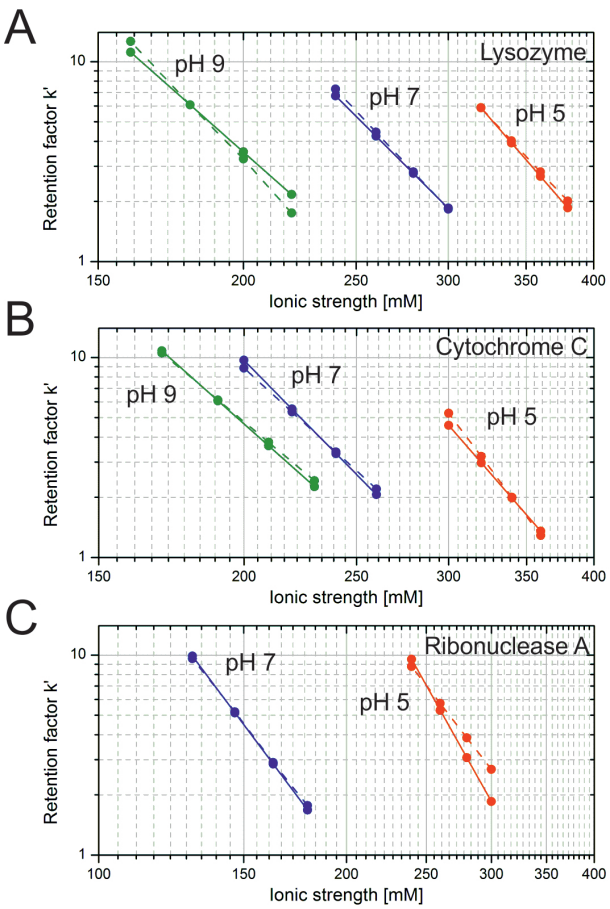

Figure 3: Isocratic elution data for lysozyme (**A**), cytochrome C (**B**) and ribonuclease A (**C**) measured experimentally (dashed lines) and calculated during fitting (solid lines).

3.2 Results of SMA parameter estimation

SMA parameters were estimated, simultaneously capturing retention volumes of pH gradient elution data at different salt concentrations and isocratic elution data at different but constant pH. Fitting quality for the isocratic data is shown in Fig. 3. The average relative error in retention volumes was 6 %, 5 % and 10 % for lysozyme, cytochrome C and ribonuclease A. The average relative error for pH gradient elution was 5.5 %, 14.5 % and 2.2 % respectively.

The resulting parameters are plotted in Fig. 4 for a pH between 4 and 12. Obviously, all predicted isoelectric points for the binding sites were above the ones for the whole proteins (∼11.6 for lysozyme and cytochrome C, ∼11 for ribonuclease A). It has been reported earlier that some proteins, including lysozyme, bind to cation exchange matrices above their pI, as a fraction of their surface can still be positively charged even above the pI (lysine are arginine both have pKa values between 10 and 12 depending on the structural environment). It should be noted

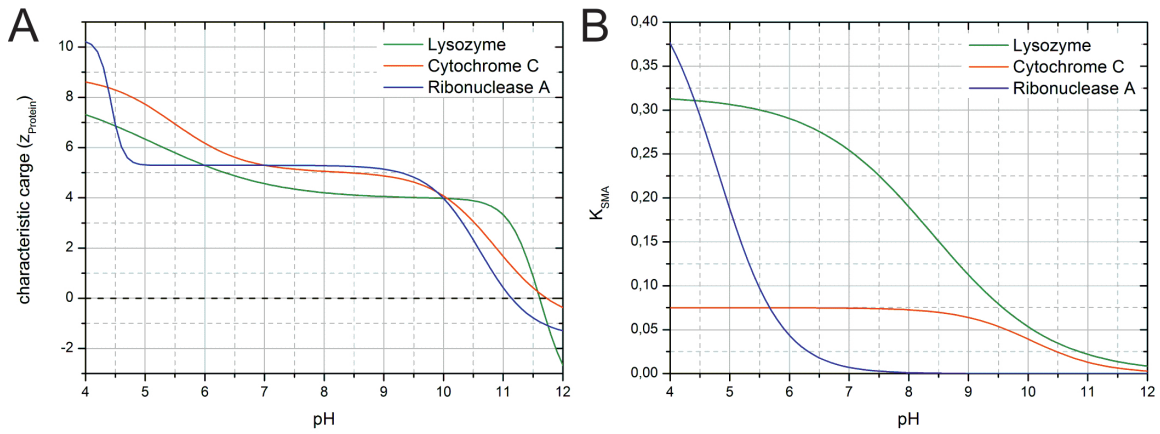

Figure 4: Characteristic charge (**A**) and K_{SMA} (**B**)for lysozyme, cytochrome C and ribonuclease A. Parameters were estimated using pH gradient and isocratic retention experiments.

that the K_{SMA} for ribonuclease A drops to values close to zero at around pH 8, resulting in retention volumes close to zero. Thus the characteristic charge curve beyond this pH might not be very accurate. In fact, ribonuclease A did not show retention at pH 9 in the isocratic elution experiments.

While cytochrome C has the highest characteristic charge of up to pH 10, lysozyme apparently binds stronger leading to higher K_{SMA} values, which explains why lysozyme always eluted last under all conditions.

3.3 Model validation

In order to validate the model for different elution modes, separation runs of all three components were performed and compared to modelling results: 1. Two isocratic elutions, 2. A linear salt gradient elution with a monotonously increasing salt concentration of 250 mM over the whole gradient, 3. bi-linear salt gradient elution at constant pH, 4. A pH gradient elution combined with a salt gradient. The model was used to identify conditions that were expected to give a reasonably good separation (resolutions around > 2) simply by calculating retention times at different salt concentrations and pH values of all three components before the experiments were performed. Resolutions were calculated for Cytochrome C only as it always eluted in the middle. Thus in a system with a resolution of 2 for cytochrome C, ribonuclease A would elute first followed by cytochrome C with at least a difference of 2 ml and then finally lysozyme again with a difference of at least 2 ml.

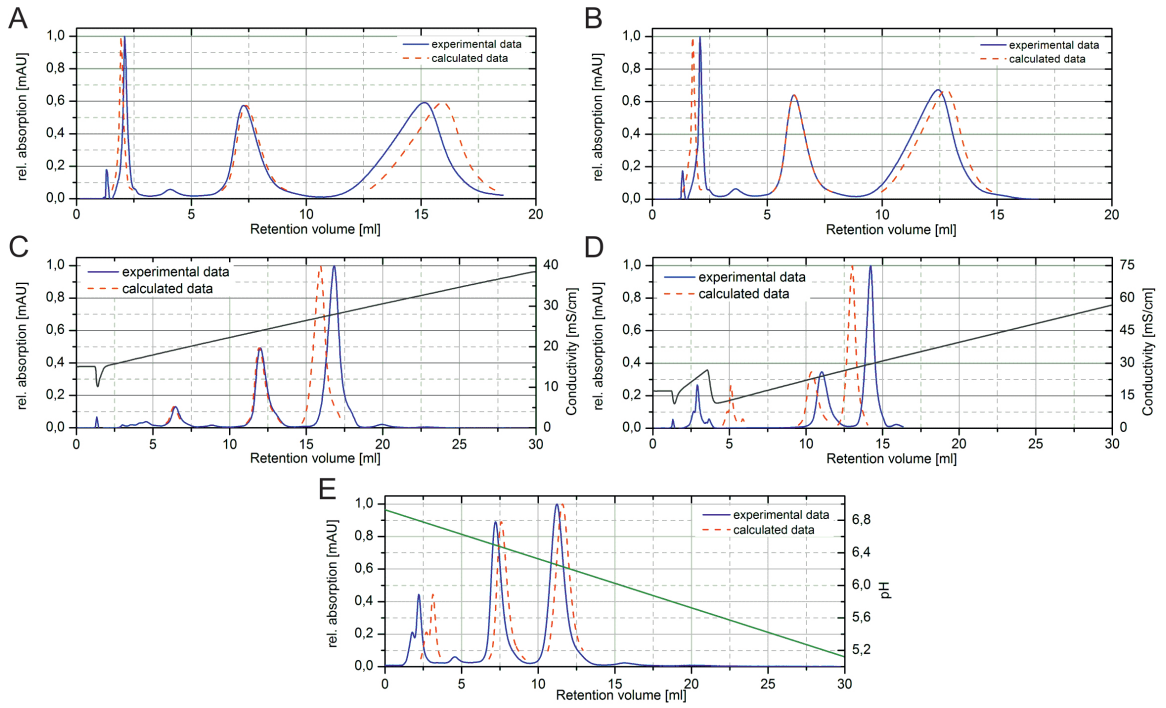

Figure 5: Elution runs at different conditions: isocratic at 225 mM (**A**) and at 235 mM (**B**) ionic strength, monotonously increasing salt gradient (**C**), bi-linear salt gradient (**D**) and a combined pH and salt gradient. Solid lines in blue refer to experimental data, dashed lines in gray are peaks shifted to retention volumes predicted by the model.

3.4 Isocratic elution

Using the SMA parameters determined, suitable elution conditions giving a selectivity of 5.5 in a total run volume of approximately 16 ml were found at a salt concentration of 225 mM salt at pH 6.92. Calculated retention volumes were 1.92 ml, 7.38 ml and 15.9 ml for ribonuclease A, cytochrome C and lysozyme. Figure 5 A shows the experimental results for these conditions, experimental retention volumes revealed a prediction error of 0.18 ml (9.4 %), 0.1 ml (1.4 %) and 0.79 ml (5.0 %). Resolution was 5.2 instead of 5.5. The peaks showing the predicted elution profiles are the measured peaks, shifted to the predicted elution volume, peak shapes were not calculated as mass-transfer and binding kinetics were not included in the model.

To experimentally validate one of the drawbacks of isocratic elution, namely a relatively low robustness due to a high sensitivity for small changes in salt concentrations, a second run was performed at the same pH but with a slightly higher salt concentration of 235 mM (instead of 225 mM). The model predicted a drop in resolution down to 4.38, mainly due to a shift in retention volume of cytochrome C from 7.38 to 6.18. Experiments again confirmed the retention volumes predicted with an error of < 0.4 ml, resolution dropped to 4.08. All details are summarized in

Table 4: Summary of model predictions and experimental validation for isocratic and gradient elution. All volumes in ml. MC indicates the results of Monte Carlo simulations.

	Parameters		**Ribonuclease A**			**Cytochrome C**			**Lysozyme**			**Resolution**	
			V_r	V_r		V_r	V_r		V_r	V_r			
		pH	calc.	exp.	Error	calc.	exp.	Error	calc.	exp.	Error	Calc.	Exp.
$c_{salt,iso}$	225 mM	6.92	1.92	2.10	0.18 9.4 %	7.38	7.28	0.10 1.4 %	15.90	15.11	0.79 5.0 %	5.46	5.18
			MC: 1.56 to 2.58			MC: 6.48 to 8.34			MC: 11.88 to 20.94				
	235 mM	6.92	1.80	2.10	0.30 16.7 %	6.18	6.18	0 0 %	12.78	12.42	0.79 2.8 %	4.38	4.08
			MC: 1.5 to 2.22			MC: 5.64 to 7.08			MC: 9.60 to 16.38				
$c_{initial,1}$ $c_{final,1}$ V_1	161 mM 410 mM 28 ml	6.32	6.36	6.45	0.09 1.4 %	11.76	11.80	0.04 0.3 %	15.84	16.72	0.88 5.6 %	4.8	4.08
			MC: 4.92 to 8.04			MC: 10.38 to 13.32			MC: 14.34 to 17.4				
$c_{initial,2}$ $c_{final,2}$ V_2 $c_{initial,3}$ $c_{final,3}$ V_3	183 mM 297 mM 2 ml 116 mM 651 mM 28 ml	6.44	5.10	2.90	2.2 43 %	10.38	11.04	0.66 6.4 %	13.02	14.23	1.21 9.3 %	2.64	3.19
			MC: 2.02 to 7.01			MC: 9.66 to 11.43			MC: 12.24 to 14.01				
$c_{initial,4}$ $c_{final,4}$ $pH_{initial,4}$ $pH_{final,4}$ V_4	200 mM 400 mM 6.93 5.15 28 ml		3.16	2.24	0.93 41.1 %	7.65	7.27	0.38 5.0 %	11.22	11.58	0.36 3.2 %	4.07	4.38
			MC: 2.78 to 3.88			MC: 6.82 to 8.59			MC: 10.51 to 12.93				

Tab. 4, chromatograms are shown in Fig. 5 B.

3.5 Salt gradient elution

The first gradient run was a linear salt gradient from 161 mM to 410 mM over 28 ml ran at pH 6.32. The gradient shape was restricted to have a monotonously increasing shape with a minimum difference of 250 mM salt between start and final concentration. The model predicted a resolution of 4.9 ml and the closest eluting component was predicted to be lysozyme. Predicted retention volumes had an error of 0.09 ml (1.4 %), 0.04 ml (0.3 %) and 0.88 ml (5.6 %), predictions are shown in Fig. 5 C (detailed results are summarized in Tab. 4). Due to the relatively high error for lysozyme, the experimental resolution was only 4.08 ml.

In the second gradient elution example (Fig. 5 D) the gradient was not restricted to any particular shape. A steep gradient from 183 mM to 297 mM over 2 ml was used to elute ribonuclease A, according to the model, followed by a relatively flat gradient from 163 mM to 631 mM over 28 ml, all at pH 6.44. The resolution was calculated to be 2.64. Retention volumes of cytochrome C and lysozyme were predicted with an error below 10 %, while the retention volume of ribonuclease A was off by about 2.2 ml (43 %) a possible explanation is given in the next section. The experimentally determined resolution was 3.19.

The third and last gradient was a linear salt gradient from 200 mM to 400 mM over 29.5 ml

while simultaneously decreasing the pH from 6.93 to 5.15. Predicted retention volumes matched the experimentally determined ones with an error of 0.93 ml (41.1 %), 0.38 ml (5.2 %) and 0.36 ml (3.2 %) as shown in Fig. 5 E. A resolution of 4.38 ml was achieved experimentally, while the predicted value was 4.07.

3.6 Error estimation

Especially for the second gradient elution, the error in retention volumes was relatively high (2.2 ml or 43 % for ribonuclease A, the highest error of all validation runs). To access whether this high error was due to a general problem with SMA parameter estimation, a series of Monte-Carlo simulations was performed, assuming a normal distributed error of up to 5 % on all salt concentrations and the pH. 20.000 simulations runs were performed for each experiment. The results are shown in Fig. 6 for the second gradient elution. This figure shows the calculated retention time on the x-axis and the calculated resolution on the y-axis, each dot is the result of one modelling run, each cluster relates to one protein. Interestingly, the retention volume of ribonuclease (cluster on the left) was most strongly affected by the artificial error and varied between 2.02 and 7.01 ml. The reason for this is the very short first salt gradient at the beginning that was supposed to speed-up ribonuclease A elution. If salt concentrations were slightly too low or pH was slightly too high, ribonuclease A would not elute in this small gradient but rather in the longer second gradient, which was actually seen in the experimental data, explaining the relatively high error of 2.2 ml measured for ribonuclease A. Actually, all measured retention volumes summarized in Tab. 4 were within the limits determined by Monte-Carlo simulations, except for two cases: 1) lysozyme in the bi-linear gradient that eluted after 14.23 ml, while the Monte Carlo simulation gave a maximum retention volume of 14.01 ml and 2) ribonuclease A elution in the inverse pH gradient.

3.7 Maximum resolution, efficiency and robustness

Looking at the Monte Carlo simulation results summarized in Tab. 4, two robustness issues become obvious:

1. As retention times increase in an isocratic elution, errors in both salt concentration and pH can lead to a wide distribution of retention volumes (e.g. 11.9 to 20.9 ml for lysozyme eluting at 225 mM (+/- error) ionic strength)).

2. For gradient elution, if the gradient shape becomes more complex, elution volumes can also significantly shift (2 to 7 ml for ribonuclease A in the bi-linear gradient elution).

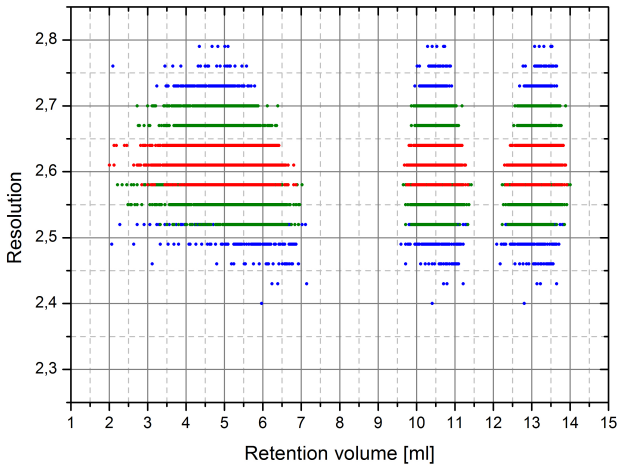

Figure 6: Sensitivity analysis for the bi-linear gradient elution for ribonuclease A (left cluster), cytochrome C (middle cluster) and lysozyme (right cluster). Errors of 5 % were put on pH and on all salt concentrations. Blue & green & red dots represent all individuals, green and red represent 95 % of all individuals and red dots represent 50 % of all individuals.

To determine what kind of gradient is a) most selective in terms of archiving the highest resolution, b) most efficient and c) most robust an in silico study was done with 250.000 random parameter combinations for isocratic, salt gradient and pH gradient elution. The results are shown in Fig. 7. As to be expected, the highest resolution could be achieved with isocratic elution, resulting in a maximum resolution of 14. Of course, for isocratic elution peak broadening is more pronounced compared to gradient elution, but as this is not accounted for in the model we used, this will also not be part of the discussion. pH gradients operated under isocratic salt conditions achieved the same maximum resolution (data not shown). In contrast to this, salt gradient mode performed worse in terms of maximum resolution, as it stayed well below 6. Surprisingly, isocratic elution also outperformed gradient elution in terms of elution efficiency (meaning a maximum resolution at minimum time). For all conditions, the same resolution could be achieved with reduced elution times. Again, the more pronounced peak broadening for an isocratic elution could have a negative impact. Figure 7 C shows that pH gradient elution combined with a salt gradient gave higher resolutions of up to 7 ml compared to a salt gradients at constant pH. When running a process in an industrial environment, process robustness becomes an issue. To study the effect of small fluctuations in salt concentrations and pH, conditions were chosen for four systems, one for each elution mode (isocratic, isocratic + pH gradient, salt gradient and salt gradient + pH gradient). Systems were chosen that gave a resolution of 5 at minimum in a total volume as low as possible, results are shown in Fig. 8. While retention volumes for all modes including a salt gradient showed deviations of about +/- 1.3 ml independent

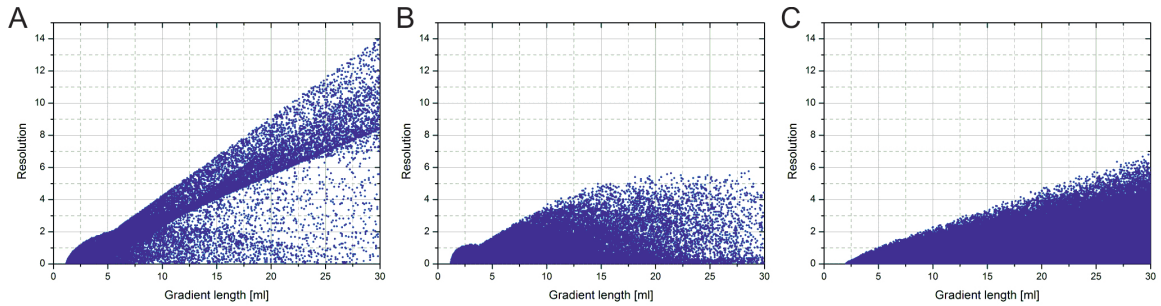

Figure 7: Resolution over gradient length plots for isocratic elution (**A**), gradient elution with a minimum increase of 250 mM (**B**) and pH gradient combined with a salt gradient with a minimum increase of 250 mM (**C**).

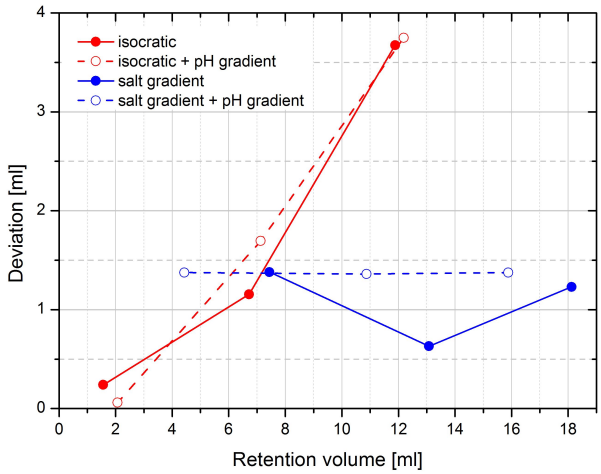

Figure 8: Range of deviations in retention volumes for four different elution modes resulting from Monte Carlo simulations (5 % error on salt concentrations and pH).

of retention volume (except cytochrome C in the salt gradient which showed significantly lower deviations), all isocratic elution modes showed an increasing deviation with increasing retention volume of up to +/- 3.75 ml for lysozyme eluting last. Furthermore, in the worst case scenario the resolution in the isocratic runs dropped down from 5 to 3.7 for constant pH and even further down to 1.8 for the pH gradient runs, compared to 4.4 for the salt gradient at constant pH and 4.8 for the salt gradient combined with a pH gradient.

Although magnitudes of the effects seen might be specific for this particular separation task, the overall trends will probably hold for other separations as well: using a salt gradient with the minimal slope necessary to achieve the needed resolution might be the best choice for a robust process. Adding a pH gradient could further increase selectivity without a negative effect on process robustness.

4 Conclusion

In this manuscript we were able to show that all possible elutions modes in ion exchange chromatography can well be captured by an equation that was originally derived for isocratic elution only. Using this equation, pH gradients in combination with a set of isocratic elutions were successfully used to determine a set of SMA parameters for a wide range of pH values for three different model proteins. Resulting SMA parameters could be used to calculate retention volumes for a number of different elution modes, predictions compared well to experimental data and deviations could be explained by Monte Carlo simulations. SMA parameters were also accurate enough to allow for an in silico optimization of separation conditions and a subsequent robustness analysis, showing that isocratic elution conditions (at constant pH or combined with a pH gradient) can yield the highest resolutions but are also more sensitive to small changes in salt concentrations and pH. Compared to that, salt gradient conditions at constant pH were more robust but less selective. Simultaneously applying a pH gradient increased maximal resolution without decreasing robustness.

Isoform separation and binding site determination of mono-PEGylated lysozyme with pH gradient chromatography

Benjamin Maiser[a,c], Frieder Kröner[a,c], Florian Dismer[a], Gerald Brenner-Weiß[b], Jürgen Hubbuch[a]

[a] *Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany*

[b] *Insitute of Functional Interfaces, Karlsruhe Institute of Technology (KIT), Eggenstein-Leopoldshafen, Germany*

[c] *These authors contributed equally to this manuscript*

Abstract

Covalent attachment of PEG to proteins, known as PEGylation, is currently one of the main approaches for improving the pharmacokinetics of biopharmaceuticals. However, the separation and characterization especially of positional isoforms of PEGylated proteins is still a challenging task. A common purification strategy uses ion exchange chromatography with increasing ionic strength by shallow salt gradients. This paper presents a method which applies a linear pH gradient chromatography to separate five of six possible isoforms of mono PEGylated lysozyme, modified with 5 kDa and 10 kDa mPEG-aldehyde. To identify the corresponding PEGylation sites a comparison of elution pH values and calculated isoelectric points of each isoform, was used. The resulting correlation showed an $R^2 > 0.99$. Fractionation, tryptic digestion and subsequent MALDI-MS analysis of each peak, verified the predicted elution order. Based on UV areas the N-terminal amine at lysine 1 exhibited the highest reactivity, followed by the lysine 33 residue.

Journal of Chromatography A (Volume 1268, December 14, Pages 102-8)

1 Introduction

Since the first PEGylation in 1977 by Abuchowski and Davis, polymer modification with poly(ethylene glycol) (PEG) has become an important method to enhance the pharmacological properties of therapeutic biopharmaceuticals [110]. The covalent attachment of PEG chains to a target molecule is well established and successfully used for numerous FDA approved proteins such as PEGylated interferon-α and erythropoietin (Pegasys® and Micera® from Hoffmann LaRoche, respectively).

Advantages of polymer modification generally include increased circulation half-life and a reduced immunogenicity of the conjugate compared to the unmodified form. Additional positive effects of PEGylation can be an increased thermal stability as well as a higher solubility which are also important for the therapeutics final formulation [111–113]. These changes in the pharmacological behaviour can mostly be explained with the increased hydrodynamic radius of the conjugate and the resulting "shielding effect" of the attached PEG, which is reviewed in detail by many publications [114–116].

For PEG attachment, various activated PEG agents and coupling strategies are commercially available. Depending on the polymer modification, the attachment takes place at different surface residues of the target molecule. A common chemistry for PEGylation targets accessible amino residues such as lysine or the N-terminus. For this, a modification with succinimidyl activated PEGs (PEG-NHS) or PEG-aldehyde as shown in Fig. 1, can be applied. While PEG-NHS is capable of binding also with histidine and tyrosine residues, the latter approach allows a main binding of PEG at the N-terminal α-amine at low pH values and was chosen by Kinstler et. al. [117, 118] to develop polymer modified granulocyte colony stimulating factor (G-CSF, Neulasta® from Amgen). Reactions with target molecules that exhibit more than one acces-

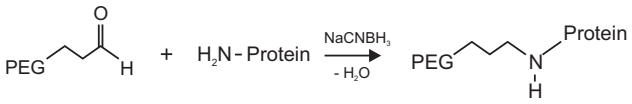

Figure 1: PEGylation reaction with PEG-aldehyde and $NaCNBH_3$ as reducing agent.

sible conjugation site, result in randomly attached PEGs and thus in heterogeneous mixtures. The resulting product consists of proteins with a different number of bound PEG and positional isoforms. Due to steric hindrance of the attached polymer, changes in PEGylation degree and binding site can have a major influence on biological activity of the different conjugates [119, 120]. With only 7 % residual activity compared to the native protein, Pegasys®, a mixture of different PEG-interferon (IFN) isoforms, shows how drastic this effect can be [121]. Additional experiments with PEG-IFN showed a range of residual activity from 6 % to 40 % depending on

the binding site of the attached PEG [122]. The loss in activity is generally compensated by the increased body residence time, but the high variance in activity underlines the influence of the PEGylation site. Consequently, mixture characterization and isoform identification is of high interest and also needed for regulatory approval. Additionally, the selective PEGylation either of a single site or the screening of reaction conditions towards favorable PEGylation sites with high residual activity are preferred.

The separation of conjugate mixtures regarding the PEGylation degree can be achieved effortlessly with size exclusion chromatography (SEC), due to the increase in size. Thus, fast monitoring of crude PEG-protein mixtures combined with an improved data evaluation, such as multivariate data analysis (MVDA) can be achieved by size based separation and can be used in a first step to screen and optimize different PEGylation conditions [123]. A less systematic approach for a lysozyme PEGylation optimization without an isoform analytic was shown for example by Moosmann et. al [124].

However, the separation of isoforms with varying attachment site and the preparative purification is challenging. Many approaches showed that ion exchange chromatography (IEC) is an effective tool and currently the method of choice for conjugate and positional isoform separation, by means of shallow salt gradients [125–127]. The different behaviour is based upon the shielding effect of attached polymer chains, and thus reduced interactions between chromatographic matrix and protein. Additionally a decreasing protein surface charge with an increasing number of bound PEG weakens the interaction with the oppositely charged resin, which was already shown by Fee in an *in silico* approach [128]. Applying classical salt gradient chromatography, different lysozyme PEGylation studies with PEG-NHS and PEG-aldehyde showed that it was not possible to separate more than three isoforms, although six isoforms are being formed during PEGylation. Even though some approaches showed promising results, only little attention was paid to alternative ion exchange chromatography with pH gradients [122, 129].

Besides the isoform separation, identification of PEG attachment sites is another challenge. Time consuming methods that are widely used comprise Edman degradation for small peptides and peptide mapping with combined MALDI-TOF analysis [130–132]. The mass spectrometric approach was applied by Lee and Park [130] for the characterization of PEGylated lysozyme. For lysine residue modification the authors used biotin-PEG-NHS, to separate PEG-peptide fragments from unmodified peptides, after tryptic digestion. Employing mass spectrometrical analysis of the peptide fragments, three positional isoforms with different reactivities were identified.

The presented work describes the separation of mono-PEGylated lysozyme isoforms with a linear pH gradient on a cation exchange column. In contrast to salt gradient chromatography runs, a significant increase in resolution could be achieved and five of six possible isoforms were

separated. Based on the assumption that every PEG conjugation to amino residues neutralizes an effective charge of the protein, a fast *in silico* approach was used to calculate the isoelectric point of each isoform. The elution pH values were correlated with the calculated isoelectric points to identify the PEGylation sites and to determine the isoform reactivities. Peptide mapping and the common mass spectrometric approach was applied to verify the results.

2 Materials and methods

2.1 Chemicals

Sodium phosphate and sodium chloride for chromatography buffer and PEGylation buffer preparation were obtained from Merck (Darmstadt, Germany). Hen egg white lysozyme and sodium cyanoborohydride ($NaCNBH_3$) were provided from Sigma-Aldrich (St. Louis, MO, USA). Methoxy-PEG-propionaldehyde (mPEG-aldehyde) with an average molecular weight of 5 kDa and 10 kDa were provided from NOF Cooperation (Tokyo, Japan). CABS (Cyclohexylaminobutansulfonic acid) with a pK_a value of 10.7 was provided from Santa Cruz Biotechnology Inc. (Santa Cruz, CA, USA) and was used as buffer substance for pH-gradient chromatography. The calibration of the pH meter was performed with high precision calibration standards from Hanna instruments (Woonsocket, RI, USA). The reagents for the proteolytic digestion of the PEGylated protein, including ammoniumbicarbonate which was used as the buffering substance, as well as dithiothreitol and iodoacetamide which were used for the reduction and alkylation of the disulfide bonds, were obtained from Sigma-Aldrich. Proteolytic digestion was prepared with proteomics grade trypsin from Sigma-Aldrich and RapiGest surfactant (Waters Corporation, Milford, MA, USA). For the hydrolysis of the acid labile surfactant trifluoracetic acid (TFA) from Sigma-Aldrich was used. For MALDI matrix preparation α-cyano-4-hydroxycinnamic acid (CHCA) from Sigma-Aldrich and LC-MS grade acetonitrile from Merck (Darmstadt, Germany) was used. Ultrapure water was generated with the Arium pro water purification system from Sartorius Stedim (Goettingen, Germany). All solutions used for chromatography were filtered using 0.2 μm cellulose acetate filters from Sartorius Stedim (Goettingen, Germany) and degassed for 20 min in an ultrasonic bath.

2.2 Batch PEGylation reaction

Lysozyme (5 mg/mL) and mPEG-aldehyde with a molar polymer to protein ratio of 6:1 were dissolved in a 25 mM sodium phosphate buffer pH 7.2. containing 20 mM sodium cyanoborohydride as reducing agent. The reaction was carried out in a continuously shaken falcon tube at room temperature, for about 3.5 h. Monitoring of the PEGylation reaction was conducted

with size exclusion chromatography (SEC) using a Superdex 200 GL10/300 column on a Äkta Ettan system (GE Healthcare, Uppsala, Sweden). For SEC 25 mM sodium phosphate pH 7.2 containing 150 mM NaCl as mobile phase was used.

2.3 Analytical protocol

2.3.1 Separation of PEGylation degree

After the PEGylation reaction an IEC was used as a first chromatographic to separate the different PEGylation degrees and to stop the reaction. The reaction mixture was separated with a Toyopearl GigaCap S-650M resin (Tosoh Bioscience GmbH, Stuttgart, Germany) packed according to the manufacturers protocol into an Omnifit glas column (25 mm x 400 mm, Diba Industries Ltd., Cambridge, UK). The resulting bed volume was 13.4 mL. The sample was diluted with ultrapure water (1:1) to reduce ionic strength. After column equilibration with running buffer (25 mM sodium phosphate, pH 7.2), 50 mL sample mixture was loaded onto the column. For elution a gradient ranging from 0 to 40 % of the elution buffer (25 mM sodium phosphate, 500 mM sodium chloride, pH 7.2) was performed over 21 column volumes. The flow rate was maintained at 1 mL/min. To obtain mono-PEG lysozyme samples, the fractionation volume was set to 5 ml. Until further measurement, the fractions were pooled and stored at -32°C.

2.3.2 Molecular weight determination

Peak fractions from IEC were analyzed by SEC with a light scattering (LS) detector. This analysis was conducted using an Äkta Ettan system from GE Healthcare (Uppsala, Sweden) in combination with a Dawn Heleos 8+ multi-angle LS detector and an Optilap rEX refractive index (RI) detector, both from Wyatt Technology (Santa Barbara, USA). The LS detector was equipped with a fused silica cell and a laser with a wavelength of 658 nm. The LS and RI detector were calibrated with toluene and NaCl, respectively. The LS detector was normalized using 2 mg/ml of bovine serum albumin (BSA) monomer from Sigma Aldrich, as reference. For SEC a Superdex 200 GL10/300 (GE Healthcare, Uppsala, Sweden) with a mobile phase of 25 mM sodium phosphate, pH 7.2, containing 150 mM NaCl was used. The flow rate was set to 0.8 ml/min. Injection volumes between 50 and 100 μl were chosen. After UV absorbance monitoring at 280 nm using the Äkta UV-900 monitor, each sample passed the LS and RI detector. Correction of detector alignment and band broadening, as well as molecular weight (M_w) calculation were done by the ASTRA software (software version 5.3.4.18)

2.3.3 Separation of isoforms

Isoform separation was conducted on a MonoS 4.6/100 column from GE Healthcare (Uppsala, Sweden). Sample volumes between 100 μl and 150 μl of previously purified mono-PEG lysozyme (modified with 5 kDa and 10 kDa mPEG) were chosen. For pH-gradient elution 20 mM CABS was used as buffer component. Running buffer A and elution buffer B were titrated with 4 M NaOH to pH 10.5 and pH 11.5, respectively. The pH was measured with a pH meter from Hanna Instruments (Woonsocket, RI, USA), calibrated from pH 10 to 12 with high precision calibration standards. The elution was carried out with a linear gradient ranging from 0 % to 100 % buffer B over 12 column volumes. The flow rate was set to 1.5 ml/min. To obtain samples for MALDI-TOF analysis the resulting peaks from mono-$PEG_{5.000}$-lysozyme isoforms were fractionated with a constant volume of 250 μl. The fractions of multiple runs were pooled to reach sufficient amount of single isoforms for subsequent mass spectrometrical peptide mapping. The results of the pH gradient separation were additionally compared with a classical salt gradient based cation exchange chromatography. For salt gradient elution the same column, injection volume as well as gradient length was used. As buffer system 25 mM sodium phosphate pH 7.2 was used. The elution was carried out with a salt gradient reaching from 0 mM to 500 mM NaCl.

2.3.4 Tryptic digestion of PEGylated lysozyme

After the chromatographic runs the collected fractions were transferred to VivaSpin 20 ultrafilters (Sartorius Stedim, Germany) with a molecular weight cutoff of 5 kDa and were diafiltrated into 50 mM ammoniumbicarbonate buffer with pH 8.0 including 0.1 % (v/v) RapiGest. Additionally the fractions were concentrated to the maximum degree and the protein concentration was determined via UV 280 nm absorption measurement on a Infinite M200 plate reader (Tecan, Maennedorf, Switzerland). The samples were then chemically reduced by the addition of dithiotreitol up to a final concentration of 20 mM and denatured for 30 minutes at 60°C on a thermoshaker (Eppendorf, Hamburg, Germany). Alkylation was carried out by adding iodoacetamide with a final concentration of 60 mM and mixing the samples for 45 min on a shaker under exclusion of light. The prepared samples were then digested with a trypsin to protein ratio of 1/2 - 1/10, depending on the sample protein concentrations. The digestion was carried out at 37°C over night. After the digestion, TFA was added to a final concentration of 0.5 % (v/v) and the samples were incubated for another 45 minutes and then centrifuged to remove the RapiGest surfactant from the solution. The samples were stored at -32°C until measurement.

2.3.5 MALDI-MS based PEGylation site identification

To prepare samples for the final MALDI-MS measurement, they were processed with C18 ZipTips (Millipore Corporation, Billerica, MA, USA). The samples were bound to preconditioned C18 ZipTips by pipetting and then eluted with the matrix solution, 10 mg/ml α-cyano-4-hydroxycinnamic acid (CHCA) in 70 % (v/v) ACN and directly spotted on a MALDI stainless steel target. After drying the sample spots, MS analytics were analyzed in a MALDI TOF/TOF 4800 analyzer (Applied Biosystems, Framingham, MA, USA). The mass spectrometer was run in positive, reflectron mode. Optimized device settings for the matrix were used. The measured data was analyzed and exported using Data Explorer Software 4.0 (Applied Biosystems, Framingham, MA, USA). To identify the PEGylation site the masses of the expected peptides were predicted with the PeptideCutter tool (Expasy - SIB Bioinformatics Resource Portal), modified by the mass of the carbamidomethylation and compared to the results of the single measurements. The PEGylated peptide fragment is heavier by the mass of the PEG molecule (5008 Da) and could therefore be identified. Due to the polydispersity of the PEG molecule the average mass of the measured fragment was determined by "gauss-fitting" the resulting signal and determining the maxima of the fitted curve with MATLAB (MathWorks, Natick, MA, USA).

2.4 pI calculation

Isoelectric point calculation of native lysozyme and PEGylated lysozyme isoforms were conducted with the freely accessible web tool 'protein continuum electrostatics' (PCE). This tool calculates pK_a values of titratable groups in proteins solving the Poisson-Boltzmann equation based on the MEAD (macroscopic electrostatics with atomic detail) program, developed by Bashford [133, 134]. For pI calculation, the structural information of native lysozyme (PDB-ID: 132L) obtained from the RSCB protein data bank was used. With resulting pK_a values, pI calculation was done applying the Henderson-Hasselbalch equation with Microsoft Excel (Microsoft, Redmond, MA, USA). The pI value of each PEG-isoform was calculated with neutralized lysine residues, involved in the conjugation.

3 Results and discussion

3.1 Lysozyme PEGylation degree

As PEGylation target the model protein lysozyme from chicken egg was used in this work. According to the manufacturers datasheet it has a molecular weight of 14.3 kDa and an isoelectric point of 11.3. Possible binding sites for the PEG-aldehyde reaction are six lysine residues and the N-terminal amino group. The three dimensional model, shown in Fig. 2, illustrates a lysozyme

molecule and depicts that all lysine residues as well as the N-terminus are located at the surface of the molecule, which can be explained with their hydrophilic character. With the additional N-terminal amino group, lysine 1 contains two binding sites. Consequently, PEGylation reactions with amino coupling PEG agents can yield six mono-PEG lysozyme isoforms. However, different PEGylation studies with PEG-NHS and PEG-aldehyde showed, that it was not possible to detect and separate more than three isoforms [127]. The PEGylation reaction was monitored

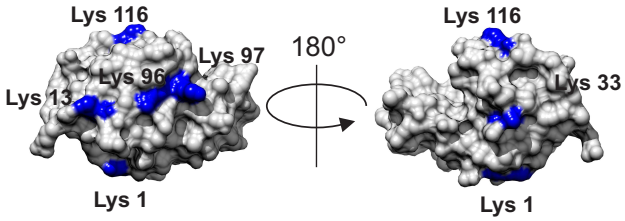

Figure 2: Two orientations of the three dimensional structure of lysozyme with labeled surface lysine residues.

using SEC, to optimize the reaction time towards a high mono-PEG-lysozyme yield. In Fig. 3 the SEC chromatograms of PEG-lysozyme mixtures modified with 5 kDa and 10 kDa PEG-aldehyde, after 3.5 hours reaction time are shown. The different elution volumes in each mixture result from the increase in size by PEG attachment and indicates different PEGylation degree. As can be seen, di-$PEG_{5.000}$- and mono-$PEG_{10.000}$-lysozyme conjugates have the same retention time in SEC, resulting from the same hydrodynamic radius. This illustrates that the increase in size is only dependent of the molecular weight and is regardless of the number of bound PEG, which was already shown by Fee and Van Alstine [135]. As PEG-aldehyde is a non UV active component, unreacted PEG was not detectable in the 280 nm UV trace.

For mono-PEG-lysozyme sample preparation, preparative purification of PEG-lysozyme mixtures with ion exchange chromatography was applied, according to Moosmann et al. [136]. In addition to PEG-lysozyme conjugate separation, this method provides a removal of unreacted native protein as well as a flow through of unreacted PEG. Fig. 4 depicts a resulting chromatogram with a 5 kDa PEG-lysozyme mixture and shows a comparable elution behaviour to SEC (see Fig. 3). After peak fractionation, SEC with combined light scattering was conducted with mono-PEG-lysozyme samples to verify purity and PEGylation degree. The overall molecular mass of pooled mono-$PEG_{5.000}$-lysozyme samples were calculated to 19.2 kDa with a protein fraction of 14.1 kDa. Calculated values of mono-$PEG_{10.000}$-lysozyme samples provided masses of 24.2 kDa and 13.9 kDa, respectively. SEC chromatograms of purified mono-PEG-lysozyme samples showed no impurities of other PEGylation degrees (data not shown).

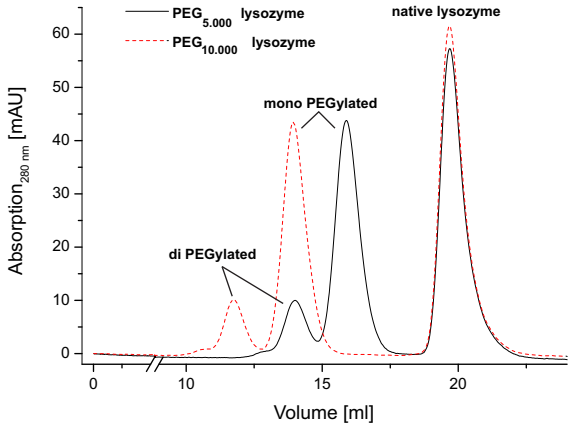

Figure 3: SEC chromatograms of 5 kDa and 10 kDa PEGylated lysozyme mixtures after 3.5 h with a Superdex 200 10/300 GL column and following elution order: di-PEG-lysozyme, mono-PEG-lysozym and native lysozyme.

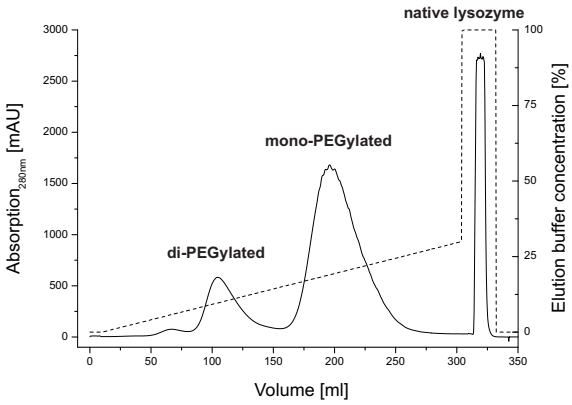

Figure 4: Chromatogram of preparative purification of 5 kDa PEG-lysozyme mixture, using a GigaCap 650S ion exchange adsorber. PEGylation degree was validated with SEC-LS.

3.2 Isoform separation

The isoform separation was conducted with purified mono-PEGylated-lysozyme samples, modified with 5 kDa and 10 kDa PEG-aldehyde. In Fig. 5 the resulting chromatograms with salt gradient and pH gradient elution are shown. Five peaks could be detected in the 280 nm trace in the pH elution chromatogram, while salt gradient elution resulted in a separation of only three peaks. The different elution volumes of each peak can be linked to altered surface charge due to different PEGylation sites. As only pure mono-PEG lysozyme samples were injected, each peak represents lysozyme with one bound PEG chain attached to a different amino residue. Even though lysozyme consists of six possible conjugation sites, as mentioned before, only five isoforms were detected. The separation of a 10 kDa PEG modified lysozyme sample showed lower elution pH values, but the same elution pattern compared to mono-$PEG_{5.000}$-lysozyme.

3.3 Isoform identification

In pH gradient chromatography the elution of proteins is based on a decreasing surface charge, which results in a decreasing strength of interaction between the adsorbed proteins and the oppositely charged adsorber. Thus, resulting elution pH values correlate sometimes with the proteins isoelectric point [137]. In classical salt gradient chromatography the proteins are eluted due to an ion exchange process. This difference leads to salt and pH gradient elution strategies, which can result in a different separation performance.
To indicate the elution order of the separated isoforms, pI calculations were performed with the

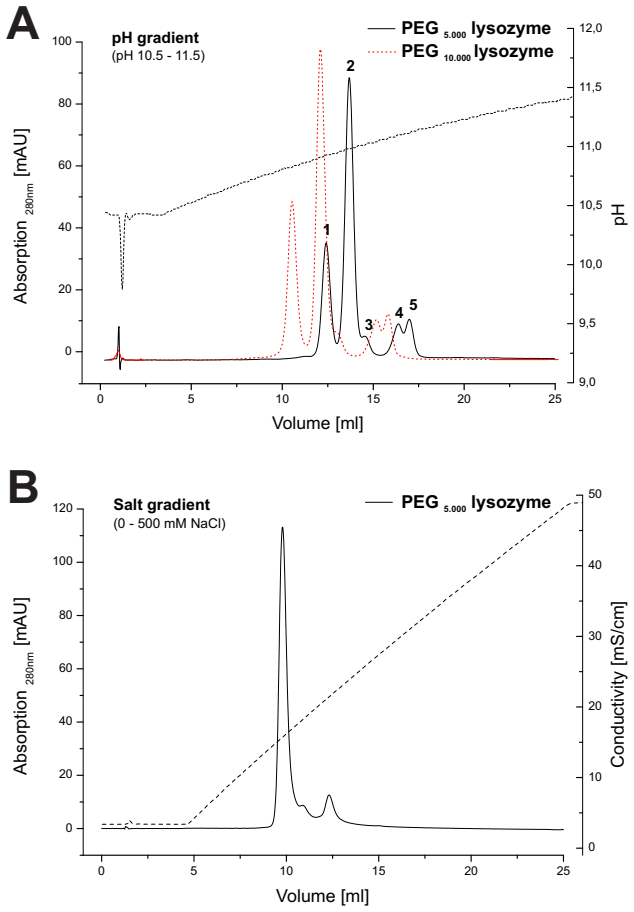

Figure 5: Chromatogragraphic separation of PEG-lysozyme isoforms using a pH gradient with a pH rage from 10.5 to 11.5 (**A**) and a salt gradient from 0 mM to 500 mM NaCl in 25 mM phosphate buffer, pH 7.2 (**B**) on a Mono S 4.6/100 column. The peak labels correspond to different positional isoforms.

PCE tool from Mitera et al. [6]. Applying the described method, the pI for native lysozyme was calculated as 11.28, with a good correlation to already published values by Ahamed et. al [137]. The altered charge of the lysine residues caused by attaching a PEG molecule was taken as a basis for the isoform pI calculations. Assuming that every PEG conjugation to the primary amine of a lysine residue neutralizes the positive charge, surface net charge, isoelectric point and also elution pH are reduced compared to native lysozyme. For isoform pI calculation of PEG-lysozyme, lysine pK_a values involved in the conjugation were ignored. Titration curve calculations, using the Henderson-Hasselbalch equation to determine the proteins pI, were applied regardless of the attached PEG molecular weight. Comparing the elution pH values with the calculated isoelectric points, a peak identification to corresponding binding sites as shown in Tab. 1 was conducted. The isoform with PEG binding site lysine 33 (in the following labeled as PEG-lys_{33}) exhibited

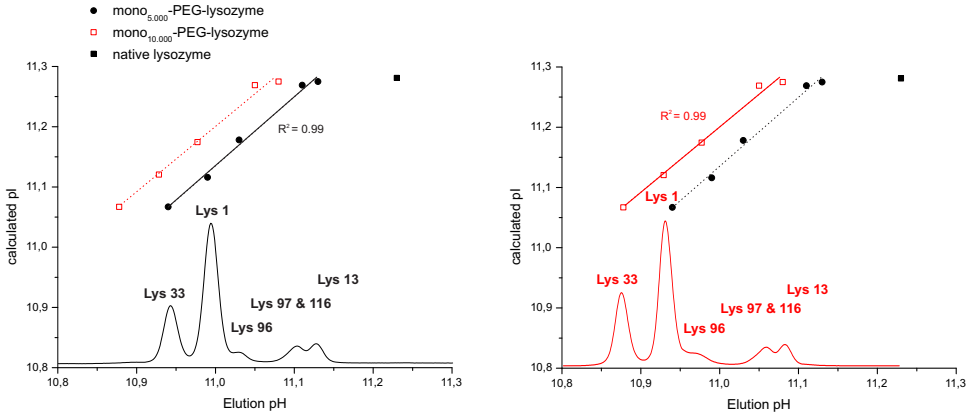

Figure 6: Plotted elution pH values of mono-PEG-lysozyme isoforms(left: modified with 5 kDa, right: modified with 10 kDa) and native lysozyme versus calculated isoelectric points.

the lowest calculated pI and was assigned to the peak with the lowest elution pH. As lysine 97 and 116 exhibited the same calculated pI, these isoforms were assumed to elute in one peak. The isoform elution sequence obtained from pI calculation results, starting with the lowest elution pH, was as follows: PEG-lys$_{33}$, PEG-lys$_{1}$, PEG-lys$_{96}$, followed by PEG-lys$_{97}$ and PEG-lys$_{116}$ eluting in one peak and PEG-lys$_{13}$ eluting last.

Fig. 6 depicts the obtained results and shows that native lysozyme eluted roughly at its isoelectric point. PEG-isoforms eluted earlier and at lower pH values than native lysozyme, as expected. Additionally the elution volume of both PEG species correlated with an $R^2 > 0.99$ to the calculated pI values. However, the elution pH of 10 kDa-PEG isoforms and 5 kDa-PEG isoforms were shifted to lower pH values, indicating a lower interaction to the adsorber matrix with increasing PEG weight. In addition to reduced net charge, attached PEG chains might weaken the interaction of the isoforms to the adsorber matrix, due to steric hindrance. This effect is probably dependent on the attached molecular PEG weight. An increased protein-resin

Table 1: Elution pH values of separated mono-PEG-lysozyme isoforms and coresponding calculated isolectric points. Calculation of isoelectric points were applied with pK_a values based from PCE-tool.

Binding site	Calculated pI	Elution pH 5 kDa	Elution pH 10 kDa
Lys 33	11.07	10.94	10.88
Lys 1	11.12	10.99	10.92
Lys 96	11.18	11.03	10.98
Lys 97 Lys 116	11.27	11.11	11.05
Lys 13	11.28	11.13	11.08
native lysozyme	11.28	11.23	

distance, might thus explain the differences in the elution behaviour of mono-$PEG_{5.000}$- and mono-$PEG_{10.000}$-lysozyme isoforms. A study by Abe et al. [119] investigated this effect for ion exchange chromatography and can be used for a detailed view of PEG-protein binding mechanisms. By comparing the UV areas of each isoform, the reactivity of the different binding sites were evaluated. Thus, PEGylation with 5 kDa and 10 kDa PEG occurs preferentially at lys1, followed by lys33 using the described reaction conditions. Only small reactivites were found for lysine 97, 116 and 96.

3.4 MALDI-MS analysis

To verify the conducted binding site identification, peptide mapping in combination with mass spectrometric analysis was applied. The common procedure involves a tryptic digestion of the isoform mixture and the comparison of the resulting peptides with the peptide pattern of a digested native protein solution. Considering trypsin is sterically blocked by attached PEG molecules, missing peptides refer to PEG-conjugation sites.
Instead of analyzing the peptide pattern, we decided to investigate the mass of the PEG-peptide fragments. Lysozyme peptides resulting from tryptic digestion, hydrolyzed at lys1 or lys97 consist of only one amino acid and are consequently difficult to detect within the signals of the used MALDI matrix. With the analysis of the heavier PEG-peptide fragments, this problem was avoided.
Analyzing unbound PEG-aldehyde samples with the described MALDI-TOF analytics the polydisperse character of the polymer was obvious, which is caused by the production process. Pure PEG samples showed a normally distributed signal which was composed of single mass peaks. Each peak showed differences in weight of 44 Da, indicating PEG chains with a different number of monomer units. Fitting the PEG sample peaks using a gaussian fit function in MATLAB®, the calculated average molecular mass was 5031 Da. The quality analysis by the PEG supplier provided an average weight of 5008 Da for the same charge, confirming the used fitting method. Four PEG-peptide fragments were detected in the separated and fractionated PEG-lysozyme isoform samples. Mass differences between the mass peaks in each detected PEG-peptide signal were 44 Da and refer consequently to PEG. The PEG-peptide fragments, detected in isoform fractions PEG-lys_{33} and PEG-lys_1 are shown in Fig. 7, with the corresponding calculated gaussian fits and peak maxima. Weight differences between the two fragment signals can be noticed, which are caused by different peptide masses due to varying PEG binding sites. Calculated theoretical PEG-peptide fragments and evaluated average masses of analyzed PEG-peptide signals are listed in Tab. 2 and assigned to their corresponding binding sites. For calculating the theoretical total PEG-peptide mass, the PEG suppliers weight specification was used and the mass of

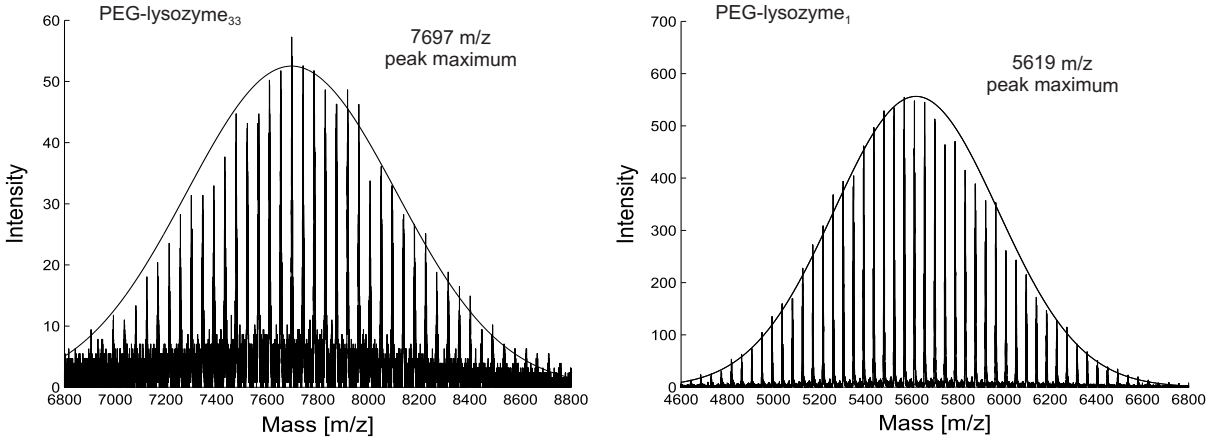

Figure 7: MALDI-TOF spectra with PEG-peptide fragments of two 5 kDa PEG-lysozyme isoforms. Spectra correspond to lysozyme with bound PEG at lysine 33 (PEG-lysozyme$_{33}$) and lysine 1 (PEG-lysozyme$_1$), respectively

.

Table 2: PEG binding site, resulting mass of peptide-fragments and measured mass of PEG-peptide fragments of mono$_{5000}$-PEG-lysozyme isoforms. Theoretical mass of each PEG-peptide-fragment was calculated with a PEG weigth of 5008 Da and corresponding peptides. The mass of a water molecule was subtracted additionally.

Peak number	**PEGylation site**	**Peptide mass [m/z]**			**PEG-peptide fragment mass [m/z]**	
					expected	measured
1	**Lys 33**	22-**33**	34-35			
		1325.63	1428.65		7744	7697
2	**Lys 1**	**1**	2-5			
		147.11	478.28		5615	5619
3	**Lys 96**	74-**96**	97			
		2508	147.11		7645	n/A
4	**Lys 116**	115-**116**	117-125			
		307.14	1045.54		6343	6352
	Lys 97	**97**	98-112			
		147.11	1675.80		6813	n/A
5	**Lys 13**	6-**13**	14	15-21		
		893.42	175.12	874.42	6933	6854

one water molecule was subtracted. A good agreement of theoretical and the measured masses of the first two peaks, labeled as PEG-lys$_{33}$ and PEG-lys$_1$ can be noticed. Considering baseline separation was provided between these peaks the proposed peak identification was validated. Peak 3, labeled as PEG-lys$_{96}$ contained no detectable PEG mass peaks. Due to small sample concentrations, a detection of the PEG-peptide signal was not possible. In addition, the small

resolution to peak 2, hindered the fractionation of sufficient sample volume.
Instead of two PEG signals in peak 4, only one PEG-peptide fragment was found. With an average weight of 6352 m/z the measured signal corresponds well to the theoretical PEG-peptide mass of 6343 m/z of the proposed lys116 binding site.
In peak 5, labeled as PEG-lys_{13} a theoretical PEG-peptide weight consisting of two peptides and attached PEG could not be found. Thus, the corresponding PEG-peptide mass was corrected assuming a sterical hindrance of PEG blocking the tryptic digestion between lys13 and arg14 additionally. The resulting expected fragment weight of 6933 m/z agreed then with the measured average weight of 6854 m/z. Comparing the mass peaks of the last isoform peak, it can be noticed that the measured mass of 6854 m/z is between the expected weight of binding site lys97 and lys13. Therefore, the proposed lys13 binding site of peak 5 could not be validated by MS analytics.
A comparison of the findings using MS analytics with the alternative *in silico* binding site identification showed a good correlation and demonstrates the applicability of this fast method. Lysozyme PEG binding site identifications were already conducted by numerous authors including Lee et. al and Tilton et al. [130, 138], using MS analytics. Studies by Lee et al. suggest lysine 33, lysine 97 and lysine 116 as major PEG binding sites, using a PEG-NHS modification. Tilton applied a PEG-aldehyde reaction and proposed a predominant N-terminal (lysine 1) modification of mono-PEG-lysozyme, with slight modifications at lysine residues 33 and 97. This suggests a major influence of the coupling reaction to the resulting PEG conjugation site. By comparing the results made in this study with results published by Tilton with the same PEG reaction, a good agreement regarding the lysine reactivities can be detected. In addition both results corresponds with the fact that main attachment of PEG-aldehyde primarily occurs at the N-terminal amino group at low buffer pH values.

4 Conclusion and outlook

In this paper the separation of five positional isoforms of mono-PEGylated lysozyme is shown, and thus represents an increase in resolution compared to published results. By applying a pH gradient elution, we were able to achieve a superior resolution in contrast to classical salt gradient runs, which underlines the high separation performance of pH gradients regarding charge variant separation. An isoform elution order and the resulting PEG conjugation sites were identified, using a fast in silico approach. The resulting reactivities of the identified lysine residues were evaluated, and showed a good agreement to comparable PEG reactions [138]. To validate the proposed binding site results, common MS analytics were applied. The usability of the *in silico* binding site identification application was only shown for PEG-lysozyme. With the presented

chromatographic method, the evaluation of lysozyme isoform reaction kinetics are possible and are already under investigation. In addition, on-column PEGylation experiments can now be conducted and analyzed for changes in lysozyme binding orientations.

Conclusions

In this work novel approaches to accelerate the development of chromatographic purification processes for low concentrated proteins contained in very complex biotechnological mixtures, are demonstrated. Due to the low concentration, the feedstock complexity as well as the lack of applicable affinity techniques, purification process development for these proteins is exceptionally challenging. In 2007 Ahamed et al. [3] demonstrated that pH gradient IEC can be used as a very useful tool for chromatographic process development.

However, establishing pH gradient based IEC methods is very challenging as no systematic approach to generate buffer systems for linear, controllable pH gradients with a preferably low ionic strength existed so far. A fast and simple methodology to routinely generate such buffer systems was necessary. Therefore, an *in silico* methodology to optimize the composition of buffer systems for pH gradient IEC (chapter I) was developed. Multicomponent buffers with constant buffer capacity, minimized ionic strength and a linear titration curve over a defined pH range (7.5 pH units) were generated, using the *in silico* approach. These were successfully applied to form linear pH gradients in AEC / CEC. After validating the buffers applicability for protein separation, pH gradient IEC was further applied to characterize model proteins and comparing the results with those of isoelectric focusing. The comparison revealed clear mechanistic differences between the separation methods (AEC / CEC / IEF), motivating a selective use for different separation challenges. With the *in silico* tool, buffer compositions in any pH range for many different purposes, e.g. for pH gradient IEC as a process developmental tool, can be easily generated.

A systematical screening approach applying pH gradient IEC to determine the information necessary for laying out purification of low titer proteins from cell lysates by IEC, is demonstrated in chapter II. The screening process is based on multidimensional fractionation of the biotechnological feedstock, applying pH gradient IEC, target specific analytics (Dot-Blot), SDS PAGE as well as protein identification using LC-MS. First the elution-pH values of the target protein are determined using long range pH gradient IEC (ΔpH = 7.5) and target specific analytics. Afterwards, critical impurities, proteins with electrostatic properties comparable to those of the target protein, are specifically fractionated using a short range pH gradient IEC (ΔpH = 2) focused on the elution range of the target protein. Fractions are further separated using SDS-PAGE, the contained proteins are identified using LC-MS based protein identification. By applying this screening procedure a cell lysate (SF9/Baculovirus) containing a recombinantly expressed autoantigen was successfully characterized. The obtained information was used to lay out a simple two-step purification based on IEC, achieving 80 % purity for the target protein.

The achieved purity underlines the usage of such an analytical screening for purification process development instead of classical, knowledge-based purification process development. The shown screening procedure based on pH gradient IEC is very time efficient, can be easily handled by a skilled lab technician, while only a small amount of sample is necessary, allowing purification process development at a very early developmental stage.
A 2nd approach for purification process development focusing on analytically acquiring parameters for *in silico* simulation / optimization of chromatographic separations of complex biotechnological mixtures with low titers, is demonstrated in chapter III. *In silico* optimization of chromatographic procedures has clear advantages over classical process development, the low developmental time, costs and the increased gain of process understanding due to the possibility to carry out a large amount of experiments *in silico*. A rapid, multidimensional fractionation approach applying IEC, RPC, sample concentrating and HT-CGE was designed and successfully applied to characterize a complex biotechnological mixture (SF9/Baculovirus cell lysate). To obtain the linear SMA parameters for each single protein contained, the lysate was fractionated by salt gradient AEC and analysed, using the described approach four times with different gradient volumes. From the retention volumes and the determined column parameters the linear SMA parameters of each protein were successfully determined [95]. Acquired parameters were finally validated by comparing experimental and predicted retention volumes of each protein contained. Also three exemplary chromatograms were predicted using the transport-dispersive model. Predicted retention volumes of all proteins, as well as predicted chromatographic elution profiles of the three chosen proteins were highly comparable to the experimentally determined ones, underlining the quality of the determined parameters. The advantages of the described approach are its velocity, the low amount of sample consumed as well as the possibility to acquire parameters, allowing for *in silico* process optimization. The described multi-dimensional fractionation approach might not only be used for parameter estimation but also for monitoring the composition of such biotechnological feedstocks. To our knowledge this is the first approach treating single protein parameter acquisition for mechanistic modelling from very complex mixtures.
The methodologies described in chapter II-III are clearly showing that analytical screenings to characterize biotechnological feedstocks can be applied as a very useful tool in chromatographic process development. Furthermore, the screening procedures are designed for the application on very complex biotechnological feedstocks containing low concentrated proteins. Both approaches can be used at an early stage of process development as all the experiments can be performed in analytical lab scale. This concludes to the fact that the major challenge, to speed up purification process development for low concentrated proteins contained in very complex mixtures, e.g. biotechnologically produced autoantigens, can be directly addressed by implementing these procedures already at the stage of upstream process development. Put together with the fact that

these analytical procedures can be performed very fast by a trained technician, a considerable decrease of the time-to-market for such proteins can be achieved.
The work described in chapter IV is focusing on a later stage of chromatographic (IEC) process development, meaning the decision on detailed chromatographic conditions (pH, salt conc, gradient operations) for the purification of proteins, also treating the robustness of a process as an important parameter. To acquire the linear SMA parameters for *in silico* process development a set of pH gradient operations at different salt concentrations (10 - 350 mM) and isocratic elution experiments at three different pH values (pH 5, 7, 9) were performed. SMA parameters were successfully determined from the acquired data by extending the applicability of the existing SMA model [5] to protein elution in pH gradient IEC. The obtained parameters were successfully validated by comparing predicted elution behaviour of the proteins in various possible elution modes to experimentally determined data. Resulting deviations could be explained by typical experimental errors (pH value, salt concentration), which was concluded from a Monte Carlo simulation expecting a defined error range for pH and salt concentration. *In silico* studies on the separation, concerning resolution, efficiency and robustness resulted in the fact that highest chromatographic resolution can be achieved with pH gradient or isocratic elution, while robustness is very low. Salt gradient IEC provides the lowest resolution, while showing very high robustness. The resolution of salt gradient IEC can be increased by adding a pH gradient to the salt gradient, while robustness of the separation is not negatively influenced hereby. Determining SMA parameters by the described set of experiments delivers very valuable datasets which allow full optimization of ion exchange chromatography including all elution modes, furthermore robustness studies can be performed *in silico* for all conditions and elution modes. To our knowledge, the approach to obtain full SMA parameter sets for a wide pH range is the first one described in literature and will surely find further applications in academic and industrial R&D for purification of proteins. The demonstrated approach might also be used for purification process development of low concentrated proteins from very complex mixtures but most possibly at a later stage in process development, as tracking the elution of many different proteins in pH gradient and isocratic IEC experiments would be very challenging. Therefore, determining the SMA parameters using the described procedure rather demands for a partly purified intermediate product instead of a complex feedstock, containing a reduced amount of different proteins.
In the last chapter a further application of pH gradient IEC in protein charge variant analysis is described. The idea of using pH gradient IEC for the analysis of PEGylated proteins emerged due to the discovery of increased chromatographic resolution achieved by pH gradient IEC (chapter IV) compared to classical salt gradient IEC. For validation, mono-PEGylated lysozyme was analytically separated using pH gradient IEC. Thereby, we were able to separate five of

six possible PEGylation isomers. PEGylation sites were identified by correlating calculated pI values of the different isoforms with the elution-pH values ($R^2 > 0.99$). Elution order was finally validated by identifying the PEGylation sites using tryptic digestion and MALDI-MS. To our knowledge the separation of three isoforms of mono-PEGylated lysozyme using chromatography was the maximum amount so far. Therefore, pH gradient IEC clearly demonstrated its strength in the separation of charge variants due to its increased chromatographic resolution. This leads to the conclusion that pH gradient IEC might also be a very useful tool for the analysis of other protein charge variants, e.g. glycosylation variants of mAbs.

It can be concluded that the set objective to develop simple, more rational methodologies to accelerate purification process development for protein diagnostics, contained with low titers in complex biotechnological feedstocks was clearly achieved. The developed analytical screening methodologies have proven their applicability as experimental approaches for regular application to accelerate and simplify purification process development for these proteins. Even though protein concentrations were very low, screening procedures are still kept on analytical scale consuming only low amounts of sample. Due to the low amount of sample consumed, these methods may be applied in lab scale at a very early process developmental stage. Additionally, a systematic approach to generate buffer systems for pH gradient IEC was developed, enabling the application of pH gradient IEC as standard separation method for proteins and other molecules. pH gradient IEC proved its usability not only as a screening tool for process development, or to obtain SMA parameters, but also as a tool for the analytical separation of protein charge variants, e.g. positional isoforms of PEGylated proteins.

Outlook

The purification process development for proteins expressed with low titers in very complex biotechnological mixtures is still a challenging subject, even though the described analytical screenings represent a big step towards rational and accelerated process development. All approaches demonstrated in this work are applying ion exchange chromatography only, for purification of these proteins, currently not taking other separation technologies into account. The next step would be to implement other chromatographic or even non-chromatographic techniques into the analytical screening approaches to extend the portfolio of applicable techniques to purify these proteins. The biggest challenges associated with the implementation of these technologies are already solved, as the analytical methodologies might be easily linked to other screening experiments.

A bigger step further would be the development of complete databases containing purification parameters of all contained components in standard expression systems, e.g. CHO, insect cell/Baculovirus expression systems. These databases would further be applicable to completely develop purification processes *in silico* for target proteins expressed in these systems. As proteome analytical technologies got more and more technically advanced and affordable in the last few years this perspective is not a simple pipe dream any more and most probably first promising scientific results in that field will be seen in the next few years. Nevertheless, obtaining full parameter databases for standard expression systems is still a long road to go.

Due to the introduction of a systematic method to develop buffer compositions for pH gradient IEC, method development for protein separation using pH gradient IEC is not a challenge any more. As it was already demonstrated that by using pH gradient IEC superior chromatographic resolution can be achieved, pH gradient IEC will most probably find further application in the separation of protein charge variants, e.g. glycosylation variants of antibodies. Further applications might be the usage of buffer systems for pH gradients in Mixed Mode Chromatography or as a universal buffer systems for pH screenings, e.g. in crystallization or other pH dependent screenings.

Bibliography

[1] C. A. Brooks and S. M. Cramer, "Steric mass-action ion exchange: Displacement profiles and induced salt gradients," *AIChE J.*, vol. 38, pp. 1969–1978, Jun 1992.

[2] I. E. W. Group, *ICH Harmonised Tripartite Guideline Validation of analytical procedures: text and methodology Q2(R1)*, 2005.

[3] T. Ahamed, B. K. Nfor, P. D. E. M. Verhaert, G. W. K. van Dedem, L. A. M. van der Wielen, M. H. M. Eppink, E. J. A. X. van de Sandt, and M. Ottens, "ph-gradient ion-exchange chromatography: an analytical tool for design and optimization of protein separations.," *J. Chromatogr., A*, vol. 1164, pp. 181–188, Sep 2007.

[4] A. A. Shukla, S. S. Bae, J. Moore, and S. M. Cramer, "Structural characteristics of low-molecular-mass displacers for cation-exchange chromatography: Ii. role of the stationary phase," *J. Chromatogr., A*, vol. 827, pp. 295–310, Jul 1998.

[5] L. Pedersen, J. Mollerup, E. Hansen, and A. Jungbauer, "Whey proteins as a model system for chromatographic separation of proteins," *J. Chromatogr., B*, vol. 790, pp. 161–173, Jun 2003.

[6] M. A. Miteva, P. Tuffery, and B. O. Villoutreix, "Pce: web tools to compute protein continuum electrostatics.," *Nucleic Acids Res.*, vol. 33, pp. 372–375, Jul 2005.

[7] E. S. Lander and et al., "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, pp. 860–921, Feb 2001.

[8] E. Pennisi, "Gene counters struggle to get the right answer," *Science*, vol. 301, pp. 1040–1041, Aug 2003.

[9] J. C. Venter and et al., "The sequence of the human genome," *Science*, vol. 291, pp. 1304–1351, Feb 2001.

[10] *Global Biopharmaceutical Market Report (2010 - 2015).* International Market Analysis Research and Consulting Group.

[11] B. Leader, Q. J. Baca, and D. E. Golan, "Protein therapeutics: a summary and pharmacological classification," *Nat. Rev. Drug. Discov.*, vol. 7, pp. 21–39, Feb 2008.

[12] *Global Immunodiagnostics Market 2011-2015.* TechNavio.

[13] E. Langer and J. Ranck, "Capacity bottleneck squeezed by downstream processes," *Bioprocess Int.*, pp. 14–17, Mar 2006.

[14] U. Gottschalk, "The renaissance of protein purification," *Int. BioPharm*, Jun 2006.

[15] S. Farid, "Economic drivers and trade-offs in antibody purification processes," *Int. BioPharm*, Mar 2009.

[16] *Protein Purification Handbook*. GE-Healthcare.

[17] G. Guiochon and L. A. Beaver, "Separation science is the key to successful biopharmaceuticals," *J. Chromatogr., A*, vol. 1218, pp. 8836–8858, Dec 2011.

[18] B. K. Nfor, P. D. Verhaert, L. A. van der Wielen, J. Hubbuch, and M. Ottens, "Rational and systematic protein purification process development: the next generation," *Trends Biotechnol.*, vol. 27, pp. 673–679, Dec 2009.

[19] B. Kelley, "Industrialization of mab production technology: The bioprocessing industry at a crossroads," *mAbs*, vol. 1, pp. 443 – 452, Sep 2009.

[20] T. Puzyn, J. Leszczynski, and M. T. Cronin, eds., *Recent advantages in QSAR studies*. Springer AG, 2010.

[21] F. Dismer and J. Hubbuch, "3d structure-based protein retention prediction for ion-exchange chromatography," *J. Chromatogr., A*, vol. 1217, pp. 1343–1353, Feb 2010.

[22] B. K. Nfor, T. Ahamed, M. W. Pinkse, L. A. van der Wielen, P. D. Verhaert, G. W. van Dedem, M. H. Eppink, E. J. van de Sandt, and M. Ottens, "Multi-dimensional fractionation and characterization of crude protein mixtures: Toward establishment of a database of protein purification process development parameters," *Biotechnol. Bioeng.*, vol. 109, pp. 3070–3083, Dec 2012.

[23] Y. Shapira, N. Agmon-Levin, and Y. Shoenfeld, "Geoepidemiology of autoimmune rheumatic diseases.," *Nat. Rev. Rheumatol.*, vol. 6, pp. 468–476, Aug 2010.

[24] P. Youinou, J.-O. Pers, M. E. Gershwin, and Y. Shoenfeld, "Geo-epidemiology and autoimmunity.," *J. Autoimmun.*, vol. 34, pp. 163–167, May 2010.

[25] A. Lleo, P. Invernizzi, B. Gao, M. Podda, and M. E. Gershwin, "Definition of human autoimmunity–autoantibodies versus autoimmune disease.," *Autoimmun. Rev.*, vol. 9, pp. 259–266, Mar 2010.

[26] S. M. Anderton, "Post-translational modifications of self antigens: implications for autoimmunity," *Curr. Opin. Immunol.*, vol. 16, pp. 753–758, Dec 2004.

[27] M. R. Lerner and J. A. Steitz, "Antibodies to small nuclear rnas complexed with proteins are produced by patients with systemic lupus erythematosus.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 76, pp. 5495–5499, Nov 1979.

[28] J. Schmitt, "Recombinant autoantigens for diagnosis and therapy of autoimmune diseases.," *Biomed Pharmacother.*, vol. 57, pp. 261–268, Sep 2003.

[29] M. Pepaj, S. R. Wilson, K. Novotna, E. Lundanes, and T. Greibrokk, "Two-dimensional capillary liquid chromatography: ph gradient ion exchange and reversed phase chromatography for rapid separation of proteins.," *J. Chromatogr., A*, vol. 1120, pp. 132–141, Jul 2006.

[30] T. Ahamed, S. Chilamkurthi, B. K. Nfor, P. D. E. M. Verhaert, G. W. K. van Dedem, L. A. M. van der Wielen, M. H. M. Eppink, E. J. A. X. van de Sandt, and M. Ottens, "Selection of ph-related parameters in ion-exchange chromatography using ph-gradient operations.," *J. Chromatogr., A*, vol. 1194, pp. 22–29, Jun 2008.

[31] J. C. Rea, G. T. Moreno, Y. Lou, and D. Farnan, "Validation of a ph gradient-based ion-exchange chromatography method for high-resolution monoclonal antibody charge variant separations," *J. Pharm. Biomed. Anal.*, vol. 54, pp. 317–323, Jan 2011.

[32] B. Maiser, F. Kröner, F. Dismer, G. Brenner-Weiß, and J. Hubbuch, "Isoform separation and binding site determination of mono-pegylated lysozyme with ph gradient chromatography," *J. Chromatogr., A*, vol. 1268, pp. 102–108, Dec 2012.

[33] L. Shan and D. J. Anderson, "Effect of buffer concentration on gradient chromatofocusing performance separating proteins on a high-performance deae column.," *J. Chromatogr., A*, vol. 909, pp. 191–205, Feb 2001.

[34] L. Sluyterman and O. Elgersma, "Chromatofocusing: Isoelectric focusing on ion-exchange columns : I. general principles," *J. Chromatogr., A*, vol. 150, pp. 17–30, Mar 1978.

[35] L. I. Tsonev and A. G. Hirsh, "Theory and applications of a novel ion exchange chromatographic technology using controlled ph gradients for separating proteins on anionic and cationic stationary phases," *J. Chromatogr., A*, vol. 1200, pp. 166–182, Jul 2008.

[36] F. Celentano, E. Gianazza, G. Dossi, and P. G. Righetti, "Buffer systems and ph gradient simulation," *Chemometr. Intell. Lab.*, vol. 1, pp. 349–358, Oct 1987.

[37] P. G. Righetti, M. Fazio, C. Tonani, E. Gianazza, and F. C. Celentano, "ph gradients generated by polyprotic buffers. ii. experimental validation.," *J. Biochem. Biophys. Methods.*, vol. 16, pp. 129–140, Jun 1988.

[38] F. C. Celentano, C. Tonani, M. Fazio, E. Gianazza, and P. G. Righetti, "ph gradients generated by polyprotic buffers. i. theory and computer simulation.," *J. Biochem. Biophys. Methods.*, vol. 16, pp. 109–128, Jun 1988.

[39] E. Giaffreda, C. Tonani, and P. G. Righetti, "ph gradient simulator for electrophoretic techniques in a windows environment," *J. Chromatogr., A*, vol. 630, pp. 313–327, Feb 1993.

[40] T. Thiel, L. Liczkowski, and S. T. Bissen, "New zwitterionic butanesulfonic acids that extend the alkaline range of four families of good buffers: Evaluation for use in biological systems," *J. Biochem. Biophys. Methods.*, vol. 37, pp. 117–129, Nov 1998.

[41] C. se Kim, F. Kueppers, P. Dimaria, J. Farooqui, S. Kim, and W. Paik, "Enzymatic trimethylation of residue-72 lysine in cytochrome c. effect on the total structure," *Biochim. Biophys. Acta*, vol. 622, pp. 144–150, Mar 1980.

[42] P. G. Righetti and T. Caravaggio, "Isoelectric points and molecular weights of proteins : A table," *J. Chromatogr., A*, vol. 127, pp. 1–28, Apr 1976.

[43] W.-O. Baek and M. Vijayalakshmi, "Effect of chemical glycosylation of rnase a on the protein stability and surface histidines accessibility in immobilized metal ion affinity electrophoresis (image) system," *Biochim. Biophys. Acta*, vol. 1336, pp. 394–402, Oct 1997.

[44] H. van der Wel and K. Loeve, "Isolation and characterization of thaumatin i and ii, the sweet-tasting proteins from thaumatococcus daniellii benth," *Eur. J. Biochem.*, vol. 31, pp. 221–225, Aug 1972.

[45] S. P. Basu, S. N. Rao, and J. A. Hartsuck, "Influence of fatty acid and time of focusing on the isoelectric focusing of human plasma albumin," *Biochim. Biophys. Acta*, vol. 533, pp. 66–73, Mar 1978.

[46] W. W. P. Chang, C. Hobson, D. C. Bomberger, and L. V. Schneider, "Rapid separation of protein isoforms by capillary zone electrophoresis with new dynamic coatings," *Electrophoresis*, vol. 26, pp. 2179–2186, Jun 2005.

[47] C. Desert, C. Guérin-Dubiard, F. Nau, G. Jan, F. Val, and J. Mallard, "Comparison of different electrophoretic separations of hen egg white proteins," *J. Agric. Food. Chem.*, vol. 49, pp. 4553–4561, Oct 2001.

[48] M. Bier, L. Terminiello, J. A. Duke, R. J. Gibbs, and F. Nord, "Investigations on proteins and polymers. x. composition and fractionation of ovomucoid," *Arch. Biochem. Biophys.*, vol. 47, pp. 465–473, Dec 1953.

[49] C. T. Supuran, "Carbonic anhydrases: novel therapeutic applications for inhibitors and activators," *Nat. Rev. Drug. Discov.*, vol. 7, pp. 168–181, Feb. 2008.

[50] S. Funakoshi and H. F. Deutsch, "Human carbonic anhydrases. i. isolation and demonstration of isozymes in erythrocytes," *J. Biol. Chem.*, vol. 243, pp. 6474–6481, Dec 1968.

[51] V. W. S. Lee, Y.-L. Chen, and L. Konermann, "Reconstitution of acid-denatured holomyoglobin studied by time-resolved electrospray ionization mass spectrometry," *Anal. Chem.*, vol. 71, pp. 4154–4159, Oct 1999.

[52] A. M. de Vos, M. Hatada, H. van der Wel, H. Krabbendam, A. F. Peerdeman, and S. H. Kim, "Three-dimensional structure of thaumatin i, an intensely sweet protein," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 82, pp. 1406–1409, Mar 1985.

[53] T. T. Franco, A. T. Andrews, and J. A. Asenjo, "Use of chemically modified proteins to study the effect of a single protein property on partitioning in aqueous two-phase systems: Effect of surface charge," *Biotechnol. Bioeng.*, vol. 49, pp. 309–315, Feb 1996.

[54] Z. T. Milena Vespalcova, Radana Holsova and D. Vranova, "Findings and indication in the page-analysis of sweet protein thaumatin," *Chem. Listy*, vol. 99, pp. 362–364, Aug 2005.

[55] M. Gabaldon, "Preparation and characterization of bovine albumin isoforms," *Int. J. Biol. Macromol.*, vol. 30, pp. 259–267, Oct 2002.

[56] R. V. Wenn and J. Williams, "The isoelectric fractionation of hen's-egg ovotransferrin," *Biochem. J.*, vol. 108, pp. 69–74, Jun 1968.

[57] J. M. Gutteridge and Y. Hou, "Iron complexes and their reactivity in the bleomycin assay for radical-promoting loosely-bound iron," *Free Radic. Res. Commun.*, vol. 2, pp. 143–151, May 1986.

[58] G. de Jong and H. G. van Eijk, "Microheterogeneity of human serum transferrin: A biological phenomenon studied by isoelectric focusing in immobilized ph gradients," *Electrophoresis*, vol. 9, pp. 589–598, Apr 1988.

[59] A. Dautry-Varsat, A. Ciechanover, and H. F. Lodish, "ph and the recycling of transferrin during receptor-mediated endocytosis," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 80, pp. 2258–2262, Apr 1983.

[60] J. P. Landers, R. P. Oda, B. J. Madden, and T. C. Spelsberg, "High-performance capillary electrophoresis of glycoproteins: The use of modifiers of electroosmotic flow for analysis of microheterogeneity," *Anal. Biochem.*, vol. 205, pp. 115–124, Aug 1992.

[61] K. Yamashita, Y. Tachibana, A. Hitoi, and A. Kobata, "Sialic acid-containing sugar chains of hen ovalbumin and ovomucoid," *Carbohydr. Res.*, vol. 130, pp. 271–288, Jul 1984.

[62] T. H. Plummer, C. H. W. Hirs, and W. the technical assistance of Anne L. Tench, "The isolation of ribonuclease b, a glycoprotein, from bovine pancreatic juice," *J. Biol. Chem.*, vol. 238, pp. 1396–1401, Apr 1963.

[63] T. Samejima, M. Kamata, and K. Shibata, "Dissociation of bovine liver catalase at low ph," *J. Biochem.*, vol. 51, pp. 181–187, Oct 1962.

[64] T. Samejima, "Splitting of catalase molecule by alkali treatment," *J. Biochem.*, vol. 46, pp. 155–159, Jun 1959.

[65] F. Khan, P. M. Legler, R. M. Mease, E. H. Duncan, E. S. Bergmann-Leitner, and E. Angov, "Histidine affinity tags affect msp142 structural stability and immunodominance in mice," *Biotechnol. J.*, vol. 7, pp. 133–147, Jan 2012.

[66] T. W. Randolph, "The two faces of his-tag: Immune response versus ease of protein purification," *Biotechnol. J.*, vol. 7, pp. 18–19, Dec 2012.

[67] M. Lienqueo and J. A. Asenjo, "Use of expert systems for the synthesis of downstream protein processes," *Comput. Chem. Eng.*, vol. 24, pp. 2339–2350, Oct 2000.

[68] E. Leser and J. Asenjo, "Rational design of purification processes for recombinant proteins," *J. Chromatogr., B*, vol. 584, pp. 43–57, Dec 1992.

[69] T. Ahamed, M. Ottens, B. K. Nfor, G. W. van Dedem, and L. A. van der Wielen, "A generalized approach to thermodynamic properties of biomolecules for use in bioseparation process design," *Fluid Phase Equilib.*, vol. 241, pp. 268–282, Jan 2006.

[70] A. Ladiwala, K. Rege, C. M. Breneman, and S. M. Cramer, "A priori prediction of adsorption isotherm parameters and chromatographic behavior in ion-exchange systems," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 102, pp. 11710–11715, Aug 2005.

[71] S. Yamamoto, "Electrostatic interaction chromatography process for protein separations: Impact of engineering analysis of biorecognition mechanism on process optimization," *Chem. Eng. Technol.*, vol. 28, pp. 1387–1393, Nov 2005.

[72] F. Kröner and J. Hubbuch, "Systematic generation of buffer systems for ph gradient ion exchange chromatography and their application," *J. Chromatogr., A*, vol. 1285, pp. 78–87, Apr 2013.

[73] G. Candiano, M. Bruschi, L. Musante, L. Santucci, G. M. Ghiggeri, B. Carnemolla, P. Orecchia, L. Zardi, and P. G. Righetti, "Blue silver: A very sensitive colloidal coomassie g-250 staining for proteome analysis," *Electrophoresis*, vol. 25, pp. 1327–1333, May 2004.

[74] H. D. Meiring, E. van der Heeft, G. J. ten Hove, and A. P. J. M. de Jong, "Nanoscale lc-ms(n): technical design and applications to peptide and protein analysis," *J. Sep. Sci.*, vol. 25, pp. 557–568, Jun 2002.

[75] W. Xu, H. Zhou, and F. E. Regnier, "Regio-specific adsorption of cytochrome c on negatively charged surfaces," *Anal. Chem.*, vol. 75, pp. 1931–1940, Mar 2003.

[76] L. P. Pathange, D. R. Bevan, and C. Zhang, "Effects of protein microstructure on the retention time of t4 lysozyme variants in cation exchange chromatography," *Anal. Chem.*, vol. 81, pp. 649–655, Dec 2009.

[77] F. Dismer and J. Hubbuch, "A novel approach to characterize the binding orientation of lysozyme on ion-exchange resins," *J. Chromatogr., A*, vol. 1149, pp. 312–320, May 2007.

[78] R. A. Silva, M. D. Urzúa, D. F. S. Petri, and P. L. Dubin, "Protein adsorption onto polyelectrolyte layers: Effects of protein hydrophobicity and charge anisotropy," *Langmuir*, vol. 26, pp. 14032–14038, Sep 2010.

[79] W. K. Chung, S. T. Evans, A. S. Freed, J. J. Keba, Z. C. Baer, K. Rege, and S. M. Cramer, "Utilization of lysozyme charge ladders to examine the effects of protein surface charge distribution on binding affinity in ion exchange systems," *Langmuir*, vol. 26, pp. 759–768, Jan 2010.

[80] N. Carinhas, A. M. Robitaille, S. Moes, M. J. T. Carrondo, P. Jenoe, R. Oliveira, and P. M. Alves, "Quantitative proteomics of *Spodoptera frugiperda* cells during growth and baculovirus infection," *PLoS ONE*, vol. 6, e26444, pp. 1–12, Oct 2011.

[81] M. Tavaria, T. Gabriele, I. Kola, and R. L. Anderson, "A hitchhiker's guide to the human hsp70 family," *Cell Stress Chaperones*, vol. 1, pp. 23–28, Apr 1996.

[82] V. S. Mikhailov, K. Okano, and G. F. Rohrmann, "The redox state of the baculovirus single-stranded dna-binding protein lef-3 regulates its dna binding, unwinding, and annealing activities," *J. Biol. Chem.*, vol. 280, pp. 29444–29453, Jun 2005.

[83] R. Gary and A. Bretscher, "Heterotypic and homotypic asssociations between ezrin and moesin, two putative membrane-cytoskeletal linking proteins," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 90, pp. 10846–10850, Nov 1993.

[84] F. Li, N. Vijayasankaran, A. Y. Shen, R. Kiss, and A. Amanullah, "Cell culture processes for monoclonal antibody production," *mAbs*, vol. 2, pp. 466–479, Sep-Oct 2010.

[85] M. Wiendahl, P. Schulze Wierling, J. Nielsen, D. Fomsgaard 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," *Chem. Eng. Technol.*, vol. 31, pp. 893–903, May 2008.

[86] J. L. Coffman, J. F. Kramarczyk, and B. D. Kelley, "High-throughput screening of chromatographic separations: I. method development and column modeling," *Biotechnol. Bioeng.*, vol. 100, pp. 605–618, Jul 2008.

[87] B. D. Kelley, M. Switzer, P. Bastek, J. F. Kramarczyk, K. Molnar, T. Yu, and J. Coffmann, "High-throughput screening of chromatographic separations: Iv. ion-exchange," *Biotechnol. Bioeng.*, vol. 100, pp. 950–963, Aug 2008.

[88] R. Bhambure, K. Kumar, and A. S. Rathore, "High-throughput process development for biopharmaceutical drug substances," *Trends Biotechnol.*, vol. 29, pp. 127–135, Mar 2011.

[89] A. Osberghaus, S. Hepbildikler, S. Nath, M. Haindl, E. von Lieres, and J. Hubbuch, "Optimizing a chromatographic three component separation: A comparison of mechanistic and empiric modeling approaches," *J. Chromatogr., A*, vol. 1237, pp. 86–95, May 2012.

[90] A. Osberghaus, K. Drechsel, S. Hansen, S. Hepbildikler, S. Nath, M. Haindl, E. von Lieres, and J. Hubbuch, "Model-integrated process development demonstrated on the optimization of a robotic cation exchange step," *Chem. Eng. Sci.*, vol. 76, pp. 129–139, Jul 2012.

[91] B. K. Nfor, T. Ahamed, G. W. van Dedem, P. D. Verhaert, L. A. van der Wielen, M. H. Eppink, E. J. van de Sandt, and M. Ottens, "Model-based rational methodology for protein purification process synthesis," *Chem. Eng. Sci.*, vol. 89, pp. 185–195, Feb 2013.

[92] T. Ishihara and S. Yamamoto, "Optimization of monoclonal antibody purification by ion-exchange chromatography: Application of simple methods with linear gradient elution experimental data," *J. Chromatogr., A*, vol. 1069, pp. 99–106, Mar 2005. 17th International Symposium on Preparative and Process Chromatography; Ion Exchange, Adsorption/Desorption Processes and Related Separation Techniques.

[93] T. Ishihara, T. Kadoya, N. Endo, and S. Yamamoto, "Optimization of elution salt concentration in stepwise elution of protein chromatography using linear gradient elution data: Reducing residual protein a by cation-exchange chromatography in monoclonal antibody purification," *J. Chromatogr., A*, vol. 1114, pp. 97–101, May 2006.

[94] S. Chan, N. Titchener-Hooker, D. G. Bracewell, and E. Sørensen, "A systematic approach for modeling chromatographic processes – application to protein purification," *AIChE J.*, vol. 54, pp. 965–977, Feb 2008.

[95] E. S. Parente and D. B. Wetlaufer, "Influence of urea on the high-performance cation-exchange chromatography of hen egg white lysozyme," *J. Chromatogr., A*, vol. 288, pp. 389–398, Apr 1984.

[96] M. E. Young, P. A. Carroad, and R. L. Bell, "Estimation of diffusion coefficients of proteins," *Biotechnol. Bioeng.*, vol. 22, pp. 947–955, Feb 1980.

[97] E. J. Wilson and C. J. Geankoplis, "Liquid mass transfer at very low reynolds numbers in packed beds," *Ind. Eng. Chem. Fund.*, vol. 5, pp. 9–14, Feb 1966.

[98] J. H. Chon and G. Zarbis-Papastoitsis, "Advances in the production and downstream processing of antibodies," *N. Biotechnol.*, vol. 28, pp. 458–463, Sep 2011.

[99] A. A. Shukla and J. Thömmes, "Recent advances in large-scale production of monoclonal antibodies and related proteins," *Trends Biotechnol.*, vol. 28, pp. 253–261, Sep 2010.

[100] J. Dolan, J. Gant, and L. Snyder, "Gradient elution in high-performance liquid chromatography : Ii. practical application to reversed-phase systems," *J. Chromatogr., A*, vol. 165, pp. 31–58, Mar 1979.

[101] L. Snyder, J. Dolan, and J. Gant, "Gradient elution in high-performance liquid chromatography : I. theoretical basis for reversed-phase systems," *J. Chromatogr., A*, vol. 165, pp. 3–30, Mar 1979.

[102] T. Sasagawa, Y. Sakamoto, T. Hirose, T. Yoshida, Y. Kobayashi, Y. Sato, and K. Koizumi, "Prediction of retention times in ion-exchange chromatography," *J. Chromatogr., A*, vol. 485, pp. 533–540, Jan 1989.

[103] S. R. Gallant, A. Kundu, and S. M. Cramer, "Modeling non-linear elution of proteins in ion-exchange chromatography," *J. Chromatogr., A*, vol. 702, pp. 125–142, May 1995.

[104] S. Gallant, S. Vunnum, and S. Cramer, "Optimization of preparative ion-exchange chromatography of proteins: linear gradient separations," *J. Chromatogr., A*, vol. 725, pp. 295–314, May 1996.

[105] V. Natarajan and S. Cramer, "A methodology for the characterization of ion-exchange resins," *Separ. Sci. Technol.*, vol. 35, pp. 1719–1742, Apr 2000.

[106] V. Natarajan, B. W. Bequette, and S. M. Cramer, "Optimization of ion-exchange displacement separations: I. validation of an iterative scheme and its use as a methods development tool," *J. Chromatogr., A*, vol. 876, pp. 51–62, Apr 2000.

[107] J. C. Bosma and J. A. Wesselingh, "ph dependence of ion-exchange equilibrium of proteins," *AIChE J.*, vol. 44, pp. 2399–2409, Apr 1998.

[108] T. Yang, M. C. Sundling, A. S. Freed, C. M. Breneman, and S. M. Cramer, "Prediction of ph-dependent chromatographic behavior in ion-exchange systems," *Anal. Chem.*, vol. 79, pp. 8927–8939, Nov 2007.

[109] B. Guelat, G. Strählein, M. Lattuada, L. Delegrange, P. Valax, and M. Morbidelli, "Simulation model for overloaded monoclonal antibody variants separations in ion-exchange chromatography," *J. Chromatogr., A*, vol. 1253, pp. 32–43, Aug 2012.

[110] A. Abuchowski, T. Vanes, N. C. Palczuk, and F. F. Davis, "Alteration of immunological properties of bovine-serum-albumin by covalent attachment of polyethylene-glycol," *J. Biol. Chem.*, vol. 252, pp. 3578–3581, Jun 1977.

[111] Y. Nodake and N. Yamasaki, "Some properties of a macromelecular conjugate of lysozyme prepared by modification with a monomethoxypolyethylene glycol derivative," *Biosci. Biotechnol. Biochem.*, vol. 64, pp. 767–774, Apr 2000.

[112] B. K. Lee, J. S. Kwon, H. J. Kim, S. Yamamoto, and E. K. Lee, "Solid-phase pegylation of recombinant interferon alpha-2a for site-specific modification: Process performance, characterization, and in vitro bioactivity," *Bioconjug. Chem.*, vol. 18, pp. 1728–1734, Oct 2007.

[113] R. S. Rajan, T. Li, M. Aras, C. Sloey, W. Sutherland, H. Arai, R. Briddell, O. Kinstler, A. M. K. Lueras, Y. Zhang, and et al., "Modulation of protein aggregation by polyethylene glycol conjugation : Gcsf as a case study," *Protein Sci.*, vol. 15, pp. 1063–1075, May 2006.

[114] F. M. Veronese and G. Pasut, "Pegylation, successful approach to drug delivery," *Drug Discov.Today*, vol. 10, pp. 1451–1458, Nov 2005.

[115] J. M. Harris, N. E. Martin, and M. Modi, "Pegylation - a novel process for modifying pharmacokinetics," *Clin. Pharmacokinet.*, vol. 40, pp. 539–551, Aug 2001.

[116] V. Gaberc-Porekar, I. Zore, B. Podobnik, and V. Menart, "Obstacles and pitfalls in the pegylation of therapeutic proteins," *Curr. Opin. Drug. Discov. Devel.*, vol. 11, pp. 242–250, Mar 2008.

[117] O. Kinstler, G. Molineux, M. Treuheit, D. Ladd, and C. Gegg, "Mono-n-terminal poly(ethylene glycol)-protein conjugates," *Adv. Drug Deliver Rev.*, vol. 54, pp. 477–485, Jun 2002.

[118] O. B. Kinstler, D. N. Brems, S. L. Lauren, A. G. Paige, J. B. Hamburger, and M. J. Treuheit, "Characterization and stability of n-terminally pegylated rhg-csf," *Pharm. Res.*, vol. 13, pp. 996–1002, Jul 1996.

[119] M. Abe, P. Akbarzaderaleh, M. Hamachi, N. Yoshimoto, and S. Yamamoto, "Interaction mechanism of mono-pegylated proteins in electrostatic interaction chromatography.," *Biotechnol. J.*, vol. 5, pp. 477–483, May 2010.

[120] G. Pasut, "Pegylated alpha-interferons: two different strategies to achieve increased efiicacy," in *PEGylated Protein Drugs: Basic Science and Clinical Applications* (F. M. Veronese, ed.), pp. 205–216, Birkhäuser Verlag, 2009.

[121] F. M. Veronese and G. Pasut, "Pegylation: Posttranslational bioengineering of protein biotherapeutics," *Drug Discov. Today*, vol. 5, pp. 57–64, Sep 2008.

[122] S. P. Monkarsh, Y. M. Ma, A. Aglione, P. Bailon, D. Ciolek, B. DeBarbieri, M. C. Graves, K. Hollfelder, H. Michel, A. Palleroni, and et al., "Positional isomers of monopegylated interferon alpha-2a: Isolation, characterization, and biological activity," *Anal. Biochem.*, vol. 247, pp. 434–440, May 1997.

[123] S. K. Hansen, B. Maiser, and J. Hubbuch, "Rapid quantification of protein-peg conjugates by multivariate evaluation of chromatographic data," *J. Chromatogr., A*, vol. 1257, pp. 41–47, Aug 2012.

[124] A. Moosmann, J. Blath, R. Lindner, E. Müller, and H. Böttinger, "Aldehyde pegylation kinetics: a standard protein versus a pharmaceutically relevant single chain variable fragment.," *Bioconjug. Chem.*, vol. 22, pp. 1545–1558, Aug 2011.

[125] S. Foser, A. Schacher, K. A. Weyer, D. Brugger, E. Dietel, S. Marti, and T. Schreitmüller, "Isolation, structural characterization, and antiviral activity of positional isomers of

monopegylated interferon alpha-2a (pegasys)," *Protein Expres. Purif.*, vol. 30, pp. 78–87, Jul 2003.

[126] C. J. Fee and J. A. Van Alstine, "Peg-proteins: Reaction engineering and separation issues," *Chem. Eng. Sci.*, vol. 61, pp. 924–939, Feb 2006.

[127] S. Yamamoto, S. Fujii, N. Yoshimoto, and P. Akbarzadehlaleh, "Effects of protein conformational changes on separation performance in electrostatic interaction chromatography: unfolded proteins and pegylated proteins.," *J. Biotechnol.*, vol. 132, pp. 196–201, Oct 2007.

[128] C. J. Fee, "Protein conjugates purification and characterization," in *PEGylated Protein Drugs: Basic Science and Clinical Applications* (F. M. Veronese, ed.), pp. 113–125, Birkhäuser Verlag, 2009.

[129] P. K. Ng, J. He, and M. Snyder, "Separation of protein mixtures using ph-gradient cation-exchange chromatography.," *J. Chromatogr., A*, vol. 1216, pp. 1372–1376, Feb 2009.

[130] H. Lee and T. G. Park, "A novel method for identifying pegylation sites of protein using biotinylated peg derivatives," *J. Pharm. Sci.*, vol. 92, pp. 97–103, Jan 2003.

[131] O. B. Kinstler, D. N. Brems, S. L. Lauren, A. G. Paige, J. B. Hamburger, and M. J. Treuheit, "Characterization and stability of n-terminally pegylated rhg-csf," *Pharm. Res.*, vol. 13, pp. 996–1002, Jul 1996.

[132] J. R. Michael and J. M. Harris, "Attachment of degradable poly(ethylene glycol) to proteins has the potential to increase therapeutic efficacy," *J. Pharm. Sci.*, vol. 87, pp. 1440–1445, Nov 1998.

[133] D. Bashford and K. Gerwert, "Electrostatic calculations of the pka values of ionizable groups in bacteriorhodopsin.," *J. Mol. Bio.*, vol. 224, pp. 473–486, Mar 1992.

[134] D. Bashford, "An object-oriented programming suite for electrostatic effects in biological molecules an experience report on the mead project," in *Scientific Computing in Object-Oriented Parallel Environments* (Y. Ishikawa, R. Oldehoeft, J. Reynders, and M. Tholburn, eds.), vol. 1343 of *Lecture Notes in Computer Science*, pp. 233–240, Springer Berlin Heidelberg, 1997.

[135] C. J. Fee and J. M. Van Alstine, "Prediction of the viscosity radius and the size exclusion chromatography behavior of PEGylated proteins," *Bioconjug. Chem.*, vol. 15, pp. 1304–1313, Oct 2004.

[136] A. Moosmann, J. Christel, H. Böttinger, and E. Müller, "Analytical and preparative separation of pegylated lysozyme for the characterization of chromatography media.," *J. Chromatogr., A*, vol. 1217, pp. 209–215, Nov 2010.

[137] T. Ahamed, B. K. Nfor, P. Verhaert, G. van Dedem, L. van der Wielen, M. Eppink, E. van de Sandt, and M. Ottens, "ph-gradient ion-exchange chromatography: an analytical tool for design and optimization of protein separations.," *J. Chromatogr., A*, vol. 1164, pp. 181–188, Sep 2007.

[138] S. M. Daly, T. M. Przybycien, and R. D. Tilton, "Adsorption of poly(ethylene glycol)-modified lysozyme to silica," *Langmuir*, vol. 21, pp. 1328–1337, Feb 2005.