

DISSERTATION

Aqueous two phase extraction of proteins: From molecular understanding to process development

Autor:
M.Sc. Stefan A. OELMEIER
aus Bonn

Referent:
Prof. Jürgen HUBBUCH
Korreferent:
Prof. Matthias FRANZREB

zur Erlangung des akademischen Grades eines
Doktors der Ingenieurwissenschaften (Dr.-Ing.)
der Fakultät für Chemieingenieurwesen und Verfahrenstechnik des
Karlsruher Institutes für Technologie (KIT)

genehmigte
DISSERTATION

Tag der mündlichen Prüfung: 28. September 2012

To You, and Us.



DANKSAGUNGEN

- Herrn Professor Jürgen Hubbuch, dem Doktorvater meiner Arbeit, möchte ich herzlich danken für die Chance, in seiner Gruppe diese Arbeit anfertigen zu können. Die große Freiheit bei der Arbeit und die fachlichen Förderungen und Forderungen weiß ich sehr zu schätzen.
- Die finanzielle Unterstützung meiner Arbeit wurde von der Firma Boehringer Ingelheim GmbH & Co. KG geleistet. Hierfür möchte ich mich herzlich bedanken. Besonderen Dank für die fachliche Unterstützung von Seiten dieses Sponsors gilt Herrn Dr.-Ing Michael Dieterle, dem direkten Ansprechpartner für dieses Projekt, der dieses Projekt stets mit Interesse und Enthusiasmus begleitet hat, sowie Herrn Michael Richter, der die unterstützenden Laborarbeiten in Biberach durchgeführt hat und immer ein offener, interessierter, netter Diskussionspartner war.
- Ganz besonderen Dank möchte ich Herrn Dr.-Ing Florian Dimer, und Frau Sigrid Hansen aussprechen. Ihr seid mir in den letzten Jahren zu guten Freunden geworden und habt mich in meinen fachlichen, persönlichen und gesundheitlichen Hochs und Tiefs begleitet.
- Mein besonderer Dank auch gilt der gesamten Arbeitsgruppe "MAB" - besonders meinen Doktorgeschwistern Patrick, Ben, Frieder, Anna und Natalie. Ich bin froh und dankbar dafür, ein Mitglied eines so tollen Teams sein zu dürfen.
- Ich möchte mich ebenso bei meinen zwei Diplomanden, Herrn Dipl.Ing. Johannes Knoll und Herrn Dipl.Ing. Christoph Ladd-Effio für ihre sehr wertvolle Mitarbeit an meinem Projekt bedanken.
- Bianca, als wichtigster Mensch in meinem Leben, hast Du Dank verdient, der fernab von jeglicher für ich in Worte zu fassender Dimension liegt. Daher bemühe ich jemand anderes Worten - Bianca, you are the reason I am, you are all my reasons.

CONTENTS

1	Abstract	2
2	Abstract in German - Zusammenfassung	5
3	Introduction	9
3.1	Purification of high value proteins in the pharmaceutical industry	10
3.2	Automation and miniaturization in pharmaceutical downstream process development	11
3.3	Aqueous two-phase extraction - Applications in batch mode	12
3.4	Modeling protein partitioning in ATPS	13
3.5	Liquid-Liquid chromatography	14
3.6	Scope of this dissertation	18
4	Research proposal	20
5	Published scientific material and manuscripts	23
5.1	Manuscript I: Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation	26
5.2	Manuscript II: Evaluation of correlations of protein properties and ATPS distribution	48
5.3	Manuscript III: Molecular dynamics simulations on aqueous two-phase systems: Single PEG-molecules in solution	64
5.4	Manuscript IV: Molecular dynamics simulations on aqueous two-phase systems: Binodal curve and tie line prediction	84
5.5	Manuscript V: HTS based selection of phases for ATPS CPC of mAbs . . .	105
5.6	Manuscript IV: A Sub-Two Minutes Method for mAb-Aggregate Quantification using Parallel Interlaced Size Exclusion HPLC	126
6	Outlook	143
7	Bibliography	144
8	Curriculum vitae	158

CHAPTER 1

ABSTRACT

This doctoral thesis centers around the molecular understanding and preparative application of aqueous two-phase systems for the purification of proteins. Aqueous two-phase systems have been a topic of research for over four decades and still the mechanistic understanding of this separation process is limited. With the lack of mechanistic understanding comes the lack of mechanistic models which might facilitate process design and the application of the separation technique. This in turn leaves only the laborious way of extensive screening and empirical optimization for the development of an aqueous two-phase extraction step. In combination with limited experience, the need for capital investment and the absence of regulatory track records leaves this separation technique to stay in a niche even though a large body of scientific work has proven its capabilities and applicability. The present doctoral thesis aims at ameliorating this situation by both developing tools to generate a mechanistic understanding of the process of distribution on a molecular level and by devising screening tools and screening methodologies targeted at the preparative application of aqueous two-phase systems and reducing the experimental and material effort needed for process development.

In order to develop a tool to reach a mechanistic understanding of the separation process in aqueous two-phase systems, a molecular dynamics approach was used. In this molecular simulation technique, every atom of a system is accounted for. In terms of complexity it ranks between coarse grain simulations, in which molecules are abstracted in some way, and quantum mechanical simulations, in which every single atomic particle is taken into account. From an initial setup kinetic energy is introduced into the system to a set temperature. The attractive and repulsive forces experienced by every single atom are calculated based on the atoms' surroundings and a set of parameters called a force field. The atoms are then moved for a certain amount of time in accordance to Newtonian laws of motion. As this way of simulating a set of molecules is not a first principles method, any new simulation approach needs to be validated against experimental data. In the case of many inorganic molecules and organic solvents as well as proteins in aqueous solutions, this has already been done and the set of parameters needed for a realistic representation of these molecules are given by the appropriate force fields. The application of a simulation capable of representing both proteins and a polymer as polyethylene glycol had not yet been shown. The simulation approach employed in this thesis used the software package Yasara in combination with the force field amber03 for the representation of proteins and the AutoSMILES algorithm to generate force field parameters for PEG. Simulations of single PEG molecules in water were conducted. Structural elements and the surface properties of the resulting 3D-structures of the polymer were investigated. It was found that the surface hydrophobicity of the PEG molecules is mainly influenced by two effects. First, the dilution of the polar hydroxyl end-groups with increasing chain length renders the longer polymer chain more hydrophobic.

Second, the partial, and dynamic formation of helical regions within the polymer chain turns the oxygen atom of the ethyl groups towards the inside of the helix thus omitting it from interactions with the surrounding water. These findings provided an understanding of the change in hydrophobicity of PEG with increasing chain length on a molecular scale. The results were validated against two sets of experimental data. First, it was possible to correlate two phase formation in the presence of phosphate to the surface hydrophobicity obtained from MD simulations. Second, solvent polarity measured by the absorption maximum shift of a solvatochromic dye was also correlated to surface hydrophobicity taken from simulations. It was thus shown that the taken simulation approach gives a representation of the PEG molecules in agreement with experimental results. The main achievement of these efforts was to provide a validated simulation platform capable of simulating both PEG molecules and proteins - the basis for further work to generate mechanistic understanding of ATPS on a molecular scale.

In the following work, mixtures of PEG and phosphate were simulated both in the stable single-phase and in the unstable two-phase region. The molecular weights of the PEG used were 300, 600, 1000, and 1450 Da. This work needed extensive investment into the development of statistical evaluation routines capable of coping with the arising large datasets. Again, it was possible to correlate the results of the simulations to experimental data. Using a set of four system descriptors derived from the simulations and three experimentally determined binodals, we were able to accurately predict a fourth binodal curve solely based on the simulation data. The descriptors used in the correlations were found to cluster into four distinct sets. Among the most prominent descriptors were the PEG surface hydrophobicity, the total energy of H₂O, van der Waals energy of PEG, and total system angle energy. The prominence of these factors point to their important role in phase formation. Due to the clustering of the descriptors however, the factors determining phase formation could not be singled out exactly. In addition to the binodal data, solvatochromic characterization of the phases was also carried out and compared to the results of the simulations. It was found that the polarity of the solvent most closely correlated to the surface hydrophobicity of the PEG molecule, which confirmed the results from the previous work. H-bonding properties measured by two different solvatochromic dyes correlated well with the according H-bonding properties observed in simulation. Finally, the distribution of the model protein lysozyme was measured in 20 different systems, 5 system compositions for each PEG MW used. Using a simplified Abraham equation, the distribution coefficient of this protein could be correlated to the three phase properties either derived from dye measurements or the simulations with equal quality.

While the work conducted on the molecular scale provides the ground for a new approach to generate a mechanistic model, the application of such a model to process development still lies in the future. In order to facilitate current process development, a high throughput screening method was set up on a robotic platform. Liquid handling of highly viscous fluids and mixing of the systems on an orbital shaker where the two most challenging process-science related properties of the screening method that were optimized. With total system volumes of 650 μ L and a processing time of 2h per 96 samples this method significantly improves throughput and reduces the need for sample material. The method formed the basis for all other ATPS related studies conducted within the scope of this thesis and was successfully transferred to the sponsoring industry partner.

Having established the screening procedure, the next goal was to investigate protein descriptors that might be correlated to the protein's behavior in ATPS. Many different attempts had previously been made to improve the understanding of ATPS and decrease the amount of screening necessary by finding correlations between phase properties, solute properties, and their distribution. While some attempts were shown to yield good results

for the dataset investigated, none had yet been shown to be transferable to an industrial separation task. With both the ATPS screening method and a newly developed precipitation screening in place, a large investigation was conducted including 7 different types of ATPSs, 50 individual system compositions, the distribution of ten model proteins in these systems and the precipitation data on these proteins in the presence of ammonium sulfate. Diverse protein descriptors including size, isoelectric point, and multiple measures of hydrophobicity were investigated for correlations to the proteins' distributions. It was found that none of the previously described correlations could be verified. One major difference of this study was the total protein concentration used to generate the distribution data. A concentration of 1 g/L was chosen as the aim of the study was to find correlations at protein concentration reaching industrially interesting levels. Even though this study was unsuccessful in finding correlations between protein descriptors and distribution, it generated two important results for the overall topic of this thesis. First, it became obvious that using protein descriptors alone in correlations is no promising approach. A more promising approach would incorporate both protein and phase descriptors. With the capability of molecular dynamics simulations to generate such phase descriptors as described above, the results encourage the further application of the simulation approach. The second important result of this study was the dependency of a protein's distribution in an aqueous two-phase system on the solubility of the protein in the two phases. This becomes especially important under preparative conditions, where a high capacity of the systems is aimed for.

The findings described above led to the establishment of a new screening approach for the preparative application of aqueous two-phase extraction. Initially, binodal and tieline data are generated on the liquid handling platform. Prior to subjecting the target protein to ATPSs, the protein's solubility in the presence of the phase forming components is measured. With the binodal, tieline, and precipitation data, appropriate systems can be selected and process windows for these systems can be chosen and further investigated. Such screenings were conducted with four different monoclonal antibodies and a corresponding host cell protein pool. Based on these results, two system compositions were chosen to be scaled up from 650 μL to a multi-step extraction step using a counter current chromatography column of 500 mL volume. The process parameters flow rate and rotation speed were optimized to yield a high ratio of stationary phase inside the column. Host cell protein depletion from an antibody-HCP mixture was then investigated while varying sample volume, protein concentration, and operating conditions. The best results were found at flow rates of 5 mL/min and 2500RPM, conditions that gave the highest ratio of stationary phase. HCP clearance was found to decrease with increasing load volume. It was however not influenced by load concentration underlining the importance of a screening methodology considering target protein solubility in the phases. While this study demonstrated the applicability of centrifugal partitioning chromatography to an aqueous two-phase extraction step, it also revealed one major obstacle to its application in protein purification. The high dilution introduced into the sample over the column needs to be reduced before it can become a viable alternative. This will be the topic of further studies combining ATPS and precipitation.

CHAPTER 2

ABSTRACT IN GERMAN - ZUSAMMENFASSUNG

Die zentralen Themen dieser Dissertation sind das molekulare Verständnis sowie die präparative Anwendung von wässrigen Zweiphasensystemen (ATPS) zur Proteinaufreinigung. Wässrige Zweiphasensysteme werden seit etwa vier Jahrzehnten beforscht. Dennoch ist das mechanistische Verständnis dieses Extraktionsschrittes weiterhin stark begrenzt. Ohne mechanistisches Verständnis können jedoch keine mechanistischen Modelle aufgestellt werden, was wiederum dazu führt, dass die Prozessauslegung auch nicht mit solchen Modellen unterstützt werden kann. Hierdurch bleibt für die Prozessentwicklung nur der recht arbeitsaufwendige Weg des extensiven Screenings und der empirischen Optimierung. Die Kombination mit den geringen Erfahrungswerten bei der industriellen Anwendung der Technik, der Notwendigkeit des Kapitalinvestments, sowie mangelnder Bekanntheit bei den regulatorischen Behörden, verhindert bisher die industrielle Anwendung diesen Aufarbeitungsschrittes, obwohl reichlich akademische Arbeiten existieren, die die Anwendbarkeit und Vorteile von wässrigen Zweiphasenextraktionen aufgezeigt haben. Hier setzt diese Dissertation an, indem sowohl die Erarbeitung einer Plattform für die Generierung mechanistischen Verständnisses auf molekularer Ebene als auch die Bereitstellung von Hochdurchsatzmethoden zur schnellen und günstigen Untersuchung großer Anzahlen von Zweiphasensystemen angestrebt werden.

Um einen Ansatz zum Generieren mechanistischen Verständnisses des Separationsprozesses in wässrigen Zweiphasensystemen zu erarbeiten, wurden Moleküldynamiksimulationen (MD) verwendet. Bei dieser Simulationsart werden alle einzelnen Atome eines Gesamtsystems betrachtet. Die Komplexität solcher Simulationen liegt zwischen den sogenannten „Coarse Grain Simulations“, bei denen die simulierten Moleküle abstrahiert betrachtet werden, und den quantenmechanischen Simulationen, bei denen alle atomaren Partikel einzeln betrachtet werden. Ausgehend von einer generierten Ursprungsanordnung der Moleküle wird dem System kinetische Energie zugefügt, bis eine vorgegebene Temperatur erreicht wird. Anschließend werden für jedes Atom die auf es wirkenden anziehenden und abstoßenden Kräfte berechnet. Hierbei kommt ein „Force Field“ genannter Satz an Parametern zum Einsatz, der für jedes Atom beschreibt, welche Kraft unter welchen Umständen auf es wirkt. Nach der Berechnung der Kräfte werden für eine vorgegebene Zeit - einen Zeitschritt lang - die Atome entsprechend der Newton'schen Gesetze bewegt. Diese Schritte werden so lange wiederholt, bis die Simulation eine vorgegebene Gesamtzeit erreicht hat. Da es sich bei dieser Simulationsart nicht um eine auf thermodynamischen Grundgleichungen aufbauende „First Order“-Simulation handelt, müssen die Ergebnisse stets mit experimentellen Resultaten verifiziert werden. Dies ist für viele anorganische und organische Moleküle sowie Proteine schon ausführlich demonstriert worden und die entsprechenden Simulations-

parameter in die dazugehörigen Force-Fields eingeflossen. Simulationen, bei denen sowohl Proteine als auch Polyethylenglycol-Moleküle realistisch repräsentiert werden, waren bisher jedoch noch nicht gezeigt worden. Die Simulationen innerhalb dieser Arbeit wurden mit dem Software-Paket Yasara durchgeführt. Als Force-Field für die Beschreibung von Proteinen wurde „amber03“ verwendet, ein Force-Field, welches speziell für die Simulation von Proteinen erstellt wurde. Um die Simulationsparameter für PEG zu bestimmen ist ein veröffentlichter Algorithmus, genannt AutoSMILES, in der verwendeten Software implementiert. Einzelne PEG Moleküle ansteigender Kettenlänge, gelöst in Wasser, wurden bis zum Erreichen eines Systems im Gleichgewicht simuliert. Dies ergab Simulationszeiten von bis zu 40 ns. Strukturelle Elemente und Oberflächeneigenschaften der resultierenden 3D-Strukturen der PEG wurden anschließend untersucht. Zunächst wurde festgestellt, dass die mit der Kettenlänge ansteigende Hydrophobizität der Moleküloberfläche auf zwei Effekte zurückzuführen ist. Zum einen wird der Einfluss der recht hydrophilen Endgruppe dieser linearen Moleküle mit ansteigender Kettenlänge geringer. Zum zweiten entstehen dynamisch helikale Bereiche in den Molekülen, bei denen der Sauerstoff der Ethylgruppe jeweils nach Innen gerichtet ist und somit von Interaktionen mit dem umgebenden Wasser ausgeschlossen ist. Somit brachten die durchgeführten Simulationen ein Verständnis für die Hydrophobizität von PEG Molekülen auf molekularer Ebene. Die *in silico* Ergebnisse wurden gegen zwei Sätze experimenteller Daten validiert. Zum einen ergaben sich Korrelationen zwischen der bei Anwesenheit bestimmter Mengen Phosphat zur Zweiphasenbildung benötigten PEG-Menge und der Oberflächenhydrophobizität der entsprechenden PEG Moleküle. Zweitens konnte die Oberflächenhydrophobizität mit Lösungsmittelpolaritäten, gemessen durch die Absorptionsverschiebung eines solvatochromen Farbstoffes, korreliert werden. Hierdurch konnte folglich gezeigt werden, dass der gewählte Simulationsansatz für PEG Moleküle Ergebnisse generiert, die in Übereinstimmung mit experimentellen Ergebnissen sind. Das Simulationsprotokoll konnte somit als validiert angesehen werden und die Ergebnisse bildeten somit die Basis für alle weiteren Simulationen mit PEG einzeln oder in Lösung mit Phosphat oder Proteinen.

In den sich anschließenden Arbeiten wurden Mischungen von PEG und Phosphat sowohl im stabilen einphasigen als auch im instabilen zweiphasigen Bereich simuliert. Hierbei wurden PEGs mit Molekulargewichten von 300, 600, 1000 und 1500 Da verwendet. Diese Arbeiten erforderten eine ausführliche Entwicklung von Datenmanagementmethoden und Methoden der statistischen Auswertung der großen anfallenden Datenmengen. Wiederum konnten die Simulationsergebnisse mit experimentellen Daten korreliert werden. Unter Benutzung von vier aus den Simulationen extrahierten Deskriptoren und den experimentellen Daten zu drei Binodalen konnte eine vierte Binodale ausschließlich basierend auf Simulationsergebnissen mit hoher Genauigkeit vorhergesagt werden. Die hierbei benutzten Deskriptoren konnten in vier verschiedene Cluster einsortiert werden. Hierbei ergaben die auch in der vorangegangenen Arbeit betrachtete Oberflächenhydrophobizität des PEG, der Energiegehalt des Lösungsmittels Wasser, die van der Waals Energie des PEG, sowie die Winkelenergie des Gesamtsystems die besten Vorhersagen. Dies deutet auf die mechanistische Bedeutung dieser Faktoren hin. Aufgrund des Zusammenfallens der Deskriptoren in Cluster konnten die die Zweiphasigkeit bedingenden Faktoren jedoch nicht genau aufgelöst werden. Zusätzlich zu den Binodalendaten wurden erneut solvatochrome Charakterisierungen der Phasen experimentell durchgeführt. Hierbei wurde festgestellt, dass die in der Simulation gemessenen Oberflächenhydrophobizitäten mit den experimentell bestimmten Polaritäten der Phasen korrelierten. Dies bestätigte ebenfalls die in der vorherigen Studie gefundenen Korrelationen. Außerdem konnten die Wasserstoffbrückenbindungseigenschaften aus Simulation und Experiment miteinander korreliert werden. Abschließend wurden Verteilungskoeffizienten des Modellproteins Lysozym in 20 Zweiphasensystemen gemessen, mit je fünf

Systemzusammensetzungen pro verwendetem PEG Molekulargewicht. Unter Verwendung einer vereinfachten Form der Abraham-Gleichung, konnten die Verteilungskoeffizienten in gleicher Güte sowohl mit den experimentell als *in silico* bestimmten Phaseneigenschaften modelliert werden.

Während die vorangegangenen Arbeiten eine Basis für die Entwicklung eines mechanistischen Verständnisses der Verteilung in wässrigen Zweiphasensystemen geben, liegt deren Anwendung im industriellen Umfeld noch in der Zukunft. Um aktuelle Prozessentwicklungen zu erleichtern wurde eine Hochdurchsatzmethode zur Evaluierungen von Zweiphasenextraktionen auf einer Pipettierstation etabliert. Hierbei waren die präzise Handhabung auch hochviskoser Flüssigkeiten sowie das zuverlässige Mischen von Zweiphasensystemen unterschiedlicher Zusammensetzung die zwei größten verfahrenstechnischen Herausforderungen. Mit Gesamtsystemvolumina von $650 \mu\text{L}$ und eine Prozesszeit von zwei Stunden für 96 Systeme ergaben sich deutliche Verbesserungen des Durchsatzes und des Materialverbrauches. Die entwickelte Methode war die Grundlage aller nachfolgenden Untersuchungen zu wässrigen Zweiphasensystemen. Sie wurde zudem erfolgreich in das Labor des industriellen Sponsors dieser Arbeit übertragen.

Das Ziel der ersten sich anschließenden Studie war es, Proteindeskriptoren zu untersuchen, die Vorhersagen über das Verhalten der Proteine in wässrigen Zweiphasensystemen erlauben könnten. Eine Anzahl von Publikationen hatte zuvor solche Zusammenhänge präsentiert und so das Verständnis um die entscheidenden Faktoren erweitert. Dennoch war die erfolgreiche Anwendung dieser Zusammenhänge im industriellen Umfeld bisher nicht demonstriert worden. Mit der ATPS-Screeningmethode und einem neu entwickelten Präzipitationsscreening wurde eine umfangreiche Reevaluierung der publizierten Korrelationen zwischen Proteindeskriptoren und Proteinverteilung durchgeführt. Hierbei wurden sieben verschiedene Arten von Zweiphasensystemen, 50 unterschiedliche Systemzusammensetzungen und zehn Modellproteine verwendet. Zusätzlich wurden Ammoniumsulfat-Präzipitationen der Modellproteine durchgeführt. Verschiedene Proteindeskriptoren wie Größe, isoelektrischer Punkt, sowie verschiedene Maße der Hydrophobizität wurden auf Korrelation zur Proteinverteilung untersucht. Es ergab sich, dass keine der zuvor publizierten Korrelationen validiert werden konnte. Hierbei bestand ein Hauptunterschied zwischen den aktuellen und den publizierten Ergebnissen in den verwendeten Gesamtproteinkonzentrationen. In den aktuellen Untersuchungen wurde eine Proteinkonzentration von 1g/L verwendet, da Korrelationen gesucht wurden, die auch unter industriell relevanten Proteinkonzentrationen standhielten. Obwohl diese Arbeiten keine solchen Korrelationen finden konnten, konnten zwei wichtige Schlussfolgerungen gewonnen werden. Somit wurde nämlich deutlich, dass Korrelationen, die ausschließlich auf Proteindeskriptoren beruhen, wenig vielversprechend sind. Ansätze, die Phasen- und Proteineigenschaften miteinander kombinieren sollten jedoch erfolgversprechender sein. Dass die zuvor beschriebenen Moleküldynamiksimulationen solche Deskriptoren generieren können, unterstreicht die Validität und Relevanz der Ergebnisse der Moleküldynamiksimulationsstudien. Als zweites wichtiges Ergebnis dieser Studie kann angesehen werden, dass der Einfluss der Proteinlöslichkeit und Proteinpräzipitation auf den Verteilungskoeffizient herausgestellt wurde. Dies ist besonders im industriellen Umfeld, wo hohen Kapazitäten und Proteinkonzentrationen angestrebt werden, von Bedeutung.

Nachdem die Bedeutung der Proteinlöslichkeit für die Proteinverteilung herausgestellt worden war, wurde darauf basierend ein neuer Screeningansatz entwickelt, der auf die präparative Anwendung von wässrigen Zweiphasensystemen ausgerichtet ist. Hierbei werden die Systeme zunächst auf ihre Binodalen und Konoden untersucht. Anschließend werden Proteinpräzipitationsversuche mit den phasenbildenden Komponenten als Präzipitanten durchgeführt. Anhand dieser Daten und den Binodalen sowie Konoden können dann Systeme und Systempunkte ausgewählt werden, die eine hohe Proteinkapazität und eine vorteilhaft

te Verteilung versprechen. Entsprechende Screenings wurden unter Verwendung von vier monoklonalen Antikörpern und den dazugehörigen Wirtsorganismen-Proteinpools durchgeführt. Auf deren Ergebnissen basierend wurden zwei Systempunkte ausgewählt, die von 650 μL auf eine mehrschrittige Extraktion unter Verwendung einer Gegenstromchromatographie mit einem Säulenvolumen von 500 mL hochskaliert. Die Prozessparameter Flussrate und Rotationsgeschwindigkeit wurden auf einen hohen Rückhalt stationärer Phase hin optimiert. Die Abreicherung von Wirtsorganismenproteine wurde unter Verwendung verschiedener Zielproteinkonzentrationen, Probenvolumina, Flussraten und Rotationsgeschwindigkeiten untersucht. Hierbei wurden die besten Ergebnisse bei 5 mL/min und 2500 RPM erreicht, Bedingungen, die zum höchsten Rückhalt stationärer Phase führten. Die Wirtszellproteinabreicherung wurde mit steigendem Beladungsvolumen verringert, während sie unabhängig von der Zielproteinkonzentration in der Probe war. Dies unterstreicht die Wichtigkeit eines Screeningansatzes, der die Löslichkeit des Zielproteins in den Phasen mit einbezieht. Die Untersuchungen zeigten zum einen, dass die Gegenstromchromatographie in Verbindung mit wässrigen Zweiphasensystemen für die Proteinreinigung geeignet ist. Sie legte jedoch auch einen Schwachpunkt dieser Technik offen. Die hohe Verdünnung der Zielproteine beim Passieren der Säule muss verringert werden, soll diese Technik eine Alternative zu bestehenden Prozessschritten darstellen. Diese Schwachstelle wird Thema folgender Arbeiten sein, bei denen wässrige Zweiphasenextraktion und Präzipitation miteinander kombiniert werden.

CHAPTER 3

INTRODUCTION

In his PhD thesis in 1960 in Uppsala, Sweden, Per-Åke Albertsson discovered that PEG-phosphate aqueous two-phase systems were well suited for the purification of chloroplasts [7]. With his work he was among the initiators of the interest in purifying biological macromolecules and particles by using aqueous two-phase extraction. Zaslavsky focused his work around ATPS, extending the knowledge on this topic to a degree that his book “Aqueous two-phase partitioning” is one of the standard references for this field of science since 1995 [195]. He was among the first to extensively consider ATPS as an analytical technique.

The last two decades have seen a continuous increase in the number of scientific articles concerned with aqueous two-phase systems (ATPS) published per year. Figure 3.1 shows the number of results of a search on the topic ATPS performed on the database of the ISI Web of KnowledgeSM [176]. While the overall increase in scientific output and the growth of the scientific community worldwide has to be taken into account, an increased interest in the topic of Aqueous two-phase systems can be concluded from the data shown in figure 3.1. This is mirrored by the perceived increased interest in alternative protein purification techniques such as ATPS-extraction or precipitation expressed by members of the pharmaceutical industry both on conferences and within the scope of collaborations.

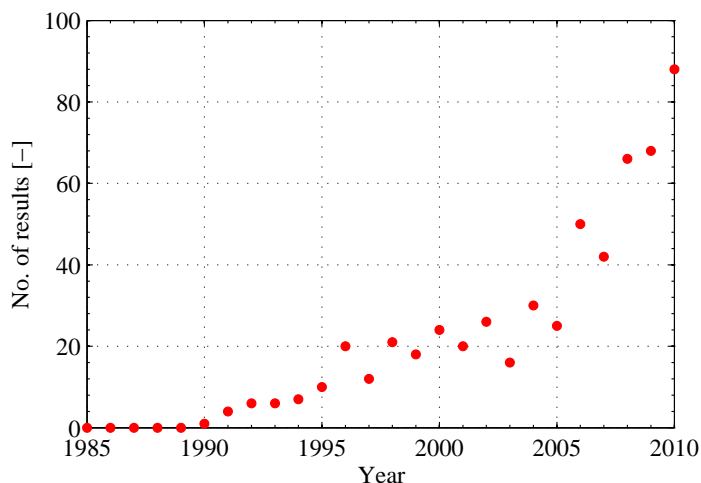


Figure 3.1: Number of database entries of the ISI Web of KnowledgeSM [176] on the topic “ATPS” plotted over the year of publication for the last two decades.

Liquid two-phase separations were among the first separation techniques used in the purification of drug molecules. Organic solvents tend to denature proteins and are thus not applicable to protein separation. In contrast, the phases in aqueous two-phase systems

3.1 Purification of high value proteins in the pharmaceutical industry

are each composed of >70% water and thereby have a better biocompatibility compared to organic extractions. Aqueous two-phase systems are currently considered as an alternative separation step to the most commonly used chromatographic separations in the downstream processing of highly valuable proteins. While capacity and waste disposal issues might arise with the use of this technique, its advantages over chromatography are the low costs of goods as compared to chromatographic resins and its ease of upscale.

In the following section, the purification of high value proteins will be discussed in the case of monoclonal antibodies. The subsequent sections then present the development of batch aqueous two-phase extraction steps and review current literature on the application ATPS for the purification of monoclonal antibodies. Following this section, past and current attempts to model and predict protein behavior in ATPSs are presented. The introduction ends with a discussion on the theory of liquid-liquid chromatography and its application to the purification of proteins.

3.1 Purification of high value proteins in the pharmaceutical industry

OKT3, an anti-CD3 antibody was introduced to the US market in 1986, thus making it the first molecule of that class to be approved as a pharmaceutical ever [54, chap. 9]. While OKT3 was produced in the ascites of mice, the majority of antibodies since then have been produced in genetically modified mammalian cell lines such as chinese hamster ovary (CHO) mouse myeloma (NS0), baby hamster kidney (BHK) cell lines [192]. Mammalian cell lines, in contrast to procaryote or yeast fermentation, produce correctly glycosylated proteins. Non-human glycosylation pattern would elicit an immune response thus diminishing drug efficacy. In order to further reduce immunological response to the drug substance, mAbs were first "humanized" by replacing the murine scaffold sequence with the human sequence. Later, mice were designed to carry a human immune system thus yielding actual human antibody sequences. While the initial cell lines yielded final product concentrations well below 1 g/L, volumetric productivities have increase 20-fold over the past 25 years [74, 192]. Downstream processing of these molecules has also seen changes of similar magnitude. The large molecular similarity of all antibodies encouraged the standardization of the downstream process layout for the entire molecule class. A so called platform process is now run by the vast majority of companies producing mAbs with the processes differing only in detail from company to company. In general, an affinity chromatography step using a ProteinA resin is used as first step after clarification of the cell culture supernatant. After this initial step, the target molecule is highly pure and highly concentrated. Host cell protein levels are reduced by a factor of 1000 and product concentration increased by a factor greater 10. Owing to the low pH elution step in ProteinA chromatography, a first viral inactivation step by low pH hold is readily incorporated into the process at that stage. This step is followed by two chromatography steps, usually either two ion exchange steps with one run in flow-through mode, or an ion exchange step followed by an hydrophobic interaction chromatography. These steps reduce DNA, host cell protein and aggregate levels to acceptable levels. After viral filtration, the processed product reaches the formulation. The standardization of the mAb production process has drastically reduced process development efforts. If it is validated that a molecule fits into the platform process the process development is reduced to optimization and the validation demanded by the regulatory authorities [159]. With this level of sophistication and standardization reached in the production process, what could motivate a large change in the process design such as the integration of a non-chromatographic unit operation? The only reasons to bring a major change into the process would be economic ones. Upstream process development continues

3.2 Automation and miniaturization in pharmaceutical downstream process development

to improve volumetric productivities which are expected to supersede 10 g/L. While this increases the overall output of the fermentation and reduces its capacity needs, it puts a burden on the downstream process. While increased product titers do not pose a problem in fermentation, chromatography unit operations are limited by the mass of a target molecule. In order to handle the increased amount of product coming into downstream processing in a timely fashion column sizes need to be increased. However, there is a limit in how large a chromatography column can be constructed. Thus, additional columns need to be installed increasing downstream process costs. As Low *et. al* [114] argue, alternative step such as precipitation, crystallization or extraction may offer economies of scale and might thus be considered even though they would increase process development time.

Whether there actually is an economic driver towards a change in the mAb production process remains a topic of debate. While the current market situation is in favor of keeping the process as it is, changes in demand or political changes to the health care systems might shift the economic incentives. However, with the abundance of material available, mAb/HCP solutions can be regarded as a both highly valuable and highly available model system for a separation task faced by the pharmaceutical industry. New technologies can be put to the test on these feedstocks and their use can be demonstrated and judged in reference to current process data. New molecule classes or the development of personalized medicine might make changes to the current downstream processes inevitable. Alternative purification techniques could be a valuable addition to the process developers' toolbox. As stated above, one such alternative protein purification technique is aqueous two-phase extraction, which is the central topic of this thesis.

3.2 Automation and miniaturization in pharmaceutical downstream process development

Traditionally, protein purification processes were optimized using bench-top sized equipment and a one-factor-at-a-time approach. The number of experiments conducted was severely limited due to both time and material restraints resulting from the relatively large scale and labor intensity of the experiments. In order to improve throughput, and decrease time, labor, and material costs, the unit operations used in protein purification were miniaturized, parallelized, and automated. Such high throughput experimentation ("HTE") approaches are now a common feature of pharmaceutical downstream process development [33]. The increase in throughput allows for the implementation of high throughput process development ("HTPD") schemes. Such schemes most often rely heavily on statistical design of experiments and can yield, in addition to a process optimization, valuable insight into the factors influencing the process. Based on these developments, regulatory agencies have recently introduced the Quality by Design ("QbD") initiative. In applying QbD, production process knowledge is used in such a way that it allows to evaluate the influence on the product quality of a deviation in the production process [142]. Ultimately, production process control could be implemented based on HTE results and mathematic modeling of the results. Obviously, deep process understanding mandates a large amount of experiments, proper statistical evaluation of the results, and accurate analytics.

While most unit operations have already been adapted to a HTE approach, the progression towards QbD is often impeded by a lack of sample throughput in the analytics. As an example, the quantification of aggregate level in antibody solutions is usually done using size exclusion chromatography. Sample throughput of the standard methods used is between two and four samples per hour. With an HTE method generating at least 96 samples within two hours, it is clear, that not all samples can possibly be analyzed. This particular problem was addressed in this dissertation by established a size exclusion chromatography

aggregate assay with a throughput of 30 samples per hour.

3.3 Aqueous two-phase extraction - Applications in batch mode

Aqueous two-phase system partitioning is the topic of a large body of scientific literature. Various biological materials such as cells [96, 125, 183], cell organelles or membranes [7, 92], and nucleic acids [65, 144] have been successfully purified using ATPS. The most often investigated class of biological molecules is however proteins. Extensive research both to the preparative and analytical application of ATPS for proteins has been conducted. Several conclusive reviews articles have been published [79, 81]. The feasibility of using ATPS for the purification of proteins up to a very large scale ($10^5 L$) was demonstrated by Raghavarao *et. al* [141].

For the development of an aqueous two-phase extraction step, numerous factors have to be taken into consideration. Apart from special cases such as thermo-separating or pH sensitive systems, there is the initial choice between polymer-polymer or polymer-salt systems. Polymer-polymer systems might be favored if large quantities of salt containing waste should be avoided. Additionally, distribution in polymer-polymer systems can be more strongly influenced by the addition of smaller concentrations of additives, than in polymer-salt systems. However, material prices might be higher and further processing of the phases might be more difficult due to high viscosity. Polymer-salt systems have the advantage of mostly lower material prices, waste treatment might, however, be in turn more expensive. Processing might be easier due to lower viscosity, but protein solubility in the presence of large amounts of salt can be reduced. Systems with large amount of buffering salts can also not as easily be evaluate at different pH values as polymer-polymer systems. The next choice to make would be both polymer and salt type as well as polymer chain length. The most commonly used phase forming components are polyethylene glycol in combination with either phosphate or dextran. With the additional complexity of choosing polymer chain length, pH, additives, and protein load, it is obvious that a large number of experiments is needed for the development of an aqueous two-phase extraction step.

While the application of ATPS for bulk biologicals such as enzymes has already been extensively discussed, its use in the production of high value pharmacological proteins has only been considered more recently. Again, monoclonal antibodies as the most prominent molecule class has mostly been the focus of such work and was also used within this thesis. Azevedo *et. al* [16] demonstrated the application of a PEG6000-phosphate systems to the separation of a polyclonal antibody from model proteins and a hybridoma cell culture supernatant. It was found that the antibody distributed into the lower phase when no NaCl was added to the systems. The addition of NaCl at concentrations up to 15% [w/w] pushed the antibody into the upper phase and increased purity. The limiting factor for the application of the system was the low achievable antibody load. Total system concentrations as low as 0.5 mg/mL lead to a decrease in recovery most likely caused by precipitation. The same research group went on to investigate PEG3350-phosphate systems using statistical design of experiments [148]. As in the previous study, it was demonstrated that addition of NaCl could shift the distribution of the target protein into the upper phase and increase purity by doing so. It was also demonstrated, that a subsequent back extraction using systems of low salt content was feasible. The limiting factor was, again, the loss in protein recovery caused by a low solubility. This factor has until now been neglected in the selection of an ATPS extraction step. Research has been focused on influencing protein partitioning and identifying conditions giving a high selectivity for the target protein. The limited protein solubility and the resulting low system capacities need however be addressed in order to yield a holistic understanding of the industrial capabilities of ATPSs.

3.4 Modeling protein partitioning in ATPS

A single unit operation in downstream processing cannot yield the purity necessary in the production of biopharmaceuticals. The compatibility of any unit operation with its preceding or subsequent unit operations needs to be considered. The integration of an aqueous-aqueous extraction step into a complete process can be a challenge. Both phases contain a large quantity of solutes that can interfere with subsequent process steps or create the need for extensive conditioning steps. The combinations of an ATPS step with both hydrophobic and cation exchange chromatography has been successfully demonstrated for the case of antibody purification [17, 18]. Thus, while both the implementation of a ATPS purification step for a pharmaceutical protein and its integration into a multi-step process have been successfully demonstrated, two main obstacles still impede the industrial application of this technique. First, capacities of the systems used in the published scientific literature were too low to be industrially applicable. Additionally, the development of an ATPS step currently relies heavily on prior knowledge and the experimenter's experience. Models reliably predicting protein distribution in ATPS are missing, although there have been numerous attempts to established such models as described in the following section.

3.4 Modeling protein partitioning in ATPS

Since the underlying physicochemical forces driving protein partitioning in aqueous two-phase systems are not yet fully understood, numerous attempts have been made to find correlations using protein or system descriptors to explain and predict protein behavior in ATPSs. The systematic search for correlations between characteristics of suspended materials and their distribution began in the early 1970s. Charge of the suspended material was identified as a factor influencing partitioning in polymer-polymer systems both for human cells [36] and proteins [184]. The molecular weight of proteins, and thus the physical size of the molecule was found to correlate with the partition coefficient for certain groups of proteins at their isoelectric point [153]. The most promising models were found when correlating protein partitioning to measures of the hydrophobicity of the protein. Berggren *et al.* [26] describe a correlation between hydrophobicities calculated based on the amino acid sequence of the proteins and their distribution. The hydrophobicities of the individual amino acids were determined by measuring the distribution coefficients of peptides composed of an increasing number of the particular amino acid investigated. In a similar study, Salgado *et al.* [152] considered measures of hydrophobicity based on the three dimensional structure of the proteins. The correlations were best when only the primary sequence was used. This is somewhat surprising and might point to the important role transient partial unfolding of the proteins caused by the high concentrations of phase forming components might play. An attempt to incorporate such effects was made by Asenjo *et al.* [9, 13]. In their studies, the amount of ammonium sulfate needed to start precipitation of a fixed concentration of protein was used as a measure of protein hydrophobicity. Within the present dissertation, these correlations were reevaluated using a much larger set of conditions.

While the correlations described above gave valuable insights into the forces at play, it needs to be stressed that until now, these correlations are limited to the proteins and systems used in the respective study. While they provide general trends that an experimenter can use to guide process development, there are currently no universally applicable models that could predict protein behavior in either polymer-polymer or polymer-salt systems. In order to devise a system that could predict protein partitioning in ATPS, a combination of protein descriptors with characteristics of the phases might be promising. One such approach was presented by Madeira *et. al* [116]. The distribution of 13 model proteins in 10 different polymer-polymer systems was correlated using solvatochromic solvent parameters and the previously determined susceptibility of the model proteins to these solvent

3.5 Liquid-Liquid chromatography

parameters. Using solvatochromic dyes, both phases of ATPSs were characterized for their dipolarity/polarizability (π^*), their H-Bond basicity (β), and H-Bond acidity (α) as described by Kamlet and Taft [94, 95, 171]. Model protein distribution was then fitted to the difference in solvent parameters between the two phases using the Abraham model [1]. The resulting linear correlation for all proteins and systems between measured and calculated distribution had an R^2 value of 0.997.

While this approach is promising, it is not possible to generate mechanistic understanding of the processes on a molecular scale. How the solvent polarity and H-bonding properties influence a protein's distribution remains unclear. In order to fully understand the processes on a molecular scale, an appropriate molecular representation of an aqueous two-phase system containing a protein would be needed. One approach to devise such a representation is through the use of molecular dynamics simulations. In such simulations, every atom and all its bonds of a system are taken into account. The forces acting on each individual atom through its vicinity are calculated in each frame of a simulation. Between two frames, the forces act on the atoms according to Newtonian law of motion for a set time period. This process is repeated until the system reaches an equilibrium. Molecular dynamics were introduced to the downstream processing by Dismer *et al.* [50]. In their work, molecular dynamics simulations were used to gauge the electrostatic interaction potentials between model proteins and a cation exchangers surface as a function of the protein's orientation. It was thus possible not only to accurately reproduce the protein's orientation that had previously been determined experimentally by Dismer *et al.* [51]. It was also possible to predict the elution behavior of a model protein in a salt gradient elution from a cation exchanger based on simulation data and a calibration using a different model protein. Thus molecular dynamics simulations have proven a valuable tool to generate understanding of protein purification phenomena on a molecular level. In this dissertation, molecular dynamics simulations were evaluated as a tool to study PEG molecules in solution, aqueous two-phase systems, and protein partitioning therein.

3.5 Liquid-Liquid chromatography

The need for continuous extraction

While the purification of biologicals in batch ATPS was shown to yield good results in some cases, a single extraction step is not sufficient for many purification tasks. In order to improve the resolution between the molecules of interest and the impurities a series of recurring extractions can be employed. In 1950 Craig [45] pioneered a countercurrent distribution ("CCD") device now dubbed the Craig apparatus. It consists of a series of glass devices connected in series in such a way that rocking the device back and forth results in a flow of the lighter phase towards the end of the apparatus. The heavier phase stays stationary in this process. Each cycle consists of mixing the two phases, letting them settle and then transferring the lighter phase into the next tube. Substances distributing into the upper phase move faster through the apparatus while substances distributing into the heavier stationary phase are held back. This CCD mechanism has since been replaced by more sophisticated technological implementations, but the basic principle of using liquid-liquid extraction in a chromatography like mode is still the same. A review of the different ways to implement liquid-liquid chromatography can be found in the publications by Grushka and Grinberg [69, Chapter 9] and Sutherland [167]. The following two paragraph briefly summaries the different concepts.

Hydrodynamic liquid-liquid chromatography

There are two ways of implementing liquid-liquid chromatography: hydrodynamic and hydrostatic columns. the hydrodynamic countercurrent chromatography (“CCC”) was pioneered by Ito *et. al* [88]. Of the many possible implementations of this technology the J-type CCC is among the most common. A tube, helically coiled around a drum, is rotated in a planetary motion. Owing to the Archimedean screw force and the centrifugal force [71, 86], each coil consists of a zone of mixing and a zone of demixing when a mobile phase is pumped through the column. If the rotation is stopped, mobile and stationary phase settle owing to the force of gravity. The settling of the phases when the rotation is stopped is the reason why this implementation of liquid-liquid chromatography is termed hydrodynamic.

Hydrostatic liquid-liquid chromatography

The hydrostatic centrifugal partitioning chromatography (“CPC”) was pioneered by Murayama *et. al* [126]. A CPC rotor consists of a series of chambers connected by capillaries in such a way that the top of one chamber is connected to the bottom of the next. The chambers are mounted on disks and the disks stacked to create the actual rotor sometimes containing over 1000 separate chambers. The rotor is spun creating a centrifugal force whose direction is along the chambers. Owing to the repetition of capillaries and chambers, the heavier phase will not settle at the bottom of the rotor when the rotor is stopped, which is why this implementation of liquid-liquid chromatography is called hydrostatic. As explained by Sutherland [167] major improvement of the chamber design were achieved by Marchal *et. al* [118]. The improvements were based on work done by van Buell *et. al* [178] who used stroboscopic measurements to improve the understanding of the effects at play in CPC. Ikehata *et. al* [84] recognized the important role of the Coriolis force for mixing in CPC.

Modeling CCC/CPC

In 2009 de Folter *et. al* [47, 60] published a software modeling CCC/CPC. The model assumes that the basic separation principle can be abstracted from the CCD apparatus. Based on the works by [31, 43, 169] they come to the following conclusions. The retention time (t_R) of a solute in CCC/CPC can be calculated based on the volumes of mobile (V_M) and stationary phase (V_S), the distribution coefficient D and the flow-rates of mobile (F_M) and stationary phase (F_S) as follows:

$$t_R = \frac{V_M + DV_S}{F_M + DF_S} \quad (3.1)$$

With the distribution coefficient D defined as the ratio of the concentration of a solute in the stationary phase (C_S) to its concentration in the mobile phase (C_M).

$$D = \frac{C_S}{C_M} \quad (3.2)$$

Sutherland *et al.* found the peak width w to obey the following equation:

$$w = 4\sqrt{t_R D \frac{V_M}{V_S}} \quad (3.3)$$

3.5 Liquid-Liquid chromatography

It can be seen from equation 3.3 that the peak width and thus the dilution of a solute in CPC increases with increasing retention time, increasing distribution into the stationary phase and increasing ratio of mobile phase in the column. To avoid dilution of a certain solute, the CCC/CPC should thus be operated using a system that gives a low distribution coefficient for the solute and a high ratio of stationary phase in the column. Care must however be taken not to reduce the residence time to a degree that sacrifices too much resolution. The general equation for the resolution between two peaks in chromatography, assuming Gaussian peak shapes, as a function of their retention times and the peak width is:

$$R_s = \frac{t_R^1 - t_R^2}{0.5 * (w_1 + w_2)} \quad (3.4)$$

For each set of peak widths a difference in retention time can be found to yield a certain resolution. Figuratively speaking, wide peaks need to be further apart from one another than narrow peaks to generate the same resolution. The resolution of two adjacent peaks depends on the selectivity α of a column, the number of theoretical plates N , the partitioning coefficient D , and the ratio of mobile to stationary phase X .

$$R_S = \frac{\sqrt{N}}{4} \times (\alpha - 1) \times \frac{k}{1 + k} \quad (3.5)$$

The retention factor k is defined by using the retention volume V_R of a solute, the volume of mobile (V_M) and stationary phase (V_S) in the column:

$$k = \frac{V_R^1 - V_M}{V_M} \quad (3.6)$$

In CCC/CPC the use of a partition coefficient K_D is more common. It is defined as:

$$K_D = \frac{V_R^1 - V_M}{V_S} \quad (3.7)$$

Thus, k can be substituted by:

$$k = K_D \times \frac{V_S}{V_M} \quad (3.8)$$

Using the above equations, equation 3.5 can be rewritten to contain the ratio of mobile to stationary phase:

$$R_S = \frac{\sqrt{N}}{4} \times (\alpha - 1) \times \frac{K_D}{K_D + \frac{V_M}{V_S}} \quad (3.9)$$

Using equation 3.9 Berthod [69, chapter 9] demonstrated the complementarity of CCC/CPC and HPLC with an example much like the following. Given are two solutes of low K_D -value ($K_D^1 = 1$, $K_D^2 = 2$). Compared to CCC/CPC HPLC columns usually have a much lower ratio of V_S to V_M . A stationary to mobile phase ratio of 0.02 would result in retention factors of $k'^1 = 0.02$ and $k'^2 = 0.04$. The two solutes would thus hardly be retained in the column. In order to generate baseline resolution ($R_S = 1.5$) according to equation 3.5 approximately 93600 theoretical plates would be needed. Such high numbers of theoretical plates are achievable with current HPLC / U(H)PLC equipment. They lie, however, far outside the numbers of theoretical plates reported for CCC/CPC which are in the region of 80-300. CCC/CPC equipment can still generate the same resolution. This is due to the much large ratio of V_S to V_M which are commonly between 0.6 and 9. Assuming an equal volume of the two phases within the column, using equation 3.7 retention factors of 1 and 2 can be calculated. Using equation 3.9 to calculate the number of theoretical plates needed

3.5 Liquid-Liquid chromatography

to generate a resolution of 1.5 we get $N = 144$. The low number of theoretical plates in CCC/CPC is counteracted by the much larger ratio of stationary to mobile phase volume as compared to HPLC. This leads to a much increased retention factor in CCC/CPC. With longer retention in the column comes dilution which means that an HPLC would generate two narrow peaks eluting very closely, while the CCC/CPC would generate much wider peaks with a much larger difference in retention times. It should be noted that the above example is a thought experiment only. HPLC operating conditions leading to retention factors of 0.02 and 0.04 would most likely not be chosen. Mobile phase composition would be chosen to yield increased retention factors and reduced need for such high numbers of theoretical plates. Also, this assumes isocratic conditions for the HPLC which is highly unusual except for size exclusion chromatography.

The severe peak broadening and thus dilution of the compounds in CPC is a consequence of the long retention times. This leads to the following conclusions: a high ratio of stationary phase is beneficial both to the resolution and to the sample dilution. Lower mobile phase volume and higher distribution coefficient lead to higher retention times and increasing peak width. In order to avoid dilution of the sample, a system should be chosen that gives a high distribution coefficient D and a high ratio of stationary to mobile phase in the column.

Application of CCC/CPC to ATPS and biologicals

In contrast to organic-organic or aqueous-organic biphasic systems, CCC and CPC stationary phase retention and adequate mixing are harder to achieve when switching to aqueous-aqueous systems. Compared to organic systems, ATPSs most often display lower differences in density of the two phases, low interfacial tension, and higher viscosity. This results in higher settling times and less mixing, leading to higher stripping of the stationary phase during CCC/CPC and a lower efficiency in terms of number of theoretical plates. High retention of stationary phase is, as shown above, desired. For ATPS in combination with J-type CCC, it has been shown in theory [85] and in practice [70] that achieving a high level stationary phase retention is mostly only possible at low flow-rates. Low flow-rates lead to a decrease in productivity. It is clear that the design of the current CCC/CPC apparatus needs to be adjusted for the use of ATPS, and research geared towards this is currently performed. Despite the pitfalls of the currently available CCC/CPC machines, both hydrodynamic and hydrostatic columns have been investigated for the purification of biological using ATPS. Guan *et. al* [72] used a J-type CCC in combination with a 12.5% K_2HPO_4 12.5% PEG1000 ATPS. They investigated the influence of system setup, flow rate, column length, and sample loading on the separation of myoglobin from lysozyme and of the dipeptides His-Gly from Val-Tyr. Baseline separation of the two proteins was only achieved at low flow rates of 0.62 mL/min (= 173 min/CV). They found that the separation of the dipeptides was better even though single component distribution coefficients would have suggested otherwise (α was larger for the proteins than for the dipeptides). Based on the results presented by Ito *et. al* [87], the authors conclude that J-type CCC machines can provide adequate mixing for ATPS, the molecular weight of proteins however leads to severe band broadening thus worsening resolution. CPC has been used for the determination of water-octanol distribution coefficients and for the purification of chemical compound by organic-organic extraction. The use of CPC for protein purification is less common. Bérot *et. al* [27] demonstrated the use of a CPC for the preparative purification of pea albumin using a butanol water system. As stated above, using ATPSs in CPC poses several challenges but has been successfully demonstrated. Sutherland *et. al* [168] used a 12.5% K_2HPO_4 12.5% PEG1000 ATPS in an Armen instruments 500 mL CPC system investigating the influence of flow-rate, rotation speed, and sample load volume on the resolution of myoglobin from lysozyme. These proteins had distribution coefficients of 0.59 and 1.91 in batch

3.6 Scope of this dissertation

ATPS. In CPC, they achieved resolutions of up to 3.2. Maximizing the load volume, they found that near baseline separation ($R_S = 1.28$) was achievable when loading an ATPS of 10% column volume with concentrations of 2.2 mg/mL for both proteins. While rotation speed had little influence on the resolution, an increase in flow rate significantly reduced resolution. The separation was run at 10 mL/min (50 min/CV). A flow rate of 20 mL/min (=25 min/CV) resulted in a complete stripping of the column. Load volumes higher than 10% had the same effect, as the proteins were loaded in an ATPS rather than in mobile phase. The authors were able to demonstrate that a simple linear scale-up approach to an Armen instruments 6.25L CPC system yielded results in excellent agreement with the smaller scale. However, even at the highest possible load volume, the proteins were diluted more than threefold. Chen *et. al* [37] demonstrated the use of a PEG6000-phosphate ATPS for the separation of BSA and lysozyme using a 230 mL CPC. A review of the published work on the purification of biologicals using liquid-liquid chromatography was written by Ian Sutherland [167]. As the same ATPS (12.5% K_2HPO_4 12.5% PEG1000) and the same model proteins (myoglobin and lysozyme) are used in most studies, the aforementioned review article gives a good overview of the current capabilities of the technology. It can be concluded that at the current state of development, CPC is the most promising option of liquid-liquid chromatography using ATPS for the separation of proteins as CPC was shown to work at higher flow rates, higher sample load volumes, greater system volume, generating better resolution.

One major obstacle for the implementation of ATPS CCC/CPC is the lack of a solvent selection system. The selection of a two-phase system for organic-organic or aqueous-organic extraction has been significantly simplified by the development of the "Arizona" solvent system [32, 61]. Twenty four different combinations of Heptane, Ethyl Acetate, Methanol, and water were chosen such that a wide range in polarity of the resulting two phases was covered. The polarity of the phases was judged by the bathochromic shift of the absorbance maximum of the Reichhardt's dye [143]. With this systematic approach the selection of an extraction system is greatly simplified as one can infer from the results of a distribution experiments which solvents might be better suitable for the separation task. For ATPS however, no such approach has yet been developed. In fact, several publications both on CCC and CPC used the exact same ATPS (12.5% K_2HPO_4 12.5% PEG1000) investigating the resolution of myoglobin from lysozyme. While this is beneficial in terms of comparability between the studies, it also underlines that a practical approach for selecting the biphasic system to use is lacking.

3.6 Scope of this dissertation

In order to allow for a more systematic approach in the selection of ATPS suitable for scale-up to CPC two measures are presented in this thesis. First, an automated screening method, generating binodal data, tieline data, protein solubility and protein distribution data is presented in section 5.1. Second, in section 5.3, results from MD simulation studies on PEG solutions and in section 5.4 from PEG-phosphate solutions are shown. The results give insights into the structure and surface properties of PEG molecules in solutions. The results were correlated with solvent polarity measurements. Thus these studies might lay the groundwork for the development of a scale of solution properties for PEG-phosphate systems similar to the "Arizona" system used for organic extractions. The HTE method developed to screen aqueous two-phase system was then used to investigate correlations between protein descriptors and protein distribution. These efforts are summarized in section 5.2 While CPC had been shown to work for the separation of two model proteins using ATPS an evaluation of this technology for a complex purification task such as the

3.6 Scope of this dissertation

separation of a biopharmaceutical target molecule from host cell proteins was missing. This is addressed in section 5.5 of this thesis. The selection of an ATPS suitable for CPC was again performed using the HTE method developed.

CHAPTER 4

RESEARCH PROPOSAL

Aqueous two-phase extraction has recently attracted increased interest both from the scientific community and from the biotech industry as one of the alternatives to chromatography. For protein purification, ATPS have been shown to yield good results for various purification tasks. Its merits include integration of solid removal and product recovery, easy upscale, and high selectivity and recovery. ATPSs have been shown to be a valuable alternative to conventional chromatography in certain cases both judging from product recovery and purity as well as from process economics. Regardless of the many proposed benefits, there is still only limited industrial scale application of this technique. This might be ascribed to the following facts. Despite substantial academic efforts, there is only limited understanding of the underlying principles, resulting in an absence of detailed physical models. This makes it hard to select - from the wide range and diversity available - a set of suitable systems to be tested when faced with a certain purification task. Manual screening on the other side is a time, labor and material intensive process. Once a suitable system is identified, current industrial processes are based on batch processing only, neglecting the potential of multi-stage processes. There are two possible approaches to making this technique more feasible for the biopharmaceutical industry. Find out the principles at work and decrease the screening costs. This research proposal aims to work on both fronts, understanding the underlying principles of ATPS and selecting suitable systems as well as an feasibility study for multi-stage ATPS processing.

Several attempts to model protein distribution either on phase characteristics or protein properties have been published. However, a verification of these correlations against an independent data set is missing. The application of such correlations to an industrial separation task has also not yet been demonstrated. Finding correlations between system or solute descriptors and their distribution is one way to enhance understanding of the systems. Another successful way of gaining detailed knowledge of the forces at play are molecular dynamics simulations. They have been shown to accurately predict the elution behavior of two model proteins from an cation exchange resins based on the 3D-structure and molecular dynamics simulations. To be able to use this approach to study protein behavior in ATPS however, the basic description of polymer and polymer/salt solutions via molecular modeling first needs to be demonstrated. The capability to describe polymer and polymer/salt solutions without proteins is thus a necessary prerequisite for studying proteins in ATPS via MD simulations.

Monoclonal antibody separation from host cell proteins is probably the most commonly investigated purification task in the context of biopharmaceutical downstream processing and a well suited model system. ATPS have been shown to be applicable for this task. However, lack of resolution and low system capacity were seen. Low system capacities need to be addressed by screening more ATPSs, specifically ATPS composed of lower molecular

weight polymers that promote higher capacities. Increased resolution, while simultaneously scaling up the extraction step, can be achieved by the use of centrifugal partitioning chromatography (CPC). This unit operation has been demonstrated to work with aqueous two-phase systems and to generate high resolution for the separation of proteins. Severe dilution of the target molecules were encountered and need to be prevented when working with pharmaceutical proteins.

With reliable models for the distribution of proteins in ATPSs lacking, fast and resource efficient screening methods constitute an important part in the development of this process step. Miniaturization and parallelization of process step developments, recently termed high-throughput process development “HTPD”, has advanced greatly in recent years and can now be considered the state-of-the-art approach for the purification process for high value protein pharmaceuticals. The most well characterized process step in this format is chromatography both in batch and in column format. While numerous other process steps, including ATPSs, have been shown to be feasible for HTPD, deeper understanding of the challenges of scale-down and their applicability to pharmaceutical process development still need to be investigated.

This research proposal has two closely associated foci relating to the statements above. As first focus, the proposed research will evaluate the capabilities of a published method of predicting protein distribution in ATPS based on protein precipitation data. Using the developed screening platform, the number of model proteins and ATPSs previously used will be expanded. If a correlation is found, it is to be used in the screenings of appropriate phases for liquid-liquid chromatography. If the correlations are not found to hold in light of a larger dataset, a new approach to predicting protein distribution in ATPS will be attempted. Molecular dynamics simulations shall be used in an attempt to characterize PEG solutions and ATPS based on PEG and phosphate *in silico*. To do so, first a simulation protocol for polymer solutions needs to be established. Then, the properties of polymer solutions *in silico* need to be correlated with experimental values. Before protein distribution in ATPS can be considered, it needs to be demonstrated, that the *in silico* representation of an ATPS without protein can be correlated to experimental values. It needs to be shown, that the representation of PEG/phosphate solutions in the molecular modeling is accurate enough to yield differences in the characteristics of single and biphasic systems, and between the two phases of an ATPS. This will lay the basis on which protein distribution might be evaluated based on molecular dynamics simulations. Creating this basis for further research is the final goal of the second focus of the proposed research.

As a second focus, the industrial application of an ATPS extraction step is to be evaluated. Purification of monoclonal antibodies from host cell proteins is to be used as model system. As explained above, the lack of an adequate model to predict protein distribution in ATPSs creates the need for a sophisticated, miniaturized, automated screening platform. This platform needs to be capable of generating in an automated fashion binodal data, tieline data, and data on protein distribution and recovery. Additionally, to allow the estimation of protein solubility limited load capacities of the systems, an automated screening method for protein precipitation by any chosen precipitant needs to be developed. HTS methods currently lack high informational content, high accuracy analytics matching the throughput of the experiments. Thus one currently has to choose between fast, low content assays and time consuming high content assays. In order to supplement the HTS methodology, ways to improve throughput in chromatographic assays run on state-of-the-art U(H)PLC equipment will be evaluated within this project. Finally, a CPC apparatus will be used in order to evaluate scalability of batch experiments to liquid-liquid chromatography with the aim of increasing resolution while keeping sample dilution to a minimum. The influence of operation parameters and choice of biphasic system on the performance of a CPC needs to

4 Research proposal

be evaluated. A screening scheme fitting the needs of liquid-liquid chromatography phase selection will be developed. As the final goal of this focus of the proposed research is the demonstration of the use of a CPC machine beginning with the selection of the phases by HTS methods and the successful transfer of the developed methodologies to the sponsoring industry partner.

In summary, this projects aims to deepen the basic understanding of aqueous two-phase systems for the purification of proteins and generate an easily applicable process development platform ranging from early stage screenings to pilot scale preparative purification.

CHAPTER 5

PUBLISHED SCIENTIFIC MATERIAL AND MANUSCRIPTS

1. ***Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation***

STEFAN OELMEIER, FLORIAN DISMER, JÜRGEN HUBBUCH

This paper describes the establishment of the ATPS screening method on a liquid handling platform. It includes a description of the liquid handling parameters needed for accurate handling of viscous fluids and a discussion on the factors influencing mixing of fluids in small cavities. It demonstrates the application of the screening method to the separation of two monoclonal antibodies from their corresponding host cell proteins by extraction in PEG4000-phosphate systems with varying system composition, pH, and NaCl addition.

*published in: **Biotechnology and Bioengineering.** 2011;108(1):69-81*

2. ***Evaluation of correlations of protein properties and ATPS distribution***

STEFAN OELMEIER, JOHANNES KNOLL, JÜRGEN HUBBUCH

In this manuscript the ATPS screening method is applied to a set of 50 different ATP-system compositions and ten model proteins. It investigates correlations between protein descriptors derived either from 3D structures or from precipitation experiments and the proteins' distributions in the ATPSs. It was found that protein solubility is a major factor influencing protein partitioning under preparative conditions.

in preparation

3. ***Molecular dynamics simulations on aqueous two-phase systems - Part I: Single PEG-molecules in solution***

STEFAN OELMEIER, FLORIAN DISMER, JÜRGEN HUBBUCH

In this manuscript, molecular dynamics simulations of single PEG molecules with increasing molecular weight in water are shown. The simulations are evaluated for structural elements of the PEG, PEG surface hydrophobicity, and PEG-water interactions. Phase formation in the presence of PO_4 and solvent polarity measurements using a solvatochromic dye are used to validate the simulation protocol against experimental data.

*published in revised form in: **BMC Biophysics** 2012 Aug 8;5(1):14*

4. ***Molecular dynamics simulations on aqueous two-phase systems: Phase Formation and Protein Partitioning***

FLORIAN DISMER, STEFAN OELMEIER, JÜRGEN HUBBUCH

Based on the validated MD simulation protocol for PEG molecules, mixtures of PEG

and phosphate are simulated both in the stable one-phase and unstable two-phase region. The results are evaluated for system and component descriptors that correlate with phase formation. A model predicting binodal curves based on a training set of experiments and independent simulation data is shown. Simulation results are validated by comparison to experimental data measured with a set of three solvatochromic dyes.

in preparation

5. ***HTS based selection of phases for ATPS CPC of mAbs***

STEFAN OELMEIER, CHRISTOPHER LADD-EFFIO, JÜRGEN HUBBUCH

This paper describes the development of a new screening method for the preparative application of ATPS with a focus on its further application in centrifugal partitioning chromatography. The screening method combines protein solubility data and ATPS characterization for the selection of appropriate phase systems. The feasibility of using ATPS in CPC for the purification of monoclonal antibodies is demonstrated. Factors influencing the purification results of the CPC step are discussed.

published in revised form in: Journal of Chromatography A. 2012;1252:104-114

6. ***A Sub-Two Minutes Method for mAb-Aggregate Quantification using Parallel Interlaced Size Exclusion HPLC***

PATRICK DIEDERICH, SIGRID HANSEN, STEFAN A OELMEIER, BIANCA STOLZENBERGER, JÜRGEN HUBBUCH

This paper demonstrates a method to reduce analysis time for monoclonal antibody-aggregate analysis by size exclusion chromatography. It evaluates new size-exclusion media suitable for higher back-pressures which reduce analysis time in single injection mode. Additionally, a way of controlling the U(H)PLC machine such that two columns can be run in parallel with both column running interlaced injections brought analysis time of a single sample down from 30 minutes to under two minutes.

published in: Journal of Chromatography A. 2011;1218:9010-9018.

MANUSCRIPT I

APPLICATION OF AN AQUEOUS TWO-PHASE
SYSTEMS HIGH THROUGHPUT SCREENING
METHOD TO EVALUATE MAB HCP
SEPARATION

Stefan A. Oelmeier, Florian Dismer, Jürgen Hubbuch*

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

* Corresponding author. Tel.: +49 721 608-42557; fax: +49 721 608-46240. E-mail address: juergen.hubbuch@kit.edu

5.1 Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation

Abstract

Aqueous two-phase systems as separation technique have regained substantial interest from the biotech industry. Biopharmaceutical companies faced with increasing product titers and stiffening economic competition reconsider ATPS as an alternative to chromatography. As the implementation of an ATPS is material, time and labor intensive, a miniaturized and automated screening process would be beneficial. In this article such a method, its statistical evaluation and its application to a biopharmaceutical separation task are shown. To speed up early stage ATPS profiling an automated application of the cloud-point method for binodal determination was developed. PEG4000-PO4 binodals were measured automatically and manually and were found to be identical within the experimental error. The ATPS screening procedure was applied to a model system and an industrial separation task. PEG4000-PO4 systems at a protein concentration of 0.75 mg/mL were used. The influence of pH, NaCl addition and tieline length was investigated. Lysozyme as model protein, two monoclonal antibodies and a host cell protein pool were used. The method was found to yield partition coefficients identical to manually determined values for lysozyme. The monoclonal antibodies were shifted from the bottom into the upper phase by addition of NaCl. This shift occurred at lower NaCl concentration when the pH of the system was closer to the pI of the distributed protein. Addition of NaCl, increase in PEG4000 concentration and pH led to significant loss of the mAb due to precipitation. Capacity limitations of these systems were thus demonstrated. The chosen model systems allowed a reduction of up to 50% HCP with a recovery of greater than 95% of the target proteins. As these values might not be industrially relevant when compared to current chromatographic procedures, the developed screening procedure allows a fast evaluation of more suitable and optimized ATPS system for a given task.

Keywords: *Aqueous two-phase system, high throughput screening, monoclonal antibody, design of experiments, multilinear regression*

1 Introduction

Aqueous two phase systems (ATPS) are long known to provide a way of separating biological components by liquid-liquid extraction [1]. There has been extensive research on ATPS and comprehensive review articles have been published on this topic [2–4]. For protein purification, ATPS have been shown to yield good results for various purification tasks. Its merits include integration of solid removal and product recovery, easy upscale, and high selectivity and recovery [5, 6]. ATPS have also been shown to be a valuable alternative to conventional chromatography in certain cases both judging from product recovery and purity as well as from process economics [7].

Regardless of these potential benefits, only a limited number of industrial scale application of this technique are currently observed. Recently however, ATPSs have attracted new interest of the biopharmaceutical industry. As current product titer improvements during cultivation shifts economic pressure onto the downstream side of a process [8–10] alternative processing steps might become favorable over conventional chromatography. While the applicability of ATPSs for the purification of human antibodies has been demonstrated [11–13], the experimental efforts needed for its implementation are still high. First, the binodal curve of the system to be used to select a reasonable screening space needs to be determined. While this information can be found in the literature for the most common ATPSs, any additive or deviation from literature conditions will lead to changes of the experimental space and thus create new systems. An automated way of gaining this information would significantly reduce resource consumption and speed up this process. Second, there is no clear understanding of the underlying physical principles, which makes it hard to select - from the wide range and diversity available - a set of suitable systems to be tested when faced with a certain purification task. Large numbers of experiments for each system are needed in order to gain sufficient insights. When performing the necessary studies on a

5.1 Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation

manual basis, substantial amounts of the target protein, usually a very limited resource, are consumed considering that a typical system size in manual screenings is 5 g [14]. Miniaturization of the screening would reduce this protein consumption dramatically. As time and labor are scarce resources in current short industrial downstream process development cycles an automation of the screening procedure and thus significant reduction in time and labor force needed would be highly beneficial. Automation and miniaturization would also allow to increase replicate number and number of conditions tested. This will lead to better and statistically more validated understanding of the screening results, which is essential to follow a QbD approach. In short, as in other areas of downstream process development, a high throughput screening strategy would be highly beneficial [15–17]. While single examples of statistically motivated manual ATPS screens have been carried out in the past and automated screening ATPS systems using single protein solutions have been shown, the application of an automatic ATPS screening method to a complex protein mixture in a DoE context in order to evaluate an industrial separation task is still missing. In this paper, such an approach is shown in detail. We first describe the optimization and extension of an earlier ATPS screening method established by our group [18]. Total system volume is decreased to 650 μ L thus reducing sample consumption. Liquid handling and sampling procedure were optimized concerning their accuracy and reproducibility. The method is statistically profiled to give a better understanding of the quality of the produced data. Additionally, a new method of automatically determining binodals of ATPS is described. The distribution of lysozyme is investigated by a design of experiments approach using a central composite face centered (CCF) design. The applicability of the screening method to an industrial purification task is demonstrated by measuring the distribution coefficient and the recovery of 2 monoclonal antibodies and a corresponding host cell protein mixture. Aqueous two-phase systems composed of PEG4000 and PO4 are used. The effect of system pH, NaCl addition and PEG4000 concentration on distribution and recovery are evaluated.

2 Materials and methods

2.1 System setup

2.1.1 Liquid handling station

In this study, a Tecan Freedom Evo 200 system (Tecan, Crailsheim, Germany) was used as liquid handling platform. It is equipped with three robotic arms: one 8-port liquid handling arm, outfitted with 4 fixed tips and 4 adapters for disposable tips, one standard robotic plate handling arm equipped with a centric gripper as well as a 96 channel liquid handling arm equipped with an eccentric gripper. Additionally, the system has an integrated centrifuge (Rotanta 46RSC, Hettich, Tuttlingen, Germany), a rotational shaker (Te-Shake, Tecan, Crailsheim, Germany) outfitted with a PreDictor frame (Tecan, Crailsheim, Germany), and an Infinite200 spectrophotometer (Tecan, Crailsheim, Germany).

2.1.2 Disposables

For spectroscopic measurements Greiner Bio-One (Kremsmünster, Austria) UV-Star plates (article no. 655801) were used. ATPS were prepared in 1.3 mL Nalgene Nunc (Rochester, NY, USA) Deep Well plates (product no. 260252). For all other purposes Greiner polypropylene flat bottom MTPs were used (article no. 655261).

2.1.3 Software

Excel 2007 (Microsoft, Redmond, WA, USA) files were used as import format and for data storage. All calculations, evaluation and visualization of data were done using Matlab R2009b (The Mathworks, Natick, ME, USA). The robotic workstation was controlled using Evoware 2.1 SP1 standard (Tecan, Crailsheim, Germany). The spectrophotometer was

5.1 Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation

controlled using Magellan 6.4 (Tecan, Crailsheim, Germany). Design of Experiments (DoE) setup and evaluation were done using Modde 8.0 (Umetrics, Umeå, Sweden) and Matlab.

2.2 Reagents, preparations and analytics

2.2.1 Stock solutions

Stock solutions were prepared in dH₂O as follows: 40% [w/w] PEG4000, 40% [w/w] NaH₂PO₄, 40% [w/w] K₂HPO₄, 25% [w/w] NaCl. PEG4000 was purchased from Carl Roth (Karlsruhe, Germany, product no. 0156). All other chemicals were bought from Sigma-Aldrich. Protein stock solutions ranged in concentration from 3.8 mg/mL to 10 mg/mL and were prepared as described in section 2.2.2. Phosphate stock solutions for ATPS were combined as follows: 52.54 g NaH₂PO₄ solution and 47.46 g K₂HPO₄ to yield pH 6.0; 29.71 g NaH₂PO₄ solution and 70.29 g K₂HPO₄ to yield pH 7.0; 11.15 g NaH₂PO₄ solution and 88.85 g K₂HPO₄ to yield pH 8.0. NaH₂PO₄ and K₂HPO₄ were chosen for the higher solubility of these salts compared to Na₂HPO₄ and KH₂PO₄.

2.2.2 Protein solution preparation

Two monoclonal antibodies (mAb) and a corresponding host cell protein (HCP) pool were supplied by Boehringer Ingelheim Pharma GmbH & Co. KG. The monoclonal antibodies were supplied as active pharmaceutical ingredient (API). If needed, protein concentration in the API solutions was increased by ultrafiltration as described in section 2.3. The mAb free flow-through obtained from a ProteinA chromatography step of clarified cell culture fluid (CCF) of a mAb producing cell line was used as HCP pool. The proteins contained therein were brought into a 20 mM Pi buffer at pH 7.0 by size exclusion chromatography (SEC) using a Sephadex G25 Medium (GE Healthcare, Little Chalfont, UK) column. Protein concentration was then increased by ultrafiltration using a 1kDa MWCO membrane. Lysozyme, used as model protein, was bought from Sigma (product no. L6876). Lysozyme stock was prepared as a 10 mg/mL protein solution in 20mM Pi buffer at pH 7.0.

2.2.3 Protein concentration analytic

Protein concentration was measured via absorption of light at 280 nm wavelength. For manual determination of protein concentration, standard UV transparent cuvettes placed in an Infinite200 spectrophotometer (Tecan, Crailsheim, Germany) were used. Protein concentration of a sample was calculated in reference to a calibration curve. Protein concentration determination on the liquid handling platform was performed using UV transparent microtiter plates placed in an Infinite200 spectrophotometer. Protein concentration was calculated by comparison to a reference sample with known concentration (internal standard).

2.3 Ultrafiltration

Ultrafiltration was done using a PureTec CP-120 system (SciLog, Middleton, WI, USA). Minimate TFF Omega 10 kDa molecular weight cut-off (MWCO) membranes (Pall, Port Washington, NY, USA) were used for mAb containing solutions, Minimate TFF Omega 1 kDa MWCO membranes (Pall, Port Washington, NY, USA) were used to increase HCP concentration.

2.4 Liquid handling calibration

Liquid handling by the 8-port liquid handling arm was calibrated by pipetting onto an analytical balance (WXTS205DU, Mettler-Toledo, Greifensee, Switzerland). Measurement of liquid density were done using a standard 5 mL pycnometer. A set of general liquid handling parameters optimized for low, medium and high viscosity fluids was used as basis for further optimization. The most prominent parameters changed as adaptation to higher

5.1 Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation

viscosities were plunger movement speed and waiting interval after plunger movement (“delay”). Liquid handling was optimized for both types of tips used (Fixed silicone coated tips and disposable tips) as well as for pipetting into air and onto a liquid. The calibrated range was 10 μL to 1000 μL .

2.5 ATPS screening process

2.5.1 Manual procedure

Systems of a total volume of 1 mL were prepared in standard 1.5 mL tubes. Stock solutions of phase forming components and dH_2O were combined as required and the systems mixed on a vortex shaker. After this initial mixing step, protein stock solution was added as last component. The obtained systems were then incubated on an end-over-end shaker for 15 minutes. Phase separation was achieved by centrifugation (17,000 x g, 3 minutes). Samples of both phases were taken, diluted 1:10 and analyzed for protein content by UV280 absorption. Samples from the bottom phase were obtained by puncturing the bottom of the tube.

2.5.2 Automated procedure

The method described herein is an extension and optimization of a method established earlier by our group [18]. First, aqueous two-phase systems of 650 μL total volume were put together, where water was dispensed first, followed by salt and polymer solutions. After an intermediate mixing step, protein solution was added as the last portion followed by a final mixing step. All liquids except polymer solutions were pipetted using the system’s fixed tips. Due to their high viscosity, polymer solutions were handled using disposable tips. Second, the systems were mixed on an orbital shaker (1300 rpm, 5 minutes) to ensure equilibrium between the two phases. To test if the chosen conditions led to the systems reaching equilibrium within the 5 minutes of mixing time, samples from the bottom phase were analyzed for Methylviolet. Phase separation was accomplished by centrifugation (1,800 x g, 3 minutes). Samples of both phases were then taken, diluted 1:10 in dH_2O and analyzed for protein and dye content. Taking a sample from the top phase was achieved by using the liquid level detection functionality of the LHS. Pipetting parameters were optimized to assure correct volume transfer of any of the differently composed phases. This was done as described above by using the top phases of four reference PEG4000 - PO_4 systems. The tip was lowered 0.25 mm below the liquid surface. The sample was drawn at a speed of 5 $\mu\text{L}/\text{s}$ with a delay time of 1000 ms after aspiration. The leading airgap in the tip was set to 30 μL . Dispense speed was 10 $\mu\text{L}/\text{s}$ with a breakoff speed of 50 $\mu\text{L}/\text{s}$. The factor between plunger movement and volume pipetted was determined to be 1.092. When sampling from the bottom phase, the pipetting tip has to be moved through the top phase. Thus, pipetting parameters were optimized for correct volume and for least carry over. Using a series of firmware commands directly sent to the liquid handling arm via visual basic scripts, 43 μL of sample were aspirated at a speed of 15 $\mu\text{L}/\text{s}$ after moving the tip through the top phase at a speed of 1 mm/s. Following a delay of 1000 ms, the tip was retracted at a speed of 20 mm/s. The tips were then lowered onto the surface of wells containing dH_2O to wash off remaining top phase stuck to the outside of the tips. 5 μL of bottom phase sample were discarded into the dH_2O containing well as it had been observed that top phase material is drawn into the tip due to capillary forces during retraction of the tip. Finally, the samples were dispensed at 133 $\mu\text{L}/\text{s}$. 30 μL sample of both phases were transferred into a UV transparent MTP containing 270 μL of dH_2O . The diluted samples were mixed (1000 rpm, 20 seconds) and then spun down (1,800 x g, X minutes) to remove any surface air bubbles. Protein and dye concentration were measured via light absorption (A280 for proteins, A586 for the dye) in relation to an internal standard. For the PEG4000 - PO_4 systems used

5.1 Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation

in this study, the internal standard was composed of 10% PEG4000 and 5% PO_4 at the appropriate pH.

2.6 Binodal and tieline determination

Binodal curves and tielines were measured to select reasonable screening ranges. Both automated and manual binodal determination were done using the “cloud point method” as described for example by Huddleston *et. al* [19]. Manual experiments were done by titration of one phase forming component into the other. Phase separation was determined by visual inspection after rigorous mixing on a vortex. While one-phase systems stay clear, two-phase systems become opaque after mixing. Determination using the robotic platform was done in UV transparent MTPs using a total volume of 300 μL per system. Systems were put together as described above, shaken at 1000 rpm on an orbital shaker while simultaneously being mixed using the robot’s 96 channel pipetting arm (30 cycles of aspiration and dispensing of 150 μL with aspiration speed of 100 $\mu\text{L}/\text{s}$ and dispense speed of 300 $\mu\text{L}/\text{s}$). Systems beyond the binodal were identified by an increase in light scattering measured by an increased absorbance of light at 230 nm wavelength. The 230 nm wavelength was chosen as the difference in absorbance caused by light scattering was largest between single and two-phase systems. In a series of increasing phase forming component concentration, one point on the binodal was identified as the mean of adjacent data points between which phase separation was observed. Systems containing a dye known to partition exclusively into the upper phase (methyl violet) were used to determine phase volumes as described before [18]. Tielines were then calculated by the lever arm rule.

2.7 Method validation

In order to validate the method, distribution coefficients and a binodal line generated by the liquid handling system were compared to manually determined values. The binodal of a PEG4000 - PO_4 system at pH 6.0 was used for comparison. Manually and automatically determined distribution coefficients of lysozyme in 4 different PEG4000 - PO_4 systems were compared. The systems compositions were: 1.: 10% PO_4 , 11.5% PEG4000, pH 6.0; 2.: 11% PO_4 , 14.9% PEG4000, pH 6.0; 3.: 10% PO_4 , 11.5% PEG4000, pH 8.0 and 4.: 11% PO_4 , 14.9% PEG4000, pH 8.0. Additionally, 33 equally composed systems were analyzed in order to evaluate the fluctuation range of liquid handling system generated data. System composition used for these experiments was 11%[w/w] PO_4 , pH 6.0, 10%[w/w] PEG4000, 0% NaCl, 0.75mg/mL lysozyme. The resulting yield and partition coefficient were combined into 11 means of triplicates.

2.8 Method application

2.8.1 DoE screening using lysozyme

As a proof of concept, the influence of tieline length (TTL), salt addition, protein concentration and pH on the distribution and recovery of lysozyme was investigated using a design of experiments approach. A reduced, face centered central composite (CCF) design was used within the boundaries given in table I. The design comprised 21 factor combinations, each measured in triplicates with two triplicates measured of the center point. Thus, a total of 66 systems was analyzed. PO_4 and PEG4000 concentrations were not treated as individual factors. Rather, PEG4000 concentration was chosen such that system volume ratio was approximately 1 at the three levels of PO_4 concentration tested. Thus, the calculated effect of the PO_4 concentration is the combined effect of changes in PO_4 and PEG4000 concentration. The three levels of PEG4000/ PO_4 thus lie on three different tielines and the measured effect relates to the effect of differing tieline length.

5.1 Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation

2.8.2 Industrial separation task

The screening method was used to evaluate the influence of PEG4000 concentration, NaCl addition and pH on the distribution and recovery of 2 monoclonal antibodies (termed mAb1 and mAb2) and an HCP pool in PEG4000 - PO_4 systems. 45 factor combinations were tested for the three protein solutions. Each data point was measured at least in triplicates. Table I shows the factor levels investigated. All factor level combinations were tested. Total protein concentration was 0.75 mg/mL in all cases.

Table I Upper part: Factor boundaries of the DoE investigation using lysozyme as protein for proof of concept. PO_4 and PEG4000 concentrations were coupled. A central composite face centred full factorial design was used. **Bottom part:** Factor levels investigated during the application of the screening method to mAb and HCP containing solutions. All possible combinations were tested.

DoE using lysozyme			
Factor	low	mid	high
% PO_4 [w/w]	10	10.5	11
% PEG4000 [w/w]	11.5	13.4	14.9
% NaCl [w/w]	0	2.5	5
pH [-]	6.0	7.0	8.0
Lysozyme [mg/mL]	0.25	0.5	0.75

HCP and mAb screening	
pH	6.0—7.0—8.0
PEG4000 % [w/w]	8—9—10—11—12
NaCl % [w/w]	0—1.75—3.5—4.25—7

3 Results

3.1 Method development

Prior to its use, the method was optimized in terms of sample volume, ease of handling, accuracy and reproducibility.

3.1.1 Liquid handling optimization

Liquid handling was optimized in terms of accuracy and reproducibility by a calibration procedure based on dispensing a series of set volumes onto an analytical balance. Reproducibility was found to be mainly influenced by three parameters: speed of liquid aspiration, speed of liquid dispensing, and waiting time after liquid aspiration (post aspiration delay). Depending on liquid characteristics such as viscosity, density, and surface tension, different values for these parameters were found to be optimal. Dispensing liquids into air rather than into liquid was found to be superior in terms of accuracy and precision. A calibration curve relating plunger movement to transferred volume was determined for all stock solutions used in the screenings. Table II summarizes the results of the liquid handling system calibration. R^2 of the determined calibration curves varied between 0.997 and 1.0. The average range r_{av} , calculated as the difference between maximum (X_{max}) and minimum volume (X_{min}) pipetted in percent of the mean volume pipetted (\bar{X}) at each of the n set volume, was below 3% for all individual stock solutions and volumes calibrated (see also equation 1). The average range was 1.5% for volumes of 25-200 μL and 0.66% for volumes of 250-990 μL . The average range over all stock solutions and all volumes was 1.1%.

5.1 Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation

$$r_{av} = 100 * \frac{1}{n} * \sum \frac{X_{max} - X_{min}}{\bar{X}} \quad (1)$$

Table II Liquid handling parameters and results of liquid handling calibration of the stock solutions used. AS = Aspiration speed, DS = Dispense speed.

Liquid	Volume [μ L]	AS [μ L/s]	Delay [ms]	DS [μ L/s]	Calibration
water	5 - 50	30	500	300	$1.026 \times x + 0.801$
	50 - 250	70	500	300	$1.026 \times x + 0.801$
	250 - 1000	70	500	300	$1.013 \times x + 0.0$
protein solution	15 - 50	30	500	300	$1.0246 \times x + 1.143$
	50 - 200	70	500	300	$1.0246 \times x + 1.143$
	200 - 1000	125	500	300	$1.0132 \times x + 3.744$
40% PEG4000	3 - 20	20	540	600	$1.042 \times x + 1.1$
	20 - 200	100	540	100	$1.0486 \times x + 4.8797$
	200 - 1000	100	750	50	$1.0374 \times x + 0.00$
40% PO_4	15 - 50	30	500	300	$1.0647 \times x + 2.9059$
	50 - 200	70	500	300	$1.0647 \times x + 2.9059$
	200 - 1000	125	500	300	$1.0291 \times x + 13.0$
25% NaCl	15 - 50	30	500	300	$1.0065 \times x + 1.0502$
	50 - 200	70	500	300	$1.0065 \times x + 1.0502$

3.1.2 Sampling the top and bottom phase

Sampling from the top and bottom phase was calibrated by pipetting a set volume of top and bottom phase material of 4 reference systems onto an analytical balance. After optimization of the respective liquid class, mean mass pipetted from the bottom phase was at 100.43% of the target with a CV of 1.62% (n=11). Mean mass pipetted from the top phase was at 98.8% of the target value with a CV of 1.04% (n=12). To check for carry over from the top phase when sampling the bottom phase, samples were taken from an ATPS containing methylviolet, a dye partitioning entirely into the top phase. Absorption values of these bottom phase samples did not differ significantly from the blind value, meaning no carry over was detected. The following parameters values were used for bottom phase sampling: aspiration speed: 45 ticks/sec, dispense speed: 400 ticks/sec, post-aspiration delay: 1000 ms, number of ticks aspirated: 129, number of ticks dispensed: 98. One tick equaled a plunger movement of 0.33 μ L.

3.1.3 System volume and mixing

Generally, the most convenient way of mixing on liquid handling stations lies in the use of orbital shakers. However, many applications of orbital shaking systems reported are hampered by insufficient power input, low shaking orbit or non-optimized parameter sets [18]. When applying orbital shaking, shaking frequency and system volume need to be optimized in concert. The first boundary condition was set by the hardware specifications of the shaker used, allowing a maximal rotation speed of 1500 rpm with a shaking orbit of 3 mm. The second boundary condition was imposed by spillage occurring during shaking when a too high liquid volume was deposited into the wells. This led to a small operational window, where liquid volumes below 600 μ L did not allow sufficient mixing while liquid volumes above 700 μ L led to spillage of liquid (visual inspection). At a system volume of 650 μ L and a rotating speed of 1300 rpm adequate mixing without spillage could be achieved.

5.1 Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation

3.2 Method validation

The distribution coefficient of lysozyme in PEG4000 - PO_4 systems was determined both manually and by use of the liquid handling system. Resulting partition coefficients are summarized in figure 1. Data is in good agreement between the two methods. Pairwise t-tests resulted in insignificant differences of the means of all conditions tested at the 5% significance level.

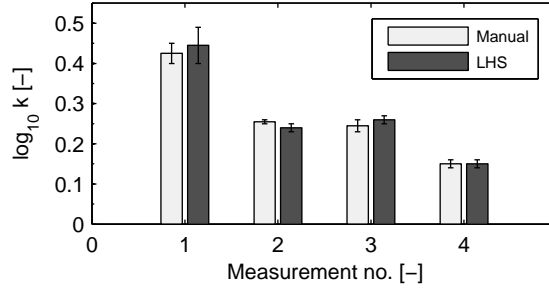


Figure 1: Comparison of distribution coefficients of lysozyme in 4 different systems both measured manually and by the liquid handling system. System composition was: Measurement 1: 10% PO_4 , 11.5% PEG4000, pH 6.0; Measurement 2: 11% PO_4 , 14.9% PEG4000, pH 6.0; Measurement 3: 10% PO_4 , 11.5% PEG4000, pH 8.0; Measurement 4: 11% PO_4 , 14.9% PEG4000, pH 8.0

Binodals were determined for the PEG4000 - PO_4 system at pH 6.0 both manually and by use of the LHS. Equation 2, also used by Merchuk *et al.* [20], was fitted to the combination of both datasets to yield the binodal line.

$$[PEG] = A * \exp(B * [PO_4]^{0.5} + C * [PO_4]^3) \quad (2)$$

The resulting coefficient values for the binodal line as well as its upper and lower 95% prediction bounds are $A = 78.1 \pm 9.0$, $B = -0.54 \pm 0.056$, and $C = (-5.31 \pm 0.76) * 10^{-04}$. R^2 of the fit is 0.997. As can be seen from figure 2, both LHS and manual data points fall within the 95% prediction bounds of the fitted curve. It was thus concluded that data is in good agreement between the two methods. Additionally, data taken from [21] was also in good agreement with the observations described herein.

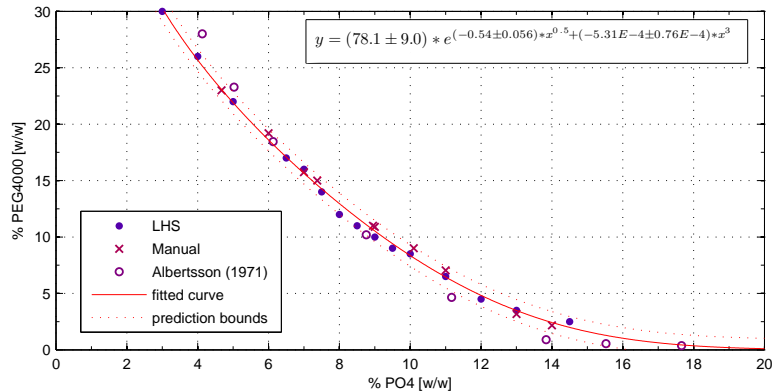


Figure 2: Binodal data of PEG4000 - PO_4 (pH 6.0) system measured manually and by the liquid handling system. The binodal line was fitted to equation 2 using the combination of both data sets. Dashed lines: 95% prediction bounds of the fit.

To determine the variation caused solely by the LHS, 11 triplicates of identically composed systems were evaluated. System composition used was 10% PO_4 , 11.5% PEG4000,

5.1 Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation

pH 6.0, 0.75 mg/mL lysozyme. The determined $\log_{10}k$ was 0.45 with a coefficient of variation (CV) of 5.5%. Mean recovery was 101.2% with a CV of 2.1%. Figure 3 shows the resulting $\log K$ and recovery values as well as their 95% confidence intervals of the 11 triplicate measurements. All values fall within the 95% confidence interval and distribute randomly around the mean over all values.

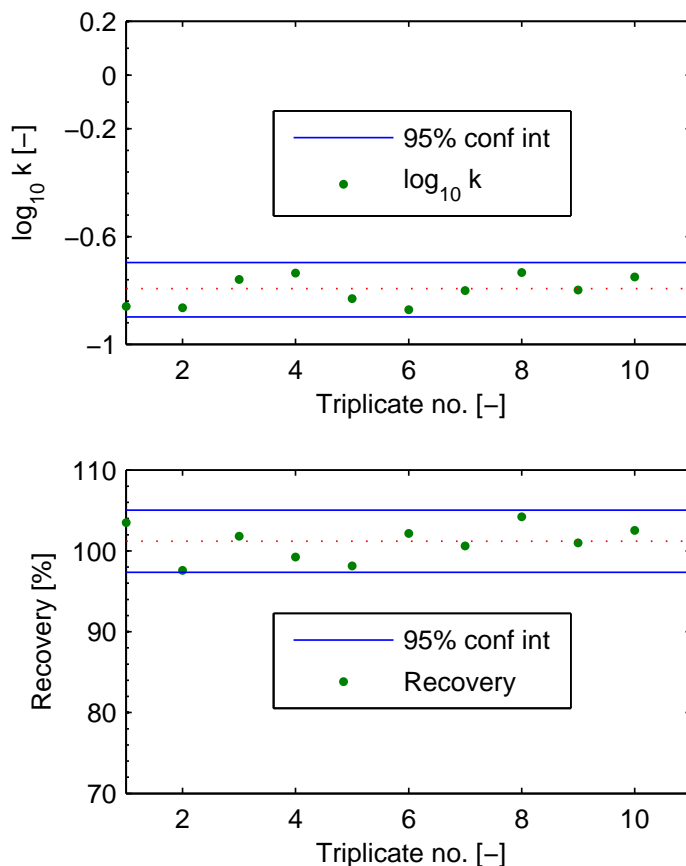


Figure 3: Variance of mean and recovery as measured by the LHS. System composition used was 10% PO_4 , 11.5% PEG4000, pH 6.0, 0.75 mg/mL lysozyme. Each data point represents the mean of a triplicate. 95% confidence intervals are shown as solid lines. 100% recovery and $\log_{10}k = 0$ are shown as dashed lines for orientation.

3.3 DoE screening using lysozyme

The screening method described above was used in concert with a DoE approach to investigate the distribution and yield obtained when investigating lysozyme behavior in PEG4000 - PO_4 systems. The combination of protein model and aqueous two phase system was chosen as a benchmark system for the developed screening method as the behavior of lysozyme in PEG-Phosphate systems builds on a broad knowledge base. A total of 66 systems was analyzed. Mass balance was consistent around 100% in all systems. A multi linear regression (MLR) of the distribution coefficient was performed. All varied parameters and their interactions were evaluated. Additionally, the squared NaCl concentration was included as a factor. The regression identified NaCl concentration as the major factor influencing lysozyme distribution within the conditions tested. $R^2 = 0.993$ was calculated for the resulting overall model. Figure 4 shows the normal probability plot of all effects and the main

5.1 Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation

effect plot of the NaCl concentration. The former was used to identify NaCl concentration as the major factor influencing the distribution of lysozyme. The effects of all factors except NaCl and $NaCl^2$ fall on a straight line in the normal probability plot. This implies that their effects are randomly distributed and thus not significant. The latter shows the increase of the distribution coefficient with increasing NaCl concentration both as measured and as modeled. The model fits well to the measured values. The mean partition coefficient increased from -0.56 to 1.19 by raising NaCl concentration from 0 to 2.5%. Adding another 2.5% NaCl further increased the mean partition coefficient to 2.03. The good fit of the model to the results and the low overall range of the distribution coefficient caused by varying all other factors at a constant level of NaCl also demonstrate the importance of NaCl concentration for the distribution of lysozyme. The automated screening method in concert with a DoE approach was thus capable of both identifying and quantifying the most important factors for the distribution of a model protein in an ATPS in this context.

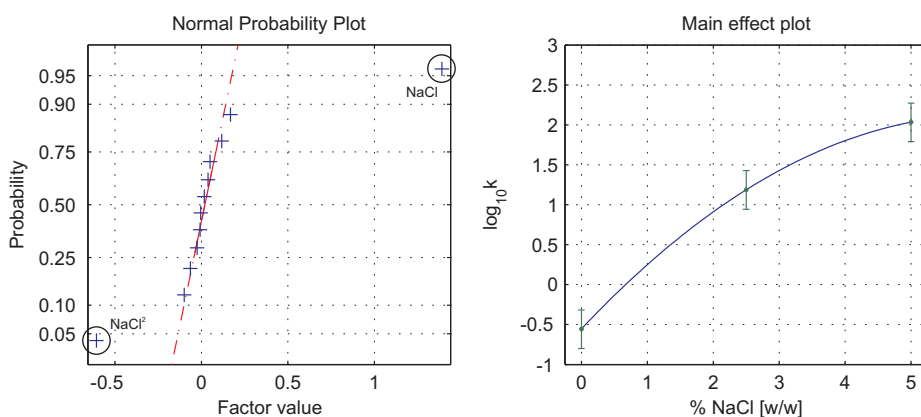


Figure 4: Left: Normal probability plot of the effects. NaCl and $NaCl^2$ do not follow normal probability distribution. NaCl concentration was thus identified as most influential on the distribution coefficient of lysozyme within the investigated range of parameters. Right: Main effect plot showing the effect of NaCl concentration on the distribution coefficient. Data points and error bars indicate the mean and overall range of distribution coefficients measured at one level of NaCl concentration. The plotted line shows the effect of NaCl concentration as calculated by the model.

3.4 Application to an industrial separation task

Influence of pH, PEG4000 concentration and NaCl addition on the distribution and recovery of two monoclonal antibodies and a host cell protein pool was investigated. In total, 143 individual systems were evaluated for mAb1, 209 systems for mAb2 and 200 systems for HCP. A total of 45 factor combinations was tested for all three protein solutions. All factor combinations were measured at least in triplicates. For each factor combination mean mass balance and distribution coefficient over all measurements were calculated. Systems resulting in $> 115\%$ mass balance were excluded from the dataset. The resulting mass balances and partition coefficient were modeled using MLR. R^2 values calculated for the fits vary between 0.77 and 0.97. Figure 5 shows the resulting contour plots for mAb1, figure 6 those for mAb2. Both antibody behave differently, but show similar trends. At all three pH values investigated, recovery reduction of $> 90\%$ was observed at the higher concentrations of PEG4000 and NaCl tested. This reduction is caused by precipitation of the antibody, as confirmed by visual inspection of the ATPSs. As pH values were increased and were thus closer to the pI of the antibodies, recovery reductions increased due to aggravated protein precipitation. At the same time, pH increase and NaCl addition cause distribution of the antibodies to shift more towards the upper phase of the systems. This is in line with published data [14]. No loss of protein due to precipitation was observed for

5.1 Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation

the HCP pool at any condition (data not shown). HCPs were distributed almost equally between the phases at all conditions tested. Figure 7 shows the distribution coefficient of HCPs and the difference of distribution coefficient of HCPs and mAbs. The three best conditions for separating HCP from each mAb while recovering $> 95\%$ of the mAb are listed in table III. In all cases listed, mAb distributes into the bottom phase while HCP distributes approximately evenly between the phases. Recovery and HCP reductions are calculated for the bottom phase of a single step process. Overall screening time for one microtiter plate containing 33 individual ATPS took between 2.5 and 3 h of robotic time and consumed 16.1 mg of protein.

Table III Best three conditions resulting in separation of HCP from POI at $> 95\%$ recovery of the POI.

POI	PEG4000	pH	NaCl	Recovery POI	Reduction HCP
mAb1	8	7	0	100.5	49.2
mAb1	8	8	0	97.2	50.2
mAb1	9	6	1.75	95.0	38.4
mAb2	9	6	1.75	95.5	38.4
mAb2	12	6	0	101.3	28.7
mAb2	8	6	3.5	97.6	40.7

Figure 7

4 Discussion

4.1 Method development

The performance of any automated screening procedure is dependent on its accuracy, precision and ease of handling. A minimum of sample volume needed and good knowledge about experimental error further determines the quality of a screening procedure. The accuracy and precision of liquid handling lies in the core of any automated liquid handling station. In this respect the most important issue and basis for every process is the ability to correctly aspirate and dispense set volumes. Closely related to pipetting as such is the special feature of sampling of liquids without any cross contamination between the samples. As can be seen from table II, optimum liquid handling parameters and plunger movement to dispense volume relation were different for each stock solution used. This is most likely caused by differing surface tension and viscosity of each stock solution and demonstrates that precise calibration of the liquid handling system is crucial for the success of a HTS method development. Coffman *et. al* [22] reported a variance of 3% when subsequently aspirating multiple liquids into one tip and dispensing them simultaneously. While this procedure speeds up the pipetting process, pipetting each liquid individually decreased the average volume range to be below 3% even for small volumes of below 50 μL . The highly viscous polymer stock solution was found to be handled with higher precision when using disposable tips instead of fixed tips. In this respect, Coffman *et. al* [22] reported, that pre-wetting of disposable tips by aspiration and dispensing of water prior to their use improved pipetting accuracy and precision. This however might lead to dilution of the dispensed liquid and was thus not applied in the current paper. Additionally, it was observed that previously used, thoroughly washed and dried disposable tips yielded less precision than virgin tips. Finally, when using disposable tips on a system liquid operated LHS the tips could be filled with system liquid prior to their use. This pretreatment decreased precision in our case and was thus decided against. Using virgin tips without any pretreatment, average volume range of pipetting 40% [w/w] PEG4000 stock solution was within 3% of mean pipetted volume.

5.1 Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation

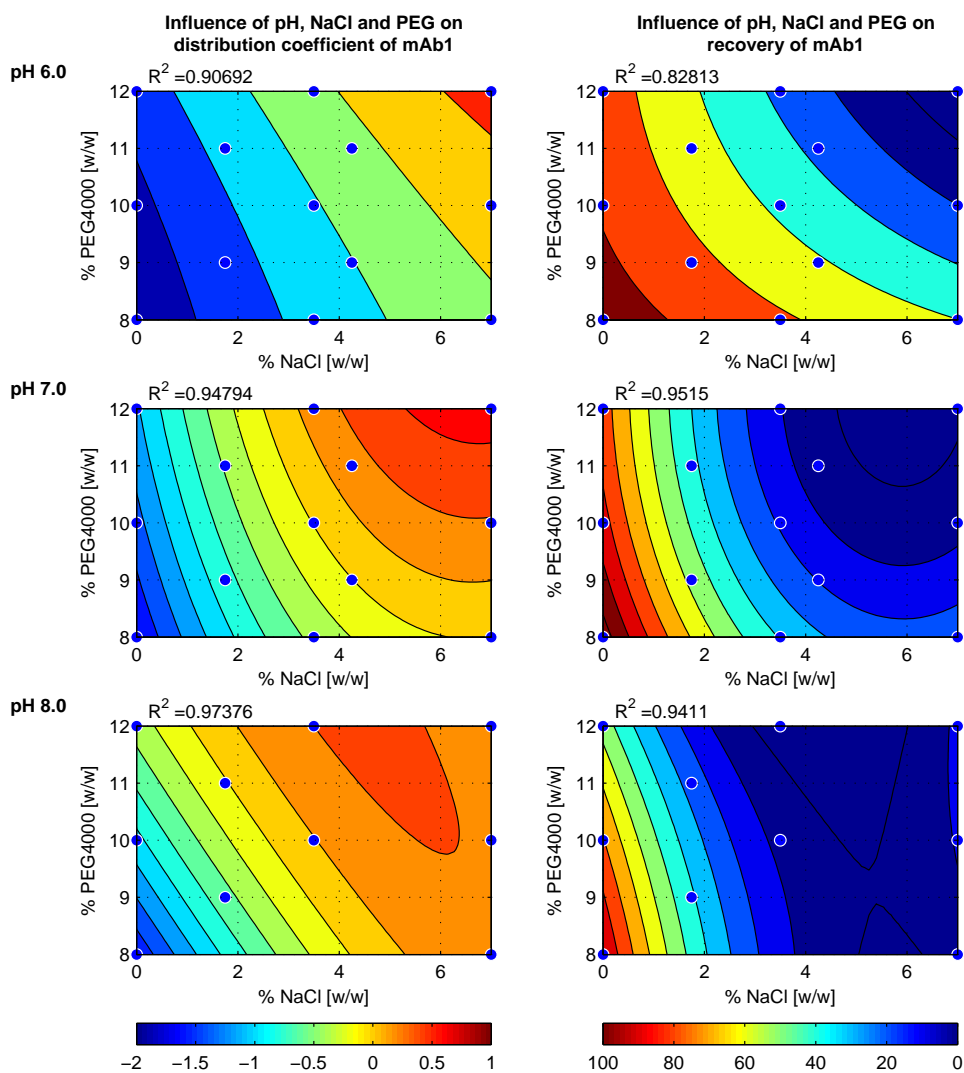


Figure 5: **Left:** \log_{10} of the distribution coefficient of mAb1. A value of 0 means equal distribution of the protein across the phases. **Right:** Recovery (Y) of mAb1 in %. Dots show factor combinations tested. Contour plot shows the level of $\log_{10}k$ and Y

5.1 Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation

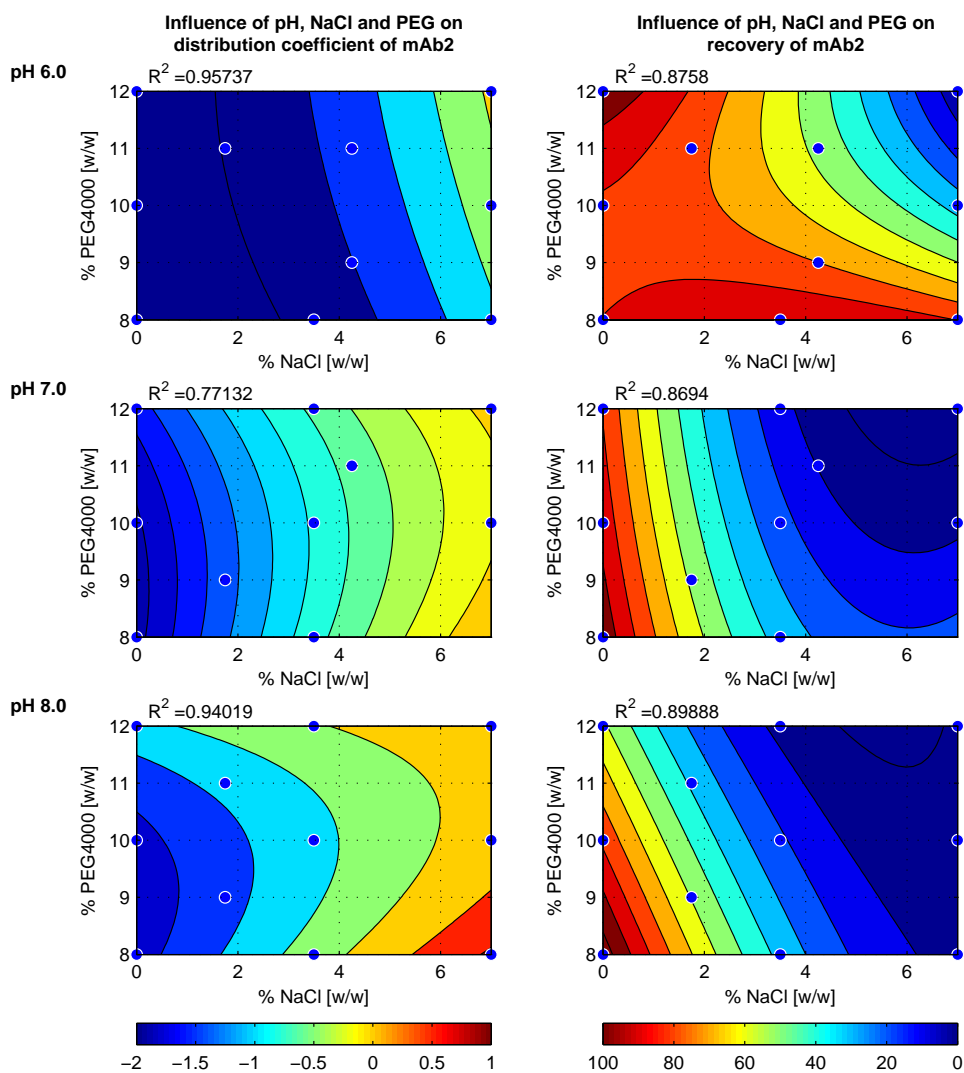


Figure 6: **Left:** \log_{10} of the distribution coefficient of mAb2. A value of 0 means equal distribution of the protein across the phases. **Right:** Recovery (Y) of mAb2 in %. Dots show factor combinations tested. Contour plot shows the level of $\log_{10}k$ and Y

5.1 Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation

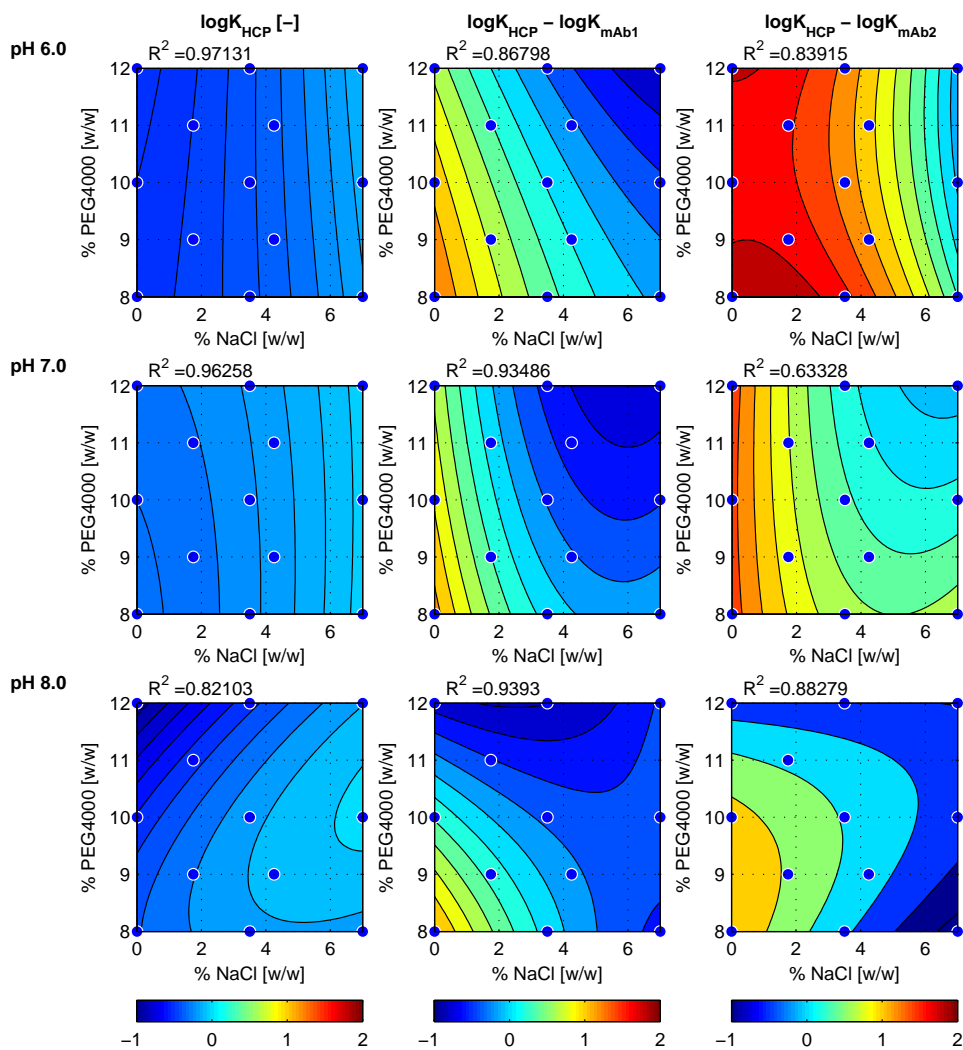


Figure 7: Left: \log_{10} of the distribution coefficient of HCP. Middle: Difference in $\log_{10}k$ of HCP and mAb1. Right: Difference in $\log_{10}k$ of HCP and mAb2. A positive difference in $\log_{10}k$ means the antibody distributed more into the bottom phase than HCP.

5.1 Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation

Mixing speed and total system volume were optimized in concert. A small operating window between 600 μL and 700 μL with an optimal parameter set of 650 μL and 1300 rpm was found. Hermann *et. al* [23] have investigated mixing of liquid in microtiter plates in the context of oxygen transfer rates for micro-scale cultivations. They have derived an equation to calculate the critical shaking frequency at which a liquid surface increase begins. Additionally, they quantified the increase in liquid height at a shaking diameter of 3 mm and a frequency of 1000 rpm to be approximately 4 mm. In our case, mixing needs to occur above this critical shaking frequency as the stock solution, pipetted on top of each other, would otherwise not mix due to their density difference and high viscosity. Following the equation of Hermann *et. al* [23], the critical shaking frequency is decrease with increasing shaking orbit, increasing liquid density, increasing filling volume, and decreased surface tension. In this study, we used an orbital shaker with an increased orbit, as well as liquid with a density considerably above 1 and a decreased surface tension cause by the addition of PEG. These three parameters were however fixed for any given system, leaving filling volume as only parameter left to optimize. With a filling volume of 700 μL and a well geometry of $8.5 * 29.1\text{mm}$ (round wells, round bottom) there are 15.4 mm head-space in each well. As spillage occurred at this volume with a shaking frequency of 1300 rpm, liquid height increase was significantly higher than the 4 mm observed by Hermann *et. al* for an aqueous buffer at 1000 rpm. The decreased surface tension and increased density of the liquid in our case might explain the higher increase in liquid height in comparison to the observations of Hermann *et. al* [23].

The “cloud point method” for binodal determination, adapted to a robotic platform, was successfully applied to a PEG4000 - PO_4 system. As can be seen from figure 2 the obtained result was in good agreement both with manually derived binodal data and data from the literature. Slight deviations from literature data can be seen towards the ends of the concentration range of phase forming components investigated. In contrast to the data presented herein, which were generated using a mixture of sodium and potassium phosphate, the binodal given in [21] was derived using only potassium phosphate. Additionally, PEG polydispersity and pH are not given in [21]. The slight deviations between the two dataset might be ascribed to these differences. Compared to the method described earlier [18], this automated “cloud point method” is significantly faster in generating the information necessary to select a reasonable screening range for the investigation of a selected ATPS. The only prerequisite to generating this data is to ensure precise liquid handling of the stock solutions as discussed above.

One advantage of an automated system over manually determined data is the ease of obtaining replicate numbers high enough for inferential statistics. One can, for example, measure the deviations caused by the LHS once with a high replicate number and infer the deviations of samples with lower replicate number from this data. This was exemplarily done for one system point of the model system used herein. 33 individual values of the distribution coefficient of lysozyme in a 10% PO_4 , 11.5% PEG4000, pH 6.0 system were investigated. Even though some outliers were identified in the 33 individual values, the 11 calculated triplicates showed normal distribution. It was concluded that outliers are reduced to a normal range when averaging over three measurements done by the LHS. This is in line with experience from other screenings with the LHS used. The ability to assess data quality to such a degree more easily is a clear advantage of a miniaturized and automated screening system over manual experiments.

5.1 Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation

4.2 Method application

4.2.1 DoE using lysozyme

The applicability of the screening method described herein combined with a typical DoE approach was demonstrated using a model system in a well characterized aqueous two-phase system. By screening a total of 66 systems, the influence of four parameters on the distribution coefficient were quantified. Interactions between the parameters were taken into account. The experimental points chosen purposely cover an area where no drastic changes in the overall performance of the systems could be expected in order to evaluate the methodology as such. Factors expected not to have significant effect on the distribution of the model protein were included to validate the method's capability to differentiate between significant and non-significant factors. As can be seen from figure 4 NaCl concentration was identified as the most influential parameter. While tieline length and pH showed little effect, no protein concentration effect was observed. These findings are in line with literature data [24]. Addition of a neutral salt to an aqueous two-phase systems shifts the relative importance of factors determining protein partitioning. In the case described herein, NaCl allows the protein to partition into the more hydrophobic top phase of the system. This is most likely caused by a shielding of the electrostatic forces which promote protein distribution into the bottom phase [21] as well as an increase of hydrophobic effects pulling the protein into the upper phase. The more NaCl is added, the less its effect, as the bottom phase gets depleted of protein. This is accounted for in the model by the quadratic term of NaCl concentration getting a negative cofactor value. The pH of an ATPS can play an important role in protein partitioning as the protein's net charge is determined by the pH of the solution. In the case of lysozyme, however, no pH effect was expected in the investigated range as the pI of lysozyme lies above pH 11. pH was however included into the screening to demonstrate the ability of the screening method in concert with DoE to differentiate between significant and non-significant factors. As all findings were in line with the expectation and reproducibility was high, it was concluded that the method described herein is a suitable screening tool for aqueous two-phase systems.

4.2.2 Application to mAbs & HCP

The established screening method was subsequently applied to an industrially relevant separation task. Two monoclonal antibodies and a HCP pool were distributed in PEG4000- PO_4 systems. The screening method was successfully applied to determine distribution and recovery of 552 single systems with changes in pH, PEG4000 and NaCl concentration. Figures 5, 6, and 7 show the results of these screenings. Without addition of NaCl, the antibodies distributed almost exclusively into the bottom phase. Addition of NaCl shifted the equilibrium more to the upper phase. Besides pushing the antibody more into the upper phase, addition of NaCl also induced precipitation. Severe loss of both antibodies due to precipitation was observed at the higher levels of NaCl tested. Changing the pH showed a similar effect. As both antibodies have a pI above 8.0, increasing the pH from 6.0 to 8.0 decreased overall charge of the molecules without changing the overall charge's sign. Both distribution to the upper phase and precipitation were increased thereby. While the two monoclonal antibodies tested gave very similar results concerning their precipitation and their trends of distribution within the systems tested, mAb2 had stronger affinity for the bottom phase. It showed lower distribution coefficients at all conditions tested than mAb1. The HCP distributed approximately evenly over the phases at all conditions. As HCP is a mixture of diverse proteins, this is not surprising. The distribution coefficient measured is a sum signal over all proteins in the pool. More specific analytics would be needed to determine which protein of the HCP pool distributes to which phase and to what extent.

Mean mass balance and partition coefficients were modeled via MLR for both mAbs

5.1 Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation

and the HCP-pool at each pH. Most calculated R^2 values are around 0.9 indicating a good fit. Single R^2 values were, however, lower. This might be ascribed to the more complex distribution between the two liquid phases and a solid phase precipitate observed when partitioning mAbs and HCP. More factors than identified in the case of lysozyme, as well as the interaction of these factors show an influence on distribution and mass balance. Mass balance was not consistently around 100% for the mAbs as precipitation occurred thus adding additional noise to the data. This noise is firstly caused by the precipitation itself. The screening procedure does not include a prolonged equilibration step. Thus, it is unlikely that an equilibrium between precipitated and soluble protein is reached. Differing precipitation kinetics might then add noise to the data. Secondly, although less often observed, data can be skewed by increased light scattering by precipitates that were not removed by centrifugation.

The screening method described herein performed well both when using single protein solutions (mAbs and lysozyme) and a complex protein mixture (HCP) as feed. Thus, the distribution of mAbs and HCP together in a single system can be inferred. Within the systems screened, best conditions would give up to 50% clearance of HCP while recovering > 95% of the antibody in a single step (see table III) . While this separation is not sufficient for a single step purification compared to current chromatographic methods [25, 26] usually employed during mAb purification, it demonstrates the usefulness of the described screening procedure. The screening method was capable of determining the HCP clearance and thus assessing the applicability of the investigated ATPSystems. Further screenings might be conducted at different molecular weights of PEG or with polymer-polymer systems.

Recently Azevedo *et. al* published a series of papers describing the purification of polyclonal and monoclonal human antibodies from an artificial mixture of impurities and cell culture supernatant via ATPS [14, 27, 28]. Using 12% PEG6000 in combination with 10% PO_4 at pH 6 and 15% NaCl they demonstrated a yield of 88% of an mAb from CHO cell culture supernatant with a purification factor of 4.1. The total mAb concentration used in their screening was 0.42 mg/mL. In comparison, we found that systems with PEG4000 and a total IgG concentration of 0.75 mg/mL were not suitable at NaCl concentration above 3.5% due to yield loss caused by precipitation. In contrast to the results by Azevedo *et al.* best conditions found in our screening were at 0% NaCl. The mAbs distributed almost entirely into the bottom phase under these conditions. Yield close to 100% and depletion of 50% HCPs would give a lower purification factor of approximately 2 than identified by Azevedo *et al.* With this result we demonstrate the importance of solubility of the target protein for the outcome of an ATPS screening. Low solubility can arise due to the high content of salts and polymer in the mixture, leading to salting out and preferential exclusion of proteins. This was also observed by Azevedo *et al.* [14]. Thus, optimization of an aqueous two-phase extraction step includes not only screening for the optimal distribution of target protein and impurities. Rather, solubility of the components to be separated needs to be considered as well. Both distribution and solubility are influenced by a multitude of parameters including phase forming components, additives such as neutral salts, tie line length, pH, and protein concentration. Protein concentration in each phase in turn depends on the protein's distribution coefficient as well as the volume ratio of the two phases. When investigating such complex phenomena, a fast and automated screening procedure is of great benefit.

To further improve the separation of a suboptimal system caused by a near equal distribution of HCP, a continuous mode of operation could be chosen such as demonstrated for PEG1000- PO_4 systems by Sutherland *et. al* [29]. The distribution of the antibodies in the batch systems measured here should predict its behavior in a continuous separation well. The sum signal HCP distribution in a batch system however cannot be used to accurately

5.1 Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation

predict the outcome of a continuous process. Nevertheless, the clearance of a continuous process should be at least that of the batch process. A higher degree of HCP clearance is highly likely as all components of the HCP pool distributing more to the upper phase than the antibody will be removed. Using a software tool recently published [30], one can see that if HCP was to behave as a single component, excellent separation should be achievable using a continuous mode of operation at the conditions identified.

5 Conclusion and outlook

In the presented work, it was shown how the previously described ATPS high throughput screening method was further optimized in respect to system volume as well as pipetting accuracy and precision. The applicability of the method in a DoE context to both single protein solutions and complex protein mixtures was demonstrated. While the ATPSs investigated herein did not yield HCP clearance comparable to chromatographic separation steps, the screening procedure has proven useful in evaluating ATPSs for a complex, industrial separation task. Future work might include screening of ATPSs with PEGs of different molecular weight as well as evaluation of upscale to continuous separations of the best conditions identified in the screenings.

6 Acknowledgements

The authors would like to thank Boehringer Ingelheim Pharma GmbH & Co. KG for supply of material, financial and scientific support. Additionally, the authors would like to thank Johannes Knoll for conducting parts of the lab work.

References

- [1] P.-A. Albertsson, Particle fractionation in liquid two-phase systems the composition of some phase systems and the behavior of some model particles in them application to the isolation of cell walls from microorganisms, *Biochim Biophys Acta* 27 (1958) 378–395.
- [2] R. Hatti-Kaul, Aqueous two-phase systems - a general overview, *Mol Biotechnol* 19 (3) (2001) 269–277.
- [3] U. B. Hansson, C. Wingren, Separation of antibodies by liquid-liquid aqueous partition and by liquid-liquid partition chromatography, *Sep Purif Methods* 27 (2) (1998) 169–211.
- [4] J. G. Huddleston, A. Lyddiatt, Aqueous 2-phase systems in biochemical recovery - systematic analysis, design, and implementation of practical processes for the recovery of proteins, *Appl Biochem Biotechnol* 26 (3) (1990) 249–279.
- [5] J. Thömmes, M. Halfar, H. Gieren, S. Curvers, R. Takors, R. Brunschier, M. R. Kula, Human chymotrypsinogen b production from *Pichia pastoris* by integrated development of fermentation and downstream processing. part 2. protein recovery, *Biotechnol Prog* 17 (3) (2001) 503–512.
- [6] C. Kepka, E. Collet, J. Persson, A. Stahl, T. Lagerstedt, F. Tjerneld, A. Veide, Pilot-scale extraction of an intracellular recombinant cutinase from *E. coli* cell homogenate using a thermoseparating aqueous two-phase system, *J Biotechnol* 103 (2) (2003) 165–181.
- [7] O. Aguilar, V. Albiter, L. Serrano-Carren, M. Rito-Palomares, Direct comparison between ion-exchange chromatography and aqueous two-phase processes for the partial purification of penicillin acylase produced by *E. coli*, *J Chromatogr, B: Anal Technol Biomed Life Sci* 835 (1-2) (2006) 77–83.

5.1 Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation

- [8] A. C. A. Roque, C. R. Lowe, M. A. Taipa, Antibodies and genetically engineered related molecules: Production and purification, *Biotechnol Prog* 20 (3) (2004) 639–654.
- [9] D. Low, R. O’Leary, N. S. Pujar, Future of antibody purification, *J Chromatogr, B: Anal Technol Biomed Life Sci* 848 (1) (2007) 48–63.
- [10] S. S. Farid, Process economics of industrial monoclonal antibody manufacture, *J Chromatogr, B: Anal Technol Biomed Life Sci* 848 (1) (2007) 8–18.
- [11] A. M. Azevedo, P. A. J. Rosa, I. F. Ferreira, M. R. Aires-Barros, Chromatography-free recovery of biopharmaceuticals through aqueous two-phase processing, *Trends Biotechnol* 27 (4) (2009) 240–247.
- [12] J. Benavides, M. Rito-Palomares, Practical experiences from the development of aqueous two-phase processes for the recovery of high value biological products, *J Chem Technol Biotechnol* 83 (2) (2008) 133–142.
- [13] D. Platis, N. E. Labrou, Application of a peg/salt aqueous two-phase partition system for the recovery of monoclonal antibodies from unclarified transgenic tobacco extract, *Biotechnol J* 4 (9) (2009) 1320–1327.
- [14] A. M. Azevedo, P. A. Rosa, I. F. Ferreira, M. R. Aires-Barros, Optimisation of aqueous two-phase extraction of human antibodies, *J Biotechnol* 132 (2) (2007) 209–17.
- [15] M. Bensch, P. S. Wierling, E. von Lieres, J. Hubbuch, High throughput screening of chromatographic phases for rapid process development, *Chem Eng Technol* 28 (11) (2005) 1274–1284.
- [16] M. Wiendahl, P. S. Wierling, J. Nielsen, D. F. Christensen, J. Krarup, A. Staby, 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* 31 (6) (2008) 893–903.
- [17] A. Susanto, K. Treier, E. Knieps-Gruenhagen, E. von Lieres, J. Hubbuch, High throughput screening for the design and optimization of chromatographic processes: Automated optimization of chromatographic phase systems, *Chem Eng Technol* 32 (1) (2009) 140–154.
- [18] M. Bensch, B. Selbach, J. Hubbuch, High throughput screening techniques in downstream processing: Preparation, characterization and optimization of aqueous two-phase systems, *Chem Eng Sci* 62 (7) (2007) 2011–2021.
- [19] J. G. Huddleston, H. D. Willauer, R. D. Rogers, Phase diagram data for several peg + salt aqueous biphasic systems at 25 c, *J Chem Eng Data* 48 (5) (2003) 1230–1236.
- [20] J. C. Merchuk, B. A. Andrews, J. A. Asenjo, Aqueous two-phase systems for protein separation studies on phase inversion, *J Chromatogr, B: Anal Technol Biomed Life Sci* 711 (1-2) (1998) 285–293.
- [21] P.-A. Albertsson, Partition of cell particles and macromolecules, New York: Wiley-Interscience, 1971.
- [22] J. L. Coffman, J. F. Kramarczyk, B. D. Kelley, High-throughput screening of chromatographic separations: I. method development and column modeling., *Biotechnol Bioeng* 100 (4) (2008) 605–18.

5.1 Application of an Aqueous two-phase systems high throughput screening method to evaluate mAb HCP separation

- [23] R. Hermann, M. Lehmann, J. Büchs, Characterization of gas-liquid mass transfer phenomena in microtiter plates., *Biotechnol Bioeng* 81 (2) (2003) 178–86.
- [24] F. Hachem, B. A. Andrews, J. A. Asenjo, Hydrophobic partitioning of proteins in aqueous two-phase systems, *Enzyme Microb Technol* 19 (7) (1996) 507–517.
- [25] P. S. Wierling, E. Knieps-Grnhagen, R. Bogumil, J. Hubbuch, High-throughput screening of packed-bed chromatography coupled with seldi-tof ms analysis: monoclonal antibodies versus host cell protein., *Biotechnol Bioeng* 98 (2) (2007) 440–50.
- [26] A. Stein, A. Kiesewetter, Cation exchange chromatography in antibody purification: ph screening for optimised binding and hcp removal., *J Chromatogr, B: Anal Technol Biomed Life Sci* 848 (1) (2007) 151–8.
- [27] A. M. Azevedo, P. A. J. Rosa, I. F. Ferreira, M. R. Aires-Barros, Integrated process for the purification of antibodies combining aqueous two-phase extraction, hydrophobic interaction chromatography and size-exclusion chromatography, *J Chromatogr, A* 1213 (2) (2008) 154–161.
- [28] A. M. Azevedo, P. a. J. Rosa, I. F. Ferreira, J. de Vries, T. J. Visser, M. R. Aires-Barros, Downstream processing of human antibodies integrating an extraction capture step and cation exchange chromatography., *J Chromatogr, B: Anal Technol Biomed Life Sci* 877 (1-21-2) (2009) 50–8.
- [29] I. a. Sutherland, G. Audo, E. Bourton, F. Couillard, D. Fisher, I. Garrard, P. Hewitson, O. Intes, Rapid linear scale-up of a protein separation by centrifugal partition chromatography., *J Chromatogr, A* 1190 (1-21-2) (2008) 57–62.
- [30] J. de Folter, I. A. Sutherland, Universal counter-current chromatography modelling based on counter-current distribution, *J Chromatogr, A* (2008) 4218–4224.

MANUSCRIPT II

EVALUATION OF CORRELATIONS OF PROTEIN PROPERTIES AND ATPS DISTRIBUTION

Stefan A. Oelmeier, Johannes Knoll, Jürgen Hubbuch*

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

** Corresponding author. Tel.: +49 721 608-42557; fax: +49 721 608-46240. E-mail address: juergen.hubbuch@kit.edu*

in preparation

Abstract

Aqueous-two phase extraction (ATPE) has been shown to be a possible alternative to chromatographic separation steps for the purification of protein drugs. The implementation of an ATPE step is, however, impeded by the little mechanistic understanding of the forces at play. This creates the need for laborious empirically driven process development, making ATPE less attractive to the pharmaceutical industry. Several reports have described correlations of certain protein attributes with their distribution in ATPS. These correlations are mostly based on a set of model proteins and their distribution in a limited number of ATPSs. Applying current high-throughput experimentation methods, we reevaluated these correlations with a substantially increased dataset. 500 combinations of model protein and system composition were evaluated. Seven classes of PEG-salt systems with and without the addition of NaCl were used. While previously published general trends in protein partitioning were confirmed, contrary to previously published studies, no strong correlations between protein descriptors and distribution were found. It was thus shown that protein distribution in aqueous two-phase systems is too complex to be modeled using protein descriptors alone. Additionally, it was found that protein solubility plays a crucial role even at low protein concentrations. This effect has to be taken into account if a correlation is to hold under preparative conditions.

Keywords: *Aqueous two-phase systems, surface hydrophobicity, correlation, solubility*

1 Introduction

Recently, aqueous two-phase extraction as a purification technique for proteins has gained renewed interest by the pharmaceutical industry. Purification processes of proteinaceous drug substances are currently highly dependent on chromatographic separations. While chromatographic separation steps are comparably expensive due to material and equipment cost, the implementation of these steps is greatly facilitated by both the large experience and the existence of detailed models describing the separation process. From simple relation such as a proteins isoelectric point to its behavior on an ion exchange resin to detailed mechanistic models and molecular simulations predicting elution times [1], there exists a large supporting framework of knowledge around chromatography. Most alternative separation methods, including aqueous two-phase extraction, lack this framework making their implementation more empirical and less attractive. In order to build such a framework, knowledge about the underlying forces that drive protein behavior in aqueous two-phase systems needs to be generated.

One way of doing so is to gather data and generate experience based general rules and guidelines for the implementation of the technique. In two review articles by Rito-Palomares [2, 3] these general rules are put into a step development concept starting with the characterization of the target molecule followed by the selection of the pH and phase forming components, and finishing with an optimization procedure including other factors such as tie line length or product load. Solution pH is stated as a highly influential parameter and the general rule is given, that molecules tend to distribute into the PEG rich phase when charged negatively. This implies that there should be a correlation between a proteins pI and its distribution coefficient. Similar rules, mostly derived from experimental experience can be found in the textbooks by Albertsson [4], Walter [5], and Hatti-Kaul [6].

Another way of facilitating aqueous two-phase extraction step development is to search for correlations between protein descriptors and their distribution in ATPSs and several such correlations have previously been described. A linear correlation between logMW of a protein and its logK at the proteins isoelectric point in a dextran/PEG two-phase system was found by Sasakawa *et al.* [7]. Johansson *et al.* [8] established a linear correlation be-

tween the net charge of a protein and its logK. They used lysozyme and ovalbumin as model proteins and a dextran/PEG ATPS to which different salts were added. A protein's surface hydrophobicity was found to influence its distribution in several different ATPSs. Asenjo *et al.* [9] used three different measures of protein hydrophobicity, retention in hydrophobic interaction chromatography, retention in reversed phase chromatography, and a scale based on protein precipitation established by Przybycien *et al.* [10]. They showed a linear correlation between the precipitation based hydrophobicity descriptor and the distribution of seven model proteins in a PEG4000-phosphate and a PEG8000-phosphate system. They also established, that the correlation was better for systems containing additional NaCl. Hachem *et al.* [11] demonstrated that the same correlation was also found when using five model proteins distributed in a PEG-dextran system. Berggren *et al.* [12] showed that for two EOPO-dextran systems the distribution of eight model proteins was best correlated to a hydrophobicity scale based on amino-acid hydrophobicity and solvent accessible surface contribution. Thus, while strong correlations were found for a range of system types and different sets of model proteins, no unifying method of predicting protein distribution has yet been found. The correlations are based on various different scales of protein descriptors that are mostly unrelated to one another.

While correlation does not imply causation, strong correlations can give hints as to which forces are at play. Additionally, if correlations are found, they might be used predictively when faced with a new separation process. To apply a correlation in this way however, they must be proven reliable when faced with a dataset different from the one they were found in. There has been no report of an implementation of an ATPE step using distribution predictions based on the correlations described above. Additionally, an investigation on the validity of such correlations under preparative conditions is missing. However, for the industrial applicability of such correlations, their validity under such conditions is crucial. With high throughput methods now available for the evaluation of protein distribution and recovery in aqueous two-phase systems [13, 14] the correlations described above can be revisited with larger datasets.

In this paper, we reevaluate the published correlations between protein descriptors and their distribution in seven different PEG-salt systems against a larger dataset. 50 different ATPS system compositions and 10 model proteins are used. Single- and multi-factor-correlations are evaluated. We evaluate both correlations within one type of ATPS and correlations over all types of ATPS used. We revisit generally accepted trends in protein distribution and discuss the influence of protein solubility on these correlations. A preparative applicability of such correlations is the underlying aim of this work.

2 Materials & methods

2.1 Chemicals

Stock solutions were prepared in dH₂O as follows: 40% [w/w] PEG4000, 40% [w/w] 40% [w/w] PEG8000, 40% [w/w] NaH_2PO_4 , 40% [w/w] K_2HPO_4 , 25% [w/w] NaCl. PEG4000 was purchased from Carl Roth (Karlsruhe, Germany, product no. 0156). All other chemicals were bought from Sigma-Aldrich. 29.71 g NaH_2PO_4 solution and 70.29 g K_2HPO_4 were combined to yield pH 7.0. NaH_2PO_4 and K_2HPO_4 were chosen for the higher solubility of these salts compared to Na_2HPO_4 and KH_2PO_4 .

2.2 Proteins

For this study, ten model proteins were obtained from Sigma-Aldrich (St. Louis, USA). The following list give the name, product number, and abbreviation of all proteins. Human

5.2 Evaluation of correlations of protein properties and ATPS distribution

serum albumin (A1653, “HSA”), bovine serum albumin (A7906, “BSA”), Cytochrome C (C2506, “Cyt”), catalase (C40, “Cata”), α -chymotrypsinogen A (C4879, “Chymo”), β -glucuronidase (G0251, “Glucu”), β -lactoglobulin (L3908, “Lacto”), hen egg white lysozyme (L6876, “Lys”), Myoglobin (M1882, “Myo”), α -amylase (10065, “Amyl”). All proteins were dissolved in 50 mM phosphate buffer (pH 7.0) to yield a final concentration of 10 mg/mL.

2.3 Aqueous two-phase systems used in this study

Seven different classes of aqueous two-phase systems were used within this study. They are referred to throughout this manuscript as “System I” to “System VII”. All systems were composed of polyethylene glycol and a salt, with two molecular weights of polyethylene glycol and two types of salt. Additionally, NaCl was added to the systems in different amounts. Several PEG4000-PO₄ system without NaCl showed very low recoveries due to precipitation. These systems were not tested with NaCl added. For each type of system, 5-10 system points were evaluated, with each system point measured at least six times, at most ten times. Table I shows all system compositions. Final protein concentration in all system was 1 g/L for all model proteins.

2.4 Liquid handling station

In this study, a Tecan Freedom Evo 200 system (Tecan, Crailsheim, Germany) was used as liquid handling platform. It is equipped with three robotic arms: one 8-port liquid handling arm with 8 fixed tips, one standard robotic plate handling arm equipped with a centric gripper as well as a 96 channel liquid handling arm equipped with an eccentric gripper. Additionally, the system has an integrated centrifuge (Rotanta 46RSC, Hettich, Tuttlingen, Germany), a rotational shaker (Te-Shake, Tecan, Crailsheim, Germany) outfitted with a PreDictor frame (Tecan, Crailsheim, Germany), and an Infinite200 spectrophotometer (Tecan, Crailsheim, Germany).

2.5 Disposables

For spectroscopic measurements Greiner Bio-One (Kremsmüister, Austria) UV-Star plates (article no. 655801) were used. ATPS were prepared in 1.3 mL Nalgene Nunc (Rochester, NY, USA) Deep Well plates (product no. 260252). For all other purposes Greiner polypropylene flat bottom MTPs were used (article no. 655261).

2.6 Software

Excel 2010 (Microsoft, Redmond, WA, USA) files were used as import format and for data storage. All calculations, evaluation and visualization of data were done using Matlab R2011a (The Mathworks, Natick, ME, USA). The robotic workstation was controlled using Evoware 2.2 standard (Tecan, Crailsheim, Germany). The spectrophotometer was controlled using Magellan 6.4 (Tecan, Crailsheim, Germany). Evaluation of 3D protein structure, energy minimization, and surface characterizations were performed using the Yasara Structure software package [15], version 10.10.29.

2.7 ATPS screening method

Aqueous two-phase system distribution and recoveries of the ten model proteins in the 50 individual systems selected for this study were performed on a liquid handling platform as described by Oelmeier *et al.* [14]. In short, the systems are put together in deep well 96 well plate by pipetting the appropriate stock solutions to yield a total system volume of 650

5.2 Evaluation of correlations of protein properties and ATPS distribution

Table I ATPS compositions used in this study. Seven different types of systems were used, termed “System I” through “System VII”. x designates systems that were evaluated.

AS	PEG4000	System I 0% NaCl		
13	7	x		
13	10	x		
13	14	x		
13	17	x		
13	19,5	x		
10	14	x		
16	14	x		
11,5	14	x		
14,5	14	x		
18	14	x		

Pi	PEG4000	System II 0% NaCl	System IV 1,5% NaCl	System VI 4% NaCl
13	10	x	x	x
13	13	x	x	x
13	15	x	x	x
13	18	x		
10	15	x	x	x
12	15	x		
14	15	x		
16	15	x		
18	15	x		
15	15		x	x

Pi	PEG8000	System III 0% NaCl	System V 1,5% NaCl	System VII 4% NaCl
12	8	x	x	x
10	15	x	x	x
15	15	x	x	x
13	10	x	x	x
13	13	x	x	x
13	15	x	x	x
13	17	x	x	x

μL . The plate is shaken to ensure equilibrium between the two phases. Phase separation by centrifugation is followed by sampling of both phases ($30\mu\text{L}$ sample each) and measuring the protein content of the diluted (10x) sample. Including blanks approximately 6000 ATPSs were analyzed in this way.

2.8 Precipitation screening method

In order to measure m^* -values of the model proteins as described by Hachem *et al.* [11] and Berggren *et al.* [12], a protein precipitation screening on the liquid handling station was developed. Buffer, precipitant stock solution, and protein stock solution are combined to yield a sequence of samples with equal protein concentration but increasing precipitant content. The precipitant used for m^* determination was ammonium sulfate. Sample volume was $300\mu\text{L}$. Sample container were standard 96 well plates made of polypropylene. Using the rotational shaker samples were mixed and incubated for at least 15 minutes, at most 24h. Different shaking / incubation times were evaluated for their influence on m^* . If the

incubation time was larger than 1 hour, plates were sealed to avoid evaporation. After incubation, plates were centrifuged at 2900g for 30 minutes at 25°C thus removing aggregates from the supernatant. Samples were taken from the solution’s surface and diluted 1:10 to yield a total volume of 300 μL . Protein content was determined by measuring sample absorption at 280 nm wavelength. Plotting the logarithm of supernatant protein concentration over precipitant concentration, two linear region are discernible. At low precipitant concentration, no protein precipitation occurs, thus protein concentration is constant. At high precipitant concentrations, protein content in the supernatant decreases exponentially as described by the Cohn equation [16]. m^* values were determined as described by Hachem *et al.* [11] as the intersection of two linear fits to these two linear regions of the plot. Values were determined in triplicates at least.

2.9 Model protein characterization

All model proteins were characterized by 6 measures. GRAVY scores were calculated based on amino acid sequence described by Doolittle *et al.* [17]. GRAVY scores are a measure of protein hydrophobicity factoring in the average probability of amino acids occurring at the protein’s surface. Isoelectric points (pI) of the proteins represented a measure to characterize a proteins charge. pI were calculated using propka [18] thus factoring in the micro environment of the amino acid side chains. Molecular weight of the proteins was used as factor describing the size of the proteins. m^* values were measured as described above. These values are considered to relate to the proteins hydrophobicity [11]. Finally, two amino acid hydrophobicity scales using the distribution of uniform peptides in two polymer-polymer aqueous two-phase systems were used to calculate a surface hydrophobicity value as described by Berggren *et al.* [12]. The hydrophobicity values for each amino acid was multiplied with the amino acids relative contribution to the solvent accessible surface. The proteins hydrophobicity value was calculated as the sum of all its amino acids’ hydrophobicity values. Relative surface contributions were determined using the Yasara software after exposing the crystal structure to an energy minimization. All factors were evaluated for their correlation with protein partitioning and protein recovery results. To cover both exponential and linear correlations, each factor was evaluated as is and logarithmized. Multi-linear correlations using up to four factors were performed to evaluate an interplay of different factors.

3 Results

3.1 Model protein characterization

Table II shows the results of the model protein characterization. While the two hydrophobicity values based on peptide slopes and relative surface contribution (“PS I” and “PS II”) are highly correlated, no other factor correlates with any other factor. Thus the factors can be considered to be individual measures of protein properties. As to be expected, there is a trend towards lower solubility at pH 7.0 the closer the pI of a protein is to this value. This is reflected in the m^* values. However, there are exceptions to this rule. For example, while BSA and Amyl show almost identical m^* values, their pI differ significantly. Cytochrome C and myoglobin did not precipitate even at the highest concentration of ammonium sulfate (3.15 M). Thus no m^* value could be determined. The three measures of protein hydrophobicity did not correlate with each other and did not result in the same order.

Table II Model proteins used in this study and their parameters used for correlation with ATPS distribution. GRAVY values calculated as described by Doolittle *et al.* [17]. pI-values were calculated using propka [18]. PS I and PS II: hydrophobicity values based on solvent accessible surface as described by Berggren *et al.* [12]. ¹: Homology model based on HSA. Model obtained from Modbase [19]. ²: No precipitation.

Protein	UniProt ID	PDB Entry	GRAVY	pI	MW [kDa]	m* [M]	PS I	PS II
Cyto	P00004	1HRC	-0,88	9,79	11,8	— ²	-0,002	0,005
BSA	P02769	— ¹	-0,43	5,68	69,3	2,48	0,011	0,020
Lacto	P02754	3BLG	-0,01	4,58	19,9	2,83	0,010	0,020
Myo	P68082	2V1F	-0,38	7,83	17,1	— ²	0,036	0,050
Glucu	P08236	1BHG	-0,32	6,81	70,3	1,26	0,036	0,057
HSA	P02768	1HK1	-0,35	5,67	69,4	2,40	0,008	0,017
Chymo	P00766	2CGA	0,05	9,59	25,7	1,89	0,020	0,031
Lys	P00705	193L	-0,15	10,75	14,3	1,93	0,016	0,034
Kata	P00432	7CAT	-0,63	6,69	59,9	1,36	0,024	0,038
Amyl	P0C1B3	2GUY	-0,21	3,64	54,8	2,49	0,021	0,035

3.2 ATPS distribution and recoveries of model proteins

Table III summarizes the distribution coefficients of the model proteins in the seven classes of ATPS used in this study. The minimum, mean, and maximum value of all system in each class is given. The proteins can be roughly clustered according to their distribution, however, no definite pattern emerged. While some proteins, such as cytochrome C strongly distribute into the bottom phase under all conditions, others, such as Chymotrypsinogen A tend to distributed more evenly and switch from preferring the bottom phase to distributing more into the upper phase depending on the system composition. Some proteins, like glucuronidase, show a strong influence on the individual phase system composition, while other, like lactoglobulin are not as strongly influenced in their distribution by this. While at no addition of NaCl PEG8000-PO₄ systems showed lower logK values than PEG4000-PO₄, the addition of NaCl reversed this order. Average logK values for PEG4000-AS systems were lowest. Average logK increased with the addition of both 1.5% and 4% NaCl in PEG8000-PO₄ systems. In contrast, averaged over all model proteins, the logK value decreases slightly with the addition of 1.5% NaCl for PEG4000-PO₄ systems and increases with the addition of 4% NaCl. While a general trend of increasing distribution coefficient with the addition of NaCl can be seen, the individual proteins react significantly different to these changes. The shift upwards in distribution coefficient with the addition of NaCl depended on the individual protein. Lysozyme, one of the most commonly used model proteins, shows a very distinct behavior in that it is most strongly influence by the addition of NaCl. The relative order of protein distribution changes depending on system type and individual system composition.

Table IV summarizes the recoveries measured of the proteins in all 50 ATPSs, clustered by system type. Mean and minimum value are presented. As for the distribution coefficient, no well defined clusters were found within this data. While some proteins, like BSA or glucuronidase, showed high recoveries in all systems, for others, like catalase or chymotrypsinogen precipitation significantly reduced protein recovery. A third group, composed of proteins such as amylase and myoglobin showed recoveries strongly depending on system type and system composition. A general trend towards lower recoveries can be seen with increasing PEG molecular weight and addition of NaCl. This however was not correlated to the trends towards higher distribution coefficients described above.

5.2 Evaluation of correlations of protein properties and ATPS distribution

Table III Min mean max LogK values per protein and system type

		Cyto	BSA	Lacto	Myo	Glucu	HSA	Chymo	Lys	Kata	Amyl
I	min	-3,0	-1,2	-1,5	-1,7	-2,0	-1,5	-0,2	-0,5	-0,7	-0,4
	mean	-3,0	-0,9	-1,2	-1,5	-0,4	-1,2	0,1	-0,3	0,0	0,0
	max	-3,0	-0,6	-1,0	-1,0	0,0	-0,9	0,5	-0,1	0,2	0,7
II	min	-2,0	-2,0	-0,9	-1,3	-0,5	-0,4	-0,3	-0,9	-0,4	-0,5
	mean	-1,7	-1,2	-0,6	-1,2	-0,2	0,0	0,0	-0,7	0,1	-0,1
	max	-1,4	-0,6	-0,5	-1,0	0,1	0,5	0,4	-0,4	0,3	0,5
III	min	-3,0	-1,7	-1,0	-1,3	-0,8	-1,2	-0,8	-1,2	-0,4	-0,2
	mean	-2,1	-1,2	-0,7	-1,0	-0,3	-0,7	-0,3	-0,8	0,0	0,4
	max	-0,6	-0,8	-0,2	-0,7	0,0	-0,4	0,1	0,1	0,2	0,8
IV	min	-1,4	-1,2	-1,0	-1,0	-0,7	-0,9	-0,9	0,1	-0,3	-0,2
	mean	-1,3	-1,1	-0,9	-0,9	-0,6	-0,8	-0,8	0,2	-0,1	-0,1
	max	-1,1	-1,0	-0,7	-0,9	-0,5	-0,7	-0,8	0,5	0,1	-0,1
V	min	-2,0	-1,7	-1,0	-1,3	-0,8	-1,1	-0,6	-0,1	-0,7	-0,2
	mean	-1,6	-1,1	-0,8	-1,0	-0,4	-0,7	-0,4	0,4	0,0	0,3
	max	-0,8	-0,6	-0,4	-0,5	-0,1	-0,5	0,0	0,8	0,3	0,6
VI	min	-2,0	-1,1	-0,9	-1,0	-0,5	-0,8	0,6	1,8	-0,2	-0,4
	mean	-1,7	-0,9	-0,8	-0,9	-0,3	-0,7	0,9	2,2	-0,1	-0,3
	max	-1,4	-0,7	-0,8	-0,8	-0,1	-0,7	1,1	2,6	-0,1	-0,2
VII	min	-3,0	-1,4	-1,2	-1,2	-0,5	-0,8	0,6	1,7	0,1	-0,1
	mean	-1,8	-0,8	-0,8	-1,0	-0,2	-0,5	1,0	2,2	0,2	0,3
	max	-0,7	-0,3	-0,4	-0,4	0,0	-0,2	1,6	2,7	0,4	0,6

Table IV mean and min yield per protein and system type

		Cyto	BSA	Lacto	Myo	Glucu	HSA	Chymo	Lys	Kata	Amyl
I	mean	98	95	102	97	57	95	34	87	48	94
	min	91	85	98	92	31	89	16	30	23	66
II	mean	99	89	110	101	68	98	55	83	47	97
	min	95	79	101	99	47	93	28	25	37	88
III	mean	81	85	94	81	81	86	70	81	52	68
	min	72	76	90	73	59	82	26	29	17	33
IV	mean	96	75	95	91	61	71	95	96	69	97
	min	93	53	88	86	46	50	91	95	42	90
V	mean	75	81	81	77	77	80	90	71	44	60
	min	70	74	75	70	51	73	74	43	21	28
VI	mean	92	97	100	97	53	92	76	95	57	94
	min	86	96	98	94	42	83	57	92	44	89
VII	mean	79	82	86	83	65	80	85	98	22	62
	min	71	74	82	72	50	70	43	78	14	25

3.3 Correlation ATPS behavior with protein attributes

3.3.1 Linear regression analysis

Linear regression of the protein attributes and their distribution in each of the 50 ATPS tested were calculated. Table V shows the resulting R^2 values of the regression clustered by the seven classes of ATPS used. Both the mean R^2 value for each class and its maximum value are given. As can be seen from table V no protein attribute can be considered to linearly correlate with protein distribution based on our measurements. Maximum values of R^2 were only slightly larger than the mean value in most cases, meaning that all individual systems of a class gave very low R^2 values. The only correlations between protein attributes and logK with R^2 values greater than 0.9 were found when using the gravity score on a subset of the model proteins. The distribution of 6 model proteins, excluding both the two highest and the two lowest scoring proteins on the gravity scale, could be linearly correlated with their gravity scores in PEG-PO₄ systems to which NaCl had been added. The R^2 values of these correlations are listed under "Gravy mid" in table V and can be seen to average at 0.85. No protein attribute could be correlated with the shift in logK caused by the addition of NaCl.

Table V Mean and maximum R^2 values for linear regression of protein attribute to its logK in all 50 ATPSs investigated. Results are clustered according to the seven different classes of ATPS used.

		I	II	III	IV	V	VI	VII
Mean	Gravy	0,35	0,21	0,23	0,07	0,24	0,32	0,30
	Gravy mid	0,46	0,16	0,28	0,97	0,89	0,71	0,83
	log MW	0,25	0,33	0,26	0,03	0,05	0,02	0,01
	$\log(1/M^*)$	0,51	0,45	0,29	0,33	0,34	0,26	0,29
	PS I	0,29	0,13	0,27	0,14	0,18	0,07	0,07
	PS II	0,32	0,13	0,26	0,21	0,25	0,12	0,12
	pI	0,04	0,11	0,24	0,00	0,01	0,21	0,12
Max	Gravy	0,46	0,33	0,36	0,12	0,37	0,38	0,39
	Gravy mid	0,74	0,31	0,82	0,98	0,98	0,76	0,96
	log MW	0,31	0,53	0,45	0,06	0,08	0,03	0,02
	$\log(1/M^*)$	0,66	0,52	0,48	0,38	0,42	0,28	0,33
	PS I	0,39	0,17	0,43	0,16	0,42	0,11	0,11
	PS II	0,42	0,18	0,42	0,24	0,54	0,16	0,17
	pI	0,08	0,23	0,40	0,01	0,04	0,28	0,28

In figure 1(a), overall average distribution coefficient is plotted over the gravity score of the proteins. As described above, leaving out the two highest and two lowest gravity scores, a strong correlation can be seen. Figure 1(b) shows the overall average distribution coefficient of the ten model proteins plotted over the proteins' $\log(1/m^*)$ value. With a correlation coefficient of 0.47, only a very rough correlation between the two values can be stated.

The overall average recovery of the model proteins in the seven classes of ATPS is plotted over the $\log(1/m^*)$ in figure 2. A trend of increased average yield with increasing m^* value can be seen. With a correlation coefficient of 0.69 this trends can however only be considered indicatively. m^* values were the only hydrophobicity descriptors to be correlated to protein recovery.

As loss in recovery can be caused by a saturation of the phases, which in turn could distort the correlations, the effect of protein recovery on the resulting correlations was tested as follows. All distribution coefficients of proteins with a recovery lower than 80% were omitted from the data. Then, correlations between the remaining protein distribution

5.2 Evaluation of correlations of protein properties and ATPS distribution

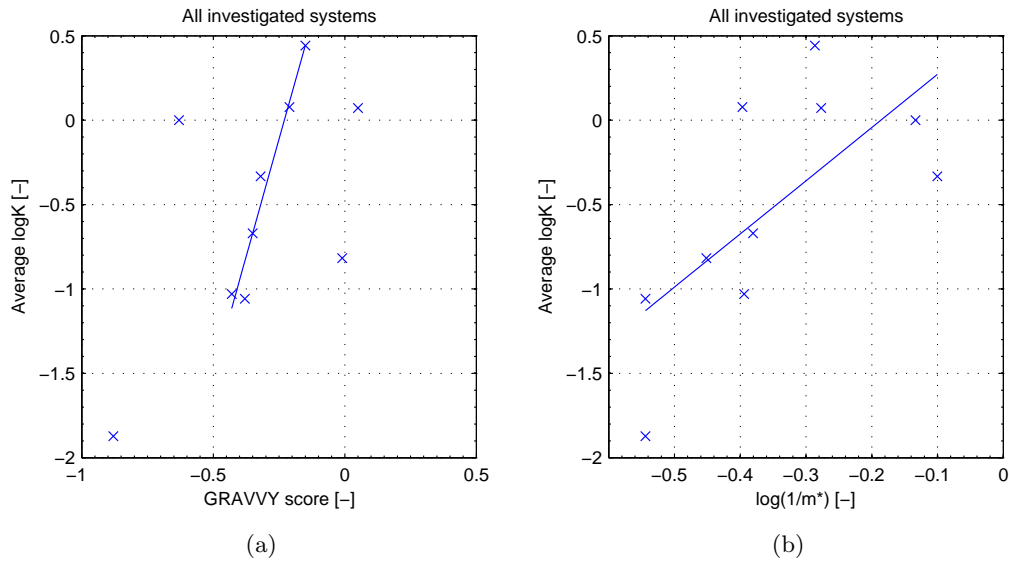


Figure 1: (a): $\log K$ of the ten model proteins averaged over all 50 ATPS used plotted against the gravity score of the proteins. While a strong linear correlation for the six medium gravity values can be seen, including all ten proteins yields a weaker correlation. The plotted line represents the linear correlation found for the medium six gravity values. (b): $\log K$ of the ten model proteins averaged over all 50 ATPS used plotted against the $\log(1/m^*)$ value the proteins. A general trend towards increasing $\log K$ values with increasing m^* values can be seen. The plotted line represents the result of a linear regression over all ten proteins.

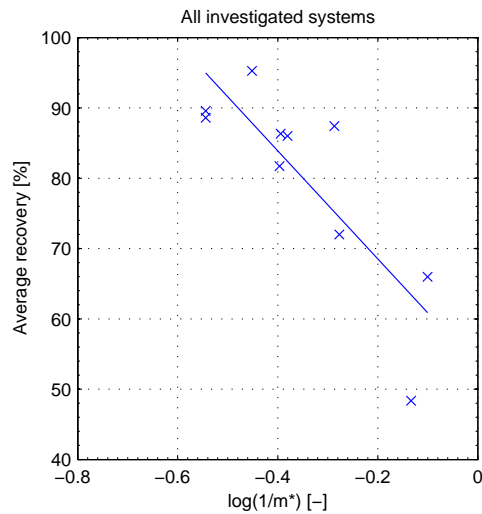


Figure 2: Recovery of the ten model proteins used in this study average over all 50 ATPS plotted against the $\log(1/m^*)$ value of the proteins. A trend towards lower recoveries with decreasing m^* value can be seen. The plotted line represents the result of a linear regression over all ten proteins.

and the protein descriptors were calculated for those systems with more than six distribution coefficients remaining. No significant improvement in the correlation coefficients was found.

3.3.2 Multi-linear regression analysis

In order to account for the interplay of multiple factors influencing protein distribution in ATPS, multi-linear regressions were performed. pI , $\log MW$, $\log(1/m^*)$, and gravity score were selected as protein descriptors. Interaction and quadratic terms were excluded from

the model as their inclusion would result in a number of factors higher than the number of results. The result of this regression analysis is summarized in table VI, in which the factors ascribed to of each descriptor are given along with their mean values and the correlation coefficients. The highest R^2 value of 0.8 was obtained for the PEG-AS systems. In general, pI and logMW were given factors close to zero in almost all models built. The average R^2 value calculated was 0.66

Table VI Results from multi-linear regressions of the distribution coefficient to four protein descriptors. The value of the factors ascribed to the individual protein descriptor are given. R^2 values of the resulting model are shown.

	const	pI	logMW	log(1/M*)	gravvy	R^2
I	-2,06	0,00	0,28	3,25	1,77	0,80
II	-0,18	-0,06	0,11	2,36	0,79	0,71
III	3,02	-0,16	-0,13	2,78	0,87	0,68
IV	4,89	-0,08	-0,38	2,87	0,09	0,44
V	4,74	-0,08	-0,32	3,15	0,75	0,54
VI	0,61	0,23	-0,07	3,06	2,25	0,77
VII	0,80	0,18	-0,03	3,45	2,20	0,68
mean	1,69	0,01	-0,08	2,99	1,24	0,66

4 Discussion

4.1 Model protein characterization

The model proteins used in this study varied in pI, size, origin, and hydrophobicity. They represent a commonly used set of model proteins. Model proteins were characterized by 6 different scales, containing three different measures of hydrophobicity. The selected protein descriptors were taken from published examples of successful correlations to protein distribution. Our aim was to validate these correlations by increasing both the number of proteins and systems used. Gravy value take into account the entire amino acid sequence but factor in the probability of an amino acid to occur at the protein’s surface. PS values used experimentally determined hydrophobicity values for each amino acid and factor in the actual contribution of an amino acid to the protein’s surface via evaluation of the 3D structure. m^* values are experimentally determined hydrophobicity values based on ammonium sulfate precipitation. While all three scales represent a protein’s hydrophobicity, they were not correlated. Proteins were ranked differently by all three scales. This underlines how challenging the task to characterize a proteins hydrophobicity is. A protein’s hydrophobicity depends not only on the amino acid sequence and its 3D structure, but also on the surrounding solution. Other hydrophobicity scales based on amino acids retention times in hydrophobic interaction chromatography or amino acid distribution in aqueous-organic systems have been evaluated for their correlation with protein distribution in ATPS. As they did not yield good correlations, they were not included in this study.

4.2 ATPS distribution and recoveries of model proteins

Using an HTS method previously described, we were able to characterize the proteins behavior in seven types of ATPS, with 50 individual system compositions, thus generating a large dataset to validate correlations between protein distribution and protein descriptors in a short amount of time using little material. Protein distribution data did not show any

obvious clusters. Proteins reacted differently to the change in system composition and the addition of salt. This already indicates that finding correlations between protein descriptors and its behavior in ATPS will be very challenging as the descriptors were constant, but the order of proteins changes.

Protein recovery is a major issue in ATPS as the high concentrations of the phase forming component can decrease protein solubility to a level at which an industrial application would be inviable [20]. The total protein concentration of 1 g/L is a commonly used level and has previously been used to evaluate correlations between protein descriptors and distribution [9]. The recoveries measured in this dataset show a decreased recovery for some model proteins. This reduction is caused by protein precipitation, which often occurs when both phases of an ATPS are saturated with the protein [21]. A saturation of the phases can lead to a distortion between measured and real distribution coefficient. This might lead to a change in the observed correlations. From a practical standpoint however, models that are only applicable to very low protein concentrations are of little value for the implementation of an industrial separation step.

4.3 Correlation ATPS behavior with protein attributes

The distribution data was evaluated for correlations to protein descriptors. While several trends can be seen within the data, no strong correlations were found. This is in contrast to previously published studies showing correlations between protein molecular weight, protein solubility, or protein hydrophobicity. Some of these studies used polymer-polymer ATPS, which might explain the discrepancy to our results. In polymer-polymer ATPS electrostatic forces and molecular weight likely play a more pronounced role than in polymer-salt systems. The disagreement with our results with a study published by Hachem *et al.* [11] is remarkable as the authors also used PEG4000 and PEG8000 phosphate systems. The discrepancy seems to arise from the difference in model protein selection and phase system composition. If however a correlation can only be found for specific phase compositions and sets of model proteins, it will most likely not be applicable to model protein distribution more generally. For example, we found a very good correlation between the gravity score and the distribution coefficient of a subset of 6 model proteins in PEG-phosphate system with added NaCl. However, while using only a subset of the model proteins resulted in high R^2 values, the validity of excluding the remaining four model proteins is questionable. As the proteins with both the highest and lowest Gravity scores were excluded, one might argue that the correlation found for the subset of proteins is valid only in a small range of the gravity scale. However, the gravity scores of the excluded proteins cannot be considered extreme values on this scale when taking the overall distribution of this value into account.

One trend that was seen for the recoveries of the model proteins was that proteins of lower m^* value generally showed lower recoveries. Both precipitation and ATPS distribution experiments were conducted at pH 7.0. Proteins with a lower solubility against ammonium sulfate can be expected to also show lower recoveries in the presence of high concentrations of PEG and phosphate. The trend towards lower recoveries with decreasing m^* value is thus not surprising. Multi-linear regression using four protein descriptors were attempted. While R^2 values were generally low, there was an agreement between all models that the coefficients ascribed to the pI and MW values were close to zero, while those ascribed to protein hydrophobicity descriptors were significantly different from zero. This is in agreement with previous reports that stated that electrostatic and size exclusion effects do not play as an important role in polymer-salt systems as in polymer-polymer systems.

To be applicable to a new separation problem, correlations between a protein descriptors and its distribution need to be strong and generalizable. Such correlations are most likely to be found if there is actual causation underlying the correlation. We doubt that any

correlation reported so far for ATPS can actually claim this. A protein's distribution in ATPS is most likely the result of a complex interplay of both protein and phase system characteristics which are not yet fully understood. If correlations can only be found for certain sets of proteins or specific system compositions, while indicative to the distribution mechanism at work, they cannot be applied with certainty to a different protein and / or system composition. As we show in this work, protein descriptors alone or in combination do not suffice. It is thus crucial to not only characterize the protein, but also the two phases of each system and the proteins susceptibility to the phases' attributes. In a recent publication Madeira *et al.* [22] start to work along these lines. Investigating distribution of 12 model proteins in three classes of polymer-polymer ATPS, with ten individual system composition in total, they were able to present a linear correlation between phase descriptors and protein distribution similar to the approach presented by Yizhak *et al.* [23] for aqueous-organic systems. The phases were characterized using solvatochromic dyes. The difference of these descriptors between the two phases was used to correlate each proteins distribution. Thus a susceptibility of each protein to the difference in phase descriptors was established. In order to better understand the actual mechanisms behind protein distribution, more systems including salt-polymer ATPSs now need to be investigated. Understanding the proteins characteristics that lead to its susceptibility to the phase characteristics might then further advance the possibilities of predicting protein behavior in ATPS.

5 Conclusion and outlook

In this manuscript, we evaluated correlations between protein attributes and protein behavior in aqueous two-phase systems. All evaluated protein attributes had previously shown to yield good correlations to protein distribution in limited sets of model proteins and ATPSs. However, none of the correlations described could be reproduced with the larger sets of model proteins and increased number of biphasic systems used in this study. While most previous publications included only small numbers of different two-phase systems, we tested a large set of 50 systems and 10 model proteins and found that no protein attribute neither in linear nor in multi-linear regression gave promising results. While a large number of protein attributes can readily be obtained either *in silico* or from experiments, it seem obvious, that factors describing the two phases of an ATPS are missing. Only limited work has been conducted trying to relate phase describing attributes to protein behavior in ATPS. As our results clearly show, using protein attributes alone is insufficient and future research will focus on combining phase and protein attributes to better model protein behavior in ATPS.

6 Acknowledgments

The data presented in this manuscript are the results of the work done by Dipl.-Ing. Johannes Knoll during the course of his diploma thesis. The authors gratefully acknowledge financial support by Boehringer Ingelheim Pharma GmbH & Co KG.

References

- [1] F. Dimer, J. J. Hubbuch, 3d structure-based protein retention prediction for ion-exchange chromatography., *Journal of chromatography. A* 1217 (8) (2010) 1343–53.
- [2] M. Rito-Palomares, Practical application of aqueous two-phase partition to process development for the recovery of biological products, *J. Chromatogr. B* 807 (2004) 3–11.

5.2 Evaluation of correlations of protein properties and ATPS distribution

- [3] J. Benavides, M. Rito-Palomares, Practical experiences from the development of aqueous two-phase processes for the recovery of high value biological products, *Journal of Chemical Technology & Biotechnology* 83 (2) (2008) 133–142.
- [4] P. Albertsson, Partition of cell particles and macromolecules: distribution and fractionation of cells, mitochondria, chloroplasts, viruses, proteins, nucleic acids, and antigen-antibody complexes in aqueous polymer two-phase systems, Wiley-Interscience, 1971.
- [5] H. Walter, D. Brooks, D. Fisher, Partitioning in aqueous two-phase systems: theory, methods, uses, and application to biotechnology, Academic Press, 1985.
- [6] R. Hatti-Kaul, Aqueous two-phase systems: methods and protocols, *Methods in biotechnology*, Humana Press, 2000.
- [7] S. Sasakawa, H. Walter, Partition behavior of native proteins in aqueous dextran-poly(ethylene glycol)-phase systems., *Biochemistry* 11 (15) (1972) 2760–5.
- [8] G. Johansson, A. Sarnesto, E. Høge-Jensen, I. Szabo-Lin, C. Guthenberg, B. Manervik, Effects of salts on the partition of proteins in aqueous polymeric biphasic systems., *Acta Chemica Scandinavica* 28b (1974) 873–882.
- [9] J. Asenjo, A. Schmidt, F. Hachem, B. Andrews, Model for predicting the partition behaviour of proteins in aqueous two-phase systems, *Journal of Chromatography A* 668 (1) (1994) 47–54.
- [10] T. M. Przybycien, J. E. Bailey, Aggregation kinetics in salt-induced protein precipitation, *AIChE Journal* 35 (11) (1989) 1779–1790.
- [11] F. Hachem, Hydrophobic partitioning of proteins in aqueous two-phase systems, *Enzyme and Microbial Technology* 19 (7) (1996) 507–517.
- [12] K. Berggren, A. Wolf, J. A. Asenjo, B. a. Andrews, F. Tjerneld, The surface exposed amino acid residues of monomeric proteins determine the partitioning in aqueous two-phase systems., *Biochimica et biophysica acta* 1596 (2) (2002) 253–68.
- [13] M. Bensch, B. SELBACH, J. J. Hubbuch, High throughput screening techniques in downstream processing: Preparation, characterization and optimization of aqueous two-phase systems, *Chemical Engineering Science* 62 (7) (2007) 2011–2021.
- [14] S. A. Oelmeier, F. Dimer, J. Hubbuch, Application of an aqueous two-phase systems high-throughput screening method to evaluate mab hcp separation., *Biotechnology and bioengineering* 108 (1) (2011) 69–81.
- [15] E. Krieger, G. Koraimann, G. Vriend, Increasing the precision of comparative models with yasara nova—a self-parameterizing force field., *Proteins* 47 (3) (2002) 393–402.
- [16] E. J. Cohn, The physical chemistry of the proteins, *Physiological Reviews* 5 (3) (1925) 349–437.
- [17] J. Kyte, R. F. Doolittle, A simple method for displaying the hydropathic character of a protein., *Journal of molecular biology* 157 (1) (1982) 105–32.
- [18] H. Li, A. D. Robertson, J. H. Jensen, Very fast empirical prediction and rationalization of protein pka values., *Proteins* 61 (4) (2005) 704–21.

5.2 Evaluation of correlations of protein properties and ATPS distribution

- [19] U. Pieper, B. M. Webb, D. T. Barkan, D. Schneidman-Duhovny, A. Schlessinger, H. Braberg, Z. Yang, E. C. Meng, E. F. Pettersen, C. C. Huang, et al., Modbase, a database of annotated comparative protein structure models, and associated resources., *Nucleic acids research* 39 (2011) D465–74.
- [20] P. a. J. Rosa, Application of central composite design to the optimisation of aqueous two-phase extraction of human antibodies, *J. Chromatogr. A* 1141 (2007) 50–60.
- [21] A. Schmidt, A. M. Ventom, J. A. Asenjo, Partitioning and purification of α -amylase in aqueous two-phase systems, *Enzyme and Microbial Technology* 16 (2) (1994) 131–142.
- [22] P. P. Madeira, C. a. Reis, A. E. Rodrigues, L. M. Mikheeva, A. Chait, B. Y. Zaslavsky, Solvent properties governing protein partitioning in polymer/polymer aqueous two-phase systems, *Journal of Chromatography A* 1218 (10) (2011) 1379–1384.
- [23] Y. Marcus, Linear solvation energy relationships. correlation and prediction of the distribution of organic solutes between water and immiscible organic solvents, *The Journal of Physical Chemistry* 95 (22) (1991) 8886–8891.

MANUSCRIPT III

MOLECULAR DYNAMICS SIMULATIONS ON
AQUEOUS TWO-PHASE SYSTEMS: SINGLE
PEG-MOLECULES IN SOLUTION

Stefan A. Oelmeier, Florian Dismer, Jürgen Hubbuch*

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

* Corresponding author. Tel.: +49 721 608-42557; fax: +49 721 608-46240. E-mail address: juergen.hubbuch@kit.edu

+ These authors contributed equally to this publication

submitted to BMC Biophysics Journal

Abstract

Molecular Dynamics (MD) simulations are a promising tool to generate molecular understanding of processes related to the purification of proteins. Polyethylene glycols (PEG) of various length are commonly used in the production and purification of proteins. The molecular mechanisms behind PEG driven precipitation, aqueous two-phase formation or the effects of PEGylation are however still poorly understood. In this paper, we ran MD simulations of single PEG molecules of variable length in explicitly simulated water. The resulting structures are in good agreement with experimentally determined 3D structures of PEG. The increase in surface hydrophobicity of PEG of longer chain length could be explained on an atomic scale. PEG-water interactions as well as aqueous two-phase formation in the presence of PO_4 were found to be correlated to PEG surface hydrophobicity. We were able to show that the taken MD simulation approach is capable of generating both structural data as well as molecule descriptors in agreement with experimental data. Thus, we are confident of having a good *in silico* representation of PEG.

Keywords: *Molecular dynamics simulation, polyethylene glycol, hydrophobicity, aqueous two-phase systems, amber03, yasara*

1 Introduction

Polyethylene glycol (PEG) is among the most commonly used chemicals in protein purification processes. Its use ranges from precipitating agent [1, 2], phase forming component in liquid-liquid extraction [3–6], displacer in hydrophobic interaction chromatography [7] to potential additive in protein formulation [8, 9] and drug modifying agent [10, 11]. Despite its frequent application, the exact nature of the polymer solvent interactions as well as the structural dynamics in solution governing its role in protein purification processes remain unclear.

One approach to understand the nature of molecules on an atomic scale are molecular dynamics (MD) simulations. MD simulations have proven a useful tool to understand the binding and elution behavior of proteins in ion exchange chromatography [12]. Its results could be correlated to experimental data and thus used predictively. Besides HTS process development methodologies, MD simulations represent another promising approach to speed up process development in protein purification.

Several molecular dynamics studies have been performed focusing on the tertiary structure of polyethylene glycol or polyethylene oxide in aqueous solutions. Lee *et al.* in 2008 [13] used a revision of the CHARMM ether force field (C35) [14] to establish a molecular understanding of the relation between hydrodynamic radius and radius of gyration of PEG, which differs in behavior from polymer theory under certain conditions. They extended their study in 2009 [15] by using a coarse grained model and were thus able to extend the simulated timespan from 20 to 800 ns. These studies were again focused on investigating the hydrodynamic properties of the polymer molecules, their hydrodynamic radius in relation to their radius of gyration and the coherence of the simulation results with polymer theory. One major goal was to understand - on a molecular scale - the influence of the three dimensional structure of polymer molecules in solution on macroscopic, hydrodynamic properties of these solutions e.g. viscosity. Other studies, such as the one conducted by Borodin *et al.* in 2001 [16] are geared towards polymer dynamics of pure polymers and the effect of added water. One main goal of this study was again to understand the physicochemical properties of polymer solutions such as viscosity.

Tasaki *et al.* performed MD simulation of a single PEG molecule of 722 Da in 1996 [17] focused more on the secondary structure and polymer-water interaction. While their study yielded interesting insights into the 3D structure of single PEG molecules in solution, the

advance in computational power since their publication now allows for longer simulations of larger PEG molecules. Another, more recent, MD simulation study using a PEG-derivate focused the importance of choosing a proper force field [18] to obtain secondary structure results in agreement with experimental data. One force field, not tested in their study, but commonly used for protein simulations, is the amber03 force field [19]. While this force field is designated to proteins, a self-parameterizing algorithm can be applied to adjust for non proteinaceous molecules [20]. The advantage of using a force field designated to proteins is that combined MD simulations of proteins and PEG could be readily run, once it is established that the self-parameterizing algorithm yield reasonable results for PEG. In order to verify the structural data resulting from such MD simulations, experimentally determined structures are needed. There have been several reports of 3D structural data of PEG [21] with the recent addition of a high quality crystallographic record [22].

As stated above, MD simulations focused on the secondary structure of single PEG molecules of fixed molecular weight as well as studies concentrating on the three dimensional structure and its implications for macroscopic solution properties have been conducted. Simulations of larger PEG molecules and an investigation of the effects of polymer molecular weight on properties of PEG such as secondary structure, surface hydrophobicity and H-bond network with the solvent are lacking. These properties most likely govern the characteristics of PEG molecules in protein purification, thus their deeper understanding on a molecular scale would be beneficial. PEG polymer molecular weight influences the effect PEG has on protein solubility or aqueous two-phase formation. To understand the influence of molecular weight of PEG molecules on their structural and surface characteristics on an atomic scale would thus be valuable.

This study's first aim is to confirm the validity of all-atom molecular dynamics simulations of PEG in water based on the amber03 force field by comparing the resulting structure to the recently published crystallographic data and previous MD results. Results from 10-30 ns long MD simulations of PEG molecules ranging in molecular weight from 300 Da to 3500 Da are shown. This study puts a focus on the effect PEG chain length has on geometric parameters, secondary structure, polymer surface characteristics and polymer solvent interactions. We aim to understand the influence on PEG molecular weight on polymer properties important for its use in protein purification. Correlations to experimental data on the phase forming behavior of 4 different PEGs in the presence of phosphate are given. Correlation between solvent polarity determined using a solvatochromic dye to hydrophobicity determined via MD simulations are shown.

2 Materials & methods

2.1 Molecular Dynamics simulation software

Molecular dynamics simulations were performed using the Yasara Structure software package [23], version 10.2.1. The software was installed in a cluster computer environment running Suse Linux Enterprise 10 as operating system. Each simulation was run using a single node of the cluster computer. Each node was equipped with 2 Intel Xeon Quad Core (X5355) processors at 2.66 GHz and 16GB local memory (2GB per processor core).

2.2 Force field

The software package used for the simulations employs an automatic parameterization algorithm (termed "AutoSMILES") to generate force field parameters for unknown structures. The method is described in detail in [19, 20]. This method was used to generate force field parameters for the polymer molecules. Van der Waals forces were truncated at a cutoff

5.3 Molecular dynamics simulations on aqueous two-phase systems: Single PEG-molecules in solution

of 10 Å. Long range Coulomb interactions were calculated using the Particle Mesh Ewald algorithm detailed by Essmann *et. al* [24]. Grid point for the PME evaluation were evenly spaced in each dimension. The amber03 force field [25] was used for all molecular dynamics simulations. As water model, TIP3P was used. The amber03 force field was chosen as following studies were to include protein molecules in the simulation and amber03 is a well-established force field for protein simulations.

2.3 MD simulation protocol

All-atom molecular dynamics simulations were run for PEGs with subunit numbers (n) between 6 and 81 (6, 7, ..., 21 subunits; 21, 23, ..., 41 subunits, and 41, 45, ..., 81 subunits). Each simulation consisted of a single PEG molecule surrounded by water molecules. Number of atoms simulated ranged from 1572 for n=6 to 17211 for n=81. PEG molecules were imported into the simulation software using an OpenBabel plug-in and SMILES file format. Thus, each simulation initially contained a perfectly linear PEG molecule with all geometric parameters set to standard values. A simulation box was put around the PEG molecule with a distance of 5 nm. The simulation box's size ranged from 28.33 x 24.23 x 23.59 Å for n=6 to 309.14 x 23.59 x 23.59 Å for n=81. Boundary conditions were set to periodic. Water density was set to 1 g/cc, temperature to 298K. The simulation box was (automatically) filled with water molecules to the set density and a short steepest decent energy minimization was run. Finally, a 5 ns long molecular dynamics simulation was started and snapshots of the simulation taken every 5 ps. Table I summarizes the parameters set for these simulations.

Table I MD simulation parameter used in this study.

Parameter	Value
Force-Field	Amber03 [25–27]
Wall boundaries	periodic
Simulation time	10-30 ns
Snapshot interval	5 ps
Density [g/l]	1.0
pH	7.0
Temperature control	rescale velocities [28, 29]

2.4 MD snapshot data evaluation

The PEG molecule in each simulation snapshot was analyzed for a series of parameters. The following geometric parameters were recorded: dihedral angle between adjacent oxygen atoms (“CC-dihedral”), dihedral angle between adjacent oxygen and carbon atoms (“OC-dihedral”), bond angles (“COC”-, “OCC”-, “CCO”-, and “HCH”-angle), bond length (“OC”-, “CC”-, and “CH”-bond length). Additionally, H-bonds (total number, number donated, number accepted, and total energy), surface characteristics (solvent accessible surface (“SAS”) of CH-groups, O-atoms, and OH-groups), and solvent accessible volume (“SAV”, 1.4Å probe radius) were measured.

2.5 Reference structure generation

Recently published crystallographic data of a PEG molecule [22] was used to build reference structures. The geometric parameters (dihedral angles, bond angles, bond lengths) of the crystallographic record were measured and PEG-molecules with 6 to 81 subunits

5.3 Molecular dynamics simulations on aqueous two-phase systems: Single PEG-molecules in solution

constructed accordingly. Surface characteristics and solvent accessible volume of these molecules were measured as described for the MD snapshots.

2.6 Data evaluation

Data evaluation was performed using Matlab[®] (The Mathworks[™], Natick, ME, USA), version 2010a. All geometric parameters were first analyzed for a trend over simulation time. If no dependency on simulation time was found, the mean value over all snapshots was used. If no dependency on the position within the molecule was found, the mean value across the molecule was used. Finally, if no dependency on PEG-length was found, the mean value over all PEG-length was calculated. To assess the degree of secondary structure formation of the simulated PEG molecules, regions of four and more consecutive CC-dihedral angles within the “gauche-conformation range” of 50° to 100° or -50° to -100° were considered to form a helical structure. The ratio of CC dihedrals forming a helical structure over the total number of CC dihedrals in the molecule was taken as a measure for the helicality of the molecule. Mean helicality over all snapshots were calculated for each PEG-length simulated. Curve fitting was done using the non-linear-least-square method for fitting and the least-average-residual (“LAR”) method for robustness.

2.7 Binodal data generation

Experimentally determined binodal curves were obtained using the approach previously described [30]. In short, a Tecan (Crailsheim, Germany) Freedom EVO[®] 200 robotic platform was used to determine phase transition points using a “cloud point method” modified for liquid handling platforms. PEG used were obtained from Sigma-Aldrich (Product numbers were: PEG300: 202371 - PEG600: 202401 - PEG1000: 202428 - PEG1500: 81214). 40%[w/w] PO_4 stock solution at pH 7.0 was composed of 11.9 g sodiumdihydrogenphosphate and 28.1 g di-potassiumhydrogenphosphate per 100 g of solution.

2.8 Polarity measurements

Determination of empirical $E_T(30)$ values were done using the solvatochromic dye Nile Red (Sigma Aldrich, Steinheim) by measuring the shift of the last absorbance maximum between 500 and 650 nm with a Perkin Elmer Lambda 35 spectrophotometer [31]. The following PEGs were investigated: PEG 200, PEG 400, PEG 600, PEG 1000 and PEG 1450 (all purchased from Sigma Aldrich, Steinheim as stated in the previous section). Samples were prepared for different molar fractions of PEG in water with the investigated range depending on the PEG molecular weight (0-44.2% for PEG 200, 0-21% for PEG 400, 0-15.8% for PEG 600, 0-7.8% for PEG 1000 and 0-3.3% for PEG 1450). 2 μl of a saturated solution of Nile Red in 75% v/v acetonitrile were added to 1ml of sample solution. Calibration of the $E_T(30)$ scale was done by plotting $E_T(30)$ values of organic solvents (ethanol and different dilutions of ethanol in water, isopropanol, acetone, toluol and octanol) over the wavelength of the absorbance maximum of Nile Red. $E_T(30)$ values were taken from literature [32]. Data was fitted with a 2nd degree polynomial function.

3 Results

3.1 PEG primary structure in solution

All simulations were started with a linear PEG molecule, with all geometric parameters set to standard values by the simulation software. Table II shows average geometric parame-

5.3 Molecular dynamics simulations on aqueous two-phase systems: Single PEG-molecules in solution

ters resulting from the MD simulations in comparison to those obtained from the crystal structure [22]. While tetrahedral angles showed almost no deviation (maximum deviation 0.45%) between the crystal structure and the structure found by our simulations, bond length and dihedral angle differ by 1.4% - 4.1% and 2.0% - 4.1% respectively. CC dihedral and OC distance showed a dependency on the number of PEG subunits. CC dihedral angle increased by 1.02% over the range of PEG subunits investigated herein (see figure 3(b)). The increase in OC bond length was 0.14%.

Table II Comparison of geometric parameter between crystal structure of PEG [22] and MD simulation results. OC distance and OCCO dihedral showed dependency on PEG-length. All other parameters were averaged over all PEG lengths. Dihedral values are given for the dominating conformation. Average dihedral of gauche(+) and gauche(-) conformation is given for MD simulation results.

Parameter	Crystal	MDs
HCH angle [°]	109,01	108,97
CCO angle [°]	109,96	109,87
COC angle [°]	115,64	115,09
OCC angle [°]	109,49	109,87
OC distance [Å]	1,43	1,45-1,452
CC distance [Å]	1,48	1,54
CC dihedral <i>gauche</i> [°]	-74,95	-73,43-74,18
OC dihedral <i>trans</i> [°]	177,41	170,13

3.2 Secondary structure

PEG secondary structure was mainly governed by the CC dihedrals settling into gauche conformation. While the average CC dihedral was found in good agreement to the crystal structure (see table II), its value increased and the ratio of CC dihedral in gauche conformation decreased with PEG chain length (see figures 3(b) and 3(d)). Concurrently, helicality, defined as the ratio of dihedrals partaking in stretches of more than three dihedrals of equal conformation, increased with PEG chain length (see figure 3(c)). Figure 4 shows a rendering of a simulation snapshot of a PEG550 (n=11). Helical regions and the surface accessible volume (“SAV”) are noted. In agreement with other publication, the PEG molecule adapts a random-coil conformation with the CC dihedrals preferring a gauche conformation. Stretches of CC dihedral in gauche conformation of equal sign form helical regions within the random coil.

3.3 Tertiary structure

As mentioned in the previous section, the PEG molecules were found to settle into a random-coil conformation in agreement with previously published studies. While the fluctuation of total system energy suggested to have reached an equilibrium after 1 ns of simulation time, values of R_g reached equilibrium only after 25 ns for the longest PEG molecules. In figure 1 the $\log R_g$ is plotted over $\log PEG_{MW}$. A linear fit using the least average residual algorithm was calculated and is shown in figure 1. The radius of gyration R_g was found to correlate with the molecular weight with an exponent v for $R_g \propto M_w^v$ not significantly different from 0.5. This is in agreement with polymer theory for molecules within the tested range of molecular weights [33, 34] when ideal chain behavior, or a theta solvent respectively, are assumed. While the exponent v was found not to be significantly

5.3 Molecular dynamics simulations on aqueous two-phase systems: Single PEG-molecules in solution

different from 0.5 the 95% confidence interval accommodates values as high as 0.535 thus also allowing for a divergence towards real-chain behavior. As helical coils were observed as a structural element in the secondary structure, a divergence from the ideal chain model would be reasonable. Longer simulation times might be needed to average out structural fluctuations of R_g thus narrowing the confidence interval around v . However, as mentioned above, as the aim of this study was not to infer from structure to macroscopic solution properties, the fluctuations of R_G are acceptable within the scope of this work. Solvent accessible volume (“SAV”) of the artificial linear PEG helices and the equivalent average volume of the PEG molecules from MD simulations did not differ significantly. SAVs of simulated PEG molecules were normally distributed around their mean with an average coefficient of variation (“CV”) of 2.4 %. It was concluded, that, while the volumes of PEG molecules in our simulations change dynamically with time, no stable tertiary structures excluding the solvent were formed. While the observed helical regions influenced surface hydrophobicity, they did not significantly change the accessible volume of the molecules.

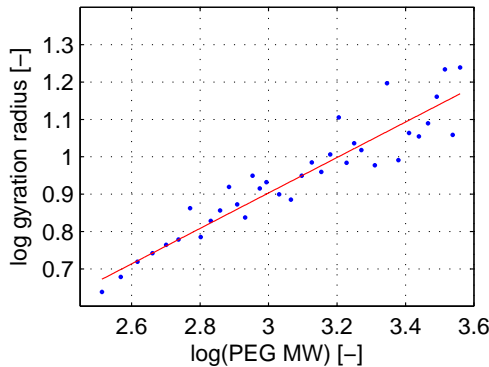


Figure 1: Dots: $\log R_G$ plotted over $\log PEG_{MW}$. Line: Linear fit using least average square algorithm. The slope of the linear fit is not significantly different from 0.5.

3.4 Structure dynamics

Figure 2 highlights three aspects of the structural dynamics observed over the course of the simulations. Figure 2(a) shows the value of two CC-dihedrals arbitrarily selected from a PEG1162 ($n=25$) over the entire course of a 5 ns simulation. The two angles are in gauche(-) or gauche(+) conformation in the vast majority of the snapshots. The angles flip irregularly between the (-) and (+) conformation. While the angles are either gauche(-) or gauche(+) most of the time, regions where both angles have the same conformation are more limited. Helical structures form in regions where multiple consecutive CC-dihedrals are of the same conformation. Figure 2(b) shows the normalized PEG dihedral energy of three PEGs of different molecular weight over the course of 0.5 ns of simulation. The higher the molecular weight of the PEG, the longer it took to reach equilibrium. Figure 2(c) shows the normalized system energy of three PEGs of different molecular weight over the course of 1 ns of simulation. All three simulations show the same trend and reach equilibrium within this time frame. Scattering of the data is lower, the higher the molecular weight of the PEG. Figure 2(d) shows the radius of gyration plotted over the simulation time of four exemplary PEG molecules. It can be seen, that the tertiary structure of the PEG reached equilibrium within the simulated timespan. Longer PEG molecules needed longer simulation times. The range over which average properties were calculated is detailed in figure 2(d). It can be seen, that the tertiary structure is in equilibrium in the time span

which was used for the calculation of average PEG properties. Both figure 2(b) and figure 2(c) suggest, that the flexibility of the PEG molecule is dependent on its molecular weight.

3.5 Surface hydrophobicity

As no tertiary structure excluding the solvent was formed during our simulations, it was concluded that the effect of PEG molecules on their surrounding solvent is governed by its solvent accessible surface. Thus, the influence of PEG chain length on the surface characteristics were investigated. Figure 3(a) shows the dependency of the surface hydrophobic fraction on the number of PEG subunits. Surface hydrophobic fraction was defined as the solvent accessible surface of CH-groups in relation to the total surface. Three dataset are compared in this figure: first, the results from the MD simulations; second, hypothetical, entirely helical PEG molecules constructed with the geometric parameters obtained from the crystal structure as describe in the section 'Reference structure generation', and third, hypothetical, entirely linear PEG molecules, constructed in the same way as the helical structures, but with dihedral angles all set to 180° . Three effects are discernible. First, there is an offset between surface hydrophobicity of the helical and the linear structure. It was concluded, that the degree of surface hydrophobicity is strongly influenced by the secondary structure of the PEG molecule. Within helical regions, oxygen atoms face inwards, away from the solvent, while CH-groups are turned towards the solvent, thus increasing surface hydrophobicity (see figure 4(a)). It should be noted that due to the compaction of the overall structure, both the hydrophobic and the hydrophilic surface area decrease in the process of helix formation. However, hydrophilic surface area decreases more than the hydrophobic surface area thus making the overall surface more hydrophobic. Second, all three curves show a dependency of the surface hydrophobic fraction on the number of PEG subunits. This effect is mostly independent of the secondary structure. By comparing the hydrophobicity of the end- and mid-groups of the various PEGs, the effect could be ascribed to the "dilution" of the end group with increasing PEG chain length. The end groups are less hydrophobic than the repetitive units in the middle of the molecule. Thus, surface hydrophobicity increases the more of these repetitive units exist. Third, surface hydrophobic fraction of the MD simulations are generally below those of the entirely helical structures. At higher numbers of subunits, the values start to converge. To explain this trend, several underlying effects need to be looked at. As helicality of a PEG molecule has a major effect on its surface hydrophobicity, and helicality increases with number of PEG subunits (see figure 3(c)), it is reasonable that MD simulation results close in on the surface characteristics of a perfectly helical molecule with increasing molecular weight. While helicality increases, helix angle also increases (figure 3(b)) and fraction of CC dihedrals in gauche conformation decreases (figure 3(d)), with both effects decreasing surface hydrophobicity. Additionally, as shown in figure 4(b), during the MD simulations, temporary structures can be formed that hardly change the solvent accessible volume but strongly reduce the solvent accessible surface of certain parts of the molecules. These regions were found randomly distributed over time and place in the MD simulations and contribute to a decrease in surface hydrophobicity as they only affect mid-groups. The relation of surface hydrophobicity to molecular weight were the result of the combination of all the underlying effects mentioned above.

3.6 PEG-solvent interaction

To quantify the influence of the change in surface hydrophobicity on the interaction of the PEG molecule with their surrounding solvent, average number of H-bonds per PEG subunit were plotted over PEG chain length and surface hydrophobicity in figure 5. Average

5.3 Molecular dynamics simulations on aqueous two-phase systems: Single PEG-molecules in solution

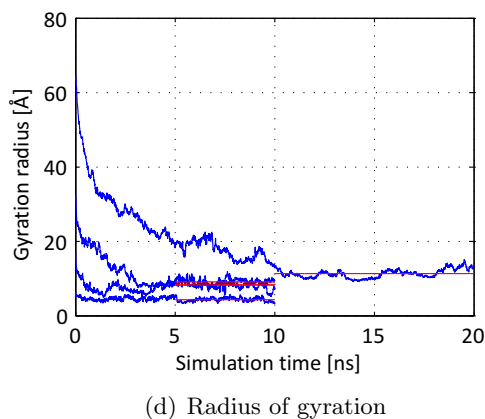
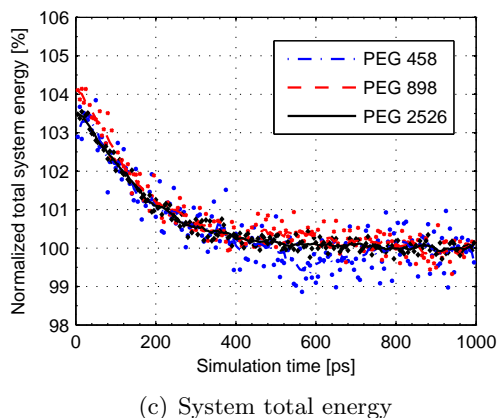
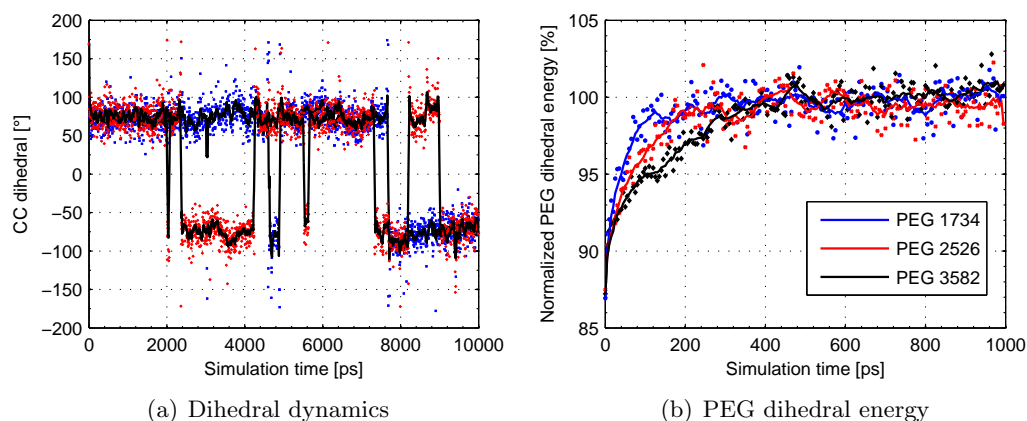


Figure 2: Structural dynamics resulting from the MD simulation. (a): Two dihedral angles of a PEG1162 over simulation time. (b): Dihedral energy of three selected PEGs over simulation time. (c): System total energy over simulation time. Single data points as well as smoothed line calculated over 9 data points are shown. The number of subunits is given in the legends. (d): Radius of gyration over simulation time. Blue lines show raw data of (from top to bottom) PEG2746, PEG1426, PEG766, and PEG326. Red lines show the timespan over which average properties of the polymer molecule were calculated.

5.3 Molecular dynamics simulations on aqueous two-phase systems: Single PEG-molecules in solution

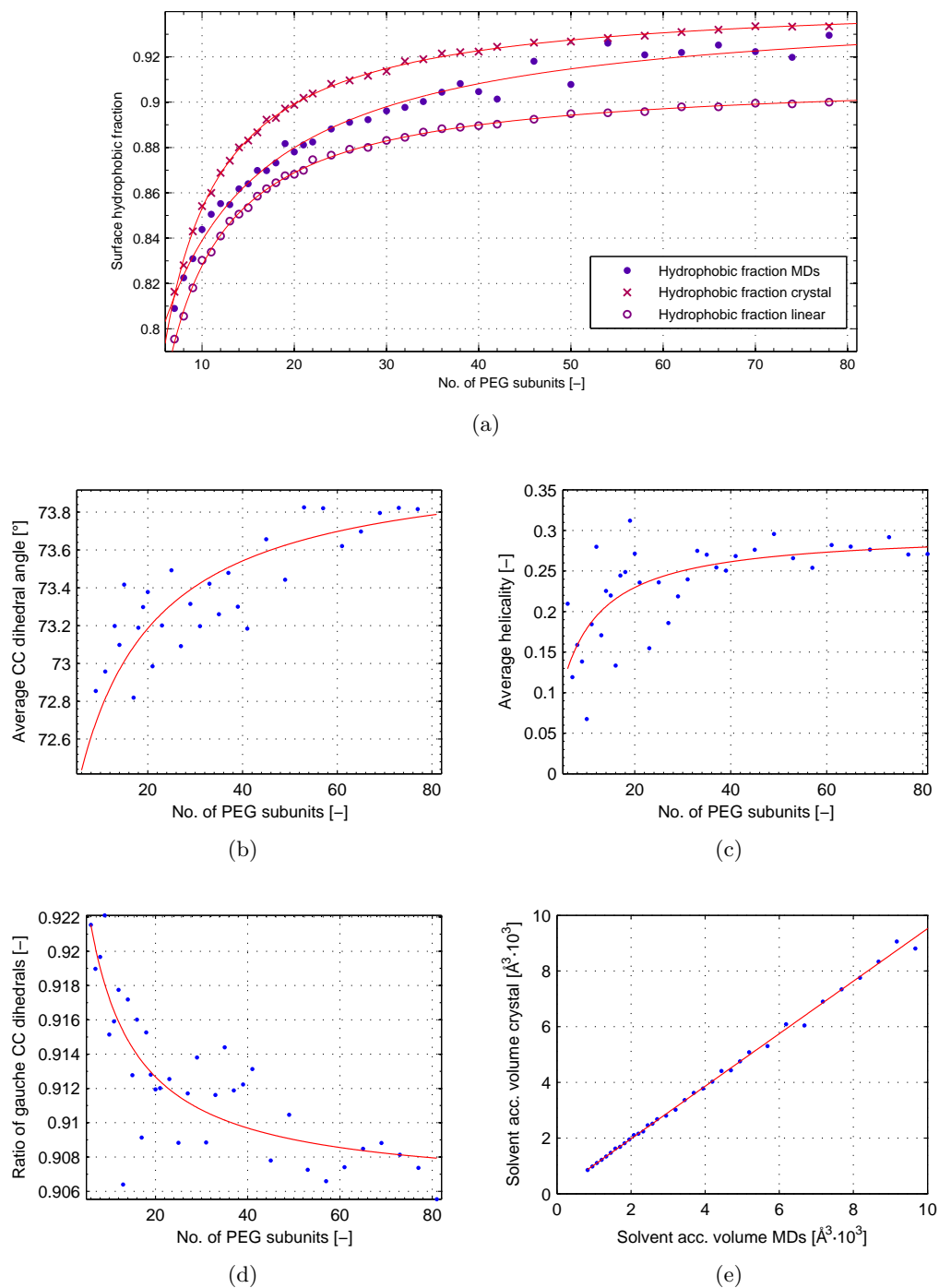


Figure 3: (a): Surface hydrophobic fraction of artificially constructed perfectly helical, perfectly linear PEGs and results of the MD simulations over PEG chain length. (b)-(d): factors influencing surface hydrophobicity over PEG chain length with (b): CC dihedral angle, (c): helicity, (d): ratio of CC dihedrals in gauche conformations. (e) solvent accessible volumes of the PEG molecules determined from MD simulations over the corresponding volume of perfectly helical molecules.

5.3 Molecular dynamics simulations on aqueous two-phase systems: Single PEG-molecules in solution

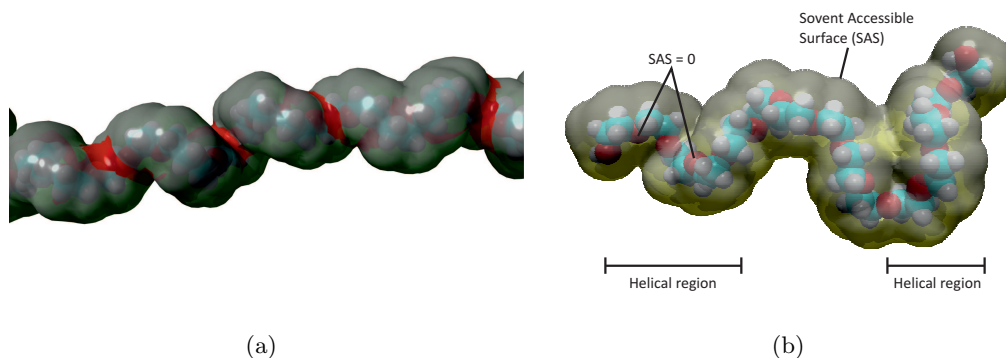


Figure 4: (a): 3D rendering of a helical region formed during the MD simulation. CH-surfaces are marked gray, O-surfaces are marked red. (b): 3D rendering of a simulation snapshot of a PEG722 including the solvent accessible surface (1.4\AA probe radius). Two oxygen atoms having near 0 solvent accessible surface as well as the helical regions are pointed out.

number of H-bond per subunit decreased with increasing PEG chain length in the form of a power function. A linear correlation between average H-bonds per subunit and surface hydrophobic fraction was found with $R^2 = 0.963$.

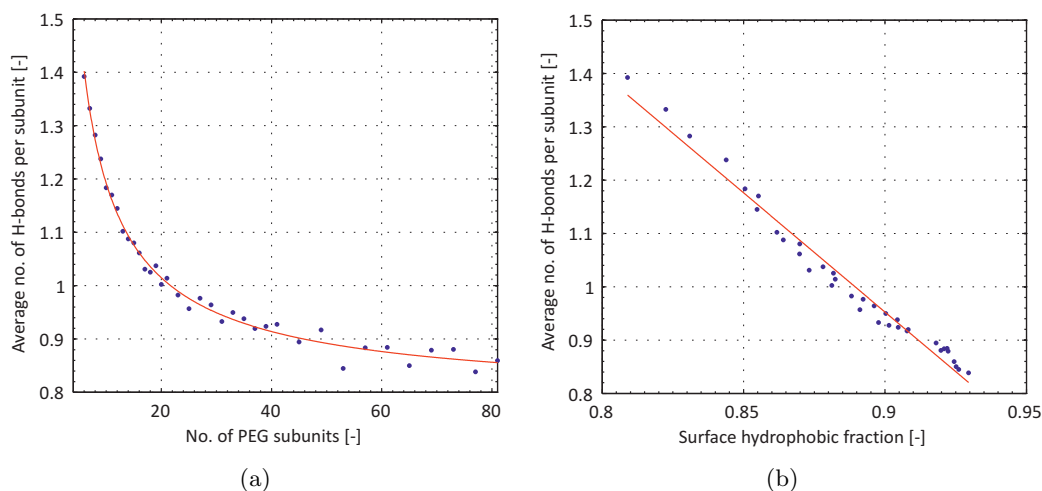


Figure 5: Average number of H-bonds per subunits over number of subunits (top) and surface hydrophobic fraction (bottom)

Binodal curves of PEG300, PEG600, PEG1000 and PEG1500 in combination with PO_4 at pH 7.0 were determined experimentally. The least PEG concentration needed for two-phase formation at four different concentrations of PO_4 were plotted against the surface hydrophobic fraction of the corresponding PEG determined in the MD simulations (figure 6). Linear correlations with an average R^2 of 0.994 were found for all four PO_4 concentrations. The good correlation between experimental and simulation data suggests that the taken simulation approach was successful in generating meaningful polymer surface properties.

5.3 Molecular dynamics simulations on aqueous two-phase systems: Single PEG-molecules in solution

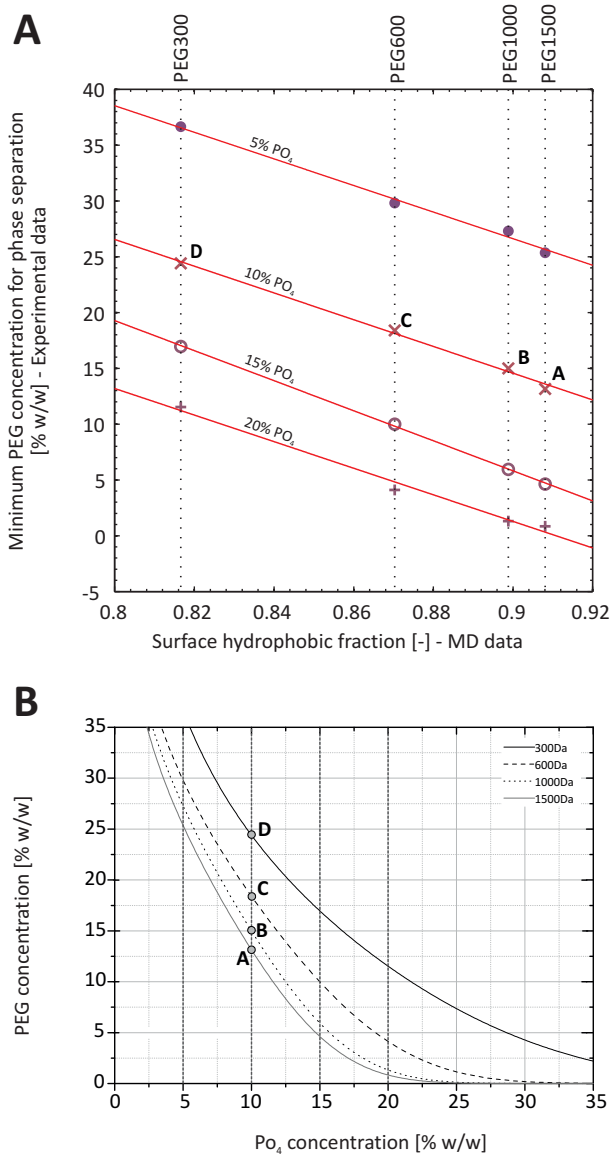


Figure 6: Experimentally determined PEG concentration needed for two-phase formation of four different PEG molecular weights at four concentrations of PO₄ plotted over the surface hydrophobic fraction of the corresponding PEG molecules determined from MD simulations. Phase formation in the presence: ● 5% PO₄, × 10% PO₄, ○ 15% PO₄, + 20% PO₄. **B:** Experimentally determined binodals of PEG-PO₄ ATP systems. Dashed lines represent the levels of PO₄ concentration at which the minimum concentration of PEG needed for phase separation was determined. The data point labelled “A”, “B”, “C”, and “D” are the same points on the binodals in both subfigures.

3.7 Polarity measurements

Figure 7 A shows $E_T(30)$ values as an empirical measure for solvent polarity of PEG in an aqueous solution (PEG 200 to 1450) at different molar fractions. Molar fractions of higher PEG molecular weights were corrected relative to PEG 200 to assure comparability of different PEGs (PEG 400 molar fractions were multiplied by 2, molar fractions of PEG 600 by 3 and so on). Polarity decreased with increasing molar fraction of each PEG. PEG with higher molecular weight was more hydrophobic at the same relative molar fraction (meaning that 2 molecules of PEG 200 were less hydrophobic than 1 molecule of PEG 400). With increasing molar fractions the decrease of $E_T(30)$ clearly deviated from linearity. For that reason the slope of the linear region was used as a basis for comparing polarities of different PEGs rather than using the extrapolated intercept with the x-axis. These values showed a linear correlation with hydrophobicity values generated on the basis of PEG structure information as shown in figure 7 B.

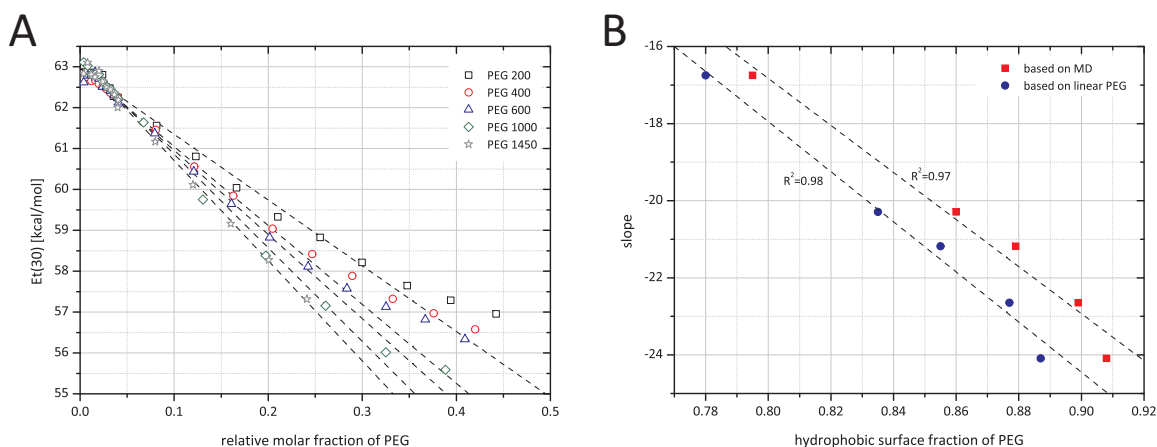


Figure 7: **A:** $E_T(30)$ values of solutions of PEG of varying molecular weight plotted over the relative mole fraction of PEG calculated as described in section “Polarity measurements”. The slopes of the linear parts of the plots were used as measure for the polarity of PEG. **B:** The slopes as determined in subfigure **A** plotted over the surface hydrophobicity calculated from MD simulation results of the corresponding PEG molecules.

4 Discussion

4.1 Primary & secondary structure

Starting from a completely linear structure with all geometric parameters set to standard values by the importing plug-in, our molecular simulations resulted in average PEG structures with geometric parameters in good agreement with both the crystal structure [22] and previous MD simulations [17].

While Tasaki *et al.* [17] found a ratio of gauche(+) to gauche(-) of 0.33 to 0.67 for the CC dihedral in their 0.5-2ns long simulations, we found equal distribution of the two conformations. As there is no molecular constraint that would lead to an uneven distribution of the conformation, we conclude that the chosen simulation time of 5 ns is sufficient to yield a representative, average structure. System total energy reaching equilibrium within 1 ns supports this conclusion.

A clear dependency of CC dihedral and the OC distance on PEG chain length was found. The increase in CC dihedral (1.02%) was significantly larger than the increase in OC bond length (0.12%). While the dependency was obvious, we consider the degree of change of

OC bond length too low to have a significant effect on the resulting surface characteristics. The reason for this dependency remains unclear.

In a recently published paper, Winger *et al.* [18] investigated the influence of force-fields on the resulting structure of a modified PEG molecule. They concluded that the development of a helical structure of the simulated PEG molecule in aqueous solution is dependent on the force-field used. Improper force fields or simulations in vacuo resulted in the PEG molecule collapsing into a random coil. In our study, PEG molecules formed helical structures and stayed elongated. The resulting geometric parameters are in good agreement with the results from previous MD simulations [17] as well as crystallographic data [22] and structural data from other sources (as summarized in [21]). We thus conclude, that the amber03 force-field in concert with the employed parameterization algorithm is well suited to run MD simulations of PEG molecules.

4.2 Structure dynamics

Structure dynamics were looked at in terms of time to reach equilibrium of total system energy, scattering of total system energy and time to reach dihedral energy equilibrium. The general conclusion drawn from these observations was, that longer PEG chains are less flexible. It was further concluded, that the increase in helicality with increasing PEG chain length is a direct consequence of the decreased flexibility. Helical regions are less prone to be broken up if the structure is changing less dynamically. Tasaki *et al.* [17] concluded that their simulation times (0.5-2 ns) were not sufficient to reach an equilibrium. In contrast, system total energy in our simulations reached equilibrium within 1 ns of simulations. We conclude, that simulation times of 1 ns are sufficient when using the simulation protocol employed herein and when the surface of the molecule is the target of the investigation. Reaching an equilibrium tertiary structure, measured as reaching an equilibrium in the radius of gyration, needed significantly longer simulation times as high as 25ns for the largest molecules investigated. Lee *et. al* in 2009 [13] used a different modeling approach to realize simulation times of up to 800 ns in order to investigate PEG tertiary structure and its hydrodynamic properties in water. Such long simulation times are currently outside the reach of all-atom MD simulations. Again, it should be pointed out that the focus of this study was on the secondary structure and PEG solvent interactions.

In 1970 Koenig *et. al* concluded from Raman spectroscopy studies that PEG molecules display a more highly ordered structure in water than in methanol. They hypothesized that the helical nature of the solid state is partly retained in the solubilized state [35]. This was also discussed by Devanand *et. al* [36] and is supported by the results presented herein. While the radius of gyration observed did not support a large deviation from ideal chain behavior, helical regions were found to form within the molecule.

4.3 Surface hydrophobicity

Surface hydrophobicity of the PEG molecules was looked at in terms of surface contribution of CH groups. Two main effects were identified. First, formation of helical regions within the molecule increased the overall hydrophobicity. Second, the effect of the hydrophilic end-group got diluted with increasing PEG chain length. These two effects could explain the overall trend in surface hydrophobicity of the simulated PEG molecules. Additionally, several underlying effects were identified, that influenced surface hydrophobicity and, in part, counteracted one another. For example, both helix angle and overall helicality increased with PEG chain length, with the former decreasing and the latter increasing surface hydrophobicity. While the interplay of the various effects is complex, the overall trend, increasing surface hydrophobicity with increasing PEG chain length showed clearly

5.3 Molecular dynamics simulations on aqueous two-phase systems: Single PEG-molecules in solution

in our simulations. Zaslavsky *et al.* [37] measured the hydrophobicity of tritium labeled PEG molecules (1.5 to 40 kDa) in terms of distribution in Ficoll-400-dextran-70 ATPS. The distribution was found to be equal within experimental certainty. The authors concluded that the investigated PEG molecules had equal hydrophobicity. While results from the MD simulations described herein suggest an increase in surface hydrophobicity, the smallest PEG molecule investigated by Zaslavsky *et al.* had a MW of 1.5 kDa. At this molecular weight (no. of subunits = 34) we found surface hydrophobicity to become less dependent on PEG chain length. Surface hydrophobicity in our simulations changed most in the range 300 Da to 1.1 kDa. Thus, it would be most interesting to obtain distribution data for PEG molecules within this range.

4.4 PEG-solvent interaction

In this study, the direct interactions of PEG molecules and the surrounding water molecules via H-bonds were quantified. It was found that the number of H-bonds per PEG subunit is a function of PEG chain length and decreases with increasing number of PEG subunits. This is in agreement with the increased hydrophobic surface fraction and the helical structure formed by the PEG molecule, in which the oxygen atoms face inwards and are thus excluded from interactions with the solvent. Tasaki *et al.* [17] discussed the average number of water molecules associated with the PEG molecule. The association of water molecules with PEG was determined by a water density function around the PEG molecule and found to be 2.9. There are several studies in which water associated with PEG molecules was quantified, the results ranging from 1 to 5 water molecules per PEG subunit. Our results are at the lower end of this range. However, it should be noted, that the interaction via H-bonds quantified here does not necessarily translate directly into the association measured in the before mentioned studies, as the nature of this association remains unclear and is governed by proximity rather than direct interaction.

In general, H-bond formation is an exothermic process. Less formation of H-bond with increasing PEG chain length might create the need for a higher ratio of order water structure around the PEG molecule, which is an entropically unfavorable process. This might have implications on the effect PEG molecules have on other dissolved entities or its behavior on aqueous two-phase systems. A direct correlation of the surface hydrophobic fraction to the experimentally determined PEG concentration needed for two-phase formation in the presence of fixed concentrations of PO_4 was found. We concluded that the simulation results are a good representation of the actual behavior of PEG molecules in solution and that the hydrophobic character of the molecule might be a driver of phase separation in PEG ATPSs. Further investigations will be based on this approach focusing on phase formation in PEG- PO_4 ATPSs.

While this article focuses on ATPSs, it should be noted that the results might also be applicable to protein precipitation by PEG. Both its structure and thus the space occupied by the PEG molecule in the solvent [1, 38] as well as PEG solvent interactions and thus its influence on solvation energy of the protein might be associated with protein solubility. This will be the scope of following investigations.

4.5 $E_T(30)$ versus relative surface hydrophobicity

To approve the calculation of relative hydrophobicity of PEG based on structural information, we compared $E_T(30)$ as an empirical measure for solvent polarity with calculation results based on PEG structure information. In accordance with theory, PEG solutions became more hydrophobic with increasing molar fraction of PEG, indicating that PEG itself is less polar than water (see figure 7 A). Over a wide range of PEG concentrations

this relationship was found to be linear, at higher PEG concentrations further addition of PEG led to an disproportional decrease of $E_T(30)$. For that reason the slope for the linear part was used as a measure for PEG polarity and correlated with relative PEG hydrophobicity based surface contributions of unpolar $-CH$ and polar $-O-$ and $-OH$ groups of 1) linear PEG molecules and 2) partly helical PEG molecules from MD simulations. In both cases a linear correlation was obtained with similar R^2 values (see figure 7 B), indicating that the structure-based hydrophobicity measure was suitable to describe changes in PEG hydrophobicity. Nevertheless, the experimental data could not be used to differentiate between linear and partly helical PEG structures.

5 Conclusion and outlook

In this study, the validity of using the amber03 force field in combination with the AutoSMILES self-parameterizing algorithm for MD simulations of PEG was confirmed. MD simulations were run on a series of PEG molecules ranging in molecular weight from 300 Da to 3500 Da. 3D data from these simulations were found in good agreement with recently published crystallographic data and published MD simulation results. It was found that PEG chain length has a major influence on the surface characteristics and solvent interaction of PEG. Surface hydrophobicity values derived from these simulations could be correlated to experimentally determined minimum PEG concentrations needed to establish two-phase systems in the presence of PO_4 . Surface hydrophobicity values from MD simulations correlated linearly with solution polarity experimentally determined via a solvatochromic dye. This work forms the basis for conducting further MD studies on phase behavior of PEG as well as simulations including both PEG and proteins such as proteins in ATPSs or simulations of PEGylated proteins.

References

- [1] D. H. Atha, K. C. Ingham, Mechanism of precipitation of proteins by polyethylene glycols. analysis in terms of excluded volume., *The Journal of biological chemistry* 256 (2323) (1981) 12108–17.
- [2] I. L. Shulgin, E. Ruckenstein, Preferential hydration and solubility of proteins in aqueous solutions of polyethylene glycol., *Biophysical chemistry* 120 (3) (2006) 188–98.
- [3] P.-Å. Albertsson, Particle fractionation in liquid two-phase systems the composition of some phase systems and the behavior of some model particles in them application to the isolation of cell walls from microorganisms, *Biochim Biophys Acta* 27 (1958) 378–395.
- [4] R. Hatti-Kaul, Aqueous two-phase systems - a general overview, *Mol Biotechnol* 19 (3) (2001) 269–277.
- [5] U. B. Hansson, C. Wingren, Separation of antibodies by liquid-liquid aqueous partition and by liquid-liquid partition chromatography, *Sep Purif Methods* 27 (2) (1998) 169–211.
- [6] J. G. Huddleston, A. Lyddiatt, Aqueous 2-phase systems in biochemical recovery - systematic analysis, design, and implementation of practical processes for the recovery of proteins, *Appl Biochem Biotechnol* 26 (3) (1990) 249–279.

5.3 Molecular dynamics simulations on aqueous two-phase systems: Single PEG-molecules in solution

- [7] K. M. Sunasara, F. Xia, R. S. Gronke, S. M. Cramer, Application of hydrophobic interaction displacement chromatography for an industrial protein purification., *Biotechnology and bioengineering* 82 (3) (2003) 330–9.
- [8] A. Gabizon, R. Catane, B. Uziely, B. Kaufman, T. Safra, R. Cohen, F. Martin, A. Huang, Y. Barenholz, Prolonged circulation time and enhanced accumulation in malignant exudates of doxorubicin encapsulated in polyethylene-glycol coated liposomes., *Cancer research* 54 (4) (1994) 987–92.
- [9] G. Gregoriadis, Engineering liposomes for drug delivery: progress and problems, *Trends in Biotechnology* 13 (1212) (1995) 527–537.
- [10] R. B. Greenwald, Y. H. Choe, J. McGuire, C. D. Conover, Effective drug delivery by pegylated drug conjugates., *Advanced drug delivery reviews* 55 (2) (2003) 217–50.
- [11] J. M. Harris, R. B. Chess, Effect of pegylation on pharmaceuticals., *Nature reviews. Drug discovery* 2 (3) (2003) 214–21.
- [12] F. Dismer, J. Hubbuch, 3D structure-based protein retention prediction for ion-exchange chromatography., *Journal of chromatography. A* 1217 (8) (2010) 1343–53.
- [13] H. Lee, R. M. Venable, A. D. Mackerell, R. W. Pastor, Molecular dynamics studies of polyethylene oxide and polyethylene glycol: hydrodynamic radius and shape anisotropy., *Biophysical journal* 95 (4) (2008) 1590–9.
- [14] I. Vorobyov, V. M. Anisimov, S. Greene, R. M. Venable, A. Moser, R. W. Pastor, A. D. MacKerell, Additive and classical drude polarizable force fields for linear and cyclic ethers, *Journal of Chemical Theory and Computation* 3 (3) (2007) 1120–1133.
- [15] H. Lee, A. H. d. Vries, S.-J. Marrink, R. W. Pastor, A coarse-grained model for polyethylene oxide and polyethylene glycol: conformation and hydrodynamics., *The journal of physical chemistry. B* 113 (40) (2009) 13186–94.
- [16] O. Borodin, D. Bedrov, G. D. Smith, A molecular dynamics simulation study of polymer dynamics in aqueous poly(ethylene oxide) solutions, *Macromolecules* 34 (16) (2001) 5687–5693.
- [17] K. Tasaki, Poly(oxyethylene)-water interactions : A molecular dynamics study, *Journal of the American Chemical Society* 118 (1996) 8459–8469.
- [18] M. Winger, A. H. de Vries, W. F. van Gunsteren, Force-field dependence of the conformational properties of α,ω -dimethoxypolyethylene glycol, *Molecular Physics* 107 (13) (2009) 1313–1321.
- [19] J. Wang, R. M. Wolf, J. W. Caldwell, P. a. Kollman, D. a. Case, Development and testing of a general amber force field., *Journal of computational chemistry* 25 (9) (2004) 1157–74.
- [20] A. Jakalian, D. B. Jack, C. I. Bayly, Fast, efficient generation of high-quality atomic charges. am1-bcc model: Ii. parameterization and validation., *Journal of computational chemistry* (2002) 1623–41.
- [21] D. Seebach, E. Zass, W. B. Schweizer, A. J. Thompson, A. French, B. G. Davis, G. Kyd, I. J. Bruno, Polymer Backbone Conformation-A Challenging Task for Database Information Retrieval, *Angewandte Chemie-International Edition* 48 (51) (2009) 9596–9598.

5.3 Molecular dynamics simulations on aqueous two-phase systems: Single PEG-molecules in solution

- [22] A. C. French, A. L. Thompson, B. G. Davis, High-Purity Discrete PEG-Oligomer Crystals Allow Structural Insight, *Angewandte Chemie-International edition* 48 (7) (2009) 1248–1252.
- [23] E. Krieger, G. Koraimann, G. Vriend, Increasing the precision of comparative models with yasara nova—a self-parameterizing force field., *Proteins* 47 (3) (2002) 393–402.
- [24] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, A smooth particle mesh ewald method, *The Journal of Chemical Physics* 103 (19) (1995) 8577.
- [25] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, et al., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *Journal of the American Chemical Society* 117 (1919) (1995) 5179–5197.
- [26] E. J. Sorin, V. S. Pande, Exploring the helix-coil transition via all-atom equilibrium ensemble simulations, *Biophysical Journal* 88 (4) (2005) 2472 – 2493.
- [27] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, et al., A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations., *Journal of computational chemistry* 24 (1616) (2003) 1999–2012.
- [28] E. Krieger, T. Darden, S. B. Nabuurs, A. Finkelstein, G. Vriend, Making optimal use of empirical energy functions: force-field parameterization in crystal space., *Proteins* 57 (4) (2004) 678–83.
- [29] H. J. C. Berendsen, J. P. M. Postma, W. F. v. Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath, *The Journal of Chemical Physics* 81 (8) (1984) 3684.
- [30] S. A. Oelmeier, F. Dimer, J. Hubbuch, Application of an aqueous two-phase systems high-throughput screening method to evaluate mab hcp separation., *Biotechnology and bioengineering* 108 (1) (2010) 69–81.
- [31] D. L. Sackett, J. Wolff, Nile red as a polarity-sensitive fluorescent-probe of hydrophobic protein surfaces, *Analytical Biochemistry* 167 (2) (1987) 228–234.
- [32] K. Dimroth, F. Bohlmann, C. Reichard, T. Siepmann, Über Pyridinium-n-phenolbetaine und ihre Verwendung zur Charakterisierung der Polarität von Lösungsmitteln, *Annalen Der Chemie-Justus Liebig* 661 (1963) 1–37.
- [33] C. Tanford, *Physical Chemistry of Macromolecules*, John Wiley & Sons, 1961.
- [34] M. Doi, S. F. Edwards, *The Theory of Polymer Dynamics*, Claredon Press, 1986.
- [35] J. L. Koenig, A. C. Angood, Raman spectra of poly(ethylene glycols) in solution, *Journal of Polymer Science Part A-2: Polymer Physics* 8 (10) (1970) 1787–1796.
- [36] K. Devanand, J. Selser, Asymptotic behavior and long-range interactions in aqueous solutions of poly (ethylene oxide), *Macromolecules* 24 (22) (1991) 59435947.
- [37] B. Zaslavsky, A. Baevskii, S. Rogozhin, A. Gedrovich, A. Shishkov, A. Gasanov, A. Masimov, Relative hydrophobicity of synthetic macromolecules i. polyethylene glycol, polyacrylamide and polyvinylpyrrolidone, *Journal of Chromatography A* 285 (1984) 63–68.

5.3 Molecular dynamics simulations on aqueous two-phase systems: Single PEG-molecules in solution

- [38] T. Arakawa, S. N. Timasheff, Mechanism of poly(ethylene glycol) interaction with proteins., *Biochemistry* 24 (2424) (1985) 6756–62.

MANUSCRIPT IV

MOLECULAR DYNAMICS SIMULATIONS ON AQUEOUS TWO-PHASE SYSTEMS: PHASE FORMATION AND PROTEIN PARTITIONING

Florian Dismer⁺, Stefan A. Oelmeier⁺, Jürgen Hubbuch*

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

* Corresponding author. Tel.: +49 721 608-2557; fax: +49 721 608-6240. E-mail address: juergen.hubbuch@kit.edu

⁺These authors contributed equally to this publication

in preparation

Abstract

Product titers have steadily increased in the biopharmaceutical industry over the past decades adding new challenges to downstream purification processes. For that reason unit operations are needed that can be operated in a continuous fashion offering high productivity, one such alternative being aqueous two phase systems. These systems were shown to have high selectivity while offering mild working conditions ensuring the integrity of biomolecules. Nevertheless, driving forces for aqueous-aqueous partitioning of proteins are not well understood and process development remains based on empirical screenings. Here we investigated the use of Molecular Dynamics (MD) simulations to determine phase system properties of polyethylen glycol (PEG) - phosphate systems that guide protein distribution. MD simulations were accurate enough to resolve differences in certain molecular properties of the ATPS components (such as the relative hydrophobicity and H-bonding properties of PEG) over a wide range of PEG and salt concentrations (0-35% w/w) as well as PEG molecular weights (282 - 1471 kDa). We successfully correlated these properties with phase formation resulting in a model that was capable of predicting binodals and tie lines for the investigated systems. We also extracted upper- and bottom phase properties (polarity, H-bonding donor and acceptor functionalities) necessary for the use of the Abraham equation which was shown to be useful in predicting partitioning coefficients of proteins. These properties were compared to experimentally determined values measured by using solvatochromic dyes. Both datasets, experimentally and *in silico* determined phase system properties showed good correlation with partitioning coefficients of lysozyme in these systems.

Keywords: *Molecular dynamics simulation, polyethylene glycol, binodal prediction, aqueous two-phase systems, amber03, yasara*

1 Introduction

The use of aqueous two-phase systems for the purification of biotechnological products has been extensively studied in the past [2, 8, 14, 22] as they offer some advantages over traditional chromatography: easy scale-up, mild working conditions, continuous operation modes, high mass-transfer rates and integrated cell debris removal. One of the ATPS components is always a water-miscible polymer (e.g. polyethylene glycol, dextrane, etc.), the second component can be again a polymer, a salt (e.g. sodium or potassium phosphate, ammonium sulfate, sodium chloride, etc.) or a detergent (e.g. polyoxyethylene detergents, Triton X100).

For biotechnological applications conditions need to be found where the differences in distribution coefficients of the target molecule and the main contaminants are big enough to yield an acceptable purity and yield. Since the distribution event itself is of a multi-modal nature and numerous system parameters are known to play a role during the partitioning event, rationally choosing the right system and the right system composition can be difficult [8]. Additionally, many fundamental mechanisms behind the formation of phases and the distribution of proteins are not well understood making it even harder to rationally adjust system parameters.

Due to these limitations, high-throughput techniques are still state-of-the-art when it comes to process development [4, 18]. To reduce experimental efforts, these screenings can be combined with Design of Experiment approaches as shown by [7, 19] and others. Parameters such as tie line length, temperature, sodium chloride concentration and polymer molecular weight are usually being looked at.

The use of mathematical models can further decrease the screening work by *in silico* identifying promising systems. For that a wide range of different approaches can be found in literature. Silverio *et al.* [23] showed that the Collander equation can be applied that allows estimating partitioning coefficients of proteins in ATP systems based on available

5.4 Molecular dynamics simulations on aqueous two-phase systems: Binodal curve and tie line prediction

data for chemically similar systems. Baughman *et al.* [3] demonstrated already in 1994 that neuronal networks combined with expert systems can be used to predict phase diagrams and protein distribution coefficients based on ATP system properties, protein properties based on secondary structure information and salt properties. The major drawback of that method, as stated by the authors, was that their network did not provide the user with information on the rationale behind the network answer, making it difficult to extend the model to problems beyond the training dataset. Salgado *et al.* [20] correlated protein hydrophobicity (based on primary and tertiary protein structures using different hydrophobicity scales) with partitioning behavior in PEG-salt systems. The model based on the hydrophobicity calculated based on the primary structure of the proteins (assuming an equal accessibility of all amino acids) outperformed the 3D structure-based approach. The same can be found for protein aggregation predictions that correlated well with primary sequence based approaches. One reason for that are structural changes by partial unfolding of the protein molecules when undergoing non-native aggregation. So our hypothesis is that the protein might also undergo some reversible structural changes when present in an ATPS so that the crystal structures used correlations with distribution coefficients are not representable for proteins under high salt and PEG concentrations. Andrews *et al.* [1] used ammonium sulfate precipitation to determine protein hydrophobicity and showed that hydrophobicity alone can be used only in 5 out of 12 systems investigated (PEG-sulfate, -phosphate, -citrate and dextrane systems with different levels of sodium chloride concentrations) for reasonably good predictions (with correlation coefficients above 0.80). Models only based on net charge or charge densities (both per area and per volume) performed even worse. This underlines the multi-modal nature of protein distribution and thus the complexity of aqueous two-phase systems when it comes to predictive modeling. The major drawback of modeling protein partitioning in such a way is that the model itself does usually not include parameters describing the properties of the phase systems. As a result the models are predictive for the calibrated phase systems only, they are e.g. not capable of describing changes in the partitioning of proteins with increasing tie line length.

In a very recent study Madeira *et al.* [12] showed that a modified version of the Abraham equation, which is commonly applied for small organic compounds, is suitable to predict protein partitioning in phase systems that were characterized using solvatochromic dyes to measure the solvent polarizability π^* , the hydrogen-bond donor acidity α and the hydrogen-bond acceptor basicity β of both the upper and lower phase of polymer-polymer systems. In a series of distribution experiments, the sensitivity of a protein to these three parameters can be determined on an empirical basis in the form of three coefficients that can be used to determine distribution coefficients in ATPS with known $\Delta\pi^*$, $\Delta\alpha$ and $\Delta\beta$ values characterizing the differences between upper and lower phase. This again emphasizes that a) protein partitioning is of multi-modal nature and that it is b) in parts based on the interaction of phase system components (water, buffer molecules and/or PEG) with the protein.

All these findings indicate that the properties of both, the phase-forming components and the protein need to be looked at, one alone does possibly not lead to good, universally applicable models. Especially the findings by Andrews *et al.* show the need for new approaches that could account for possible structural changes. One such approach is the use of force-field based molecular dynamics simulation including all components present in such a system. In the work presented here, we studied the applicability of this tool, in a first step by simply looking at the formation of phases in PEG-phosphate systems and the effects of different PEG molecular weights on both, the position of the binodal curve and the slope of the tie lines within the phase diagram. In the second part of the presented work we extracted $\Delta\pi^*$, $\Delta\alpha$ and $\Delta\beta$ values from MD simulations as well as from experimental

5.4 Molecular dynamics simulations on aqueous two-phase systems: Binodal curve and tie line prediction

data based on solvatochromic dyes, compared the trends observed in the data and used both datasets to build a predictive model based on the publication by Madeira *et al.* [12].

2 Materials & methods

2.1 Molecular Dynamics simulation software

Molecular dynamics simulations were performed using Yasara Structure [11], version 10.2.1. The software was installed in a cluster environment running Suse Linux Enterprise 10 as operating system. Each simulation was run using 16 nodes of the cluster computer. Each node was equipped with 2 Intel Xeon Quad Core (X5355) processors at 2.66 GHz and 16 GB local memory (2 GB per processor core).

2.2 Force field and PEG parametrization

The amber03 force field [6] was used for all molecular dynamics simulations. PEG molecules were parameterized using AutoSMILES to generate force field parameters and validated as described by Oelmeier *et al.* [17]. Phosphate ions were built with Yasara and also parametrized using AutoSMILES [9, 24]. Long range cutoff was set to 10 Å.

2.3 MD simulation design

Simulations were performed for four different PEGs with increasing molecular weight: 282 Da, 590 Da, 1031 Da and 1471 Da. The size of the simulation cell was 45 Å in both x and y direction, the z dimension was adjusted according to the length of a single PEG molecule, ranging from 22.5 Å for the 282 Da PEG to 112 Å for the 1471 Da PEG. For each PEG, 64 MD simulations were run with PEG and phosphate concentrations equally spaced between 0-35% [w/w] (eight different PEG concentrations at eight different phosphate concentrations). The following molecules were added to the simulation cell in the listed order: PEG, acidic phosphate buffer component, basic phosphate buffer component, potassium ions and water molecules. The number of molecules per volume was determined by the final density of the system. System density was considered to be a function of PEG and phosphate concentrations. The PEG stock solution density was measured by a pycnometer as 1.2 g/mL, independent of molecular weight and phosphate concentration (calculated for different concentrations according to Schiel *et al.* [21] at pH 7). Final densities of the systems were in the range of 1.00 - 1.15 g/mL. The ratio of acidic to basic buffer components was adjusted to pH 7 using the Henderson-Hasselbalch equation 1.

$$pH = pK_a + \log \frac{c(A^-)}{c(HA)} \quad (1)$$

Potassium ions were added to keep electroneutrality. The total number of atoms per simulation cell was in the range of 3916 to 22.764 atoms. At the lowest PEG concentration, a minimum of 5 molecules were added to the simulation to account for intramolecular effects. PEG molecules were energy minimized before adding them to the simulation cell, and were evenly distributed within the cell.

2.4 MD simulation parameters

Each of the 64 all-atom simulations per PEG were run with a total simulation time of 1 ns, with the 'rescale velocity' setting of Yasara. Snapshots were taken and analyzed with 2.5 ps timesteps.

2.5 Data handling

Raw data was preprocessed by averaging to reduce the amount of data. All calculated energies were averaged over the last 500 ps of each simulation, surface properties of PEG (hydrophobic and hydrophilic surface) and H-bonds data were averaged over the last 250 ps. The averaged dataset was then normalized in two ways to generate two datasets for modeling:

- relative to a volume of 1 nm³ and
- relative to the number of a) molecules for water and buffer components, b) PEG units for all PEG related descriptors and c) atoms for all system energies

Both approaches are being discussed later.

To increase data density for modeling and tie line prediction, individual energies was fitted over the PEG and phosphate concentration by a 3rd degree polynomial surface function and data points were interpolated to a grid of 35 PEG and 35 phosphate concentrations over the same range used for the MD simulations (0-35 % [w/w]).

2.6 Modeling of binodals

Models were build using a multivariate clustering approach that is implemented in Matlab as part of the 'classify' command, using the mahalanobis distance for cluster analysis. For modeling, data was split into a teaching set consisting of three different PEG molecular weights and a test-set including the fourth PEG. Each model was generated 4 times, with changing teaching- and test-sets, so that each PEG was in the test set once. This leave-one-out approach was shown to be a useful in terms of validating predictive models [5, 10]. Altogether 45 input parameters were used: 38 different energy contributions (bond, angle, dihedral, coulomb, Van-der-Waals and total energy) for all components in the system (PEG, water, acidic and basic buffer components, ions) as well as 3 parameters describing H-bond formation of PEG and 4 parameters characterizing PEG hydrophobicity.

The experimentally determined binodals for each PEG in the teaching-set were used to assign each simulated system to either the one-phase or the two-phase region of the phase diagram. This information was then used to build a model based on clustering of certain descriptor combinations. The most complex models were build by using four different descriptors, model screening was done in a full-factorial approach.

2.7 Model error and prediction quality

Models were ranked according to their error and their predictive quality. Model error was calculated as the ratio between misclassified data points in the teaching set to the total number of data points used to build the model, e.g. if 2 data points were assigned to the 1-phase region of a phase diagram but were experimentally determined to be in the 2-phase region, the model error was calculated as 2/64 = 0.031 (or 3.1 %). Since each model was built four times with a different PEG molecular weight being the test set, the model error was averaged over these four models.

To determine the model quality, the number of phases for each data point in the test-set was predicted, and points of phase transition were determined (the last 1-phase and the first 2-phase data point along a line of constant PEG concentration and increasing PO₄ concentration). These data points were fitted with a function representing a binodal curve proposed by Merchuk *et al.* [16]:

$$f_{c_{PEG}} = a * exp^{(b*c_{PO_4}^{0.5} + c*c_{PO_4}^3)} \quad (2)$$

5.4 Molecular dynamics simulations on aqueous two-phase systems: Binodal curve and tie line prediction

The experimental and predicted binodal were compared by calculating the relative difference between both within the experimental range (0-35 % [w/w] PEG and phosphate). As with the model error, quality values were averaged over all four models. Both, model error and prediction quality averages were then combined to a score value by simply multiplying both values, so that a model with an average error of 20 % and an average deviation of 10 % between predicted and experimental binodals of all four PEGs would have a score of:

$$\text{Score} = \text{Error} * \text{Quality} = 0.20 * 0.10 = 0.02 \quad (3)$$

2.8 Tie line prediction

Given a point P (that is defined by a phosphate concentration $c_{P,phosphate}$ and a PEG concentration $c_{P,PEG}$) in the 2-phase region of a phase diagram, the corresponding tie line is defined by this point and a negative slope STL in the form of:

$$c_{P,PEG} = STL * c_{P,phosphate} + a \quad (4)$$

with a being the intercept with the y-axis. By knowing the slope at a given point P , a can easily be determined. The following steps were performed using simulation data to determine the slope:

1. Definition of a point P .
2. Definition of a series of slopes to investigate (here: slopes between -0.5 and -2 in steps of 0.025).
3. Determination of the left and right intercept with the binodal for each slope (as they define the upper and lower phase composition the system would split into).
4. Calculation of descriptor values for these systems (upper and lower phase) for each slope using the 3rd-degree polynomial surface used for interpolation.
5. Finding the slope at which both descriptor value reach a maximum or minimum (depending on type the descriptor).
6. This slope was then compared with experimental data published by Zaslavsky [25] for PEG 600, PEG 1000 and PEG 1500.

2.9 Solvatochromic dye measurements

In order to determine α , β and π^* values for different phase systems (PEG 300, 600, 1000 and 1500), three different dyes were used: 4-nitrophenol for the determination of β , 4-nitroanisole for the determination of π^* and Nile red for the determination of α . All three dyes were calibrated by measuring their absorption maximum in different organic solvents with known α , β and π^* values as published by Marcus [13]. The solvents used for calibration were: water, ethanol, different mixtures of water and ethanol, acetonitrile, acetic acid, formic acid, octanol, isopropanol and acetone. All dyes and solvents were purchased from Sigma Aldrich (Taufkirchen, Germany). The following equations 5 to 7 were obtained after calibration:

$$\text{4-Nitroanisole: } \pi^* = -11.91 + 0.041 * \lambda \quad (5)$$

$$\text{4-Nitrophenol: } \beta = 0.32 * \left(35.6 - \frac{10000}{\lambda} \right) - 0.84 * \pi^* \quad (6)$$

5.4 Molecular dynamics simulations on aqueous two-phase systems: Binodal curve and tie line prediction

$$\text{Nile Red: } \alpha = -4.08e^{-4} * \lambda^2 + 0.48 * \lambda - 0.76\pi^* - 140.34 \quad (7)$$

For each PEG molecular weight, 48 systems equally distributed in the one-phase region of the phase diagram were prepared on a TECAN station (Crailsheim, Germany) in 96-well UV plates (Greiner bio-one, Frickenhausen, Germany) in duplicates with a volume of 300 μl each. 100 μl were transferred into a 384-well plate (Greiner bio-one, Frickenhausen, Germany). 2 μl of dye were added to the systems and absorption spectra were measured with an Infinite 200 spectrophotometer (Tecan, Crailsheim, Germany). To speed up measurement two dyes were combined in one sample (nile red and 4-nitroanisole, nile red and 4-nitrophenol). All α , β and π^* values were fitted with polynomial surfaces with the lowest degree possible still giving a reasonably good R^2 in order to extrapolate the parameters to the experimentally determined binodals which is necessary for the calculation of the Δ values.

2.10 Protein partitioning coefficients

Lysozyme partitioning coefficients were determined for all four PEG molecular weights on the TECAN station as published by Oelmeier *et al.* [18] for total lysozyme concentrations of 0.1 mg/ml. At this concentration no influence of the protein concentration on the partitioning coefficient was observed. Lysozyme concentrations in both phases were measured using an HPLC method with a gradient elution from 40 % to 100 % acetonitrile (+ 0.1 % TFA) over 2 min on a reversed-phase column (Chromolith RP-18, Merck, Darmstadt) with a flow rate of 4 ml/min. The analytical method was calibrated with an injection volume of 4 μl and lysozyme concentrations of 0.01 - 1 mg/ml. All samples with higher concentration were diluted.

3 Results & discussion

The general idea behind clustering of data is to find a link between multiple observations for a number of individuals and their respective properties. Here, the individuals were different ATPSs, each defined by a PEG and a phosphate concentration as well as a PEG molecular weight and the property of interest was the number of phases formed, which is of course connected to the position of each individual system within the phase diagram. The observations were 45 descriptors extracted from MD simulations:

- 31 Energies of single components in this particular system (water, PEG, buffer components)
- 4 PEG surface descriptors
- 3 PEG H-bonding properties
- 7 System energies

All individuals cluster into two sub-populations: those in the 1-phase region (below the binodal curve) and those in the 2-phase region (above the binodal curve). The goal was to find descriptors that could, alone or in combination with others, be linked to phase formation and thus be used for binodal predictions based on MD data.

In the simplest case (a 2-descriptor model) the descriptors could be e.g. the PEG and the phosphate concentration. One could then determine the center points of both clusters calculate by calculating the arithmetical means of both descriptors for all 1-phase individuals and all 2-phase individuals. By defining certain boundaries, e.g. a maximal euclidean

5.4 Molecular dynamics simulations on aqueous two-phase systems: Binodal curve and tie line prediction

distance of a system to one of the center points, each new individual that is characterized by a PEG and a phosphate concentration can then be assigned to one of the two clusters. If the training set is big enough, and if only one PEG molecular weight is used, this simple model could probably give a reasonably good prediction quality. Nevertheless, this model would fail to explain effects of increasing PEG molecular weight and it would not be able to generate mechanistic understanding. As a result, better or at least more descriptors are needed. With increasing number of descriptors, clustering becomes a multidimensional problem. In that case the mahalanobis instead of the euclidean distance can be calculated and used to specify the boundaries. For more details, please refer to McLachlan *et al.* [15].

At this point we want to emphasize that the scope of this paper is not the characterization of the process of phase formation itself, it is rather the end-point of the process that we were interested in and the question whether MD simulations are accurate enough to resolve the differences between stable systems in the 1-phase region and instable systems in the 2-phase region.

3.1 Data normalization, interpolation and random models

In the process of data pre-processing, two different approaches for data normalization were chosen: a) per volume and b) per molecules. Normalization was necessary for predictive modeling since neither the number of molecules nor the volumes of the simulation boxes were constant over all systems. Both procedures generally yield different information: data normalized per volume gives information on the total contribution of one type of molecules to the system while data normalized per molecule gives information about changes of the energetical state of a certain type of molecules. For example the energy x of a component y normalized per molecule can remain constant at different concentrations while the same energy normalized per volume changes due to an increasing or decreasing number of molecules possibly leading to a changed contribution of that component to the total energy of the system.

To judge model validity data was generated from random 3rd degree polynomial surfaces for each PEG similar to the ones used for data interpolation. Modeling was performed in the exact same way as used for MD based data and was repeated for 100,000 random datasets to give a representative probability distribution. All MD based models had significantly lower scores than the random models, actually no 4-descriptor model based on random data had a score < 0.03 (the lower the score, the better the model).

Interpolation of the MD data to increase data density (from 5 % (w/w) to 1 % (w/w) steps for PEG and phosphate) lead to an overall increase in the probability for a good score, although it did not positively affect the best 20 % of the random models (for 4 descriptors, the probability decreased, which was different for 1-3 descriptors where no effect of high data density was found). Only 20 % of the random models were capable of differentiating between 1-phase and 2-phase systems for each PEG, for MD based data 75 % of all models met these criteria.

3.2 Model complexity

Figure 1A shows the best model score for each level of model complexity and the influence of 1) data interpolation and 2) the effect of two different normalization procedures (as described in the 'Data handling' section). As expected, models became better with increasing number of descriptors used. All models shown here (except the 1 descriptor model for the 'per volume' normalized and interpolated data) had a probability of < 0.001 % to occur from random data. Adding one descriptor to the model increased model quality by approximately 45 % which can be translated into a decrease in model error and an increase in

5.4 Molecular dynamics simulations on aqueous two-phase systems: Binodal curve and tie line prediction

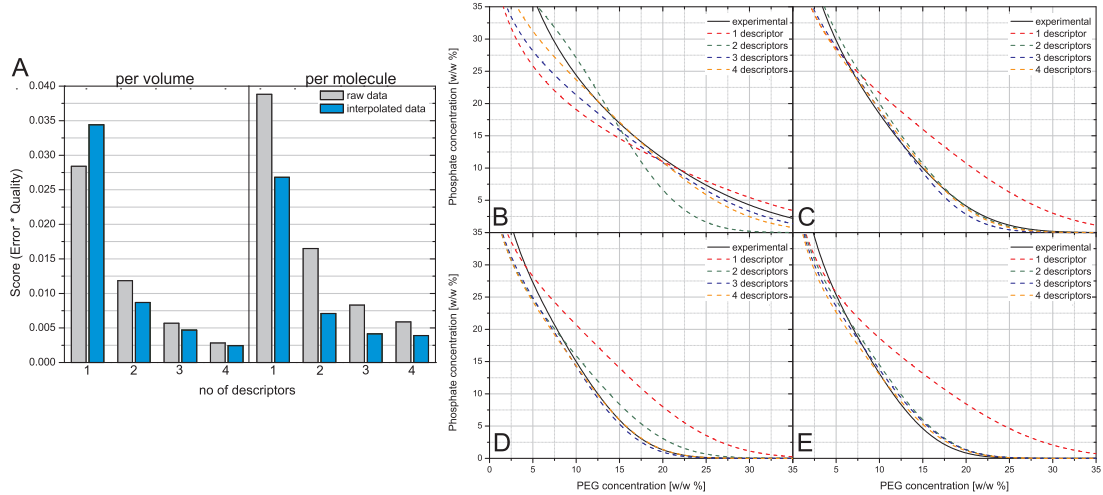


Figure 1: Model scores and results for the prediction of binodals based on MD data. **A.** Best model scores of models with increasing complexity and the influence of data normalization and interpolation. **B-E.** Prediction results (dashed lines) using the best models in comparison with experimental data (solid lines) for all PEGs between 300 Da (**B**), 600 Da (**C**), 100 Da (**D**) and 1500 Da (**E**).

prediction quality by about 25 % each. Extrapolating to 5 descriptors showed only a minor score improvement and was thus not considered, especially when taking into consideration, that for 5 descriptor models, about 1.2 billion descriptor combinations would have to be evaluated.

Data interpolation improved the overall model quality for the following reasons:

1. Interpolation averaged out noise from the MD simulation data. This also explains why interpolation had a stronger effect on the per molecule normalized data, since the R^2 values for data fitting were generally lower.
2. An increase in data density led to higher resolution of data points close to the binodal, thus training of the models became precise.
3. Higher resolution data also improved binodal curve fitting to the predicted data.

The overall best model was found for data normalized by volume with a score of 0.0024 (with a model error of about 0.08 and a prediction quality of 0.03). The predicted binodals

Complexity	Descriptors	Score
1 descriptor	PEG: accepted H-bonds	0.027
2 descriptors	PEG: hydrophobic surf. PEG: accepted H-bonds	0.007
3 descriptors	H ₂ O: bond energy H ₂ O: angle energy PEG: rel. hydrophobicity	0.004
4 descriptors	PEG: vdW energy H ₂ O: total energy PEG: rel. hydrophobicity System: angle energy	0.002

Table I: Summary of the best models and the descriptors used for these models

5.4 Molecular dynamics simulations on aqueous two-phase systems: Binodal curve and tie line prediction

are shown in figure 1 B. Due to the leave-one-out approach for model building, the predicted PEG was not part of the teaching set. 1-descriptor models could not clearly distinguish between different PEG molecular weights whereas 2-descriptor models performed reasonably good for PEG 600 and PEG 1500 with some deviations for PEG 1000 and strong deviations for PEG 300. Adding more descriptors mostly affected prediction quality for PEG 300 and 4-descriptor models showed good agreement between experimental and predicted binodals with the highest deviation at low phosphate concentrations.

The descriptor combinations that gave the best models are summarized in table I. PEG and H₂O descriptors seemed to dominate over others. All PEG descriptors were linked to either the hydrophobicity or the formation of H-bonds with surrounding water molecules, both previously found to be good discriminators between PEGs with different molecular weights [17]. Trying to identify crucial descriptors based on the best models alone is not a valid approach to generate mechanistic understanding since other models with almost the same score used different descriptors. Thus looking at the interplay of different descriptors would be the key to understanding phase formation.

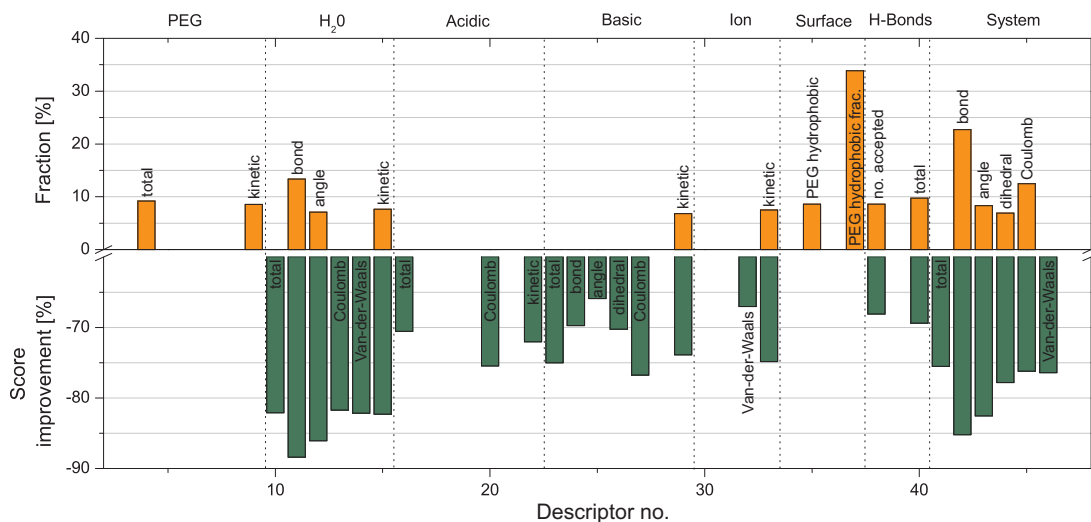


Figure 2: Upper half: Frequency of occurrence of each descriptor within the 5 % best models. Only those descriptors are shown that occurred more frequently than the average over all descriptors (6.6 %). **Lower half:** Average relative score improvement for each descriptors. The average relative decrease in score when adding an additional descriptor was by 65 %. The numbers close to the bars indicate the cluster this descriptor belonged to. Descriptor clustering is discussed at a later point.

3.3 Global descriptor assessment

In general only about 21 % of all models that were built with descriptors of only one component (e.g. only PEG descriptors) were better than average indicating that looking at PEG properties and energy contributions only is not sufficient for good predictions endorsing the use of MD with explicit solvent and buffer components. To generate mechanistic understanding, we looked at two different measures for the importance of a certain descriptor: 1) frequency of occurrence of each descriptor within the best 5 % of all models and 2) average score improvement when a certain descriptor is added to a model.

The results are summarized in figure 2. The upper half clearly shows that the relative hydrophobicity of PEG (descriptor 37) is one of the most frequently occurring descriptors (in 34 % of all models). Interestingly this frequency went up with increasing number of descriptors from 3 % for 2-descriptor models up to 53 % for 4-descriptor models. Looking at the predicted binodals in figure 1 B, the models with 1 and 2 descriptors could hardly

5.4 Molecular dynamics simulations on aqueous two-phase systems: Binodal curve and tie line prediction

capture the effects of increasing PEG molecular weight, especially for the 300 kDa PEG, thus the role of descriptor 37 is mainly to discriminate between different PEGs rather than being responsible for predicting the shape of the binodal. The reason, why this descriptor did not appear in the lower half of the plot showing the average model improvement is the specific interplay of this descriptor, that improved models by 90 % when combined with water energies 11, 12 and 15 (bond, angle and kinetic energy), but when combined with other descriptor model improvement was rather poor.

The second most frequent descriptor was the system bond energy (42) that occurred in 23 % of all models with a decreasing tendency (43 % down to 12 %). The same decreasing trend could be observed for the coulomb energy of the system (descriptor 45, from 19 % down to 9 %). All these system energies showed also a good average model improvement. System energies capture the trend in a certain energy over all components. With increasing number of descriptors system energies were being replaced by single component energies giving more detailed information (bond energy of water (11) and PEG (4) both went up by about 8 % each, coulomb energy for acidic and basic buffer components both went up by 3 % each).

Other descriptors showing a rather high frequency were the kinetic energies of all individual components (15,29 and 33) and descriptors characterizing the H-bond formation of PEG as an H-bond acceptor (38,40).

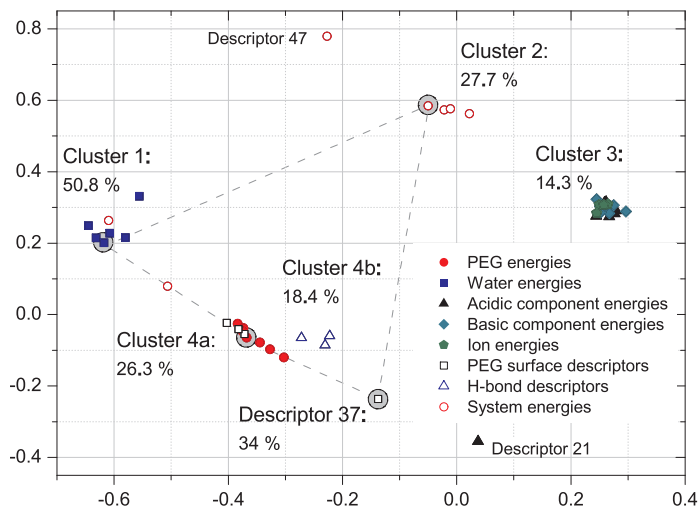


Figure 3: Cluster analysis of all descriptors. The Euclidean distance is a rough representation of the similarity of the descriptors: close points are similar in terms of modeling, distant points are different. The numbers below the cluster names indicate the frequency, with which energies from this cluster appear in the top 5 % models. The dashed gray line shows the descriptor combination of the overall best model.

Problematic was the interpretation of the role of descriptors that showed a significant reduction of model score without appearing in the best 5 % of all models (13 out of 25, mostly buffer related energies). To fully understand this, the informational content of the MD based descriptors was analyzed by a clustering approach to reveal differences in the informational content of each descriptor relative to all others. We ran a regression analysis of each combination of 2 descriptors by a 2nd degree polynomial function. The resulting R^2 were used to generate figure 3. Each point in the plot represents a descriptor and the distance between two points is a measure for their similarity in terms of a 2n degree polynomial correlation (represented by $1-R^2$ from the regression), meaning that close points have a high and distant ones a low similarity. Four clusters formed during the analysis indicating that there were groups of descriptors sharing a similar informational content.

5.4 Molecular dynamics simulations on aqueous two-phase systems: Binodal curve and tie line prediction

All buffer related energies were located in one cluster (cluster 3) explaining why only very few buffer descriptors were present in the top models as they all provide similar information. Apparently there are also descriptors not located in one of the cluster as they contained unique information; descriptor 37 was one of them again pointing out the special role PEG hydrophobicity. The dashed line in gray highlights the descriptors used for the overall best model. All were located in different clusters, maximizing the informational content of the model.

3.4 Mechanistic understanding

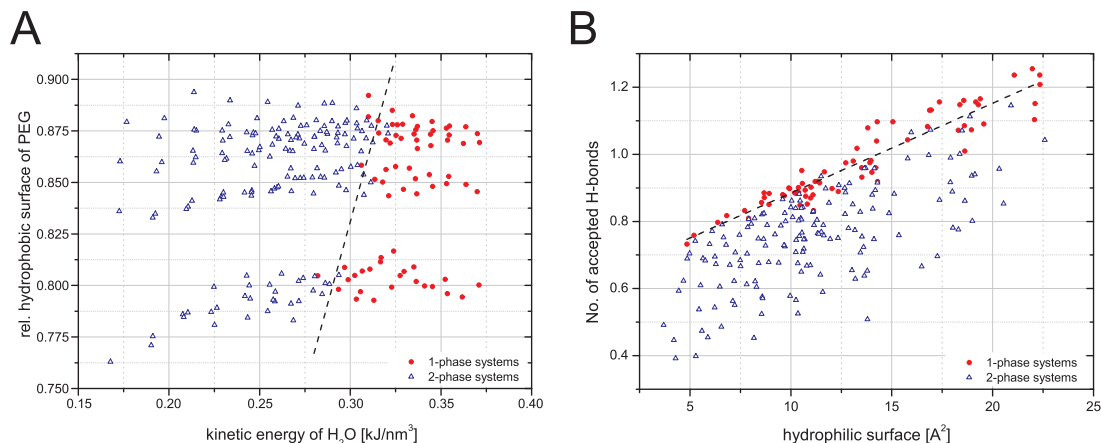


Figure 4: **A.** Critical kinetic energy of water (normalized per volume) as a function of the relative PEG hydrophobicity. **B.** Number of accepted H-bonds per PEG unit in relation to the hydrophilic surface ($-OH$ and $-O-$ contributions) of PEG per unit. The dashed line represents a linear fit of the 1-phase system datapoints with an $R^2 = 0.88$.

Since the cluster analysis showed that some descriptors could be grouped by similar informational content the following discussion should be viewed in light of that finding as other descriptors from the same cluster may result in models with a similar score. For example the total energy of H-bonds between water and PEG is in the same cluster as the total number of accepted H-bonds per PEG. Both descriptors are a measure for a similar property, namely the ability of PEG to form H-bonds. Thus when discussing the amount of H-bonds per PEG unit later in this section, discussing the total energy of H-bonds of PEG would yield similar insights.

3.4.1 Relative hydrophobicity of PEG

The only descriptor that did not cluster and still occurred in about 34 % of all top models was the relative hydrophobicity of PEG surface that was already identified in a previous publication [17] as one of the key discriminators for PEGs with different molecular weights. Oelmeier *et al.* showed, that with increasing molecular weight the surface of PEG becomes more hydrophobic due to the 'dilution' of hydrophilic endgroups and the formation of helical secondary structures (that was also more pronounced for higher molecular weight PEGs). Fig. 4 A shows its correlation with phase formation. It was generally relative hydrophobicity increased with PEG concentration and decreased with phosphate concentration for all PEGs with an average differences between minimum and maximum hydrophobicity of 3 % (PEG 300: 0.765 - 0.805, PEG: 600: 0.825 - 0.86, PEG 1000: 0.855 - 0.88 and PEG 1500: 0.87 - 0.89). At high PEG concentrations intramolecular connections between single PEG molecules were observed, excluding water molecules in that region. This mainly affected

5.4 Molecular dynamics simulations on aqueous two-phase systems: Binodal curve and tie line prediction

hydrophilic areas which explains the increase of relative hydrophobicity with increasing PEG concentration. With increasing phosphate concentration the helicality of PEG molecules decreased from about 58 % down to 45 % for PEG 1500 (30 % down to 19 % for PEG 300). As found by Oelmeier *et al.* [17] a decrease in helicality might correlate with a decrease in hydrophobicity as the surface contribution of oxygen atoms of PEG units increases.

3.4.2 Kinetic energy of water

Figure 4 A also shows the kinetic energy of water plotted versus the relative hydrophobicity of all four PEGs. There seems to be a critical kinetic energy of water that separates the 1-phase systems from the 2-phase systems, and this critical energy is a function of the hydrophobicity of PEG. From a thermodynamic point of view only the 1-phase systems are actually stable systems. Each system in the 2-phase region would separate into two systems in the 1-phase region both lying on the binodal curve. The kinetic energy of water represents entropic effects in the system, meaning that the formation of two stable phases is driven by a gain in entropy.

3.4.3 H-bond formation

Since H-bonds were not explicitly accounted for in the simulations but rather calculated after each MD run by screening for hydrogens of water molecules within a certain distance to H-bond acceptor atoms of PEG (mostly oxygen atom of PEG units), the observed overall trend was determined by the ratio of water to PEG molecules and the accessibility of the acceptor atoms. The general trend was again similar for all PEGs with a maximum number of H-bonds per PEG unit at low PEG and phosphate concentrations and a decrease with both, increasing PEG and phosphate concentration. Higher molecular weight PEGs formed less H-bonds per unit (0.4 to 0.9) than smaller PEGs (0.55 to 1.4). A possible explanation for that is again the helicality of PEG and thus the accessibility of oxygen atoms that act as H-bond acceptors (see discussion above). Supporting this theory figure 4 B shows the number of H-bonds versus the amount of hydrophilic surface per PEG unit. The 1-phase systems showed a very narrow distribution with a more or less fixed ratio between H-bonds and hydrophilic area (the slope is 0.0267 H-bonds/per A^2 surface), while the 2-phase systems all showed lower ratios with a wider distribution.

3.5 Tie line prediction

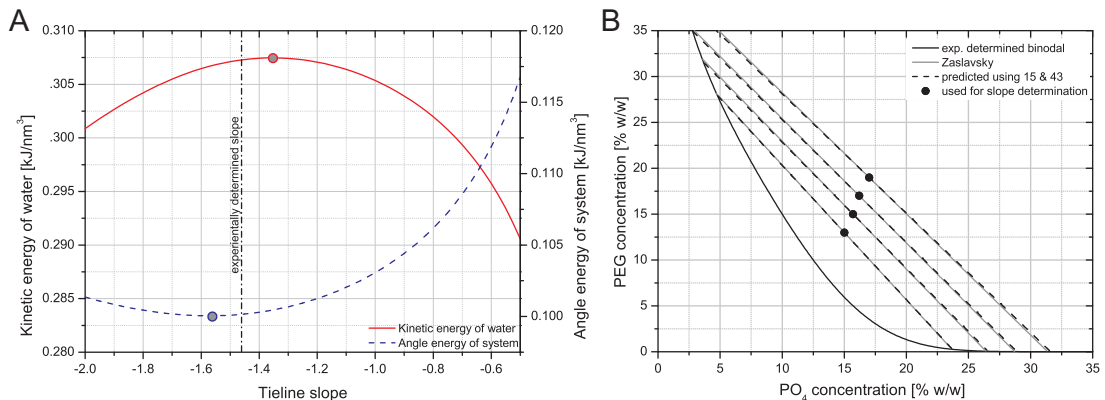


Figure 5: **A.** Kinetic energy (solid line) and bond energy (dashed line) of water as a function of the tie line slope for a point in the 2-phase region ($c_{PEG} = 13\%$, $c_{PO_4} = 15\%$) of PEG 1000. **B.** Predicted (dashed line) and experimentally determined (solid gray line, [25]) tie lines for PEG 1000.

5.4 Molecular dynamics simulations on aqueous two-phase systems: Binodal curve and tie line prediction

Each system with a total PEG and phosphate concentration in the 2-phase region, separates into a PEG-rich upper phase and a phosphate-rich lower phase, both lying on the binodal curve. When connecting these points, one obtains a tie line. All systems on a tie line have the exact same composition of upper and lower phase but with different volume ratios, according to the lever arm rule. Since the chemical composition of the phases does not change along a tie line, the distribution coefficient of a molecule A remains constant as well. Thus knowing the tie lines can be crucial for process development.

Theoretically, the system could separate into numerous different combinations of upper and lower phase compositions. Practically, from a thermodynamic point of view, it will separate into two systems that are energetically more favorable than the others. Since we had identified a series of descriptors that could be linked to phase formation, we examined these looking for compositions that were more favorable than other. For each PEG, four points were chosen in the 2-phase region for which experimental data was published by Zaslavsky [25] (data was only available for PEG 600, 1000 and 1500).

Of all descriptors only five descriptors revealed such a behavior: 11 (bond energy of water), 15 (kinetic energy of water), 38 (number of accepted H-bonds of PEG), 40 (H-bond energy of H-bonds formed with PEG) and 43 (total angle energy of the system), and among these five only descriptors 15 and 43 showed an average relative error below 10 %. A representative plot of descriptor values is shown in figure 5 A. This plot was generated for PEG 1000, a PEG concentration of 13 % and a phosphate concentration of 15 %. Both descriptors clearly showed an optimum that was in good agreement with the findings discussed above: 1) the kinetic energy of water reached a maximum at a slope of -1.35, which could be interpreted as a maximum in entropy of water, and 2) the total angle energy of the system reached a minimum at a slope of -1.58 which could be interpreted as a state, where the all molecules were in an energetically favorable conformation. Using the kinetic energy of water, the slopes showed an average error of -8.3 %, those determined by the total angle energy of the system showed an error of +9.6 %. Using the average of both slopes yielded a relative error of only 2.3 % over all PEGs, the results for PEG 1000 are shown in figure 5 B. (the average error for this PEG was 1.9 %). The calculated slopes confirmed a trend that was also found in the experimental data: the tie line slope became less negative with increasing distance from the critical point of the binodal, thus tie lines were not absolutely parallel. This was observed for all PEGs.

4 Protein distribution

As mentioned in the introduction, the main goal of this work was to evaluate MD simulations as a tool to characterize and predict protein distribution in ATPS. In a recent publication Madeira and Zaslavsky showed [12] that solvent properties of ATPS can be linked to distribution coefficients by using a modified version of the Abraham equation. To determine properties of the upper and bottom phase of a given ATPS, they used different solvatochromic dyes in order to characterize phase acidity (H-bond donor functionality, α), basicity (H-bond acceptor functionality, β) and polarity (π). Their manuscript was limited to polymer/polymer systems so we needed to validate their approach for polymer/salt systems. A high-throughput screening process was established in 384-well plates as described in the materials and methods section to determine α , β and π .

4.1 H-bond donor properties (α)

The α property relates to the H-bond acceptor property of the polymer-salt systems. In a PEG molecule, the only groups that can act as H-bond acceptors are the hydrophilic,

5.4 Molecular dynamics simulations on aqueous two-phase systems: Binodal curve and tie line prediction

terminal hydroxyl-groups so it was to be expected that the number of donated H-bonds per PEG units is relatively low and decreases with increasing chain length of PEG. Fig. 6. H-bond formation can be influenced by secondary structure formation. Nevertheless, helix formation should mainly influence the repetitive units rather than the terminal ones, so the effect on H-bond donor properties should be small.

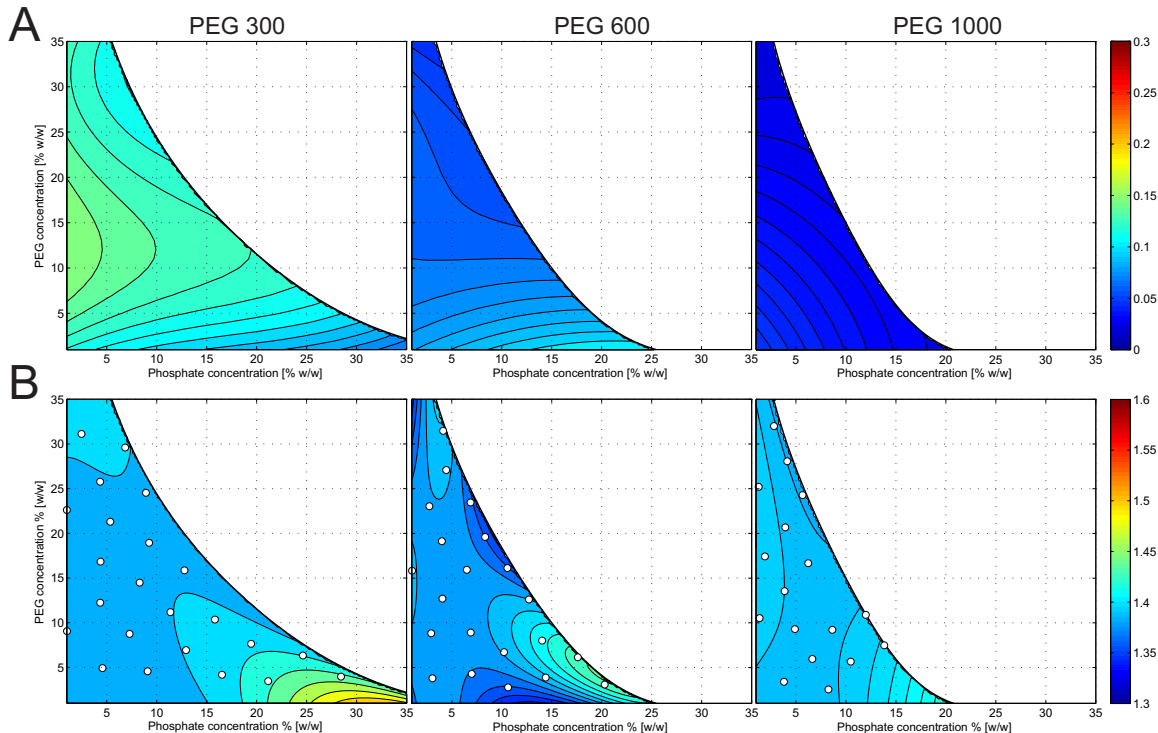


Figure 6: H-bond donor functionality of PEG in water: **A.** Alpha values determined by solvatochromic dye measurements in PEG-phosphate systems below the respective binodals. The symbols indicate data points used for regression. **B.** Number of H-bonds between PEG and water with PEG being the donor. Values are per PEG unit.

The simulation data revealed all these trends: generally low values compared to the number of accepted H-bonds (see following section), almost no change throughout the phase diagram since secondary structure formation has no effect but a very slight decrease with increasing PEG chain length. When comparing the simulation data to experimentally determined α values, similar trends could be observed, although the slight decrease with increasing PEG molecular weight could not be resolved.

4.2 H-bond acceptor properties (β)

The β property relates to the H-bond acceptor functionality and is expected to be more pronounced for PEG since each repetitive unit can act as an H-bond acceptor. Nevertheless, from correlations with phase formation that were discussed earlier, we knew that helix formation of PEG plays a crucial role here as it influences the accessibility of the oxygen atom that is responsible for H-bond formation. Increasing helicality with increasing molecular weight of PEG should therefore decrease the number of accepted H-bonds.

Again, the simulation results showed these trends. Additionally a decrease in the number of accepted H-bonds could be observed with increasing PEG concentration. This might be explained by the decreasing ratio of water to PEG molecules with increasing PEG concentration and the fact, that at higher concentrations intermolecular contacts appeared more often, excluding water from the PEG surface. The experimental data showed similar

5.4 Molecular dynamics simulations on aqueous two-phase systems: Binodal curve and tie line prediction

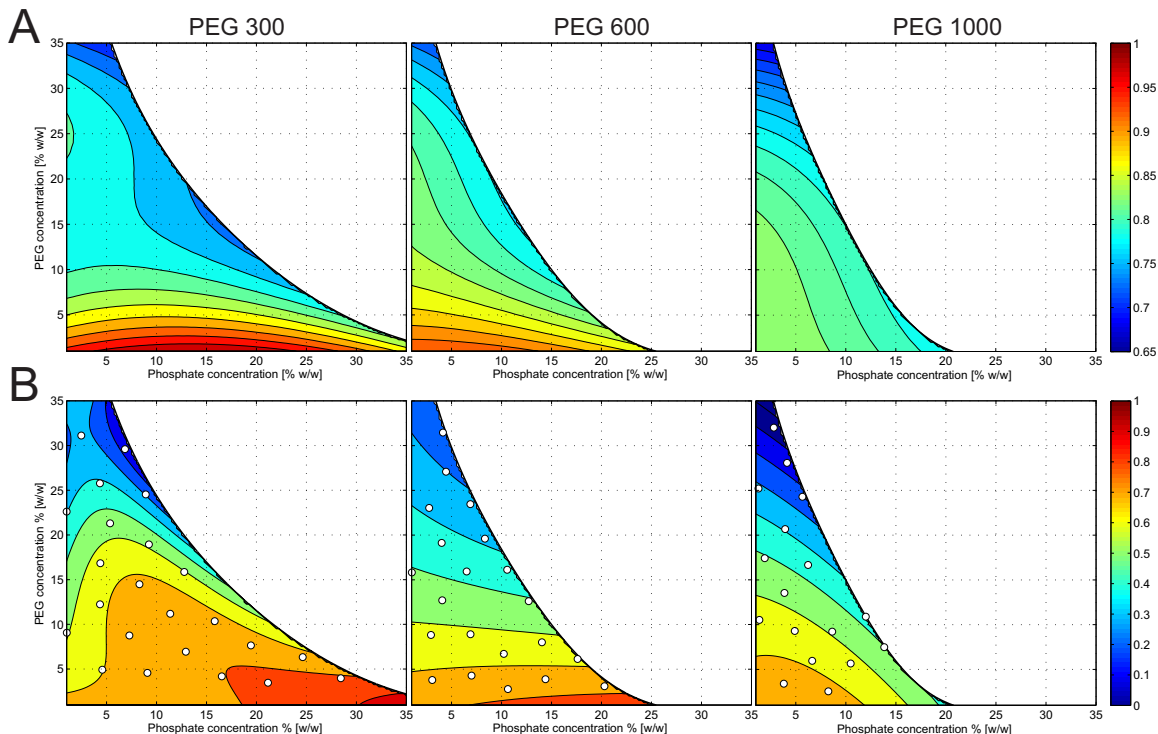


Figure 7: H-bond acceptor functionality of PEG in water: **A.** Beta values determined by solvatochromic dye measurements in PEG-phosphate systems below the respective binodals. The symbols indicate data points used for regression. **B.** Number of H-bonds between PEG and water with PEG being the acceptor. Values are per PEG unit.

trends: highest beta values at low PEG concentrations and a decrease with increasing PEG molecular weight.

4.3 Abraham equation for protein distribution

Generally, the agreement between solvatochromic dye measurements and MD simulation data was very good in terms of the influence of PEG and phosphate concentration as well as. Absolute values are not comparable between both datasets since the scales for α , β and π are arbitrary (for each parameter one solvent was chosen to have a value of 1, all other solvents have values relative to this reference system). The qualitative trends for experimentally determined pi values and rel. hydrophobicities from MD simulations were similar (data not shown). They showed only little differences throughout the phase diagram but slight differences with increasing PEG molecular weight. When building a model using the Abraham equation we found, that van-der-Waals interactions of PEG with surrounding molecules as a measure for polarity (or polarizability) yielded better models. The final model quality can be found in Fig. 8. The model quality increased by using the values normalized by volume rather than per molecule, although models could be built for both.

5 Conclusions and outlook

In this paper we have shown that it is possible to generate predictive models to determine binodal curves for PEGs with varying molecular weights from molecular dynamics simulations data. The resulting binodals were in good agreement with experiments. We also evaluated the effects of data preprocessing (data interpolation, normalization) and in-

5.4 Molecular dynamics simulations on aqueous two-phase systems: Binodal curve and tie line prediction

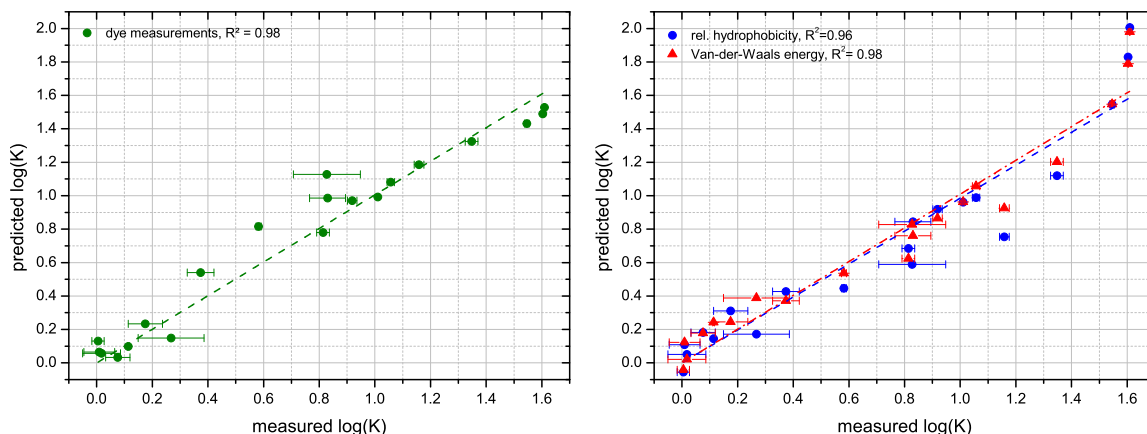


Figure 8: Correlation of phase system properties with lysozyme distribution coefficients for all four different PEG molecular weights. **A.** Correlation using solvatochromic dye measurements. **B.** Correlation using MD data. Two models are shown: a model based on the rel. hydrophobicity of PEG as third descriptor (blue) and a model based on Van-der-Waals energies (red).

creasing model complexity on both, model error and prediction quality. Data interpolation proved valuable as it improved the overall model score by about 30 % on average. With this finding, the number of MD simulations could probably be reduced from 64 down to 16 simulations per PEG (just enough for data interpolation). Data normalization also had an effect, although good models could be generated with both approaches. As data normalized to a constant simulation cell volume showed higher R^2 values when fitted for interpolation, this approach would be our choice for future studies. Energy clustering was successfully applied to identify codependent descriptors, helping to identify critical descriptors for phase separation. The clustering also pointed out why drawing mechanistic conclusions from such a modeling approach (and common QSPR and QSAR approaches as well) can yield misleading results, since some descriptors can be exchangeable if they have a similar informational content. Descriptors playing a key role during phase formation were identified; namely the relative PEG hydrophobicity, the kinetic energy of water and energies characterizing the geometrical state of the molecules in the simulations. Two of these descriptors, the kinetic energy of water and the total angle energy of the system could also be linked to the slope of the tie lines showing a correlation that was good enough to actually predict the slopes based on these two descriptors with an average error of 2.4 %. The scope of this paper was to show that molecular dynamics simulations are a suitable tool to study ATPS, and PEG-phosphate systems in particular. This might be an interesting approach to study protein distribution in such systems and to extract protein specific parameters influencing partitioning behavior. We observed changes in hydrophobicity of PEG and its ability for form H-bonds with surrounding molecules. According to the Abraham model both properties influence protein partitioning. We think that MD simulations might outmatch traditional 3D structure based prediction approaches as they have the potential to capture structural changes at high phosphate and PEG concentrations and they can additionally account for solvent effects, which might also be crucial in understanding protein partitioning.

References

- [1] B. A. Andrews, A. S. Schmidt, and J. A. Asenjo. Correlation for the partition behavior of proteins in aqueous two-phase systems: Effect of surface hydrophobicity and charge. *Biotechnology and Bioengineering*, 90(3):380–390, 2005.
- [2] A. M. Azevedo, P. A. J. Rosa, I. F. Ferreira, and M. R. Aires-Barros. Chromatography-free recovery of biopharmaceuticals through aqueous two-phase processing. *Trends in Biotechnology*, 27(4):240–247, 2009. 432.
- [3] D. R. Baughman and Y. A. Liu. An expert network for predictive modeling and optimal-design of extractive bioseparations in aqueous 2-phase systems. *Industrial & Engineering Chemistry Research*, 33(11):2668–2687, 1994.
- [4] M. Bensch, B. Selbach, and J. Hubbuch. High throughput screening techniques in downstream processing: Preparation, characterization and optimization of aqueous two-phase systems. *Chemical Engineering Science*, 62(7):2011–2021, 2007.
- [5] K. C. Chou and C. T. Zhang. Prediction of protein structural classes. *Critical Reviews in Biochemistry and Molecular Biology*, 30(4):275–349, 1995.
- [6] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, and et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(1919):5179–5197, 1995.
- [7] J. T. de Faria, F. C. Sampaio, A. Converti, F. M. L. Passos, V. P. R. Minim, and L. A. Minim. Use of response surface methodology to evaluate the extraction of *Debaryomyces hansenii* xylose reductase by aqueous two-phase system. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences*, 877(27):3031–3037, 2009.
- [8] R. Hatti-Kaul. Aqueous two-phase systems - a general overview. *Molecular Biotechnology*, 19(3):269–277, 2001.
- [9] A. Jakalian, D. B. Jack, and C. I. Bayly. Fast, efficient generation of high-quality atomic charges. am1-bcc model: Ii. parameterization and validation. *Journal of computational chemistry*, 23(16):1623–41, 2002.
- [10] J. T. Kent, K. V. Mardia, and S. R. Jammalamadaka. Characterization of the uniform-distribution on the circle. *Annals of Statistics*, 7(4):882–889, 1979.
- [11] E. Krieger, G. Koraimann, and G. Vriend. Increasing the precision of comparative models with yasara nova—a self-parameterizing force field. *Proteins*, 47(3):393–402, 2002.
- [12] P. P. Madeira, C. A. Reis, A. E. Rodrigues, L. M. Mikheeva, and B. Y. Zaslavsky. Solvent properties governing solute partitioning in polymer/polymer aqueous two-phase systems: Nonionic compounds. *Journal of Physical Chemistry B*, 114(1):457–462, 2010.
- [13] Y. Marcus. The properties of organic liquids that are relevant to their use as solvating solvents. *Chemical Society Reviews*, 22(6):109–416, 1993.
- [14] P. G. Mazzola, A. M. Lopes, F. A. Hasmann, A. F. Jozala, T. C. V. Penna, P. O. Magalhaes, C. O. Rangel-Yagui, and A. Pessoa. Liquid-liquid extraction of biomolecules:

5.4 Molecular dynamics simulations on aqueous two-phase systems: Binodal curve and tie line prediction

- an overview and update of the main techniques. *Journal of Chemical Technology and Biotechnology*, 83(2):143–157, 2008. 261.
- [15] G. McLachlan and D. Peel. *Finite Mixture Models*. John Wiley & Sons, New York, 2000.
- [16] J. C. Merchuk, B. A. Andrews, and J. A. Asenjo. Aqueous two-phase systems for protein separation studies on phase inversion. *Journal of Chromatography B*, 711(1-2):285–293, 1998.
- [17] S. Oelmeier, F. Dismer, and J. Hubbuch. Molecular dynamics simulations on aqueous two-phase systems: Single peg-molecules in solution. *BMC Biophysics*, submitted, 2012.
- [18] S. A. Oelmeier, F. Dismer, and J. J. Hubbuch. Application of an aqueous two-phase systems high-throughput screening method to evaluate mab hcp separation. *Biotechnology and bioengineering*, 108(1):69–81, Jan 2011.
- [19] D. M. Pericin, S. Z. Maddarev-Popovic, and L. M. Radulovic-Popovic. Optimization of conditions for acid protease partitioning and purification in aqueous two-phase systems using response surface methodology. *Biotechnology Letters*, 31(1):43–47, 2009.
- [20] J. C. Salgado, B. A. Andrews, M. F. Ortuzar, and J. A. Asenjo. Prediction of the partitioning behaviour of proteins in aqueous two-phase systems using only their amino acid composition. *Journal of Chromatography A*, 1178(1-2):134–144, 2008.
- [21] J. E. Schiel and D. S. Hage. Density measurements of potassium phosphate buffer from 4 to 45 degrees c. *Talanta*, 65(2):495–500, 2005.
- [22] J. Schindler and H. G. Nothwang. Aqueous polymer two-phase systems: Effective tools for plasma membrane proteomics. *Proteomics*, 6(20):5409–5417, 2006. 101.
- [23] S. C. Silverio, O. Rodriguez, J. A. Teixeira, and E. A. Macedo. Solute partitioning in polymer-salt atps: The collander equation. *Fluid Phase Equilibria*, 296(2):173–177, 2010.
- [24] J. Wang, R. M. Wolf, J. W. Caldwell, P. a. Kollman, and D. a. Case. Development and testing of a general amber force field. *Journal of computational chemistry*, 25(9):1157–74, 2004.
- [25] B. Y. Zaslavsky. *Aqueous two-phase partitioning : physical chemistry and bioanalytical applications*. M. Dekker, New York, 1995.

MANUSCRIPT V

HTS BASED SELECTION OF PHASES FOR ATPS CPC OF MABS

Stefan A. Oelmeier, Jürgen Hubbuch*

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

* Corresponding author. Tel.: +49 721 608-42557; fax: +49 721 608-46240. E-mail address: juergen.hubbuch@kit.edu

*published in revised form in: **Journal of Chromatography A**. 2012;1252:104-114*

Abstract

Aqueous two-phase systems have been demonstrated to be a possible alternative to chromatographic separations during the industrial purification of proteins. While convenient high throughput screening methods were shown to drastically reduce experimental effort for the evaluations of ATPS as a unit operation, the selection of which phases to investigate is currently guided largely by prior knowledge. Correlations between protein descriptors and distribution were found, but the general applicability of such correlations especially under conditions of high protein load is questionable, as currently no correlation take the saturation of the phases with protein into account. In this manuscript, we demonstrate how precipitation experiments using the phase forming components can guide the selection of both system type and tieline length for the purification of monoclonal antibodies. Good qualitative correlations between precipitation data and both distribution and recovery of the target molecule were found. Most promising system were selected for upscale to a 500 mL CPC. Influence of operation condition on the column and on HCP clearance was investigated. An increase in HCP clearance of more than threefold compared to batch extractions was observed. The importance of load protein concentration underlined the value of using a screening approach that incorporated target protein solubility data.

Keywords: *centrifugal partitioning chromatography, monoclonal antibody, host cell protein, process development, high throughput screening*

1 Introduction

Aqueous two-phase systems (ATPS), a biopurification technique known for many decades, has recently received increased interest as an alternative separation technique for the biotechnological industry. ATPSs promise a low cost, easily scalable unit operation with high selectivity and is discussed as an alternative for chromatographic separations [1]. One major obstacle to the implementation of aqueous two-phase systems for the preparative purification of proteins is the lack of predictive models. The distribution in ATPS is governed by multiple factors such as protein hydrophobicity, charge, and size as well as system characteristics such as PEG molecular weight, ion distribution, and pH. Several correlations between protein or system descriptors and protein distribution have been demonstrated. For certain conditions, correlations were found between molecular weight and distribution [2], charge and distribution [3], as well as hydrophobicity and distribution [4–6]. At conditions under which no phase is saturated with protein, the distribution is influenced solely by physicochemical factors such as those detailed above. If one phase is however saturated with protein, the measurable apparent distribution starts to deviate and becomes a function of total system protein concentration. The apparent distribution under such conditions is strongly influenced by the solubility of the distributed molecule in the two phases. For preparative purification of proteins a high system protein load is desirable. Thus the effect of saturating one or both phases with protein needs to be taken into account by the selection scheme. As the underlying effects governing distribution in ATPS are little understood, neither scale nor selection theme currently exists and process development is largely guided by experience and prior knowledge. The application of PEG-phosphate systems for the purification of monoclonal antibodies (mAb) has previously been demonstrated using PEG4000 systems [7, 8]. Conditions were found under which a reduction of host cell proteins by approximately 50% was achieved while fully recovering the target protein. Even though total system mAb load was only up to 0.75 mg/mL in these studies, mAb recovery was below 90% for most tested conditions. While the application of HTS platforms greatly decreases the experimental effort for the evaluations of ATPS, a rational scheme guiding the selection of the conditions to be investigated could further decrease and systematize

the experimental work. To aid the system selection for a preparative application of ATPS and to improve capacity, target protein solubility would need to be taken into account.

One way to improve resolution of aqueous two-phase extractions while concurrently scaling the extraction up is to use centrifugal partitioning chromatography (CPC). CPC has been demonstrated to be applicable to ATPS. Sutherland *et al.* [9] used a PEG1000-PO4 system to separate two model proteins (lysozyme and myoglobin) using a 500 mL CPC system. With distribution coefficients of 1.3 and 0.7 respectively, a single batch extraction step would hardly generate any resolution of the two proteins. In CPC however, baseline separation was achieved under these conditions. As CPC experiments are high both in time and material consumption, a scheme to select the phases to upscale from batch to CPC on a rational basis needs to be devised. For organic extractions, polarity scales and the “Arizona” system greatly simplifies phase selection [10]. No similar scale or selection scheme currently exist for aqueous two-phase systems.

The aim of this study is to devise a ATP-system selection scheme that takes into account the effect of overloaded systems that occur when maximizing system protein load. The selection scheme thus needs to consider the influence of protein solubility on the apparent distribution of the protein. Additionally, selecting phases for their application in centrifugal partitioning chromatography is guided by the data generated. All data needed for phase selection are generated on a high throughput screening platform. PEG-PO4 systems and the separation of monoclonal antibodies from host cell proteins are used in this study. Scale-up from 650 μ L batch system to a 500 mL CPC system is demonstrated and the influence of operating parameters on HCP clearance, mAb dilution and recovery are evaluated.

2 Materials & methods

2.1 Chemicals

Stock solutions were prepared in 50 mM phosphate buffer pH 6.0 as follows: 70% [w/w] PEG400, 70% [w/w] PEG600, 70% [w/w] PEG1000, 60% [w/w] PEG1450. Phosphate stock solutions were prepared in dH_2O as 40% [w/w] NaH_2PO_4 , and 40% [w/w] K_2HPO_4 . All chemicals were bought from Sigma-Aldrich. 52.54 g NaH_2PO_4 solution and 47.46 g K_2HPO_4 were combined to yield pH 6.0. NaH_2PO_4 and K_2HPO_4 were chosen for the higher solubility of these salts compared to Na_2HPO_4 and KH_2PO_4 .

2.2 Proteins

Four monoclonal antibodies were used in this study. All mAbs were derived from CHO cell culture and were of the same subtype. mAb stock solutions varied in protein concentration between 3.6 g/L and 16 g/L. As impurities, the clarified supernatant of a mock fermentation was used. This host cell protein (HCP) solution was concentrated by a factor of 10 to yield the HCP stock solution used in this study. UV absorption gave a HCP concentration of 4.1 g/L for the 10x solution.

2.3 Liquid handling station

In this study, a Tecan Freedom Evo 200 system (Tecan, Crailsheim, Germany) was used as liquid handling platform. It is equipped with three robotic arms: one 8-port liquid handling arm with 8 fixed tips, one standard robotic plate handling arm equipped with a centric gripper as well as a 96 channel liquid handling arm equipped with an eccentric gripper. Additionally, the system has an integrated centrifuge (Rotanta 46RSC, Hettich, Tuttlingen, Germany), a rotational shaker (Te-Shake, Tecan, Crailsheim, Germany) outfitted with

a PreDictor frame (Tecan, Crailsheim, Germany), and an Infinite200 spectrophotometer (Tecan, Crailsheim, Germany).

2.4 Disposables

For spectroscopic measurements Greiner Bio-One (Kremsmüster, Austria) UV-Star plates (article no. 655801) were used. ATPS were prepared in 1.3 mL Nalgene Nunc (Rochester, NY, USA) Deep Well plates (product no. 260252). For all other purposes Greiner polypropylene flat bottom MTPs were used (article no. 655261).

2.5 Software

Excel 2010 (Microsoft, Redmond, WA, USA) files were used as import format and for data storage. All calculations, evaluation and visualization of data were done using Matlab R2011a (The Mathworks, Natick, ME, USA). The robotic workstation was controlled using Evoware 2.2 standard (Tecan, Crailsheim, Germany). The spectrophotometer was controlled using Magellan 6.4 (Tecan, Crailsheim, Germany).

2.6 ATPS screening method

Aqueous two-phase system distribution and recoveries of four monoclonal antibodies and a host cell protein pool were performed on a liquid handling platform as described by Oelmeier *et al.* [8]. In short, the systems are put together in deep well 96 well plate by pipetting the appropriate stock solutions to yield a total system volume of 650 μL . The plate is shaken to ensure equilibrium between the two phases. Phase separation by centrifugation is followed by sampling of both phases (30 μL sample each) and measuring the protein content of the diluted (10x) sample. Including blanks approximately 2600 ATPSs were analyzed for this study. Four different PEG molecular weight were used in this study: 400, 600, 1000 and 1450 Da. All ATPS were composed of PEG and phosphate buffer with a pH of 6.0. Binodal data was fitted to equation 1 as described by Merchuk *et al.* [11]. Protein distribution and phase volumes were measured at least in triplicate, blank values at least in duplicate. Mean values of protein distribution and recovery are reported throughout this manuscript.

2.7 Precipitation screening method

A protein precipitation screening method on the liquid handling station was developed. Buffer, precipitant stock solution, and protein stock solution are combined to yield a sequence of samples with equal protein concentration but increasing precipitant content. Phase forming components were used as precipitants. Sample volume was 300 μL . Sample container were standard 96 well plates made of polypropylene. Using the rotational shaker samples were mixed and incubated for at least 15 minutes, at most 24h. Different shaking / incubation 15 minutes and 24 hours were evaluated for their influence on m^* . If the incubation time was larger than 1 hour, plates were sealed to avoid evaporation. After incubation, plates were centrifuged at 2900g for 30 minutes at 25°C thus removing aggregates from the supernatant. Samples were taken from the solution's surface and diluted 1:10 to yield a total volume of 300 μL . Protein content was determined by measuring sample absorption at 280 nm wavelength. Plotting the logarithm of supernatant protein concentration against the precipitant concentration two linear region are discernible. At low precipitant concentration, no protein precipitation occurs, thus protein concentration is constant. At high precipitant concentrations, protein content in the supernatant decreases exponentially as described by the Cohn equation [12]. Precipitations were conducted in triplicate at least and mean values are reported throughout this manuscript.

2.8 Centrifugal partitioning chromatography

Centrifugal partitioning chromatography (“CPC”) experiments were performed on an Armen Instruments 500 machine. An Armen Instruments HPLC pump delivered solvent flow to the column. An Äkta prime (GE Healthcare) system was used for online UV and conductivity signals and fraction collection. All column experiments were conducted in ascending mode, using the upper phase as mobile phase. 2.5 kg of an ATPS were made and settled in a separating funnel for approximately 60 minutes. The pump inlet tubing was put into the separating funnel to draw lower phase. For column equilibration with the stationary phase, the column was operated at 500 rpm and a flow of 50 mL/min was delivered for 13 minutes. Subsequently, the lower phase was drained from the separating funnel. Flow rate and rotation was adjusted to the set values and mobile phase delivered to the column. Fraction collection was started together with delivery of mobile phase to the column. Fraction size was 10 mL. Samples were injected 10 minutes after the start of the elution. Samples of 10 mL volume were injected using a manual Rheodyne injection valve. Samples of higher volume were injected by pausing the pump, transferring the pump inlet tubing into the sample container, and restarting the pump. The elution was continued for at least 90 minutes. Samples were analyzed for stationary phase content using graduated fraction containers. mAb and HCP content were analyzed as described in section 2.9

2.9 Protein content analytics

In the high throughput screenings, protein content was measured by UV-absorption at 280 nm with appropriate blanks subtracted. Distribution coefficients were calculated as the ratio of protein concentration in the upper phase to the concentration in the lower phase. CPC fractions were analyzed by ProteinA-HPLC for mAb concentration. HCP content in CPC fractions were measured using a proprietary ELISA-like assay.

3 Results

3.1 Precipitation by phase forming components

Precipitation screenings were performed using the phase forming components. Figure 1 shows the results of precipitations using PEG of all four molecular weights, the four mAbs, and HCP. In general, protein solubility decreases with increasing PEG MW. Each mAb showed a distinct solubility. mAb2 had the highest solubility with mAb6, mAb1, and mAb4 following in decreasing order. As an example, supernatant recovery of the mAbs at a concentration of 30 % PEG600 was 72.4% for mAb2, 34.7% for mAb6, 13.8% for mAb1, and 2.7% for mAb4. HCP generally precipitated less than the mAbs, but HCP precipitation also increased with the molecular weight of the PEG. Maximum ratio of precipitated HCP found during these experiments was with PEG1000 and PEG1450 at concentrations above 30 % [w/w]. 25% reduction of supernatant HCP concentration was observed under these conditions. mAb concentration was reduced by at least 80 % under these conditions. Certain conditions might allow precipitating the antibody while keeping most of the HCPs in solution. For example at a level of 18% PEG1450, 98% of mAb4 and 94% of mAb1 precipitated, while 88% of HCPs stayed in solutions.

Three additional sets of precipitation screenings were performed: precipitations using phosphate buffer, precipitations using PEG with 5% [w/w] phosphate buffer added, and precipitations using phosphate buffer with 5% PEG added. The latter two were conducted due to the following rational: in aqueous two-phase systems, no phase is completely void of any of the two phase forming components. At short tielines, the level of the non-dominating

5.5 HTS based selection of phases for ATPS CPC of mAbs

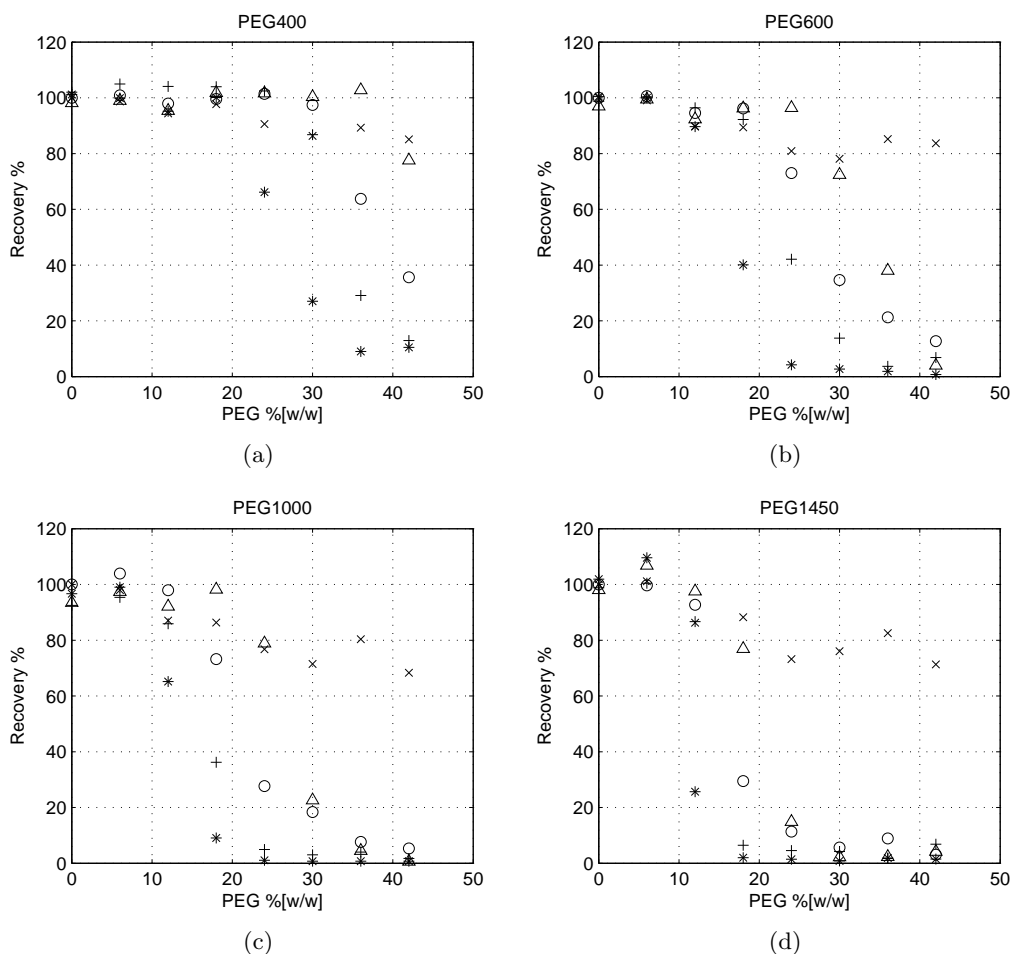


Figure 1: Supernatant recovery in percent of total mAb amount after precipitation with PEG. mAb1: +, mAb2: Δ , mAb4: *, mAb6: \circ , HCP: \times

component is higher than at longer tielines. With decreasing PEG molecular weight, this concentration also tends to increase. Binodal and tieline data suggested, that systems composed of PEG400 and phosphate had a phosphate concentration of approximately 5%. The aim of this investigation was to evaluate if precipitation data of target proteins can be used to predict the proteins behavior in aqueous two-phase systems under overloading conditions. Thus, in order to judge, which precipitation screen yield the best predictions, precipitations with and without the addition of 5% of the non-dominating phase forming component were conducted. The addition of 5% phosphate to PEG solutions increased mAb solubility significantly for all mAbs. Data density does not allow for an exact quantification of this effect, protein precipitation however set in one data point later, which equals additional 6% of precipitant. Likewise, addition of 5% PEG also increased mAb solubility, with the effect decreasing with increasing PEG molecular weight. Addition of PEG400 increased the starting precipitating phosphate concentration by an average of 3%. The effect was significantly lower for PEG600. It was thus decided to use data from precipitations by combined phase forming components only for the prediction of protein behavior in PEG400 systems.

Table I Binodal curve equation coefficients A, B, C of PEG-phosphate systems and the average tieline slope (TLS) determined in these systems.

PEG MW	A	B	C	TLS
400	72,02	-0,34	-6,89E-05	-1,28
600	68,87	-0,37	-1,54E-04	-1,38
1000	95,80	-0,54	-1,96E-04	-1,65
1450	88,72	-0,55	-2,24E-04	-1,69

3.2 Protein distribution and recovery in ATPS

3.2.1 ATPS characterization and system point selection

All four systems were initially characterized by determination of the binodal curve using a cloud point method as described in [8]. Tielines were determined using the lever arm rule and measured phase volumes. The average tieline slope was calculated based on five to ten different system compositions. Table I shows the resulting coefficients A , B , and C from equation 1 as well as the average tieline slope.

$$[PEG] = A * exp(B * [PO_4]^{0.5} + C * [PO_4]^3) \quad (1)$$

The influence of contact time of the proteins with the ATPSs before centrifugation was evaluated. Precipitation kinetics were found to differ depending on system composition. Precipitation rates were found sufficiently low as to show differences in recovery and distribution over the course of 2 hours (data not shown). It was decided to implement an additional incubation time of 45 minutes, thus increasing the total contact time of the proteins with the ATPS to 60 minutes to mimic realistic contact times in a theoretical production scenario.

Based on the average tieline slope, five systems point were selected per PEG molecular weight, with increasing tieline length and a volume ratio of approximately one. Protein distribution and recovery of all four mAbs were determined at least at total protein concentrations 0.5 mg/mL, 1 mg/mL, and 1.5 mg/mL. System with closed mass balances at 1.5 mg/mL were additionally tested with total protein concentration of 2.5 mg/mL, 3.5 mg/mL, and 6 mg/mL. Host cell protein distribution and recovery was tested at a total concentration of 0.41 mg/mL, mimicking the cell culture supernatant level.

3.2.2 Correlation of ATPS distribution and recovery to precipitation data

Figures 2 to 5 show the results of the four mAbs tested. The binodal curve and the tielines of the tested systems are shown. At the intersection of the tielines with the binodal a dot plot represents the recovery of the mAb in the upper and lower phase respectively. The distribution of the mAbs can be estimated from the recoveries in the two phases. Bar plots along the X- and Y axes show the results of precipitation screening using the phase forming components. Initial concentration of both precipitation and ATPS screening was 1 g/L. For PEG400-PO4 systems, precipitation results of PEG400 with 5% PO4 as precipitant are shown.

Several general trends were seen in the ATPS distribution and recovery data of the mAbs. Distribution coefficients increased with increasing tieline length, as long as no precipitation occurred. Distribution coefficients however decreased with decreasing recovery and with increasing PEG molecular weight. Recovery decreased both with increasing tieline length and with increasing PEG molecular weight.

5.5 HTS based selection of phases for ATPS CPC of mAbs

Table II The twenty system points ("SP") chosen based on binodal, tieline, and solubility data for the four classes of ATPS. Tieline length increases from SP1 to SP5 in each system.

PEG 400	SPI	SP2	SP3	SP4	SP5
PEG %[w/w]	15	16,5	18	20	22,5
PO4 %[w/w]	17,15	17,45	17,7	18	18,35
PEG 600	SPI	SP2	SP3	SP4	SP5
PEG %[w/w]	11,6	13,7	15,5	17,2	18,7
PO4 %[w/w]	16	16	16,1	16,3	16,6
PEG 1000	SPI	SP2	SP3	SP4	SP5
PEG %[w/w]	12,2	13,3	14,8	16,2	18,2
PO4 %[w/w]	12,6	12,8	13,1	13,5	14,1
PEG 1450	SPI	SP2	SP3	SP4	SP5
PEG %[w/w]	12,1	13,2	14,3	15,3	16,4
PO4 %[w/w]	11,6	11,9	12,1	12,4	12,7

From figures 2 to 5 it can be seen that the precipitation data could be used as a predictor for the distribution and recovery of the mAbs in all four system types tested. For PEG400 and PEG600, phosphate levels in the bottom phase were too high to allow protein distribution into the bottom phase. PEG concentrations in the upper phase were low enough to permit the mAbs to be soluble in the upper phase. The resulting distribution in these systems was very one sided with K values above 50. As tieline length increases, PEG concentrations in the upper phase reached levels that precipitated the mAbs, thus decreasing the overall recovery. With increasing molecular weight, bottom phase phosphate concentration decrease. With PEG1450 as phase forming component, bottom phase phosphate concentrations for the first two system points tested were sufficiently low to allow for higher concentrations of mAb to partition in this phase. Concurrently, PEG1450 concentrations in the upper phase were too high to admit distribution of mAbs into the upper phase. Thus the distribution was reversed in these systems.

In figure 6, the concentrations in the top and bottom phase as well as the resulting distribution coefficients are plotted against the total system protein concentration for the least soluble mAb (mAb4) in a PEG400-phosphate system (SP1). While mAb concentration in the bottom phase was approximately constant independent of the total system concentration, mAb concentration increased linearly with an increasing total protein concentration. The resulting distribution coefficient increased accordingly. This suggests, that the bottom phase was already saturated with mAb at a very low concentration (approximately 0.05mg/mL). Any additional mAb added to the system distributed into the upper phase. All tested total protein concentrations yielded closed mass balances, with the highest top phase concentration at 12.64 mg/mL. Identical trends were seen for other system points and proteins. This leads to the following two conclusions. One, the systems were overloaded, the distribution of the mAbs under these conditions was a function of the solubility of the mAbs in the two phases rather than an equilibrium of the attractive and repelling forces that guide distribution under non-overloaded conditions. Two, the precipitation data could be used as a good predictor for the behavior of the mAbs in the systems.

Host cell proteins distributed more evenly across the two phases than antibody. Dis-

5.5 HTS based selection of phases for ATPS CPC of mAbs

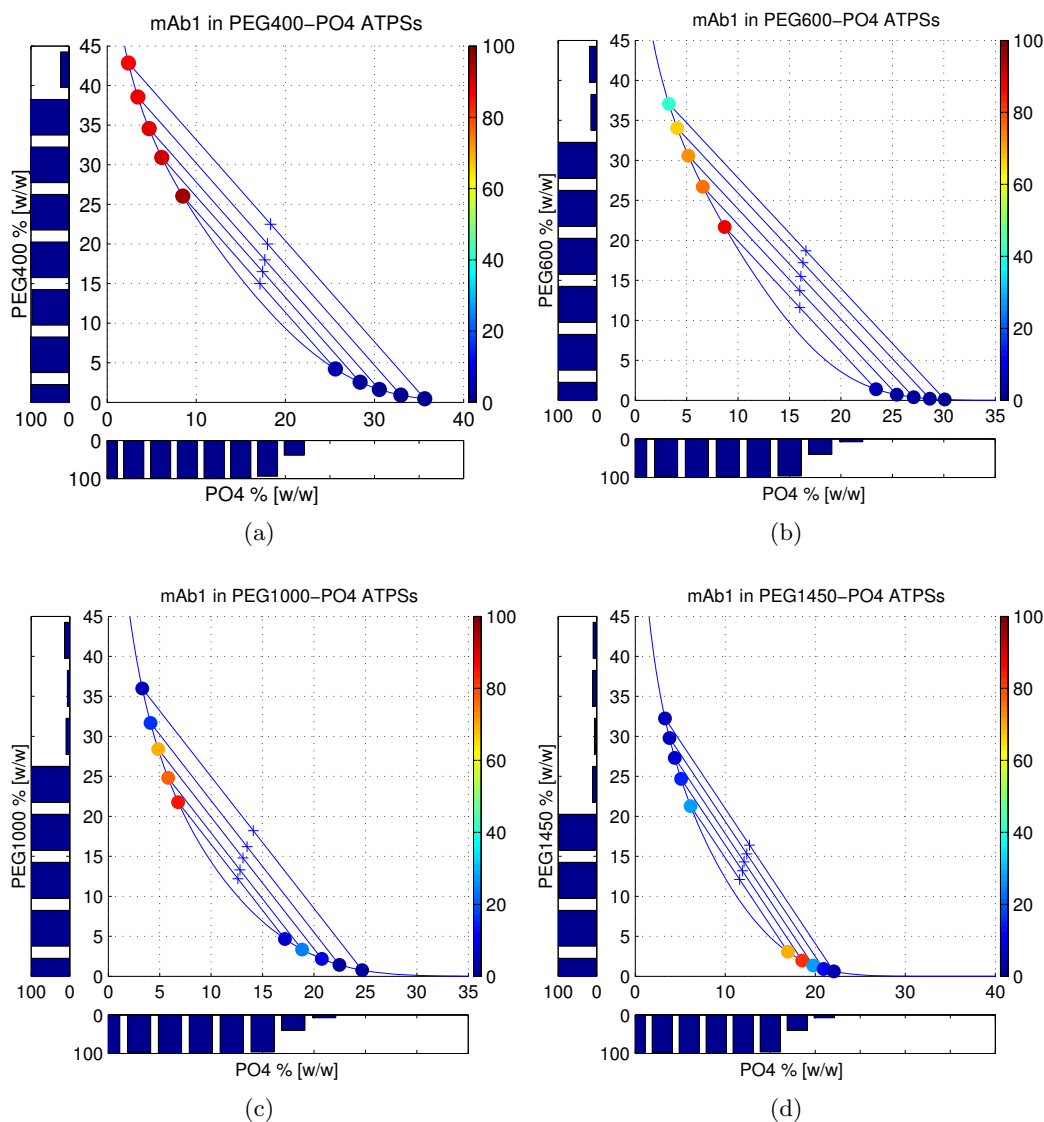


Figure 2: Result of precipitation and ATPS distribution screenings using mAb1. Bar plots along the axes represent results of precipitation screenings with phase forming components. *marks the total system compositions tested. Dot plots at the top and bottom phase composition represent mAb yield in the respective phase, coded by color.

tribution coefficients increased with increasing tieline length from 3.9 to 5.6 in PEG400 systems, from 2.1 to 2.8 in PEG600 systems, and from 1.0 to 1.1 in PEG1000 systems. In PEG1450 systems distribution coefficient of HCP was 0.7 regardless of tieline length. The trend of decreasing distribution coefficient with increasing PEG molecular weight was thus also seen for HCP. The effect of tieline length on the distribution coefficient decreased with increasing PEG molecular weight. Recoveries of HCP were generally high, but decreasing with increasing tieline length. Average recoveries were 88% for PEG400 systems, 72% for PEG600 systems, 100% for PEG1000 systems, and 95% for PEG1450 systems. However, no conditions were found in which mAbs and HCP distributed different enough to yield high reductions of HCP levels in a single batch step, thus justifying the attempt to increase resolution by use of CPC.

Based on the above data, three system points were chosen for upscale to centrifugal partitioning chromatography: SP1 and SP2 of the PEG400-phosphate systems and SP1 of

5.5 HTS based selection of phases for ATPS CPC of mAbs

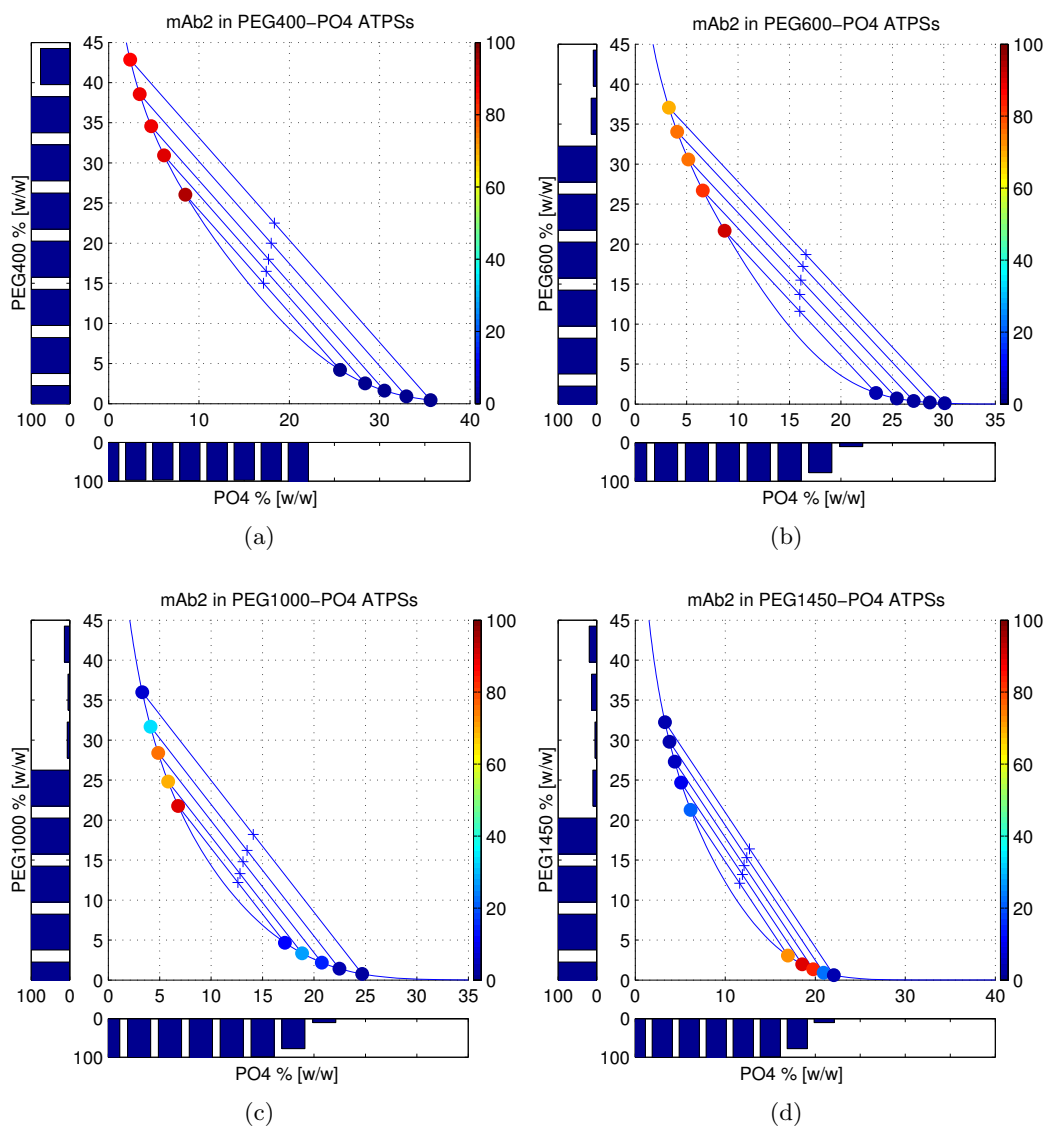


Figure 3: Result of precipitation and ATPS distribution screenings using mAb2. Bar plots along the axes represent results of precipitation screenings with phase forming components. *marks the total system compositions tested. Dot plots at the top and bottom phase composition represent mAb yield in the respective phase, coded by color.

the PEG600-phosphate systems.

3.3 CPC column characterization

In order to select operating conditions for centrifugal partitioning chromatography, the composition and stability of a column in dependence of the operating conditions needs to be judged. The systems PEG400 SP2 and PEG400 SP3 were selected for this evaluation. The systems were selected as they promised both a robustness due to the length of the tieline, a favorable mAb/HCP distribution, and high mAb capacities. A two-factor, two level full factorial designed experiment was conducted with flowrates between 5 mL/min and 15 mL/min and rotation speed between 1500 RPM and 2500 RPM (equaling 125 to 350-g on this rotor). The influence of the aforementioned factors on column stability, column composition at mobile phase breakthrough, column composition at the steady state,

5.5 HTS based selection of phases for ATPS CPC of mAbs

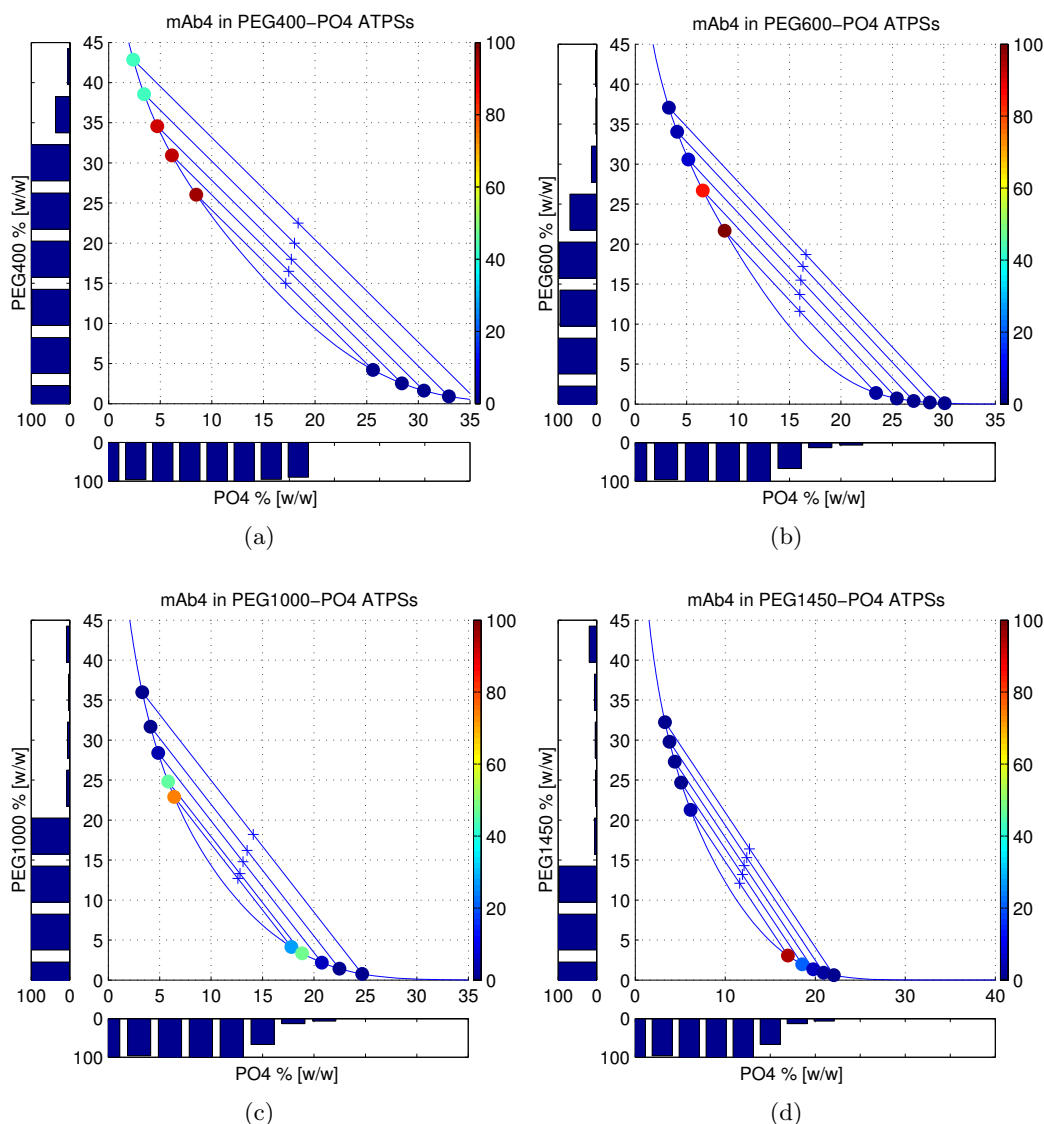


Figure 4: Result of precipitation and ATPS distribution screenings using mAb4. Bar plots along the axes represent results of precipitation screenings with phase forming components. *marks the total system compositions tested. Dot plots at the top and bottom phase composition represent mAb yield in the respective phase, coded by color.

and column back-pressure was investigated. Column stability was found to be high after reaching a steady state. Ratio of mobile to stationary phase in the eluate was above 100 for all conditions and both systems tested. Thus hardly any column bleeding was observed. Other investigated factors varied with flow-rate and RPM. Column back pressure varied ranged from 19 to 56 bar for SP2 and from 24.5 bar to 67.9 bar for SP3. Highest back pressure was observed at 2500 RPM and 5 mL/min. Lowest backpressure was measured at 1500 RPM and 5 mL/min. Column backpressure was thus found to be mainly influenced by rotation rate. Column ratio at the stable state, expressed as volume of mobile phase to total system volume ranged from 0.37 at 5 mL/min and 1500 RPM to 0.65 at 15 mL/min and 2500 RPM for SP2 and from 0.27 at 5 mL/min and 2500 RPM to 0.55 at 15 mL/min and 2500 RPM for SP3. Column composition was thus found to be a result of the interactions of the two operating parameters. For visual representation, the results were fitted using a cubic spline interpolation. Contour plots thereof are shown in figure 7.

5.5 HTS based selection of phases for ATPS CPC of mAbs

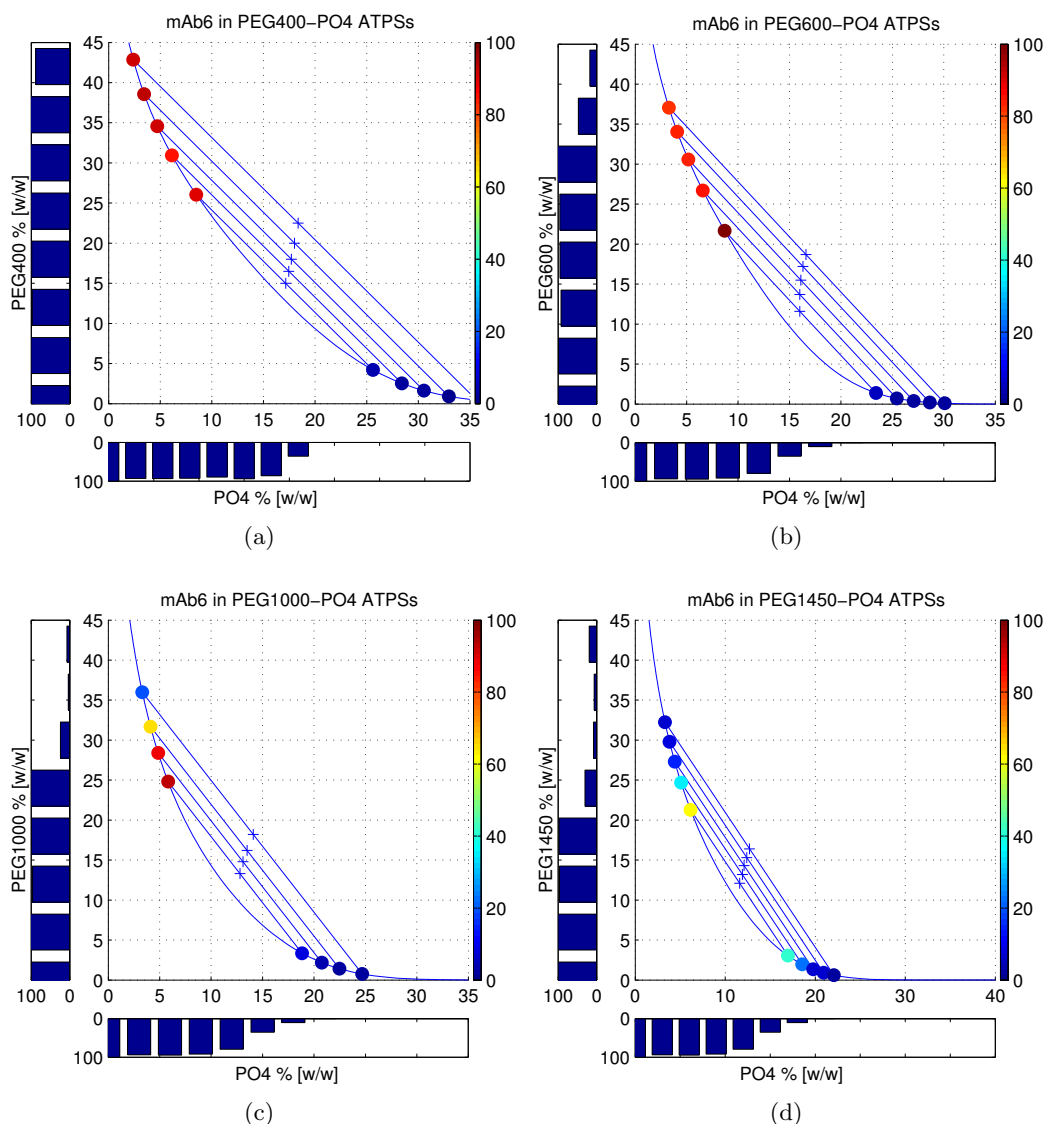


Figure 5: Result of precipitation and ATPS distribution screenings using mAb6. Bar plots along the axes represent results of precipitation screenings with phase forming components. *marks the total system compositions tested. Dot plots at the top and bottom phase composition represent mAb yield in the respective phase, coded by color.

3.4 mAb elution on CPC column

For mAb/HCP separation on the CPC column, the PEG400-SP2 system was chosen. This system promised higher possible loads and generated less backpressure. Less backpressure allows for higher rotation speed, which can be beneficial to the number of theoretical plates of the column [13]. Additionally, the systems were selected such that the target protein distributed strongly into the mobile phase in order to reduce target protein dilution. The model mAb used had distribution coefficients above 200 in the systems used, while HCP distribution coefficient was 4.8 in SP2. mAb concentration in the load was 1 mg/mL, HCP was diluted to yield 1x concentration. Load volume was 10 mL. The column operating conditions also used in the stability studies were again used to investigate their influence on HCP clearance and mAb recovery. Figure 8(a) shows the resulting HCP clearances plotted against the achieved mAb yield in dependence of the pooling of the fractions. Application

5.5 HTS based selection of phases for ATPS CPC of mAbs

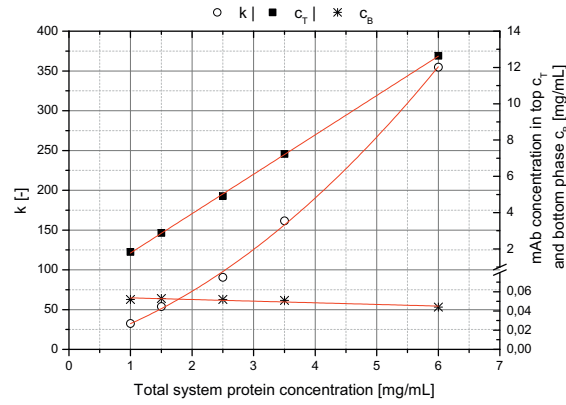


Figure 6: Top (c_T), bottom (c_B) phase concentration, and distribution coefficient of mAb4 in the PEG400-SP1 system as a function of total system mAb concentration. While the bottom phase concentration stays approximately constant, the top phase concentration and thus the distribution coefficient increase with increasing total system mAb concentration.

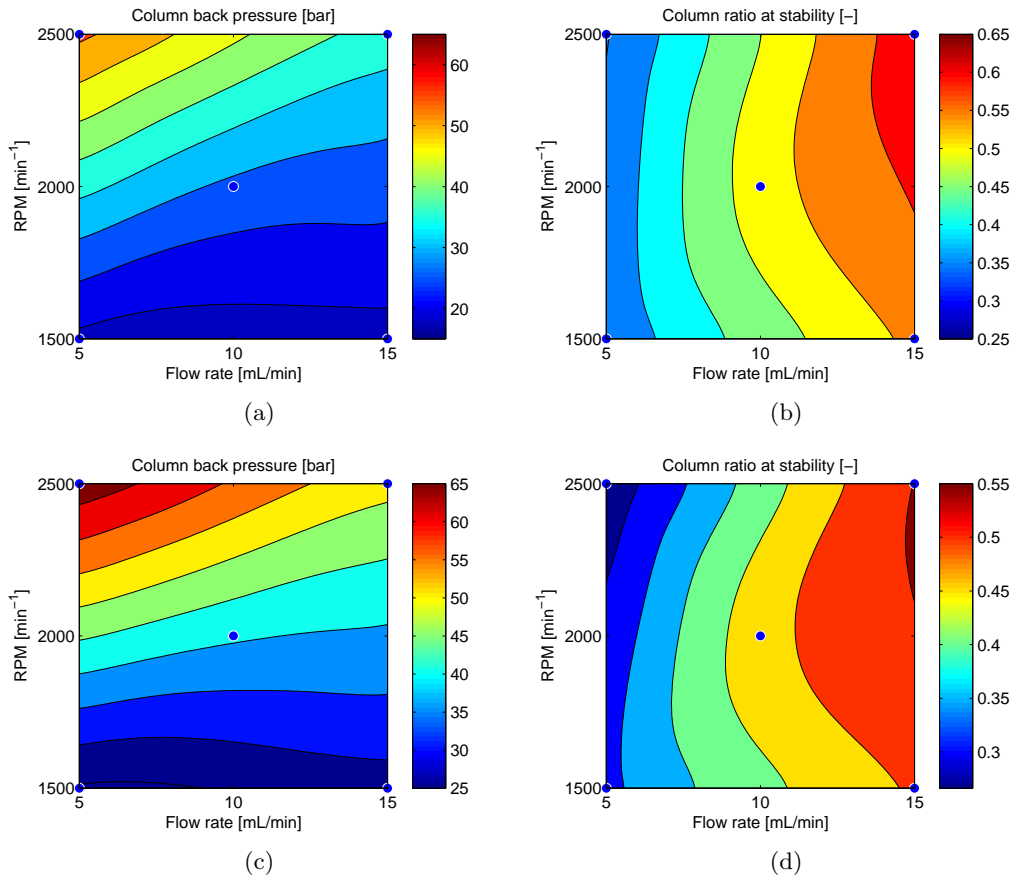


Figure 7: Results of column stability studies using the PEG400-SP2 and PEG400-SP3 system. a&c) column backpressure as a function of rotation speed and flow rate for SP2 & SP3. b&d) Ratio of stationary phase after column stabilization as a function of rotation speed and flow rate SP2 & SP3.

of the mAb/HCP mixture approximately triple the HCP clearance effect of the ATPS in comparison to a batch system without sacrificing yield. If a loss of mAb of 10% is acceptable, HCP clearance of 75% can be reached. From figure 8(a) it can also be seen,

that the operating conditions clearly influence the results. High rotation speeds and low flowrates are beneficial to the purification.

The operating conditions that gave the best purification (5 mL/min, 2500RPM) were subsequently chosen to investigate the influence of load volume and load mass on the purification. Figure 8(b) shows the results of this investigation. It was found that an increase of the load volume from 10 mL to 20 mL did not influence the purification. Further increase to 50 mL and 100 mL however decreased the HCP clearance significantly. The dilution of the target protein was also evaluated as a function of load volume. The results are shown in table III, calculated for a yield of 98% mAb. Dilution of the target molecule decreased significantly with increasing load volume. Loading 10 mL resulted in a dilution of 2.9x, while a load volume of 100 mL resulted in a dilution of only 1.25x. To evaluate the influence of load mass on the purification, samples of 10 mL volume with 1 mg/mL and 5 mg/mL mAb were compared. No difference in the purification was observed (data not shown).

Table III Dilution of mAb as a function of load volume. Sample was loaded onto the CPC running the PEG400 SP2 system at 5 mL/min and 2500 RPM. Resulting dilutions calculated for mAb yields of > 98%.

Load [mL]	10	20	50	100
mAb dilution [-]	2.9	2.3	1.5	1.25

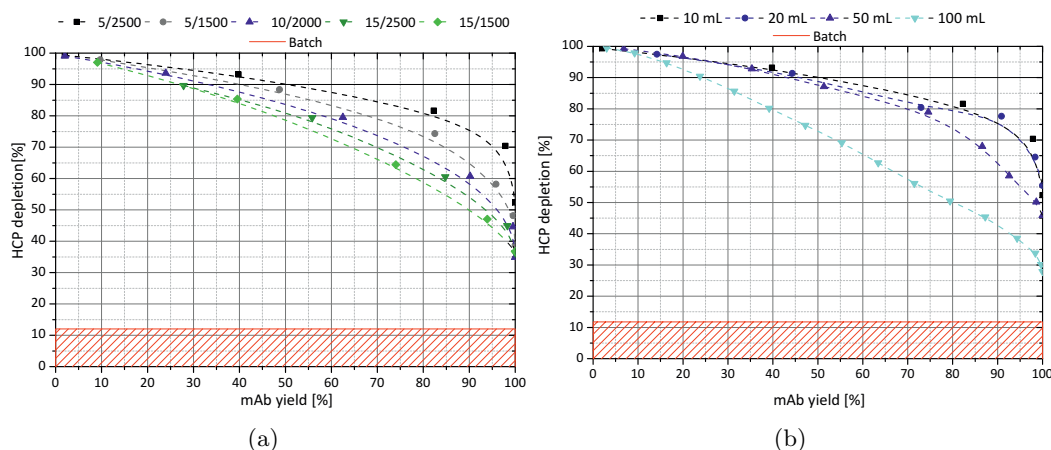


Figure 8: Influence of operating conditions (a) and load volume (b) on the purification. HCP clearance is plotted against mAb yield. The PEG400 SP2 was used in all cases. The dashed red area represents the HCP clearance achieved with a single batch separation.

4 Discussion

4.1 Precipitation screenings

Results of the precipitation screenings revealed that all four mAbs used in this study showed a distinct precipitation behavior when subjected to high levels of the phase forming components. This is remarkable as mAbs are a comparably large protein and share a very similar scaffold. Difference in the proteins are mostly restricted to the variable regions and especially the complementarity determining regions. The differences in these regions are sufficient to change the proteins' susceptibility to precipitation by both PEG and phosphate. All mAbs used in this study were derived from CHO cell culture and belonged to the same subclass. Glycosylation can thus be ruled out as factor determining the difference

in the precipitation experiments. Increasing PEG molecular weight was found to encourage precipitation both of the mAbs and the HCPs. This is in agreement with previous findings by Atha *et al.* [14] who found that the precipitation strength increases threefold for human albumin when PEG molecular weight is increased from 400 Da to 20 kDa.

Initial protein concentration can have an influence on precipitation results depending on the proteins used [15, 16]. For one mAb Ahamed *et al.* [17] showed no such influence. As the intended use of the precipitation data was not to yield an exact correlation of the distribution coefficient, but rather to qualitatively predict protein behavior in ATPS, precipitation screenings using one set initial protein concentration were conducted.

It was found that the addition of 5% [w/w] of phosphate to the samples increased mAb solubility in PEG-rich solutions. A cooperative effect might have been expected when two precipitants are combined. In the cases described here however, one precipitant counteracted the effect of the other. Protein precipitation by polymers and salts are most likely based on different mechanisms [15, 18]. Melander and Horvath [19] described the precipitation of a protein by a salt as the results of two opposing effects, an electrostatic effect increasing solubility (“salting in”) and a hydrophobic effect decreasing solubility (“salting out”). For precipitations using PEG Bhat *et al.* [18] reported that the precipitating action of polymer is mostly based on its steric exclusion from the protein domain. In contrast to the competitive effect of a salt increasing and decreasing a proteins solubility, Middaugh *et al.* [20] found no solubility increasing effect when PEGs are used as precipitant. Thus, the increased solubility of the mAbs with the addition of 5% phosphate buffer to PEG solutions, can be ascribed to a salting in effect. The mechanism behind the increased solubility against phosphate when 5% PEG is added remains unclear.

The precipitation experiments also revealed potential conditions to precipitate the target protein while keeping impurities in solution. Precipitation as a purification technique has found extensive use [21] and has also been demonstrated for the purification of mAbs [22, 23]. If the protein of interest can be redissolved in an active state, precipitating the protein of interest is attractive as it drastically reduces process volume. However, from the experiments detailed above, it remains unclear, whether the mAbs re-dissolution can be achieved after precipitation. Additionally, precipitation of the mAb might lead to co-precipitation of the impurities thus reducing the purification effect. While the precipitation screenings were conducted to predict protein behavior in ATPS, it is clear that the information gained from these experiments might also be valuable in such circumstances.

Overall, the precipitation screenings successfully generated valuable data using little amount of protein and time and allowed for the selection of ATPS systems to be screened.

4.2 ATPS distribution and recovery

The main goal of this work was to establish an aqueous two-phase system selection scheme usable for the preparative application of ATPS. Similar to non-linear chromatography, preparative application of ATPS under conditions of maximum protein load lead to the saturation of at least one of the phases. The apparent distribution coefficient measured for proteins in ATPSs under these conditions thus becomes a function of total system protein concentration and the protein’s solubility in the two phases. The results in section 3.2.2 show a good qualitative correlation between the results of protein precipitation experiments and the protein’s behavior in the ATPS. While single data points diverge, such as recovery of mAb1 in the PEG400-PO4 system with the longest tieline, the general trends both in distribution and in recovery of the mAbs could be predicted using data from precipitation experiments. It is thus plausible to select both system type and a range of tieline length to further investigate from the precipitation data in concert with binodal curves and tielines.

Several concepts to aid the implementation of an aqueous two-phase extraction step have been proposed either based on experience or based on the characterization of the protein. An experience based development concept was suggested by Rito-Palomares *et al.* [24] and Benavides *et al.* [25]. Basic target protein characteristics such as isoelectric point and molecular weight guide the initial selection of phase forming components and pH of the ATPS. This initial selection is based on general rules of thumb concerning the distribution of proteins and is thus largely based on the experimenters experience. If the initial selection proves successful, the system is subsequently optimized concerning tieline length and protein load. A different approach to facilitating ATPS process step development is to measure a characteristic of the target protein that correlates with its distribution. Several such correlations have been published both for polymer-polymer [2, 3, 26] systems and polymer-salt systems [5]. Asenjo *et al.* [5] demonstrated a linear correlation between a protein hydrophobicity descriptor based on ammonium sulfate precipitations [27]. While such correlations give valuable insights into the force influencing protein distribution in ATPSs, they are unlikely to be applicable for the implementation of a preparative separation step. The major obstacle for their practical use is the influence of total protein load on the distribution coefficient in overloaded systems as detailed above. For preparative application of an ATPS, the proteins solubility in the phases needs to be taken into account. The approach described in this publication uses target protein specific data generated from precipitation screenings with the phase forming components in combination with phase systems specific data, binodal and tielines, to guide the selection of a system type that promises both high capacity and a favorable distribution. It thus neither relies on extensive experimental experience on the developers side nor does it neglect the influence of protein solubility on the distribution and recovery. Experimental effort to use this scheme is generally low. Binodal and tieline data can either be taken from literature or generated, for example using HTS platforms [8]. Precipitation screenings, which can also be performed on a HTS platform, using the phase forming components as precipitants then give the basis for a selection of systems points to be further investigated. In some cases, as for PEG400 in this work, the amount of phosphate in the upper and PEG400 in the lower phase might need to be taken into account during the precipitation experiments. A salting-in effect of 5% phosphate during PEG precipitations was observed. A similar effect was also observed when 5% PEG400 was added during PO₄ precipitations. The amount of other phase forming component in the PEG and PO₄ rich phases depends on the position of the binodal curve and the tieline length chosen. As binodal curves move towards lower concentrations with increasing PEG molecular weight, this effect is reduced in a way that is unnecessary to take into account for PEG molecular weights of 600 Da and above.

Capacity of ATPS for mAbs was previously shown to be critically low in PEG4000-PO₄ systems. Oelmeier *et al.* [8] and Rosa *et al.* [7] showed that mAb concentrations as low as 1 g/L lead to recovery loss due to precipitation in PEG4000-PO₄ systems. Attempts to shift mAb distribution towards the upper phase by the addition of NaCl largely failed due to precipitation. With precipitation experiments as those shown above, this capacity limitation could have been predicted as PEG4000 concentrations in the upper phase are too high for the mAbs to be soluble at high concentrations in the upper phase.

The presented process step development approach thus constitutes a systematic way of selecting promising conditions for the preparative application of ATPS.

4.3 CPC column characterization

Centrifugal partitioning chromatography as a means to scale-up an extraction while concurrently improving resolutions [28, 29] was evaluated for the presented purification task. Column stability using both PEG400 systems was generally high. Stripping of stationary

phase was below 1% after reaching a stable state. Thus, column stability was significantly higher than that of PEG1000-PO4 systems reported by Sutherland *et al.* [9] and PEG4000-PO4 systems tested by the authors. Similar trends were seen for stationary phase ratio. PEG400-PO4 systems showed a significantly increased ratio of stationary phase under equal conditions when compared to PEG1000 and PEG4000-PO4 systems. A general trend towards both higher column stability and higher ratios of stationary phase with decreasing PEG molecular weight can be deduced. Tieline length was also found to influence column composition and backpressure. A longer tieline lead to a higher ratio of stationary phase and higher backpressures. Armstrong *et al.* [30] investigated the influence of mobile phase flow rate, density difference between the two phases, the viscosity, and centrifugal force on the backpressure generated over the column using organic/aqueous two-phase systems. He found that the backpressure increased with increasing density difference and viscosity. Our findings for ATPS are in line with these observations. Under the same operating conditions, a system of higher density difference and viscosity - a system with a longer tieline - generated a higher backpressure and retained more stationary phase.

4.4 mAb elution on CPC column

mAb/HCP separation was evaluated for the PEG400 SP2 system under various operation conditions and different load volumes. HCP clearance was best at a flow-rate of 5 mL/min and 2500 RPM. These conditions gave the highest stationary phase retention. It had previously been shown, that high retention of stationary phase increases resolution in CCC type liquid-liquid chromatography [31]. Our results for CPC are thus in line with previously published results. Load volume was found to decrease sample dilution in the expense of HCP clearance. Again, these finding are in line with previously published results [9]. In contrast to load volume, an increase of mAb concentration in the load did not decrease mAb/HCP resolution. Thus, to maximize system load, load volume should only be increased to a certain extent, while mAb concentration in the load should be maximized. This underlines the importance of conducting screening experiments under preparative conditions and considering the solubility of the mAb in the presence of the phase forming conditions such as conducted in this work.

As the target protein eluted within one mobile phase volume of the column, it is reasonable to suggest that the observed dilution of the target protein is the minimal dilution achievable with the used load volume. As also demonstrated by Sutherland *et al.* [9] we found that higher load volumes can reduce dilution. Loss of resolution with increasing load volume was however observed. An increase in process volume by a factor of 4 or above cannot be tolerated, especially at an early phase of a process. Thus a compromise between clearance and dilution must be selected when CPC is considered as a unit operation. According to CPC theory [32], to improve resolution, a higher ratio of stationary phase would be beneficial. A higher ratio of stationary phase could be achieved by further decreasing flow rate. This might however lead to industrially in-applicably long processing times and would increase dilution of the target molecule. Additionally, column backpressure will most likely increase with a further decrease of flow-rate thus approaching the pressure limit of the equipment.

In order to produce a sample to be injected onto the CPC column, a batch ATPS was used. The upper phase of this batch system was used as sample. The benefit of generating the sample in this way is that the target protein will be in mobile phase at the time of injection. Sutherland *et al.* [9] observed detrimental effects on the column stability if an entire ATPS is injected directly onto the column. Load volume above 10% column volume lead to a collapse of the column. By injecting the sample contained in mobile phase, we showed that load volumes of 100 mL (20% column volume) had no effect on the column

compositions. Higher load volumes are most likely possible, due to the decreasing HCP clearance however, of little applicability. Additionally, by using a single batch step to generate the load, target protein concentration can be increased before injection onto the CPC column and higher load masses might be achievable. A strong distribution of the target protein into the mobile phase is needed in order to be able to use this approach.

5 Conclusion and outlook

In this publication, a selection theme for preparative application of aqueous two-phase extraction for the purification of proteins was demonstrated. PEG-PO4 systems were used. As industrially relevant purification task the separation of four monoclonal antibodies from host cell proteins was chosen. The selection of the phases is based on binodal and tieline data, as well as protein precipitation experiments performed on a high throughput screening platform. The phase selection took into account the high protein loads desirable for preparative application of an ATPS and its scale-up to a centrifugal partitioning mode. It was shown that precipitation experiments performed with the phase forming components gave a good qualitative correlation to protein behavior in ATPSs. Upscale from 650 μ L batch systems to a 500 mL CPC was successfully demonstrated. Influence of process parameters on HCP clearance and target protein dilution were investigated. Conditions leading to high retention of stationary phase yielded the highest HCP clearance. Dilution of the target protein could be reduced, at the expense of HCP clearance, by increasing the load volume. Using the CPC, HCP clearance was improved more than threefold compared to batch separations with recoveries of the target protein above 98%. While the target protein dilution needs to be addressed, it was shown that CPC can be a valuable addition to the protein purification toolbox.

6 Acknowledgments

The authors gratefully acknowledge supply of CPC equipment by **Armen Instrument** and material supply and financial support by **Boehringer Ingelheim Pharma GmbH & Co. KG**

References

- [1] A. M. Azevedo, P. A. J. Rosa, I. F. Ferreira, M. R. Aires-Barros, Chromatography-free recovery of biopharmaceuticals through aqueous two-phase processing., *Trends in biotechnology* 27 (4) (2009) 240–7.
- [2] S. Sasakawa, H. Walter, Partition behavior of native proteins in aqueous dextran-poly(ethylene glycol)-phase systems., *Biochemistry* 11 (15) (1972) 2760–5.
- [3] G. Johansson, A. Sarnesto, E. Høge-Jensen, I. Szabo-Lin, C. Guthenberg, B. Mannervik, Effects of salts on the partition of proteins in aqueous polymeric biphasic systems., *Acta Chemica Scandinavica* 28b (1974) 873–882.
- [4] K. Berggren, A. Wolf, J. A. Asenjo, B. A. Andrews, F. Tjerneld, The surface exposed amino acid residues of monomeric proteins determine the partitioning in aqueous two-phase systems., *Biochimica et biophysica acta* 1596 (2) (2002) 253–68.

- [5] J. A. Asenjo, A. S. Schmidt, F. Hachem, B. A. Andrews, Model for predicting the partition behaviour of proteins in aqueous two-phase systems., *Journal of Chromatography. A* 668 (1) (1994) 47–54.
- [6] B. A. Andrews, A. S. Schmidt, J. A. Asenjo, Correlation for the partition behavior of proteins in aqueous two-phase systems: effect of surface hydrophobicity and charge., *Biotechnology and bioengineering* 90 (3) (2005) 380–90.
- [7] P. A. J. Rosa, Application of central composite design to the optimisation of aqueous two-phase extraction of human antibodies., *Journal of Chromatography. A* 1141 (2007) 50–60.
- [8] S. A. Oelmeier, F. Dismer, J. Hubbuch, Application of an aqueous two-phase systems high-throughput screening method to evaluate mab hcp separation., *Biotechnology and bioengineering* 108 (1) (2011) 69–81.
- [9] I. A. Sutherland, G. Audo, E. Bourton, F. Couillard, D. Fisher, I. Garrard, P. Hewitson, O. Intes, Rapid linear scale-up of a protein separation by centrifugal partition chromatography., *Journal of Chromatography. A* 1190 (1-2) (2008) 57–62.
- [10] A. Berthod, M. Hassoun, M. J. Ruiz-Angel, Alkane effect in the arizona liquid systems used in countercurrent chromatography., *Analytical and bioanalytical chemistry* 383 (2) (2005) 327–40.
- [11] J. C. Merchuk, B. A. Andrews, J. A. Asenjo, Aqueous two-phase systems for protein separation. studies on phase inversion., *Journal of chromatography. B* 711 (1998) 285–93.
- [12] E. J. Cohn, The physical chemistry of the proteins, *Physiological Reviews* 5 (3) (1925) 349–437.
- [13] A. Foucault, *Centrifugal partition chromatography*, Chromatographic science, M. Dekker, 1995.
- [14] D. H. Atha, K. C. Ingham, Mechanism of precipitation of proteins by polyethylene glycols. analysis in terms of excluded volume., *The Journal of biological chemistry* 256 (23) (1981) 12108–17.
- [15] Y. C. Shih, J. M. Prausnitz, H. W. Blanch, Some characteristics of protein precipitation by salts., *Biotechnology and bioengineering* 40 (10) (1992) 1155–64.
- [16] H. V. Iyer, T. M. Przybycien, Protein precipitation: Effects of mixing on protein solubility., *AIChE Journal* 40 (2) (1994) 349–360.
- [17] T. Ahamed, B. N. A. Esteban, M. Ottens, G. W. K. v. Dedem, L. A. M. v. d. Wielen, M. A. T. Bisschops, A. Lee, C. Pham, J. Thommes, Phase behavior of an intact monoclonal antibody., *Biophysical journal* 93 (2) (2007) 610–9.
- [18] R. Bhat, S. N. Timasheff, Steric exclusion is the principal source of the preferential hydration of proteins in the presence of polyethylene glycols., *Protein science : a publication of the Protein Society* 1 (9) (1992) 1133–43.
- [19] W. R. Melander, C. Horvath, Salt effects on hydrophobic interactions in precipitation and chromatography of proteins: An interpretation of the lyotropic series., *Archives of biochemistry and biophysics* 183 (1) (1977) 200215.

- [20] R. C. Middaugh, W. A. Tisel, R. N. Haire, A. Rosenberg, Determination of the apparent thermodynamic activities of saturated protein solutions., *Journal of Biological Chemistry* 254 (2) (1979) 367.
- [21] A. Kumar, I. Galaev, B. Mattiasson, *Precipitation of proteins: Nonspecific and specific.*, CRC Press, 2003, Ch. 7.
- [22] F. Perosa, R. Carbone, S. Ferrone, F. Dammacco, Purification of human immunoglobulins by sequential precipitation with caprylic acid and ammonium sulphate., *Journal of immunological methods* 128 (1) (1990) 9–16.
- [23] J. Persson, P. Lester, Purification of antibody and antibody-fragment from e. coli homogenate using 6,9-diamino-2-ethoxyacridine lactate as precipitation agent., *Biotechnology and bioengineering* 87 (3) (2004) 424–34.
- [24] M. Rito-Palomares, Practical application of aqueous two-phase partition to process development for the recovery of biological products, *Journal of Chromatography. B* 807 (2004) 3–11.
- [25] J. Benavides, M. Rito-Palomares, Practical experiences from the development of aqueous two-phase processes for the recovery of high value biological products, *Journal of Chemical Technology & Biotechnology* 83 (2) (2008) 133–142.
- [26] F. Hachem, Hydrophobic partitioning of proteins in aqueous two-phase systems, *Enzyme and Microbial Technology* 19 (7) (1996) 507–517.
- [27] T. M. Przybycien, J. E. Bailey, Aggregation kinetics in salt-induced protein precipitation., *AIChE Journal* 35 (11) (1989) 1779–1790.
- [28] I. A. Sutherland, Recent progress on the industrial scale-up of counter-current chromatography., *Journal of Chromatography. A* 1151 (1-2) (2007) 6–13.
- [29] S. Bérot, E. Le Goff, A. Foucault, L. Quillien, Centrifugal partition chromatography as a tool for preparative purification of pea albumin with enhanced yields., *Journal of Chromatography. B* 845 (2) (2007) 205–9.
- [30] D. W. Armstrong, Theory and Use of Centrifugal Partition Chromatography, *Journal of Liquid Chromatography* 11 (12) (1988) 2433–2446.
- [31] I. a. Sutherland, J. de Folter, P. Wood, Modelling CCC Using an Eluting Countercurrent Distribution Model, *Journal of Liquid Chromatography & Related Technologies* 26 (9&10) (2003) 1449–1474.
- [32] J. d. Folter, I. A. Sutherland, Universal counter-current chromatography modelling based on counter-current distribution., *Journal of Chromatography. A* 1216 (19) (2009) 4218–24.

MANUSCRIPT IV

A SUB-TWO MINUTES METHOD FOR MAB-AGGREGATE QUANTIFICATION USING PARALLEL INTERLACED SIZE EXCLUSION HPLC

Patrick Diederich⁺, Sigrid K. Hansen⁺, Stefan A. Oelmeier⁺, Bianca Stolzenberger, Jürgen Hubbuch*

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

⁺ These authors contributed equally to this publication and appear in alphabetical order

* Corresponding author. Tel.: +49 721 608-42557; fax: +49 721 608-46240. E-mail address: juergen.hubbuch@kit.edu

5.6 A Sub-Two Minutes Method for mAb-Aggregate Quantification using Parallel Interlaced Size Exclusion HPLC

Abstract

In process development and during commercial production of monoclonal antibodies (mAb) the monitoring of aggregate levels is obligatory. The standard assay for mAb aggregate quantification is based on size exclusion chromatography (SEC) performed on a HPLC system. Advantages hereof are high precision and simplicity, however, standard SEC methodology is very time consuming. With an average throughput of usually two samples per hour, it neither fits to high throughput process development (HTPD), nor is it applicable for purification process monitoring. We present a comparison of three different SEC columns for mAb-aggregate quantification addressing throughput, resolution, and reproducibility. A short column (150 mm) with sub-two micron particles was shown to generate high resolution (~ 1.5) and precision (coefficient of variation (cv) < 1) with an assay time below six minutes. This column type was then used to combine interlaced sample injections with parallelization of two columns aiming for an absolute minimal assay time. By doing so, both lag times before and after the peaks of interest were successfully eliminated resulting in an assay time below two minutes. It was demonstrated that determined aggregate levels and precision of the throughput optimized SEC assay were equal to those of a single injection based assay. Hence, the presented methodology of parallel interlaced SEC (PI-SEC) represents a valuable tool addressing HTPD and process monitoring.

Keywords: *Monoclonal antibody, aggregates, Size Exclusion Chromatography, high throughput analytics, interlaced injection, PI-SEC*

1 Introduction

Aggregate levels in monoclonal antibody drugs are a critical quality attribute due to their potential immunogenicity [1, 2]. Aggregates of monoclonal antibodies are often the most abundant product related impurity. The purification process needs to ensure that aggregate levels are reduced to an acceptable level in the final drug product. While the first two steps in a standard mAb downstream process are readily capable of depleting three highly abundant process related impurities, host cell protein, DNA, and water, the reduction of aggregate levels to acceptable levels is often challenging. Thus, monitoring aggregate levels is critical in process development.

One way to reduce process development costs is to increase development throughput. Various process steps have been scaled down to fit into a high throughput process development (HTPD) scheme [3–6]. Additionally, platform processes have been implemented for monoclonal antibody based products, further reducing the efforts needed from process development down to process verification [7]. These improvements have created an analytical bottleneck in process development. To match throughput of the experimentation, reasonably short analysis times need to be achieved.

Size exclusion chromatography (SEC) is the standard method for mAb-aggregate analysis. The standard SEC assay with a throughput of two samples per hour [8, 9] does however not suit a HTPD approach. Several measures are thus in the spotlight to increase throughput in HPLC without changing the analytical technique as such: parallelization and interlacing sample injection. While parallelization using multiple HPLC stations is currently the most often used approach, it is for obvious reasons also the most expensive. Parallelization of multiple columns on a single detector via column switching valves is a way to reduce parallelization cost and has been successfully demonstrated [10]. Most often in this approach, the elution and the regeneration of a chromatographic analysis are separated such that one column regenerates while the other column performs an analysis [11]. In contrast to gradient elution, column regeneration is however not necessary in SEC. Another approach to improve throughput is to run a single column in an interlaced mode. In interlaced chromatography a sample is injected onto the column before the preceding analysis has been completed. This approach requires isocratic conditions. Farnan *et. al* [12] successfully demonstrated its use for aggregate analysis of mAbs and were able to reduce assay time per sample by more than a factor of two from 30 minutes to 14 minutes.

Finally, HPLC equipment capable of higher back pressures has been implemented (most often termed UHPLC) [13]. Shorter columns with smaller column volume and smaller particle sizes can be used with this equipment, thus reducing assay time without sacrificing resolution. While one of the most often used columns for mAb-aggregate analysis has a pressure limit of 7.2 MPa (Tosoh TSKgel[®] 3000 SWxl), two

5.6 A Sub-Two Minutes Method for mAb-Aggregate Quantification using Parallel Interlaced Size Exclusion HPLC

new SEC columns suitable for higher back pressures of 24.1 MPa (Zenix™ SEC-250 (Sepax Technologies)) to 41.4 MPa (ACQUITY UPLC® BEH200 SEC (Waters Corporation)) recently became commercially available.

In this paper, we compare mAb-aggregate analysis performed on these three SEC columns. The columns are compared in terms of assay throughput, resolution, and precision. We demonstrate the application of ACQUITY UPLC® BEH200 SEC columns (Waters Corporation) in an interlaced mode as well as by interlaced injections on two columns run in parallel. We demonstrate how throughput can be increased by a factor of 10-15 compared to a standard analysis using a TSKgel® 3000 SWxl column. Advantages and disadvantages of the methodology are discussed.

1.1 Theory - Increasing Throughput by Interlacing and Parallelization

While the presented methodology can be applied universally to any type of SEC-column, differences arise in the use of (U)HPLC equipment and the actual pressure rating of the respective SEC-columns and adsorbents. To implement the method developed in this study to its full potential, a prerequisite lies in the use of an (U)HPLC system which is equipped with two independent flow switching valves. An *inlet valve* directs the flow to the columns and autosampler and an *outlet valve* directs the flow from the column outlets to the detector and waste. For maximum throughput two SEC columns can thus be run in parallel applying interlaced injections on each of the two identical columns. The idea of parallel interlaced (PI-) SEC methodology is to eliminate every region of a chromatogram which is not providing any relevant data (e.g. antibody aggregate and monomer). In a first step, data of a single chromatographic SEC analysis therefore serve as a benchmark for the estimation of analysis time and method development as described in the following:

Single Injection

In figure 1 A and 2 A typical chromatograms of common mAb SEC analysis are displayed. The chromatograms can be divided into four main phases. The first phase after sample injection is the initial lag phase (t_{lag}). The time span in which aggregate species and monomer elute is referred to as information phase (t_{inf}). In this work, protein fragments are not considered as species of interest and are not included in t_{inf} . The third phase between monomer peak and the eluting salt fraction is referred to as hold phase (t_{hold}). It is assumed that no protein elute later than the salt fraction of the injected sample. The elution region of salt species is referred to as t_{salt} .

An single chromatogram of the sample material provides the user with the retention times of every elution phase for the column used at the specific flow rate. The total time required for the analysis of n samples can be stated as:

$$t_{total} = n \cdot (t_{lag} + t_{inf} + t_{hold} + t_{salt}) \quad (1)$$

Given these retention times, the first step to increase analysis throughput is to eliminate t_{lag} from the resulting chromatograms as explained below.

Interlaced Injection

Farnan *et al.* [12] has described the methodology of interlaced SEC in detail. In a brief, the methodology is based on injecting a subsequent sample before the ongoing analysis of a sample has completed. The subsequent information phase begins immediately after the salt fraction of the preceding sample has eluted. Figure 3 A and B show the transition from a mode of single injection to interlaced injection. By the use of a second timebase (see section 2.2), a separate control program for data acquisition ("program DAD") facilitates distinct chromatograms for each injection and corresponding sample. In figure 3 B it is demonstrated that the lag phase can thus be eliminated from analysis. The total time required for the analysis of n samples can be stated as:

$$t_{total} = t_{lag} + n \cdot (t_{inf} + t_{hold} + t_{salt}) \quad (2)$$

5.6 A Sub-Two Minutes Method for mAb-Aggregate Quantification using Parallel Interlaced Size Exclusion HPLC

Parallel Interlaced Injection

A further increase in throughput can be achieved when applying interlaced injections on two columns which are operated in parallel. Starting from interlaced chromatography, in parallel interlaced SEC the assay time is further reduced by t_{hold} , as is demonstrated in figure 3 B and C. Two switching valves are used to direct the flow alternately between autosampler, two columns and the detector, thus enabling the elimination of t_{lag} , t_{hold} and t_{salt} . In figure 3 D a scheme of the valve switching is displayed. The use of two columns and switching valves require two distinct programs assigned to *timebase 1*, on which pumps, autosampler and column compartment including the switching valves are controlled. The programs contain the same commands, but differ in the direction of both valves switching. As for interlaced chromatography, data acquisition is performed separately by using a second timebase (*timebase 2*) for the detector, now only recording phase t_{inf} of each injected sample.

For programming PI-SEC, three possible cases need to be considered, since elution profiles of a single injection analysis differ in t_{lag} , t_{hold} and t_{salt} depending on column type and sample material. For reason of simplicity, it is assumed that $t_{hold} > t_{salt}$, which is the common case in SEC analysis of antibody samples.

Case-1. $t_{inf} > t_{hold}$:

The first sample is injected on column 1 at:

$$t_1 = 0 \quad (3)$$

The second sample is injected on column 2 at:

$$t_2 = t_1 + t_{inf} \quad (4)$$

The subsequent samples are alternately injected on column 1 and column 2 at times:

$$t_{n,inj} = t_{n,inj-1} + t_{inf} \quad (5)$$

The total assay time for the analysis of n samples can hence be calculated by equation 6. This equation gives the theoretically possible increase in throughput which can be gained via PI-SEC using one single detector.

$$t_{total} = t_{lag} + n \cdot (t_{inf}) + t_{hold} + t_{salt} \quad (6)$$

The *outlet valve* is switched as soon as the information phase of a sample from one column has passed the detector. At that time, the salt peak has completely eluted from the other column. Samples are alternately injected on the two columns and analyzed without any interference of eluting salt fractions. As an example, figure 1 shows a schematic drawing of PI-SEC methodology for the case of $t_{hold} < t_{inf}$.

Case-2. $k \cdot t_{inf} < t_{hold}$:

If $k \geq 1$, one or more informational phases fit into t_{hold} and k additional injections (rounding down of k to whole numbers) on one column become feasible before switching to the second column. The injection times and the time needed for the analysis of n samples can be estimated using the same equations 3 - 6 as given in case one. Figure 2 shows a schematic drawing of the PI-SEC methodology applied for a case 2 elution profile where $1 < k < 2$. Now, two salt peaks elute from one column within the time two information phases elute from the other column.

Although time benefit is the same as in case one, it should be noted that in this mode proteins of multiple, subsequently injected samples pass the salt fraction of the preceding injected samples,

5.6 A Sub-Two Minutes Method for mAb-Aggregate Quantification using Parallel Interlaced Size Exclusion HPLC

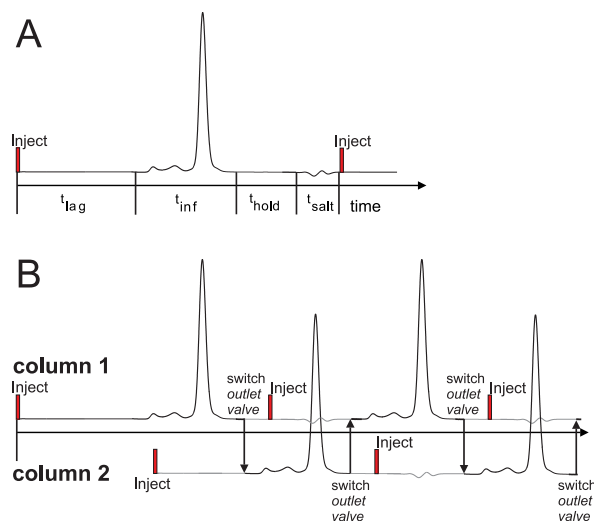


Figure 1: Schematic of PI-SEC methodology applicable in the case of $t_{inf} > t_{hold}$. A: chromatogram of a mAb sample analyzed in single injection mode. Using the elution phases t_{lag} , t_{inf} , t_{hold} , and t_{salt} , a PI-SEC program can be set up (B). In this case, samples are injected alternately on two columns, while the *outlet valve* directs the flow from the column outlet to the detector.

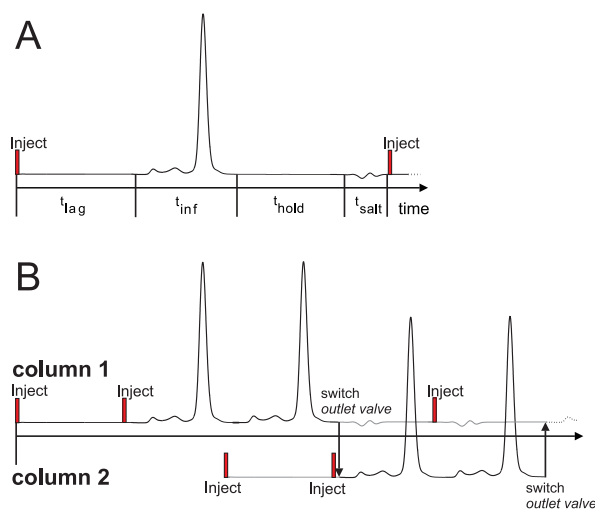


Figure 2: Schematic of PI-SEC methodology applicable in the case of $t_{inf} < t_{hold}$. A: chromatogram of a mAb sample analyzed in single injection mode. Using the elution phases t_{lag} , t_{inf} , t_{hold} , and t_{salt} , a PI-SEC program can be set up (B). In this case, two samples are subsequently injected per column before switching to the second column.

5.6 A Sub-Two Minutes Method for mAb-Aggregate Quantification using Parallel Interlaced Size Exclusion HPLC

whereas for case one the salt fraction of each sample always elute earlier from the column than does the information phase. Multiple injections on one column is further only applicable, if no species of lower molecular weight than the monomer species is present in the sample material. Otherwise the species of lower molecular weight will elute within the information phase of the subsequent sample injected on the same column.

In the case that $k < 1$ and the *outlet valve* is switched instantly after the information phase of a sample from one column has passed the detector, the salt fraction of the preceding sample has not eluted yet from the second column. Therefore, some additional time (t_{add}) must be added before switching the *outlet valve*. The sum of $t_{add} + t_{inf}$ needs to be greater than $t_{hold} + t_{salt}$. The time needed for the analysis of n samples can be estimated using equation 6, while including t_{add} (9). This delay needs also to be factored in the injection times of the interlaced mode of each column. When the first injection at t_1 is performed, the second injection takes place at:

$$t_2 = t_1 + t_{inf} + t_{add} \quad (7)$$

The injection time of sample n can be hence given by:

$$t_{n,inj} = t_{n,inj-1} + t_{inf} + t_{add} \quad (8)$$

The total assay time for n samples can be calculated using:

$$t_{total} = t_{lag} + n \cdot (t_{inf} + t_{add}) + t_{hold} \quad (9)$$

From a practical aspect it should be mentioned that, if t_{inf} is slightly smaller or exactly equals the sum of $t_{hold} + t_{salt}$, the *outlet valve* is switched just when salt is detected or just arrives at the detector. The baseline determination and an autozero processing of the absorbance signal is hence affected and might lead to imprecise peak integration.

Regarding all described scenarios case one marks the optimal condition for PI-SEC since information phases of samples injected alternately on two columns neither interfere with eluting salt fractions nor are additional times required. With an increasing ratio of t_{inf}/t_{hold} , the benefit of using two columns in parallel over interlaced injection decreases. For the purpose of method robustness, in any of the above described cases additional time for switching the inlet and outlet valves should be implemented: Switching the *inlet valve* should occur a few seconds before the injection takes place and switching of the *outlet valve* should occur a few seconds before the high molecular weight species elute. Thus, baseline determination and peak integration become more precise. To set up the control program, sampling and washing times need to be taken into account. The duration of sampling and washing depends strongly on the used (U)HPLC equipment and might significantly slow down the assay if it exceeds the duration of the information phase. Furthermore, differences in column packing and hence retention times need to be considered.

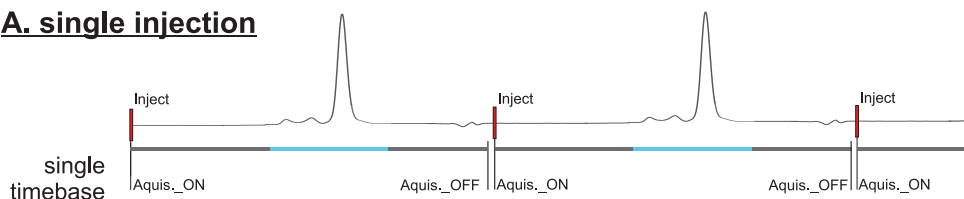
2 Materials & Methods

2.1 SEC columns

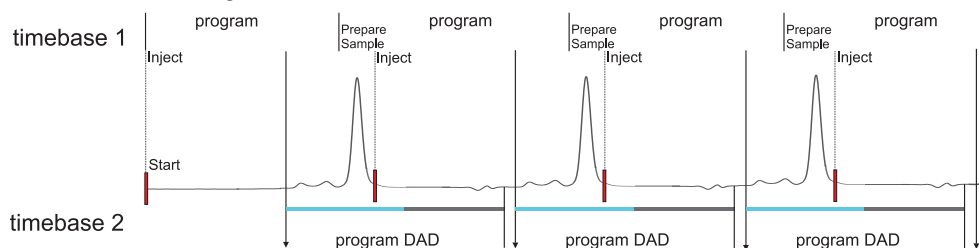
SEC columns from three vendors were used in this work: 1. TSKgel 3000 SWxl (Tosoh Corporation, Tokio, Japan) 2. ACQUITY UPLC[®]BEH200 SEC (Waters Corporation, Milford, MA, USA) 3. Zenix SEC-250 (Sepax Technologies, Newark, DE, USA). Columns were fitted with 0.2 μm inlet filter (Opti-Solv[®]EXP[™], Optimize Technologis, Oregon City, OR, USA). In table I the column properties are listed. The columns differ in macroscopic as well as microscopic dimensions.

5.6 A Sub-Two Minutes Method for mAb-Aggregate Quantification using Parallel Interlaced Size Exclusion HPLC

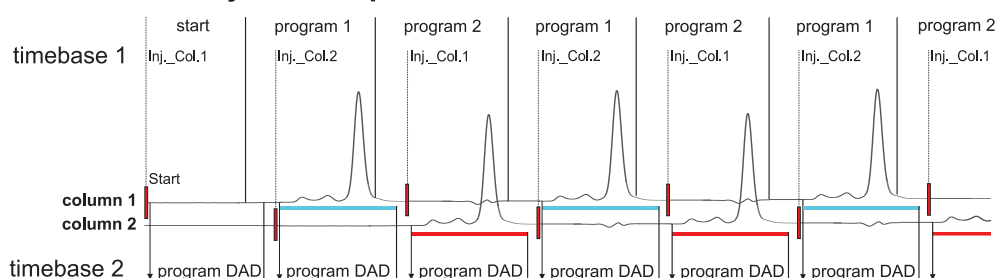
A. single injection



B. interlaced injection



C. interlaced injection in parallel



D. valve configuration

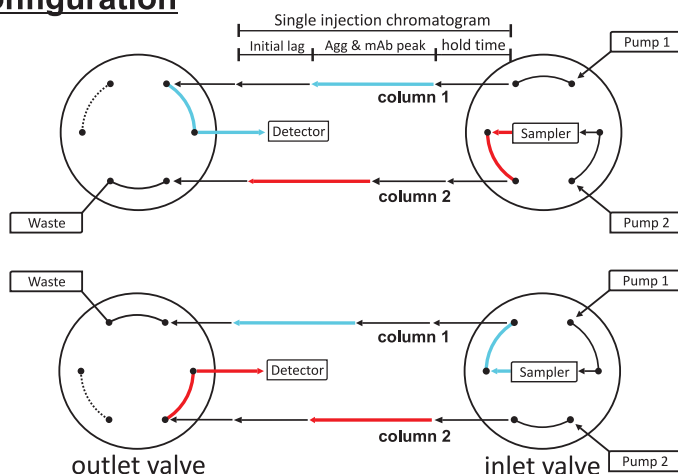


Figure 3: Three modes of operating SEC analysis are displayed. Based on single run chromatograms (**A**) throughput can be improved by interlacing sample injections (**B**) on one SEC column. By using a second timebase (*timebase 2*) for data acquisition, a dedicated chromatogram is generated for every sample injection. Each timebase is controlled by separate programs. Using a second column run in parallel and two timebases (**C**), throughput can be pushed to its theoretical maximum by performing interlaced injections on both columns. Hereby, two programs on *timebase 1* are implemented differing only in the switching direction of the switching valves. A schematic of the configuration of two six-port-valves (**D**) demonstrates the switching procedure which has to be implemented in the control programs 1 and 2.

5.6 A Sub-Two Minutes Method for mAb-Aggregate Quantification using Parallel Interlaced Size Exclusion HPLC

Table I Specifications of the HPLC SEC columns used in this study.

vendor description	column dimension	pore size	particle size	maximum pressure	column volume	void
ACQUITY UPLC BEH200 SEC	4.6x150 mm	200 Å	1.7 µm	41.5 mPa	2.5 mL	1.97 mL
Zenix SEC-250	4.6x250 mm	300 Å	3.0 µm	24.1 mPa	4.2 mL	3.45 mL
TSKgel 3000 SWxl	7.8x30 0mm	250 Å	5.0 µm	7.8 mPa	14.3 mL	12.23 mL

2.2 UHPLC setup

An UltiMate3000 RSLC x2 Dual system from Dionex (Sunnyvale, CA, USA) was used for UHPLC analysis. The system was composed of two HPG-3400RS pumps, a WPS-3000TFC-analytical autosampler and a DAD3000RS detector. The autosampler was equipped with a sample loop of 5 µl or 20 µl, respectively. The volume of the injection needle was 15 µl, the syringe size was 250 µl. In all experiments, full loop injections were performed. The system included a TCC-3000RS column thermostat to enclose two columns, which were connected to two six-port column switching valves. The *inlet valve* directs the flow between autosampler outlet and column inlets, hence controlling to which column a sample is injected. The *outlet valve* directs the flow between column outlets and UV-detector, hence controlling from which column outlet the UV signal is measured. All column experiments were conducted at 25 °C. For SEC analysis performed in interlaced and parallel-interlaced mode, the system was split in two virtual parts by using two separate timebases. *Timebase 1* controlled pumps, autosampler, valves and column compartment and *Timebase 2* controlled the UV detector. The two timebases were physically linked by connecting a relay assigned to *timebase 1* with an input assigned to *timebase 2*. Switching of the relay in *timebase 1* triggered an input signal in *timebase 2*. This input signal was then used to trigger the UV signal acquisition. By this setup, it was possible to record the information phase of each sample separately.

2.3 Software

Matlab2010a (The Mathworks Natick, ME, USA) was used for data analysis. Chromeleon®(6.80 SR10) was used to control the UHPLC equipment and to integrate the elution peaks in the chromatograms. The Chromeleon software was extended to include two timebases.

2.4 Buffer and Sample

SEC analysis were performed using a 0.2 M potassium phosphate buffer at pH 6.2 containing 0.25 M potassium chloride. Buffers were filtered through 0.2 µm filters (Sartorius, Germany) prior to use. When two pumps were used simultaneously (parallel-interlaced protocol), the same buffer preparation was apportioned in two bottles. A proteinA pool of a CHO expressed IgG was used as mAb sample. The concentration was set to a concentration of 1 g/L by dilution with dH2O.

2.5 Aggregate Level and Chromatographic Resolution

For each single injection run, the aggregate level and the resolution was determined. For all interlaced and parallel-interlaced runs only the aggregate level was determined. The aggregate level was defined as the percentage of the species in the mAb sample eluting prior to the monomer. The achieved chromatographic resolution of the mAb monomer and the smallest aggregate (dimer) was calculated based on the EP norm:

$$R = 1.18 \cdot \frac{t_{monomer} - t_{dimer}}{W_{50\%,monomer} + W_{50\%,dimer}} \quad (10)$$

5.6 A Sub-Two Minutes Method for mAb-Aggregate Quantification using Parallel Interlaced Size Exclusion HPLC

2.6 Single Injection SEC Protocols

The TSKgel column was loaded with 20 μL of sample and the analysis was run at flow rates between 0.235 mL/min and 1.5 mL/min (30 - 188 cm/h). The ACQUITY column was loaded with 5 μL of sample and run at flow rates between 0.05 mL/min and 0.5 mL/min (18 - 181 cm/h). The Zenix column was loaded with 5 μL of sample and run at flow rates between 0.05-0.96 mL/min respectively (18 - 347 cm/h). The exact flow rates are listed in table II.

2.7 Interlaced SEC protocol

For interlaced SEC experiments the chromatography system was split in two virtual parts as described in section 2.2. It should be noted, that this is not a necessary prerequisite in interlaced chromatography, but rather a convenience for the experimenter. By splitting the instrument and running dedicated programs for UV signal acquisition, the relation of chromatogram and injected sample is facilitated. The methodology described in section 1.1 was applied to the use of ACQUITY columns. A single chromatographic run at a flow rate of 0.4 mL/min was used to determine the initial lag phase (t_{lag}) (see fig. 1 A).

In the adapted method, the data acquisition program on *timebase 2* was triggered by switching a relay on *timebase 1* at $t = t_{lag}$ after injection. The withdrawal of the subsequent sample (pulled-loop mechanism) was triggered 51 seconds prior to injection by using the "PrepareNextSample"- command. This avoided additional hold phases between subsequent control programs.

2.8 Parallel-Interlaced SEC Protocol

To improve throughput further, a second column was run in parallel to the first column using two switching valves directing the flow to the columns and to the detector, respectively. The eluate of one column was directed to the waste right after the monomer peak has passed the detector. The eluate of the second column was then directed to the detector, while the salt peak eluted from the first column into the waste. By running both columns simultaneously in an interlaced mode, the maximum possible throughput of the system was realized (section 1.1). In this work, two ACQUITY columns were used at a flow rate of 0.4 mL/min. The time for sample withdrawal was adjusted to 27 seconds (pulled-loop mechanism). Thoroughly washing of the sample loop and the injection needle was set to be performed within 90 seconds.

2.9 Aggregate Spiking Studies

Aggregate spiking studies were conducted in order to evaluate the linearity of aggregate determination of the presented parallel-interlaced methodology. Two solutions containing different levels of aggregate were mixed to control the level of aggregate in the samples. In order to obtain a solution with a high aggregate content, aggregate was isolated from the proteinA pool. This was done by loading the mAb sample onto a Poros 50 HS (GE Healthcare, Germany) column. Before loading the column, the mAb sample had been adjusted to a conductivity of 15 mS/cm and a pH of 5.5. These conditions had been found to provide high selectivity for mAb aggregates compared to mAb monomer. The elution was performed with a sodium chloride gradient from 10-150 mM in 20 mM MES buffer at pH 5.5. The eluate was collected in fractions, analyzed by SEC and merged to create an aggregate pool with approximately 50 % aggregate. Seventeen aggregate levels were tested ranging from 2.1 to 48.7 %. The samples were first analyzed on two different ACQUITY columns in single injection mode, where each sample was measured sixfold. Subsequently, the presented parallel-interlaced assay was applied, using the same two columns and the same samples which were measured sixfold each. The results were compared in terms of coincidence of the linear regression between expected aggregate level and aggregate level determined via the different approaches.

3 Results & Discussion

SEC columns from three different vendors with different particle size, pore size, and length were applied for mAb aggregate quantification. In contrast to the TSKgel column, the ACQUITY and the Zenix columns have entered the market recently. The TSKgel column has been on the market for almost 25 years and a literature survey revealed a marked preference for this particular column in relation with mAb analysis (data not shown). The chosen columns were compared in terms of generated chromatographic resolution, throughput and precision of aggregate quantification. Based on the results, the best suited column and flow rate was chosen and used to establish a in throughput optimized assay by combining interlaced injections with parallel operation of two SEC columns.

3.1 Single Injections

Three different columns were used to analyze identical mAb samples. Figure 4a shows all three resulting chromatograms. The applied flow rates were 108 cm/h for the ACQUITY, 116 cm/h for the Zenix column, and 126 cm/h for the TSKgel column. For comparability, the chromatograms were normalized with respect to void volume of the respective column (figure 4b). The void volume of each column was defined as the elution volume of the sample buffer. These are listed in table I.

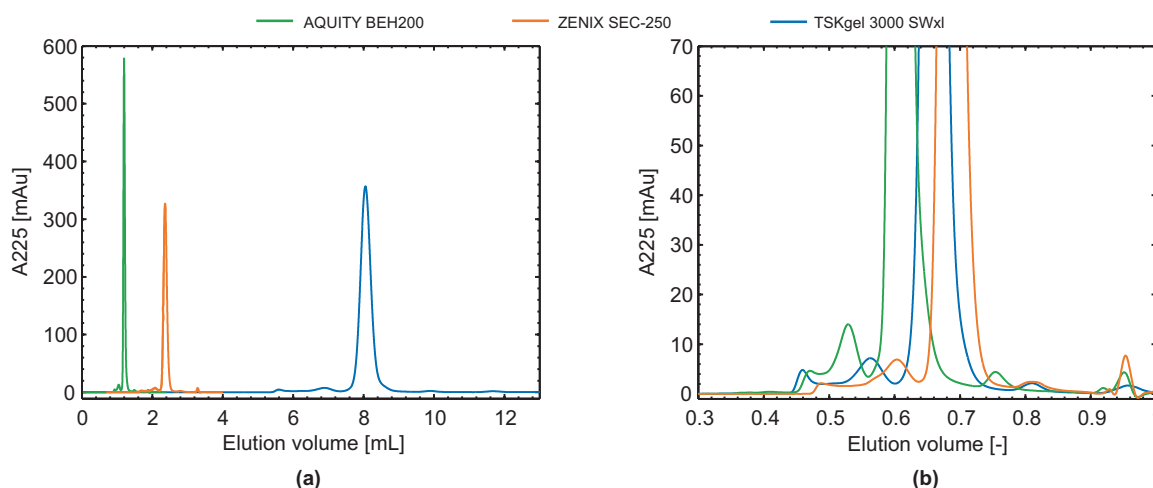


Figure 4: a: Overlay of single injection chromatograms of the mAb sample (1.0 g/L) analyzed on three different SEC columns. b: For comparability, elution volumes were normalized to column void volumes.

The normalized chromatograms revealed similar elution patterns for all columns in which the mAb species eluted over a range from approximately 0.45 to 0.85 void volumes. The elution order, based on normalized elution volume of the monomer species from the three different columns ($V_{ACQUITY} < V_{TSKgel} < V_{Zenix}$) correlated with the decreasing pore size of the column material (ACQUITY: 200 μm , TSKgel: 250 μm , Zenix: 300 μm). The elution profiles generated by the Zenix and the TSKgel column exhibited a more widely stretched elution of the aggregate species. At very low flow rates, these two columns also revealed a third aggregate species in the mAb sample which eluted in between the two main aggregate species (data not shown). However, if an analytical assay aims for the total aggregate level, a resolution of single aggregate species is not necessary. In such a case, the most important parameter is the resolution of the smallest mAb aggregate species (dimer) and the mAb monomer. Hence, in the following the term resolution will refer only to the resolution of mAb monomer and dimer species.

5.6 A Sub-Two Minutes Method for mAb-Aggregate Quantification using Parallel Interlaced Size Exclusion HPLC

3.1.1 Aggregate Levels and Precision

The determined resolution, aggregate level, and coefficient of variation (*cv*) for each applied flow rate and column are listed in table II. The columns were shown to generate different results regarding aggregate level, even though the same mAb sample was analyzed. Using the TSK column, the highest and most stable aggregate level (4.80 % \pm 0.08) over the tested range of flow rates was determined. Using the ACQUITY column, a lower mean aggregate level was determined (4.17 % \pm 0.44) and further the determined aggregate levels exhibited an increase with increasing flow rate (3.79 % - 5.02 %). However, the precision resulting from each tested flow rate was comparable to the accuracy obtained with the TSKgel column ($cv_{mean,TSKgel} = 0.91$, $cv_{mean,ACQUITY} = 0.87$). The overall aggregate level determined using the Zenix column (4.20 % \pm 0.35) was similar to the one obtained with the ACQUITY column, however the accuracy of the results was lower compared to both other columns ($cv_{mean,Zenix} = 1.38$). As for the ACQUITY column, the aggregate level determined with the Zenix column exhibited an increase with increasing flow rate (3.69 % - 4.54 %). For all columns, a tendency of higher precision at medium flow rates was observed.

Table II Aggregate levels determined for a mAb sample using three different columns. Each column was operated at several different flow rates. All displayed results are based on six replicates.

TSKgel® 3000 SWxl				
Flow rate (cm/h)	Flow rate (mL/min)	Aggregate (%)	cv (%)	resolution (-)
30	0.235	4.74	1.91	1.85
44	0.352	4.87	0.85	1.77
63	0.50	4.87	0.60	1.71
94	0.75	4.84	0.27	1.59
126	1.00	4.83	0.52	1.50
157	1.25	4.79	0.48	1.41
188	1.50	4.64	1.75	1.34
ACQUITY UPLC® BEH200 SEC				
Flow rate (cm/h)	Flow rate (mL/min)	Aggregate %	cv %	resolution (-)
18	0.05	3.79	1.94	1.66
27	0.075	3.90	1.00	1.60
36	0.10	3.90	0.99	1.61
72	0.20	4.00	0.48	1.56
108	0.30	4.16	0.27	1.52
144	0.40	4.36	0.94	1.47
181	0.50	5.07	1.52	1.45
Zenix™ SEC-250				
Flow rate (cm/h)	Flow rate (mL/min)	Aggregate %	cv %	resolution (-)
18	0.05	3.69	2.33	1.35
27	0.075	3.96	0.97	1.33
36	0.10	4.11	0.58	1.30
116	0.32	4.28	0.97	1.14
231	0.64	4.62	1.53	1.01
347	0.96	4.54	1.91	0.92

3.1.2 Resolution vs. Analysis Time

The main objective of the presented work, was to establish an ultra-rapid SEC assay for mAb aggregate quantification. Due to the different column dimensions, the correlation between resolution and flow rate does not transmit directly to analysis time. To give an overview of the direct relation between analysis time and chromatographic resolution, the resolution generated for each flow rate and column was plotted as a function of the required time per analysis (figure 5). The evaluation was performed

5.6 A Sub-Two Minutes Method for mAb-Aggregate Quantification using Parallel Interlaced Size Exclusion HPLC

in sequential mode, thus time per analysis equals time needed for processing a single column volume (CV). In general, the decrease in resolution correlated with the particle size of the column material. We found that at assay times above 20 min, the TSKgel column achieved the highest resolution of the columns tested. The resolution achieved under these conditions ranged from 1.59 to 1.85. However, in most cases, a resolution of 1.5 is sufficient for precise quantification. Hence, the high resolution achieved by the TSKgel column at the lower end of the tested flow rates will in some cases be disadvantageous as an unnecessary low throughput is the consequence of the achieved yet dispensable resolution. At lower assay times (increased flow rates) the resolution achieved with the TSK column was shown to decrease faster compared to the ACQUITY column. Of all columns, the ACQUITY column was shown to generate the highest resolution at assay times below 20 min. This finding correlates with the smaller particle size of the ACQUITY column. The tested Zenix column was outperformed by the TSKgel and ACQUITY columns with respect to resolution at all tested assay times. One advantage of the Zenix column was the potentially lower assay time, but the low resolution under these conditions were shown to generate imprecise results (see table II) . However, assay times down to 13 min generated adequate precision ($cv < 1$) despite the low resolution. Hence, taking the relative low cost for the Zenix column compared to the TSKgel and the ACQUITY column into consideration (which exhibits a factor of 1:1.5:2), this column could pose a favorable alternative to the otherwise comprehensive use of the TSKgel column.

Sufficient resolution (~ 1.5) and precision ($cv < 1$) was shown feasible with the ACQUITY column even at very low analysis times. This clearly favours the ACQUITY column for development of a high throughput parallel-interlaced SEC assay. A flow rate of 0.4 ml/min was chosen, both to guarantee sufficient accuracy and also not to operate the column close to maximal flow rate.

The findings presented above are based on measurements performed with only one column per column type. Hence, the conclusions do not take batch and packing variability into consideration. This influence is shown in the studies below. Further, a buffer optimization was not in the scope of this work and changes in performance under other buffer conditions can not be ruled out.

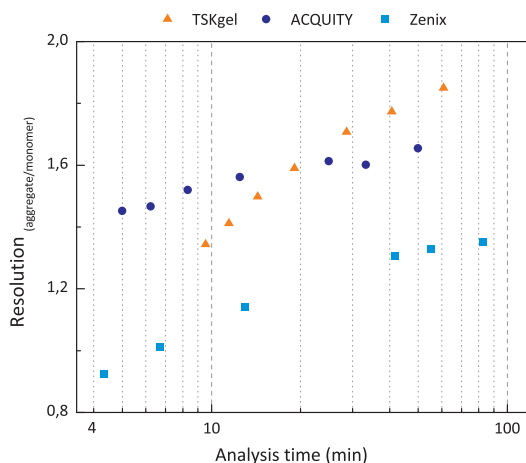


Figure 5: Achieved chromatographic resolution for each tested column displayed as a function of the analysis time. Each data point represents the mean value of six measurements.

3.2 Interlaced SEC

Twenty five injections of the same load material were performed on three ACQUITY columns in interlaced mode. Average analysis time per sample was 3:27 minutes at a flow rate of 0.4 ml/min. Figure 6A shows the resulting A225 trace from the detector. It can be seen that the initial lag time was successfully cut from the analysis time. In this mode of operation 1.43 samples were analyzed per column

5.6 A Sub-Two Minutes Method for mAb-Aggregate Quantification using Parallel Interlaced Size Exclusion HPLC

volume. While aggregate levels resulting from all three columns were in the same range and normally distributed around their mean, pairwise t-tests ($\alpha = 0.01$) showed that all results differed statistically significantly from one another. The first column resulted in a mean aggregate level of 5.08 % with a standard deviation of 0.04. The second column yielded mean 5.02 % with a standard deviation of 0.05. The third column yielded mean 4.91 % with a standard deviation of 0.04.

By interlacing injections and switching to a column of smaller volume and particle size, the assay time was reduced from 14 minutes reported by Farnan [12] to 3:27 minutes. The obvious advantage of using interlaced injections lies in the improved throughput. However, special care has to be taken in order to correctly relate sample and chromatogram. By splitting the instrument into two virtual parts (timebases) a comfortable solution to this problem can be achieved. While throughput was increased, there was still room for optimization. First, column utilization is not optimal as only the initial lag phase is eliminated by interlaced injections. Second, the next sample was not injected until 15 seconds after the salt fraction of the preceding sample had eluted.

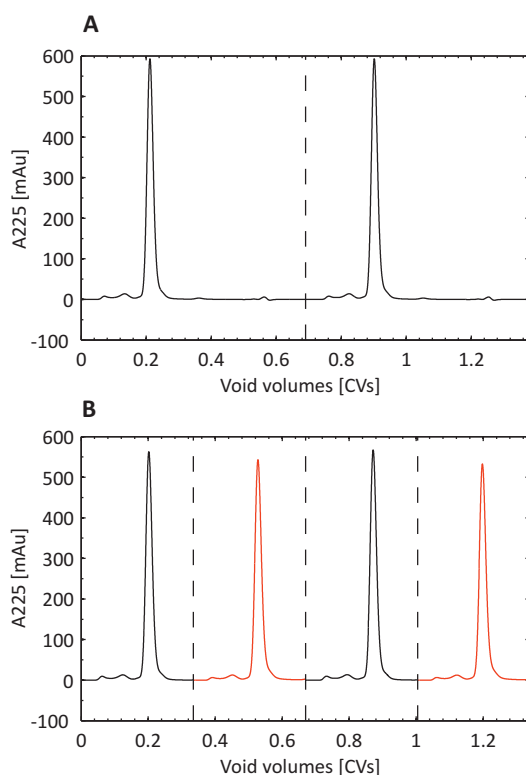


Figure 6: A: A225 absorption data of two injections run in interlaced mode on the ACQUITY column. The dashed line represents the limit between the two samples. 1.43 samples could be analyzed per CV in this mode of operation. B: A225 absorption data of four injections run in parallel-interlaced mode on two ACQUITY columns. The dashed lines represent the moments of switching the column outlet valve to the detector for the subsequent sample. Separate result files are generated for each sample as delimited by the dashed lines. Equally colored lines represent samples analyzed over the same column. Three samples were analyzed per CV in this mode of operation.

3.3 Parallel-Interlaced SEC

Program Parameters

By parallelization of two ACQUITY columns operated with interlaced sample injections, chromatograms containing only the aggregate and monomer areas could be generated. As described in chapter 1.1, the control program was set up based on a single run at a flow rate of 0.4 ml/min. The operation commands of the Chromeleon® software and the corresponding times in the control programs of *timebase 1* and *timebase 2* are summarized in table III. t_{lag} was set to 2:00 minutes. t_{inf}^{min} , the minimal possible analysis time was 1:12 minutes. Twenty-four seconds were added to t_{inf}^{min} to make the method more robust against changes in sample composition. t_{inf} used for programming the method was thus 1:36 minutes. The determined t_{hold} was 1:18 minutes. A sequence of samples was first started with a dummy run in which the first sample is injected but no protein elutes. DAD data acquisition thus generated a blank sample. Immediately after DAD data acquisition has ended, the *outlet valve* was switched. Fifteen seconds were added to the method to ensure a stable baseline after switching the *outlet valve* ($t_{add}^1 = 0:15$). Next, the *inlet valve* was switched. Three seconds were added to the method to flush the autosampler prior to injection ($t_{add}^2 = 0:03$). Triggering the data acquisition was performed three seconds after the sample injection by using the following commands: after the **Inject** command triggered sample injection in *timebase 1*, a **Relay.State = ON** command switched a relay which was connected to an input via cable. A **wait Input.State = ON** as first command in the control method for *timebase 2* triggered the start of this control method and thus of DAD data acquisition as soon as relay 3 was switched. 1:27 minutes later the next sample withdrawal was started using the **PrepareThisSample** command. 0:09 minutes afterwards, DAD data acquisition was stopped thus closing one cycle of sample injection and detection. The process of sample withdrawal took 27 seconds and was performed during the last nine seconds of t_{inf} of the preceding sample and the t_{add}^1 and t_{add}^2 after switching the *outlet valve* and *inlet valve*.

In general, the operating speed of the autosampler was found to be an important factor when programming the control method. Slower autosampling equipment might hinder the implementation of the method. Compared to the data presented, faster autosampling procedures, for example by using an inline split-loop autosampler instead of the used pulled-loop would take the method closer to its theoretical minimum of 1:12 minutes.

To analyze a batch of samples, two batch files were created, one for each timebase. The batch file for *timebase 1* contained two different control programs with each used for every other sample. The two control programs were equal but for the valve switching commands. The batch file for *timebase 2* consisted of a sequence of the DAD control program. The two batch files were started simultaneously.

Method Performance

Fifty injections (25 on each column) of the same mAb load material were performed in parallel-interlaced mode. The analysis time for this batch was 1:57 minutes per sample. Figure 6B shows the resulting detector signal at a wavelength of 225 nm of four consecutive samples. Compared to the standard analytic (single injections, TSKgel column), throughput was improved by 10-15x. Compared to single injections on the same column type, throughput was increased approximately 3x. In accordance to equation 6 the analysis time per sample for n samples can be calculated as follows:

$$t_{analysis} = \frac{t_{lag}}{n} + (t_{inf} + \sum t_{add}) + \frac{t_{hold}}{n} \quad (11)$$

which in our case amounts to

$$1:57 = \frac{2:00}{50} + (1:36 + (0:15 + 0:03)) + \frac{1:18}{50} \quad (12)$$

It is obvious that t_{lag} and t_{hold} do not contribute substantially to the overall analysis time when running the columns in parallel interlaced mode.

A statistical analysis of the results was performed and two data points differing more than 3 standard deviations from the mean value were excluded from further analysis. Average aggregate content detected

5.6 A Sub-Two Minutes Method for mAb-Aggregate Quantification using Parallel Interlaced Size Exclusion HPLC

was 5.03% with a standard deviation of 0.26 This rather large standard deviation was due to differing results from the two separate columns used. Mean aggregate level determined on the first column was 5.27% with a standard deviation of 0.06. Mean aggregate level determined on the second column was 4.78% with a standard deviation of 0.05. While both columns yielded aggregate levels normally distributed around their mean value, results from both column differed statistically significantly as determined by a t-test ($p < 0.1\%$).

Table III Control parameters used to control *timebase 1* (TB1; autosampler, pumps, column compartment including switching valves) and *timebase 2* (TB2; DAD). The commands for injecting five samples are shown. The initial flow path was: sampler → column 1 → DAD. Column “Time” shows the actual time during the analysis. Columns “TB 1” and “TB 2” show the time points programmed into the control programs for timebase 1 and timebase 2. The “action” columns adjacent to the “TB 1” and “TB 2” columns contain the commands used at the corresponding time point. Column “Sample” shows the time during which a sample is on a specific column. The first data acquisition on *timebase 2* generates a chromatogram (‘dummy #’) that only contains the t_{lag} of the first sample. (The two control programs of *timebase 1* differ only in switching valve commands. The data acquisition program on *timebase 2* is started by switching a relay ON.)

Time	Sample	TB 1	Action	Flow path	Action	TB 2
00:00		-0:27	Prepare sample #1			
00:27		0:00	Inject + Start Wash		Wait Input.state = ON	00:00
00:30		0:03	Relay.State = ON		Data Acquisition On	
01:54	sample #1 / column 1	1:27	Pump Acquisition OFF		dummy #	
01:57		1:30	method end			
01:57		-0:27	Prepare sample #2			
02:06		-0:18	switch outlet valve	column 1 → DAD		Data Acquisition Off
02:12					method end	1:42
02:21		-0:03	switch inlet valve	sampler → column 2		
02:24		0:00	Inject + Start Wash		Wait Input.state = ON	00:00
02:27		0:03	Relay.State = ON		Data Acquisition On	
03:51	sample #2 / column 2	1:27	Pump Acquisition OFF		sample #1	
03:54		1:30	method end			
03:54		-0:27	Prepare sample #3			
04:03		-0:18	switch outlet valve	column 2 → DAD		Data Acquisition Off
04:09					method end	01:42
04:18		-0:03	switch inlet valve	sampler → column 1		
04:21		0:00	Inject + Start Wash		Wait Input.state = ON	00:00
04:24		0:03	Relay ON		Data Acquisition On	
05:48	sample #3 / column 1	1:27	Pump Acquisition OFF		sample #2	
05:51		1:30	method end			
05:51		-0:27	Prepare sample #4			
06:00		-0:18	switch outlet valve	column 1 → DAD		Data Acquisition Off
06:06					method end	1:42
06:15		-0:03	switch inlet valve	sampler → column2		
06:18		0:00	Inject + Start Wash		Wait Input.state = ON	00:00
06:21		0:03	Relay.State = ON		Data Acquisition On	
07:45	sample #4 / column 2	1:27	Pump Acquisition OFF		sample #3	
07:48		1:30	method end			
07:48		-0:27	Prepare sample #5			
07:57		-0:18	switch outlet valve	column 2 → DAD		Data Acquisition Off
08:03					method end	01:42
08:12		-0:03	switch inlet valve	sampler → column 1		
08:15		0:00	Inject + Start Wash		Wait Input.state = ON	00:00
08:18		0:03	Relay.State = ON		Data Acquisition On	
09:42	sample #5 / column 1	1:27	Pump Acquisition OFF		sample #4	
09:45		1:30	method end			

The presented method was shown to achieve large improvements of throughput for the particular analysis investigated. Certain prerequisites for achieving these improvements for any given chromatographic assay should be noted. First, the method works for isocratic elutions only, which is the case for SEC and some IEC/HIC analytics. Second, the improvement in assay throughput is related to the ratio of the information to the non-information phases of the chromatogram as only those parts containing no valuable information can be eliminated from the chromatogram. In the case described here, the information phase was approximately 24% of the entire chromatogram. Samples and analysis tasks

5.6 A Sub-Two Minutes Method for mAb-Aggregate Quantification using Parallel Interlaced Size Exclusion HPLC

making use of a larger portion of the chromatogram are amenable to the methodology as described in section 1.1 but might not yield throughput improvements as high as those reported here.

Reliability, robustness, and quantitiveness are the hallmarks of analytical SEC chromatography for mAb-aggregate quantification. Thus, it is preferred over other, even faster analytical methods such as capillary gel electrophoresis. The presented methodology increased sample throughput to an extent that it matches the speed of high throughput experimentation without changing the robust, underlying analytical principle. More detailed studies of aggregation and aggregate depletion during process development and production of mAb based pharmaceuticals can thus be performed.

3.4 Aggregate spiking studies

Aggregate spiking studies resulted in a linear response of the detected aggregate level to the expected aggregate level in the sample throughout the entire range tested (2.1% to 48.7%). The linear regression of measured aggregate level versus expected aggregate level were compared for the two separate columns used and two modes of operation (single and parallel-interlaced injection mode). The linear regression results were found to coincide, slope and intercepts were found to be statistically not different. The overall regression of expected versus measured value was resulted in a R^2 value of 0.9993 with an intercept fixed at 0 and a resulting slope of 1.01. This underlines our conclusion that the method presented herein can replace the standard method of running SEC columns for mAb-aggregate analysis and that the column used is well suited for the analysis task investigated. In theory, increasing aggregate levels could have increased the aggregate peak area to an extent where either monomer-aggregate peak resolution would decrease or where column valve switching times might have had to be adjusted. However, neither was found leading to the conclusion that the presented method is robust regarding aggregate levels of up to 48.7%. Aggregate levels below 2.1% were not investigated owing to the sample material at hand. However, the authors find no reason to believe that lower aggregate levels would pose a problem to the method.

4 Conclusion

In case of total mAb aggregate quantification, we find the ACQUITY column to be the best suited choice of the tested columns, as it enables more than a twofold improvement in throughput when compared to the TSKgel column (assay time comparison at a resolution of 1.5, see figure 5 and table II). Further, due to the relatively low influence of flow rate on the separation which was found for the ACQUITY column, assay throughput can be increased further without compromising resolution significantly. The ACQUITY column also offers the benefits of lower buffer consumption and lower sample volume, latter being of great importance when performing HTPD.

A new methodology to improve throughput for SEC mAb analysis applied in biopharmaceutical science was demonstrated in this paper. By combining interlaced injections with parallel operation of two columns, near optimum utilization of SEC columns for the quantification of monomer and aggregate of a monoclonal antibody solution was achieved. Assay time was reduced to 1:57 minutes per sample as compared to 20-30 minutes using standard analytical protocols. Resulting aggregate levels were found to be comparable between different columns and different modes of operation. As an added benefit, heterogeneity between separate columns is factored into the results by using this method. With analysis times in the range of 2 minutes per sample the method presented in this paper is well suited for current high throughput pharmaceutical process development and process monitoring.

References

- [1] M. Vázquez-Rey, D. a. Lang, Aggregates in monoclonal antibody manufacturing processes., *Biotechnology and bioengineering* 108 (7) (2011) 1494–508.

5.6 A Sub-Two Minutes Method for mAb-Aggregate Quantification using Parallel Interlaced Size Exclusion HPLC

- [2] A. S. Rosenberg, Effects of protein aggregates: an immunologic perspective., *The AAPS journal* 8 (3) (2006) E501–7.
- [3] J. L. Coffman, J. F. Kramarczyk, B. D. Kelley, High-Throughput Screening of Chromatographic Separations: I. Method Development and Column Modeling, *Biotechnology and Bioengineering* 100 (4) (2008) 605–618.
- [4] P. S. Wierling, R. Bogumil, E. Knieps-Grünhagen, J. Hubbuch, High-Throughput Screening of Packed-Bed Chromatography Coupled With SELDI-TOF MS Analysis : Monoclonal Antibodies Versus Host Cell Protein, *Biotechnology and Bioengineering* 98 (2) (2007) 440–450.
- [5] M. Wiendahl, P. S. Wierling, J. Nielsen, D. F. Nielsen, J. Krarup, A. Staby, J. Hubbuch, High Throughput Screening for the Design and Optimization of Chromatographic Processes - Miniturization, Automation, and Parallelization of Breakthrough and Elution Studies, *Chemical Engineering & Technology* 6 (31) (2008) 893–903.
- [6] S. Oelmeier, F. Dimer, J. Hubbuch, Application of an Aqueous Two-Phase Systems High-Throughput Screening Method to Evaluate mAb HCP Separation, *Biotechnology and Bioengineering* 108 (1) (2011) 69–81.
- [7] B. Kelley, Industrialization of mab production technology: the bioprocessing industry at a crossroads., *mAbs* 1 (5) (2009) 443–52.
- [8] J. F. Kramarczyk, B. D. Kelley, J. L. Coffman, High-throughput screening of chromatographic separations: Ii. hydrophobic interaction., *Biotechnology and bioengineering* 100 (4) (2008) 707–20.
- [9] Tosoh Bioscience LLC, Analysis of human immunoglobulins by size exclusion chromatography, application note G00951.
- [10] R. King, C. Miller-Stein, D. Magiera, J. Brann, Description and validation of a staggered parallel high performance liquid chromatography system for good laboratory practice level quantitative analysis by liquid chromatography/tandem mass spectrometry., *Rapid communications in mass spectrometry : RCM* 16 (1) (2002) 43–52.
- [11] M. Nelson, J. Dolan, Parallel chromatography - double your money, *LC GC North America* 22 (4) (2004) 338–343.
- [12] D. Farnan, G. Moreno, J. Stults, A. Becker, G. Tremintin, M. van Gils, Interlaced size exclusion liquid chromatography of monoclonal antibodies., *Journal of chromatography. A* 1216 (51) (2009) 8904–9.
- [13] M. Swartz, Uplc: An introduction and review, *Journal of Liquid Chromatography & Related Technologies* 28 (7) (2005) 1253–1263.

CHAPTER 6

OUTLOOK

This doctoral thesis was centered on aqueous two-phase partitioning of pharmaceutical proteins. A high throughput screening method was implemented and applied to an industrial separation task. Using this platform, previously reported correlations between protein descriptors and distribution were evaluated and it was found that target protein solubility constitutes the limiting factor to the application of these correlations under preparative conditions of high protein load. In accordance, a different approach to screen ATPS for their industrial application was devised and put to use for the selection of ATPSs used in centrifugal partitioning chromatography. To improve the understanding of aqueous two-phase partitioning, a new modeling approach based on molecular dynamics was set up. This approach was initially validated using single PEG molecules in solution. Next, the approach was extended to mixtures of polyethylene glycol and phosphate and finally to lysozyme in PEG/PO₄ mixtures.

While this thesis was able to close some important gaps in the knowledge in both the preparative application and the molecular understanding of aqueous two-phase extraction, it also provides a basis for further investigations. First, it was found that the applicability of centrifugal partitioning chromatography for the preparative purification of pharmaceutical proteins is hampered by the strong dilution of the target protein caused by this process step. Further evaluation of this technology might show ways to augment this problem. Higher ratios of stationary phase might lead to both less dilution and better separation. Additionally, alternative modes of operating the CPC should also be investigated. Centrifugal partitioning chromatography offers the possibility of switching stationary and mobile phase during a run. The target molecule could thus be trapped inside the column first and then stripped of impurities akin to a cross flow extraction. In a second step, a reversal of mobile and stationary phase would lead to the elution of the target protein at the column inlet. Initial experiment in this direction were conducted and showed promising results. As CPC was shown to be capable of generating good resolution with the disadvantage of concurrently diluting the sample, a combination of CPC and precipitation might also be promising. With the process stream already containing high concentrations of polymer or salt, the addition of a small amount of a precipitant might force the target molecule out of solution. This would solve both the dilution issue as well as the separation of the target molecule from the phase forming components. Finally, the integration of the process step into a larger process with subsequent filtration and chromatography steps and the economic evaluation of these constructs could be an interesting topic of investigation and can be conducted on the basis of the results and implementations achieved within the scope of this thesis.

CHAPTER 7

BIBLIOGRAPHY

- [1] Abraham MH, Chadha HS, Whiting GS, and Mitchell RC. Hydrogen Bonding. 32. An Analysis of Water-Octanol and Water-Alkane Partitioning and the $\Delta \log P$ Parameter of Seiler. *Journal of Pharmaceutical Sciences*, **83**(8):1085–1100, 1994
- [2] Aguilar O, Albiter V, Serrano-Carreón L, and Rito-Palomares M. Direct comparison between ion-exchange chromatography and aqueous two-phase processes for the partial purification of penicillin acylase produced by *E. coli*. *Journal of Chromatography B*, **835**(1-2):77–83, 2006
- [3] Aguilar O, Glatz CE, and Rito-Palomares M. Characterization of green-tissue protein extract from alfalfa (*Medicago sativa*) exploiting a 3-D technique. *Journal of Separation Science*, **32**(18):3223–3231, 2009
- [4] Ahamed T, Esteban BNA, Ottens M, Dedem GWKv, Wielen LAMvd, Bisschops MAT, Lee A, Pham C, and Thommes J. Phase behavior of an intact monoclonal antibody. *Biophysical journal*, **93**(2):610–619, 2007
- [5] Al-Marzouqi I, Levy MS, and Lye GJ. Hydrodynamics of PEG-Phosphate Aqueous Two-Phase Systems in a J-Type Multilayer Countercurrent Chromatograph. *Journal of Liquid Chromatography & Related Technologies*, **28**(9):1311–1332, 2005
- [6] Albertsson PÅ. Particle fractionation in liquid two-phase systems The composition of some phase systems and the behavior of some model particles in them application to the isolation of cell walls from microorganisms. *Biochimica et Biophysica Acta*, **27**:378–395, 1958
- [7] Albertsson PÅ. *Partition of cell particles and macromolecules : distribution and fractionation of cells, viruses, microsomes, proteins, nucleic acids, and antigen-antibody complexes in aqueous polymer two-phase systems*. Ph.D. thesis, Uppsala, 1960
- [8] Albertsson PÅ. *Partition of cell particles and macromolecules: distribution and fractionation of cells, mitochondria, chloroplasts, viruses, proteins, nucleic acids, and antigen-antibody complexes in aqueous polymer two-phase systems*. Wiley-Interscience, 1971. ISBN 9780471020479
- [9] Andrews BA, Schmidt AS, and Asenjo JA. Correlation for the partition behavior of proteins in aqueous two-phase systems: Effect of surface hydrophobicity and charge. *Biotechnology and bioengineering*, **90**(3):380–390, 2005
- [10] Arakawa T and Timasheff SN. Mechanism of poly(ethylene glycol) interaction with proteins. *Biochemistry*, **24**(24):6756–6762, 1985

7 Bibliography

- [11] Armstrong DW. Theory and Use of Centrifugal Partition Chromatography. *Journal of Liquid Chromatography*, **11**(12):2433–2446, 1988
- [12] Asenjo JA. Phase separation rates of aqueous two-phase systems: correlation with system properties. *Biotechnology and bioengineering*, **79**:217–223, 2002
- [13] Asenjo JA, Schmidt AS, Hachem F, and Andrews BA. Model for predicting the partition behaviour of proteins in aqueous two-phase systems. *Journal of Chromatography A*, **668**(1):47–54, 1994
- [14] Atha DH and Ingham KC. Mechanism of precipitation of proteins by polyethylene glycols. Analysis in terms of excluded volume. *The Journal of biological chemistry*, **256**(23):12108–12117, 1981
- [15] Aubert JH and Tirrell M. Flow Rate Dependence of Elution Volumes in Size Exclusion Chromatography: A Review. *Journal of Liquid Chromatography & Related Technologies*, **6**(9):219–249, 1983
- [16] Azevedo AM, Rosa PA, Ferreira IF, and Aires-Barros MR. Optimisation of aqueous two-phase extraction of human antibodies. *Journal of Biotechnology*, **132**(2):209–217, 2007
- [17] Azevedo AM, Rosa PAJ, Ferreira IF, and Aires-Barros MR. Integrated process for the purification of antibodies combining aqueous two-phase extraction, hydrophobic interaction chromatography and size-exclusion chromatography. *Journal of Chromatography A*, **1213**(2):154–161, 2008
- [18] Azevedo AM, Rosa PAJ, Ferreira IF, and Aires-Barros MR. Chromatography-free recovery of biopharmaceuticals through aqueous two-phase processing. *Trends in Biotechnology*, **27**(4):240–247, 2009. 432
- [19] Azevedo AM, Rosa PAJ, Ferreira IF, de Vries J, Visser TJ, and Aires-Barros MR. Downstream processing of human antibodies integrating an extraction capture step and cation exchange chromatography. *Journal of Chromatography B*, **877**(1-21-2):50–58, 2009
- [20] Baughman DR and Liu YA. An Expert Network for Predictive Modeling and Optimal-Design of Extractive Bioseparations in Aqueous 2-Phase Systems. *Industrial & Engineering Chemistry Research*, **33**(11):2668–2687, 1994
- [21] Benavides J and Rito-Palomares M. Practical experiences from the development of aqueous two-phase processes for the recovery of high value biological products. *Journal of Chemical Technology & Biotechnology*, **83**(2):133–142, 2008
- [22] Bensch M, Selbach B, and Hubbuch J. High throughput screening techniques in downstream processing: Preparation, characterization and optimization of aqueous two-phase systems. *Chemical Engineering Science*, **62**(7):2011–2021, 2007
- [23] Bensch M, Wierling PS, von Lieres E, and Hubbuch J. High throughput screening of chromatographic phases for rapid process development. *Chemical Engineering & Technology*, **28**(11):1274–1284, 2005
- [24] Berek D. Size exclusion chromatography—a blessing and a curse of science and technology of synthetic polymers. *Journal of separation science*, **33**(3):315–335, 2010

7 Bibliography

- [25] Berendsen HJC, Postma JPM, Gunsteren WFv, DiNola A, and Haak JR. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, **81**(8):3684–3690, 1984
- [26] Berggren K, Wolf A, Asenjo JA, Andrews Ba, and Tjerneld F. The surface exposed amino acid residues of monomeric proteins determine the partitioning in aqueous two-phase systems. *Biochimica et biophysica acta*, **1596**(2):253–268, 2002
- [27] Bérot S, Le Goff E, Foucault A, and Quillien L. Centrifugal partition chromatography as a tool for preparative purification of pea albumin with enhanced yields. *Journal of Chromatography B*, **845**(2):205–209, 2007
- [28] Bérot S, Le Goff E, Foucault A, and Quillien L. Centrifugal partition chromatography as a tool for preparative purification of pea albumin with enhanced yields. *Journal of Chromatography B*, **845**(2):205–9, 2007
- [29] Berthod A and Cardabroch S. Determination of liquid-liquid partition coefficients by separation methods. *Journal of Chromatography A*, **1037**(1-2):3–14, 2004
- [30] Berthod A, Friesen JB, Inui T, and Pauli GF. Elution-countercurrent chromatography: theory and concepts in metabolic analysis. *Analytical chemistry*, **79**(9):3371–3382, 2007
- [31] Berthod A and Hassoun M. Using the liquid nature of the stationary phase in countercurrent chromatography. IV. The cocurrent CCC method. *Journal of chromatography A*, **1116**(1-2):143–8, 2006
- [32] Berthod A, Hassoun M, and Ruiz-Angel MJ. Alkane effect in the Arizona liquid systems used in countercurrent chromatography. *Analytical and bioanalytical chemistry*, **383**(2):327–340, 2005
- [33] Bhambure R, Kumar K, and Rathore AS. High-throughput process development for biopharmaceutical drug substances. *Trends in biotechnology*, **29**(3):127–35, 2011
- [34] Bhat R and Timasheff SN. Steric exclusion is the principal source of the preferential hydration of proteins in the presence of polyethylene glycols. *Protein science: a publication of the Protein Society*, **1**(9):1133–1143, 1992
- [35] Borodin O, Bedrov D, and Smith GD. A Molecular Dynamics Simulation Study of Polymer Dynamics in Aqueous Poly(ethylene oxide) Solutions. *Macromolecules*, **34**(16):5687–5693, 2001
- [36] Brooks D and Seaman G. Detection of differences in surface-charge-associated properties of cells by partition in two-polymer aqueous phase systems. *Nature*, 1971
- [37] Chen J, Ma G, and Li D. HPCPC separation of proteins using polyethylene glycol-potassium phosphate aqueous two-phase. *Preparative Biochemistry & Biotechnology*, **29**(4):371–383, 1999
- [38] Cheng W and Hollis D. Flow-rate effect on elution volume in size-exclusion chromatography. *Journal of Chromatography A*, **408**:9–19, 1987
- [39] Chou KC and Zhang CT. Prediction of Protein Structural Classes. *Critical Reviews in Biochemistry and Molecular Biology*, **30**(4):275–349, 1995

7 Bibliography

- [40] Coffman JL, Kramarczyk JF, and Kelley BD. High-Throughput Screening of Chromatographic Separations: I. Method Development and Column Modeling. *Biotechnology and bioengineering*, **100**(4):605–618, 2008
- [41] Cohn E and Edsall J. *Proteins, Amino Acids and Peptides As Ions and Dipolar Ions*. Reinhold Publishing, New York, 1943
- [42] Cohn EJ. The Physical Chemistry of the Proteins. *Physiological Reviews*, **5**(3):349–437, 1925
- [43] Conway W. *Countercurrent chromatography: apparatus, theory, and applications*. VCH, 1990. ISBN 9780895733313
- [44] Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA, and et al. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*, **117**(19):5179–5197, 1995
- [45] Craig LC. Partition Chromatography and Countercurrent Distribution. *Analytical Chemistry*, **22**(11):1346–1352, 1950
- [46] de Faria JT, Sampaio FC, Converti A, Passos FML, Minim VPR, and Minim LA. Use of response surface methodology to evaluate the extraction of *Debaryomyces hansenii* xylose reductase by aqueous two-phase system. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences*, **877**(27):3031–3037, 2009
- [47] de Folter J and Sutherland IA. Probabilistic model for immiscible separations and extractions (PromISE). *Journal of chromatography A*, **1218**(36):6009–14, 2011
- [48] Devanand K and Selser J. Asymptotic behavior and long-range interactions in aqueous solutions of poly(ethylene oxide). *Macromolecules*, **24**(22):5943–5947, 1991
- [49] Dimroth K, Bohlmann F, Reichard C, and Siepmann T. Über Pyridinium-N-Phenol-Betaine und ihre Verwendung zur Charakterisierung der Polarität von Lösungsmitteln. *Annalen Der Chemie - Justus Liebig*, **661**:1–37, 1963
- [50] Dimer F and Hubbuch J. 3D structure-based protein retention prediction for ion-exchange chromatography. *Journal of chromatography A*, **1217**(8):1343–1353, 2010
- [51] Dimer F and Hubbuch JJ. A novel approach to characterize the binding orientation of lysozyme on ion-exchange resins. *Journal of chromatography A*, **1149**(2):312–20, 2007
- [52] Doi M and Edwards SF. *The Theory of Polymer Dynamics*. Claredon Press, 1986
- [53] Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, and et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of computational chemistry*, **24**(1616):1999–2012, 2003
- [54] Dübel S. *Handbook of therapeutic antibodies: Approved therapeutics*. Handbook of Therapeutic Antibodies. Wiley-VCH, 2007. ISBN 9783527314539

7 Bibliography

- [55] Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, and Pedersen LG. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, **103**(19):8577–8593, 1995
- [56] Fahrner RL, Knudsen HL, Basey CD, Galan W, Feuerhelm D, Vanderlaan M, and Blank GS. Industrial purification of pharmaceutical antibodies: development, operation, and validation of chromatography processes. *Biotechnology & genetic engineering reviews*, **18**(July):301–327, 2001
- [57] Farid SS. Process economics of industrial monoclonal antibody manufacture. *Journal of Chromatography B*, **848**(1):8–18, 2007
- [58] Farnan D, Moreno G, Stults J, Becker A, Tremintin G, and van Gils M. Interlaced size exclusion liquid chromatography of monoclonal antibodies. *Journal of Chromatography A*, **1216**(51):8904–8909, 2009
- [59] Ferreira IF, Azevedo AM, Rosa PAJ, and Aires-Barros MR. Purification of human immunoglobulin G by thermoseparating aqueous two-phase systems. *Journal of Chromatography A*, **1195**(1-2):94–100, 2008
- [60] Folter Jd and Sutherland IA. Universal counter-current chromatography modelling based on counter-current distribution. *Journal of chromatography A*, **1216**(19):4218–4224, 2009
- [61] Foucault A. *Centrifugal partition chromatography*. Chromatographic science. M. Dekker, 1995. ISBN 9780824792572
- [62] Foucault A and Nakanishi K. Comparison of Several Aqueous two Phase Solvent Systems (ATPS) for the Fractionation of Biopolymers by Centrifugal Partition Chromatography (CPC). *Journal of Liquid Chromatography & Related Technologies*, **13**(12):2421–2440, 1990
- [63] Franco TT, Andrews AT, and Asenjo JA. Use of chemically modified proteins to study the effect of a single protein property on partitioning in aqueous two-phase systems: Effect of surface hydrophobicity. *Biotechnology and bioengineering*, **49**(3):300–308, 1996
- [64] French AC, Thompson AL, and Davis BG. High-Purity Discrete PEG-Oligomer Crystals Allow Structural Insight. *Angewandte Chemie-International edition*, **48**(7):1248–1252, 2009
- [65] Frerix A, Schönwald M, Geilenkirchen P, Müller M, Kula MR, and Hubbuch JJ. Exploitation of the coil-globule plasmid DNA transition induced by small changes in temperature, pH salt, and poly(ethylene glycol) compositions for directed partitioning in aqueous two-phase systems. *Langmuir*, **22**:4282–90, 2006
- [66] Gabizon A, Catane R, Uziely B, Kaufman B, Safra T, Cohen R, Martin F, Huang A, and Barenholz Y. Prolonged circulation time and enhanced accumulation in malignant exudates of doxorubicin encapsulated in polyethylene-glycol coated liposomes. *Cancer research*, **54**(4):987–992, 1994
- [67] Greenwald RB, Choe YH, McGuire J, and Conover CD. Effective drug delivery by PEGylated drug conjugates. *Advanced drug delivery reviews*, **55**(2):217–250, 2003

7 Bibliography

- [68] Gregoriadis G. Engineering liposomes for drug delivery: progress and problems. *Trends in Biotechnology*, **13**(12):527–537, 1995
- [69] Grushka E and Grinberg N. *Advances in Chromatography*. Number 47 in Advances in Chromatography. Taylor and Francis, 2009. ISBN 9781420060362
- [70] Guan Y, Bourton E, Hewitson P, Sutherland I, and Fisher D. The importance of column design for protein separation using aqueous two-phase systems on J-type countercurrent chromatography. *Separation and Purification Technology*, **65**(1):79–85, 2009
- [71] Guan Y, Fisher D, and Sutherland IA. Model for spiral columns and stationary phase retention in synchronous coil planet centrifuges. *Journal of Chromatography A*, **1151**(1-2):136–141, 2007
- [72] Guan YH, Fisher D, and Sutherland IA. Protein separation using toroidal columns by type-J synchronous counter-current chromatography towards preparative separation. *Journal of Chromatography A*, **1217**(21):3525–3530, 2010
- [73] Hachem F. Hydrophobic partitioning of proteins in aqueous two-phase systems. *Enzyme and Microbial Technology*, **19**(7):507–517, 1996
- [74] Hacker DL, De Jesus M, and Wurm FM. 25 Years of Recombinant Proteins From Reactor-Grown Cells - Where Do We Go From Here? *Biotechnology advances*, **27**(6):1023–1027, 2009
- [75] Hansson UB and Wingren C. Separation of antibodies by liquid-liquid aqueous partition and by liquid-liquid partition chromatography. *Separation and Purification Methods*, **27**(2):169–211, 1998
- [76] Harris JM and Chess RB. Effect of pegylation on pharmaceuticals. *Nature reviews Drug discovery*, **2**(3):214–221, 2003
- [77] Harris LJ, Larson SB, Hasel KW, and McPherson a. Refined structure of an intact IgG2a monoclonal antibody. *Biochemistry*, **36**(7):1581–1597, 1997
- [78] Hatti-Kaul R. *Aqueous two-phase systems: methods and protocols*. Methods in biotechnology. Humana Press, 2000. ISBN 9780896035416
- [79] Hatti-Kaul R. Aqueous two-phase systems - A general overview. *Molecular Biotechnology*, **19**(3):269–277, 2001
- [80] Hermann R, Lehmann M, and Büchs J. Characterization of gas-liquid mass transfer phenomena in microtiter plates. *Biotechnology and bioengineering*, **81**(2):178–186, 2003
- [81] Huddleston JG and Lyddiatt A. Aqueous 2-Phase Systems in Biochemical Recovery - Systematic Analysis, Design, and Implementation of Practical Processes for the Recovery of Proteins. *Applied Biochemistry and Biotechnology*, **26**(3):249–279, 1990
- [82] Huddleston JG, Veide A, Köhler K, Flanagan J, Enfors SO, and Lyddiatt A. The molecular basis of partitioning in aqueous two-phase systems. *Trends in biotechnology*, **9**(111):381–388, 1991

7 Bibliography

- [83] Huddleston JG, Willauer HD, and Rogers RD. Phase Diagram Data for Several PEG + Salt Aqueous Biphasic Systems at 25 C. *Journal of chemical & engineering data*, **48**(5):1230–1236, 2003
- [84] Ikehata JI, Shinomiya K, Kobayashi K, Ohshima H, Kitanaka S, and Ito Y. Effect of Coriolis force on counter-current chromatographic separation by centrifugal partition chromatography. *Journal of Chromatography A*, **1025**(2):169–175, 2004
- [85] Ito Y. New horizontal flow-through coil planet centrifuge for counter-current chromatography : I. Principle of design and analysis of acceleration. *Journal of Chromatography A*, **188**(1):33–42, 1980
- [86] Ito Y. Speculation on the Mechanism of Unilateral Hydrodynamic Distribution of Two Immiscible Solvent Phases in the Rotating Coil. *Journal of Liquid Chromatography*, **15**(15-16):2639–2675, 1992
- [87] Ito Y, Matsuda K, Ma Y, and Qi L. Toroidal coil counter-current chromatography study of the mass transfer rate of proteins in aqueous-aqueous polymer phase system. *Journal of Chromatography A*, **802**(2):277–283, 1998
- [88] Ito Y, Weinstein M, Aoki I, Harada R, Kimura E, and Nunogaki K. The Coil Planet Centrifuge. *Nature*, **212**(5066):985–987, 1966
- [89] Iyer HV and Przybycien TM. Protein precipitation: Effects of mixing on protein solubility. *AIChE Journal*, **40**(2):349–360, 1994
- [90] Jakalian A, Jack DB, and Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of computational chemistry*, pages 1623–1641, 2002
- [91] Jin M, Szapiel N, Zhang J, Hickey J, and Ghose S. Profiling of host cell proteins by two-dimensional difference gel electrophoresis (2D-DIGE): Implications for downstream process development. *Biotechnology and bioengineering*, **105**(2):306–316, 2010
- [92] Johansson G, Joelsson M, Olde B, and Shanbhag VP. Affinity partitioning of biopolymers and membranes in ficoll-dextran aqueous two-phase systems. *Journal of Chromatography A*, **331**:11–21, 1985
- [93] Johansson G, Sarnesto A, Høge-Jensen E, Szabo-Lin I, Guthenberg C, and Mannervik B. Effects of Salts on the Partition of Proteins in Aqueous Polymeric Biphasic Systems. *Acta Chemica Scandinavica*, **28b**:873–882, 1974
- [94] Kamlet M and Taft R. The solvatochromic comparison method. I. The beta-scale of solvent hydrogen-bond acceptor (HBA) basicities. *Journal of the American Chemical Society*, **98**(2):377–383, 1976
- [95] Kamlet MJ, Abboud JL, and Taft RW. The solvatochromic comparison method. VI. The pi* scale of solvent polarities. *Journal of the American Chemical Society*, **99**(18):6027–6038, 1977
- [96] Karr LJ, Shafer SG, Harris J, Alstine JMV, and Snyder RS. Immuno-affinity partition of cells in aqueous polymer two-phase systems. *Journal of Chromatography A*, **354**:269–282, 1986. <m:note/>
- [97] Kelley B. Industrialization of mAb production technology: the bioprocessing industry at a crossroads. *mAbs*, **1**(5):443–452, 2009

7 Bibliography

- [98] Kelley BD, Switzer M, Bastek P, Kramarczyk JF, Molnar K, Yu T, and Coffman J. High-throughput screening of chromatographic separations: IV. Ion-exchange. *Biotechnology and bioengineering*, **100**(5):950–963, 2008
- [99] Kent JT, Mardia KV, and Jammalamadaka SR. Characterization of the Uniform-Distribution on the Circle. *Annals of Statistics*, **7**(4):882–889, 1979
- [100] Kepka C, Collet E, Persson J, Stahl A, Lagerstedt T, Tjerneld F, and Veide A. Pilot-scale extraction of an intracellular recombinant cutinase from *E. coli* cell homogenate using a thermoseparating aqueous two-phase system. *Journal of Biotechnology*, **103**(2):165–181, 2003
- [101] King R, Miller-Stein C, Magiera D, and Brann J. Description and validation of a staggered parallel high performance liquid chromatography system for good laboratory practice level quantitative analysis by liquid chromatography/tandem mass spectrometry. *Rapid communications in mass spectrometry*, **16**(1):43–52, 2002
- [102] Koenig JL and Angood AC. Raman spectra of poly(ethylene glycols) in solution. *Journal of Polymer Science Part A-2: Polymer Physics*, **8**(10):1787–1796, 1970
- [103] Konagurthu AS, Whisstock JC, Stuckey PJ, and Lesk AM. MUSTANG: A Multiple Structural Alignment Algorithm. *Proteins*, **64**(3):559–574, 2006
- [104] Kramarczyk JF, Kelley BD, and Coffman JL. High-throughput screening of chromatographic separations: II. Hydrophobic interaction. *Biotechnology and bioengineering*, **100**(4):707–720, 2008
- [105] Krieger E, Darden T, Nabuurs SB, Finkelstein A, and Vriend G. Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins*, **57**(4):678–683, 2004
- [106] Krieger E, Koraimann G, and Vriend G. Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins*, **47**(3):393–402, 2002
- [107] Krieger E, Nielsen JE, Spronk CaEM, and Vriend G. Fast empirical pKa prediction by Ewald summation. *Journal of molecular graphics & modelling*, **25**(4):481–486, 2006
- [108] Kumar A and Galaev I. *Precipitation of proteins: Nonspecific and specific.*, chapter 7. CRC Press, 2003. ISBN 0-8247-4759-3
- [109] Kyte J and Doolittle RF. A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology*, **157**(1):105–132, 1982
- [110] Lee H, Venable RM, Mackerell AD, and Pastor RW. Molecular dynamics studies of polyethylene oxide and polyethylene glycol: hydrodynamic radius and shape anisotropy. *Biophysical journal*, **95**(4):1590–1599, 2008
- [111] Lee H, Vries AHd, Marrink SJ, and Pastor RW. A coarse-grained model for polyethylene oxide and polyethylene glycol: conformation and hydrodynamics. *The journal of physical chemistry B*, **113**(40):13186–13194, 2009
- [112] Li H, Robertson AD, and Jensen JH. Very fast empirical prediction and rationalization of protein pKa values. *Proteins*, **61**(4):704–721, 2005

7 Bibliography

- [113] Liu K and Parsons J. Solvent effects on preferred conformation of poly(ethylene glycols). *Macromolecules*, **2**(5):529–533, 1969
- [114] Low D, O’Leary R, and Pujar NS. Future of antibody purification. *Journal of Chromatography B*, **848**(1):48–63, 2007
- [115] Macko T and Berek D. Pressure Effects in Hplc: Influence of Pressure and Pressure Changes on Peak Shape, Base Line, and Retention Volume in Hplc Separations. *Journal of Liquid Chromatography & Related Technologies*, **24**(9):1275–1293, 2001
- [116] Madeira PP, Reis Ca, Rodrigues AE, Mikheeva LM, Chait A, and Zaslavsky BY. Solvent Properties Governing Protein partitioning in Polymer/Polymer Aqueous two-phase Systems. *Journal of Chromatography A*, **1218**(10):1379–1384, 2011
- [117] Madeira PP, Reis CA, Rodrigues AE, Mikheeva LM, and Zaslavsky BY. Solvent Properties Governing Solute Partitioning in Polymer/Polymer Aqueous Two-Phase Systems: Nonionic Compounds. *Journal of Physical Chemistry B*, **114**(1):457–462, 2010
- [118] Marchal L, Foucault AP, Patissier G, Rosant JM, and Legrand J. Chapter 5 Centrifugal partition chromatography: an engineering approach. In A Berthod (editor), *Countercurrent Chromatography*, volume 38 of *Comprehensive Analytical Chemistry*, pages 115–157. Elsevier, 2002
- [119] Marcus Y. Linear solvation energy relationships. Correlation and prediction of the distribution of organic solutes between water and immiscible organic solvents. *The Journal of Physical Chemistry*, **95**(22):8886–8891, 1991
- [120] Mazzola PG, Lopes AM, Hasmann FA, Jozala AF, Penna TCV, Magalhaes PO, Rangel-Yagui CO, and Pessoa A. Liquid-liquid extraction of biomolecules: an overview and update of the main techniques. *Journal of Chemical Technology and Biotechnology*, **83**(2):143–157, 2008. 261
- [121] McLachlan G and Peel D. *Finite Mixture Models*. John Wiley & Sons, New York, 2000
- [122] Melander WR and Horvath C. Salt effects on hydrophobic interactions in precipitation and chromatography of proteins: An interpretation of the lyotropic series. *Archives of biochemistry and biophysics*, **183**(1):200–215, 1977
- [123] Merchuk JC, Andrews BA, and Asenjo JA. Aqueous two-phase systems for protein separation Studies on phase inversion. *Journal of Chromatography B*, **711**(1-2):285–293, 1998
- [124] Middaugh C, Tisel W, Haire R, and Rosenberg A. Determination of the apparent thermodynamic activities of saturated protein solutions. *Journal of Biological Chemistry*, **254**(2):367–370, 1979
- [125] MIHEEVA L, ZASLAVSKY B, and ROGOZHIN S. Choice of an aqueous polymer two-phase system for cell partition. *Biochimica et Biophysica Acta - General Subjects*, **542**(1):101–106, 1978
- [126] Murayama W, Kobayashi T, Kosuge Y, Yano H, Nunogaki Y, and Nunogaki K. A new centrifugal counter-current chromatograph and its application. *Journal of Chromatography A*, **239**:643–649, 1982

7 Bibliography

- [127] Nelson M and Dolan J. Parallel Chromatography - Double Your Money. *LC GC North America*, **22**(4):338–343, 2004
- [128] Nfor BK, Verhaert PDEM, van der Wielen LAM, Hubbuch J, and Ottens M. Rational and systematic protein purification process development: the next generation. *Trends in Biotechnology*, **27**(12):673–679, 2009
- [129] Oelmeier SA, Dismer F, and Hubbuch J. Application of an aqueous two-phase systems high-throughput screening method to evaluate mAb HCP separation. *Biotechnology and bioengineering*, **108**(1):69–81, 2010
- [130] Olivera-Nappa A, Lagomarsino G, Andrews BA, and Asenjo JA. Effect of electrostatic energy on partitioning of proteins in aqueous two-phase systems. *Journal of Chromatography B*, **807**(1):81–86, 2004
- [131] Paul W and Smith GD. Structure and dynamics of amorphous polymers: computer simulations compared to experiment and theory. *Reports on Progress in Physics*, **67**(7):1117–1185, 2004
- [132] Pericin DM, Maddarev-Popovic SZ, and Radulovic-Popovic LM. Optimization of conditions for acid protease partitioning and purification in aqueous two-phase systems using response surface methodology. *Biotechnology Letters*, **31**(1):43–47, 2009
- [133] Perosa F, Carbone R, Ferrone S, and Dammacco F. Purification of human immunoglobulins by sequential precipitation with caprylic acid and ammonium sulphate. *Journal of immunological methods*, **128**(1):9–16, 1990
- [134] Persson J and Lester P. Purification of antibody and antibody-fragment from *E. coli* homogenate using 6,9-diamino-2-ethoxyacridine lactate as precipitation agent. *Biotechnology and bioengineering*, **87**(3):424–434, 2004
- [135] Pieper U, Webb BM, Barkan DT, Schneidman-Duhovny D, Schlessinger A, Braberg H, Yang Z, Meng EC, Pettersen EF, Huang CC, and et al. ModBase, a database of annotated comparative protein structure models, and associated resources. *Nucleic acids research*, **39**:D465–D474, 2011
- [136] Platis D and Labrou NE. Application of a PEG/salt aqueous two-phase partition system for the recovery of monoclonal antibodies from unclarified transgenic tobacco extract. *Biotechnology Journal*, **4**(9):1320–1327, 2009
- [137] Podo F, Ray A, and Nemethy G. Structure and hydration of nonionic detergent micelles - high-resolution nuclear magnetic-resonance study. *Journal of the American Chemical Society*, **95**(19):6164–6171, 1973
- [138] Popovici S and Schoenmakers P. Fast size-exclusion chromatography—theoretical and practical considerations. *Journal of Chromatography A*, **1099**(1-2):92–102, 2005
- [139] Przybycien TM and Bailey JE. Aggregation kinetics in salt-induced protein precipitation. *AIChE Journal*, **35**(11):1779–1790, 1989
- [140] Przybycien TM, Pujar NS, and Steele LM. Alternative bioseparation operations: life beyond packed-bed chromatography. *Current opinion in biotechnology*, **15**(5):469–478, 2004

7 Bibliography

- [141] Raghavarao K, Rastogi N, Gowthaman M, and Karanth N. Aqueous Two-Phase Extraction for Downstream Processing of Enzymes/Proteins. In SL Neidleman and AI Laskin (editors), *Advances in Applied Microbiology*, volume 41 of *Advances in Applied Microbiology*, pages 97–171. Academic Press, 1995
- [142] Rathore AS. Roadmap for implementation of quality by design (QbD) for biotechnology products. *Trends in biotechnology*, **27**(9):546–53, 2009
- [143] Reichardt C. Solvatochromic Dyes as Solvent Polarity Indicators. *Chemical Reviews*, **94**(8):2319–2358, 1994
- [144] Ribeiro SC, Monteiro Ga, Cabral J, and Prazeres DMF. Isolation of plasmid DNA from cell lysates by aqueous two-phase systems. *Biotechnology and bioengineering*, **78**(4):376–84, 2002
- [145] Ricker RD and Sandoval LA. Fast, reproducible size-exclusion chromatography of biological macromolecules. *Journal of Chromatography A*, **743**(1):43–50, 1996
- [146] Rito-Palomares M. Practical application of aqueous two-phase partition to process development for the recovery of biological products. *J Chromatogr B*, **807**:3–11, 2004
- [147] Roque ACA, Lowe CR, and Taipa MA. Antibodies and genetically engineered related molecules: Production and purification. *Biotechnology Progress*, **20**(3):639–654, 2004
- [148] Rosa PAJ. Application of central composite design to the optimisation of aqueous two-phase extraction of human antibodies. *Journal of Chromatography A*, **1141**:50–60, 2007
- [149] Rosa PAJ, Ferreira IF, Azevedo AM, and Aires-Barros MR. Aqueous two-phase systems: A viable platform in the manufacturing of biopharmaceuticals. *Journal of Chromatography A*, **1217**(16):2296–2305, 2010
- [150] Rosenberg AS. Effects of protein aggregates: an immunologic perspective. *The AAPS journal*, **8**(3):E501–E507, 2006
- [151] Sackett DL and Wolff J. Nile Red as a Polarity-Sensitive Fluorescent-Probe of Hydrophobic Protein Surfaces. *Analytical Biochemistry*, **167**(2):228–234, 1987
- [152] Salgado JC, Andrews BA, Ortuzar MF, and Asenjo JA. Prediction of the partitioning behaviour of proteins in aqueous two-phase systems using only their amino acid composition. *Journal of Chromatography A*, **1178**(1-2):134–144, 2008
- [153] Sasakawa S and Walter H. Partition behavior of native proteins in aqueous dextran-poly(ethylene glycol)-phase systems. *Biochemistry*, **11**(15):2760–2765, 1972
- [154] Schiel JE and Hage DS. Density measurements of potassium phosphate buffer from 4 to 45 degrees C. *Talanta*, **65**(2):495–500, 2005
- [155] Schindler J and Nothwang HG. Aqueous polymer two-phase systems: Effective tools for plasma membrane proteomics. *Proteomics*, **6**(20):5409–5417, 2006. 101
- [156] Schmidt A, Ventom AM, and Asenjo JA. Partitioning and purification of α -amylase in aqueous two-phase systems. *Enzyme and Microbial Technology*, **16**(2):131–142, 1994

7 Bibliography

- [157] Seebach D, Zass E, Schweizer WB, Thompson AJ, French A, Davis BG, Kyd G, and Bruno IJ. Polymer Backbone Conformation-A Challenging Task for Database Information Retrieval. *Angewandte Chemie-International Edition*, **48**(51):9596–9598, 2009
- [158] Shih YC, Prausnitz JM, and Blanch HW. Some characteristics of protein precipitation by salts. *Biotechnology and bioengineering*, **40**(10):1155–1164, 1992
- [159] Shukla AA, Hubbard B, Tressel T, Guhan S, and Low D. Downstream processing of monoclonal antibodies—application of platform approaches. *Journal of chromatography B*, **848**:28–39, 2007
- [160] Shulgin IL and Ruckenstein E. Preferential hydration and solubility of proteins in aqueous solutions of polyethylene glycol. *Biophysical chemistry*, **120**(3):188–198, 2006
- [161] Silverio SC, Rodriguez O, Teixeira JA, and Macedo EA. Solute partitioning in polymer-salt ATPS: The Collander equation. *Fluid Phase Equilibria*, **296**(2):173–177, 2010
- [162] Smith G and Bedrov D. A molecular dynamics simulation study of the influence of hydrogen-bonding and polar interactions on hydration and conformations of a poly(ethylene oxide) oligomer in dilute aqueous solution. *Macromolecules*, **35**(14):5712–5719, 2002
- [163] Sorin EJ and Pande VS. Exploring the Helix-Coil Transition via All-Atom Equilibrium Ensemble Simulations. *Biophysical Journal*, **88**(4):2472–2493, 2005
- [164] Stein A and Kiesewetter A. Cation exchange chromatography in antibody purification: pH screening for optimised binding and HCP removal. *Journal of Chromatography B*, **848**(1):151–158, 2007
- [165] Sunasara KM, Xia F, Gronke RS, and Cramer SM. Application of hydrophobic interaction displacement chromatography for an industrial protein purification. *Biotechnology and bioengineering*, **82**(3):330–339, 2003
- [166] Susanto A, Treier K, Knieps-Gruenhagen E, von Lieres E, and Hubbuch J. High Throughput Screening for The Design and Optimization of Chromatographic Processes: Automated Optimization of Chromatographic Phase Systems. *Chemical Engineering & Technology*, **32**(1):140–154, 2009
- [167] Sutherland IA. Recent progress on the industrial scale-up of counter-current chromatography. *Journal of Chromatography A*, **1151**(1-2):6–13, 2007
- [168] Sutherland IA, Audo G, Bourton E, Couillard F, Fisher D, Garrard I, Hewitson P, and Intes O. Rapid linear scale-up of a protein separation by centrifugal partition chromatography. *Journal of Chromatography A*, **1190**(1-2):57–62, 2008
- [169] Sutherland Ia, de Folter J, and Wood P. Modelling CCC Using an Eluting Countercurrent Distribution Model. *Journal of Liquid Chromatography & Related Technologies*, **26**(9&10):1449–1474, 2003
- [170] Swartz M. UPLC: An Introduction and Review. *Journal of Liquid Chromatography & Related Technologies*, **28**(7):1253–1263, 2005

7 Bibliography

- [171] Taft RW and Kamlet MJ. The solvatochromic comparison method II. The alpha-scale of solvent hydrogen-bond donor (HBD) acidities. *Journal of the American Chemical Society*, **98**(10):2886–2894, 1976
- [172] Tanford C. *Physical Chemistry of Macromolecules*. John Wiley & Sons, 1961
- [173] Tasaki K. Poly(oxyethylene)-Water Interactions : A Molecular Dynamics Study. *Journal of the American Chemical Society*, **118**:8459–8469, 1996
- [174] Thommes J and Etzel M. Alternatives to chromatographic separations. *Biotechnology progress*, **23**(1):42–5, 2007
- [175] Thommes J, Halfar M, Gieren H, Curvers S, Takors R, Brunschier R, and Kula MR. Human chymotrypsinogen B production from *Pichia pastoris* by integrated development of fermentation and downstream processing. Part 2. Protein recovery. *Biotechnology Progress*, **17**(3):503–512, 2001
- [176] Thomson Reuters. *ISI Web of Knowledge*, 2011
- [177] Tosoh Bioscience LLC. *Analysis of human immunoglobulins by size exclusion chromatography*. Application note G00951
- [178] van Buel MJ, van Halsema FED, van der Wielen LAM, and Luyben KCAM. Flow regimes in centrifugal partition chromatography. *AIChE Journal*, **44**(6):1356–1362, 1998
- [179] van Deemter JJ, Zuiderweg FJ, and Klinkenberg A. Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. *Chemical Engineering Science*, **5**(6):271–289, 1956
- [180] Vorobyov I, Anisimov VM, Greene S, Venable RM, Moser A, Pastor RW, and MacKerell AD. Additive and Classical Drude Polarizable Force Fields for Linear and Cyclic Ethers. *Journal of Chemical Theory and Computation*, **3**(3):1120–1133, 2007
- [181] Vázquez-Rey M and Lang Da. Aggregates in monoclonal antibody manufacturing processes. *Biotechnology and bioengineering*, **108**(7):1494–508, 2011
- [182] Walter H, Brooks D, and Fisher D. *Partitioning in aqueous two-phase systems: theory, methods, uses, and application to biotechnology*. Academic Press, 1985. ISBN 9780127338613
- [183] Walter H, Krob EJ, Garza R, and Ascher GS. Partition and countercurrent distribution of erythrocytes and leukocytes from different species. *Experimental cell research*, **55**(1):57–64, 1969
- [184] Walter H and Sasakawa S. Partition of closely related proteins in aqueous two-polymer phase systems. Human hemoglobin variants and hemoglobins from different species. *Biochemistry*, **10**(1):108–13, 1971
- [185] Wang J, Wolf RM, Caldwell JW, Kollman Pa, and Case Da. Development and testing of a general amber force field. *Journal of computational chemistry*, **25**(9):1157–1174, 2004
- [186] Wensel DL, Kelley BD, and Coffman JL. High-throughput screening of chromatographic separations: III. Monoclonal antibodies on ceramic hydroxyapatite. *Biotechnology and bioengineering*, **100**(5):839–54, 2008

7 Bibliography

- [187] Westoby M, Rogers J, Haverstock R, Romero J, and Pieracci J. Modeling industrial centrifugation of mammalian cell culture using a capillary based scale-down system. *Biotechnology and bioengineering*, **108**(5):989–998, 2010
- [188] Wiendahl M, Völker C, Husemann I, Krarup J, Staby A, Scholl S, and Hubbuch JJ. A novel method to evaluate protein solubility using a high throughput screening approach. *Chemical Engineering Science*, **64**(17):3778–3788, 2009
- [189] Wiendahl M, Wierling PS, Nielsen J, Nielsen DF, Krarup J, Staby A, and Hubbuch J. High Throughput Screening for the Design and Optimization of Chromatographic Processes - Minuturization, Automation, and Parallelization of Breakthrough and Elution Studies. *Chemical Engineering & Technology*, **6**(31):893–903, 2008
- [190] Wierling PS, Bogumil R, Knieps-Grünhagen E, and Hubbuch J. High-Throughput Screening of Packed-Bed Chromatography Coupled With SELDI-TOF MS Analysis : Monoclonal Antibodies Versus Host Cell Protein. *Biotechnology and bioengineering*, **98**(2):440–450, 2007
- [191] Winger M, de Vries AH, and van Gunsteren WF. Force-field dependence of the conformational properties of α,ω -dimethoxypolyethylene glycol. *Molecular Physics*, **107**(13):1313–1321, 2009
- [192] Wurm FM. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nature biotechnology*, **22**(11):1393–1398, 2004
- [193] Zaslavsky B, Baevskii A, Rogozhin S, Gedrovich A, Shishkov A, Gasanov A, and Masimov A. Relative hydrophobicity of synthetic macromolecules I. Polyethylene glycol, polyacrylamide and polyvinylpyrrolidone. *Journal of Chromatography A*, **285**:63–68, 1984
- [194] Zaslavsky B, Bagirov T, Borovskaya A, Gulaeva N, Miheeva L, Mahmudov A, and Rodnikova M. Structure of water as a key factor of phase separation in aqueous mixtures of two nonionic polymers. *Polymer*, **30**(111):2104–2111, 1989
- [195] Zaslavsky BY. *Aqueous two-phase partitioning: physical chemistry and bioanalytical applications*. Marcel Dekker, Inc., 1995. ISBN 0824794613

CHAPTER 8

CURRICULUM VITAE

Education

- May 2008 to June 2011 **PhD** at the Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology.
Development of aqueous two-phase extraction and precipitation screening procedures for the purification of pharmaceutical proteins. Basic research on ATPS including molecular dynamics simulations.
- Jul 2007 to May 2008 **M.Sc. thesis** at the PBC Group, BiogenIdec, San Diego, CA, USA. "Evaluation of Displacement Chromatography for the industrial scale purification of a complex fusion protein"
High throughput screening of displacer-adsorber combinations. Evaluation and optimization of HIC and AIEX displacement chromatography.
- Oct 2005 to Jul 2007 Study of biotechnology (M. Sc.), RWTH-University in Aachen, Germany. Focus on industrial biotechnology, genetics and microbiology, structural biology, plant biotechnology.
- Aug 2005 to Oct 2005 Internship at the Institute for Molecular Biotechnology (Bio7), RWTH-University in Aachen, Germany.
Genetic modification of *A. tumefaciens*. Transient expression of allergens in tobacco leaves by means of vacuum infiltration. Purification and characterization of the allergens produced.
- Apr 2005 to Jul 2005 **B.Sc. thesis** at the Institute for Biotechnology (Bio6), RWTH-University in Aachen, Germany. "Characterization of the promoter and first gene of the *phe*-operon of *Geobacillus stearothermophilus*"
mark: very good
overall mark: 1,1 (excellent)
Genetic modification of *E. coli*. Small scale fermentation of *E. coli*. Purification and characterization of the enzyme produced.
- Oct 2002 to Apr 2005 Study of biotechnology (B. Sc.), RWTH-University in Aachen, Germany.

Work Experience

- July 2011 to present PostDoc at the chair of Biomolecular Separation Engineering, Karlsruhe Institute of Technology.
- May 2008 to June 2011 PhD student at the chair of Biomolecular Separation Engineering, Karlsruhe Institute of Technology.
- Jul 2007 to Jan 2008 Temporary research assistant, Process Biochemistry group of BiogenIdec Inc., SanDiego, USA.
- Apr 2007 to Jun 2007 Intern at the Institute for Biotechnology ("Avdelningen för Bioteknik") at the University of Lund, Sweden.
- Oct 2005 to Mar 2007 Student assistant at the "Institute for Biotechnology" (Bio6) at the RWTH-University in Aachen, Germany.

Languages

German	Native speaker.
English	Fluent (written and spoken). Level C1 certificate ("Effective Operational Proficiency") obtained.
French	Intermediate. 6 years of schooling.
Swedish	Basic.

Softskills

Teaching	Realization of a literature seminar for students, development and supervision of an experiment for a mandatory student practical
Mentoring	Mentoring of two student research projects, three diploma students, and one PhD student
Workshops	Attended the following workshops with the Karlsruhe House of Young Scientists (KHYS), the professional development center for academics at the Karlsruhe Institute of Technology: Scientific writing, Communication skills, Conflict management, Working in a team, Project management, Leadership

Publications

Posters

BPP2009	Stefan Oelmeier, Jürgen Hubbuch Screening and optimization of Aqueous Two Phase Systems for the purification of proteins using a robotic platform <i>15th International Conference on BioPartitioning and Purification – West London - 2009</i>
ISPPP2010	Stefan Oelmeier, Florian Dimer, Jürgen Hubbuch Characterizing phase formation in aqueous two phase systems by molecular dynamics simulations <i>International Symposium on the Separation of Proteins, Peptides & Polynucleotides – Bologna - 2010</i>
HTPD2010	Stefan Oelmeier, Jürgen Hubbuch Statistical Profiling of an automated screening method in the case of an ATPS screening tool <i>1st International Conference on High-Throughput Process Development – Krakow - 2010</i>
BPP2011	Florian Dimer, Stefan Oelmeier, Jürgen Hubbuch Robotics aided solvatochromic characterization of PEG-PO4 ATPS <i>16th International Conference on BioPartitioning and Purification – Puerto Vallarta - 2011</i>
BPP2011	Stefan Oelmeier, Florian Dimer, Jürgen Hubbuch Evaluation of Centrifugal Partitioning Chromatography for the purification of monoclonal antibodies <i>16th International Conference on BioPartitioning and Purification – Puerto Vallarta - 2011</i>

Publications (continued)

Talks

- ECAB2011 Stefan Oelmeier, Florian Dismer, Jürgen Hubbuch
Gaining mechanistic understanding of aqueous two-phase systems for bioseparation
1st European Congress of Applied Biotechnology – Berlin - 2011
- BPI2011 Stefan Oelmeier, Jürgen Hubbuch
High throughput analytics
BioProcess International conference – Long Beach, CA, USA - 2011

Paper

- BioBio2011 Stefan Oelmeier, Florian Dismer, Jürgen Hubbuch
Application of an aqueous two-phase systems high-throughput screening method to evaluate mAb HCP separation.
Biotechnology and bioengineering, 108(1), pp.69-81
- JCA2011 Patrick Diederich, Sigrid Hansen, Stefan Oelmeier, Bianca Stolzenberger, Jürgen Hubbuch
A sub-two minutes method for monoclonal antibody-aggregate quantification using parallel interlaced size exclusion high performance liquid chromatography.
Journal of Chromatography A, 1218(50), pp.9010-9018
-
-