COMPUTATIONAL AND EXPERIMENTAL CHARACTERIZATION OF PROTEINS WITH RESPECT TO PROTEIN-SOLVENT INTERACTIONS

---

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

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

genehmigte
DISSERTATION

von
Dipl.-Ing. Sven Matthias Amrhein
geboren in Würzburg

Referent:	Prof. Dr. Jürgen Hubbuch
Korreferent:	Prof. Dr. Matthias Franzreb
Tag der mündlichen Prüfung:	9. Mai 2016

Danksagung

Ich möchte mich ganz herzlich bei Prof. Dr. Jürgen Hubbuch für die Möglichkeit bedanken, diese Dissertation innerhalb seiner Forschungsgruppe anzufertigen. Jürgen, vielen Dank für deine offene und freundschaftliche Art, dein stetes Interesse und Vertrauen und die vielen gestalterischen Freiräume, ohne welche ich mich persönlich wie auch professionell in dieser Form nicht entwickeln hätte können.

Vielen Dank an Prof. Dr. Matthias Franzreb für die freundliche Übernahme des Korreferats und die Möglichkeit, einige 3D-Drucke in seiner Arbeitsgruppe anzufertigen.

Ich möchte mich bei Jonas Wohlgemuth für die Verwandlung vieler meiner Ideen mittels 3D-Druck in kleine Plastikteile bedanken.

Die vorliegende Arbeit wurde im Rahmen des Projektes „Molecular Interaction Engineering: From Nature's Toolbox to Hybrid Technical Systems“ mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF) unter dem Förderkennzeichen 031A095B gefördert. Vielen Dank für die finanzielle Unterstützung.

Ich möchte mich auch bei Florian Dismer bedanken, der das Grundvorhaben des Projektes in der ursprünglichen Form gestaltete und die ersten Entwicklungen des Projektes mit Interesse verfolgte.

I want to give my special thanks to Dave Smithson, Martha Tse, Samir Sane, Dana Anderson, Krista Lee, Kimberly Ellis, and Ben Tran for the warm welcome and the great time at Genentech.

Bedanken möchte ich mich auch beim IT-Team Jan Müller, Ralph Lange und Simon Woll für den ungebrochenen Einsatz auch zu unmenschlichsten Zeiten alle IT-Probleme schnellstmöglich zu lösen.

Vielen Dank an all meine Arbeitskollegen und den Doktoranden der ersten Generation der Arbeitsgruppe, welche mich während dieser Arbeit unterstützt und diese Zeit zu einer sehr wertvollen und lustigen gemacht haben. Vielen Dank für die tolle Zeit.

Besonderer Dank geht an Marion Krenz, Iris Perner-Nochta, Michael Wörner und Margret Meixner für ihren Einsatz uns Doktoranden von Papierarbeit und administrativen Aufgaben soweit wie möglich zu entlasten.

Vielen lieben Dank an meine Kollegen Lara Galm, Katharina Lang, Katharina Bauer, Marie-Therese Schermeyer, Carsten Radtke, Frank Hämmerling, Josefine Morgenstern und Heike Sigloch. Es war mir immer eine große Freude mit euch zu arbeiten.

Ich bedanke mich bei Lara Galm und Katharina Bauer für die tolle und sehr fruchtbare Zusammenarbeit.

Katha, vielen Dank für deine Unterstützung und die vielen unterhaltsamen Gespräche und Kaffeerunden.

Für die Unterstützung im Labor, den tollen Einsatz und die schöne Zeit bedanke ich mich bei Marc Hoffmann, Marie-Luise Schwab, Cathrin Dürr und Susanna Suhm.

Herzlich bedanken möchte ich mich bei Patrick Diederich. Patrick, vielen Dank für deine Unterstützung, den offenen Austausch und das stete Interesse. Deine Arbeitsweise und deine kritischen Fragen zur richtigen Zeit waren für mich sehr inspirierend.

Ganz besonders bedanken möchte ich mich bei meinen Bürokollegen Lara Galm und Katharina Lang für die einzigartige Atmosphäre, euren Humor und Selbstironie und für die fachliche wie auch persönliche Unterstützung in lustigen wie auch in niederschlagenden Zeiten. Ihr beide seid so grundoptimistische, verständnisvolle, emphatische und starke Menschen, die mir sehr viel Freude in die Arbeit gebracht haben und sehr viel mehr als Arbeitskollegen wurden.

Kathy, vielen Dank für deine sehr direkte Art, deine unerschöpfliche Geduld und die vielen unterhaltsamen nerdigen Gespräche.

Lara, ich bin wahnsinnig froh, dass ich mit dir einen so enthusiastischen, aufmerksamen und integren Menschen kennen lernen durfte.

Ganz herzlich bedanke ich mich bei meiner Familie für die unvergleichbare ungebrochene Unterstützung und Liebe, die ihr mir zuteilwerden lasst. Zu jeder Zeit konnte ich mich auf eure Unterstützung und euren Rückhalt blind verlassen. Ich möchte, dass ihr wisst, wie unschätzbar wertvoll eure Unterstützung für mich ist. Ohne euch wäre ich nicht der Mensch, der ich heute bin.

Ganz herzlich möchte ich meiner Tochter Jule danken, welche erst in der letzten Phase der Arbeit in unser Leben trat und mir eine unglaubliche Motivation zur Fertigstellung dieser Arbeit war.

Aus tiefstem Herzen bedanke ich mich bei meiner wundervollen Frau Felicitas für so unglaublich viel Verständnis, Geduld und Liebe. Ein großer Teil dieser Arbeit entstand an Wochenenden, Nächten oder anderen einer Partnerschaft unzuträglichen Zeiten. Du warst die wichtigste und stärkste Unterstützung, welche diese Arbeit ermöglichte. Worte können meine Dankbarkeit für alles was du machst und bist nicht ausdrücken.

The bad news is time flies. The good news is you're the pilot.

Michael Altshuler

Contents

CHAPTER 1

Abstract

The industrial use of biotechnology redefined the medical landscape by innovative and highly potent biopharmaceuticals such as antibodies or hormones. The complexity of biopharmaceuticals requires intensive protein characterization and a thorough study of production processes to ensure process efficiency, product quality, and most notably patient safety.

The present work focuses on the investigation of the interactions between proteins and aqueous solvents. In this scope two main emphases can be formulated: The development of appropriate methods for assessing protein characteristics and the investigation of process related issues and protein phase behavior and protein partitioning in aqueous two phase systems (ATPS) in particular. The protein-solvent interactions are described by protein characteristics and are predominating factors for the aggregation of protein, which is described as protein phase behavior, and for the partitioning in ATPSs. Both major emphases were studied by means of experimental high throughput techniques as well as by means of computational methods and molecular dynamics (MD) simulations in particular.

The behavior of proteins in solution is highly complex since the proteinogenic spatial structure is a dynamic network of substructures, composed of flexible and partly titratable residues and interacting with covalent and non-covalent bonds. The protonation of the residues is influenced in turn by physicochemical solvent properties and temperature. The knowledge of protein behavior, i.e. protein association to aggregates, is crucial to ensure safety and quality of biopharmaceuticals. The formation of aggregates, visible and sub-visible (<100 µm) particles can occur in every stage during the production process and its monitoring and characterization became a major challenge in drug development and manufacturing.

The protein-solvent interactions influence the stability of proteins to a large extent and are dominated by the protein hydrophobicity in case of aqueous solvents. However, the

characterization of protein hydrophobicity is highly challenging. A number of hydrophobicity scales were developed in the last decades allowing the assessment of protein hydrophobicity based on its primary sequence or on the amino acid composition of the protein surface in case of more advanced approaches. However, these theoretical approaches are not capable of considering protein dynamics, complexity, or external factors such as solvent chemistry and temperature. These major drawbacks were faced by the development of an *in-silico* approach based on MD simulations for the characterization of proteinogenic surfaces with respect to hydrophobicity. MD simulations serve as a highly effective methodology to investigate protein dynamics on an atomic level of detail and brick the gap between experimental observations and molecular dynamics. The developed approach exploits the varying affinity between polar water molecules and non-polar tracer molecules to hydrophilic and hydrophobic surface areas on proteinogenic surfaces. These varying affinities were determined by means of the modified radial distribution function (mRDF) and allowed to characterize the proteinogenic surface regarding hydrophobicity. The approach was developed by means of a set of eight custom designed peptides varying in the composition of nominally hydrophobic amino acids. The derived *in-silico* results were in excellent agreement with retention behavior of the peptides in reversed phase (RP) chromatography, which is one of the most established experimental approaches to assess peptide hydrophobicity. The retention behavior exposed a stronger correlation to the derived *in-silico* results than to one of the most applied hydrophobicity scales, namely the Kyte and Doolittle scale. The transfer of this MD simulation based approach to protein structures required an appropriate experimental hydrophobicity measure for the purpose of validation. The quantification of hydrophobicity of larger proteinogenic structures by means of RP chromatography is critical since the integrity of the protein structure and the protein hydration shell can be affected by the highly hydrophobic stationary phase and the organic solvent involved in RP chromatography. Alternative experimental approaches are based on hydrophobicity sensitive dyes, partitioning studies in organic two phase systems, or precipitation studies and bear similar integrity issues. To avoid an impairment of the protein conformational integrity, a non-invasive approach was developed which is based on the correlation of protein hydrophobicity and the effect of the protein on the surface tension. The protein hydrophobicity thereby highly influences the surface tension increment which means that an increase in protein hydrophobicity results in a more pronounced decreasing effect on the surface tension. In order to evaluate the surface tension in dependency of protein concentration, protein species, and pH, a stalagmometric device was developed which can determine surface tensions highly accurately (standard deviation < 0.1 ‰ for wa-

ter) and relatively fast (3.5 minutes per sample). The developed setup can be operated automatically by using industrial liquid handling stations. The surface tensions in dependency of protein concentration were analyzed to characterize bovine serum albumin (BSA), α-lactalbumin, lysozyme from chicken egg white, and human lysozyme at pH 3, 5, 7, and 9 with respect to hydrophobicity. All proteins exposed a pH dependency in terms of hydrophobicity. The results could be validated by absorption difference spectroscopy using bromophenol blue. This experimental method was used for the purpose of experimental validation to extend the MD simulations based approach for the characterization of peptides with regard to hydrophobicity by proteins. The surface hydrophobicity of α-lactalbumin, lysozyme from chicken egg white, and human lysozyme was studied *in-silico* at pH 5 and 7. As mentioned above, the modified radial concentration profiles (mRDF) were derived from extensive full atom MD simulations and were used to characterize surface hydrophobicity on an atomic level of detail at the respective pH. The protein surface could be characterized with regard to hydrophobicity and the influence of pH on the surface hydrophobicity could be pointed out. The computationally derived hydrophobic character of the proteins was found to be in good agreement with experimental results obtained by the developed non-invasive stalagmometric approach described above. Thus, the developed *in-silico* approach was shown to be capable of considering the influence of the physicochemical solvent properties on protein surface hydrophobicity. Thus, proteins were extensively studied in terms of protein-solvent interactions or protein hydrophobicity respectively by means of computational and experimental approaches. Peptide and protein hydrophobicity could be assessed on an atomic level of detail by means of the developed MD simulations based approach and a non-invasive experimental approach for hydrophobicity assessment based on surface tension measurements was established.
The developed approaches for characterizing proteins in solutions were further transferred to downstream processing applications, and protein aggregation and protein partitioning in aqueous two phase systems in particular. Protein aggregation is of major interest in every stage of drug development and production. Aggregation of proteins, which is generally described as protein phase behavior, can be a desired event in case of crystallization as a purification or formulation step or can occur as an undesired event in case of non-native aggregation which potentially influences drug activity and immunogenicity. Investigations focusing on the evaluation of protein aggregation are usually performed by cost, time, and material intensive experimental screenings. The experimental effort can be minimized by means of miniaturized, automated, and parallelized high throughput screenings (HTS) in combination with statistically optimized designs of experiment (DoE). In this thesis the protein phase behavior, experimentally deter-

mined by high throughput screenings, was studied by means of MD simulations and experimentally derived protein surface characteristics. The predictive power of protein characteristics, and surface zeta potential, hydrophobicity, and conformational flexibility in particular, for the protein phase behavior of α-lactalbumin, lysozyme from chicken egg white, and human lysozyme was investigated. The surface zeta potential measurement was determined by electrophoretic light scattering (ELS), hydrophobicity was assessed by means of the non-invasive stalagmometric approach mentioned above and conformational flexibility was studied by calculating the root-mean-square-fluctuation (RMSF) of the proteins' backbone atoms during extensive full atom MD simulations. The MD simulations were performed for the absence and presence of two salt species (sodium chloride and ammonium sulfate) in varying concentrations up to 1.3 mol/L. The combination of experimental investigations and MD simulations gave an unique insight into the complex interactions involved in protein phase behavior and might help every experimenter with the interpretation of experimental results. Based on the conformational flexibility, protein species could be distinguished according to their aggregation propensity. The impact of sodium chloride and ammonium sulfate on the conformational flexibility was in good agreement with the Hofmeister series, according to which ammonium sulfate has a much more pronounced destabilizing effect in comparison to sodium chloride. The protein specific aggregation propensity as a function of pH could not be correlated to solely conformational flexibility, and electrostatics and surface hydrophobicity had to be taken into account.

In addition, the protein behavior in more complex solutions, and the protein partitioning behavior in aqueous two phase systems in particular, was investigated. Aqueous two phase systems expose mixing gaps of two polymer solutions or a polymer and a salt solution. The two coexisting phases can be used for a very cost efficient protein extraction under mild conditions. Due to the excellent scalability of ATPSs, processes can be developed experimentally by high throughput screenings. However, there is still a lack of understanding of protein partitioning which is one of the main reasons for the industrial reluctance for these promising bioseparation operation. In order to brick this gap, a MD simulations based approach was developed, which serves for building predictive models for protein partitioning.

For an accurate determination of the phase diagram of investigated ATPSs an experimental method was developed which combines fully automated liquid handling stations and lab-on-a-chip technology. A so called *Tip2World* interface was developed which enables to supply liquid from a liquid handling station into standard capillary tubings and thus enables a sample supply to connected instruments. The *Tip2World* interface ex-

pands the portfolio of instruments integrable into liquid handling stations enormously. A micro Coriolis density sensor could thus be operated automatically for the determination of phase densities of ATPSs. These densities were used to evaluate tie lines of the ATPSs which redundantizes a number of simplifications typically done in high throughput ATPS screening. By means of this development phase compositions and phase densities of coexisting phases of the ATPSs composed of potassium phosphate and PEG with molecular weights of 300, 400, 600, and 1000 Da were derived. Starting from this experimental data set, MD simulations of selected ATPSs were performed to investigate the partitioning behavior of a selection of amino acids, peptides, and proteins. The solutes were simulated in coexisting phases separately and analyzed with respect to polar and non-polar interaction energies. By applying the linear interaction energy method, these interaction energies can be used to compute free energy differences and thus the partitioning behavior can be modeled. Thus, the derived interaction energies serve for model building in combination with experimentally determined protein partitioning properties.
In summary, both two main emphases of this work were investigated thoroughly and gave highly valuable insight into protein characteristics and protein behavior in solution. MD simulations were successfully used for peptide and protein assessment regarding solute-solvent interactions and were applied on process related scientific issues. Experimental setups were developed for the protein characterization and for physicochemical analysis of solvents regarding composition and density in case of aqueous two phase systems and surface tension which can be applied across research disciplines.

CHAPTER 2

Abstract in German - Zusammenfassung

Der industrielle Einsatz von Biotechnologie hat die medizinische Landschaft durch innovative und hoch wirksame Biopharmazeutika wie beispielsweise Antikörper oder Hormone neu definiert. Die Komplexität der Biopharmazeutika erfordert jedoch eine intensive Proteincharakterisierung sowie Untersuchungen des Herstellungsprozesses, um Prozesseffektivität, Produktqualität und vor allem die Sicherheit des Patienten zu gewährleisten. Die vorliegende Arbeit befasst sich mit der Untersuchung der Interaktionen zwischen Proteinen und wässrigen Lösemitteln. In diesem Zusammenhang können zwei Schwerpunkte der Arbeit formuliert werden: Die Entwicklung von Methoden zur Charakterisierung von Proteinen in Lösung und die Beschreibung von prozessrelevantem Proteinphasenverhalten und von Proteinverteilungsverhalten in wässrigen Zwei-Phasen-Systemen (aqueous two phase systems, ATPS). Beide Schwerpunkte wurden mittels experimenteller Hochdurchsatzverfahren sowie mittels *in-silico* Verfahren unter Verwendung molekuardynamischer (MD) Simulationen untersucht.

Das Proteinverhalten in Lösung ist höchst komplex, da die räumliche Proteinstruktur ein dynamisches Netzwerk aus Substrukturen ist, welche wiederum aus flexiblen und teilweise titrierbaren Aminosäuren bestehen und mittels kovalenter und nicht-kovalenter Bindungen wechselwirken. Die Protonierung der titrierbaren Aminosäuren ist wiederum von den physikochemischen Eigenschaften des Solvents und der Temperatur abhängig. Das Wissen über Proteinverhalten in Lösung und von Proteinaggregatbildung im Speziellen ist entscheidend, um die Sicherheit und Qualität des Biopharmazeutikums sicherzustellen. Die Ausbildung von Aggregaten, sichtbaren und nicht sichtbaren (<100 µm) Partikeln kann in praktisch jedem Prozessschritt auftreten und deren Überwachung und Charakterisierung wurde zu einer der großen Herausforderungen in der biopharmazeutischen Prozessentwicklung.

Die Proteinstabilität in Lösung hängt zu einem großen Teil von der Protein-Solvent-Interaktion ab und damit, im Falle von wässrigen Lösungen, von der Proteinhydropho-

bizität. Allerdings ist die messtechnische Erfassung der Proteinhydrophobizität äußerst schwierig. Eine Reihe von Hydrophobizitätsskalen wurde in den letzten Jahrzehnten entwickelt, welche die Berechnung von Proteinhydrophobizitäten auf Basis der Primärsequenz oder der Aminosäurenzusammensetzung der Proteinoberfläche ermöglichen. Allerdings ist keine dieser theoretischen Skalen in der Lage, der Proteindynamik, Komplexität und äußeren Einflussfaktoren Rechnung zu tragen. Daher wurde in dieser Arbeit ein *in-silico* Ansatz basierend auf MD Simulationen entwickelt, um proteinogene Oberflächen bezüglich der Hydrophobizität zu charakterisieren. MD Simulationen sind eine höchst effektive Methode, um die Proteindynamik auf atomarer Ebene zu untersuchen und damit die Lücke zwischen experimentellen Beobachtungen und molekulardynamischen Prozessen zu schließen. Der entwickelte Ansatz untersucht die unterschiedlichen Affinitäten von polaren Wassermolekülen und unpolaren Tracermolekülen zu hydrophilen bzw. hydrophoben Oberflächenbereichen proteinogener Strukturen. Die Affinitäten wurden mittels einer modifizierten radialen Verteilungsfunktion (modified radial distribution function, mRDF) ermittelt. Der Ansatz wurde auf eigens entworfene Peptide angewendet, die aus einer konstanten Sequenz polarer Aminosäuren und einer variablen nominell hydrophoben Sequenz bestanden. Die aus dem *in-silico* Ansatz ermittelten Werte standen in sehr guter Übereinstimmung mit dem Retentionsverhalten der Peptide in der Reversed Phase (RP) Chromatographie, welches eines der etabliertesten Verfahren zur Abschätzung der Hydrophobizität von Peptiden darstellt. Das Retentionsverhalten zeigte eine stärkere Korrelation mit den hier ermittelten MD Simulationsdaten als mit einer der meist verwendeten Hydrophobizitätsskalen nach Kyte und Doolittle.
Der Transfer dieser auf MD Simulationen basierenden Methode von Peptiden auf Proteine erforderte zu Validierungszwecken ein geeignetes experimentelles Verfahren zur Messung der Proteinhydrophobizität. Die Hydrophobizitäts-Quantifizierung größerer proteinogener Strukturen mittels RP Chromatographie ist kritisch zu betrachten, da die Integrität der Proteinstruktur durch die Wechselwirkungen mit der sehr hydrophoben stationären Phase sowie durch den Einsatz organischer Lösemittel beeinträchtigt werden kann. Alternative experimentelle Ansätze basieren auf speziellen hydrophobizitätssensitiven Farbstoffen, Verteilungsstudien innerhalb organischer Zwei-Phasen-Systemen oder Präzipitationsstudien und weisen die gleiche Integritätsproblematik auf. Um diese Integritätsproblematik zu beheben wurde in der vorliegenden Arbeit ein nicht-invasiver Ansatz entwickelt, welcher auf der Korrelation zwischen Proteinhydrophobizität und dem Einfluss des Proteins auf die Oberflächenspannung der Proteinlösung beruht. Die Proteinhydrophobizität beeinflusst dabei in hohem Maße das Oberflächenspannungsinkrement, was bedeutet, dass je hydrophober das Protein ist, desto stärker wird die

Oberflächenspannung durch das Protein verringert. Um die Oberflächenspannung in Abhängigkeit von Proteinkonzentration, Proteinspezies und pH Wert zu ermitteln, wurde ein stalagmometrischer Aufbau entwickelt, mit dem Oberflächenspannungen sehr genau (Standardabweichung $< 0{,}1$ ‰ für Wasser) und relativ schnell (3,5 Minuten pro Probe) ermittelt werden konnten. Der entwickelte Aufbau kann mittels voll automatisierter industrieller Pipettierstationen betrieben werden. Die Oberflächenspannung in Abhängigkeit von der Proteinkonzentration wurde zur Charakterisierung der Hydrophobizität von bovinem Serumalbumin (BSA), α-Lactalbumin, Lysozym aus Hühnereiklar und humanem Lysozym bei pH 3, 5, 7 und 9 verwendet. Alle Proteine wiesen eine pH-Abhängigkeit ihrer Hydrophobizität auf. Die Ergebnisse konnten mittels Absorptionsdifferenzspektroskopie unter Verwendung von Bromophenolblau bestätigt werden. Dieser nicht-invasive Ansatz ermöglicht die Ermittlung der Proteinhydrophobizität direkt im gewünschten Solvent und verbessert die Übertragbarkeit von Hydrophobizitätsmessungen. Diese experimentelle Methode wurde eingesetzt zur experimentellen Validierung des MD Simulations basierten Ansatzes zur Charakterisierung von proteinogenen Oberflächen bezüglich Hydrophobizität, erweitert auf Proteine. Die Oberflächenhydrophobizität von α-Lactalbumin, Lysozym aus Hühnereiklar und humanem Lysozym wurde bei pH 5 und 7 untersucht. Die Proteinoberflächen konnten auf atomarer Ebene charakterisiert und pH-Einflüsse deutlich gemacht werden. Die *in-silico* ermittelte Proteinhydrophobizität konnte experimentell mittels dem bereits beschriebenen nichtinvasiven Verfahren bestätigt werden. Damit konnte gezeigt werden, dass der entwickelte *in-silico* Ansatz in der Lage ist, den Einfluss der physikochemischen Solventeigenschaften zu berücksichtigen. Somit wurden die Protein-Solvent Interaktionen, und die Poteinhydrophobizität im Speziellen, ausführlich mittels eigens entwickelter *in-silio* und experimenteller Verfahren untersucht. Die entwickelten Ansätze zur Proteincharakterisierung wurden im Weiteren auf Aufarbeitungsfragestellungen und im Speziellen auf das Proteinphasenverhalten und die Proteinverteilung in wässrigen Zwei-Phasen-Systemen angewendet. Proteinphasenverhalten ist für praktisch jeden Schritt der Wirkstoffentwicklung und -herstellung von großem Interesse. Proteinaggregation, welche im Allgemeinen als Proteinphasenverhalten beschrieben wird, kann im Falle von Kristallation als Aufreinigungs- und Formulierungsschritt angewendet werden oder im Falle von nicht nativer Aggregation ein unerwünschtes Proteinverhalten sein, dass die Wirkstoffaktivität und -immunogenität deutlich beeinflussen kann. Untersuchungen zum Proteinphasenverhalten werden meist mittels kosten-, zeit- und materialintensiver experimenteller Screenings durchgeführt. Der experimentelle Aufwand kann durch miniaturisierte, automatisierte und parallelisierte Hochdurchsatzverfahren (High Throughput Screenings,

HTS) in Kombination mit statistisch optimierter Versuchsplanung (design of experiment, DoE) reduziert werden.
Im Rahmen dieser Arbeit wurde das Proteinphasenverhalten, welches experimentell mittels HTS evaluiert wurde, mithilfe von MD Simulationen und experimentell bestimmten Proteineigenschaften untersucht. Die Vorhersagekraft von Proteineigenschaften, wie Zetapotential, Hydrophobizität und konformative Flexibilität, auf das Phasenverhalten von α-Lactalbumin, Lysozym aus Hühnereiklar und humanem Lysozym wurde untersucht. Das Zetapotential wurde mittels elektrophoretischer Lichstreuung (electrophoretic light scattering, ELS) gemessen, die Hydrophobizität wurde über den oben beschriebenen experimentellen nicht-invasiven Ansatz bestimmt und die konformative Flexibilität über das Quadratmittel der Fluktuation (root-mean-square-fluctuation, RMSF) des Proteinrückgrats während umfangreicher MD Simulationen. Die MD Simulationen wurden ohne Salz und in Anwesenheit von zwei Salzspezies (Natriumchlorid und Ammoniumsulfat) in einem Konzentrationsbereich von bis zu 1,3 mol/L durchgeführt. Die Kombination aus experimentellen Verfahren und MD Simulationen ermöglichte einen einzigartigen Einblick in die komplexen Interaktionen, welche ins Proteinphasenverhalten involviert sind und können jedem Experimentator bei der Interpretation experimenteller Ergebnisse helfen. Anhand der Proteinflexibilität konnten die Proteine bezüglich ihrer Aggregationsneigung unterschieden werden. Der Einfluss von Sodiumchlorid und Ammoniumsulfat auf die Proteinflexibilität folgte der Hofmeister Serie, wonach Ammoniumsulfat einen stärker ausgeprägten destabilisierenden Effekt aufweist als Sodiumchlorid. Die proteinspezifische Aggregationsneigung als Funktion des pH Wertes konnte nicht alleine mit der Flexibilität beschrieben werden und die Elektrostatik sowie die Hydrophobizität mussten mit einbezogen werden.
Zusätzlich wurde das Proteinverhalten in komplexeren Lösungen und im Speziellen das Verteilungsverhalten von Proteinen in wässrigen Zwei-Phasen-Systemen untersucht. Wässrige Zwei-Phasen-Systeme weisen eine Mischungslücke zwischen zwei Polymerlösungen oder zwischen einer Polymer- und einer Salzlösung auf. Die koexistenten Phasen können zur kostengünstigen Proteinextraktion unter milden Bedingungen verwendet werden. Aufgrund der exzellenten Skalierbarkeit dieser Systeme können Prozesse experimentell mittels miniaturisierter Hochdurchsatz-Screenings entwickelt werden. Die zugrunde liegenden Mechanismen der Proteinverteilung sind allerdings noch nicht ausreichend verstanden, was einer der Gründe ist, warum wässrige Zwei-Phasen-Systeme in der biopharmazeutischen Industrie kaum Anwendung finden. Um ein besseres Verständnis des Proteinverteilungsverhaltens zu generieren, wurde ein *in-silico* Ansatz basierend auf MD Simulationen entwickelt, welcher zur Bildung prädiktiver Modelle dient. Zur Erfas-

sung der Zusammensetzung und Dichten der coexistenten Phasen von wässrigen Zwei-Phasen-Systemen wurden voll automatisierte Liquid Handling Stations und lab-on-a-chip Technologie kombiniert. Ein sogenanntes *Tip2World* Interface wurde entwickelt, welches es ermöglicht Flüssigkeiten mittels Liquid Handling Stations in Standardkapillaren zu injizieren und damit die Proben verbundenen Instrumenten zuzuführen. Das *Tip2World* Interface erweitert damit das Portfolio an Instrumenten, welche in eine Liquid Handling Station integriert werden können, enorm. Ein Micro-Coriolis Dichtesensor konnte damit vollautomatisiert zur Dichtebestimmung der zu untersuchenden wässrigen Zwei-Phasen-Systeme betrieben werden. Die Dichten wurden zur Ermittlung der Konoden eingesetzt, deren Bestimmung im Hochdurchsatz bisher mit Vereinfachungen verbunden war, die damit hinfällig werden. Das Verhalten einer Auswahl an Aminosäuren, Peptiden und Proteinen in wässrigen Zwei-Phasen-Systemen bestehend aus Kaliumphosphat und PEG mit Molekulargewichten von 300, 400, 600 und 1000 Da wurden *in-silico* untersucht. Eine Datenbank von Interaktionsenergien wurde mittels MD Simulationen generiert, welche zur Bildung von prädiktiven Modellen in Kombination mit experimentell ermitteltem Verteilungsverhalten dient.
Zusammenfassend wurden beide Schwerpunkte der Arbeit ausführlich untersucht und es konnten sehr wertvolle Einblicke in Proteincharakteristika und in das Proteinverhalten in wässrigen Lösungen gewonnen werden. Molekulardynamische Simulationen wurden erfolgreich zur Charakterisierung der Solvent-Interaktionen mit Peptiden und Proteinen eingesetzt und auf prozessrelevante Fragestellungen angewendet. Experimentelle Verfahren wurden zur Proteincharakterisierung und zur physikochemischen Analyse von Lösungen bezüglich Zusammensetzung und Dichte im Falle der wässrigen Zwei-Phasen-Systeme und bezüglich der Oberflächenspannung entwickelt, welche in unterschiedlichsten Forschungsdisziplinen Anwendung finden können.

CHAPTER 3

Introduction

The ground breaking decipheration of the DNA molecule by Watson and Crick in 1953 [1] paved the way for biotechnology. Only twenty years later in 1972 a research team organized by Stanford biochemist Paul Berg created the first recombinant DNA molecule by splicing together two separate DNA sequences but did not place the recombinant molecule into a cell for potential protein expression. This step was done in the following year by Stanford medical professor Stanley Cohen and biochemist Herbert Boyer from the University of California (San Francisco, CA, USA) introducing a technique for manipulating the expression of genes and the manufacture of gene products in microbes [2]. They were able to insert a recombinant ringlet of DNA, called a *plasmid*, containing functional genes of cells of Xenopus laevis into E. coli bacteria and thus created the world's first recombinant organism. This innovative methodology promised the ability to direct genetically modified organisms to synthesize desired proteins in large amounts. While Cohen stayed in academia, Boyer and venture-capitalist Robert Swanson set up Genentech, the world's first biotechnology company in 1976 and launched the very first biopharmaceutical product, namely insulin, which was licensed to *Eli Lilly and Company* in 1982. During the next decades, the biopharmaceutical industry experienced a remarkable growth and untold numbers of biopharmaceutical products for diagnostics or therapeutic purpose have been tested and hundreds have been approved for marketing. Modern medicine without these products is inconceivable. Biopharmaceuticals range from very small molecules such as blood factors, growth hormones, or insulin to highly complex and high molecular weight molecules such as monoclonal antibodies (mAb), interferon, or vaccines. Monoclonal antibodies and antibody-based therapeutic agents such as bispecific antibodies (bsFab), heavy chain only antibodies (HcAb), bispecific Fab fragments, antibody-drug-conjugates (ADC) or single-domain antibodies (sdAb) are seen as high potential candidates for the treatment of several severe diseases including autoimmune, cardiovascular and infectious diseases, cancer and inflammation [3–5]. A number

of biopharmaceuticals require post-translational modification such as glycosylation or the formation of disulfid bonds, which restricts the expression system. The most popular organisms for producing recombinant glycosylated proteins are mammalian cells [6]. Starting from a crude fermentation broth containing a relatively low concentration of the molecule of interest (MOI), a number of purification steps are necessary to purify the MOI from process related impurities such as host cell proteins, cell debris, buffer and medium ingredients or potential viral contaminants as well as protein specific impurities such as aggregates, cleaved proteins, or isoforms. The general purification process can typically be structured into several bioseparation steps, namely a capture step, purification steps, and polishing steps, which increase in separation resolution but decrease in throughput in the given order. The final purified drug substance is further formulated into the drug product. Chromatography is one of the most important industrial bioseparation techniques exploiting the heterogenity of the proteins in terms of physicochemical properties such as differences in size (size exclusion chromatography, SEC), hydrophobicity (reversed phase chromatography, RP, hydrophobic interaction chromatography, HIC), charge and charge distribution (ion exchange chromatography, IEC), or affinity for example immobilized metal ion affinity chromatography, IMAC, or immobilized molecule specific ligands (antigen, protein A, protein G, protein L)). physicochemical properties of proteins in general are described in detail in section 3.1.

To increase cost efficiency and flexibility of manufacturing, biotechnology companies strive for integrated continuous manufacturing processes comprising upstream and continuous capture and purification steps. Continuous manufacturing processes are already considered in a recently published draft guidance from the FDA [7] and according to PricewaterhouseCoopers (PwC) [8] most pharmaceuticals will be manufactured continuously by 2020. Continuous approaches benefit from a reduction of the equipment footprint and elimination of several non-value-added unit operations, such as clarification and intermediate hold steps [9–11]. A continuous bioseparation using conventional chromatography columns is possible using simulated moving bed (SMB) technology [9] and is widely used in food and pharmaceutical industries. Despite all observed challenges regarding complexity and robustness [12], a subsidiary of Sanofi, and Genzyme Corporation in particular, could file a patent application for a platform technology that will integrate upstream and downstream processing in a fully continuous end-to-end biomanufacturing in 2014 [13]. A separation technique predestinated for continuous bioprocessing and capable to overcome complexity and robustness issues is the use of aqueous two phase systems (ATPSs), also known as aqueous biphasic systems (ABSs) or aqueous two phase extraction (ATPE), for protein extraction. Here, protein specific phase affinities are ex-

ploited for bioseparation. These extraction systems are described in section 3.2.
The development of a bioseparation process in industry is driven by time-to-market and quality by design (QbD) requirements. Speeding up the process development and simultaneously increasing the knowledge about the process is technically realized by high throughput approaches which are guided by automation, miniaturization, and an enormous parallelization, allowing to gain an increased amount of data within a sufficient range of time. In combination with a statistically optimized experimental design (DoE), sophisticated analytics, and advanced data processing, high throughput screening approaches serve as powerful tools for process development. In addition, microfluidics and lab-on-a-chip technologies mature and offer unique opportunities to combine multiple analysis and operation steps on a highly miniaturized spatial format. These unique properties attracted the interest of the biopharmaceutical research and development. A lab-on-a-chip technology was used within this work to derive experimental data of aqueous two phase systems and was implemented into the experimental work-flow by the employment by robotic liquid handlers. The operability of the device was enabled by means of custom-made constructions, which were manufactured by 3D printing and were applied for different purposes in this work. 3D printing technology and open-source designs emerged as powerful tools and have significantly affected the experimental work of this thesis. A basic insight into 3D printing technology is given in section 3.4.
Beside experimental strategies, *in-silico* approaches have matured with growing computational power over the last decades. For instance, structural bioinformatics are applied for protein structure prediction and the prediction of protein–protein interactions in the *Rosetta@home* project [14]. Quantitative structure–activity relationship (QSAR/QSPR) models are generated to conclude from protein biochemical properties such as electrostatic, geometric, or steric properties to a response variable of interest, for example biological activity. One highly valuable *in-silico* methodology allowing to gain insight into the dynamics of biological systems is the use of molecular dynamics simulations. This methodology is focused in detail in section 3.3.

3.1 Conformational and Biochemical Properties of Proteins

This section focuses on fundamental conformational and biochemical properties of peptides and proteins, starting from the elementary levels of protein structure through biochemical properties dominating the behavior of proteinogenic molecules in solution.

3.1.1 From Primary Sequence to the Complex Spatial Structure of Proteins

Proteins are biomolecules consisting of one or more chains of amino acid residues called polypeptides. Commonly, short chains of amino acid residues (20-30) are called peptides. The amino acids within polypeptide are covalently linked by peptide bonds originating from the condensation reaction illustrated in Figure 3.1. One amino acid loses a hydrogen and oxygen from its carboxyl group (COOH) and the other one loses a hydrogen from its amino group (NH_2) with the accomplished production of one water molecule and the creation of the peptide bond [15].

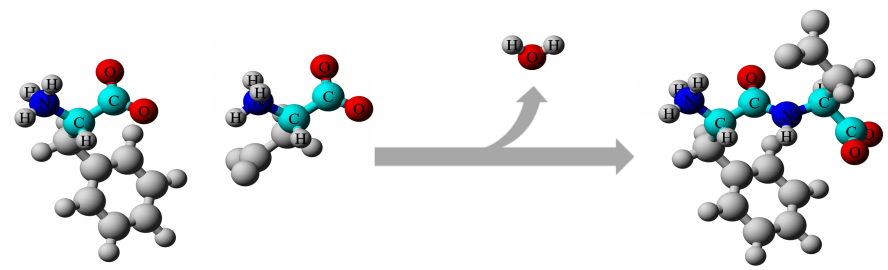


Figure 3.1: Schematic illustration of the condensation of a peptide bond between the carboxyl group of one amino acid (phenylalanine) and the amino group of the adjacent amino acid (aspartic acid). Side chains are colored in gray, backbone atoms are colored and labeled according to the element. Molecular graphic was created with YASARA (www.yasara.org) and POVRay (www.povray.org).

Overall, there are 21 proteinogenic α-amino acids which can be found in eukaryotes varying in polarity, charge, aromatic or aliphatic structure, and in the ability to undergo special interactions such as disulfide bridging in case of cystein. The sequence of the amino acids in the polypeptide are coded on DNA or RNA molecules respectively and thus can be produced by recombinant gene techniques. The sequence of a protein's amino acids is defined as the primary sequence. This polypeptide forms three dimensional secondary structures, for example randomized coils, turns or highly ordered, hydrogen bond induced structures and α-helices, 3_{10}-helices, and β-sheets in particular as illustrated in Figure 3.2. The geometric shape containing one or more secondary structures as mentioned above is called tertiary structure and is called protein. The quaternary structure of a protein refers to the three dimensional shape of an arrangement of multiple tertiary structures, which is stabilized by covalent bonds as well as non covalent bonds such as disulfide bridges, hydrogen bonds, or salt bridges.

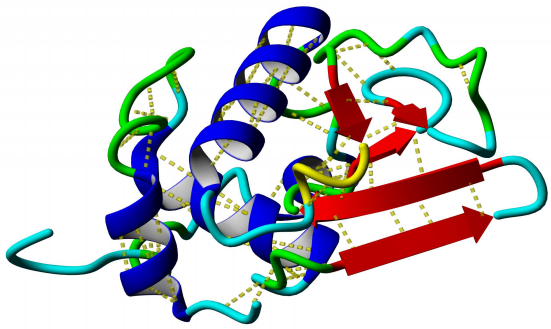


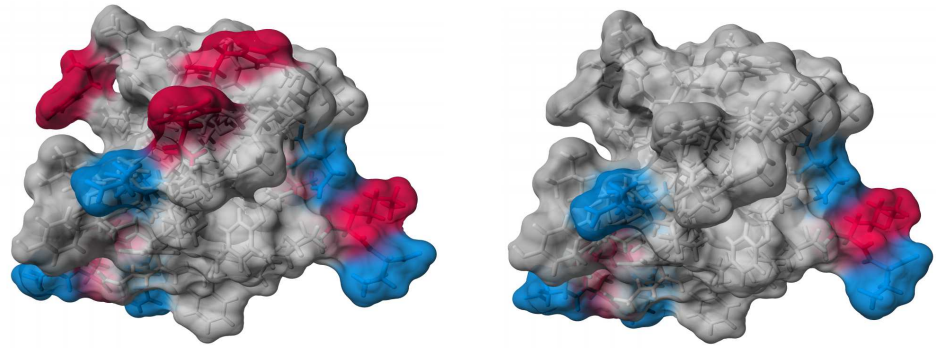
Figure 3.2: Secondary structures illustrated by means of apo-bovine alpha-lactalbumin (PDB: 1F6R [16]). α-helices are colored in blue, β-sheets are colored in red, turns are colored in green, 3_{10}-helix is colored in yellow, and coil structures are colored in cyan. H-bonds are illustrated as dotted yellow lines. Molecular graphic was created with YASARA (www.yasara.org) and POVRay (www.povray.org).

The most popular technique for structure assessment in a high resolution is the use of X-ray crystallography, which determines the three-dimensional density distribution of electrons in crystallized proteins. In addition, nuclear magnetic resonance (NMR) techniques are applied for structure analysis but are challenging and time consuming for large protein complexes. However, in contrast to X-ray crystallography NMR offers the possibility to gain information on protein dynamical properties, even it yields in less precise structures. Spatial structural data of proteins and nucleic acids are available in a freely accessible repository called Protein Data Bank (PDB) and are described in a PDB text file format including atomic coordinates, sidechain rotamers, secondary structure assignments, as well as atomic connectivity. Based on this huge amount of structural data available by PDB a number of predictive bioinformatic approaches for protein structure were developed. The most established method is the homology modeling using experimentally derived structures of related family members as templates to derive the spatial structure for a primary sequence of interest [17–19].

3.1.2 Biochemical Properties of Proteins

An intrinsic property of proteins consisting of a chain of amino acids is the pH depending charge. Apart from the terminal carboxyl (COOH) and amino group (NH_2), there are titratable amino acid residues. Basic and acidic amino acids are protonated according to their pKa values and the pH value of the solvent. The pKa values are influenced mainly

by inductive, electrostatic, and electron-delocalization (mesomeric) effects, together with contributions from hydrogen bonding, conformational differences, and steric factors [20]. Thus the ratio, number, and distribution of basic and acidic amino acids as well as the three-dimensional structure is crucial for protein polarity and thus the interactions with the solvent, which dominate the protein stability in turn. The influence of a pH shift on the charge distribution and the protonation is illustrated in Figure 3.3 by human insulin, where histidine and a N-terminal phenylalanine are deprotonated by shifting the pH from 5 to 8.



(a) (b)

Figure 3.3: Comparison of charge distribution of human insulin (PDB ID: 3W7Y) at (a) pH 5 and (b) pH 8. Positive charges are colored in red, negative charges are colored in blue. Molecular graphics were created with YASARA (www.yasara.org) and POVRay (www.povray.org).

Apart from protein net charge and the charge distribution on the protein surface, the exposed hydrophobic amino acids influence the protein-solvent interactions. In a polar solvent, the protein exposes charged groups to its surface, whereas uncharged amino acids are often found in the core of a globular protein in order to reduce the contact area to the polar solvent. After decades of research protein hydrophobicity and hydrophobic interactions are not fully understood. A number of hydrophobicity scales have been developed as a measure for protein hydrophobicity based on the protein's primary sequence or amino acid composition respectively. However, there is no consistent hydrophobicity scale used in literature, which is illustrated in Figure 3.4 by color coding the surface of human insulin according to widely used hydropbobicity scales developed by Kyte & Doolittle [21], Sweet & Eisenberg [22], and Abraham & Leo [23]. Moreover, the hydrophobicity scales are not able to consider pH induced charge variations on the protein surface.

The change of electrostatic properties by inducing or elimination of point charges within residues capable for protonation affects its polarity and thus the solvent-interactions.

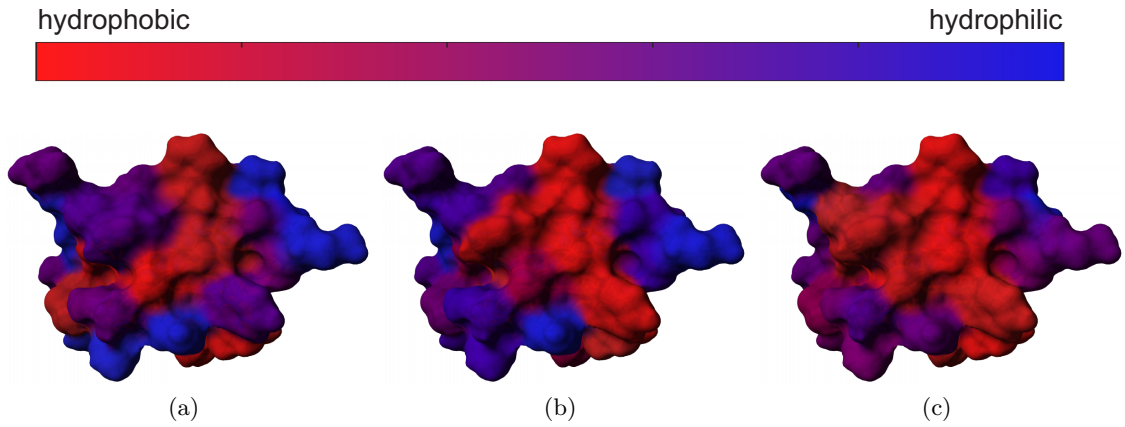


Figure 3.4: Comparison of hydrophobicity scales by color coding the molecular surface of human insulin (PDB ID: 3W7Y). The color codes are normalized to the full value range of the respective hydrophobicity scales by (a) Kyte & Doolittle [21], (b) Sweet & Eisenberg [22], and (c) Abraham & Leo [23]. Molecular graphics were created with YASARA (www.yasara.org) and POVRay (www.povray.org).

Apart from the *in-silico* approaches for estimating protein hydrophobicity based on scales, experimental methods were applied, which are based on the correlation between hydrophobicity and retention in reversed phase (RP) or hydrophobic interaction chromatography (HIC) or on the solubility in presence of precipitants such as salt or polymers. In addition, special dyes like 8-anilinonaphthalene-1-sulfonic acid (ANS) [24], cis-parinaric acid (CPA), 6-propionyl-2-(N,N-dimethylamino)naphthalene (PRODAN), or bromophenol blue [25] offer a spectrometric approach to assess hydrophobicity of the protein surface. The use of organic solvents or high salt concentrations in case of RP chromatography or HIC as well as the adsorption process to the functionalized stationary phase implies a modification in the protein hydration shell and thus is considered as insufficient for the assessment of protein surface regarding hydrophobicity in its native form in the solution of interest. The spectrometrically derived hydrophobicity values were shown to be of limited quality since the chemical nature of the dye such as aromatic or aliphatic substructures and charge is highly affecting these measurements [26]. Furthermore, these methods are invasive and thus are potentially harmful for protein conformational and biochemical integrity.

The combination of hydrophobic interactions as well as electrostatics plays a major role in protein stability in solution and for protein aggregation. Protein aggregation is a highly complex and not fully understood phenomenon which challenges the development and production of biopharmaceuticals. The reversible formation of multimeric associations of proteins in their native forms into highly structured crystals or less structured precipitate, known as native aggregation, is an appreciated bioseparation and formulation step in biopharmaceutical industry. Thereby, the molecule of interest can be selectively crystallized or precipitated or contaminants can be removed from the solution by induced aggregation. Since predictive approaches for protein aggregation are missing, the development of precipitation or crystallization steps is performed heuristically. A multimeric association of proteins in non-native states, known as non-native aggregation, typically implies at least a partial unfolding of the protein structure prior to aggregation and can be an irreversible process. Such irreversible associations of the MOI can occur throughout the production process and result in a loss of product activity and potential product immunogenicity. Thus, a thorough control and monitoring of protein aggregation is crucial to warrant product quality and to ensure patient safety.

3.2 Aqueous Two Phase Systems

An aqueous two phase system (ATPS) is formed by mixing two polymers such as the commonly used polyethylene glycol (PEG) and dextran or a polymer and a salt like PEG in combination with ammonium sulfate or potassium phosphate. It can be used as a mild, scalable, economic, and simple in terms of instrumental complexity, extraction system especially for biomolecules. More sophisticated ATPS forming smart polymers were developed which change their conformation as a response to a change in the environment (e.g. pH or temperature) [27]. These smart polymers allow to further control phase separation and thus optimize the recovery of the target molecule and the polymer in the opposite phase. Other smart polymers expose magnetic properties and thus enable to support phase separation and enhance process throughput [28].
Aqueous two phase systems are among the leading so-called integrated processes since they combine several tasks into one unit operation, and concentration, biomass removal, and product purification in particular. Additionally, the combination of ATPSs with cell disruption [29] and the combination of production and product recovery known as extractive bioconversion [30, 31] were reported.
The first extraction of biomaterial was observed by Per-Åke Albertsson in 1995, when

he mixed a potassium phosphate solution and PEG for cleaning chromatography material from adsorbed chloroplast particles. Albertsson could demonstrate the selectivity of the potassium phosphate/PEG system by means of partitioning of microorganisms, cell walls, chloroplasts, and chloroplast vesicles [32]. Further on, Albertsson studied driving forces of partitioning and applied ATPSs on a number of molecule classes such as proteins and viruses, isolated ribosomes from rat brain microsomes, concentrated and purified viruses, or single-stranded and double-stranded DNA [32–38]. Today, the aqueous two phase extraction has been established as a cost efficient and highly potent downstream processing method in biotechnology, even though there is still a reluctance from industry towards these systems and the commercial use of ATPSs is still limited, even it has been proven for commercial purification of human insulin–like growth factor I (IGF–I) [39].

Aqueous two phase systems are more suitable for bioseparation than conventional water-organic solvent systems since both phases contain mainly water and thus provide gentle conditions for labile biomaterial. Moreover, the interfacial tension of the disperse phase of aqueous to phase systems is very low compared to water-organic solvent systems. This property is highly beneficial in terms of the mass transfer due to an increased interfacial contact area and in terms of recovery of biomaterial due to reduces interfacial stress. These systems have excellent prospects for scale-up [40–42] which allows an efficient optimization of ATPS in a bench-scale and the easy transfer on pilot and production scale with performance data retained [41]. In addition, ATPSs are operateable continuously using traditional liquid–liquid extractors or more sophisticated instruments such as stagewise, differential, or centrifugal contactors [43], which could overcome some of the drawbacks that conventional chromatographic separations are facing. Recycling strategies for salt and polymer rich phases have been described [44, 45].

3.2.1 The Phase Diagram

The phase diagram of an ATPS, as illustrated in Figure 3.5, characterizes the physical properties in terms of mixing gap and composition and phase ratio of formed phases. The mixing gap and the critical concentration of two polymers, or polymer and salt respectively, forming two phases are defined by the binodal or binodal curve. A system S within the two phase region above the binodal curve separates into a top phase T and a bottom base B according to the respective tie line. The ratio of the tie line segments $\overline{ST}$ and $\overline{SB}$ corresponds to the mass ratio of the phases. This implies that systems lying

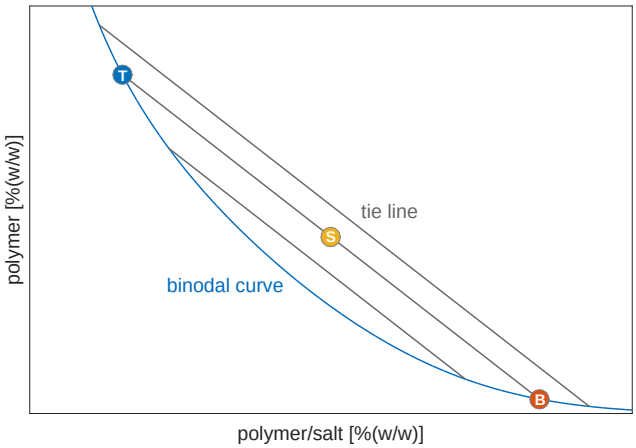


Figure 3.5: Schematic illustration of a phase diagram of an aqueous two phase system containing the binodal curve, tie lines, and an exemplary system point S and the corresponding top phase T and bottom phase B.

on the same tie line form two phases, which are identical in composition and vary only in mass ratio according to the lever arm rule as described in 3.1.

$$\frac{m_T}{m_B} = \frac{V_T \rho_T}{V_B \rho_B} = \frac{\overline{SB}}{\overline{ST}} \tag{3.1}$$

V, m, and ρ are the volume, mass and density of top (T) and bottom (B) phase and $\overline{SB}$ and $\overline{ST}$ are the segments of tie lines extracted from the phase diagram. The length of the complete tie line $\overline{TB}$ reflects the difference in composition of top and bottom phase and is widely used to express the effect of system composition on partitioning [46].
The phase diagram as described in Figure 3.5 serves as a basis for selecting the operational point. The operational point can be selected to reach extreme phase volume ratios to combine a concentration step accomplished with bioseparation but should also consider operational robustness. For instance, small variations of system point composition on very short tie lines result in dramatic changes in phase volume ratio. This implies that scaling of aqueous two phase systems should take preparation accuracy into account in order to derive a robust purification step.

3.2.2 Protein Partitioning

The partitioning of proteins between the coexisting phases underlies a complex balance of a matrix of interactions between system components, and proteins, polymers, salts, and water molecules in particular, and is described by the partitioning coefficient k_i for the partitioning solute i according to equation 3.2

$$k_i = c_{i,TP}/c_{i,BP}, \tag{3.2}$$

where c is the concentration within the top phase TP or the bottom phase BP respectively. The partitioning is dominated by the solute's phase affinities for low concentrations and is referred as true partitioning in the following. This implies a constant partitioning coefficient for low solute concentrations of systems along the tie line, since the composition of the coexisting phases is identical. At increased process related concentrations precipitation at the interface and in suspension can occur and the partitioning coefficient k_i highly depends on the higher levels of solubility of the solute in each of the phases and is referred as apparent partitioning in the following. Since the ratio of phase volumes vary along the tie line in accordance to the lever arm rule (equation 3.1), the apparent partitioning coefficient k_i is highly dependent on the position along the tie line. Thus, the significance of the partitioning coefficient as a measure of solute specific phase affinities is highly dependent on the studies solute concentration and solute stability.
The partitioning of proteins in ATPSs is affected by system and protein properties and can be dominated individually by one or a combination of a number of characteristics and hydrophobicity, electrostatics, molecular size, solvent accessibility, or biospecific affinities in particular [47–50]. From a process perspective, this properties can be influenced in turn by the system composition (eg. salt and polymer concentration), the chosen molecular weight of the polymer, pH value, and the use of additives such as NaCl, which increase the system capability to separate solutes according to their hydrophobicities. The effects can vary strongly dependent on the chosen system and used compounds and are shortly discussed by means of polymer-salt induced ATPSs used within this work in the following. A profound review is provided by Asenjo et al. [51].
In polymer-salt composed ATPSs both phases are generally hydrophilic, whereas the polymer rich top phase is relatively more hydrophobic compared to the salt rich bottom phase. Due to hydrophobicity driven partitioning, this phase property favors the partition of hydrophobic solutes towards the top phase. The hydrophobicity of the phases is directly related to the concentration and chemical nature of the system compounds. For

example, the hydrophobicity of polyethylene glycol correlates to its molecular weight [52] and thus the hydrophobic resolution can be manipulated by the used polymer molecular weight. For process relevant increased protein concentration, the precipitating effect of the phase components further affects the apparent partitioning. Based on the hydrophobic solute property, different attempts were made to correlate protein partitioning to its hydrophobicity, which was assessed experimentally by HIC, RP chromatography or precipitation or theoretical derived measures based on amino acis composition or 3D structure [50, 53, 54]. As described in section 3.1.2, these measurements are associated with simplifications and cannot consider the protein complexity and dynamic to a large extent, but were shown to result in sufficient models within the respective batch of systems. In PEG-based ATPS hydrophobic interaction were shown to be one of the main driving forces within the partitioning process [48–50].

Apart from hydrophobic interactions, the electrostatic interactions can affect the partitioning. Since the protonation state of amino acids and the electrostatics on the protein surface in turn is a function of the pH, temperature, and ionic strength, these physicochemical system properties are important parameters to manipulate protein surface properties and its partitioning [55]. The addition of electrolytes such as NaCl can further interfere the complex electrostatic interactions. Due to electrical neutrality of each phase of equilibrated ATPSs, the distribution of any charged group, such as the charged protein, is affected by the movement of any ion from one phase to the other. To maintain electroneutrality in both phases, a movement of an ion into the other phase includes the transfer of a sufficient number of counter-ions or counter-transfer of co-ions [56].

As the partitioning solute has a distinctive geometrical shape and molecular weight, steric effects can contribute to the partitioning mechanism. For the solvation of the analyte in the respective phase liquid, a cavity for the solute has to be created. The creation of this cavity results in additional forces and the interfacial tension of the solvent against the protein surface as well as the osmotic pressure originated from the concentration gradient. Based on these considerations, the solute partitioning of a solute is additionally affected by its shape and geometry.

The correlations were extensively studied and gave a highly valuable insight into the complexity and the dominating driving forces of the underlying partitioning mechanisms [48–50, 53, 54]. The studied systems could be sufficiently modeled but a mechanistic understanding on a molecular level of detail is still missing. The investigation of the interaction complexity contributing the true solute partitioning requires a thorough investigation on molecular interactions of the phase components and the solute and needs to take molecular dynamics into account. Molecular dynamics (MD) simulations,

described in section 3.3, are applied for this purpose within this work as highly promising *in-silico* methodology for the study of molecular systems on a atomic level of detail.

3.3 Molecular Dynamics Simulation

Computer-assisted virtual screening attracts an enormous interest in pharmaceutical research and majorly aims at the prediction of binding strengths between ligands and receptors or complex structures of new ligands using available data or empirically optimized models. These *in-silico* approaches assist drug design and lead optimization without synthesizing novel compounds and there are already a number of computationally designed drugs on the market [57, 58]. The efforts of computational drug designers can be roughly grouped into three different methodologies, which are shortly described in order of increasing computational effort, required structural input, and predictive power in the following [59]. Quantitative structure activity relationship (QSAR) aims at the prediction of experimental observables such as binding affinities from simple structural, chemical, and physical properties of molecules, such as polarity, octanol-water partition coefficient, surface and geometrical properties, or hydrophobicity. Sophisticated statistical regression techniques are applied to explore correlations between the molecular descriptors and the observables. QSAR models are well established in drug design due to the throughput and the incorporation of diverse experimental data. Ligand docking calculations focus on the evaluation of binding geometries and energies of ligands to a known receptor structure with no or little prior knowledge of the binding modes by heuristic ligand placement algorithms and fast empirical scoring functions, which highly minimize the computational effort. Most docking algorithms do not consider the flexibility of the receptor structure, whereas there are current attempts to face that drawback [60–62]. Free energy calculations by means of Monte Carlo (MC) or molecular dynamics (MD) simulations aim at the computation of free energies for molecular systems based on the principles of statistical thermodynamics. Extensive MC and MD simulations, which are of magnitudes more computational demanding than the QSAR and docking approaches, are applied to gain a deep insight into drug interactions. MD simulations provide an unique opportunity to gain invaluable knowledge about protein dynamics and structure in full atomic detail. Beside the *in-silico* drug design [63] MD simulations find a number of applications in biopharmaceutical science such as for protein structure prediction [64], analysis of protein stability and folding processes [65–67], and structure refinement [68]. The basic framework of computing the dynamics of a molecular system

is described in section 3.3.1 and the basic approaches for gaining free energies by means of MD simulations are described in section 3.3.2.

3.3.1 Mathematical Framework

The calculation of molecular dynamics, and of atom positions as a function of time respectively, is based on a mathematical framework, which combines the Newton's classical equation 3.3 and equation 3.6. The Newton's equation correlates the force F_i acting on each atom i to the atom's mass m_i and the atom's acceleration a_i. The atom's acceleration a_i can be substituted by the second derivative of position r_i with respect to time t, resulting in equation 3.3.

$$F_i = m_i a_i = m_i \frac{d^2 r_i}{dt^2} \tag{3.3}$$

For a constant acceleration $a_i = dv_i/dt$, the velocity v_i can be expressed as $v_i = a_i t + v_{i,0}$. Since the velocity v_i is the first derivative of position r_i with respect to time t, the coordinate r_i can be expressed as equation 3.4.

$$r_i = v_i t + r_{i,0} \tag{3.4}$$

The combination of equation 3.4 with the expression for the velocity $v_i = a_i t + v_{i,0}$ gives the value of the coordinate r_i at time t as a function of the acceleration a, the initial position r_0, and the initial velocity $v_{i,0}$ according to equation 3.5.

$$r_i = \frac{1}{2} a_i t^2 + v_{i,0} t + r_{i,0} \tag{3.5}$$

In addition, the force F_i acting on each atom is correlated to inter- and intramolecular interactions in dependency of its position, which are compromised in the potential energy function V according to equation 3.6.

$$F_i = -\frac{dV}{dr_i} \tag{3.6}$$

Consequently, the acceleration a_i is given as the derivative of the potential energy with respect to the position r_i as stated in equation 3.7.

$$a_i = -\frac{1}{m_i}\frac{dV}{dr_i} \tag{3.7}$$

The time evolution of the atom coordinates r and the potential energy function V is called a trajectory. With a known potential energy function V and given coordinates r of a starting structure and a set of velocities, the force acting in each atom can be evaluated and a new set of coordinates and velocities can be derived by combining equation 3.3 and 3.6. This new information again serves as basis for a new calculation of coordinates and velocities. By this repetitive procedure a trajectory can be calculated corresponding to the dynamic evolution of the system in time.
In detail, the functional form of the potential energy function V is a superposition of terms representing intermolecular non-covalent (V_{elec}, V_{vdW}) and intramolecular covalent interactions (V_{bond}, V_{angles},$V_{torsion}$) [69]. The non-covalent terms can be described according to the Coulomb's law for electrostatic interactions V_{elec} of atom-centered point charges, as stated in equation 3.8, and the Lennard-Jones potential for the Van der Waals V_{vdW} interactions, as shown in equation 3.9.

$$V_{elec} = \sum_{i<j} \frac{q_i q_j}{\epsilon_0 R_{ij}} \tag{3.8}$$

$$V_{vdW} = \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} \right] \tag{3.9}$$

The covalent bonds can be described according to a model of an ideal spring and the harmonic force (equation 3.10), its energy due to the geometry of electron orbitals involved in covalent bonding (equation 3.11), and its energy for twisting a bond (equation 3.12).

$$V_{bond} = \sum_{\text{bonds}} K_b(r - r_0)^2 \tag{3.10}$$

$$V_{angle} = \sum_{\text{angles}} K_a(\theta - \theta_0)^2 \tag{3.11}$$

$$V_{torsion} = \sum_{\text{dihedrals}} \frac{1}{2} V_n \left[1 + cos(n\phi - \gamma)\right] \tag{3.12}$$

The values of all relevant parameters such as force constants (K_a and K_b), equilibrium bond lengths, angles and charges are defined for every single atom and bond type in so called force fields and were originated from experimental data such as vibrational analysis and X-ray structural data.

3.3.2 Free Energy Calculations and Their Applications

Molecular dynamics simulations can be used for the calculation of free energies or free energy differences between two states respectively. This is applied for chemical or biochemical processes requiring a thorough calculation of the underlying free energy changes related to solvation [70, 71], protein-ligand binding [72–74], proton and electron transfer reactions [75, 76], transport and partitioning phenomena [77–79] or polypeptide folding and stability [80, 81]. All these complex reactions cannot be understood or even predicted without knowing the accomplished free energy changes.
Even if physical and chemical constants can currently be evaluated with a reasonable level of reliability, the investigation of realistic biological systems is still highly challenging for modern *in-silico* chemistry due to limitations in computational performance. However, with an enormous progress of parallel architectures and processing performance in combination with the optimization of software algorithms, sophisticated and modern sampling methods, free energy determinations be performed in a cost-efficient, precise, and accurate way. The computation of free energies can help to gain a profound understanding for experimental observations and can serve for developing predictive models.
Molecular dynamics simulations applied for free energy calculations became an appreciated tool for drug design and discovery and even in process development in biopharmaceutical academia and is used in biopharmaceutical industry [82–84]. In scope of biopharmaceutical research solvation free energy, binding free energy and conformational free energy are of high interest for drug design and pharmacology studies. As touched in the following, absolute free energies can be gained theoretically from MD simulations. In general terms, a molecular system can be described as a *Hamilton* function or operator which is often expressed as the Hamiltonian $H(p, q)$ of the generalized coordinates q and its momenta p [85]. Thus, a classical molecular ensemble of n atoms can be expressed by means of the Cartesian coordinates of the atoms $r \equiv (r_1, r_2, ..., r_n)$ and its momenta $p \equiv (p_1, p_2, ..., p_N)$ with the Hamiltonian according to equation 3.13

$$H(p,r) = \sum_{i=1}^{N} \mathrm{p}_i^2/2m_i + V(r_1, r_2, ..., r_N), \tag{3.13}$$

where m_i is the mass of atom i and $V(r)$ reflects the potential-energy function, describing covalent and non-covalent interactions of the atoms according to equation 3.6. In a canonical ensemble where atom number n, system volume v and system temperature T are constant, the Hamilton can be applied for calculating the Helmholtz free energy F according to equation 3.14

$$F(n,v,T) = -k_B T \ln \left[h^{-3n} \int \int \exp\left\{-H(p,r)/k_B T) d\mathrm{p} d\mathrm{r}\right\} \right], \tag{3.14}$$

where k_B is the *Boltzmann's* constant, h the *Planck's* constant, v the system volume, t the absolute temperature, and it is assumed that the n atoms are distinguishable [86]. In practice, absolute free energies according to equation 3.14 can only be calculated for a small number of model systems with limited space of states due to insufficient sampling during finite length simulations.

Due to this limitation, only free energy differences between two states of a system can be gained from molecular dynamics simulations. There are a number of existing methods for the determination of differences in free energy ranging from simple empirical, statistical or knowledge-based scoring functions, to computationally demanding statistical mechanical approaches such as the thermodynamic integration (TI) or the related free energy perturbation (FEP) method, and methods based on probability distribution functions and sampling probabilities. However, the inverse correlation between computational speed and accuracy of energy evaluation has to be considered and thus the application of the computationally demanding statistical mechanical approaches are of limited use for highly complex biochemical assemblies and biological macromolecules. To overcome this drawback, a number of simplified approximate methods were developed such as linear response theory, potential of mean force or linear interaction energy approaches. Computational effort is additionally reduced by simplified solvent treatment [87] or coarse grained simulations [88].

The linear interaction energy (LIE) approach is applied in the MD simulations based study of protein partitioning in aqueous two phase systems in this work and thus will be described in the following. The LIE approach was pioneered by Johan Åqvist's group for the calculation of binding energies between receptors and ligands and became an attrac-

tive and powerful approach in drug design and lead optimization since it offers a good compromise between speed and accuracy [89–91]. In contrast to the so called alchemy studies, to whom the thermodynamic integration and related free energy perturbation approaches belong to, the LIE method is an endpoint method, which attempts to compute the free energy difference between two states from simulations of these states only, with no consideration of either physical or non-physical intermediates like in the FEP or TI methods [59]. The binding free energy between interacting compounds is assessed on base of the differences in electrostatic and van der Waals interactions between two states and the difference between the bound and unbound state in particular. As a more generalized form, this free energy difference between the bound state A and and the free unbound state B can be expressed as a sum of the corresponding polar and non-polar components of the free energy in accordance to equation 3.15

$$\begin{aligned}\Delta G_{binding} =& \alpha \left\langle V_A^{vdW} - V_B^{vdW} \right\rangle + \beta \left\langle V_A^{elec} - V_B^{elec} \right\rangle + \gamma \\ =& \alpha \Delta \left\langle V_{A-B}^{vdW} \right\rangle + \beta \Delta \left\langle V_{A-B}^{elec} \right\rangle + \gamma, \end{aligned} \tag{3.15}$$

where α and β are empirically derived scaling factors for the non-polar and polar binding energy contributions and γ is a constant. In the first application of the LIE method, these three parameters were set to 0.16, 0.5 and 0.0, but have been reparametrized in more recent studies [92, 93]. Since the potential energy function splits the non-bonded potential energies into electrostatic and non-electrostatic components, the polar and non-polar components can be gained from energy evaluation of the trajectories.
The calculation of binding free energies by means of the LIE approach might be applied for the investigating of protein partitioning in aqueous two phase systems since the process of ligand binding to a biological macromolecule can be interpreted as a partition problem, in which the ligand, or the molecule of interest respectively, is transferred from one medium or state A into another B. Thus, the LIE approach can be applied to study partitioning of solutes between two phases as pointed out by Jorgensen and coworkers who used the linear interaction energy method for calculating hydration free energies and partition coefficients between water and chloroform [94, 95]. A proper parametrization of the weighting factors α and β are essential for the accuracy of the model and thus can be further refined [92] according to the two states A and B as expressed in equation 3.16. Experimentally derived data can serve for a proper parametrization and a refinement of the LIE approach.

$$\Delta G_{binding} = \alpha_A \left\langle V_A^{vdW} \right\rangle - \alpha_B \left\langle V_B^{vdW} \right\rangle + \beta_A \left\langle V_A^{elec} \right\rangle - \beta_B \left\langle V_B^{elec} \right\rangle + \gamma. \tag{3.16}$$

Sulea et al. could show by means of an extensive set of molecules, that the LIE approximation is at least as accurate as the FEP approach with a reduction of computing time by at least one order of magnitude [96], while additional terms were implemented into the LIE formalism. Thus the LIE method was chosen as an appropriate method to investigate the partitioning of amino acids, peptides, and proteins within aqueous two phase systems by means of molecular dynamics simulations within this thesis.

3.4 3D Printing

The 3D printing technology has matured over the last years and completely changed the mindset in laboratory work and accelerated science [97, 98]. Typical applications are the production of functional models, visual aids for engineering, education, toolmakers, and for rapid prototyping [99, 100]. In the field of process engineering, numberless applications of 3D printing technology can be found starting from the modeling of custom tailored porous media [101] or surface functionalization with proteins, multi-enzyme systems, DNA/RNA assemblies, or even cells for diagnostics or biosensors [102–105]. The fast-growing 3D printing community, open source coding and free CAD software further boost that technology. The progress in printing biocompatible and biological material potentially paves the way for personalized medicine ranging from medical devices, implants, to complete organs [106–112]. Recent advances have enabled 3D printing of biocompatible materials, cells, and supporting components into complex 3D functional living tissues and implants [113–115]. However, it should be noted that the term bio-printing in scope of surface functionalization is often used for structured coating with traditional coating techniques such as flexographic printing, gravure printing, rotary screen printing or 2D inkjet printing [116] instead of 3D printing. Basically, 3D printing builds parts in thin layers which enables to design highly complex multi-component assemblies with sophisticated internal features and passages, which are impossible to machine. Approaches to model a layer are quite different and a selection is schematically illustrated in Figure 3.6. Fused filament fabrication (FFF) as shown in Figure 3.6a, or also known as plastic jet printing (PJP) or fused deposition modeling (FDM) trademarked by Stratasys Inc. (Eden Prairie, Minnesota, USA), is one of the commonly used manufacturing

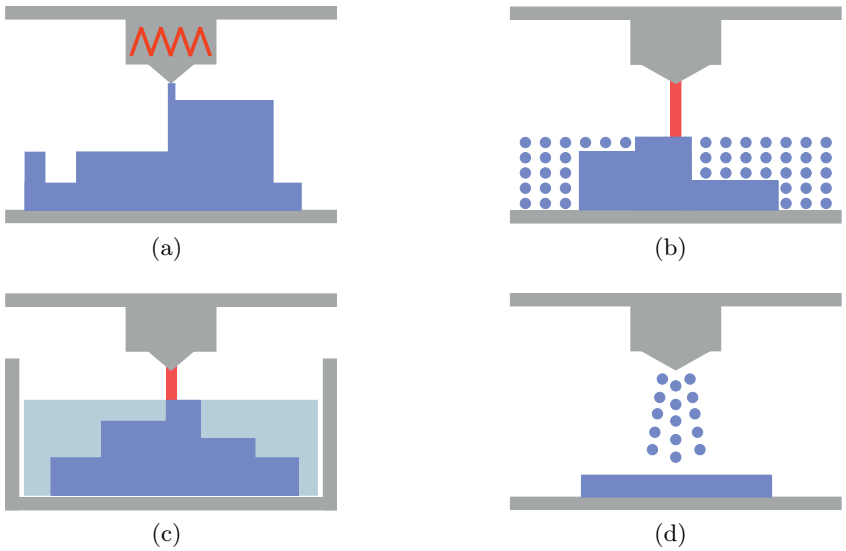


Figure 3.6: Schematic illustration of different 3D printing technologies: (a) Fused filament fabrication, (b) selective laser sintering, (c) stereolithography, and (d) multi-jet modeling.

technologies. In this approach a model is build layer by layer by extruding small flattened strings of molten material, which hardens rapidly at deposition. Selective laser sintering (SLS), as shown in Figure 3.6b, uses a laser to sinter powdered material for layer building. Stereolithography (SL or SLA), as shown in Figure 3.6c, builds the part by using an UV laser to solidify liquid photopolymer resins. Similar to SLA is multi-jet modeling (MJM), which hardens thin liquid layers of UV-sensitive photopolymers, as schematically illustrated in Figure 3.6c. A number of materials are applied for 3D printing ranging from thermoplastics, metals to biomaterials. Typically used thermoplastics are acrylonitrile butadiene styrene (ABS), polyphenylsulfone (PPSF or PPSU), polyphenylene sulfide (PPS), polycarbonate (PC), and polyamide (i.e. Nylon). For printing metal parts a number of metals and alloys such as bronze, stainless steel, or inconel 625 and 718 are used [117, 118].

CHAPTER 4

Research Proposal

The ground breaking discovery of DNA function and the following development of genetic engineering techniques paved the way for the production of biopharmaceutical drugs, which offer unique opportunities and bear hope for treating and healing severe diseases. Due to the complexity of pharmaceutical proteins in terms of structure, solubility, stability, and pharmacokinetics, the accurate assessment of protein properties and protein behavior in solution is of paramount importance. A more profound knowledge of protein behavior in solution enables an economic and efficient development of robust production procedures and increases the safety of the drug product.

The assessment of protein-solvent interactions in solution, which is described by protein hydrophobicity, is mandatory to understand the process relevant protein behavior such as protein aggregation and partitioning. There are a number of existing computational and experimental methods to estimate protein hydrophobicity. Computational methods use theoretical scales to calculate protein hydrophobicity on basis of amino acid composition, which is very cost and time efficient but cannot consider the complexity and dynamics of the protein structure and the electrostatic dynamics as a function of environmental properties such as pH or ionic strength. Experimental approaches are based on colorimetric or fluorescence probing of the protein surface, precipitation studies, or on the evaluation of the protein retention behavior in chromatographic analytics with hydrophobic resolution such as hydrophobic interaction and reversed phase chromatography. However, these experimental methods are invasive and bear the risk of harming the protein conformational integrity. Thus, the development of appropriate computational and experimental methods are of high interest and are focused in this work.

A methodology which has been shown to be a highly promising tool to investigate protein dynamics is the computational molecular dynamics (MD) simulation. Overcoming the drawbacks of conventional hydrophobicity scales, a MD simulations based approach for characterizing peptides' and proteins' surfaces regarding hydrophobicity is of high

interest and should be evaluated in scope of this project. Due to the invasiveness and limitations in buffer compatibility of conventional experimental procedures for the assessment of protein hydrophobicity, an alternative experimental approach for assessing protein-solvent interactions is highly desirable and should be focused in scope of this project. Since protein material is a limiting factor during the bioprocess development, developed methods should be integrable into a high throughput work flow.
Understanding of protein-solvent interactions could be a key for comprehending the protein phase behavior and protein partitioning behavior in aqueous two phase systems (ATPSs). ATPSs have been shown to be cost efficient, highly selective, and can be performed continuously. Moreover, these systems practically do not have any limitations regarding scalibility and are instrumentationally simple. However, the use of ATPSs is still limited to academic laboratories to the largest extent, which is mainly due to a lack of understanding. To understand the partitioning phenomena of biomolecules in ATPSs the investigation of associated free energy changes is required, which can be derived by molecular dynamics simulations. There are a number of designs and approaches to derive free energy changes by means of MD simulations, which vary in accuracy and computational demand. The combination with experimental data derived from high throughput data require a certain throughput of the *in-silico* approach and thus the design and approach for the MD simulations has to be chosen accordingly. A further process relevant issue, which is highly affected by protein-solvent interactions, is the multimeric association of protein molecules to aggregates. The mechanisms of the formation of aggregates are not fully understood and their investigation is conducted empirically by means of time and material intensive experimental studies. High throughput screenings can minimize the effort by miniaturization of scale, automation, and parallelization and are valid and highly potent for the empirical assessment of protein phase behavior. In order to understand protein phase behavior in diluted solutions, the protein-solvent interactions and protein properties have to be investigated. Therefore, protein surface properties i.e. electrostatics and hydrophobicity in combination with a measure for the conformational flexibility, which can be derived by molecular dynamics simulations, can be focused.
In summary, this project aims at the development of sufficient experimental and computer aided molecular dynamics simulations based approaches for the molecular assessment of proteins in aqueous solutions which are capable to consider the complex interplay of solvent and protein in order to deepen the understanding of protein behavior in solution. The findings and the thereby developed approaches could be used to increase the understanding of process related issues such as protein phase behavior and protein partitioning in aqueous two phase systems.

CHAPTER 5

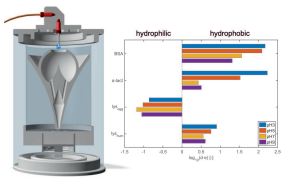
Comprehensive Overview of Publications & Manuscripts

This section provides an overview of the manuscripts compiled in scope of the presented thesis and puts the single subprojects into context. The manuscripts are ordered according to subjects and not chronologically. As mentioned earlier, the present thesis consists of two major pillars. The first one's goal is the assessment of protein properties and the second one aims at the process related knowledge building and in particular protein behavior with respect to colloidal stability and partitioning in two phase extraction.
The manuscript titled *Non-Invasive High Throughput Approach for Protein Hydrophobicity Determination Based on Surface Tension* aims at the development of a non-invasive approach for the characterization of proteins with respect to protein hydrophobicity. A stalagmometric setup was implemented into a liquid handling station by means of the *Tip2World* interface and thus could be operated fully automatically. The hydrophobicity of proteins was determined as a function of pH and served for the validation of further *in-silico* attempts as well as process related studies. The automated determination of surface tension was further refined in terms of throughput and sample consumption in the study titled *Determination of Surface Tension of Liquids by Means of Liquid Handling Stations*. This work highlights the utility of liquid handling stations for automated, accurate, and precise surface tension assessment with a simple experimental setup. A molecular dynamics simulations based approach to characterize proteinogenic structures regarding hydrophobicity is described in the manuscripts titled *Molecular Dynamics Simulations Approach for the Characterization of Peptides With Respect to Hydrophobicity* and *Molecular Dynamics Simulations Approach for the Characterization of Proteins With Respect to Hydrophobicity*. This MD simulations based approach was developed by means of simple, short, custom designed peptides and was experimentally validated by reversed phase (RP) chromatography. This *in-silico* approach was refined in terms of throughput and applicability to larger protein structures and could experimentally be validated with the experimental non-invasive approach based on surface

tension profiling mentioned above. By these works a comprehensive assessment of protein characteristics in terms of hydrophobicity could be performed.
The second major pillar of the present thesis focuses on process related questions, i.e. the protein phase behavior and the protein behavior in aqueous two phase systems. The protein phase behavior is focused by the manuscript titled *Predictive Approach for Protein Aggregation: Correlation of Protein Surface Characteristics and Conformational Flexibility to Protein Aggregation Propensity.* The study to predict protein aggregation propensity focuses on the assessment of parameters which are sufficient to predict the complex protein behavior in solution in absence and presence of precipitating salts and sodium chloride and ammonium sulfate in particular. Experimentally and *in-silico* derived protein parameters were shown to be of high value for protein phase behavior in solutions. The interactions between highly complex solvents rich in salt and polymer dominate the protein partitioning in salt/polymer aqueous two phase systems at diluted protein concentrations. The computational investigation of the partitioning mechanisms by means of molecular dynamics simulations, required an experimental characterization of the respective aqueous two phase systems regarding composition and density. This was realized as described in the manuscript titled *Characterization of Aqueous Two phase Systems by Combining Lab-on-a-Chip Technology With Robotic Liquid Handling Stations.* This manuscript focuses on a technical development which enables the combination of industrial relevant robotic liquid handling stations and lab-on-a-chip devices in order to characterize aqueous two phase systems and served the data base for the *in-silico* study of partitioning in aqueous two phase systems. In addition, the technical implementation by means of the so called *Tip2World* interface plays a key role in the projects for experimental protein assessment with respect to hydrophobicity described above. The protein specific phase affinities within aqueous two phase systems have been studied by molecular dynamics simulations, which is described in the manuscript titled and by the study titled *Molecular Dynamics Simulations on Protein Partitioning in Aqueous Two Phase Systems.* Extensive full-atom molecular dynamics simulations were performed to gain partitioning free energies applying an approximate method, namely the linear interaction energy (LIE) method. The comprehensive data set serves for building predictive models for protein partitioning in combination with experimentally determined partitioning coefficients.

Non-Invasive High Throughput Approach for Protein Hydrophobicity Determination Based on Surface Tension

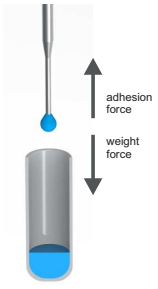
Sven Amrhein*, Katharina Christin Bauer*, Lara Galm*, and Jürgen Hubbuch
Biotechnology and Bioengineering 112.12 (2015): 2485-2494.
**contributed equally*



A non-invasive high throughput approach for protein assessment in terms of hydrophobicity has been developed in this work, which bases on the surface tension of protein solution in dependency of the protein concentration and does not require any buffer exchanges or use of additives. The surface tension was measured by a stalagmometric setup integrated into a fully automated liquid handling station. This setup is able to measure surface tensions with a high accuracy. A protein hydrophobicity scale was derived from the surface tension increment of a number of model proteins for a pH range starting from 3 to 9 and was validated with one of the most widely used invasive methods using absorption difference spectroscopy of bromophenol blue.

Determination of Surface Tension of Liquids by Means of Liquid Handling Stations

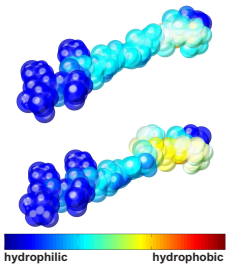
Sven Amrhein, Susanna Suhm, and Jürgen Hubbuch
Engineering in Life Sciences (2016), accepted



The determination of surface tension is of major interest throughout a number of disciplines and applications such as pharma, micro and nano fluidics, cosmetics, food and beverage, medical implants or coating applications. This technical study presents a high throughput method for the determination of surface tension based on the drop weight method. This method for surface tension determination was especially designed for liquids, which expose challenging handling properties and are limited in material. Repetitive drop generating procedures enable to create virtually an unlimited number of drops from a minimum of sample liquid and thus a highly optimized yield of statistical significance.

Molecular Dynamics Simulations Approach for the Characterization of Peptides With Respect to Hydrophobicity

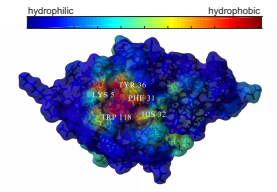
Sven Amrhein, Stefan Alexander Oelmeier, Florian Dismer, and Jürgen Hubbuch
The Journal of Physical Chemistry B 118.7 (2014): 1707-1714.



In this work an *in-silico* approach was developed in order to characterize peptides with regard to hydrophobicity. A set of custom designed short peptides containing ten amino acids each were used for molecular dynamics simulations in the presence of hydrophobic molecules, namely triethylamine. Based on the distribution of this hydrophobic molecules around the protein surface in dependency of simulation time, the hydrophobicity of the peptides could be evaluated. The results were validated with the retention in reversed phase chromatography which is a valid measure for hydrophobicity. The developed method is not exclusively able to derive an overall hydrophobicity of the total peptide, but was capable to resolve the peptides' surface hydrophobicity on an atomic level of detail.

Molecular Dynamics Simulations Approach for the Characterization of Proteins With Respect to Hydrophobicity

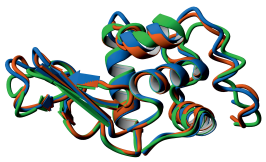
Sven Amrhein and Jürgen Hubbuch
in preparation



This work focuses on the characterization of the protein surface with respect to hydrophobicity by means of molecular dynamics simulations. A previously developed procedure described in the manuscript above was further optimized in terms of throughput in order to be applicable to larger protein structures. The surfaces of three proteins, namely α-lactalbumin, human lysozyme, and lysozyme from chicken egg white, were characterized with respect to hydrophobicity at pH 5 and pH 7. The MD based approach could be shown to be capable of considering protein complexity and solvent effects and pH in particular.

Predictive Approach for Protein Aggregation: Correlation of Protein Surface Characteristics and Conformational Flexibility to Protein Aggregation Propensity

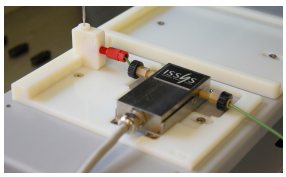
Lara Galm*, Sven Amrhein*, and Jürgen Hubbuch
Biotechnology and Bioengineering (2016), accepted
**contributed equally*



In this work protein aggregation propensity was studied by means of experimental protein assessment and theoretical examination by molecular dynamics simulations. In particular, the aggregation behavior of human lysozyme, lysozyme from chicken egg white, and α-lactalbumin as a function of protein concentration, pH, precipitant species and concentration was studied to investigate the predictive power of experimentally derived zeta potential and protein hydrophobicity and conformational flexibility assessed by molecular dynamics simulations. The phase behavior was found to correlate with zeta potential, protein hydrophobicity as well as *in-silico* derived protein properties.

Characterization of Aqueous Two phase Systems by Combining Lab-on-a-Chip Technology With Robotic Liquid Handling Stations

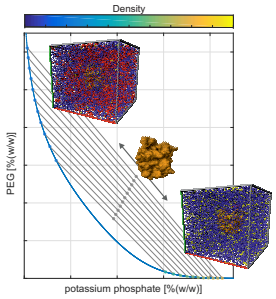
Sven Amrhein, Marie-Luise Schwab, Marc Hoffmann, and Jürgen Hubbuch
Journal of Chromatography A 1367 (2014): 68-77.



This work focuses on an innovative high throughput compatible and non-invasive approach for the characterization of aqueous two phase systems with respect to tie line determination. A generic interface which enables to integrate flow through devices into industrial relevant liquid handling stations was developed and produced by means of 3D printing. Using this so called *Tip2World* interface a micro Coriolis density sensor could be operated by liquid handlers and tie lines of aqueous two phase systems were evaluated based on the determined phase densities.

Molecular Dynamics Simulations on Protein Partitioning in Aqueous Two Phase Systems

Sven Amrhein and Jürgen Hubbuch
in preparation



Aqueous two phase systems are highly sufficient alternatives to chromatographic bioseperation procedures but mechnisms dominating the partitioning of biomolecules are barely understood. In this study, extensive molecular dynamics simulations of biomolecules with varying complexity starting from single amino acids over short peptides to complex proteins in aqueous two phase systems were performed in order to derive affinities to phases. Phase affinity was derived by energy considerations and a theoretical partitioning behavior could be described. Top and bottom phase of aqueous two phase systems composed of potassium phosphate at pH 7 and polyethylene glycol with molecular weights of 300, 400, 600, and 1000 Da were studied. The derived data could serve for building predictive models for protein partitioning.

CHAPTER 6

Non-Invasive High Throughput Approach for Protein Hydrophobicity Determination Based on Surface Tension

Sven Amrhein[1], Katharina Christin Bauer[1], Lara Galm[1], and Jürgen Hubbuch

[1] these authors contributed equally

Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Science, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany

Biotechnology and Bioengineering 112.12 (2015): 2485-2494.

Abstract

The surface hydrophobicity of a protein is an important factor for its interactions in solution and thus the outcome of its production process. Yet most of the methods are not able to evaluate the influence of these hydrophobic interactions under natural conditions. In the present work we have established a high resolution stalagmometric method for surface tension determination on a liquid handling station, which can cope with accuracy as well as high throughput requirements. Surface tensions could be derived with a low sample consumption (800 µL) and a high reproducibility (< 0.1 ‰ for water) within a reasonable time (3.5 min per sample). This method was used as a non-invasive HTP compatible approach to determine surface tensions of protein solutions dependent on protein content. The protein influence on the solutions' surface tension was correlated to the hydrophobicity of lysozyme, human lysozyme, BSA, and α-lactalbumin. Differences in proteins' hydrophobic character depending on pH and species could be resolved. Within this work we have developed a pH dependent hydrophobicity ranking, which was found to be in good agreement with literature. For the studied pH range of 3 to 9 lysozyme from chicken egg white was identified to be the most hydrophilic. α-lactalbumin at pH 3 exhibited the most pronounced hydrophobic character. The stalagmometric method occurred to outclass the widely used spectrophotometric method with bromophenol blue sodium salt as it gave reasonable results without restrictions on pH and protein species.

Keywords: *Stalagmometer; Bromophenol Blue; Hydrophilicity; Protein Solution; Protein-Solvent Interaction*

6.1 Introduction

Hydrophobic interactions play a key role in the outcome of the production process of biopharmaceutical therapeutics passing through fermentation, the purification process, formulation and storage. During protein expression in fermentation hydrophobic forces regulate the formation of the globular protein molecule [119, 120]. The resulting surface characteristics and the protein concentration govern its solubility for all following production steps. Already small changes in the hydrophobic character can provoque changes in solubility and in the aggregation tendency of the molecule [121, 122]. Undesired aggregation during the process or storage can cause denaturation and thus product loss. However, for the purification process this changes regarding solubility can also be turned into advantage in terms of protein crystallization or precipitation as purification steps. The hydrophobic character of a protein can additionally be exploited to separate complex protein mixtures by using aqueous two phase systems (ATPS) [51, 53, 123], reversed phase (RP) chromatography or hydrophobic interaction chromatography (HIC) [124]. The knowledge of protein surface hydrophobicity therefore helps to predict, control and manipulate the influence of hydrophobic interactions during processing and storage. Within the last decades a huge research effort has been spend on the development of experimental and *in-silico* approaches for assessing protein surface hydrophobicity and the identification of highly hydrophobic proteins. In experiments most frequently protein hydrophobicity has been adapted from the hydrophobicity of single amino acids, that are ranked in different hydrophobicity scales [21, 125–128]. This amino acid hydrophobicity has been measured mainly in terms of their solubility in organic and denaturant solutions [129–134] or their partition between an aqueous and an organic phase [135, 136]. Efforts have been made on hydrophobicity rankings of whole proteins based on partitioning in aqueous two phase systems [137], retention factors in HIC [138] or using hydrophobic dyes [25, 139–142]. Yet most experimental methods are invasive and not adequate to consider the influence of solution characteristics on the protein. In most cases organic liquids are inevitable which may influence the tertiary structure of the protein or even denature it.

In-silico methods apply experimentally determined hydrophobicity scales to quantify protein hydrophobicity. These methods calculate the protein hydrophobicity either based on the amino acid sequence [143] or based on the three-dimensional structure [144–146]. Salgado et al. [147, 148] found that methods based on the three-dimensional structure show a better predictive performance for protein adsorption mechanisms in HIC than

the ones based on the amino acid sequence. Sophisticated approaches based on quantitative structure property relationship (QSAR) [149, 150] or molecular dynamics (MD) simulations [151, 152] give the most detailed description of the protein surface. However, these sophisticated *in-silico* approaches require either an enormous computational effort or use theoretical hydrophobicity scales as described above and thus cannot account for the influence of the solvent. Generally, approaches using theoretically derived hydrophobicity scales are highly influenced by the selected hydrophobicity scale [153, 154].
One promising way of considering the environmental dependency is to measure hydrophobicity via surface tension. This experimental method is able of capturing the environmental dependency because it is non-invasive. The existence of a correlation between surface or interfacial tension and hydrophobicity of single amino acids or macromolecules is known for a long time [138, 139, 155, 156]. Bull and Breese [155] pointed out the potency of sorting amino acids by hydrophobicity according to their surface tension increment. According to this, hydrophobic amino acids reduce the surface tension with increasing concentration. Conversely, hydrophilic amino acids increase the surface tension.
Keshavarz and Nakai [138] could show a significant negative correlation between the effective hydrophobicities of bovine serum albumin, ovalbumin, lysozyme, γ-globulin, myoglobin, β-lactoglobulin, trypsin, conalbumin, and α-chymotrypsin and interfacial tension. The more hydrophobic the protein, the greater the depression in the interfacial tension. Recently Genest et al. [157] found a correlation between the depression of the surface tension and the polymer hydrophobicity.
There are quite a number of methods for the determination of surface tensions such as the Capillary rise method, the Wilhelmy plate method, pendant and sessile drop methods, and the stalagmometric method. This method combines low sample consumption and compatibility to high throughput liquid-handling devices. Additionally, we are confident that this gravimetric approach is superior to imaging approaches in terms of precision and robustness.
In the present work we chose the stalagmometric method and developed a high resolution experimental setup which can be integrated into high throughput (HTP) work flow by a liquid handling station. This method was used as a non-invasive high-throughput compatible approach to determine protein hydrophobicity on base of the proteins' surface tension increments. Lysozyme, human lysozyme, BSA, and α-lactalbumin were characterized regarding hydrophobicity dependent on pH value. This set of proteins covers a wide range of molecular weight, isoelectric points and includes two similar lysozyme species differing in amino acid composition. The hydrophobicity values for BSA derived from this approach were found to be in good agreement to values obtained by the widely

used method of absorption difference spectroscopy of bromophenol blue sodium salt (BPB Na). Instead of this pH and protein species dependent absorption method the developed stalagmometric method was able to cover the full pH range in a completely automated way by considering the environmental protein complexity.

6.2 Materials and Methods

6.2.1 Stalagmometric Method for Determination of Surface Tension

The stalagmometric method was chosen for the determination of the surface tension due to its ability to be transformed into a fully automated procedure. In this method the specific fluid is purged very slowly through a needle in a vertical direction in order to form drops. The drop grows up to a specific maximum volume and falls onto a high precision mass balance. The Tate's law [158] expresses the correlation between the drop's weight (W_{drop}) and its surface tension (γ) at the moment the drop falls from the needle with an outer radius r:

$$W_{drop} = 2\pi r\gamma. \tag{6.1}$$

However, the instrumental setup has an influence on the drop size and the registered weight will be W'_{drop}. W_{drop} and W'_{drop} are correlated by an instrumental setup correction factor f_{inst}. Considering this correction factor the surface tension γ can be expressed as

$$\gamma = \frac{m_{drop}g}{2\pi r f_{inst}}, \tag{6.2}$$

where m_{drop} represents the mass of the drop, g the acceleration of gravity, r the radius of the needle, and f_{inst} the correction factor. By using a liquid with a known surface tension γ_{ref}, resulting in a specific drop mass $m_{drop,ref}$, the correction factor can be evaluated and the surface tension γ_{sample} of the liquid of interest can be determined by applying equation 6.3 to its specific drop mass $m_{drop,sample}$.

$$\frac{\gamma_{ref}}{m_{drop,ref}} = \frac{\gamma_{sample}}{m_{drop,sample}}. \tag{6.3}$$

6.2.2 Automation of Stalagmometric Method for the Purpose of HTP

In the following the automation of this stalagmometric method, which principle can be found in subsection 6.2.1, is described. This method was optimized towards analysis speed, accuracy and precision. The validation was conducted using liquids with a wide range of surface tension values.

6.2.2.1 Automation Using a Liquid Handling Station

This method was established on a fully automated robotic liquid handling station, namely Freedom EVO® 100 purchased from Tecan (Crailsheim, Germany), equipped with stainless steel fixed tips and 1 mL dilutors. The liquid handling station was controlled using Evoware 2.4 SP3. The complete experimental setup, as is illustrated schematically by a partially section view in Figure 6.1, consisted of two major subunits, namely a *Tip2World* interface mounted on a docking station and the stalagmometric setup, interconnected via a capillary tubing (PEEK, ID: 0.25 mm). The *Tip2World* interface enables to supply liquid samples with a robotic liquid handling arm via connected standard capillary tubings and is described in detail elsewhere [159]. The stalagmometric system setup consisted of the distribution block, connecting the capillary tubing coming from the *Tip2World* interface with the drop generating needle, and a 250 mL container to collect the drops (Azlon®Specimen, SciLabware Limited, Stoke-on-Trent, United Kingdom). This container was placed on an Excellence WXTS205DU high performance balance unit (Mettler Toledo, Greifensee, Switzerland). The distribution block possesses a standard 10-32 coned input port and a vertically aligned standard 10-32 coned output port, where the drop generating stainless steel needle was fixed. This needle had an inner diameter of 0.5 mm and an outer diameter of 1.6 mm. Both ports were connected by an internal capillary (ID: 0.5 mm). The container held a drop trap to collect the falling drops in a gentle way. This was realized by a customized design shown in Figure 6.1 and enhanced the balance signal stability. To prevent the drop from evaporation a cylindrical evaporation trap was developed. This consisted of a cylinder, which was placed on a bottom ring with a water trough and a lid. All components were custom designed and 3D printed. Apart from bottom and lid (Sculpteo, Issy-les-Moulineaux, France) they were manufactured by a high resolution 3D printer (Stratasys, Minneapolis, USA). The construction work was carried out with the 2D/3D CAD software SolidEdge (Siemens PLM Software, Plano, USA). For the surface tension determination all samples were stored at

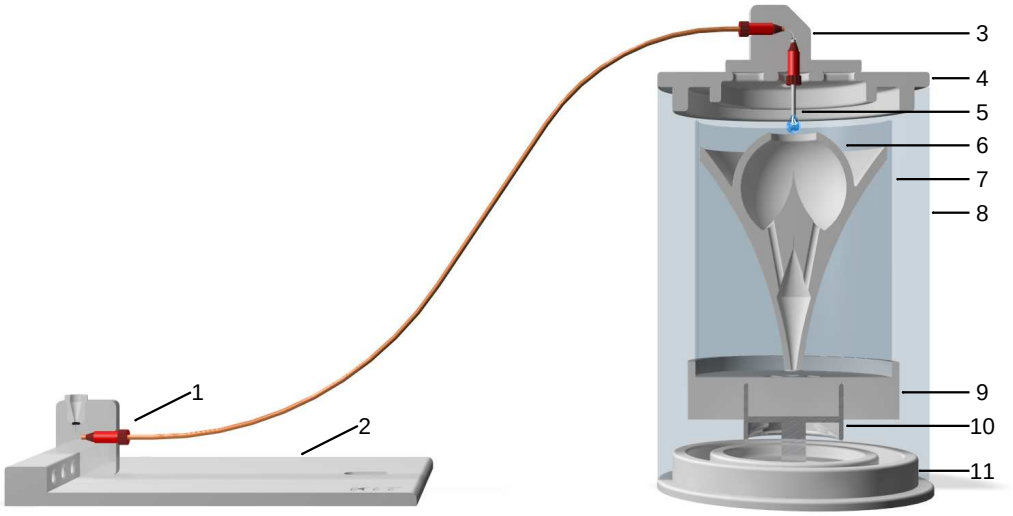


Figure 6.1: Illustration in partially section view of the setup of the stalagmometer, consisting of (1) *Tip2World* interface, (2) docking station, (3) the distribution block with a standard input 10-32 port, (4) a lid for fixation and centering of the distribution block, (5) a stainless steel needle in vertical orientation, (6) a drop trap, (7) a trough to collect and weight the drops, (8) a glass cylinder for evaporation protection, (9) a carrier, (10) balance unit, and (11) bottom part with water trough.

room temperature in 1.3 mL 96-well Deep Well plates (Nalgene Nunc, Rochester, NY, USA) and sealed with a pre-slit well cap (Nalgene Nunc, Rochester, NY, USA), in order to minimize evaporation. The sample to be analyzed was pumped with a flow rate of 5 µL/s using the *Tip2World* interface to the stalagmometer through a PEEK capillary tubing with a diameter of 0.25 mm. The falling drop masses were recorded continuously via serial communication realized by automated routines written in Matlab® R2013b (The MathWorks, Natick, ME, USA). By evaluating the step profile of the balance signal, the distinguished masses of each drop were calculated. In order to increase the accuracy, instabilities of the weight signal were sorted out. Processing of the experimental results of the stalagmometric method was performed by means of fully automated routines written in Matlab® R2013b.

6.2.2.2 Validation of Surface Tension Determination

The stalagmometric approach for surface tension determination was tested with respect to reproducibility by measuring 70 samples of ultrapure water with a volume of 800 µL each. For the validation of the procedure, surface tension measurements were performed with aqueous mixtures of sodium chloride (NaCl) and ethylene glycol solutions in order to cover a wide range of surface tensions. In particular, NaCl solutions were used with a molar fraction of 0 % to about 10 %. Ethylene glycol solutions were used within a molar fraction range of 0 % to 26 %. NaCl, analysis grade, was purchased from Merck KGaA (Darmstadt, Germany). Ethylene glycol, analysis grade, was purchased from VWR International (Radnor, PA, USA). Ethylene glycol and NaCl solutions were prepared by mixing the respective masses of ethylene glycol or NaCl with ultrapure water in order to reach the desired mass fractions. Ultrapure water was used as reference liquid. Cross contamination of samples was minimized by purging the tubing with air. All liquids were analyzed 8-fold with a volume of 800 µL each.

6.2.3 Correlation of Surface Tension and Hydrophobicity

In order to test the applicability of using the surface tension increment of an analyte to deduce its hydrophobicity, polyethylene glycol (PEG) with varying molecular weights were used. PEG is known to expose increased hydrophobicity with increasing molecular weight [52]. The polymer sample preparation is described in detail in 6.2.3.1. The method then was applied to protein solutions. These solutions varied in protein species, pH and buffer composition and the derived hydrophobicity measures were compared to hydrophobicity values derived from an orthogonal colorimetric method using BPB Na as described in section 6.2.3.2.

6.2.3.1 Stalagmometric Determination of Polymer Hydrophobicity

The surface tension increments of PEG species with a molecular weight of 200, 300, 400, 600 and 1000 Da were determined by measuring aqueous PEG solutions with molar fractions varying from 0.02 to 0.11 %. These concentrations correlate to mass fractions from 0.2 to 1.2 %(w/w) for PEG200, PEG300 and PEG400; from 0.4 to 2 %(w/w) for PEG600 and from 0.8 to 4 %(w/w) for PEG1000. Samples were prepared by mixing

the respective masses of PEG with ultrapure water. 800 µL of each solution were measured 8-fold. All PEG species were of analysis grade and purchased from Merck KGaA (Darmstadt, Germany).

6.2.3.2 Determination of Protein Hydrophobicity

Sample Preparation

The used buffer substances were citric acid (Merck, Darmstadt, Germany) and sodium citrate (Sigma-Aldrich, St. Louis, MO, USA) for pH 3, sodium acetate (Sigma-Aldrich, St. Louis, MO, USA) and acetic acid (Merck, Darmstadt, Germany) for pH 5, MOPSO (AppliChem, Darmstadt, Germany) for pH 7 and Bis-Tris propane (Molekula, Dorset, UK) for pH 9. Buffer capacity was 100 mM for all buffers. Hydrochloric acid and sodium hydroxide for pH adjustment were obtained from Merck (Darmstadt, Germany). pH adjustment was performed using a five-point calibrated pH-meter (HI-3220, Hanna Instruments, Woonsocket, RI, USA) equipped with a SenTix® 62 pH electrode (Xylem Inc., White Plains, NY, USA). pH was adjusted with the appropriate titrant with an accuracy of ± 0.05 pH units. All buffers were filtered through 0.2 µm cellulose acetate filters (Sartorius, Göttingen, Germany). Buffers were used at the earliest one day after preparation and repeated pH verification. Lysozyme from chicken egg white was purchased from Hampton Research (Aliso Viejo, CA, USA), human lysozyme, bovine serum albumin (BSA), and calcium depleted α-lactalbumin were purchased from Sigma-Aldrich (St. Louis, MO, USA). To set up the protein stock solutions, protein was weighed in and dissolved in the appropriate buffer yielding the desired concentration. All protein solutions were filtered through 0.2 µm syringe filters with cellulose acetate membrane (VWR, Radnor, PA, USA) and further desalted via size exclusion chromatography using a HiTrap Desalting Column (GE Healthcare, Uppsala, Sweden) on an AEKTAprime^{TM} plus system (GE Healthcare, Uppsala, Sweden). The desired concentration was achieved by using Vivaspin centrifugal concentrators (Sartorius, Göttingen, Germany) with PES membranes. Protein concentration determination of the collected fractions was conducted using a NanoDrop2000c UV-Vis spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). (Extinction coefficients were $E^{1\%}(280\ nm)_{lysozyme}$= 22.00, $E^{1\%}(280\ nm)_{human\ lysozyme}$= 16.00, $E^{1\%}(280\ nm)_{BSA}$= 6.70, $E^{1\%}(280\ nm)_{\alpha\text{-lactalbumin}}$= 16.81).

Stalagmometric Determination of Protein Hydrophobicity

Buffers and protein stock solutions of lysozyme, human lysozyme, BSA, and α-lactalbumin

were prepared as described above. Protein solutions were prepared by dilution of the protein stock solutions with the respective buffer. The following solutions were laid eightfold in a 1.3 mL 96-well Deep Well plate: for reference ultrapure water and the respective buffer, protein solutions varying in protein molar fractions from $5.6 \cdot 10^{-6}$ to $1 \cdot 10^{-2}$ % for lysozyme and from $1.4 \cdot 10^{-6}$ to $3.9 \cdot 10^{-5}$ % for human lysozyme, BSA, and α-lactalbumin. 800 µL of each of the 96 samples were pumped to the stalagmometer with a flow rate of 5 µL/s. The surface tension of all samples was analyzed according to 6.2.2.1 in 8-fold replicates.

Spectrophotometric Determination of Protein Hydrophobicity

The determination of protein hydrophobicity using bromophenol blue sodium salt (BPB Na) as a hydrophobicity sensitive dye was conducted for lysozyme, human lysozyme, BSA, and α-lactalbumin. Buffers and protein stock solutions were prepared as described earlier. BPB Na, purchased from Sigma-Aldrich (St. Louis, MO, USA), was dissolved in the respective buffer to reach a dye concentration of 0.02 mg/mL for the BPB Na stock solution. The spectrophotometric method was conducted according to Bertsch et al. [25]. No measurements were conducted at pH 3, because BPB Na is a well known pH indicator dye within the pH range of 3.0 to 4.6. Protein stock solutions were diluted using the respective buffers and mixed with a fixed volume of BPB Na stock solution, yielding protein solutions with a BPB Na concentration of 7.99 µM and protein molar fractions up to $3.3 \cdot 10^{-5}$ % for BSA and up to $4.9 \cdot 10^{-3}$ % for lysozyme, human lysozyme, and α-lactalbumin. The absorption spectra of a BPB Na solution without protein and of the BPB Na solutions with added protein were measured between 550 and 650 nm in 1 nm steps using an Infinite® M200 microplate reader (Tecan, Crailsheim, Germany). The measurements of each solution were conducted in triplicate. In contrast to Bertsch et al. we examined the shift of the absorption maximum ΔA_{max} in dependency of the protein concentration instead of the absorption difference at 620 nm ΔA_{620}. The value of the absorption maximum shift was calculated as mean value from the triplicate spectrophotometric measurements. The shift of the absorption maximum ΔA_{max} (in nm) was fitted using a Box Lucas model

$$\Delta A_{max} = a \cdot (1 - e^{-b \cdot \tilde{x}}), \tag{6.4}$$

where $\tilde{x}$ reflects the protein molar fraction, a and b are adjustable constants. Parameter a describes the limit of absorption maximum shift, $a \cdot b$ can be interpreted as relative surface

hydrophobicity. It reflects the slope of the fitted curve for a protein molar fraction $\tilde{x}$ approaching zero. The higher $a \cdot b$ the higher the protein surface hydrophobicity.

6.3 Results

6.3.1 Development of HTP Compatible Stalagmometric Method for Surface Tension Determination

We were able to create HTP compatibility of the stalagmometric method by connecting the stalagmometer to the prior developed *Tip2World* interface [159]. As mentioned in section 6.2.2.2 the HTP compatible stalagmometric method was tested for reproducibility using 70 samples of ultrapure water. These measurements showed a standard deviation of less than 0.1 ‰ which correlates to a 99.7 % confidence interval ($\pm$ 3σ), which equals a deviation of less than $\pm$ 0.2 mN/m for water. The validation, described in section 6.2.2.2, was conducted by comparison of derived surface tensions with published data [160–162]. The experimentally obtained surface tensions were in good agreement with reported data as illustrated in Figure 6.2. Ethylene glycol exposed a decreasing impact on the surface tension, NaCl increased the surface tension following a linear trend within the studied concentration interval. Thus the developed HTP compatible stalagmometric method showed high accuracy and covered a wide range of surface tensions from 57 to 82 mN/m. The developed method required about 3.5 minutes per sample resulting in less than 5.5 hours for analyzing a set of 96 samples without requiring sophisticated instrumentation and man power.

6.3.2 Stalagmometric Determination of Hydrophobicity

6.3.2.1 Hydrophobicity of Polymers

The influence of polyethylene glycol (PEG) on the surface tension was studied to assess hydrophobicity. PEG species of different molecular weights starting from 200 to 1000 Da were studied. All surface tensions were normalized on ultrapure water. As illustrated in Figure 6.3 all surface tensions could be derived precisely with relative standard deviations less than 1.5 ‰ throughout all PEG samples. It is apparent from Figure 6.3 that all

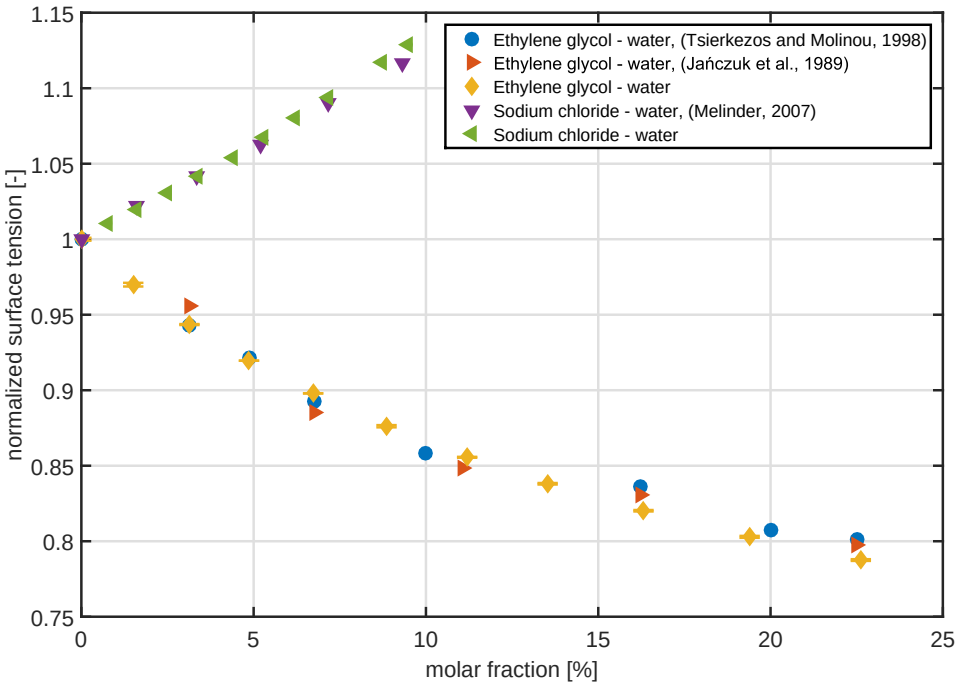


Figure 6.2: Comparison of experimentally determined surface tensions with literature values [160–162]. The error bars refer to the 95 % confidence interval ($\pm\ 2\sigma$).

studied PEG species followed a decreasing trend with increasing concentration. The higher the PEG molecular weight, the higher the decrease in surface tension.

6.3.2.2 Hydrophobicity of Proteins

Surface tensions of lysozyme, human lysozyme, BSA, and α-lactalbumin were analyzed in dependency of protein concentration and pH in order to characterize their hydrophobicity within the respective buffer system. Figure 6.4 illustrates the normalized profiles of the surface tensions of these proteins. The surface tensions were normalized on the respective pure buffer. Human lysozyme, BSA as well as α-lactalbumin decreased the surface tension even at very low molar fractions of less than $4{\cdot}10^{-5}$ %. It is apparent from Figure 6.4 that the decrease of surface tension is highly influenced by the pH value.
In particular, for BSA pH 3 had the highest influence on surface tension followed by pH 5, pH 7, and pH 9. The normalized surface tension profiles of pH 3 and pH 5 and the ones of pH 7 and pH 9 were close to each other. Regarding α-lactalbumin, the most

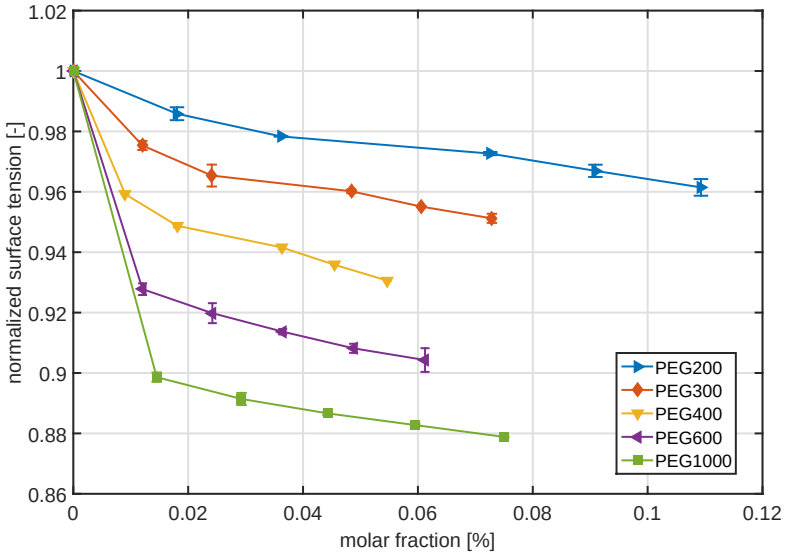


Figure 6.3: Normalized surface tension profiles of PEG varying in molecular weight. Surface tensions were normalized to ultrapure water. The error bars correlate to the 95 % confidence interval ($\pm$ 2σ).

distinctive decrease of the surface tension was observed at pH 3, followed by pH 5. The slopes of the normalized surface tensions at pH 7 and pH 9 were similar and slightly negative. In contrast to pH 3 and pH 5, the normalized surface tension of pH 7 and pH 9 followed a rather linear trend within the studied concentration range. In comparison to BSA and α-lactalbumin the influence of pH on the surface tension of human lysozyme was clearly lower, whereas the strongest decrease was observed at pH 3 again.

Lysozyme in contrast to BSA, human lysozyme, and α-lactalbumin, and human lysozyme showed less impact on the surface tension. For lysozyme a significant reduction in surface tension could only be observed at molar fractions 300 times higher. The profiles were similar for all pH values with the strongest reduction of surface tension to 0.92 at pH 3 for a molar fraction of 10^{-2} % within the studied concentration range.

In order to derive a hydrophobicity scale from the surface tension profiles they were fitted to equation 6.5

$$\gamma_{norm} = 1 - \frac{d \cdot e \cdot (\tilde{x} + c)}{1 + e \cdot (\tilde{x} + c)}, \tag{6.5}$$

where γ_{norm} stands for the normalized surface tension, $\tilde{x}$ represents the molar protein fraction. The fitting parameters d and e could be calculated for all profiles within

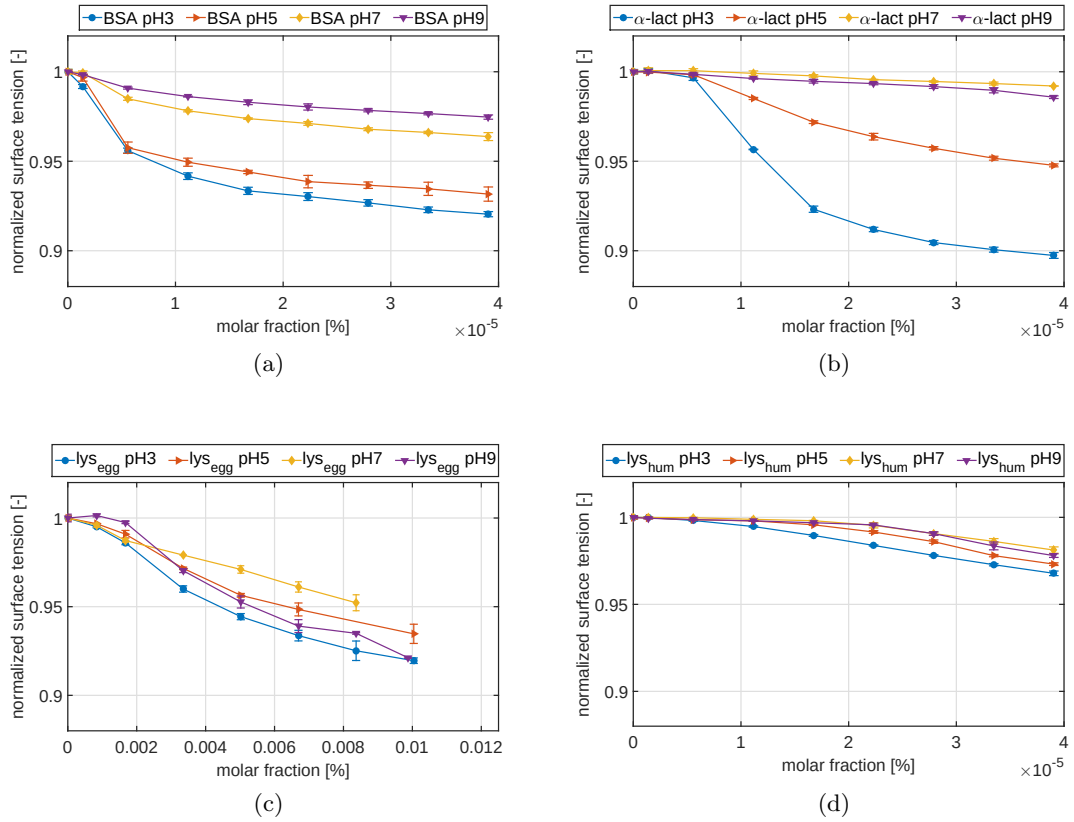


Figure 6.4: Surface tension profiles of (a) bovine serum albumin (BSA), (b) α-lactalbumin (α-lact), (c) lysozyme (lys_{egg}), and (d) human lysozyme (lys_{hum}). The error bars refer to the 95 % confidence interval ($\pm\, 2\sigma$).

this study with a coefficient of correlation larger than 0.98. Exclusively in case of α-lactalbumin the normalized surface tension of pure buffer and the lowest concentrated sample were excluded from the calculation, which was considered by setting the parameter c properly. For all other proteins, parameter c was set to zero. Equation 6.5 describes a saturation function of the normalized surface tension regression, where $d \cdot e$ can be interpreted as the surface activity of the protein and e as the maximal regression of the normalized surface tension. By means of the surface activity, the studied proteins can be ranked according to their hydrophobic character as illustrated in Figure 6.5.

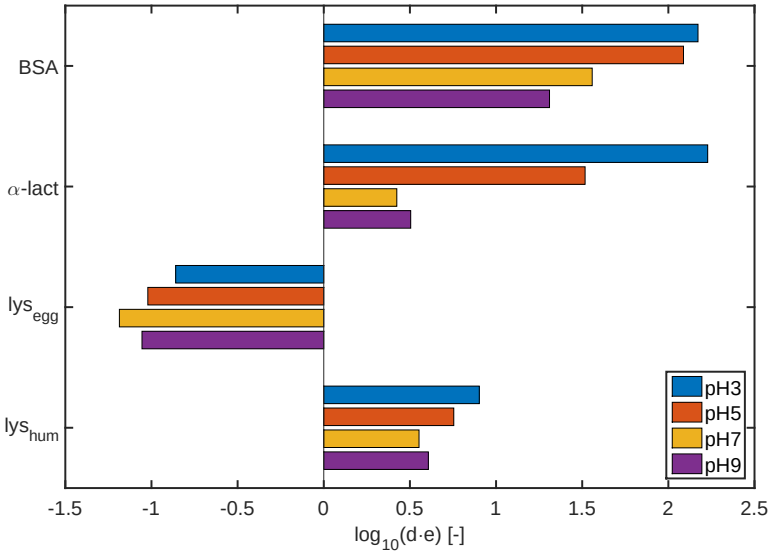


Figure 6.5: Hydrophobicity ranking of bovine serum albumin (BSA), α-lactalbumin (α-lact), lysozyme (lys_{egg}), and human lysozyme (lys_{hum}) by means of the surface activity.

6.3.3 Spectrophotometric Determination of Hydrophobicity

The shift of the absorption maximum ΔA_{max} of BPB Na in presence of BSA was measured at pH 5, pH 7, and pH 9 and is illustrated in Figure 6.6. There is a clear pH dependency of ΔA_{max} as function of the BSA molar fraction. The change of ΔA_{max} in the range between 0 and $1 \cdot 10^{-5}$ % was strongest for pH 5, followed by pH 7 and 9. The same order applied for the upper limit of ΔA_{max}. The data points could be fitted using equation 6.4 with a coefficient of correlation better than 0.97. This dye method was also used to compare different proteins. In addition to BSA, human lysozyme, α-lactalbumin, and lysozyme were investigated at pH 5, 7, and 9. For these proteins the molar fractions had to be increased 150 fold compared to BSA in order to identify significant shifts of the absorption maximum. For all of the investigated proteins a pH dependency of the absorption maximum shift was detected. A comparison between the four proteins at pH 7 is exemplarily shown in Figure 6.7. BSA showed a significantly steeper slope and higher upper limit of the fitted curve compared to human lysozyme, α-lactalbumin, and lysozyme. Apart from BSA the highest values of ΔA_{max} were reached by human lysozyme and α-lactalbumin. However, the slope of the fitted curve of lysozyme was steepest.

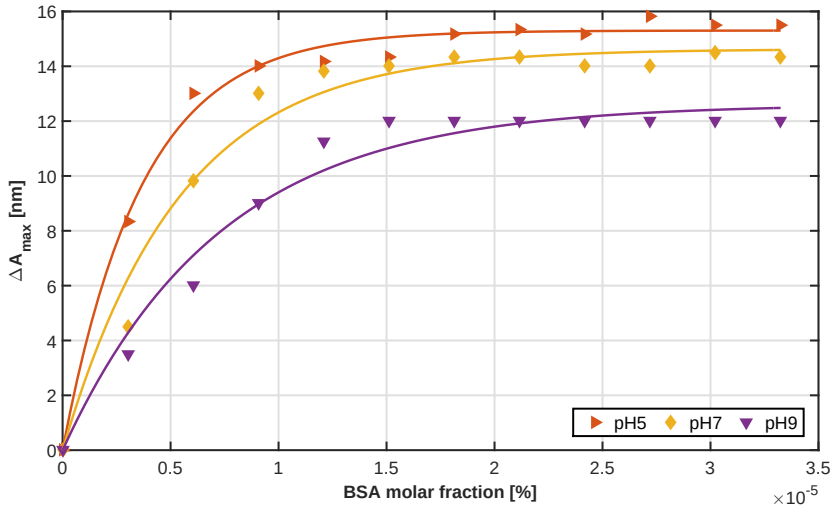


Figure 6.6: Comparison of absorption maximum shift (ΔA_{max}) of BPB Na in presence of bovine serum albumin (BSA) at pH 5, 7, and 9 and its corresponding fits.

6.4 Discussion

6.4.1 Development of HTP Compatible Stalagmometric Method for Surface Tension Determination

In the presented work a highly accurate, robust and HTP compatible stalagmometric method was established onto a liquid handling station using a high precision mass balance and two custom made subunits, the *Tip2World* interface and the stalagmometric setup in particular. Both subunits can be produced using 3D printing technology with a high precision and to a reasonable price. The method was validated with solutions containing NaCl or ethylene glycol. According to literature NaCl increases the surface tension of water [163]. Due to its lower polar character compared to water ethylene glycol exposes attractive interactions with the air-water interface and thus reduces its surface tension. With these two additives we were able to show the validity of the stalagmometric method for a wide range of surface tension values from 57 to 82 mN/m. The sample throughput of the used design was about 18 samples per hour. As mentioned before this results in an operation time of 5.5 hours for analyzing a set of 96 samples. It is important to note, that in this work we only used one tip of the liquid handling arm at a time and the setup

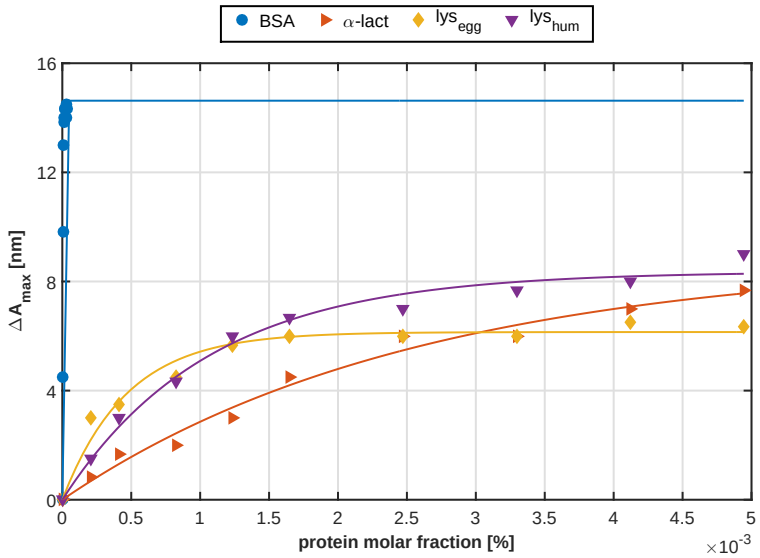


Figure 6.7: Comparison of absorption maximum shift (ΔA_{max}) of BPB Na in presence of bovine serum albumin (BSA), α -lactalbumin (α-lact), lysozyme (lys_{egg}) and human lysozyme (lys_{hum}) at pH 7 and its corresponding fits.

allows the integration of multiple stalagmometric devices. Thus, using the full span of 8 tips of the liquid handling arm in combination with eight stalagmometric devices in parallel would speed up the measurement 8 fold resulting in a throughput of 144 samples per hour.

6.4.2 Stalagmometric Determination of Hydrophobicity

6.4.2.1 Polymers' Hydrophobicity

To examine the correlation between surface tension and hydrophobicity PEG of different molecular weight was used. PEG species of higher molecular weight are known to be more hydrophobic [52]. From our stalagmometric measurements we observed a stronger decrease of surface tension with increasing molecular weight and thus conclude that this is caused by the increase of hydrophobicity. This is in agreement with the findings of Genest et al. [157] on hydrophobically modified polyelectrolytes that also exhibited lower surface tension values the more hydrophobic they were. Thus, the presented stalagmometric method is capable of deducing an analytes' hydrophobicity from its influence on

the surface tension. Distinct differences in surface tensions and thus in hydrophobicity could even be identified for small changes in molecular weight by 100 Da, as could be seen for PEG200, PEG300 and PEG400 in Figure 6.3. This highlights the sensitivity of this stalagmometric approach. Moreover, the change of surface tension dependent on PEG molar fraction demonstrates the high resolution and precision of the developed HTP stalagmometric method. As it could be proved that the here designed HTP stalagmometric setup is able to capture polymer hydrophobicity with a high sensitivity, resolution and precision and was applied to protein solutions.

6.4.2.2 Proteins' Hydrophobicity

In order to estimate the influence of protein species and pH we studied the surface tension increments of lysozyme, human lysozyme, BSA and α-lactalbumin at pH 3, 5, 7, and 9. For each protein we observed an influence of pH on the surface tension increment. This is reasonable as the pH influences the protonation of ionizable amino acids and thus the charge distribution on the protein surface what causes an impact on hydrophobic surface patches. Like for PEG an increase in hydrophobic surface character favors the interaction of the protein molecule with the less polar air-water interface. This results in a decrease of the surface tension. The derived ranking for the investigated proteins and pH values starts with α-lactalbumin at pH 3 as the sample with the most pronounced hydrophobic character. This pronounced hydrophobic character of α-lactalbumin at pH 3 can be explained by transformation of α-lactalbumin into a so called molten globule state where the protein is partially unfolded and inner hydrophobic patches are exposed on the protein surface [164]. The ranking continues with decreasing hydrophobic character as can be seen in Figure 6.5. For each of the investigated proteins the hydrophobic character is highest at pH 3. This might be due to a modification of the charge distribution on the protein surface resulting from the protonation of amino acids. For example at pH 3 the amino acids Glu and Asp are protonated in contrast to pH 5. The protonation of the respective amino acids results in a loss of a negative charge and leads to a more pronounced hydrophobic character of the surface patches and a less favored hydration [165]. In addition, the protonation of these amino acids influences the H-bond-network and thus could lead to minor effects on the formation of secondary structures such as α-helix and β-sheet. These minor structural changes might result in slightly different interactions with the solvent, which we were able to resolve with the presented approach. Major structural changes such as partial unfolding due to the acidic pH which would po-

tentially result in a much stronger influence on hydrophobicity and thus surface tension could be excluded by FT-IR analysis (data not shown). Highest hydrophobic character was not always exhibited nearest to the isoelectric points of the investigated proteins as was shown for lysozyme and human lysozyme. Both exhibited the highest hydrophobic character far away from their isoelectric points. Thus, protein surface charge also influences surface tension but is not the critical factor.
It is important to point out that the differences between the surface tension profiles and thus the pH dependent hydrophobic character of lysozyme were much smaller compared to BSA, α-lactalbumin and human lysozyme. Additionally, significant changes for the normalized surface tension could be found only at molar fractions 300 times higher than for the other proteins. Hence, lysozyme occurred to be much less hydrophobic compared to BSA, α-lactalbumin and human lysozyme. This significant different hydrophobic character of lysozyme is in agreement with Bigelow [166] who found BSA and α-lactalbumin to be very hydrophobic and lysozyme to be only weakly hydrophobic. The discrepancy between the structural similar lysozyme species (root means square deviation of backbone atoms less than 1 Å) underlines the strong impact of the amino acid composition on the protein surface (sequence identity less than 65%).
In summary, the presented HTP stalagmometric setup emerged as promising approach for deducing the hydrophobic character of whole protein molecules in their three-dimensional conformation and in aqueous solution. Differences in the hydrophobic character depending on pH and protein species could be resolved.

6.4.3 Spectrophotometric Determination of Hydrophobicity

The dye measurements for BSA resulted in the highest hydrophobicity at pH 5, followed by pH 7 and 9. This hydrophobicity order is in agreement with our findings using the stalagmometric method. For lysozyme, human lysozyme and α-lactalbumin we had to increase the molar fraction 150 fold to see a distinct and reproducible shift of the absorption maximum. Despite the high molar fractions ΔA_{max} of BPB Na caused by lysozyme, human lysozyme and α-lactalbumin reached only half the value of ΔA_{max} for BSA. Additionally, the course of ΔA_{max} depending on lysozyme, human lysozyme and α-lactalbumin molar fraction were very similar.
Though BPB and BPB Na are well-known polarity sensitive dyes [25] they were so far mainly used for investigation of BSA and HSA [25, 167–173]. It has been shown before, that the affinity for BPB and BPB Na is highly dependent on protein species [167, 173,

174]. Cao et al. [167] observed a great difference between the magnitude of the redshift of the BPB absorption maximum by BSA and Chitosan, a biopolymer 6 times smaller than BSA. However, protein size is not the only factor influencing the BPB-protein interaction. Investigations of Flores [174] and Waldmann-Meyer and Schilling [173] showed that even big proteins like ovalbumin and γ-globulin exhibited a significantly lower affinity for BPB and thus a significantly lower redshift in consequence. Likewise, small proteins like cytochrome c and myoglobin [174, 175] were shown to cause pronounced redshifts. Consequently, the BPB Na method seems to be feasible for protein molecules with a high affinity to BPB Na like BSA. In this case a pH dependent hydrophobic character could be resolved. For lysozyme, human lysozyme and α-lactalbumin the dye method turned out to be inappropriate. For these proteins the dye method yielded in very similar results independent of the protein species and a low hydrophobic character compared to BSA, which is in disagreement with literature [166] and the results of our stalagmometric method. Thus, spectrophotometric determination of protein hydrophobicity is highly dependent on protein-dye interactions and not an universally applicable method [26]. In contrast, the stalagmometric method is able to characterize small hydrophobic and hydrophilic proteins and no limitations regarding solution composition need to be considered.

6.5 Conclusions

In the present work we have developed a high throughput stalagmometric method which is able to measure surface tensions in a highly accurate way and can be operated by liquid handling stations and thus be integrated into high throughput work flow. This method was used to develop an innovative non-invasive approach for characterization of protein hydrophobicity on base of its impact on surface tension. The correlation between surface activity and hydrophobicity was validated using PEG of different molecular weights and applied to four different proteins, namely lysozyme, human lysozyme, BSA, and α-lactalbumin at four pH values. It was possible to rank protein hydrophobicity in dependency of the pH. Lysozyme was found to be hydrophilic, whereas α-lactalbumin turned out to be the most hydrophobic at pH 3. The derived ranking was in good agreement with literature and theoretical considerations regarding pH depending charge distributions.

The stalagmometric method was compared to an orthogonal and established spectrophotometric method for estimating protein hydrophobicity. Results of the spectrophotomet-

ric method regarding pH dependency of BSA were in agreement with the stalagmometric method. Spectrophotometric results for lysozyme, human lysozyme, and α-lactalbumin could not be used to derive protein hydrophobicity, as only BSA caused reasonable shifts of the absorption maximum of BPB Na. Dye based methods are often restricted by pH and protein size and are highly influenced by the aromatic and aliphatic nature of the dye molecule.
Contrary, the stalagmometric method is non-invasive and not restricted by pH and protein size. Differences in the hydrophobic character of proteins depending on protein species and pH could be resolved. Thus, we are convinced that the presented approach is an appropriate way to determine protein hydrophobicity in a non-invasive and highly accurate way.

6.6 Acknowledgment

We thank Matthias Franzreb and Jonas Wohlgemuth for performing the 3D printing. This research work is part of the projects 'Molecular Interaction Engineering: From Nature's Toolbox to Hybrid Technical Systems' and 'Proteinaggregation bei der Herstellung moderner Biopharmazeutika', which is funded by the German Federal Ministry of Education and Research (BMBF)(031A095B and 0315342B).

6.7 References

21. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157,** 105–132 (1982).

25. Bertsch, M., Mayburd, a. L. & Kassner, R. J. The identification of hydrophobic sites on the surface of proteins using absorption difference spectroscopy of bromophenol blue. *Anal. Biochem.* **313,** 187–95 (2003).

26. Alizadeh-Pasdar, N. & Li-Chan, E. C. Comparison of protein surface hydrophobicity measured at various pH values using three different fluorescent probes. *J. Agric. Food Chem.* **48,** 328–334 (2000).

51. Asenjo, J. A. & Andrews, B. A. Aqueous two-phase systems for protein separation: A perspective. *J. Chromatogr. A* **1218,** 8826–8835 (2011).

52. Harris, J. M. *Poly(Ethylene Glycol) chemistry: Biotechnical and biomedical applications* (Springer, 1992).

53. Andrews, B. A., Schmidt, A. S. & Asenjo, J. A. Correlation for the partition behavior of proteins in aqueous two-phase systems: Effect of surface hydrophobicity and charge. *Biotechnol. Bioeng.* **90,** 380–390 (2005).

119. Tanford, C. Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J. Am. Chem. Soc.* **84,** 4240–4247 (1962).

120. Dill, K. A. Dominant forces in protein folding. *Biochemistry* **29,** 7133–55 (1990).

121. Brems, D. N., Plaisted, S. M., Havel, H. a. & Tomich, C. S. Stabilization of an associated folding intermediate of bovine growth hormone by site-directed mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **85,** 3367–3371 (1988).

122. Nieba, L., Honegger, A., Krebber, C. & Pluckthun, A. Disrupting the hydrophobic patches at the antibody variable/constant domain interface: Improved in vivo folding and physical characterization of an engineered scFv fragment. *Protein Eng.* **10,** 435–44 (1997).

123. Diederich, P., Amrhein, S., Hämmerling, F. & Hubbuch, J. Evaluation of PEG/phosphate aqueous two-phase systems for the purification of the chicken egg white protein avidin by using high-throughput techniques. *Chem. Eng. J.* **104,** 945–956 (2013).

124. Janson, J.-C. *Protein purification: Principles, high resolution methods, and applications* (John Wiley & Sons, 2012).

125. Janin, J. Surface and inside volumes in globular proteins. *Nature* **277,** 491–492 (1979).

126. Eisenberg, D. Three-dimensional structure of membrane and surface proteins. *Annu. Rev. Biochem. Biochem.* **53,** 595–623 (1984).

127. Black, S. D. & Mould, D. R. Development of hydrophobicity parameters to analyze proteins which bear post- or cotranslational modifications. *Anal. Biochem.* **193,** 72–82 (1991).

128. Rose, G. D. & Wolfenden, R. Hydrogen bonding, hydrophobicity, packing, and protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **22,** 381–415 (1993).

129. Whitney, P. L., Tanford, C., W, P. L. & Tanford, C. Solubility of amino acids in aqueous urea solutions and its implications for the denaturation of proteins by urea. *J. Biol. Chem.* **237,** 1735–1737 (1962).

130. Nozaki, Y. & Tanford, C. The solubility of amino acids and related compounds in aqueous urea solutions. *J. Biol. Chem.* **238,** 4074–4081 (1963).

131. Nozaki, Y. & Tanford, C. The solubility of amino acids and related compounds in equeous ethylene glycol solutions. *J. Biol. Chem.* **240,** 3568–3573 (1965).

132. Nozaki, Y. & Tanford, C. The solubility of amino acids, diglycine, and triglycine in aqueous guanidine hydrochloride solutions. *J. Biol. Chem.* **245,** 1648–1652 (1970).

133. Nozaki, Y. & Tanford, C. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. *J. Biol. Chem.* **246,** 2211–2217 (1971).

134. Dooley, K. H. & Castellino, F. J. Solubility of amino acids in aqueous guanidinium thiocyanate solutions. *Biochemistry* **11,** 1870–1874 (1972).

135. Fendler, J. H., Nome, F. & Nagyvary, J. Compartmentalization of amino acids in surfactant aggregates. *J. Mol. Evol.* **6,** 215–232 (1975).

136. Radzicka, A. & Wolfenden, R. Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry* **27,** 1664–1670 (1988).

137. Shanbhag, V. P. & Axelsson, C.-G. Hydrophobic interaction determined by partition in aqueous two-phase systems. *Eur. J. Biochem.* **60,** 17–22 (1975).

138. Keshavarz, E. & Nakai, S. The relationship between hydrophobicity and interfacial tension of proteins. *Biochim. Biophys. Acta* **576,** 269–279 (1979).

139. Kato, A. & Nakai, S. Hydrophobicity determined by a fluorescence probe method and its correlation with surface properties of proteins. *Biochim. Biophys. Acta* **624,** 13–20 (1980).

140. Cardamone, M. & Puri, N. K. Spectrofluorimetric assessment of the surface hydrophobicity of proteins. *Biochem. J.* **282,** 589–93 (1992).

141. Hendriks, J. *et al.* Transient exposure of hydrophobic surface in the photoactive yellow protein monitored with Nile Red. *Biophys. J.* **82,** 1632–1643 (2002).

142. Hawe, A., Sutter, M. & Jiskoot, W. Extrinsic fluorescent dyes as tools for protein characterization. *Pharm. Res.* **25,** 1487–1499 (2008).

143. Salgado, J. C., Rapaport, I. & Asenjo, J. A. Is it possible to predict the average surface hydrophobicity of a protein using only its amino acid composition? *J. Chromatogr. A* **1075,** 133–143 (2005).

144. Miller, S., Janin, J., Lesk, A. M. & Chothia, C. Interior and surface of monomeric proteins. *J. Mol. Biol.* **196,** 641–56 (1987).

145. Lijnzaad, P., Berendsen, H. J. & Argos, P. A method for detecting hydrophobic patches on protein surfaces. *Proteins* **26,** 192–203 (1996).

146. Chennamsetty, N. *et al.* Design of therapeutic proteins with enhanced stability. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 11937–11942 (2009).

147. Salgado, J. C., Rapaport, I. & Asenjo, J. A. Predicting the behaviour of proteins in hydrophobic interaction chromatography: 1: Using the hydrophobic imbalance (HI) to describe their surface amino acid distribution. *J. Chromatogr. A* **1107,** 110–119 (2006).

148. Salgado, J. C., Rapaport, I. & Asenjo, J. A. Predicting the behaviour of proteins in hydrophobic interaction chromatography: 2. Using a statistical description of their surface amino acid distribution. *J. Chromatogr. A* **1107,** 120–129 (2006).

149. Mahn, A., Lienqueo, M. & Asenjo, J. Effect of surface hydrophobicity distribution on retention of ribonucleases in hydrophobic interaction chromatography. *J. Chromatogr. A* **1043.** 23rd International Symposium on the Separation of Proteins, Peptides and Polynucleotides, 47–55 (2004).

150. Chen, J. *et al.* Investigation of protein retention in hydrophobic interaction chromatographic (HIC) systems using the preferential interaction theory and quantitative structure property relationship models. *Reactive and Functional Polymers* **67.** Special Issue Dedicated to Professor Helfferich, 1561–1569 (2007).

151. Amrhein, S., Oelmeier, S. A., Dismer, F. & Hubbuch, J. Molecular dynamics simulations approach for the characterization of peptides with respect to hydrophobicity. *J. Phys. Chem. B* **118,** 1707–14 (2014).

152. Reißer, S., Strandberg, E., Steinbrecher, T. & Ulrich, A. S. 3D hydrophobic moment vectors as a tool to characterize the surface polarity of amphiphilic peptides. *Biophys. J.* **106,** 2385–2394 (2014).

153. Trinquier, G. & Sanejouand, Y.-H. Which effective property of amino acids is best preserved by the genetic code? *Protein Eng.* **11,** 153–169 (1998).

154. Biswas, K. M., DeVido, D. R. & Dorsey, J. G. Evaluation of methods for measuring amino acid hydrophobicities and interactions. *J. Chromatogr. A* **1000,** 637–655 (2003).

155. Bull, H. B. & Breese, K. Surface tension of amino acid solutions: A hydrophobicity scale of the amino acid residues. *Arch. Biochem. Biophys.* **161,** 665–670 (1974).

156. Absolom, D. R., Zingg, W. & Neumann, A. W. Protein adsorption to polymer particles: Role of surface properties. *J. Biomed. Mater. Res.* **21,** 161–71 (1987).

157. Genest, S. *et al.* Characterization of highly substituted, cationic amphiphilic starch derivatives: Dynamic surface tension and intrinsic viscosity. *Starch/Stärke* **65,** 999–1010 (2013).

158. Tate, T. XXX. On the magnitude of a drop of liquid formed under different circumstances. *Philos. Mag.* **27,** 176–180 (1864).

159. Amrhein, S., Schwab, M.-L., Hoffmann, M. & Hubbuch, J. Characterization of aqueous two phase systems by combining lab-on-a-chip technology with robotic liquid handling stations. *J. Chromatogr. A* **1367,** 68–77 (2014).

160. Tsierkezos, N. G. & Molinou, I. E. Thermodynamic properties of water + ethylene glycol at 283.15, 293.15, 303.15, and 313.15 K. *J. Chem. Eng. Data* **43,** 989–993 (1998).

161. Jańczuk, B., Białopiotrowicz, T. & Wójcik, W. The components of surface tension of liquids and their usefulness in determinations of surface free energy of solids. *J. Colloid Interface Sci.* **127,** 59–66 (1989).

162. Melinder, Å. *Thermophysical properties of aqueous solutions used as secondary working fluids* PhD thesis (School of Industrial Engineering and Management, Royal Institute of Technology, KTH Stockholm, 2007).

163. Pegram, L. M. & Record, M. T. Hofmeister salt effects on surface tension arise from partitioning of anions and cations between bulk water and the air-water interface. *J. Phys. Chem. B* **111,** 5411–5417 (2007).

164. Permyakov, E. A. & Berliner, L. J. alpha-Lactalbumin: Structure and function. *FEBS Lett.* **473,** 269–274 (2000).

165. Guo, D. *et al.* Prediction of peptide retention times in reversed-phase high-performance liquid chromatography I. Determination of retention coefficients of amino acid residues of model synthetic peptides. *J. Chromatogr. A* **359,** 499–518 (1986).

166. Bigelow, C. C. On the average hydrophobicity of proteins and the relation between it and protein structure. *J. Theor. Biol.* **16,** 187–211 (1967).

167. Cao, W. G. *et al.* Mechanism of the interaction between bromophenol blue and bovine serum albumin. *Spectrosc. Lett.* **36,** 197–209 (2003).

168. Wei, Y.-J., Li, K.-A. & Tong, S.-Y. The interaction of bromophenol blue with proteins in acidic solution. *Talanta* **43,** 1–10 (1996).

169. Tayyab, S. & Qasim, M. Binding of bromophenol blue to bovine serum albumin and its succinylated forms. *Int. J. Biol. Macromol.* **12,** 55–58 (1990).

170. Murakami, K., Sano, T. & Yasunaga, T. Kinetic studies of the interaction of bromophenol blue with bovine serum albumin by pressure-jump method. *Bull. Chem. Soc. Jpn.* **54,** 862–868 (1981).

171. Kragh-Hansen, U., Møller, J. V. & Lind, K. E. Relation between binding of phenolsulfophtalein dyes and other ligands with a high affinity for human serum albumin. *Biochim. Biophys. Acta* **365,** 360–371 (1974).

172. Bjerrum, O. J. Interaction of bromphenol blue and bilirubin with bovine and human serum albumin determined by gel filtration. *Scand. J. Clin. Lab. Investig.* **22,** 41–48 (1968).

173. Waldmann-Meyer, H. & Schilling, K. The interaction of bromophenol blue with serum albumin and γ-globulin in acid medium. *Arch. Biochem. Biophys.* **64,** 291–301 (1956).

174. Flores, R. A rapid and reproducible assay for quantitative estimation of proteins using bromophenol blue. *Anal. Biochem.* **88,** 605–611 (1978).

175. Mayburd, A. L., Tan, Y. & Kassner, R. J. Complex formation between chromatium vinosum ferric cytochrome c' and bromophenol blue. *Arch. Biochem. Biophys.* **378,** 40–44 (2000).

CHAPTER 7

Determination of Surface Tension of Liquids by Means of Liquid Handling Stations

Sven Amrhein, Susanna Suhm, and Jürgen Hubbuch

Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Science, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany

Engineering in Life Sciences (2016), accepted

Abstract

The characterization of liquid surfaces with respect to surface tension is of major interest throughout a number of disciplines. In life science technologies and in pharmaceutical production in particular, the surface tension and drop size of liquids are predominating parameters throughout the production process, starting from foaming during fermentation processes, formulation by spray drying or the drug application by aerosol inhalators. The profiling of surface tension can be further applied for physicochemical drug assessment with predictive power for the compound's pharmacology. In the present study, a high throughput approach for the determination of surface tension integrated into a fully automated liquid handling station was developed. The method bases on the accurate gravimetric determination of masses of drops generated at a liquid handler tip using a high precision balance. By means of repetitive sample conserving drop generating procedures, huge numbers of drops and thus statistical significance can be created from a minimal sample volume of a few 100 µL. The developed approach excels in instrumental simplicity, accuracy, precision, and the minimal sample consumption. The fully automated setup was validated for a broad range of surface tensions starting from about 25 to 75 mN/m. Eightfold determinations of sample liquids exposed standard deviations of less than 0.5 % which demonstrates excellent precision. Further potential revisions of the stalagmometric approach for the determination of interfacial tension between two liquids are described in detail. The employment of liquid handling stations enables the integration of the developed method into the high throughput screening paradigm and thus adds high value to the laboratory work flow.

Keywords: *Automation; High Throughput; Liquid Handling Station; Interfacial Tension; Surface Tension*

7.1 Introduction

Surface tension is a predominating parameter for the behavior of liquids in a number of phenomena and processes found in a huge range of industry fields such as adhesion and coating, cleaning, foaming, surface treatment, wetting, surface rheology, food and beverages, or medical and pharma. The investigation of surface tensions of drug solutions arose high interest in pharmaceutical research since the surface activity properties have been shown to give unique insight into drug properties and its pharmacology along the ADMET disposition starting from absorption, distribution, metabolism, excretion, and toxicity with predictive power. In this scope, surface tension profiling has been successfully applied for the determination of drug solubility [176] or even for pharmacological properties such as blood–brain barrier penetration [177–179], oral in-vivo efficacy [180, 181], and drug-release [182, 183]. The physicochemical properties of compounds can be assessed by profiling the surface tension as a function of compound concentration. Based on this profiling high valuable amphiphilicity parameters such as true surface area (TSA), critical micelle concentration (CMC), and hydrophobicity can be derived from further application of adsorption models. Moreover, surface tension profiling was applied for molecular protein assessment regarding protein surface hydrophobicity, which is a predominating parameter for the behavior of biopharmaceuticals in aqueous solutions [184]. In summary, the surface tension is a highly important parameter for numberless applications and throughout research disciplines. There are different approaches to determine the surface tension of a solution. For instance, there is the du Noüy ring, Wilhelmy plate, or rod method, which correlate the liquid's surface tension to the force needed to withdraw an optimally wettable ring, plate, or rod respectively from the solution. The bubble pressure method derives the surface tension based on the maximum internal pressure of a gas bubble which is formed in a liquid by means of a capillary. The contact angle of a drop on a solid surface or the shape and the volume of a drop formed at a vertical capillary can be optically assessed and used to derive the surface tension which are known as the sessible drop or pendant drop methods. However, the accurate optical assessment of the drop shape requires sophisticated algorithms. Due to the instrumental simplicity of a stalagmometer we have developed a stalagmometric setup in a previous study [184], which determines the surface tension on base of the mass of drops which are generated at a vertical capillary by means of a fully automated liquid handling station (LHS) and are measured by a high performance balance. The drive for optimization of this method in terms of minimization of sample consumption and instrumental require-

ments results in this study which aims at the development of an instrumentally simple high throughput approach for the determination of surface tension by means of a LHS. These robotic instruments are widely spread in academic and industrial laboratories since fully automated, miniaturized, and highly parallel high throughput screening methods became well-established and appreciated [185]. High performance balances are simple to implement into a LHS for a reasonable price. The developed setup and experimental procedure are described and validated in detail in the following.

7.2 Materials and Methods

7.2.1 Considerations for the Measurement of Surface Tension Based on Drop Mass

The developed setup is able to determine surface tensions of liquids. The approach bases on the correlation between drop mass and its interfacial tension with the surrounding medium as described in the following. The assessment of the surface tension bases on the equilibrium of the weight force F_W and the adhesive force F_A when the drop tears off. The weight force of the drop F_W at the moment of drop detachment from the tip with an outer radius r of the sample solution can be correlated to its surface tension γ according to the Tate's law [158] in equation 7.1.

$$F_W = F_A = 2\pi r\gamma \tag{7.1}$$

Since the instrumental setup influences the drop size and the drop detachment, the registered weight force of the actual drop falling from the capillary on the balance will be $F_W^{'}$ which differs from the ideal weight force F_W. $F_W^{'}$ and F_W are correlated by an instrumental setup correction factor f_inst. The correction factor was shown to be applicable over a range of $0.54 < f_{inst} < 0.94$ by Harkins and Brown [186] and Wilkinson [187]. By means of the correction factor the surface tension γ can be expressed as stated in equation 7.2

$$\gamma = \frac{F'_W}{2\pi r f_{inst}} = \frac{m_{drop} g}{2\pi r f_{inst}}, \tag{7.2}$$

where m_{drop} is the mass of the drop, g is the acceleration of gravity, and r is the outer radius of the drop generating capillary or liquid handler tip, respectively. By means of a reference sample with a known surface tension γ_{ref} resulting in a specific drop mass $m_{drop,ref}$ the correction factor can be determined and the surface tension γ_{sample} of the liquid of interest can be determined by applying equation 7.3 to its specific drop mass $m_{drop,sample}$.

$$\gamma_{sample} = m_{drop,sample} \frac{\gamma_{ref}}{m_{drop,ref}} \tag{7.3}$$

7.2.2 Experimental Procedure

The measurement of the surface tension bases on the accurate determination of drop masses and the application of equation (3). The drops were generated by liquid handler tips and their masses were recorded by a high precision balance. A high precision weight unit (WXSS205DU DualRange, Mettler Toledo, Greifensee, Switzerland) was transformed into a stalagmometer by constructing a carrier using the 3D CAD software Solid Edge ST8 (Siemens PLM Software, Plano, TX, USA). The carrier enables to mount a standard micro titer plate on the balance port. It can be fixed onto a weighting pan supplied by Mettler Toledo by means of three M3 screws. The pan is fixed into the weighting cell. The setup is illustrated in exploded view in Figure 7.1. The carrier was designed with specific focus on weight minimization and was 3D printed in Nylon (PA) using a fused deposition modeling (FDM) Fabb-It MK2 printer (Fabb-It, Lörrach-Brombach, Germany). The weight of the carrier of less than 20 g enables the use a sufficient sample volume in combination with standard deep well micro titer plates within the high resolution balance operation range in which the balance has a readability of 0.01 mg. The high performance balance was integrated into a fully automated liquid handling station Freedom EVO® 200 (Tecan, Crailsheim, Germany) equipped with stainless fixed tips and 1 mL dilutors. The sample liquid was aspirated with 50 µL/s from a 1 mL 96 round bottom deep well Nunc™ micro titer plate (Thermo Fisher Scientific, Waltham, MA, USA) mounted onto the weigh unit. The tip was slowly withdrawn to 30 mm above

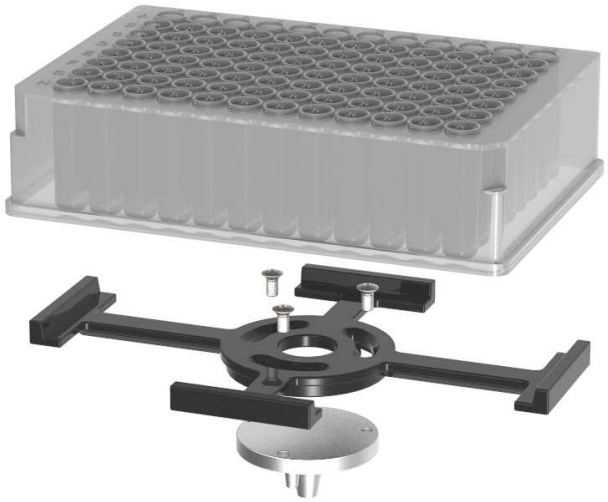


Figure 7.1: Schematic illustration of the construction works in order to transform the WXSS205DU DualRange balance unit into a stalagmometer in exploded view. The carrier holding a standard micro titer plate (96 deep well plate) is fixed onto the weighting pan by means of three M3 screws. The pan is fixed into the weighting cell which is not illustrated here.

the sample liquid level and dispensed with 5 µL/s with tracking the liquid level. The liquid level tracking ensures a constant drop falling height and is thus highly beneficial to the consistent stability of the balance signal. After the full dispense, the liquid was aspirated again and the procedure was repeated. The looped aspirate and dispense procedure minimizes the sample consumption and enables the generation of huge numbers of drops and thus increased sampling sufficiency and statistical validity. The balance signal was recorded continuously during the dispense step using fully automated routines written in Matlab® R2015b (The MathWorks, Natick, MA, USA) applying serial communication protocols with a signaling rate of 20 Hz. As illustrated in Figure 7.2 this drop generating procedure results in sharp steps within the balance profile which enables an easy extraction of drop masses. In order to minimize evaporation during the experiment, the samples were stored in sealed containers and were transferred into the analyzing plate on the balance immediately prior to its analysis and were removed after each analysis from the analyzing plate. Ultra-pure water, purified with a PURELAB™ Ultra (ELGA LabWater, Bucks, UK), dispensed in air was used as reference with a supposed surface tension of 72.62 mN/m [188]. The precision of the drop mass determination was investigated by analyzing 50 samples of ultra-pure water with a volume of 300 µL each and 5-fold repetitive aspiration and dispensing procedures. These extensive data

including minor temperature fluctuations were used to test for normal distribution of the drop mass. For the purpose of validation the developed setup in terms of accuracy, a number of reference liquids with surface tension reported in literature were analyzed. In particular, aqueous ethanol and aqueous 2-propanol solutions with a mass concentration ranging from 10 %(w/w) to 100 %(w/w) [189], aqueous ethylene glycol mixture with a mass concentration ranging from 10 %(w/w) to 90 %(w/w) [161], and aqueous sodium chloride solutions with a mass concentration of 5 %(w/w), 10 %(w/w), and 15 %(w/w) [162] were measured. In addition, pure ethanol [189–193] and ethylene glycol [161, 194] were analyzed. Ultra-pure water and chemicals purchased from Merck (Darmstadt, Germany), were used for sample preparation. Aqueous solutions were prepared by mixing the respective masses of ultra-pure water and the respective additive in order to reach the desired concentration. The masses were detected grammatically by means of an analytical balance BP221S purchased form Sartorius (Göttingen, Germany). The literature surface tension values of ethylene glycol and sodium chloride in dependency of the concentration were fitted to a second degree polynomial to extract the exact literature values [161, 162] for the prepared solution and thus to take preparation inaccuracies into account. Reference values for aqueous ethanol and aqueous 2-propanol solutions [189] in dependency of the concentration and temperature were linearly interpolated to extract values for room temperature. All samples were measured at least 4-fold with a volume of 300 µL and 5 repetitive drop generating steps each.

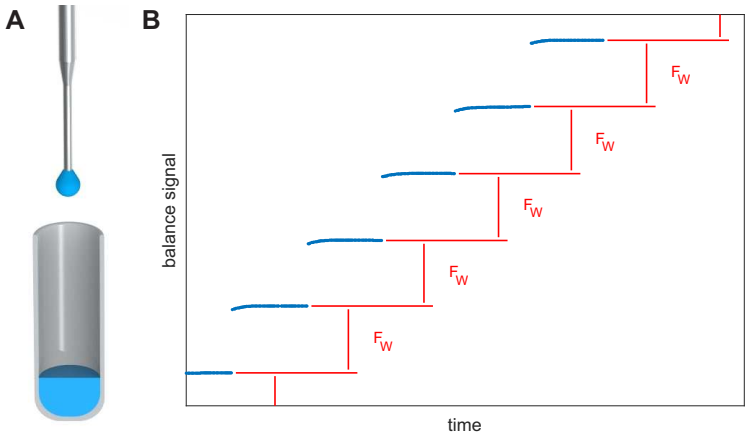


Figure 7.2: Schematic illustration of the method to determine surface tension (A) by means of drops generated above the well with (B) the corresponding balance profile.

7.3 Results

The probability distribution of generated drop masses was investigated by analyzing an experimentally determined data set containing more than 2500 single water drop masses as illustrated in a histogram in Figure 7.3A. For graphical data analysis Figure 7.3B illustrates the quantiles of the drop masses plotted against quantiles of the standardized normal distribution. The normality of generated drop masses has to be verified in order to warrant the validity of confidence intervals for standard deviation and the embedding of outlier analysis and removal within the evaluating algorithm. By means of the Lilliefors test the experimental data were shown to come from a normally distributed population at the 1 % significance level with a mean value of 20.42 mg and standard deviation of σ<0.3 mg, which corresponds to a relative standard deviation of 1.4 %. The developed setup for surface tension determination was successfully validated with

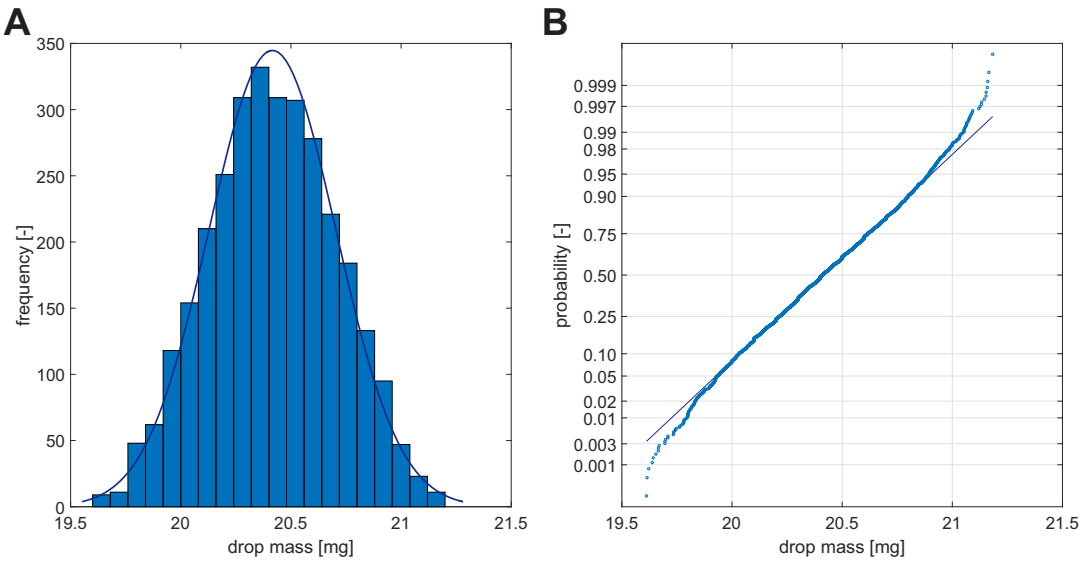


Figure 7.3: Illustration of normal distribution of water drop mass by means generated against air by (A) histogram and (B) probability plot for more than 2500 single water drop masses.

reference values reported in literature as illustrated in Figure 7.4. The error bars refer to 99.73 % confidence interval ($\pm 3\sigma$). Multiple literature values are plotted separately. The derived data exposes a highly sufficient linear correlation (R^2>0.99) with reported literature values with a slope of 1 and an offset of 0.6 mN/m. Small deviation of literature values from the hereby determined surface tension and the offset value are likely due

to minor deviations in temperature and used experimental methods between literature values and hereby derived surface tensions. Generally, the surface tension decreases with the increase of temperature as empirically described by Eötvös [195]. The comparison to the literature values demonstrates the validity of the developed approach over a wide range of surface tensions starting from about 25 into 75 mN/m. The deviations of the eightfold determinations were less than 0.5 % throughout all liquids, which demonstrates the excellent precision.

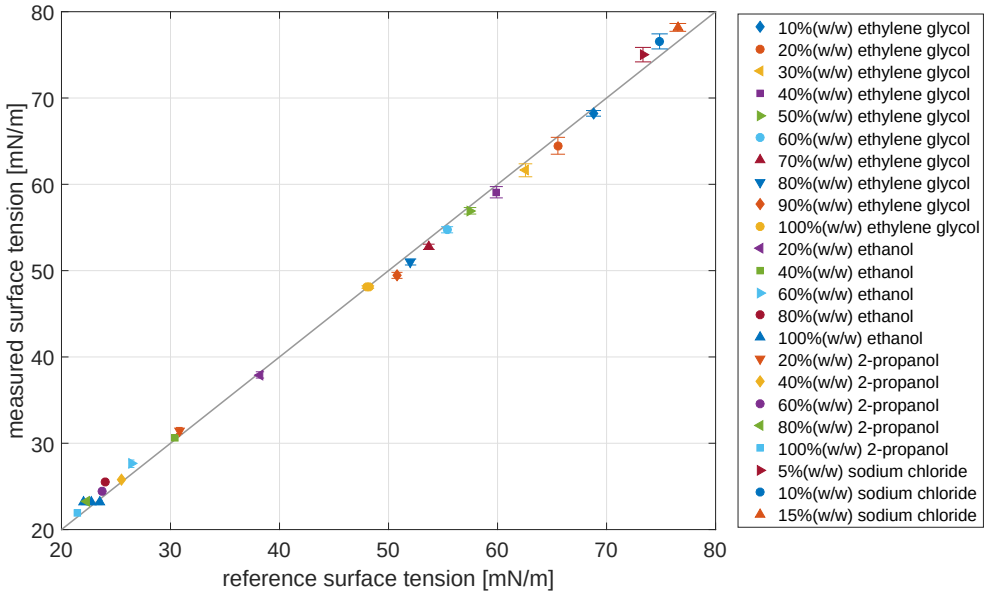


Figure 7.4: Comparison of surface tensions determined in this study and reported in literature. The error bars refer to 99.73 % confidence interval ($\pm 3\sigma$).

7.4 Discussion

It could be shown that the experimentally simple setup is capable to determine surface tensions of liquids accurately and precisely in a fully automated stalagmometric approach based on drop mass measurements. The use of separate experimental compartments in standard micro titer plate format avoids cross contamination and enables a sample recycling and thus the generation of statistically extensive data from a minimum of samples solution. Theoretically, the required sample volume is the volume of one complete drop. This volume can be repetitively aspirated and dispensed. However, due to evaporation

and concerns regarding a proper liquid transfer during aspiration and dispense steps, a small volume excess is required. The dead volume of the micro titer plate is highly dependent on a proper parameterization of the lab ware within the liquid handling station and the geometry of the bottom of the wells. In this case the dead volume can be easily considered to be less than 10 µL for the used round bottom 96 deep well plate. However, it is important to note that the integrity of shear sensitive samples could be affected by repetitive aspirate and dispense procedures. If the shear affects the surface tension, it could be monitored by analyzing the drop mass of the particular sample throughout the experiments. With the chosen flow rate of 5 µL/s and 5 drop generating repetitions per sample (300 µL) the analysis time was less than 6 minutes per sample but can further be reduced by decreasing the number of repetitions and sample volume. However, it has to be considered to which extent a reduction in generated drop number might be at the expense to the statistical significance. The analysis at temperatures varying from room temperature could be realized by using a thermally insulated and controlled weighing chamber. The approach can potentially be adapted for the determination of interfacial tensions between liquids. Therefore a drop of heavier bottom phase liquid has to be generated within the lighter top phase [196, 197]. Thus, an additional force and the buoyancy force (F_B) in particular counteracts the weight force and has to be taken into account when balancing the equilibrium of forces to derive the adhesive force as shown in equation 7.4. Buoyancy force (F_B) depends on the volume of the generated drop (V_{drop}) and the density of the top phase liquid (ρ_{TP}).

$$F_W = F_A + F_B = 2\pi r\gamma + gV_{drop}\rho_{TP} \tag{7.4}$$

The acting forces can be extracted from the resulting balance profile as illustrated in Figure 7.5. However, a significant deviation in density of the two liquids is indispensable in order to evaluate the balance profiles sufficiently. This implies that the tracking of the tip during drop forming within the top phase has to be sufficiently accurate to be capable of compensating the buoyancy force acting on the submersed liquid handler tip. The buoyancy force acting on the drop leads to larger drops in comparison to the drop generated within air. This has to be considered during scaling the phase system and used lab ware. The time point of drop detachment could be extracted from the balance profile as illustrated in Figure 7.5 since the difference between buoyancy force acting during the drop growing phase and the gravity of the drop after sinking from the tip within the well results into a small step within the balance profile.

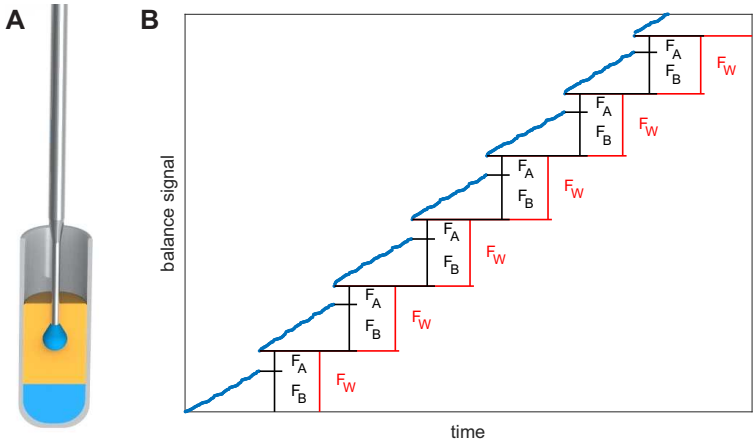


Figure 7.5: Schematic illustration of the method to determine interfacial tensions (A) by means of drops generated within the top phase with (B) the corresponding balance profile.

7.5 Concluding remarks

In this study a fully automated approach for the determination of surface tension was developed, which excels in precision and accuracy. The great asset of this method is the virtual non-existence of instrumental complexity and the possibility to generate a huge number of statistical data of drop masses from a minimum sample volume by repetitive drop generating procedures. The separate experiment containment for each sample using the micro titer plate format in combination with extensive washing and cleaning procedures of the drop generating tip avoids cross contamination of samples and preserves each sample for possible further analytics. As already pointed out, the application of surface tension profiling is of high value for pharmaceutical research and development and this approach can be realized with a minimal instrumental effort and can be integrated into the high throughput screening paradigm employing liquid handling stations. We are confident, that the developed approach will benefit laboratory work flow and underlines the huge potential of liquid handling stations in terms of applicability and utility.

7.6 Acknowledgment

This research work is part of the project „Molecular Interaction Engineering: From Nature's Toolbox to Hybrid Technical Systems", which is funded by the German Federal Ministry of Education and Research (BMBF), funding code 031A095B. We gratefully acknowledge the financial support. The authors bear the complete responsibility for the content of this publication.

7.7 References

106. Rengier, F. *et al.* 3D printing based on imaging data: Review of medical applications. English. *Int. J. Comput. Assist. Radiol. Surg.* **5,** 335–341 (2010).

110. Mironov, V. *et al.* Organ printing: Computer-aided jet-based 3D tissue engineering. *Trends Biotechnol.* **21,** 157–161 (2003).

158. Tate, T. XXX. On the magnitude of a drop of liquid formed under different circumstances. *Philos. Mag.* **27,** 176–180 (1864).

161. Jańczuk, B., Białopiotrowicz, T. & Wójcik, W. The components of surface tension of liquids and their usefulness in determinations of surface free energy of solids. *J. Colloid Interface Sci.* **127,** 59–66 (1989).

162. Melinder, Å. *Thermophysical properties of aqueous solutions used as secondary working fluids* PhD thesis (School of Industrial Engineering and Management, Royal Institute of Technology, KTH Stockholm, 2007).

176. Heikkila, T. *et al.* 96-Well plate surface tension measurements for fast determination of drug solubility. *Lett. Drug Des. Discov.* **5,** 471–476 (2008).

177. Petereit, A. C. *et al.* Prediction of blood-brain barrier penetration of poorly soluble drug candidates using surface activity profiling. *Eur. J. Pharm. Biopharm.* **75,** 405–410 (2010).

178. Seelig, A., Gottschlich, R. & Devant, R. M. A method to determine the ability of drugs to diffuse through the blood-brain barrier. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 68–72 (1994).

179. Suomalainen, P., Johans, C., Söderlund, T. & Kinnunen, P. K. Surface activity profiling of drugs applied to the prediction of blood-brain barrier permeability. *J. Med. Chem.* **47,** 1783–1788 (2004).

180. Crandall, D. *et al.* Characterization and comparative evaluation of a structurally unique PAI-1 inhibitor exhibiting oral in-vivo efficacy. *J. Thromb. Haemost.* **2,** 1422–1428 (2004).

181. Finholt, P. & Solvang, S. Dissolution kinetics of drugs in human gastric juice - the role of surface tension. *J. Pharm. Sci.* **57,** 1322–1326 (1968).

182. Jaitely, V., Karatas, A. & Florence, A. T. Water-immiscible room temperature ionic liquids (RTILs) as drug reservoirs for controlled release. *Int. J. Pharm.* **354,** 168–173 (2008).

183. Vladisavljević, G. T., Shimizu, M. & Nakashima, T. Production of multiple emulsions for drug delivery systems by repeated SPG membrane homogenization: Influence of mean pore size, interfacial tension and continuous phase viscosity. *J. Membr. Sci.* **284,** 373–383 (2006).

184. Amrhein, S., Christin Bauer, K., Galm, L. & Hubbuch, J. Non-invasive high throughput approach for protein hydrophobicity determination based on surface tension. *Biotechnol. Bioeng.* **112,** 2485–2494 (2015).

185. Mayr, L. M. & Bojanic, D. Novel trends in high-throughput screening. *Curr. Opin. Pharmacol.* **9.** Anti-infectives/New technologies, 580–588 (2009).

186. Harkins, W. D. & Brown, F. The determination of surface tension (free surface energy), and the weight of falling drops: The surface tension of water and benzene by the capillary height method. *J. Am. Chem. Soc.* **41,** 499–524 (1919).

187. Wilkinson, M. Extended use of, and comments on, the drop-weight (drop-volume) technique for the determination of surface and interfacial tensions. *J. Colloid Interface Sci.* **40,** 14–26 (1972).

188. Richards, T. W. & Coombs, L. B. the surface tensions of water, methyl, ethyl and isobutyl alcohols, ethyl butyrate, benzene and toluene. *J. Am. Chem. Soc.* **37,** 1656–1676 (1915).

189. Vazquez, G., Alvarez, E. & Navaza, J. M. Surface tension of alcohol water+ water from 20 to 50. degree. C. *J. Chem. Eng. Data* **40,** 611–614 (1995).

190. Sell, P.-J. & Renzow, D. Bestimmung des Benetzungsverhaltens von Pigmenten. *Prog. Org. Coat.* **3,** 323–348 (1975).

191. Fowkes, F. M. Attractive forces at interfaces. *Ind. Eng. Chem.* **56,** 40–52 (1964).

192. Wolf, K. & Wolff, R. Oberflächenspannung und Oberflächenaktivität. *Houben-Weyl: Methoden der organischen Chemie. Bd. III Teil* **1,** 449–479 (1955).

193. Marwedel, G. Zusammenhänge zwischen Oberflächenspannung. *Dichte und Viskosität organischer (besonders lacktechnische) Flüssigkeiten bei übereinstimmenden Temperaturen, Farbe und Lack* **69,** 437–451 (1963).

194. Della Volpe, C. & Siboni, S. Some reflections on acid-base solid surface free energy theories. *J. Colloid Interface Sci.* **195,** 121–136 (1997).

195. Eötvös, R. Über den Zusammenhang der Oberflächenspannung der Flüssigkeiten mit ihrem Molecularvolumen. *Annalen der Physik* **263,** 448–459 (1886).

196. Yadav, J. *Advanced practical physical chemisty* (Krishna Prakashan).

197. Chattoraj, D. *Adsorption and the Gibbs Surface Excess* (Springer US, 2012).

CHAPTER 8

Molecular Dynamics Simulations Approach for the Characterization of Peptides With Respect to Hydrophobicity

Sven Amrhein[a], Stefan Alexander Oelmeier[a], Florian Dismer[b], and Jürgen Hubbuch[a]

[a] *Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Science, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany*
[b] *Novo Nordisk AS, DK-2880 Bagsværd, Denmark*

The Journal of Physical Chemistry B 118.7 (2014): 1707-1714.

Abstract

It has been shown, that molecular dynamics (MD) simulations are a powerful tool to generate knowledge about complex interactions in the field of bioprocess technologies at the atomic level. In this field one of the most important non-specific interactions is the hydrophobic interaction which is still not fully understood after nearly 30 years of research. To date established hydrophobicity scales, which base mostly on proteins' primary structure, are used to estimate the overall hydrophobicity. The structural complexity and the influence of the protein's environment cannot be accommodated with these scales. In this work free solution molecular dynamics simulations were used to investigate the hydrophobic character of low molecular weight peptides. Therefore local densities of a small hydrophobic tracer molecule and unprotonated triethylamine (TEA) in particular were used to localize and quantify hydrophobic patches among the peptide surface. Comparisons between local densities and the retention behavior in reversed phase chromatography showed significant correlations. Moreover, neighbor effects caused by charges could be identified. We were able to show that the developed *in-silico* method is applicable to characterize peptides in respect to hydrophobicity in agreement with experimental data. We are confident to apply this method to larger protein structures.

Keywords: *Reversed Phase HPLC, Tracer Probing, Radial Distribution Function (RDF), In-Silico, Hydropathy*

8.1 Introduction

Hydrophobicity is at the heart of our understanding how proteins behave in solutions in respect to solubility, folding, aggregation or precipitation. In the field of bio-pharmaceuticals hydrophobicity is a key factor from the down streaming processing over the formulation to the final drug application. In downstream processing and formulation, hydrophobicity dominates the retention behavior in hydrophobic interaction and reversed phase chromatography, the partitioning in extraction processes, exemplary aqueous two phase systems (ATPS), and the tendency to forming oligomers in a preferred manner in crystallization steps or in undesirable manner of aggregation and precipitation, which leads to product loss and bears the risk of immune reactions upon applications. The bioavailability of a drug is influenced by solubility and membrane permeability, which depend on hydrophobicity in turn.

Thus the quantification of hydrophobicity is highly relevant for the processibility as well as the pharmacological property of a potential drug substance. Especially with regard to time to market demands and reduction of time and investment for candidate selection, an *in silico* approach for measuring hydrophobicity is highly desirable. One method for the experimental determination of hydrophobicity of a protein is precipitation [49, 198]. In this case proteins' solubility is influenced by the concentration of a precipitation species. The critical concentration at which protein precipitation begins is called m^*. The protein's hydrophobicity can be expressed by $1/m^*$. The precipitation or aggregation process respectively is highly complex and is influenced by a number of parameters such as pH, protein concentration, temperature or additives. Hence the transferability of the hydrophobicity determined by the precipitation method to other environmental conditions is limited.

A second experimental approach to determine a surface hydrophobicity is the use of extrinsic fluorescent dyes such as ANS, Bis-ANS, Nile Red, or Tioflavin T [199]. This method focuses on the hydrophobic patches on the protein's surface. The sensitivity of the dyes' optical properties to their environment's polarity, basicity or acidity respectively are used for protein characterization. The mechanisms are described elsewhere in detail [199]. These experimental procedures require high investments in time and material and are limited with regard to transferability to other systems because of the environmental influences mentioned above. This is why hydrophobicity scales based on structural data would be beneficial.

In the last decades a number of scales were developed to characterize proteins regarding

hydrophobicity [21, 136, 155, 200–202]. Partition coefficients between polar and apolar solvents, retention behavior in reversed phase chromatography or thermodynamic approaches were used to allocate a hydrophobicity value to every amino acid. Some of the scales, like the probably most widely used scales of Kyte and Doolittle, implicate the probability of exposure in a globular protein. Globular proteins prone to form a hydrophobic core to minimize the hydrophobic solvent accessible surface in polar solvents like water [203]. The probability of an amino acid being exposed on the surface is accounted by Kyte and Doolittle by using the results from Chothia et al [204]. Based on the primary sequence of a protein, it is possible to estimate the overall hydrophobicity by summing up the hydropathy values for each residue and dividing them by the length of the sequence, according to Kyte and Doolittle [21].

But estimating a protein's hydrophobicity solely based on its amino acid sequence cannot account for the complexity of its structure in solution. The structure of a protein is highly influenced by environmental properties like the nature of the solvent, temperature, co-solvents, additives or pH, which induces charge variations in turn. The allocation of single hydropathy values for each amino acid can give an impression of the proteins hydrophobicity but is not suitable to compare the protein properties depending on its environment.

Molecular dynamics (MD) simulations were shown to be an appropriate approach to accommodate for the proteins' complexity. For instance Cramer et al. [205] have used MD simulations to localize potential binding sites of a mixed mode ligand to various Ubiquitin mutants. In a previous work we have shown that MD simulations are suitable to simulate aqueous two phase systems [206]. By a number of tools coming from bioinformatics such as homology modeling [207, 208] it is possible to model a three dimensional structure, without the actual need to crystallize that particular protein. Thereby MD simulations can be performed in the very early stages of drug development, namely drug screening and candidate selection.

In the presented work we developed an *in silico* method to characterize small peptides in aqueous solutions with respect to hydrophobicity and its distribution on the surface. The basic intention of the presented work is the development of an *in silico* approach to characterize proteins with respect to surface hydrophobicity. We are convinced, that a first step to reach these ambitious aim is the use of short and clearly defined amino acid structures which don‘t expose complex secondary or tertiary structures. Therefore we have focused on small custom designed peptides for method development and validation in the presented work. The developed approach is potentially capable to be transferred to proteins. Explicit solvent molecular dynamic simulations containing hydrophobic tracer

molecules were performed for eight different peptides. A modified radial distribution function (mRDF) was used to generate a hydrophobicity index. Based on the mRDF the peptides' overall hydrophobicity could be quantified and the distribution of hydrophobic areas could be identified. Experimentally the overall hydrophobicity could be confirmed by reversed phase ultra high performance liquid chromatography (RP-UHPLC). It could be shown that this *in silico* method yielded a better correlation to experimental data than the most commonly used scale of hydrophobicity presented by Kyte and Doolittle.

8.2 Materials and Methods

Peptides

To develop a method to describe hydrophobicity based on MD simulations eight custom designed peptides, each with a length of ten amino acids, were purchased from Thermo Fisher Scientific GmbH (Ulm, Germany). The primary sequences of these peptides are listed in one-letter-code in table 8.1. All peptides include a nominally hydrophilic patch with duplicates of Lys and Asp and a nominally hydrophobic residue separated by neutral Gly. In case of the hydrophobic residue Leu three peptides with a different content of Leu were designed. The peptides were designed by following considerations:

- The peptide should not form complex secondary or tertiary structure like beta sheets or alpha helices.
- The hydrophobic substitution residue should not carry an additional charge, thus should not constitute the C-terminus.
- The hydrophobic substitution residue should be accessible for the C18 ligand of the reversed phase chromatography matrix. Thus very small amino acids namely Gly were set next to the substitute to minimize steric hindrance to the C18 ligand.
- The peptide should have local charges to increase water solubility.

Table 8.1: Primary sequences of used model peptides.

name	primary sequence
Pep_{1Leu}	NH_2-Asp-Lys-Asp-Lys-Gly-Gly-Gly-Gly-Leu-Gly-OH
Pep_{2Leu}	NH_2-Asp-Lys-Asp-Lys-Gly-Gly-Gly-Leu-Leu-Gly-OH
Pep_{3Leu}	NH_2-Asp-Lys-Asp-Lys-Gly-Gly-Leu-Leu-Leu-Gly-OH
Pep_{1Cys}	NH_2-Asp-Lys-Asp-Lys-Gly-Gly-Gly-Gly-Cys-Gly-OH
Pep_{1Ile}	NH_2-Asp-Lys-Asp-Lys-Gly-Gly-Gly-Gly-Ile-Gly-OH
Pep_{1Met}	NH_2-Asp-Lys-Asp-Lys-Gly-Gly-Gly-Gly-Met-Gly-OH
Pep_{1Phe}	NH_2-Asp-Lys-Asp-Lys-Gly-Gly-Gly-Gly-Phe-Gly-OH
Pep_{1Val}	NH_2-Asp-Lys-Asp-Lys-Gly-Gly-Gly-Gly-Val-Gly-OH

8.2.1 Reversed Phase HPLC

The retention factors of the peptides on a reversed phase UHPLC column were used as experimental measure of the peptides' hydrophobicity. These measurements were performed on a Dionex UltiMate 3000 RSLC x2 Dual system consisting of an HPG-3400RS binary rapid separation LC pump module, a WPS-3000TRS autosampler, a TCC-3000RS column compartment and a DAD3000(RS) diode array detector.
To avoid limitations by mass transfer the monolithic silica based HPLC column *Chromolith Performance RP-18 endcapped* (100x2 mm) from Merck KGaA (Darmstadt, Germany) was used. Each peptide was solubilized in ultra pure water to a concentration of 1 mg/ml. Samples were injected with a volume of 5 μl. The column temperature was kept at 25°C. The peptides were eluted with a gradient starting from 100 % filtered, degassed and ultra pure water to 40 % degassed acetonitril, purchased from Merck KGaA (Darmstadt, Germany) with a slope of 1 CV/min and a flow of 3 ml/min. The system dead time (t_{Sys}) was measured by injecting 5 μl of 50 $mmol$ citric acid. The columns dead time (t_0) was measured by injecting 5 μl of 50 $mmol$ citric acid onto the column.The retention factor $k_i^{'}$ was calculated by equation 8.1

$$k_i^{'} = \frac{t_{R,i}^{'} - t_0}{t_0} \tag{8.1}$$

$$t_{R,i}^{'} = t_{R,i} - t_{Sys} \tag{8.2}$$

where $t_{R,i}^{'}$ is the absolute retention time of the peptide, t_{Sys} is the systems dead time and t_0 is the retention time of a non-retained analyte. All measurements were performed as quintets.

8.2.2 MD Simulations

The explicit MD simulations were performed on two distributed memory parallelized Intel Xeon Quad Core sockets with 2.66 GHz frequency with the software *YASARA Structure* (Version 10.10.29) [209, 210]. A cubic simulation cell with dimensions of about 57 Å and periodic boundaries was used. The temperature was kept at 298 K (isotherm) by rescaling the atom velocities using a Berendsen thermostat [211]. Simulations were performed with constant atom number, constant volume and constant temperature.
The AMBER03 force field [212] was applied to describe the interactions of the proteins, which has been adequate for the simulation of peptides and proteins. The TIP3P model was used for water. The particle mesh Edwald (PME) algorithm treatment [213] was used for longrange electrostatics with a 7.86 Å cutoff for non-bonded interactions. These cutoff of 7.86 Å was selected due to concerns regarding the performance/accuracy trade-off. As the software YASARA is optimized towards a cutoff that is a multiple of 2.62 Å and the chosen cutoff applies to the Van der Waals forces only, we selected a cuttof of 7.86 Å. The simulation cell included one initially centered peptide of interest, 33 copies of the neutrally charged tracer molecule triethylamine (TEA) and over 6000 water molecules. The protonation states of residues corresponded to a pH of 7. At this pH the N-terminal Asp has a negatively charged carboxyl group and a positively charged N-terminus, Asp 3 and C-terminal Gly are negatively charged and Lys 2 and Lys 4 are positively charged resulting in a neutrally charged peptide in total. Consequently, no counter ions were added for cell neutralization.
The tracer TEA was parametrized using the AutoSMILES algorithm [214] implemented in the MD simulation software. In order to remove bumps and to correct the covalent geometry, every simulation was initially energy-minimized with the AMBER03 force field [212]. After removal of conformational stress by a short steepest descent minimization, the procedure continued by simulated annealing until convergence was reached with a time step of 2 fs and scaled down atom velocities by 0.9 every 10th step. Convergence was reached when the energy improved by less than 0.05 kJ/mol per atom during 200 steps.
To gain information about the distribution of the hydrophobic tracer molecules or the local densities respectively, a modified radial distribution function (mRDF) of each protein atom was calculated. In this work the modified radial distribution function represents the concentration of the tracer atoms in dependency of the distance from an atomic core, which is illustrated in Figure 8.1. In contrast to the generally used radial distribution function the solvent accessible volume is considered in the mRDF. This modification is

necessary in order to be able to distinguish between an atom with hydrophobic character with minor solvent accessibility and an atom with a neutral character which is exposed and completely solvent accessible. The mRDF corresponds to the probability of finding tracer molecules around a certain peptide atom. A high probability of finding a tracer molecule in proximity of a certain peptide surface region in turn corresponds to a high hydrophobicity of that region.

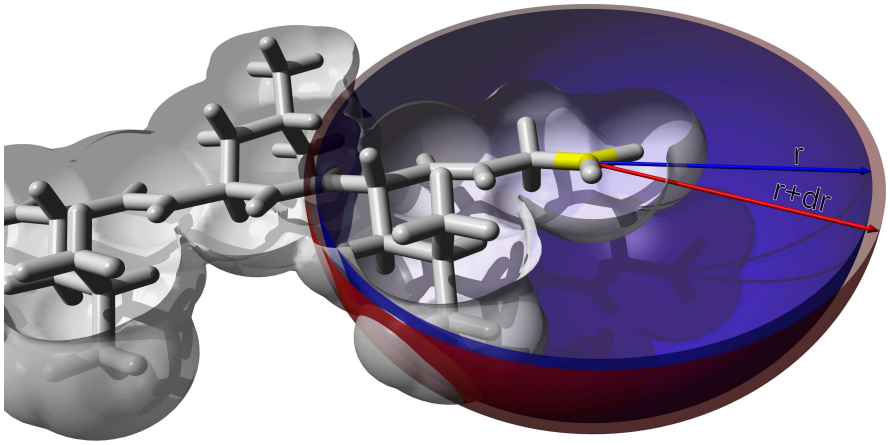


Figure 8.1: Schematic illustration of the procedure of calculating local distribution of each proteinogenic atom core using the example of the C-terminal carbon atom of Pep_{3LEU}. The atom of interest is colored in red, the peptide's Van-der-Waals-surface is shown in gray and the balancing shells, limited by the radius r and $r+dr$, are shown in blue or red respectively. Molecular graphic was created with *YASARA* (www.yasara.org) and POVRay (www.povray.org).

Data evaluation was done in Matlab® R2012a (The MathWorks, Natick, ME, USA) and determined how many tracer atoms were within a shell limited by the inner radius r and the outer radius r+dr around each atom core of the protein. The solvent accessible volume (probe radius of 1.4 Å) of each shell was calculated by a numerical approximation as follows. Each shell was filled with 5000 (N_t) evenly distributed grid points and the number of solvent accessible points (N_a) was evaluated. The solvent accessible volume of the shells (V_a) was determined by equation 8.3.

$$V_a = \frac{4}{3}\pi(r_a^3 - r_i^3)\frac{N_a}{N_t} \tag{8.3}$$

With knowledge about the number of tracer atoms and the solvent accessible volume of each shell the concentration was calculated. For each atom core the concentration was determined within a distance of 10 Å by evaluating 40 shells with a width of 0.25 Å each. Simulations were evaluated every 25 ps within a time range from 10 ns to 150 ns. The results of all simulation snapshots were averaged. By this averaging over a relative long simulation time the modified radial distribution profiles correspond to the probability of a tracer atom being in a specific region of the respective atom.

8.3 Results

8.3.1 Experimental Results

The retention factors on a monolithic reversed phase column of all peptides were measured to get a scale for their overall hydrophobicity. The peptides eluted in the following order and their retention factors are listed in table 8.2: $\mathrm{Pep_{1Cys}} < \mathrm{Pep_{1Val}} < \mathrm{Pep_{1Met}} < \mathrm{Pep_{1Ile}} < \mathrm{Pep_{1Leu}} < \mathrm{Pep_{1Phen}} < \mathrm{Pep_{2Leu}} < \mathrm{Pep_{3Leu}}$. All retention factors could be determined with a standard deviation below 3.2 %.

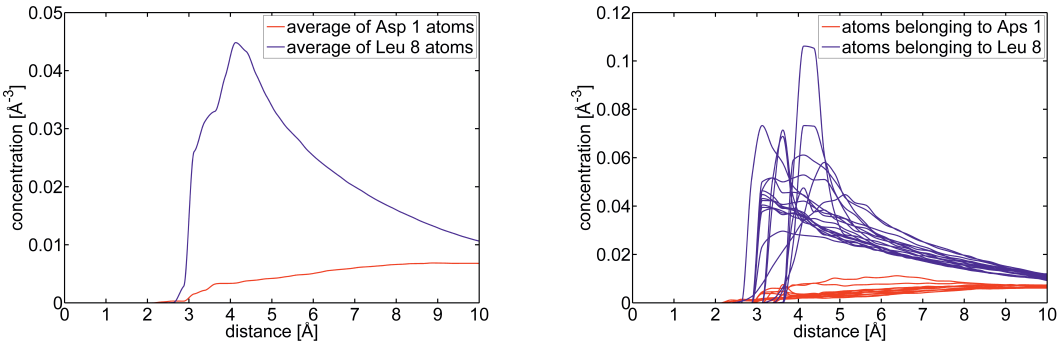
Table 8.2: Retention factors k' and the relative standard deviation σ_r of peptide $\mathrm{Pep_{Substitute}}$ obtained by reversed phase ultra high performance liquid chromatography.

Substitute	**1Cys**	**1Ile**	**1Leu**	**1Met**	**1Phen**	**1Val**	**2Leu**	**3Leu**
$k'[-]$	0.52	4.73	4.80	2.27	6.11	1.86	11.97	16.22
$\sigma_r[\%]$	0.66	3.10	2.83	0.22	1.28	0.36	0.16	0.04

8.3.2 In silico Results

The evaluation of the modified radial distribution function of each atom was focused on the characterization of each atom with respect to hydrophobicity and the peptide in total respectively. In general, the determined mRDF varied strongly within the primary sequence. The atoms of hydrophobic residues showed a peak shaped radial distribution which approximates the tracer's bulk concentration after reaching the maximal peak height. In contrast atoms of a charged residue showed no or low peaks, which is illustrated

by means of Figure 8.2. Figure 8.2a illustrates the averaged mRDF of all atoms belonging to Asp 1 and Leu 8. Figure 8.2b reveals the mRDF of every single atom of these amino acids. These simplified illustrations indicate, that the hydrophobicity of the peptides' atoms and amino acids correlates with the shape of the corresponding mRDF.



(a) (b)

Figure 8.2: Comparison of density profiles of atoms belonging to Asp 1 and Leu 8. The distribution functions were averaged over a time range of 10 ns to 150 ns.

All determined mRDF on this atomic level of the investigated peptides are presented in Figure 8.3 to illustrate the basic observations. Within these subfigures the primary sequence of the respective peptide is located on the x-axis as one-letter-code. All atoms belonging to one residue are colored consistently.

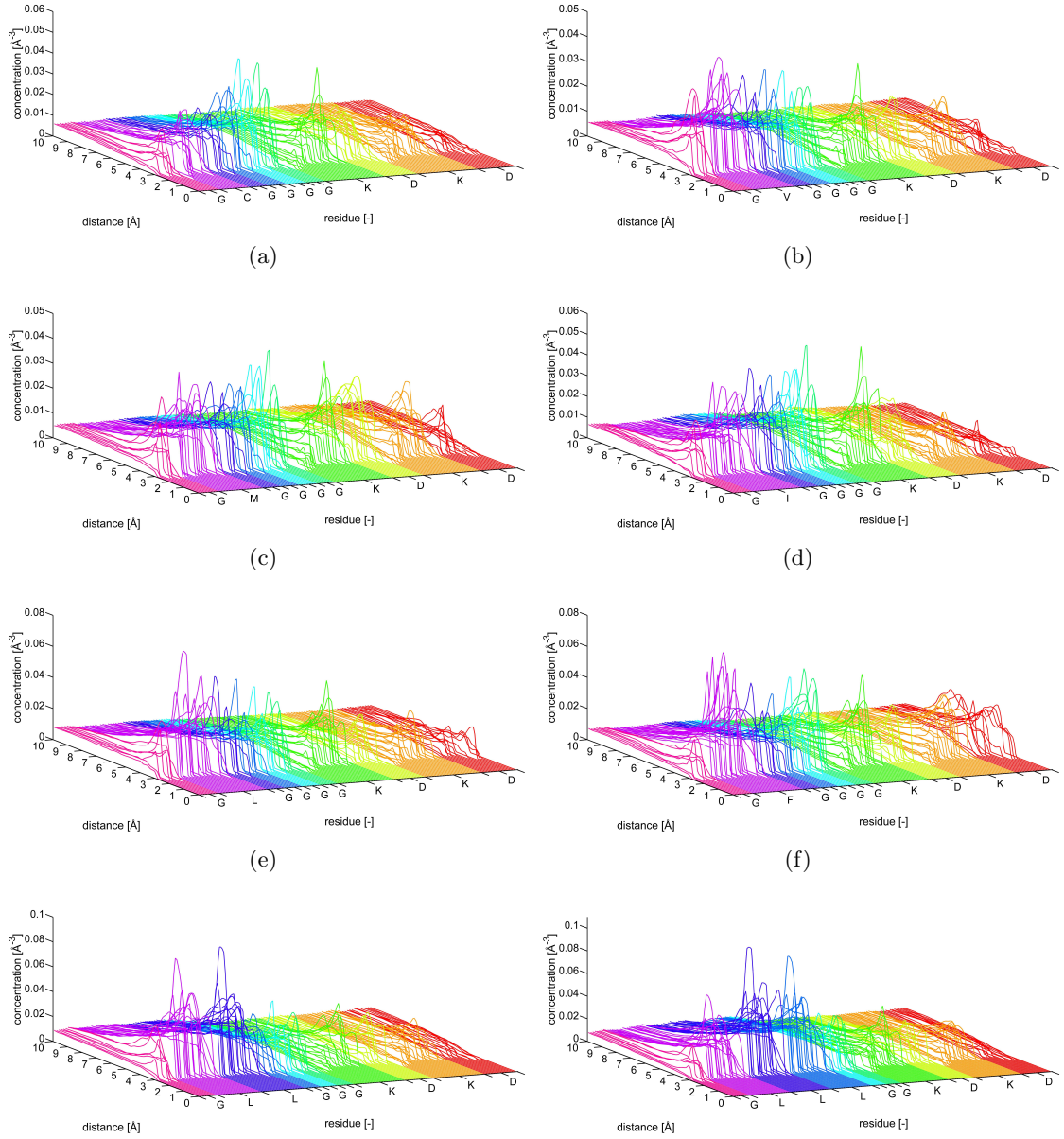


Figure 8.3: Compilation of all evaluated mRDF sorted by increasing retention factor (a) Pep_{1Cys}, (b) Pep_{1Val}, (c) Pep_{1Met}, (d) Pep_{1Ile}, (e) Pep_{1Leu}, (f) Pep_{1Phen}, (g) Pep_{2Leu}, and (h) Pep_{3Leu}.

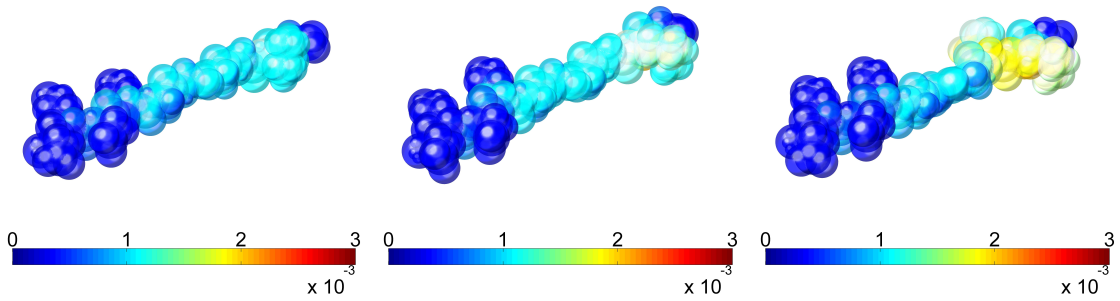
As illustrated in Figure 8.3 the averaged modified radial distribution function (mRDF) of all peptides varied strongly within their primary sequence. All modified distribution

density functions revealed no tracer atoms within a certain distance from the respective protein atom core due to steric hindrance. This distance was equal to the VdW-radius of the respective atom plus the smallest VdW-radius of the tracer atoms. The comparison of all peptides revealed a fingerprint-character of the density profiles. Profiles were unique for every peptide. While the profiles correlated well to the nature of the residue, the atom belongs to, they surpassed the common amino acid based scale in resolution. For example, the charged and nominally hydrophilic Asp 4 and especially its backbone carbon atom revealed in all mRDF some peaks with a height of about 0.04 $Å^{-3}$, equal to the hydrophobicity of the adjacent Gly and thus showing the influence of a hydrophobic residue on its surrounding. Comparing the nominally hydrophobic part of the sequence consisting of Gly and the substitute it was obvious, that the peak heights increased with the retention factor of the respective peptide. Comparing the Leu containing peptides, it was found that the peak heights of mRDF within the Leu residues increased with the number of Leu (compare Figure 8.3e, Figure 8.3g, and Figure 8.3h) and the distance to the negatively charged C-terminal Gly (see Figure 8.3h). To determine the distribution or the localization of hydrophobic surfaces a descriptor H_{atom} was extracted from the mRDF ($c(r)$) for each atom according to the following equations.

$$H_{atom} = A_{Core} \cdot \frac{A_{Core}}{(\sum_{i=1}^{n_{atoms}} A_{Core,i})} \tag{8.4}$$

$$A_{Core} = \int_{r=r_{vdW}}^{r=r_{vdW}+6\text{Å}} c(r)dr \tag{8.5}$$

The area A_{Core} of one core was calculated by integrating the mRDF of the respective atom. The integration of the mRDF corresponds to the possibility of tracer atoms being nearby the respective atom within the region limited by r_{vdW} and $r_{vdW} + 6$Å. The hydrophobicity value of an atom (H_{atom}) was calculated by weighting the respective area (A_{Core}) with its ratio to the sum of all areas within the peptide.
This atomic hydrophobicity descriptor H_{atom} was used for mapping the peptides' surfaces to characterize the distribution of hydrophobic areas. This mapping revealed a good correlation with expected hydrophobicity distribution and quantification. Exemplary the peptides $\text{Pep}_{1\text{Val}}$, $\text{Pep}_{1\text{Leu}}$ and $\text{Pep}_{2\text{Leu}}$ mapped by H_{atom} are illustrated in Figure 8.4.



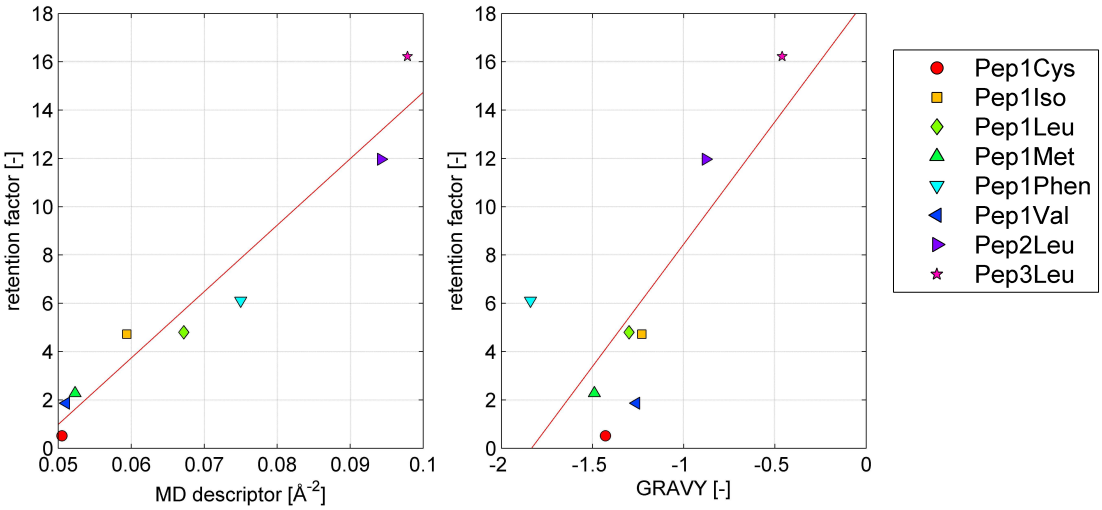
(a) (b) (c)

Figure 8.4: Characterization of the hydrophobicity distribution by mapping the peptide's surface according to the atomic descriptor H_{atom} by means of (a) $\text{Pep}_{1\text{Val}}$, (b) $\text{Pep}_{1\text{Leu}}$ and (c) $\text{Pep}_{2\text{Leu}}$. Colder colors refer to hydrophilic areas, whereas warm colors refer to hydrophobic areas.

Hydrophilic surfaces, colored in dark blue, can be precisely assigned to the charged residues. Hydrophobic surfaces, shown in warm colors, can be allocated to nominally neutral Gly and the nominally hydrophobic amino acids. However, the transitions within the sequence are smooth. The C-backbone in all peptides from Asp 3 to the C-terminus revealed a rather neutral character.
Each peptide's overall hydrophobicity (H_{Pep}) was calculated by summing up the hydrophobicity values H_{atom} of each of its atoms according to equation 8.6.

$$H_{Pep} = \sum_{i=1}^{n_{atoms}} H_{atom,i} \tag{8.6}$$

The overall peptides' hydrophobicity was tested experimentally by a RP-HPLC. As presented in Figure 8.5, there was a good linear correlation between the calculated hydrophobicity values (H_{Pep}) and the retention factors, obtained from the reversed phase chromatography, with a coefficient of determination (R^2) of 0.94. In comparison to the widely used GRAVY-value (grand average of hydropathy), which bases on the scale of Kyte and Doolittle, correlated with a R^2 of 0.60. It is important to note that the order of elution and the order of the H_{Pep} values were equal which was in contrast to the GRAVY-value. However, the focus of the presented work was not to generate an universal hydrophobicity scale comparable to Kyte and Doolittle but to develop an *in silico* method to generate an impression of the hydrophobicity of the peptide of interest in its specific environment.



(a) (b)

Figure 8.5: Correlation of the retention factor k' and (a) the descriptor obtained from MD simulations ($R^2 = 0.94$) and (b) the GRAVY value ($R^2 = 0.60$) for the eight model peptides.

8.4 Discussion

8.4.1 Reversed Phase HPLC

To compare *in silico* generated hydrophobicity values, an experimental scale of this property was needed. We decided on using small peptides not forming stable secondary structures as initial benchmark for the *in silico* parameter generation. As experimental measure of the peptides' hydrophobicities, a RP-HPLC assay was successfully established for one monolithic column. Retention in RP-HPLC is strongly dependent on the hydrophobicity of the analyte and the stationary phase [154]. However, we feel that the validation of a calculated overall hydrophobicity by retention factors in reversed phase chromatography is justified as the order of elution is a good experimental factor for the analyte's hydrophobicity, if size exclusion effects, diffusion limitations or steric hindrance of ligand-analyte binding can be considered insignificant. In our case we avoided any impact of size exclusion and diffusion limitations by using a monolithic chromatography column. Small and highly flexible peptides were designed in order to provide fully accessible hydrophobic residues. Thus steric hindrance of ligand interaction, which would

be influenced by the ligand length, can be considered negligible. Charge effects were minimized by using an endcapped RP phase and by keeping the charge of each peptide constant. We are thus confident that the established assay is a valid representation of the peptides' hydrophobicity. The peptides did not elute according to the nominal hydrophobicity of Kyte and Doolittle whereupon the elution order is supposed to be as follows:
$Pep_{1Phen} < Pep_{1Met} < Pep_{1Cys} < Pep_{1Leu} < Pep_{1Val} < Pep_{1Ile} < Pep_{2Leu} < Pep_{3Leu}$.
This was contradictory to the experiments which revealed the following elution order:
$Pep_{1Cys} < Pep_{1Val} < Pep_{1Met} < Pep_{1Ile} < Pep_{1Leu} < Pep_{1Phen} < Pep_{2Leu} < Pep_{3Leu}$.

8.4.2 MD Simulations

The present work was designed to determine hydrophobic surfaces of peptides using molecular dynamics simulations. The evaluation of the MD simulation aims at the modified distribution function in respect of a hydrophobic tracer molecule.
The concentration of TEA was optimized towards the maximum number of tracer molecules to reduce the necessary simulation time and to ensure system stability. It was considered that the necessary time to get an impression of the peptides hydrophobicity can be reduced by increasing the tracer number, as it increases the probability of the peptide meeting a tracer molecule. In contrast, phase separation was observed in the presence of too many tracer atoms. Therefore an optimization was performed. Experimental phase separation could be observed using a cloud point method at about 1.3 mol% or 0.06 g/ml respectively. A phase separation could be observed at 0,05 g/ml *In silico*, which equals to 1.0 mol%. Even the basic character of TEA is not represented in MD simulations, this concentration was within the scope of the critical concentration where phase separation could be observed in MD simulation. As a comparison the concentration for the final MD simulations was 0.55 mol%.
The necessary simulation time to get a stable mRDF is illustrated in Figure 8.6 by means of a carbon atom (number 113) belonging to the phenol ring of the peptide Pep_{1Phen}. Starting 10 *ns* into the simulation, the average mRDF was calculated for increasing simulation durations. The length of the period over which the average was calculated is color coded in Figure 8.6. With increasing duration (with warmer color) the calculated average becomes increasingly stable. Figure 8.6 clearly shows that the chosen period of 10 *ns* to 150 *ns* is sufficient to yield a stable, representative average mRDF.

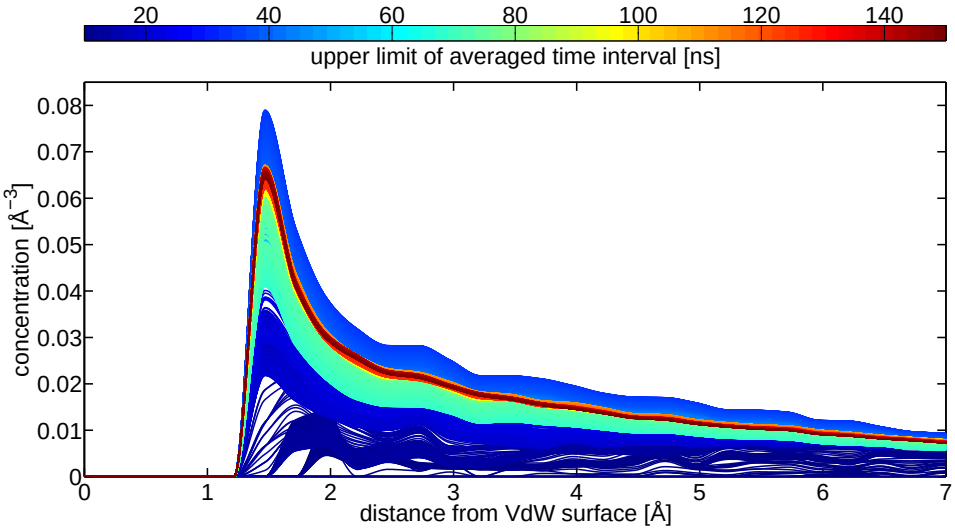


Figure 8.6: Stability of the mRDF illustrated by means of on C-atom belonging to the phenol ring (atom 113) of the peptide $\text{Pep}_{1\text{Phen}}$. The colorbar represents the the time interval which was averaged to calculate the mRDF.

As shown in the results section the profiles of the mRDF indicate the hydrophobic or hydrophilic property of each protein atom. Charged atoms and atoms with nearby charges revealed low attractions to the hydrophobic tracer. This is due to the electrostatic interactions with the dipolar water molecules and hence the preference for forming hydration shells around charged atoms [215, 216]. Charged or polar solvent molecules displace hydrophobic tracer molecules from charged peptide areas.

By contrast the mRDF profiles of atoms belonging to hydrophobic amino acids indicated an accumulation of hydrophobic tracer molecules as it is thermodynamically preferred to minimize the exposed hydrophobic area. The comparison of the mRDF profiles of the Leu containing peptides exhibit the influence of the nearby C-terminal charge and the cumulative effects of hydrophobic amino acids, as Leu 7 and Leu 8 accumulate more tracer molecules compared to Leu 9.

This shows, that this molecular dynamic approach is sensitive enough to identify hydrophobic areas of the peptides and can consider adjacency effects like decreasing hydrophobicity by nearby charges or increasing hydrophobicity by forming hydrophobic patches consisting of a number of hydrophobic amino acids. Potentially this approach could be used to identify aggregating prone areas for proteins that follow a native aggregation pattern.

A descriptor H_{Pep} was generated according to equation 8.6 on base of the evaluated

mRDF of each atom as an overall characteristic value with respect to hydrophobicity. It is important to note that the choice of the descriptor to be correlated to experimental data was guided by the underlying physical principles. It was thus decided to calculate the mRDF to a distance of the VdW radius plus 6 Å as this represents the space in which the relevant forces predominantly act. While a slightly better correlation of MD data to experimental data could be derived by using other descriptors, we are confident that correlations based on reasoning rather than statistics are more likely to have predictive power. The descriptor correlates to the probability of finding a tracer molecule within a region limited by the van der Waals radius and the used cut-off where the van der Waals interactions are considered. It seems to be capable to characterize the peptides' surface with regard to hydrophobicity in terms of localization and strength. This descriptor seems to be sensitive enough to picture effects caused by adjacent charges on an atomic level. The overall descriptor H_{Pep} showed a satisfying agreement with the retention data.

Compared to the developed *in silico* descriptor, the GRAVY values correlated less well with the retention factors. This is not surprising considering the short length of the peptides. The terminal charges have a relatively high influence on the peptides' hydrophobicity, which is not accounted in the calculation of the GRAVY value.

Kim et al [217] have already observed, that there is no correlation between elution times and the GRAVY score. Although peptide retention is not supposed to correlate to the GRAVY score, this index appears as an useful scale for distinguishing hydrophobic from hydrophilic peptides [218]. A proteomics approach for predicting peptides retention was developed by Krokhin et al [219]. We are aware that the tracer substance triethylamine is not present only in neutral form due to the basic character. However, this tracer molecule could meet our demands with respect to stability of the *in silico* system, accessibility due to small size and distinctive hydrophobic character.

The used peptides did not show a stable secondary structure, therefore the influence of the hydrophobic tracer molecule on the peptides' conformation was considered negligible. Molecular dynamics simulations of systems under identical conditions but without tracer molecules were performed to compare the conformations of the peptides. The root-mean-square deviation (RMSD) of the backbone of each peptide to its initial linear structure was calculated for all snapshots. Neither significant variations of the RSMD can be observed in the presence of tracer atoms, nor preferred conformations could be identified (data not shown). This is due to the high flexibility of the peptides. However, for larger proteins it must be considered, that the presence of the hydrophobic tracer molecules could lead to minor conformational changes. The simulation methodology used in the

presented work is not adequate to simulate major structural conformational changes like folding or unfolding as the simulation time range is not appropriate to observe larger conformational changes. It cannot be excluded that minor structural conformational changes could still be induced by the presence of tracer molecules. For instance hydrophobic residue nearby the exposed surface could be induced to move slightly to the surface attracted by a tracer molecule. For the transfer to larger proteins, we thus recommend an equilibration without tracer molecules and constraining the backbone after adding the tracer molecules.

8.5 Conclusion

This study has shown that molecular dynamic simulations can be used for the characterization of small peptides with respect to hydrophobicity with good experimental agreement. In this work an *in silico* method was developed, which can cope with peptides complexity. In particular the overall hydrophobicity of the peptides could be evaluated from molecular dynamics simulations and validated by reversed phase chromatography. Even on atomic level the hydrophobic distribution could be determined, which can be helpful to identify aggregation prone surfaces for proteins following a native aggregation pattern. Compared to the established hydrophobicity scales and Kyte and Doolittle scale (GRAVY) in particular, we found a better agreement of the *in silico* descriptor with the experiments. We are confident that this molecular dynamics approach can be transferred to larger proteins even when a reduction of computational effort is needed. This could be done by reducing the number of mRDF evaluations by calculating the mRDF of every amino acid in contrast to every atom.
Finally this approach could be used to characterize chromatographic ligands of HIC, RP or mixed mode adsorber with respect to hydrophobicity.

8.6 References

21. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157,** 105–132 (1982).

49. Hachem, F., Andrews, B. & Asenjo, J. Hydrophobic partitioning of proteins in aqueous two-phase systems. *Enzyme Microb. Technol.* **19,** 507–517 (1996).

136. Radzicka, A. & Wolfenden, R. Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry* **27,** 1664–1670 (1988).

154. Biswas, K. M., DeVido, D. R. & Dorsey, J. G. Evaluation of methods for measuring amino acid hydrophobicities and interactions. *J. Chromatogr. A* **1000,** 637–655 (2003).

155. Bull, H. B. & Breese, K. Surface tension of amino acid solutions: A hydrophobicity scale of the amino acid residues. *Arch. Biochem. Biophys.* **161,** 665–670 (1974).

198. Asenjo, J., Schmidt, A., Hachem, F. & Andrews, B. Model for predicting the partition behaviour of proteins in aqueous two-phase systems. *J. Chromatogr. A* **668.** 8th International Conference on Partitioning in Aqueous Two-Phase Systems, 47–54 (1994).

199. Hawe, A., Sutter, M. & Jiskoot, W. Extrinsic fluorescent dyes as tools for protein characterization. English. *Pharm. Res.* **25,** 1487–1499 (2008).

200. Hessa, T. *et al.* Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433,** 377–381 (2005).

201. Wimley, W. C. & White, S. H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Mol. Biol.* **3,** 842–848 (1996).

202. Guy, H. Amino acid side-chain partition energies and distribution of residues in soluble proteins. *Biophys. J.* **47,** 61–70 (1985).

203. Moelbert, S., Emberly, E. & Tang, C. Correlation between sequence hydrophobicity and surface-exposure pattern of database proteins. *Protein Sci.* **13,** 752–762 (2004).

204. Chothia, C. The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol.* **105,** 1–12 (1976).

205. Freed, A. S., Garde, S. & Cramer, S. M. Molecular simulations of multimodal ligand-protein binding: Elucidation of binding sites and correlation with experiments. *J. Phys. Chem. B* **115,** 13320–13327 (2011).

206. Oelmeier, S., Dismer, F. & Hubbuch, J. Gaining mechanistic understanding of aqueous two-phase systems for bioseparation. *Chem. Ing. Tech.* **84,** 1292–1292 (2012).

207. Bordoli, L. & Schwede, T. English. in *Homology Modeling* (eds Orry, A. J. W. & Abagyan, R.) 107–136 (Humana Press, New York, USA, 2012).

208. Cavasotto, C. N. & Phatak, S. S. Homology modeling in drug discovery: Current trends and applications. *Drug Discov. Today* **14,** 676–683 (2009).

209. Krieger, E. *et al.* Making optimal use of empirical energy functions: Force-field parameterization in crystal space. *Proteins* **57,** 678–83 (2004).

210. Krieger, E., Koraimann, G. & Vriend, G. Increasing the precision of comparative models with YASARA NOVA–a self-parameterizing force field. *Proteins* **47,** 393–402 (2002).

211. Berendsen, H. J. C. *et al.* Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3684–3690 (1984).

212. Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24,** 1999–2012 (2003).

213. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103,** 8577–8593 (1995).

214. Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **23,** 1623–1641 (2002).

215. Teeter, M. M. Water structure of a hydrophobic protein at atomic resolution: Pentagon rings of water molecules in crystals of crambin. *Proc. Natl. Acad. Sci. U. S. A.* **81,** 6014–6018 (1984).

216. Némethy, G. & Scheraga, H. A. The structure of water and hydrophobic bonding in proteins. III. The thermodynamic properties of hydrophobic bonds in proteins. *J. Phys. Chem.* **66,** 1773–1789 (1962).

217. Kim, J. *et al.* Phosphopeptide elution times in reversed-phase liquid chromatography. *J. Chromatogr. A* **1172,** 9–18 (2007).

218. Blonder, J. *et al.* A detergent- and cyanogen bromide-free method for integral membrane proteomics: Application to halobacterium purple membranes and the human epidermal membrane proteome. *Proteomics* **4,** 31–45 (2004).

219. Krokhin, O. *et al.* An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC: Its application to protein peptide mapping by off-line HPLC-MALDI MS. *Mol. Cell. Proteomics* **3,** 908–919 (2004).

CHAPTER 9

Molecular Dynamics Simulations Approach for the Characterization of Proteins With Respect to Hydrophobicity

Sven Amrhein and Jürgen Hubbuch

Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Science, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany

in preparation

Abstract

Protein hydrophobicity contributes majorly to the protein-solvent interaction and is a predominating factor for protein solubility and colloidal stability and thus effects the manufacturing process and the formulation stability of biopharmaceuticals enormously. However, the assessment of protein hydrophobicity is challenging since established experimental approaches are invasive and thus potentially affect the protein integrity. Traditional *in-silico* approaches based on hydrophobicity scales significantly neglect the protein conformational structure and dynamics as well as the influence of the solvent on protein characteristics such as pH depending charge distributions.
In this work, we present an *in-silico* approach based on molecular dynamics (MD) simulations for the assessment of the protein surface regarding hydrophobicity. The methodology was originally developed by means of short peptides previously [151] and was revised thoroughly particularly with regard to sampling speed in order to apply this approach to larger proteinogenic structures. The approach bases on tracking the positions of hydrophobic tracer molecules in an aqueous solvent during a molecular dynamic simulation with regard to the protein surface, which is evaluated as a modified radial distribution function (mRDF). The competition between water and tracer molecules enables a comprehensive characterization of the protein surface on an atomic level of detail. By means of this approach three proteins, namely human lysozyme, lysozyme from chicken egg white and α-lactalbumin, were characterized regarding surface hydrophobicity at pH 5 and 7. The proteins exhibit a high structural similarity but differ significantly in the amino acid composition. The data yielded by the MD simulation based characterization provides a strong pH dependency of protein surface hydrophobicity as well as clear discrepancy between protein species despite the structural similarity. The interference on the protein structure by the presence of added tracer molecules was taken into account by constraining the protein backbone in position. The *in-silico* derived hydrophobicity was validated against a non-invasive experimental approach, which based on the influence of proteins on surface tension and was developed previously [184]. Thus, we are convinced that this MD simulation based approach correctly assesses protein surface hydrophobicity while considering protein structural complexity.

Keywords: *Tracer Probing; Radial Distribution Function; In-Silico; Hydropathy; Protein-Solvent Interactions*

9.1 Introduction

Hydrophobicity of proteins is one of the most important parameters for protein colloidal stability, solubility and pharmaceutical potency but at the same time a highly challenging protein property to assess. There is a variation of experimental as well as theoretical approaches to assess protein hydrophobicity. Experimentally, protein hydrophobicity is derived by means of precipitation studies, specific dyes, adsorption to stationary phases such as reversed phase (RP) or hydrophobic interaction chromatography (HIC) [49, 165, 220, 221] or partitioning studies in two phase systems composed of a polar and a non-polar phase. Protein precipitation is typically induced by salt or polymers and the minimal concentration needed to decrease the soluble protein concentration (m^*) [49] as well as the sensitivity of the solubility at concentrations above $m*$ determined by the slope of the precipitation curve are used as hydrophobicity measure. Special dyes such as bromophenol blue (BPB) or 8-anilinonaphthalene sulfonate (ANS) [26, 142, 222, 223] are used for absorption or emission difference spectroscopy to assess protein hydrophobicity. In case of HIC and RP chromatography the retention behavior is correlated to the hydrophobic character of the protein based on increased retention in case of pronounced protein hydrophobicity. The partitioning coefficient, often described in the logarithmic form as log$P$, within aqueous non-polar solvent systems reflects the ratio of compound concentrations in the two immiscible phases at equilibrium and is correlated to the compound hydrophobicity. A typical system is composed of water and hydroxylic solvents such as 1-octanol [136, 224, 225]. However, ensuring the conformational integrity of the protein in presence of precipitates, binding dyes, or in presence of organic or high salt solvents is crucial. In addition, changing the solvent chemistry affects the charge distribution by influencing the protonation states of amino acids capable for ionization.
To overcome these experimental shortcomings we previously developed a non-invasive approach to characterize proteins with respect to hydrophobicity based on their surface tension increment [184]. This approach enables the assessment of the protein hydrophobicity without modifying the solvent and thus measuring the protein in the desired environment.
In order to characterize the protein surface with regard to hydrophobicity more detailed it is so far only possible to apply theoretical scales, which were developed based on established dye, retention, or precipitation based studies as mentioned above. However, none of the scales is capable of describing solvent effects or can consider the dynamic and structural complexity of a protein.

With growing computational performance due to technical developments advancing the integrated circuit technology, the *in-silico* approach of molecular dynamics (MD) simulations arises increasing interest in biopharmaceutical research which can provide an unique insight into protein dynamics on an atomic level of detail and enables to observe virtually nature at work. This highly potential methodology was used in order to investigate surface hydrophobicity of short peptides in a previous study [151]. A MD simulations based approach was developed in order to assess peptide surface hydrophobicity by means of tracking hydrophobic tracer molecules, and triethylamine (TEA) in particular, during simulation. The movement of the tracer molecules in relation to the peptide surface was analyzed by means of high resolution radial concentration profiles of each peptide atom. The profiles exposed an increased probability of tracer molecules close to hydrophobic surface patches compared to hydrophilic surface areas. It was possible to transform the profiles into a descriptor to assess the peptides' overall hydrophobic characters which were found to be in excellent agreement with retention behavior in RP chromatography and outperformed in this respect one of the most established theoretical hydrophobicity scales by Kyte and Doolittle [21].
This promising approach was applied in the present work on larger proteinogenic structures namely α-lactalbumin, human lysozyme and lysozyme from chicken egg white. These proteins are highly similar in structure (root mean square deviation <1.2 Å) but expose a low sequence identity (<65 %) according to an evaluation applying the MUSTANG method [226]. These, in terms of structure and size, highly similar proteins were evaluated *in-silico* at pH 5 and 7 regarding surface hydrophobicity.

9.2 Materials and Methods

9.2.1 Experimental Assessment of Protein Hydrophobicity

The hydrophobic character of α-lactalbumin, human lysozyme, and lysozyme from chicken egg white as a function of pH was assessed bases on its surface tension increments as described in detail previously [184]. It was evaluated by means of parameters extracted from surface tension as function of the protein concentration. The surface tension profile γ_{norm}, normalized to the respective buffer without protein, in dependency of molar protein fraction $\tilde{x}$ was fitted to the model as stated in equation 9.1

$$\gamma_{norm} = 1 - \frac{d \cdot e \cdot (\tilde{x} + c)}{1 + e \cdot (\tilde{x} + c)}. \tag{9.1}$$

The fitting parameters d and e could be calculated for all profiles within this study with a coefficient of correlation larger than 0.98. Exclusively in case of α-lactalbumin the normalized surface tension of pure buffer and the lowest concentrated sample were excluded from the calculation, which was considered by setting the parameter c properly. For all other proteins, parameter c was set to zero. Equation 9.1 describes a saturation function of the normalized surface tension regression, where $d \cdot e$ can be interpreted as the surface activity of the protein and was shown to correlate with the protein surface hydrophobicity by applying absorption difference spectroscopy of bromophenol blue. The $\log_{10}(de)$ is referred as hydrophobicic character in this work for simplifications and is illustrated in Figure 9.1. Negative values represent a rather hydrophilic character whereas positive

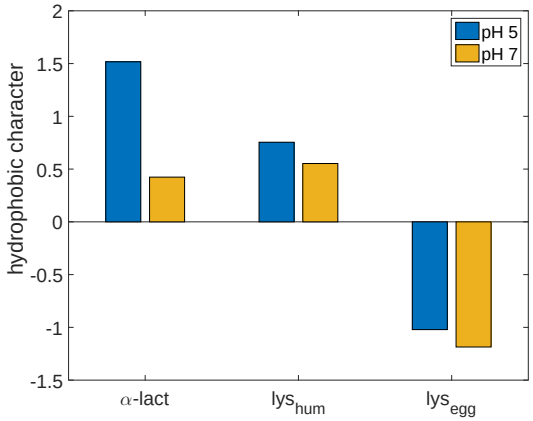


Figure 9.1: Hydrophobicity ranking of α-lactalbumin (α-lact), lysozyme from chicken egg white (lys$_{egg}$), and human lysozyme (lys$_{hum}$) by means of the surface activity. Data was reported previously [184].

values stand for a rather hydrophobic character. This experimentally derived protein hydrophobicity was used for comparison with data derived from MD simulations described in section 9.2.2.

For this study solely pH 5 and 7 were considered since more extreme pH values might cause minor conformational changes, which cannot be considered sufficiently by the performed molecular dynamics simulations. Extremely acidic pH values might cause partial unfolding as reported for α-lactalbumin, which forms a molten globule state at extremely acidic pH values [227]. In addition, the 3D structure used in the MD simulations of

α-lactalbumin (pdb entry 1F6R) was derived from crystals grown at pH 6.0 [16] and the used protein structure of lysozyme from chicken egg white (pdb entry 2VB1) was derived by crystallization at pH 4.7 [228]. Thus, we are confident, that the used protein structures are well suited for the pH values of 5 and 7.

9.2.2 Molecular Dynamics Simulations

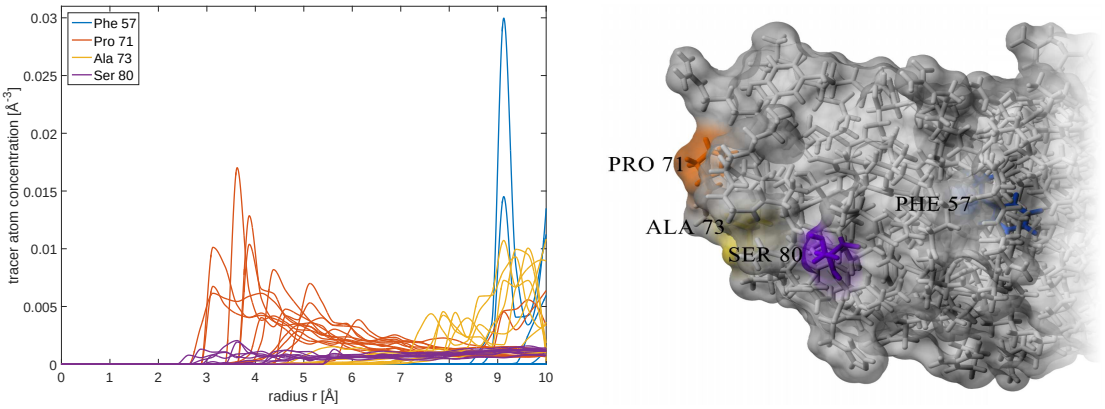
Molecular dynamic simulations were performed by means of YASARA Structure (version 14.6.23) [209]. The AMBER03 force field [212] was applied in combination with calculating electrostatic Coulomb forces without cutoff using the particle mesh Ewald (PME) algorithm [213]. Van der Waals forces were calculated with a cutoff of 7.86 Å, since YASARA is optimized to cutoffs multiple of 2.62 Å. The TIP3P model was used for water and the hydrophobic tracer molecule triethylamine was parameterized by the software implemented AutoSMILES algorithm [214], which was shown to be valid for small molecules. Guided by sequence completeness and resolution, three-dimensional structural data 1F6R [16] for α-lactalbumin, 2NWD [229] for human lysozyme, and 2VB1 [228] for lysozyme from chicken egg white were used. These proteins are very similar in structure and size, but vary strongly in composition and hydrophobicity as determined in a previous work [184].
Protein structures were prepared by the following procedure. Each protein structure was protonated according to default pk_a values and the respective solute pH value followed by an energy minimization. This includes a removal of conformational stress by a short steepest descent minimization and a subsequent annealing simulation until energy improved by less than 0.05 kJ/mol per atom during 200 steps. The resulting structures were further processed applying the H++ algorithm [230] in order to optimize the protonation states.
These optimized structures were used for explicit full atom molecular dynamics simulations on a high performance distributed memory parallel computer (InstitutsCluster II) at the Steinbuch Centre for Computing (SCC) of the Karlsruhe Institute of Technology (KIT, Karlsruhe, Germany). The temperature was set to 293.15 K and controlled by a rescaling of atom velocities using a Berendsen thermostat [211]. A cubic simulation cell with periodic boundaries and dimensions 30 Å larger than the protein's largest dimension was chosen, which ensures the cell to be large enough for the current cutoff, that protein atoms do not interact with their periodic images. The cell was filled with water

to reach a density of 0.997 g/mL. Tracer molecules were added in randomized position and orientation to reach a concentration of 0.03 mg/mL. For electrical neutrality of the cubic simulation cell respective numbers of sodium and chloride ions were added. The molecular dynamics simulations were performed for 100 ns and the trajectories were saved every 25 ps. The trajectories were analyzed for modified radial distribution function (mRDF) starting from 10 ns. In contrast to the generally known radial distribution function, the mRDF reflects the concentration in dependency of the radial distance and thus considers the solvent accessibility within the radial distance.
The trajectories were processed fully automated by means of highly parallelized routines written in Matlab® R2015a (The MathWorks, Inc., Natick, MA, USA). The mRDF of each protein atom was evaluated within a distance starting from 0 into 10 Å with 0.25 Å steps. The solvent accessible volume of each shell was determined by placing 1000 grid points evenly distributed on a sphere with the respective radius in each shell and each point was analyzed for solvent accessibility with a probe radius of 1.4 Å.
In order to investigate if the presence of the tracer molecules does effect the protein structure and thus falsify the obtained protein characteristics, all simulations were performed with and without constraining the backbone atoms of the protein as well as with and without tracer molecules. The flexibility and conformational integrity of the protein backbone atoms namely C_α, C, and N atoms were evaluated by means of the root-mean-square-fluctuations (RMSF) and root mean-square-deviation (RMSD) with respect to the initially energy minimized starting structure.

9.3 Results

The effect of the tracer molecules on the protein structure and molecular flexibility was not entirely shown to be negligible, since the addition of tracer molecules exposed party an impact on the RMSF and the RMSD profiles but did not follow a clear trend nor a significant impact on the RMSF and the RMSD profiles in a general manner. Thus, the following results and the evaluated modified radial distribution functions (mRDF) refer to the simulations of the backbone constrained proteins.

The mRDF reflects the concentration of tracer atoms at a certain distance from the respective protein atom core, which was averaged over 10 to 100 ns. Thus, this profile reflects the probability of a tracer atom being at a certain distance to the protein atom. The mRDF highly differs throughout the protein network of atoms. Exemplarily, the



(a) (b)

Figure 9.2: Illustration of (a) the modified radial distribution function (mRDF) of atoms belonging to Pro 71, Ala 73, Ser 80, and Phe 57 of human lysoyzme at pH 7 and (b) the respective 3D protein structure in stick style and the molecular protein surface. Molecular graphic was created with YASARA (www.yasara.org) and POVRay (www.povray.org).

mRDF profiles of human lysozyme at pH 7 differ significantly by the polarity of the respective amino acid and the polarity of its environment as shown in Figure 9.2. Atoms of residues carrying parital or absolute charges such as Ser 80 are primarily surrounded by polar water atoms during the simulation time whereas atoms of nominally hydrophobic residues such as Pro 71 show an accumulation of tracer atoms. The mRDF profiles of nearby amino acids, are affected by their neighborhood as illustrated by Ala 73, which is affected by the nearby Pro 71. Additionally, amino acids buried in the hydrophobic core are shielded by adjacent amino acids and their profiles are shifted towards larger radii by their distance to the protein surface as seen by Phe 57, which is part of a rather hydrophobic cavity with multiple aromatic contributing residues.
The mRDF profiles revealed a high fluctuation in dependency of pH and protein, despite the structural similarity. As a measure of the hydrophobicity of each atom ($H_{atom,i}$), the probability of finding a tracer atom within a certain solvation shell limited by the VdW radius (r_{VdW_i}) and VdW radius +6 Å ($r_{VdW_i} + 6\mathring{A}$) was used as a hydrophobicity measure. This probability and hydrophobicity measure was evaluated by the integration of the mRDF profiles fitted with shape-preserving piecewise cubic Hermite interpolation (PCHIP) within the mentioned limits according to equation 9.2. This fit preserves monotonicity and the shape of the data.

$$H_{atom,i} = \int_{r_{VdW_i}}^{r_{VdW_i}+6\mathring{A}} mRDF_i(r)dr \tag{9.2}$$

The hydrophobicity of each atom $H_{atom,i}$ serves for a heat map of hydrophobicity on the protein surface. The three proteins are illustrated in a compiled form in Figure 9.3, Figure 9.4, and Figure 9.5 and are discussed in the following. The values of $H_{atom,i}$ were normalized on 0.1, the maximal value of $H_{atom,i}$ derived in this study.
For α-lactalbumin at pH 5 no clear hydrophobic hot spot can be identified. Only minor pronounced hydrophobic spots can be found at two patches formed by Ile 41, Tyr 50, Asp 78, and Leu 81 or Pro 24 and Leu 115, respectively. The pH shift to pH 7 induces the formation of a very hydrophobic hot spot to which Phe 31, His 32, Tyr 36, and Trp 118 majorly contribute to. Three additional minor pronounced spots can be identified. One is contributed by Leu 105 and Ala 106, one by Gly 17, Tyr 18, Gly 19, Gly 20, and Val 21 and one by Ser 69 and Ser 70. Thus, α-lactalbumin seems to be more hydrophobic at pH 7 in comparison to pH 5. For human lysozyme at pH 5 and pH 7 a pronounced hot spot is formed by Val 2, Trp 34, Ser 36, Gly 37, Tyr 38, and Asn 39. Human lysozyme at pH 5 exposes a second significant hot spot formed by Arg 107, Arg 113, and Asn 114. Minor pronounced hydrophobic surface patches are contributed by Cys 6 and Cys 128, which are connected by a disulfide bond, and a second patch contributed by Asp 87 and Asn 88. Human lysozyme at pH 7 exposes two minor hydrophobic spots, one formed by Tyr 45, Asn 46, and Ala 47 and one formed by Gly 129 and Val 130. Thus, human lysoyzme at pH 7 is considered slightly more hydropbobic at pH 5 than pH 7. Lysozyme from chicken egg white does not expose significant hot spots at pH 5 and pH 7 and thus exposes a rather hydrophilic character throughout the protein surface. Based on the magnitude and number of hot spot formations, human lysozyme can be identified as the most hydrophobic protein at pH 5 and 7 followed by α-lactalbumin at pH 7. α-lactalbumin at pH 5 and lysozyme from chicken egg white at pH 5 and 7 are characterized as rather hydrophilic.

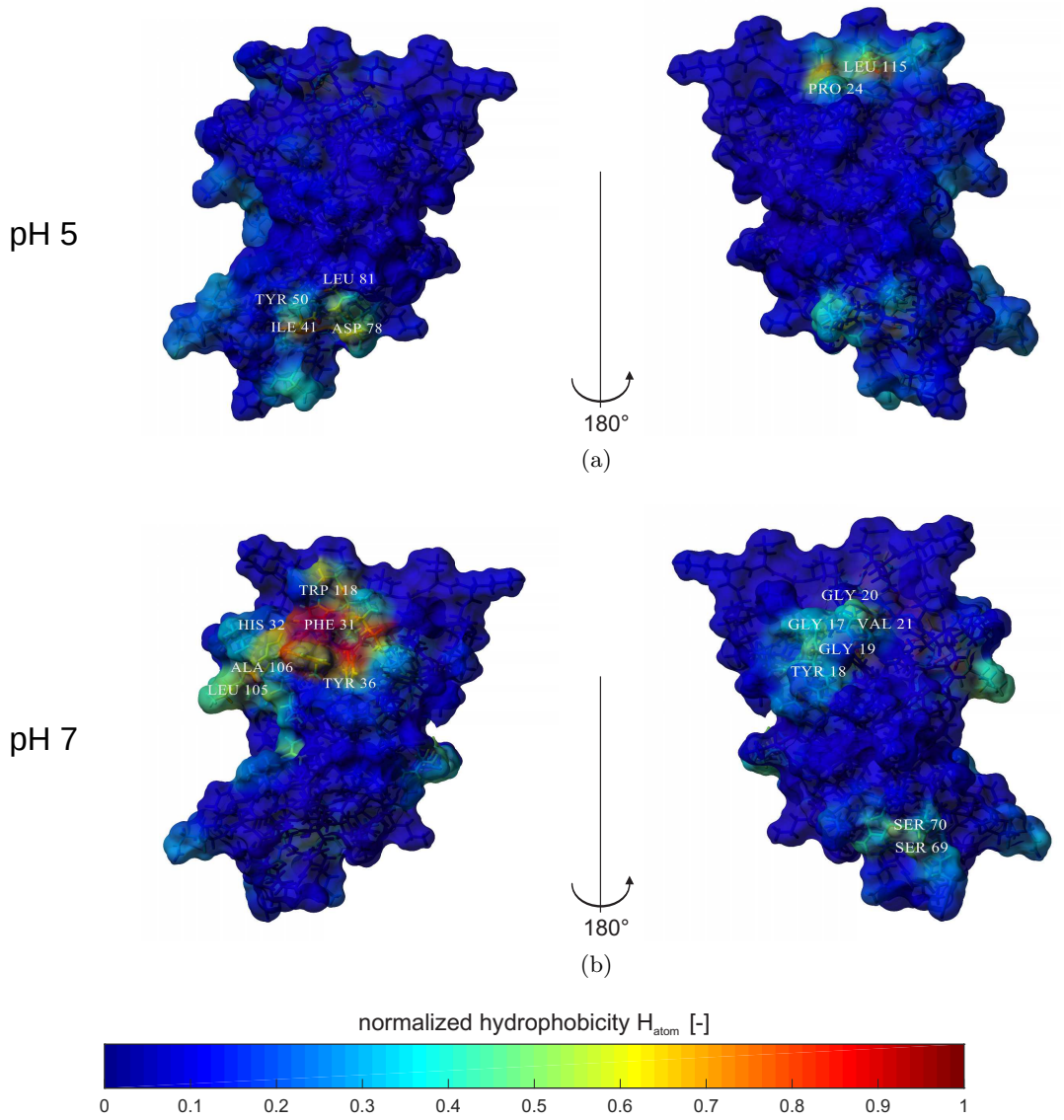


Figure 9.3: Color coded molecular surface of α-lactalbumin according to normalized atomic descriptor $H_{atom,i}$ (a) pH 5 and (b) pH 7. Warm colors refer to more pronounced hydrophobicity. Molecular graphics were created with YASARA (www.yasara.org) and POVRay (www.povray.org).

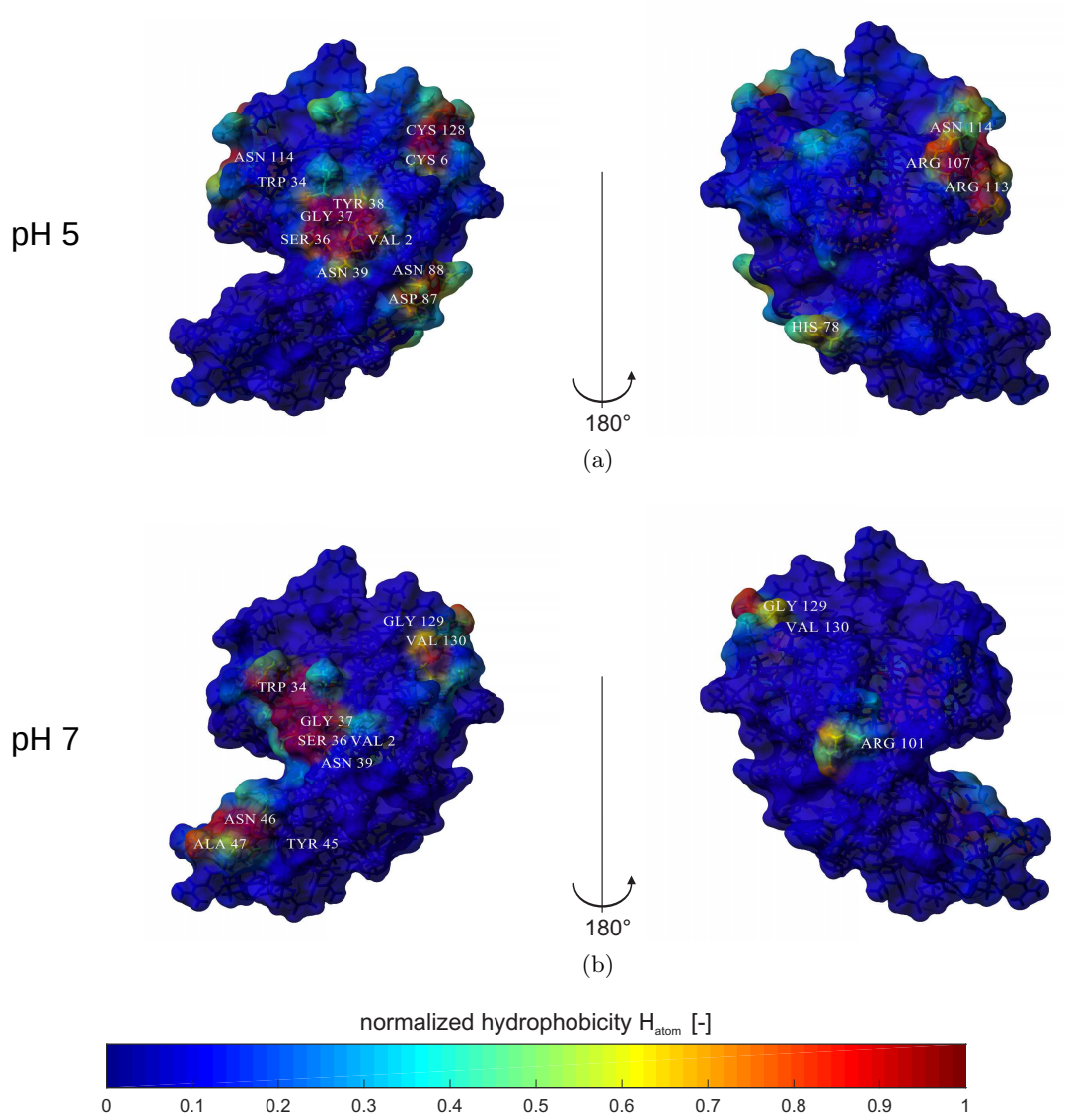


Figure 9.4: Color coded molecular surface of human lysozyme according to normalized atomic descriptor $H_{atom,i}$ (a) pH 5 and (b) pH 7. Warm colors refer to more pronounced hydrophobicity. Molecular graphics were created with YASARA (www.yasara.org) and POVRay (www.povray.org).

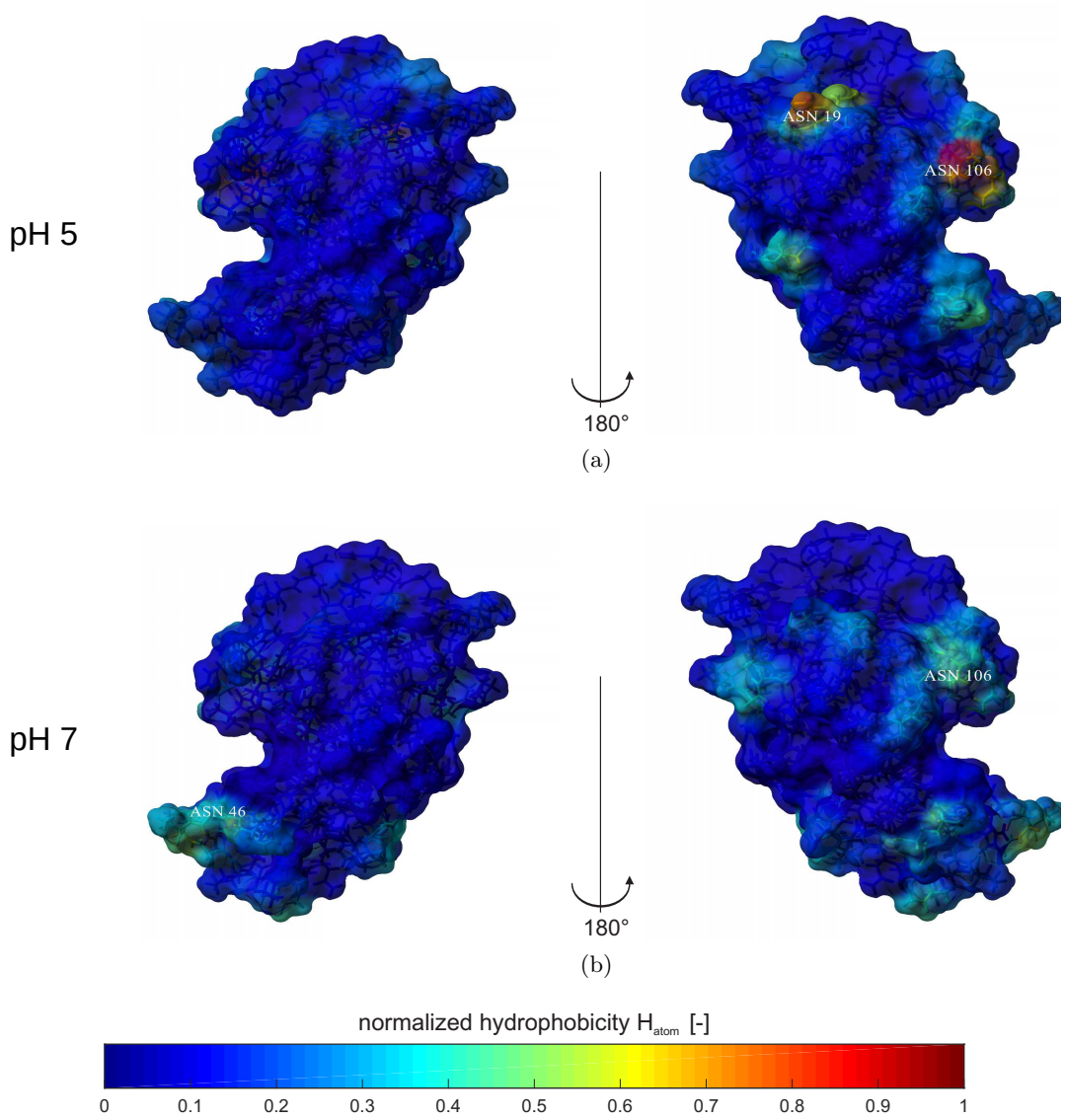


Figure 9.5: Color coded molecular surface of lysozyme from chicken egg white according to normalized atomic descriptor $H_{atom,i}$ (a) pH 5 and (b) pH 7. Warm colors refer to more pronounced hydrophobicity. Molecular graphics were created with YASARA (www.yasara.org) and POVRay (www.povray.org).

The charge distribution is a major factor for protein hydrophobicity, and highlights the dynamics of the protein surface regarding the distribution of hydrophobic patches. The pH induced protonation and deprotonation of respective amino acids capable for ionization leads to a loss or an induction of charge and thus into changing hydration behavior. This effect is reflected by the mRDF profiles and the evaluated hot spots, which can be pointed out by means of α-lactalbumin as illustrated if Figure 9.6. Exemplarily, a pH shift from 5 to 7 leads to a deprotonation of histidin and thus a loss of charge. In dependency of the amino acid composition adjacent to histidin, a rather hydrophobic patch is created or additional charges conserve the hydration and avoid the accumulation of hydrophobic tracer molecules. In case of α-lactalbumin a deprotonation of His 32 leads to the formation of a hydrophobic hot spot by adjacent hydrophobic amino acids and Phe 31, His 32, Tyr 36, and Trp 118 in particular, even though the nearby Lys 5 carries a charge on its terminal.

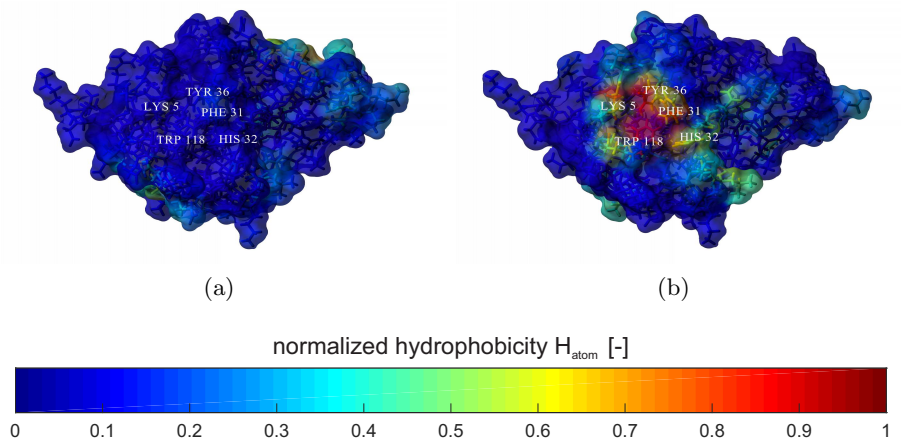


Figure 9.6: The effect of a pH induced deprotonation by means of His 32 of α-lactalbumin by shifting from (a) pH 5 to (b) pH 7. The colors reflect $H_{atom,i}$. Molecular graphics were created with YASARA (www.yasara.org) and POVRay (www.povray.org).

9.4 Discussion

The MD simulations approach for assessing the protein surface hydrophobicity was developed by means of short, highly flexible peptides and was applied to larger proteins in

this study. The scale-up to proteinogenic structures was accompanied by the impact of tracer molecules onto the molecular structure. This effect was evaluated by assessing the RMSD and RMSF profiles in absence and presence of tracer molecules and was found to be not negligible. In order to consider this effect, the backbone of the proteins was constrained to conserve the energy minimized structure. This implies, that the constrainment prevents potential conformational changes during the MD simulation, but we are confident that the protein structures which were energy minimized and equilibrated in absence of tracer molecules are reasonable representative of the dynamic conformations. In addition, the MD simulation setup is not capable to model major conformational changes such as denaturation or partial unfolding, which would influence the hydrophobic character of the protein by the associated exposure of rather hydrophobic residues buried in the core of the native structure. Thus, the restriction in the backbone structure dynamic is considered to have minor impact on the surface hydrophobicity of the protein structure that can be considered during the MD simulations.

The hot spot analysis enables the surface characterization on an atomic level of detail. The identified hot spots are predominately formed by nominally hydrophobic amino acids such as Trp, Tyr, and Phe or nominally neutral amino acids such as His, Gly and Ser. More polar amino acids are hydrated by polar water molecules and thus no hydrophobic tracer molecule approaches this surfaces.

In contrast to theoretical hydrophobicity scales or the experimental assessment of zeta potential and the interpretation along the DLVO theory, which result in one comprehensive experimental protein parameter, this approach was shown to evaluate the surface topology with respect to hydrophobic patches and considers the influence of pH induced shifts of charges. This approach was capable of considering protein surface topology and amino acid composition.

The MD simulation based approach ranked human lysozyme to be the most hydrophobic studied protein at pH 5 and 7 followed by α-lactalbumin at pH 7. This is in contrast to experimental hydrophobicity derived by surface tension profiling. However, α-lactalbumin tends to partially unfold at low pH values. It forms a molten globule state at extreme pH values [227] and it is thus likely that a partial unfolding already occurs at pH 5 to a certain extent and thus hydrophobic amino acids are exposed, which are buried in the hydrophobic core in the native state. The exposure of hydrophobic amino acids by partial unfolding and the protein flexibility cannot be considered in the used MD simulations. The native state of α-lactalbumin should be more hydrophobic at pH 5 than at pH 7 because of the increased charge and polarity due to the DLVO theory. The hydrophobicity of α-lactalbumin at pH 7 is slightly less pronounced than that of human lysozyme

at pH 5 and 7, which is in accordance to the experimental results. The MD simulations revealed lysozyme from chicken egg white at pH 5 and 7 as rather hydrophilic. This is in good agreement with the experimentally derived hydrophobicity.
In summary, the *in-silico* derived protein surface hydrophobicity can be confirmed by the experimentally determined protein hydrophobicity apart from α-lactalbumin at pH 5 which is probably due to unfolding processes which cannot be considered by this MD simulation approach.

9.5 Conclusion

The MD simulation based approach enabled the identification of hydrophobic hot spots on the protein surface based on the competition of polar water molecules and hydrophobic tracer molecules. The *in-silico* derived protein hydrophobicity confirmed the experimentally assessed protein hydrophobicity on base of surface tension profiling. The presented approach was shown to be capable of considering solvent effects and the effect of pH induced protonations on the protein surface in particular. This feature outperforms traditional hydrophobicity scales. We are confident that the presented approach is of high value for the *in-silico* molecular protein assessment.

9.6 Acknowledgment

This research work is part of the project "Molecular Interaction Engineering: From Nature's Toolbox to Hybrid Technical Systems", which is funded by the German Federal Ministry of Education and Research (BMBF), funding code 031A095B. We would like to thank Cathrin Dürr from our work group for her helpful contribution to parts of the code writing of the molecular dynamics simulations.

9.7 References

16. Chrysina, E. D., Brew, K. & Acharya, K. R. Crystal structures of apo-and holobovine α-lactalbumin at 2.2-Å resolution reveal an effect of calcium on inter-lobe interactions. *J. Biol. Chem.* **275,** 37021–37029 (2000).

21. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157,** 105–132 (1982).

26. Alizadeh-Pasdar, N. & Li-Chan, E. C. Comparison of protein surface hydrophobicity measured at various pH values using three different fluorescent probes. *J. Agric. Food Chem.* **48,** 328–334 (2000).

49. Hachem, F., Andrews, B. & Asenjo, J. Hydrophobic partitioning of proteins in aqueous two-phase systems. *Enzyme Microb. Technol.* **19,** 507–517 (1996).

136. Radzicka, A. & Wolfenden, R. Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry* **27,** 1664–1670 (1988).

142. Hawe, A., Sutter, M. & Jiskoot, W. Extrinsic fluorescent dyes as tools for protein characterization. *Pharm. Res.* **25,** 1487–1499 (2008).

151. Amrhein, S., Oelmeier, S. A., Dismer, F. & Hubbuch, J. Molecular dynamics simulations approach for the characterization of peptides with respect to hydrophobicity. *J. Phys. Chem. B* **118,** 1707–14 (2014).

165. Guo, D. *et al.* Prediction of peptide retention times in reversed-phase high-performance liquid chromatography I. Determination of retention coefficients of amino acid residues of model synthetic peptides. *J. Chromatogr. A* **359,** 499–518 (1986).

184. Amrhein, S., Christin Bauer, K., Galm, L. & Hubbuch, J. Non-invasive high throughput approach for protein hydrophobicity determination based on surface tension. *Biotechnol. Bioeng.* **112,** 2485–2494 (2015).

209. Krieger, E. *et al.* Making optimal use of empirical energy functions: Force-field parameterization in crystal space. *Proteins* **57,** 678–83 (2004).

211. Berendsen, H. J. C. *et al.* Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3684–3690 (1984).

212. Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24,** 1999–2012 (2003).

213. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103,** 8577–8593 (1995).

214. Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **23,** 1623–1641 (2002).

220. Mant, C. T., Zhou, N. E. & Hodges, R. S. Correlation of protein retention times in reversed-phase chromatography with polypeptide chain length and hydrophobicity. *J. Chromatogr. A* **476,** 363–375 (1989).

221. McCue, J. T. in *Guide to Protein Purification, 2nd Edition* (eds Burgess, R. R. & Deutscher, M. P.) 405–414 (Academic Press, 2009).

222. Haskard, C. A. & Li-Chan, E. C. Hydrophobicity of bovine serum albumin and ovalbumin determined using uncharged (PRODAN) and anionic (ANS-) fluorescent probes. *J. Agric. Food Chem.* **46,** 2671–2677 (1998).

223. Chao, C.-C., Ma, Y.-S. & Stadtman, E. R. Modification of protein surface hydrophobicity and methionine oxidation by oxidative systems. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 2969–2974 (1997).

224. Sangster, J. *Octanol-water partition coefficients: Fundamentals and physical chemistry* (John Wiley & Sons, 1997).

225. Fauchere, J. & Pliska, V. Hydrophobic parameters-pi of amino-acid side-chains from the partitioning of n-acetyl-amino-acid amides. *Eur. J. Med. Chem.* **18,** 369–375 (1983).

226. Konagurthu, A. S., Whisstock, J. C., Stuckey, P. J. & Lesk, A. M. MUSTANG: A multiple structural alignment algorithm. *Proteins: Struct., Funct., Bioinf.* **64,** 559–574 (2006).

227. Kuwajima, K. The molten globule state of alpha-lactalbumin. *FASEB J.* **10,** 102–109 (1996).

228. Wang, J. *et al.* Triclinic lysozyme at 0.65 Å resolution. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **63,** 1254–1268 (2007).

229. Durek, T., Torbeev, V. Y. & Kent, S. B. Convergent chemical synthesis and high-resolution X-ray structure of human lysozyme. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 4846–4851 (2007).

230. Gordon, J. C. *et al.* H++: A server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **33,** W368–W371 (2005).

CHAPTER 10

Predictive Approach for Protein Aggregation: Correlation of Protein Surface Characteristics and Conformational Flexibility to Protein Aggregation Propensity

Lara Galm[1], Sven Amrhein[1], and Jürgen Hubbuch

[1] these authors contributed equally

Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Science, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany

Biotechnology and Bioengineering (2016), accepted

Abstract

The aggregation of proteins became one of the major challenges in the development of biopharmaceuticals since the formation of aggregates can affect drug quality and immunogenicity. However, aggregation mechanisms are highly complex and the investigation requires cost, time and material intensive experimental effort. In the present work the predictive power of protein characteristics for the phase behavior of three different proteins which are very similar in size and structure was studied. In particular, the surface hydrophobicity, zeta potential, and conformational flexibility of human lysozyme, lysozyme from chicken egg white, and α-lactalbumin at pH 3, 5, 7, and 9 were assessed and examined for correlation with experimental stability studies focusing on protein phase behavior induced by sodium chloride and ammonium sulfate.
The molecular dynamics (MD) simulation based study of the conformational flexibility without precipitants was able to identify highly flexible protein regions which could be associated to the less regular secondary structure elements and random coiled and terminal regions in particular. Conformational flexibility of the entire protein structure and protein surface hydrophobicity could be correlated to differing aggregation propensities among the studied proteins and could be identified to be applicable for prediction of protein phase behavior in aqueous solution without precipitants. For prediction of protein phase behavior and aggregation propensity in aqueous solution with precipitants, protein flexibility was further studied in dependency of salt concentration and species by means of human lysozyme. Even though the results of the salt dependent MD simulations could not be shown to be sufficient for prediction of salt depending phase behavior, this study revealed a more pronounced destabilizing effect of ammonium sulfate in comparison to sodium chloride and thus was found to be in good agreement with theoretical considerations along the Hofmeister series as well as experimentally evaluated phase behavior.

Keywords: *Molecular Dynamics Simulation; Hydrophobicity; Zeta Potential; Root-Mean-Square-Fluctuation; Protein Phase Behavior*

10.1 Introduction

Understanding the mechanisms of protein aggregation and the cause for increased aggregation propensities are of major interest in biopharmaceutical industry as protein aggregation on the one hand is a highly selective and effective purification or formulation step but on the other hand complicates product handling, might cause immunogenic reactions or a loss of bioactivity if occurring uncontrollably. Protein aggregation can occur due to production related reasons such as mechanical stress during agitation, shear stress during filtration, pumping, and filling, or by interaction with surfaces [231–233]. More important, protein aggregation propensity in aqueous solution highly depends on the solvent conditions. Conformational and colloidal protein stability is dominated by protein-solvent and protein-protein interactions depending on solvent physiochemical properties such as pH value, salt concentration, salt type, and protein concentration [232, 234]. Protein characteristics change as function of pH since electrostatics as well as hydrophobicity are influenced by the solvent pH. Electrostatics, i.e. the net charge of the protein, depend on the composition and accessibility of amino acids capable of ionizing and its isoelectric point (pI) in particular. According to the colloidal DLVO theory the protein exhibits its lowest solubility in a polar solvent with a pH value identical to the pI [235]. For pH values below the isoelectric point, the protein is positively charged whereas it is negatively charged for pH>pI. Protein hydrophobicity has been described to be pH dependent both in the native state and in pH induced conformational variations [26, 184, 236]. Especially at extremely acidic and extremely alkaline pH values proteins tend to (partially) unfold resulting in an exposure of rather hydrophobic amino acids and thus increased hydrophobicity of the proteins [236, 237]. The presence of salt in protein solutions changes the aggregation propensity depending on salt species and salt concentration. Up to a certain salt concentration, the salts might stabilize proteins but will destabilize the solution for higher salt concentrations due to electrostatic shielding. Based on the stabilizing and destabilizing impact on conformational and colloidal stability of proteins, salt iones were classified by Hofmeister [238]. According to the Hofmeister series kosmotropic salts (e.g. PO_4^{3-}, SO_4^{2-}, and NH_4^{+}) promote hydrophobic interactions and protein precipitation whereas chaotropic salts (e.g. SCN^{-}, Ba^{2+}) decrease the strength of hydrophobic interactions and thus increase proteins' conformational and colloidal stability [239]. According to the salt contribution to protein stability the effect is known as salting-in or salting-out effect respectively. Protein aggregation during a protein purification process is likely to occur as solvent conditions frequently change throughout

the process. Salt concentration applied in chromatographic steps can raise up to 1 M or 2 M during IEX or HIC respectively [240]. During low pH viral inactivation steps pH values are adjusted to pH≤3.8 [241]. Hereby, especially irreversible aggregation leads to product loss and complicates product handling as the level of protein aggregates has to be closely monitored as critical quality attribute [242]. Conditions for induction of aggregation, if applied as purification or formulation step, are mainly found by trial and error, whereas undesired protein aggregation often is discovered not until process setup. Possibilities to predict protein aggregation propensity in aqueous solution either for application as process step or for selective process control (e.g. inhibition) would reduce cost, time, and product consumption in comparison to classical empirical methods. Though there is general knowledge about factors influencing protein aggregation propensity, such as pH and thus protein charge, protein hydrophobicity, protein concentration, and salt type as well as salt concentration, general approaches for prediction of aggregation propensity are missing. Predictive approaches for protein aggregation propensity are rare due to the complexity of protein molecules and the diversity of protein-solvent and protein-protein interactions. For examples of the few *in-silico* approaches for prediction of protein aggregation propensity see for instance [243–245]. Conformational flexibility as well as electrostatic and hydrophobic properties of proteins have been described to be of major importance for protein aggregation propensity [244–246]. However, different studies disagree about electrostatics [247, 248] or hydrophobicity [146, 249, 250] being the main determinant. According to Valerio et al. [245] a combination of conformational flexibility and of electrostatic and hydrophobic protein surface properties is supposed to be best suited for a predictive approach for protein aggregation propensity.

Thus, this study aims at developing an approach for prediction of protein aggregation propensity exemplary for three similar proteins namely α-lactalbumin (α-lact), human lysozyme (lys_{hum}) and lysozyme from chicken egg white (lys_{egg}) using conformational flexibility, electrostatic and hydrophobic protein surface properties. Conformational flexibility was determined by an *in-silico* approach using molecular dynamics (MD) simulation, whereas protein surface characteristics were determined experimentally. Conformational flexibility was assessed by the root-mean-square-fluctuation (RMSF) of the backbone atoms (C, C_α, and N) of the proteins. To evaluate protein surface characteristics, the zeta potential and the hydrophobic character were determined as function of pH. To correlate conformational flexibility and surface characteristics to protein aggregation propensity, protein phase states were determined in a protein concentration range between 2.5 and 21.75 mg/mL as function of pH. Additionally, the influence of sodium

chloride and ammonium sulfate on the aggregation propensity and its correlation to conformational flexibility was investigated.

10.2 Materials and Methods

10.2.1 Selection of Proteins

α-lactalbumin from bovine milk in its calcium depleted form (L6010, α-lact) and human lysozyme (L1667, lys_{hum}) were purchased from Sigma-Aldrich (St. Louis, MO, USA), lysozyme from chicken egg white (HR7-110, lys_{egg}) was purchased from Hampton Research (Aliso Viejo, CA, USA). These proteins were chosen for investigation due to their similar size but differing isoelectric points. α-lact has a molecular weight of 14.175 kDa [251] and an isoelectric point between 4.25 and 4.5 [252]. lys_{hum} has a molecular weight between 14 and 15.6 kDa [253] and an isoelectric point between 10.0 and 11.0 [254, 255]. lys_{egg} has a molecular weight of 14.3 kDa [251] and an isoelectric point between 10.0 and 11.0 as well [253, 256]. This selection of proteins enables to exclude size depending effects while studying pH depending surface properties and charge distributions respectively. The similarity of the secondary and tertiary structures between the three proteins is

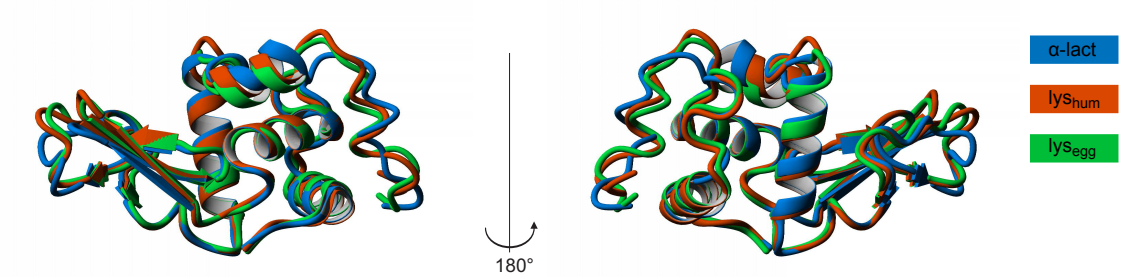


Figure 10.1: Superposition of the crystal structures of α-lactalbumin (pdb entry 1F6R) in green, human lysozyme (pdb entry 2NWD) in blue and lysozyme from chicken egg white (pdb entry 2VB1) in orange, all depicted in ribbon style. Protein alignment was conducted applying the MUSTANG method [226]. Molecular graphics were created with YASARA Structure (www.yasara.org) and POVRay (www.povray.org).

displayed by superposition of their crystal structures in Figure 10.1. The sequence iden-

tities of the primary sequences are less than 65 % according to an evaluation applying the MUSTANG algorithm [226].

10.2.2 Preparation of Buffers and Protein Solutions

α-lact, lys_{hum}, and lys_{egg} were investigated at pH 3, 5, 7, and 9. Used buffer substances were citric acid (Merck, Darmstadt, Germany) and sodium citrate (Sigma–Aldrich, St. Louis, MO, USA) for pH 3, sodium acetate (Sigma–Aldrich, St. Louis, MO, USA) and acetic acid (Merck, Darmstadt, Germany) for pH 5, MOPSO (AppliChem, Darmstadt, Germany) for pH 7 and bis-tris propane (Molekula, Dorset, UK) for pH 9. Buffers were prepared as described in detail in Baumgartner et al. [257]. The buffer capacity was 100 mM. The pH was adjusted with the appropriate titrant with an accuracy of ± 0.05 pH units. Proteins were weighed in, dissolved in the respective buffer, filtered through 0.2 µm syringe filters with PTFE membranes (VWR, Radnor, PA, USA) to remove particulates and desalted using a HiTrap Desalting Column (GE Healthcare, Uppsala, Sweden) on an AEKTAprime® plus system (GE Healthcare, Uppsala, Sweden). A subsequent protein concentration step, if necessary, was performed using Vivaspin centrifugal concentrators (Sartorius, Goettingen, Germany) with PES membranes and a molecular weight cutoff of 3 kDa.

10.2.3 Protein Surface Characterization

The zeta potential and the hydrophobic character of the surface of α-lact, lys_{hum}, and lys_{egg} were determined at different pH values, as charge and hydrophobicity of the proteins are thought to be of major importance for aggregation propensity. These measurements were conducted without any additional salts apart from buffer components in solution.

10.2.3.1 Determination of Zeta Potential

Determination of the zeta potential of α-lact, lys_{hum}, and lys_{egg} at pH 3, 5, 7, and 9 was performed with a Zetasizer Nano ZSP (Malvern Instruments Ltd, Malvern, United Kingdom). The measurements were conducted using the so-called diffusion barrier technique. Therefore, folded capillary cells (DTS1070, Malvern Instruments Ltd, Malvern,

United Kingdom) were filled completely with the respective buffer and hereafter 30 µL of the protein solution were injected into the measurement zone of the buffer filled cell using a Corning® Costar gel-loading pipet tip (4853, Corning Incorporated, Corning, NY, USA). To ensure conformational as well as colloidal integrity of the proteins during zeta potential measurements, a measurement procedure was set up with a dynamic light scattering measurement for determination of the protein size first, followed by the zeta potential measurement and a dynamic light scattering measurement again. During the zeta potential measurement the light scattered by the protein undergoing electrophoresis is combined with a reference beam. The rate of change of phase of the combined beam is measured and used to calculate the electrophoretic mobility and thus the zeta potential by using the Smoluchowski equation [258]. A significant Joule heating was avoided by setting the voltage to limit the current to a maximum of ±5 mA and by limiting the number of runs, i.e. number of phase changes, to maximum 15. The measurements were conducted at least threefold.

10.2.3.2 Determination of Hydrophobic Character

Determination of the hydrophobic character of α-lact, lys_{hum}, and lys_{egg} at pH 3, 5, 7, and 9 was conducted as described by Amrhein et al. [184]. There, it could be shown that the surface tension increment as a function of protein concentration is related to the pH depending hydrophobic character of proteins. Surface tension increments were determined using a stalagmometric approach and were used to rank the proteins according to their hydrophobic character.

10.2.4 Molecular Dynamics Simulations Based Protein Characterization

MD simulations of the proteins of interest within solvent varying in pH, salt concentration and salt species were performed in order to estimate protein conformational flexibility using the root-mean-square-fluctuation (RMSF) of atomic positions in particular as a measure of conformational flexibility. The complete MD simulation workflow starting from protein structure preparation over the calculation of the initial spatial solvent configuration to the final MD simulation and evaluation is explained in detail in the following.

10.2.4.1 Starting Structures

The selection of protein structure data was driven by sequence completeness and resolution. The PDB IDs 1F6R [16], 2NWD [229], and 2VB1 [228] were chosen for α-lactalbumin, human lysozyme, and lysozyme from chicken egg white, respectively. The starting structures were initially prepared with the MD software package YASARA Structure 14.6.23 [209] by applying force field parameters of AMBER03 [212]. Default pKa values were assigned to derive a rough protonation, followed by an optimization of the hydrogen bonding network and a simulated annealing energy minimization until the energy deviated by less than 10 J/mol per atom during 200 steps. These preprocessed protein structures then were optimized in terms of protonation and structure to match the buffer pH by applying the H++ algorithm [230]. The force field parameters of ammonium and sulfate were assigned by the build-in AutoSMILES algorithm [214]. The force field parameters of the counter ions sodium and chloride were taken from the AMBER03 force field [212].

10.2.4.2 Initial Solvation

The dimension of the cubic MD simulation cell was chosen to be 30 Å larger than the protein's largest spatial dimension according to the van der Waals radii defined in the AMBER03 force field [212]. The solvent volume and mass were calculated by means of the protein's molecular volume and the total box volume in combination with the respective density at 293.15 K to fit the desired solvent composition. The densities as a function of the salt concentrations were extracted from profiles fitted to literature values using a second-degree polynomial [259–261]. Sodium chloride and ammonium sulfate concentrations varied from 0 to 1.3 mol/L. The TIP3P model was used for water molecules. Fully randomized solvation assembles were derived by applying the PACKMOL algorithm [262, 263] in 6-fold determination in absence of salt and in 3-fold determination in presence of salt for each protein and pH value in order to realize a sufficient sampling.

10.2.4.3 Molecular Dynamics Simulations

Starting with the thus derived packed assembles the final explicit MD simulations were performed via YASARA Structure on a high performance distributed memory parallel computer (InstitutsCluster II) at the Steinbuch Centre for Computing (SCC) of the

Karlsruhe Institute of Technology (KIT). The AMBER03 force field [212] was applied. Electrostatic Coulomb forces were calculated without a cutoff using the particle mesh Ewald (PME) algorithm [213]. The van der Waals forces were calculated with a cutoff of 7.86 Å since YASARA Structure is optimized to cutoffs multiple of 2.62 Å. The temperature was set to 293.15 K and controlled by a rescaling of atom velocities using a Berendsen thermostat [211]. The final MD simulations were performed over 20 ns.

10.2.4.4 Assessment of Conformational Flexibility

The molecular dynamics simulations trajectories were analyzed for root-mean-square-fluctuations (RMSF) of the backbone atoms (C, C_α, and N) with respect to the energy minimized and equilibrated starting structure. The RMSF of atom i was determined by equation 10.1, where j runs over the three cartesian components x, y and z of the atom position vector P, and k runs over the N coordinate sets.

$$RMSF_i = \sqrt{\sum_{j=1}^{3}(\frac{1}{N}\sum_{k=1}^{N}P_{ikj}^2 - \overline{P_{ij}}^2)} \tag{10.1}$$

The data processing was performed by automated routines written in Matlab® R2015a (The MathWorks, Natick, MA, USA) employing YASARA Structure. RMSF values of C, C_α, and N backbone atoms were averaged.

10.2.5 Generation of Protein Phase Diagrams

Protein phase diagrams were generated for pH 3, 5, 7 and 9 using sodium chloride and ammonium sulfate as precipitants. Sodium chloride was obtained from Merck (Darmstadt, Germany) and ammonium sulfate was from VWR (Radnor, PA, USA). The precipitant stock solutions were prepared as described in detail by Baumgartner et al. [257]. Phase diagrams for lys_{hum} and lys_{egg} were generated earlier [257]. α-lact concentration in the protein stock solutions at pH 3, 7, and 9 was adjusted to 43.5±1 mg/mL. At pH 5 α-lact could only be concentrated to 7±1 mg/mL due to solubility limitations. The protein phase diagrams were generated by adding 12 µL of diluted protein solution to 12 µL of diluted precipitant solution on MRC under Oil 96 Well Crystallization Plates (Swissci, Neuheim, Switzerland) by using a Freedom EVO® 100 (Tecan, Maennedorf, Switzerland) automated liquid handling station. Resulting α-lact concentrations on the

crystallization plates ranged between 2.5 and 21.75 mg/mL for pH 3, 7, and 9 and between 0.4 and 3.5 mg/mL for pH 5, sodium chloride concentrations ranged between 0 and 2.5 M, ammonium sulfate concentrations ranged between 0 and 1.5 M. Crystallization plates were then centrifuged for 1 min at 1000 rpm to remove air bubbles and covered using optically clear and UV compatible sealing tape (HDclear^{TM} sealing tape, ShurTech Brands, Avon, OH, USA). The sealed plates were stored in a Rock Imager 54 (Formulatrix, Waltham, MA, USA) automated imaging device for 40 days at 20 °C that collected microscopic images of the individual wells of the crystallization plates. The images were examined after 40 days. Six possible phase states were classified: clear solution, crystallization, precipitation, skin formation, gelation and phase separation. For further details on generation of protein phase diagrams in microbatch experiments see Baumgartner et al. [257].

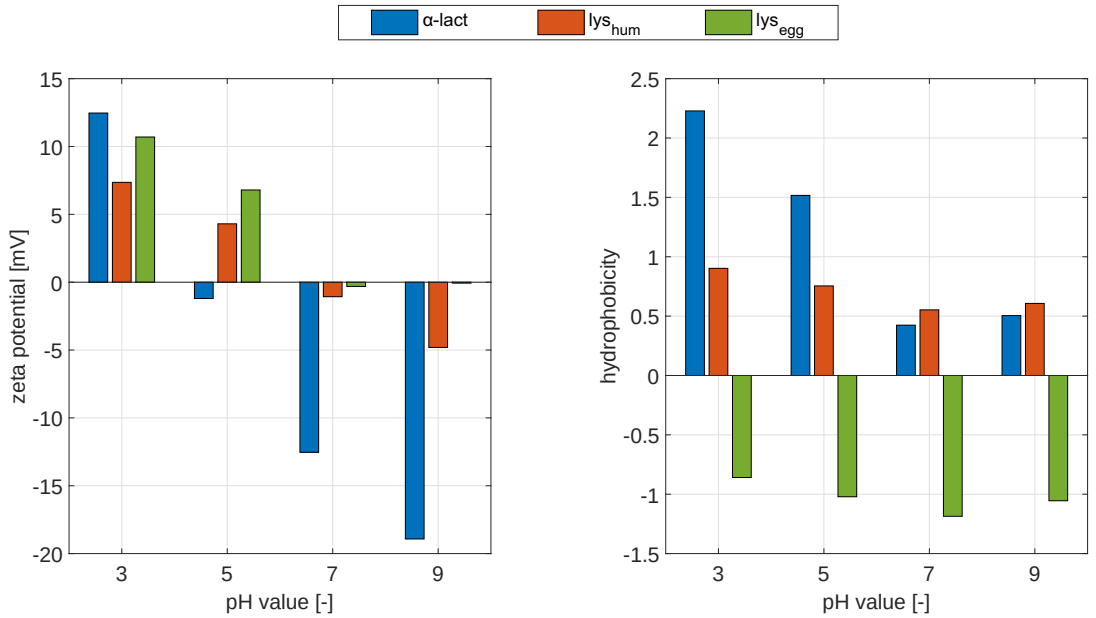
10.3 Results

10.3.1 Protein Assessment in Absence of Precipitant

10.3.1.1 Zeta Potential and Hydrophobic Character

Figure 10.2a shows the values of the zeta potential of α-lact, lys_{hum}, and lys_{egg} at pH 3, 5, 7, and 9. Zeta potentials could be determined with a maximum standard deviation of $\pm$2.6 mV. α-lact exhibited a positive zeta potential for pH 3 and a negative zeta potential for pH 5, 7, and 9. The course of the zeta potential as function of pH is very similar for lys_{hum} and lys_{egg}. lys_{hum} and lys_{egg} exhibited a positive zeta potential at pH 3 and pH 5 and a negative zeta potential for pH 7 and pH 9, though the isoelectric points are expected to be located between pH 10.0 and pH 11.0 [253–256]. However, values for the zeta potential of lys_{hum} and lys_{egg} at pH 7 and pH 9 are only slightly negative. α-lact obtained the highest absolute value of the zeta potential at pH 9 where it reached 18.9 mV. lys_{hum} and lys_{egg} obtained the highest absolute value of the zeta potential at pH 3 with 7.4 mV for lys_{hum} and 10.7 mV for lys_{egg}. Figure 10.2b shows the ranking of the hydrophobic character of α-lact, lys_{hum}, and lys_{egg} at pH 3, 5, 7, and 9 as it was determined in Amrhein et al. [184]. Originally the hydrophobic character was evaluated by means of parameters extracted from a surface tension profile fit as a function of protein concentration [184] and is referred to as hydrophobicity for simplification. Negative values represent a rather hydrophilic character whereas positive values stand

for a rather hydrophobic character. The strongest hydrophobic character among the investigated proteins was obtained by α-lact at pH 3 and pH 5, followed by lys_{hum} at pH 3 and pH 5. Though, the hydrophobic character of α-lact at pH 3 and pH 5 is significantly higher than the hydrophobic character of lys_{hum} at pH 3 and pH 5. α-lact and lys_{hum} both obtained their lowest hydrophobic character among the investigated pH values at pH 7. Negative values for the hydrophobic character, as obtained by lys_{egg}, indicated lys_{egg} to be hydrophilic instead of being hydrophobic. Here, the most pronounced hydrophilic character of lys_{egg} was obtained at pH 7 and the least pronounced hydrophilic character was obtained at pH 3.



(a) (b)

Figure 10.2: (a) Zeta potential and (b) surface hydrophobicity of α-lact, lys_{hum}, and lys_{egg} at pH 3, 5, 7, and 9 without salt. Hydrophobicity data were taken from Amrhein et al. [184].

10.3.1.2 Protein Conformational Flexibility Without Precipitants

Figure 10.3 shows the RMSF of the C_α, C, and N backbone atoms of α-lact, lys_{hum}, and lys_{egg} at pH 3, 5, 7, and 9 plotted against the primary sequence in 1-letter code. The

RMSF as function of the primary sequence was calculated as mean value from the 6-fold simulations with included outlier analysis. High RMSF values in this context correlate with a high protein flexibility. It is apparent from Figure 10.3 that there are areas in the primary sequence of α-lact, lys_{hum}, and lys_{egg} with enhanced flexibility and that the degree of flexibility, at least in the peak regions, is pH dependent. Despite the structural similarity of the three proteins, the RMSF profile of α-lact differs significantly from the RMSF profiles of lys_{hum} and lys_{egg}. Generally, the sequence based analysis identified the β-sheet and α-helix structures as highly conserved whereas random coil structures and the terminals were highly flexible.
For further investigation a pH dependent flexibility ranking was developed, what caused the need to reduce the information from Figure 10.3a, 10.3b, and 10.3c to only one parameter per protein and pH value. Therefore, the mean value of all RMSF values of the primary sequence at fixed pH was calculated. The pH dependent flexibility ranking is shown in Figure 10.3d. Conformational flexibility of α-lact was highest amongst the investigated protein species and conditions followed by lys_{egg}. The highest mean RMSF value for α-lact was 0.77 Å at pH 3. lys_{egg} exposed the highest flexibility at pH 3 as well, with a mean RMSF of 0.69 Å. The mean RMSF values of lys_{egg} at pH 5 and 9 are both 0.65 Å and are slightly lower at pH 7 with 0.62 Å. lys_{hum} exhibited the lowest conformational flexibility amongst the investigated proteins. The highest mean RMSF value and thus the highest flexibilities of lys_{hum} were observed at pH 9 and pH 3 (0.6 Å). The lowest flexibility of lys_{hum} was observed at pH 5 (0.57 Å). Generally, it has to be mentioned that mean RMSF values within one protein species as function of pH only vary slightly whereas mean RMSF values of different protein species vary significantly.

10.3.1.3 Protein Phase Behavior Without Precipitants

The protein phase states of α-lact, lys_{hum} and lys_{egg} at pH 3, 5, 7, and 9 in the absence of ammonium sulfate or sodium chloride can be gathered from Figure 10.4 or Figure 10.5 respectively. Phase transitions in the investigated protein concentration range between 2.5 and 21.75 mg/mL were observed for α-lact at pH 3 and pH 5. At pH 3 α-lact solutions gelated for α-lact concentrations of 19 and 21.75 mg/mL. At pH 5 α-lact precipitated for concentrations from 2.17 to 3.5 mg/mL. As described before, at pH 5 the α-lact stock solution could only be concentrated up to 7 ± 1 mg/mL. lys_{hum} and lys_{egg} stayed soluble in the investigated pH and protein concentration range in the absence of sodium chloride and ammonium sulfate.

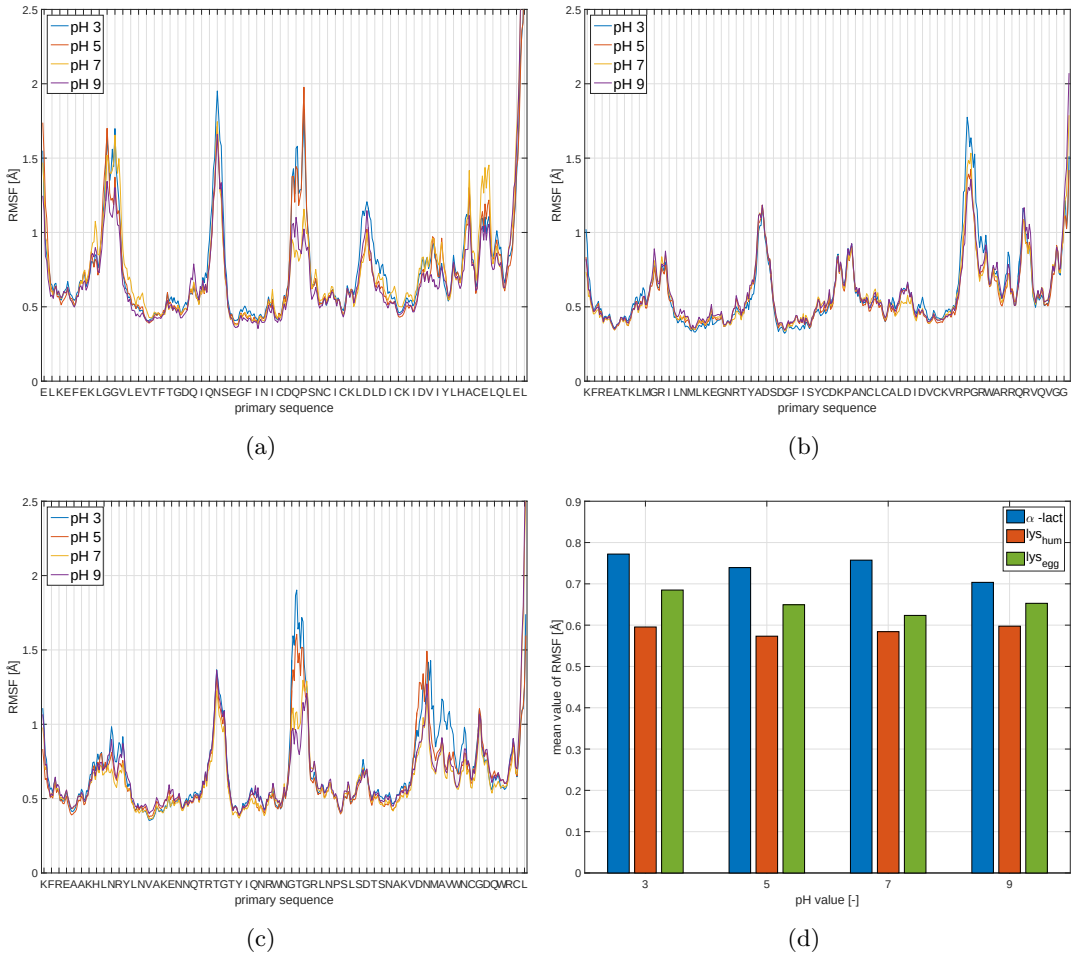


Figure 10.3: Root-mean-square-fluctuation (RMSF) of the C_α, C, and N backbone atoms of (a) α-lact, (b) lys_{hum}, and (c) lys_{egg} at pH 3, 5, 7, and 9 plotted against the primary sequence in 1-letter code. Only every second amino acid is indicated for clarity. Subfigure (d) illustrates the averaged RMSF of α-lact, lys_{hum}, and lys_{egg} in dependency of pH.

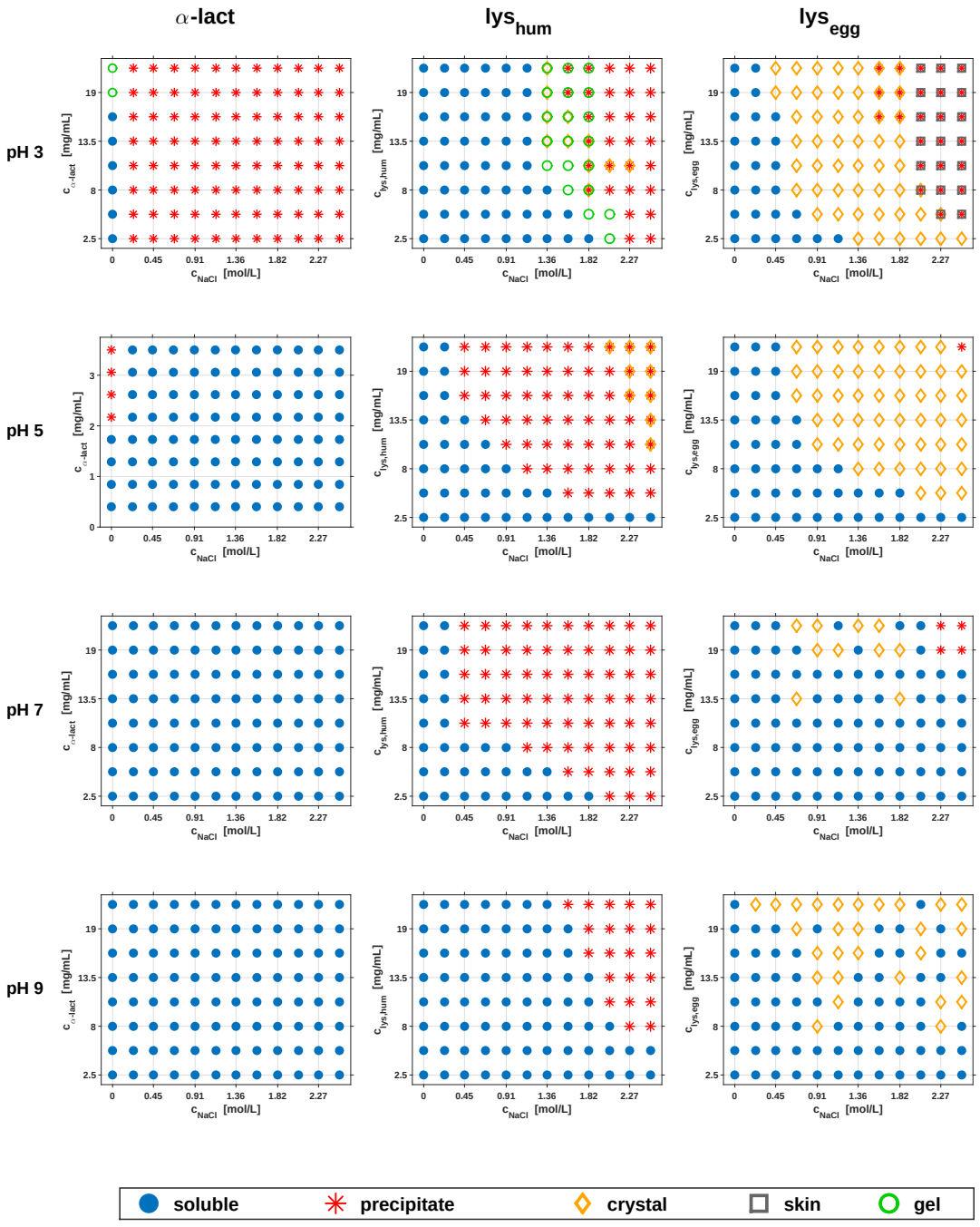


Figure 10.4: Phase diagrams of α-lact, lys_{hum}, and lys_{egg} at pH 3 (first row), pH 5 (second row), pH 7 (third row), and pH 9 (fourth row) using sodium chloride as precipitant. Data for the phase diagrams of lys_{hum} and lys_{egg} was taken from Baumgartner et al. [257].

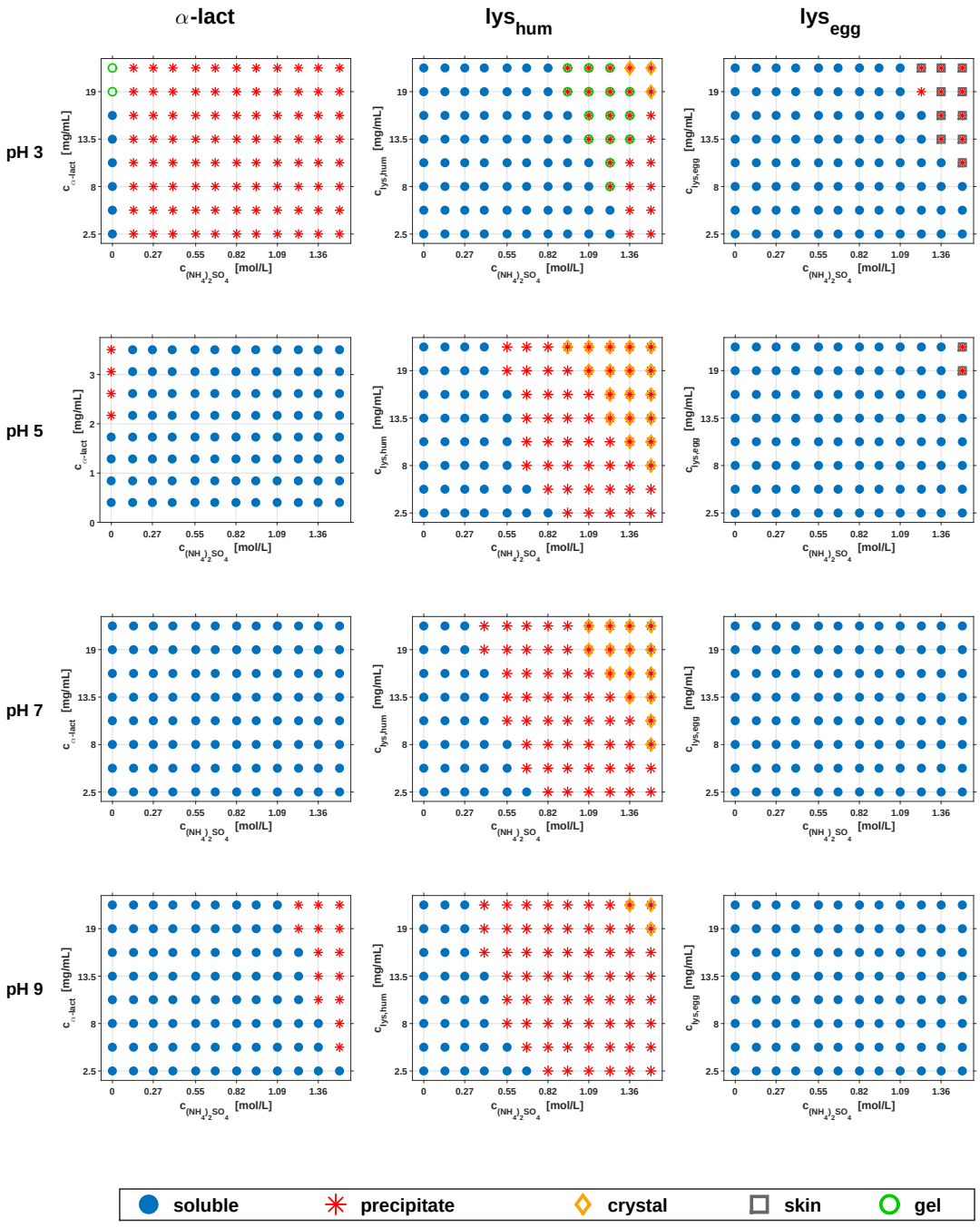


Figure 10.5: Phase diagrams of α-lact, lys_{hum}, and lys_{egg} at pH 3 (first row), pH 5 (second row), pH 7 (third row), and pH 9 (fourth row) using sodium chloride as precipitant. Data for the phase diagrams of lys_{hum} and lys_{egg} was taken from Baumgartner et al. [257].

10.3.2 Protein Assessment in Presence of Precipitant

10.3.2.1 Protein Phase Behavior Using Sodium Chloride and Ammonium Sulfate as Precipitant

Figure 10.4 and Figure 10.5 show the phase diagrams of α-lact, lys_{hum}, and lys_{egg} at pH 3 (first row), pH 5 (second row), pH 7 (third row), and pH 9 (fourth row) with sodium chloride and ammonium sulfate as precipitants. Phase behavior of α-lact is similar for pH 3 and pH 5 regardless of using sodium chloride or ammonium sulfate as precipitant. At pH 3 the first phase transition of α-lact occurred for 0.23 M sodium chloride and for 0.14 M ammonium sulfate. At pH 5 addition of sodium chloride and ammonium sulfate stabilized α-lact as the former precipitated states could be transformed to soluble phase states. At pH 9 phase transitions of α-lact could only be observed for ammonium sulfate as precipitant. At pH 9 addition of 1.23 M ammonium sulfate induced the first phase transition of α-lact. No influence of both precipitants on α-lact phase behavior could be observed at pH 7 and no influence of sodium chloride could be observed at pH 9. For lys_{hum} phase transitions induced by sodium chloride and ammonium sulfate were observed at pH 3, 5, 7, and 9. Using sodium chloride as precipitant the first phase transition of lys_{hum} was observed for 1.36 M sodium chloride at pH 3, for 0.45 M sodium chloride at pH 5 and pH 7, and for 1.59 M sodium chloride at pH 9. Using ammonium sulfate as precipitant the first phase transition of lys_{hum} was observed for 0.95 M ammonium sulfate at pH 3, for 0.55 M ammonium sulfate at pH 5, and for 0.40 M ammonium sulfate at pH 7 and pH 9. For lys_{egg} phase transitions were observed at pH 3, 5, 7, and 9 using sodium chloride as precipitant and at pH 3 and pH 5 using ammonium sulfate as precipitant. The first phase transition of lys_{egg} induced by sodium chloride was observed for 0.45 M sodium chloride at pH 3, for 0.68 M sodium chloride at pH 5 and pH 7, and for 0.23 M sodium chloride at pH 9. The first phase transition of lys_{egg} induced by ammonium sulfate was observed for 1.23 M ammonium sulfate at pH 3 and for 1.5 M ammonium sulfate at pH 5.

For α-lact the predominant phase state beside the soluble state was precipitate independent of the precipitant used. The same accounted for lys_{hum}, whereas here for very high sodium chloride or ammonium sulfate concentrations in some cases crystals and precipitate coexisted. Something special to mention is that at pH 3 for α-lact and lys_{hum} gelation was observed. The predominant phase state beside the soluble one for lys_{egg} depends on the precipitant used. For sodium chloride as precipitant lys_{egg} readily crys-

tallized whereas for ammonium sulfate as precipitant skin formation and precipitation were observed.

10.3.2.2 Conformational Flexibility of Human Lysozyme in the Presence of Sodium Chloride and Ammonium Sulfate

As described above the RMSF of the C_α, C, and N backbone atoms was used as a measure for conformational flexibility. By means of lys_{hum} the influence of ammonium sulfate and sodium chloride as precipitants was studied *in-silico* in dependency of pH and the concentration of sodium chloride and ammonium sulfate respectively. The RMSF profiles of human lysozyme are illustrated in Figure 10.6 and Figure 10.7, where flexible and less flexible regions can be identified throughout pH and salt concentration. The influence of sodium chloride significantly differs from ammonium sulfate. Sodium chloride seemed to reduce the RMSF of highly flexible regions throughout the sequence with increasing concentration at all simulated pH values in comparison to the system in absence of salt. Contrary, ammonium sulfate increased the RMSF at pH 5, 7 and 9. Interestingly, a reduction in RMSF could be observed for ammonium sulfate at pH 3 at certain rather flexible regions.

10.4 Discussion

10.4.1 Aggregation Propensity of α-Lactalbumin, Human Lysozyme, and Lysozyme From Chicken Egg White Without Precipitants and Its Correlation to the Zeta Potential, the Hydrophobic Character and Conformational Flexibility

In the following, aggregation propensity will be estimated by the minimal protein concentration at which a phase transition, i.e. aggregation, of α-lact, lys_{hum}, and lys_{egg} without precipitants was observed. In the investigated protein concentration range between 2.5 and 21.75 mg/mL phase transitions could only be observed for α-lact at pH 3 and pH 5, which is why aggregation propensity without precipitants is highest for α-lact amongst the investigated proteins. Here, α-lact was identified to have the strongest hydrophobic character of all protein species and pH values investigated. Thus, aggregation propensity correlates with the hydrophobic character of the proteins, i.e. the higher the hydrophobic

character, the higher the aggregation propensity without precipitants. However, comparing α-lact at different pH values, its aggregation propensity was clearly higher at pH 5 than at pH 3 though the hydrophobic character of α-lact at pH 3 is stronger than at pH 5. This discrepancy between the hydrophobic character and aggregation propensity might be explained by the differing zeta potential of α-lact at pH 3 and pH 5. At pH 3 the absolute value of the zeta potential of α-lact is 12.5 mV whereas it is 1.2 mV and thus close to zero at pH 5. Thus, at pH 3 electrostatic repulsion seems to be strong enough to delay aggregation. However, it is important to mention that according to the classical DLVO theory for colloids all measured absolute zeta potentials throughout the proteins are smaller than $\pm$30 mV and thus are considered to be insufficient for stabilization of colloidal suspensions [264]. Therefore, the zeta potential in any of the investigated solutions is not high enough to stabilize the proteins in solution solely by electrostatic repulsion.

Protein flexibility as well correlates with aggregation propensity as conformational flexibility of α-lact at pH 3 and pH 5 is significantly higher than for all other investigated protein species and pH values. Analogous to hydrophobicity, the conformational flexibility of α-lact at pH 3 was determined to be higher than conformational flexibility of α-lact at pH 5, assuming a more pronounced aggregation propensity of α-lact at pH 3 compared to pH 5. This discrepancy between protein flexibility and aggregation propensity might be explained by the occurrence of α-lact gelation at pH 3 as gelation represents a very special aggregate state. Denaturation is considered to be a prerequisite to gelation [265] and denaturation is a profound conformational change that can neither be reflected in the performed MD simulations nor in protein flexibility. No statement about the predictive power of protein surface characteristics and conformational flexibility for pH dependent aggregation propensities of $\mathrm{lys_{hum}}$ and $\mathrm{lys_{egg}}$ was possible as both proteins stayed soluble in the investigated protein concentration and pH range.

In summary, differences regarding aggregation propensities of α-lact, $\mathrm{lys_{hum}}$, and $\mathrm{lys_{egg}}$ in aqueous solutions without precipitants at pH 3, 5, 7, and 9 could be related to the hydrophobic character and to conformational flexibility of these proteins. Hydrophobic surface characteristics and protein flexibility seem to be sufficient to predict differing aggregation propensities for different protein species in aqueous solutions without precipitants. The pH dependent discrepancy in aggregation propensities within one protein species might not be predicted using solely hydrophobic surface characteristics and protein flexibility, but by additionally taking electrostatic surface properties into account.

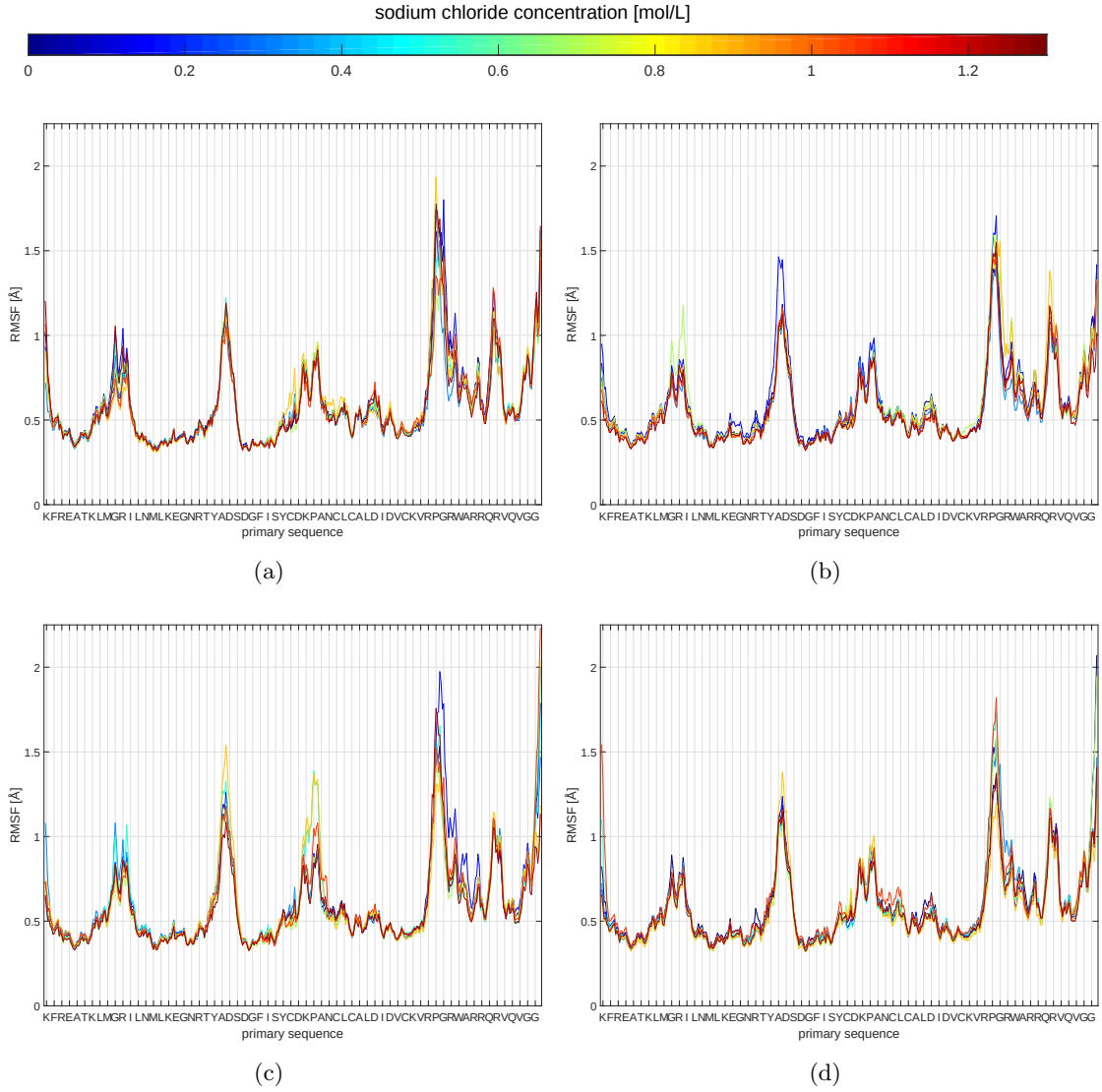


Figure 10.6: Root-mean-square-fluctuation (RMSF) of the C_α, C, and N backbone atoms of lys_{hum} at (a) pH 3, (b) pH 5, (c) pH 7, and (d) pH 9 in dependency of sodium chloride up to 1.3 mol/mL. The colorbar refers to the molar sodium chloride concentration and only every second amino acid is indicated for clarity.

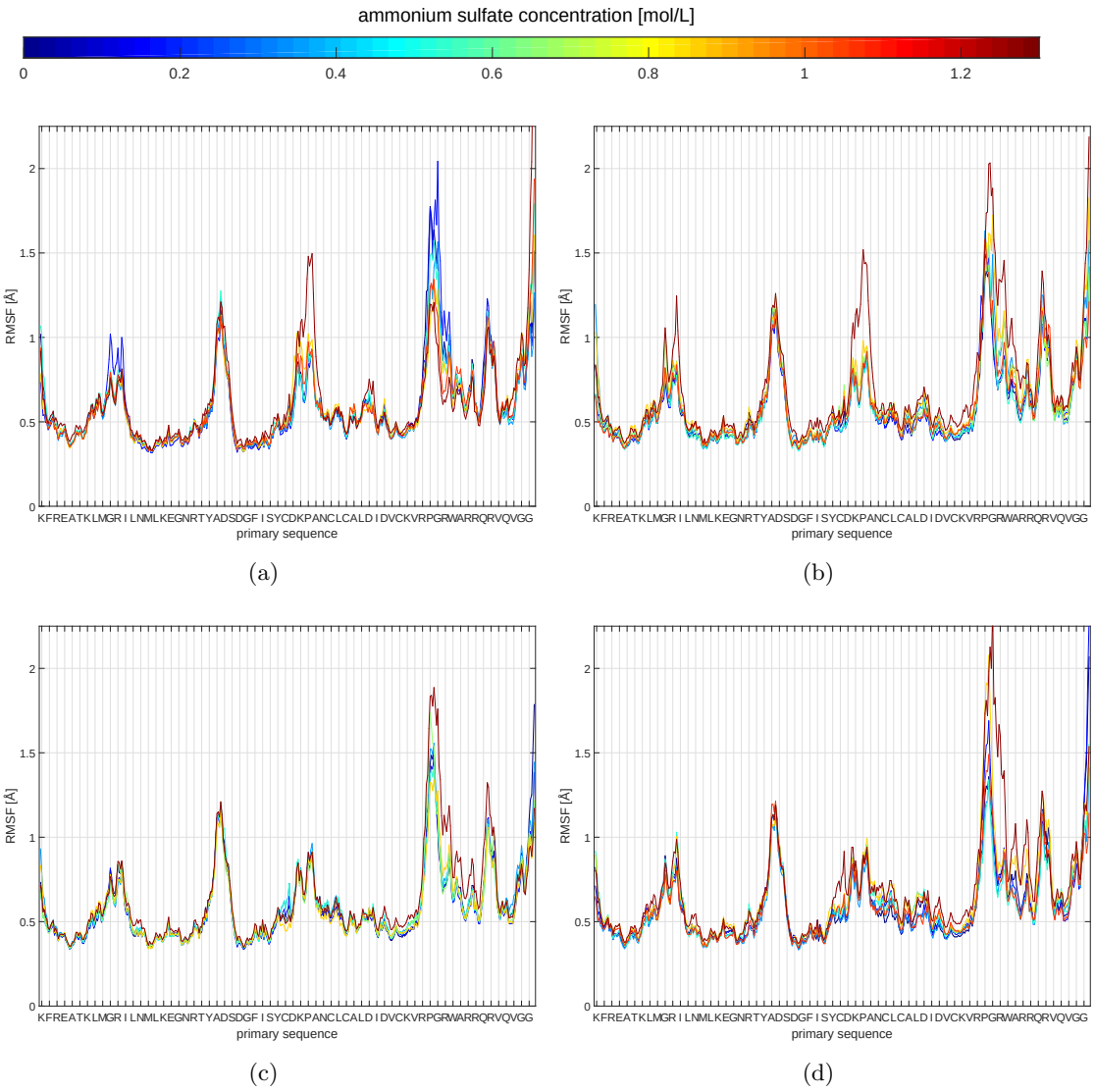


Figure 10.7: Root-mean-square-fluctuation (RMSF) of the C_α, C, and N backbone atoms of lys_{hum} at (a) pH 3, (b) pH 5, (c) pH 7, and (d) pH 9 in dependency of ammonium sulfate up to 1.3 mol/mL. The colorbar refers to the molar ammonium sulfate concentration and only every second amino acid is indicated for clarity.

10.4.2 Aggregation Propensity of α-Lactalbumin, Human Lysozyme, and Lysozyme From Chicken Egg White Using Sodium Chloride and Ammonium Sulfate as Precipitants and Its Correlation to the Zeta Potential, the Hydrophobic Character and Conformational Flexibility

As described above, protein surface characteristics and protein flexibility of α-lact, lys_{hum}, and lys_{egg} were determined in solutions without precipitants. In the previous section it could be shown, that these electrostatic and hydrophobic surface characteristics and protein flexibility could be directly correlated to aggregation propensity without precipitants. In addition to probing the predictive power of protein surface characteristics and protein flexibility for aggregation propensity in solutions without precipitants, their predictive power for the aggregation propensity of protein solutions with sodium chloride or ammonium sulfate as precipitants was examined as well.
Long-range electrostatic repulsion between charged proteins is reduced by the presence of salt ions. These salt ions screen the electrostatic fields of proteins and short-range attractive forces become noticeable [257]. As could be shown in Baumgartner et al. [257] for lysozyme from chicken egg white and sodium chloride concentrations above 0.23 M the effect of electrostatic repulsion seems to vanish. Additionally, in the present study the zeta potential in any of the investigated solutions seems to be not high enough to significantly influence the aggregation propensity of the proteins in solution and will be further reduced by addition of sodium chloride and ammonium sulfate. This is why electrostatic repulsion due to the zeta potential of proteins will be neglected in the following.
The location of the isoelectric point (pI) of a protein and the pH value of the aqueous solution the protein is dissolved in are of high importance. Salts were observed to invert their influence on protein stability for pH values above and below the pI [266]. The pI of α-lact is located between pH 4.25 and 4.5 [252]. This might explain why sodium chloride as well as ammonium sulfate induce precipitation of α-lact at pH 3 and stabilize α-lact against precipitation at pH 5. As the isoelectric points of lys_{hum} and lys_{egg} are located between pH 10.0 and 11.0 [253–256], all pH values investigated in this study are below their isoelectric point. In order to exclude effects related to the location of the pI, α-lact was for the following examination only considered at pH 3, as then protein aggregation propensity is compared for pH values below the isoelectric points of the three proteins.

10.4.2.1 Predictive Power of Hydrophobic Surface Character and Conformational Flexibility for Estimation of Aggregation Propensity Using Sodium Chloride

The predicitive power of protein hydrophobicity and protein flexibility, determined in systems without precipitants, for protein aggregation propensity in presence of sodium chloride is discussed in the following. Aggregation propensities are estimated by the minimal sodium chloride concentration at which a phase transition, i.e. aggregation, of α-lact, lys_{hum}, and lys_{egg} was observed. Aggregation propensity of these three proteins are compared at fixed pH values. At pH 3 α-lact showed the highest aggregation propensity among the three proteins, followed by lys_{egg} and lys_{hum}. As mentioned above, at pH 5, 7, and 9 only lys_{egg} and lys_{hum} will be considered. At pH 5 and pH 7, the aggregation propensity of lys_{hum} is higher than the aggregation propensity of lys_{egg}. At pH 9, the aggregation propensity of lys_{egg} is higher than the aggregation propensity of lys_{hum}. Thus, for pH 3 and pH 9 aggregation propensity as function of protein species (for pH<pI) directly correlates with protein conformational flexibility and for pH 5 and pH 7 it directly correlates with protein hydrophobicity. Furthermore, the results indicate that the classification of a protein as hydrophobic or hydrophilic strongly influences protein phase behavior under addition of salt. For α-lact and lys_{hum} the predominant phase transition induced by addition of sodium chloride was from soluble to precipitate. For lys_{egg} instead the first phase transition induced by sodium chloride was from soluble to crystal. lys_{egg} thus seems to adopt a special behavior as it was the only hydrophilic protein investigated in this work and the only one that readily crystallized under addition of sodium chloride. The two proteins with a rather hydrophobic character instead readily precipitated and crystals could only be found in coexistence with precipitate for high precipitant concentrations. For very high hydrophobicities, as observed for α-lact and lys_{hum} at pH 3, those proteins readily gelled. As for the aggregation propensity without precipitants, no correlation between protein surface hydrophobicity or protein flexibility within one protein species as function of the pH value could be found.

Overall, the results lead to the conclusion that the assessment of protein characteristics in salt free systems is well-suited for estimation of aggregation propensity of different proteins under addition of sodium chloride. Though, estimation of aggregation propensity could not be reduced to solely one protein property. Protein flexibility seems to be best suited for estimation of aggregation propensity at extreme pH values (pH 3 and pH 9), whereas surface hydrophobicity seems to be best suited for moderate to physiological pH values (pH 5 and pH 7). Protein surface hydrophobicity furthermore might indicate the

aggregate state to evolve as hydrophilicity could be related to crystallization and high hydrophobicity to gelation under addition of sodium chloride.

10.4.2.2 Predictive Power of Hydrophobic Surface Character and Conformational Flexibility for Estimation of Aggregation Propensity Using Ammonium Sulfate

The predicitive power of protein hydrophobicity and protein flexibility, determined in systems without precipitants, for protein aggregation propensity in presence of ammonium sulfate will examined in the following. Aggregation propensity will in the following be estimated by the minimal ammonium sulfate concentration at which a phase transition, i.e. aggregation, of α-lact, lys_{hum}, and lys_{egg} was observed. At pH 3 α-lact exhibited the highest aggregation propensity among the three proteins, followed by lys_{hum} and lys_{egg}. At pH 5 the aggregation propensity of lys_{hum} is higher than the aggregation propensity of lys_{egg}. At pH 7 and pH 9, the aggregation propensity of lys_{hum} is clearly higher than the aggregation propensity of lys_{egg}, as lys_{hum} readily precipitated, whereas lys_{egg} stayed soluble. Thus, for all pH values aggregation propensity as function of protein species (for pH<pI) directly correlates with protein hydrophobicity. This fits the general observation that ammonium sulfate as cosmotropic salt strengthens hydrophobic interactions [239], resulting in increased aggregation propensities for the proteins with hydrophobic character (for pH<pI) compared to a precipitant with chaotropic properties. The predominant phase transition for α-lact and lys_{hum} induced by addition of ammonium sulfate was from soluble to precipitate and crystals could only be found in coexistence with precipitate for high precipitant concentrations. For the protein with hydrophilic character the cosmotropic salt induced skin formation which is according to Zeelen [267] synonymous with protein denaturation. Thus, proteins with hydrophobic character tend to precipitate by addition of cosmotropic salts, whereas proteins with a hydrophilic character tend to crystallize by addition of chaotropic salts but denaturate by addition of cosmotropic salts.
In summary, the results lead to the conclusion that the assessment of protein characteristics, and surface hydrophobicity in particular, in salt free systems seems to be capable to serve as predictive parameter for the phase behavior of different proteins induced by ammonium sulfate. Though, again no clear correlation between surface hydrophobicity or protein flexibility and aggregation propensity as function of pH within one protein species could be found.

10.4.2.3 Predictive Power of Salt Depending Conformational Flexibility for Estimation of Aggregation Propensity of Human Lysozyme

As mentioned above, the transferability of protein properties derived in absence of precipitants was limited to comparison of different protein species regarding aggregation propensity. Differences depending on pH between the protein flexibility in absence of precipitants could not be sufficiently resolved. Thus, the salt impact on protein flexibility was studied in detail by means of lys_{hum} in dependency of salt species as well as salt concentration. lys_{hum} was chosen for investigation as here phase transitions could be observed for sodium chloride and ammonium sulfate at all pH values investigated. It is important to note that MD simulations were conducted for a fixed and diluted protein concentration and no influence of the protein concentration on protein flexibility was considered. This means, that on the one hand protein-protein interactions can not be respected by the chosen design of the MD simulations. On the other hand, the *in-silico* derived data refer to the experimental data with the lowest protein concentration ($c_{protein}$=2.5 mg/mL). Furthermore, the influence of sodium chloride and ammonium sulfate on the RMSF and on phase behavior are compared in the same salt concentration range between 0 and 1.3 mol/L.

The flexibility study of lys_{hum} revealed decreasing flexibility with increasing sodium chloride concentration which indicates a stabilizing effect of sodium chloride. The stabilizing effect is in good agreement with experimental phase behavior, where lys_{hum} is soluble at low protein concentrations up to 1.3 mol/L sodium chloride. This leads to the assumption that the protein concentration, i.e. protein-protein interaction, dominates the phase behavior of lys_{hum} in presence of sodium chloride and is less depending on protein flexibility. Contrary, the flexibility of lys_{hum} was increased with increasing ammonium sulfate concentrations for pH 5, 7, and 9. Here, precipitation occurred even at low protein concentrations within the *in-silico* salt range. Thus, conformational flexibility dominates the phase behavior of lys_{hum} in presence of ammonium sulfate for pH 5, 7, and 9. An RMSF decreasing and stabilizing effect using ammonium sulfate could be observed only for certain regions at pH 3. At this pH value the highest solubility and lowest aggregation propensity respectively as a function of precipitant concentration was observed. The formation of gel was induced by higher protein concentrations. This leads to the assumption, that the formation of gel at pH 3 for lys_{hum} induced by ammonium sulfate is not driven by partial unfolding but rather in conserving conformational stability of certain regions.

Comparing sodium chloride and ammonium sulfate the effect of ammonium sulfate on

the flexibility is more pronounced. Contrary to sodium chloride ammonium sulfate promotes protein flexibility. This is reflected in the phase behavior since solely ammonium sulfate induced phase transitions of diluted lys_{hum} solutions ($c_{protein}$=2.5 mg/mL) for salt concentrations below 1.3 mol/L. This is in good agreement with the Hofmeister series where ammonium sulfate is more chaotropic than sodium chloride and thus has a stronger impact on protein solubility and the conformational stability.
In summary, the flexibility study in dependency of salt species and concentration revealed a more detailed insight into protein stability and phase behavior in presence of salt. Sodium chloride and ammonium sulfate could be clearly distinguished according to their impact on protein solubility and conformational stability by MD simulations and the salt depending flexibility studies were in good agreement to experimental phase behavior.

10.5 Conclusions and Outlook

α-lactalbumin, human lysozyme, and lysozyme from chicken egg white, three proteins very similar in size and secondary and tertiary structure, could be shown to exhibit very different surface characteristics regarding electrostatics and hydrophobicity and very different values for conformational flexibility. In protein solutions without precipitants it could be shown that differences in the aggregation propensity among the different protein species directly correlate to protein surface hydrophobicity and protein flexibility. In order to describe pH dependent differences in the aggregation propensity within one protein species as function of pH the protein net charge needs to be taken into account additionally. Thus, protein net charge, protein surface hydrophobicity and protein flexibility could be identified to be suitable parameters for prediction of aggregation propensities of α-lact, lys_{hum}, and lys_{egg} in aqueous solutions without precipitants. Protein surface hydrophobicity and protein flexibility, determined in salt free solutions, could additionally be shown to be applicable to forecast protein aggregation propensities of different proteins (at fixed pH values) in presence of sodium chloride and ammonium sulfate.
The *in-silico* study of protein conformational flexibility in presence of sodium chloride or ammonium sulfate revealed a more detailed insight into salt depending protein aggregation propensity. The two salt species sodium chloride and ammonium sulfate could be distinguished according to their impact on protein conformational flexibility.
In summary, the study revealed the use of surface characteristics regarding electrostatics

and hydrophobicity and conformational flexibility to be sufficient to predict the aggregation propensity among the different protein species. The impact of sodium chloride and ammonium sulfate on the aggregation behavior could barely be predicted based on protein properties assessed in absence of these salts. However, the comprehensive MD simulation based study of protein conformational flexibility in dependency of salt concentration and species was proved highly promising to describe the salt impact on aggregation sufficiently.

10.6 Acknowledgment

We thank Kai Baumgartner and Juliane Schütz for providing data for the phase diagrams of human lysozyme and lysozyme from chicken egg white. We gratefully acknowledge the financial support by the Federal Ministry of Education and Research (BMBF) (0315342B, 031A095B).

10.7 References

16. Chrysina, E. D., Brew, K. & Acharya, K. R. Crystal structures of apo-and holobovine α-lactalbumin at 2.2-Å resolution reveal an effect of calcium on inter-lobe interactions. *J. Biol. Chem.* **275,** 37021–37029 (2000).

26. Alizadeh-Pasdar, N. & Li-Chan, E. C. Comparison of protein surface hydrophobicity measured at various pH values using three different fluorescent probes. *J. Agric. Food Chem.* **48,** 328–334 (2000).

146. Chennamsetty, N. *et al.* Design of therapeutic proteins with enhanced stability. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 11937–11942 (2009).

184. Amrhein, S., Christin Bauer, K., Galm, L. & Hubbuch, J. Non-invasive high throughput approach for protein hydrophobicity determination based on surface tension. *Biotechnol. Bioeng.* **112,** 2485–2494 (2015).

209. Krieger, E. *et al.* Making optimal use of empirical energy functions: Force-field parameterization in crystal space. *Proteins* **57,** 678–83 (2004).

211. Berendsen, H. J. C. *et al.* Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3684–3690 (1984).

212. Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24,** 1999–2012 (2003).

213. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103,** 8577–8593 (1995).

214. Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **23,** 1623–1641 (2002).

226. Konagurthu, A. S., Whisstock, J. C., Stuckey, P. J. & Lesk, A. M. MUSTANG: A multiple structural alignment algorithm. *Proteins: Struct., Funct., Bioinf.* **64,** 559–574 (2006).

228. Wang, J. *et al.* Triclinic lysozyme at 0.65 Å resolution. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **63,** 1254–1268 (2007).

229. Durek, T., Torbeev, V. Y. & Kent, S. B. Convergent chemical synthesis and high-resolution X-ray structure of human lysozyme. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 4846–4851 (2007).

230. Gordon, J. C. *et al.* H++: A server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **33,** W368–W371 (2005).

231. Cromwell, M. E., Hilario, E. & Jacobson, F. Protein aggregation and bioprocessing. *AAPS J.* **8,** E572–E579 (2006).

232. Mahler, H.-C., Friess, W., Grauschopf, U. & Kiese, S. Protein aggregation: Pathways, induction factors and analysis. *J. Pharm. Sci.* **98,** 2909–2934 (2009).

233. Philo, J. S. & Arakawa, T. Mechanisms of protein aggregation. *Curr. Pharm. Biotechnol.* **10,** 348–351 (2009).

234. Chi, E. Y., Krishnan, S., Randolph, T. W. & Carpenter, J. F. Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation. *Pharm. Res.* **20,** 1325–1336 (2003).

235. Scopes, R. K. *Protein purification: Principles and practice* (Springer Science & Business Media, 2013).

236. Goto, Y. & Fink, A. L. Conformational states in. beta.-lactamase: Molten-globule states at acidic and alkaline pH with high salt. *Biochemistry* **28,** 945–952 (1989).

237. Wang, W., Li, N. & Speaker, S. External factors affecting protein aggregation. *Aggregation of therapeutic proteins,* 119–204 (2010).

238. Hofmeister, F. Zur Lehre von der Wirkung der Salze (about the science of the effect of salts). *Archiv für experimentelle Pathologie und Pharmakologie* **25,** 1–30 (1888).

239. Wu, S.-L. & Karger, B. L. in *High Resolution Separation and Analysis of Biological Macromolecules Part A: Fundamentals* (ed Barry L. Karger, W. S. H.) 27–47 (Academic Press, 1996).

240. Scopes, R. K. in *Protein Purif. Princ. Pract.* 3rd, 71–101 (Springer-Verlag New York, Inc., 1994).

241. Shukla, A. A. *et al.* Downstream processing of monoclonal antibodies - Application of platform approaches. *J. Chromatogr. B* **848.** Polyclonal and Monoclonal Antibody Production, Purification, Process and Product Analytics, 28–39 (2007).

242. Paul, R. *et al.* Structure and function of purified monoclonal antibody dimers induced by different stress conditions. *Pharm. Res.* **29,** 2047–2059 (2012).

243. Brunsteiner, M., Flock, M. & Nidetzky, B. Structure based descriptors for the estimation of colloidal interactions and protein aggregation propensities. *PloS one* **8** (2013).

244. Lauer, T. M. *et al.* Developability index: A rapid in silico tool for the screening of antibody aggregation propensity. *J. Pharm. Sci.* **101,** 102–115 (2012).

245. Valerio, M. *et al.* Early events in protein aggregation: Molecular flexibility and hydrophobicity/charge interaction in amyloid peptides as studied by molecular dynamics simulations. *Proteins: Struct., Funct., Bioinf.* **58,** 110–118 (2005).

246. Sousa, R. Use of glycerol, polyols and other protein structure stabilizing agents in protein crystallization. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **51,** 271–277 (1995).

247. Yadav, S. *et al.* The influence of charge distribution on self-association and viscosity behavior of monoclonal antibody solutions. *Mol. Pharm.* **9,** 791–802 (2012).

248. Yadav, S., Shire, S. J. & Kalonia, D. S. Viscosity analysis of high concentration bovine serum albumin aqueous solutions. *Pharm. Res.* **28,** 1973–1983 (2011).

249. Kumar, V., Dixit, N., Zhou, L. L. & Fraunhofer, W. Impact of short range hydrophobic interactions and long range electrostatic forces on the aggregation kinetics of a monoclonal antibody and a dual-variable domain immunoglobulin at low and high concentrations. *Int. J. Pharm.* **421,** 82–93 (2011).

250. Zhang, J. & Topp, E. M. Protein G, protein A and protein A-derived peptides inhibit the agitation induced aggregation of IgG. *Mol. Pharm.* **9,** 622–628 (2012).

251. Smith, R. D. *et al.* New developments in biochemical mass spectrometry: Electrospray ionization. *Anal. Chem.* **62,** 882–899 (1990).

252. Robbins, F. M., Kronman, M. J. & Andreotti, R. E. Inter- and intramolecular interactions of α-lactalbumin V. The effect of amidination on association and aggregation. *Biochim. Biophys. Acta, Biophys.* **109,** 223–233 (1965).

253. Lüllig, H. *et al.* Isolation and characterization of human pneumocytes. *Prax. Klin. Pneumol.* **37,** 842 (1983).

254. Lundblad, G. & Lind, J. Studies on lysozyme from human leucemic urine by isoelectric focusing. *Acta Chem. Scand* **26,** 1771–1713 (1972).

255. Righetti, P. G. & Caravaggio, T. Isoelectric points and molecular weights of proteins: A table. *J. Chromatogr. A* **127,** 1–28 (1976).

256. Righetti, P. G., Tudor, G. & Ek, K. Isoelectric points and molecular weights of proteins: A new table. *J. Chromatogr. A* **220,** 115–194 (1981).

257. Baumgartner, K. *et al.* Determination of protein phase diagrams by microbatch experiments: Exploring the influence of precipitants and pH. *Int. J. Pharm.* **479,** 28–40 (2015).

258. Hunter, R. Zeta potential in colloid science academic. *New York,* 69 (1981).

259. Haynes, W. M. *CRC handbook of chemistry and physics* (CRC press, 2014).

260. Tans, A. A new type of nomogram. Aqueous ammonium sulfate solutions. *Ind. Eng. Chem.* **50,** 971–972 (1958).

261. VDI Gesellschaft and VDI-Gesellschaft Verfahrenstechnik und Chemieingenieurwesen. *VDI heat atlas* (Springer Science & Business Media, 2010).

262. Martínez, J. M. & Martínez, L. Packing optimization for automated generation of complex system's initial configurations for molecular dynamics and docking. *J. Comput. Chem.* **24,** 819–825 (2003).

263. Martínez, L., Andrade, R., Birgin, E. G. & Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **30,** 2157–2164 (2009).

264. Greenwood, R. Review of the measurement of zeta potentials in concentrated aqueous suspensions using electroacoustics. *Adv. Colloid Interface Sci.* **106,** 55–81 (2003).

265. Ziegler, G. R. & Foegeding, E. A. The gelation of proteins. *Adv. Food Nutr. Res.* **34,** 203–298 (1990).

266. Boström, M. *et al.* Why forces between proteins follow different Hofmeister series for pH above and below pI. *Biophys. Chem.* **117,** 217–224 (2005).

267. Zeelen, J. P. in *Protein Cryst.* (ed Bergfors, T. M.) 2nd, 175–194 (Internat'l University Line, 2009).

CHAPTER 11

Characterization of Aqueous Two phase Systems by Combining Lab-on-Chip Technology With Robotic Liquid Handling Stations

Sven Amrhein, Marie-Luise Schwab, Marc Hoffmann, and Jürgen Hubbuch

Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Science, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany

Journal of Chromatography A 1367 (2014): 68-77.

Abstract

Over the last decade, the use of design of experiment approaches in combination with fully automated high throughput (HTP) compatible screenings supported by robotic liquid handling stations (LHS), adequate fast analytics and data processing has been developed in the biopharmaceutical industry into a strategy of high throughput process development (HTPD) resulting in lower experimental effort, sample reduction and an overall higher degree of process optimization. Apart from HTP technologies, lab-on-chip technology has experienced an enormous growth in the last years and allows further reduction of sample consumption. A combination of LHS and lab-on-chip technology is highly desirable and realized in the present work to characterize aqueous two phase systems with respect to tie lines. In particular, a new high throughput compatible approach for the characterization of aqueous two phase systems regarding tie lines by exploiting differences in phase densities is presented. Densities were measured by a standalone micro fluidic liquid density sensor, which was integrated into a liquid handling station by means of a developed generic *Tip2World* interface. This combination of liquid handling stations and lab-on-a-chip technology enables fast, fully automated, and highly accurate density measurements. The presented approach was used to determine the phase diagram of ATPSs composed of potassium phosphate (pH 7) and polyethylene glycol (PEG) with a molecular weight of 300, 400, 600 and 1000 Da respectively in the presence and in the absence of 3 %(w/w) sodium chloride. Considering the whole ATPS characterization process, two complete ATPSs could be characterized within 24 h, including four runs per ATPS for binodal curve determination (less than 45 min/run), and tie line determination (less than 45 min/run for ATPS preparation and 8 h for density determination), which can be performed fully automated over night without requiring man power.
The presented methodology provides a cost, time and material effective approach for characterization of ATPS phase diagram on base on highly accurate and comprehensive data. By this means the derived data opens the door for a more detailed description of ATPS towards generating mechanistic based models, since molecular approaches such as MD simulations or molecular descriptions along the line of QSAR heavily rely on accurate and comprehensive data.

Keywords: *ATPS; Tie line; Density; High throughput; Phase Separation; Lab-on-a-chip; Microfluidics*

11.1 Introduction

The development of biopharmaceuticals is a cost and time consuming procedure which is driven by time to market demands and administrative requirements in regard to process robustness and product quality. In the theme of high throughput process development (HTPD) several methods have been developed which can cope with the industrial relevant issues regarding time to market demands, material consumption, cost efficiency and process robustness according to quality-by-design (QbD) requirements [123, 268–271].
Seeking for alternative drug purification technologies aqueous two phase systems (ATPSs) have arouse enormous interest due to cost efficiency, tolerance of solid particles, solid particle removal, scalability and gentle conditions required for biopharmaceuticals [123, 272]. Aqueous two phase systems are composed of two polymers (e.g. polyethylene glycol (PEG) and dextran) or of polymer and salt (e.g. PEG and phosphate). These systems expose a miscibility gap which is defined by the respective binodal curve. A system point located in the two phase region leads to phase separation along the tie line, whose ends show the compositions of the top and bottom phases that exist in equilibrium with each other. A number of thermodynamic predictive approaches were reported for the calculation of physiochemical properties, density, and activity [273, 274]. However, due to complexity of ATPSs the application of these models on industrial relevant ATPSs processes might be challenging. There is still a lack of understanding of ATPS phase formation and protein partitioning mechanisms. Thus process development and optimization is still realized heuristically requiring cost and time intensive experimental screening. In order to react on these shortcomings a number of studies focused on the screening for adequate ATPSs for a number of highly selective purification tasks for valuable biopharmaceutical products in the last years [27, 45, 123, 275].
To shed some light in the mechanistic understanding of phase formation but also to improve ATPS screening for industrial challenges, the initial characterization of ATPSs is highly important regarding the determination of binodal curve and tie lines which comprise the most important system properties such as phase composition and mass ratio of top and bottom phase. These properties could provide an insight into the mechanisms responsible for the protein partition [272] and could be used for molecular dynamics simulations or quantitative structure–activity relationship (QSAR) models, which have been shown to be potential approaches for mechanistic modeling protein partitioning [276, 277]. The binodal curves are used to design the screening space as well as for tie line determination. The tie lines enable the calculation of phase composition of top

and bottom phase. This information can be highly important for the comparability to precipitation by phase forming component or conditioning processes for further chromatography. The commonly used binodal curves are evaluated by cloud point method also known as turbidity titration which exploits the transition from a clear to a turbid solution when entering the two phase region. This methodology was established as a high throughput technique by Bensch et al. [278]. A number of different techniques for tie line determination have been reported previously, such as gravimetric analysis [279] or the use of conductivity (salt) and refractive index (PEG) analysis [272, 280, 281]. However these techniques are not able to cope with high throughput requirements. Bensch et al. [278] have reported on extensive high throughput screening techniques in downstream processing with respect to the preparation, characterization and optimization of ATPS. For evaluating the tie lines a dye is dissolved in the ATPS, which exclusively partitioned into the top phase and can be quantified by absorbance measurement within a few minutes. The top phase volume correlates to the concentrating factor of the dye. The tie lines can be calculated approximately by applying the lever arm rule on the phase volumes under the consideration that density differences of phase solutions are negligible. This approach has been shown to be sufficient for fast ATPS process development [123, 268, 278]. However, there are some requirements to the selected ATPS with respect to dye stability and partitioning. Moreover, a serious drawback is the assumption of negligible densities alluding to an increasing simplification with increasing tie line length. Liquid handling inaccuracies at top and bottom phase sampling as well as meniscus forming depending on phase compositions can be error sources for absorbance measurement.
To get a higher accuracy in characterizing ATPSs additional measurement methods seem necessary. Apart from HTS techniques, another promising technology is the 'micro total analysis system', also called *lab-on-a-chip* system, which has experienced an enormous growth over the last years [282, 283] and offers unique advantages in sample handling, reagent mixing, separation and detection [284]. This technology is applied in a number of different fields from standard operations like cell cultivation, reactors and mixers or cell counting and flow cytometry to applications in clinical diagnostics, cancer research and drug discovery and screening [285]. Lab on chip systems seem promising to fulfill the requirements of to be compatible to HTS demands as they require a low sample volume. Unfortunately, a number of *lab-on-a-chip* devices are not compatible to liquid handling stations (LHSs) because they do not meet design requirements to be supplied with samples using a robotic manipulator arm (RoMa) or using a liquid handling arm (LHA). A combination of *lab-on-a-chip* technologies with liquid handling stations can improve the process development and process optimization enormously and is highly desirable.

Waldbaur et al. [286] have reported previously on a generic microfluidic interface design which enables the use of microfluidic chips on LHSs.
In this work we have combined both technologies in order to characterize ATPSs with respect to tie lines. We present a new high throughput compatible approach for tie line determination by means of phase densities. Phase densities of top and bottom phase are determined in a HTS method by combining LHSs and micro fluidic density sensor and are used for evaluating tie lines. This offers the unique advantage of considering differences in phase densities and gaining density data within the screening, which are important information with respect to feasibility and processivity issues downstream and further development of thermodynamic models for ATPS. A standalone lab-on-chip micro fluidic liquid density sensor was converted into a integrated fully automated liquid handling station by using a *Tip2World* interface we present and explain in detail. This *Tip2World* interface is capable to transform standalone lab-on-chip systems along other analytical devices into fully automated instruments integrated into liquid handling stations. ATPSs composed of low molecular polyethylenglycols and potassium phosphate at pH 7 with and without NaCl were characterized with respect to binodal curves, phase densities and tie lines by applying the new approach described in detail in this work.

11.2 Materials and Methods

11.2.1 Preparation of Stock Solutions

Polyethylenglycol of analytical grade with different molecular weights was purchased from Sigma Aldrich (St. Louis, MO, USA). Stock solutions of polymers were prepared by mixing masses of polymer and water, purified by an Ariums® pro UV system (Sartorius Stedim Biotech, Göttingen, Germany), in order to reach the desired polymer concentration, which was 70 %(w/w) (PEG300, PEG400), 60 %(w/w) (PEG600) and 40 %(w/w) (PEG1000). Dipotassium phosphate and potassium phosphate of analysis grade were purchased from BDH Prolabo (Radnor, PA, USA). The added mass of potassium was included into the calculation of the total mass ratio of 40 %(w/w) potassium phosphate solutions. Masses of the basic and acid component of potassium phosphate were combined in order to reach a pH of 7.0 at room temperature. Masses of sodium chloride of analysis grade (Merck KGaA, Darmstadt, Germany) and ultra pure water were combined in order to reach a sodium chloride stock solution with 25 %(w/w) NaCl.

11.2.2 Liquid Handling Stations

All experiments were performed in a high throughput format using two fully automated liquid handling stations (LHS), namely Freedom EVO® 200 and Freedom EVO® 100 stations (Tecan, Crailsheim, Germany). Freedom EVO® 200 was used for ATPS preparation and binodal determination. This LHS exposed an 8-channel liquid handling arm with teflon coated fixed tips connected to 1 ml dilutors. In addition, this LHS was equipped with a centric gripper, and a orbital shaker (*Te-shake®* , Tecan, Crailsheim, Germany). Tie line determination and density screening respectively were performed by means of a Freedom EVO® 100. This LHS was equipped with 8-channel liquid handling arm with stainless steel fixed tips connected to 1 ml dilutors. In addition, this LHS was equipped with a sensor for liquid density determination, which is explained in detail in subsection 11.2.2.1. Compatibility between LHS and this sensor regarding sample supply by the liquid handling arm were realized by means of the *Tip2World* interface explained in detail in subsection 11.2.2.2. Both LHSs were controlled by Evoware 2.5 SP2. Data transfer with respect to pipetting volumes was automatically performed in Evoware programs employing Excel® (Microsoft, Redmond, WA, USA) files.

11.2.2.1 Optimization of Liquid Handling

The precision of liquid handling in terms of aspiration and dispensing was optimized on the Freedom EVO® 200 by adjusting all relevant parameters like dispensing and aspiration speed and height, air gap size and hold times for each solution. The pipetted volume was calibrated via the dispensed volume using an analytical balance (WXTS205DU, Mettler-Toledo, Greifensee, Switzerland). The dispensed masses were correlated with the respective density at the temperature measured at the time of experiment in order to calculate the dispensed volume. The densities as a function of solution temperature were derived from a non-bypass version of a flow through micro Coriolis liquid density sensor (MicroLDS) purchased from Issys (Integrated Sensing Systems, Inc., Ypsilanti, USA). The heart of the MicroLDS is a U-shaped, vacuum-sealed resonating silicon microtube and is described in detail elsewhere [287, 288]. The tube filled with fluid is driven into resonance electrostatically and its motion is sensed. This enables the determination of liquid density ρ by means of equation 11.1:

$$\rho = \frac{1}{V} \cdot \left[\left(\frac{K_S}{4\pi^2 f^2} - m_t \right) \right], \tag{11.1}$$

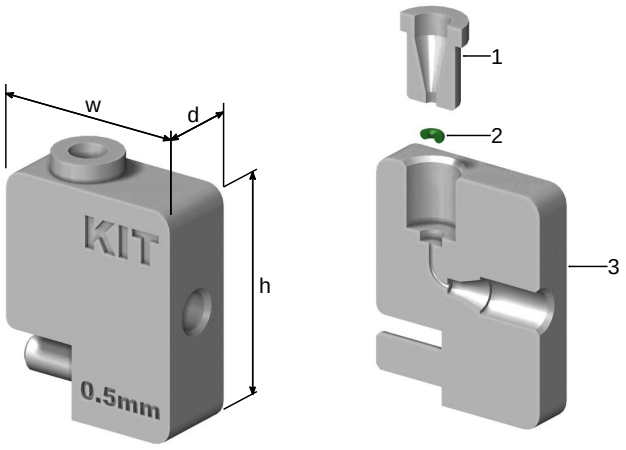
where V represents the internal volume of the resonant tube, K_S is the spring constant of the tube, m_t is the tube mass and f is the frequency of the tube. This lab-on-chip technology allows to derive a highly accurate and temperature dependent density of low volume samples.
In order to determine the temperature depending density all stock solutions were cooled down to 8°C and supplied to the cooled MicroLDS. The MicroLDS was slowly heated to 30°C over a time range of at least 4 h to minimize temperature gradients within the sensor. The density and temperature were recorded using Matlab® R2013b (The MathWorks, Natick, ME, USA). The density-temperature profiles were fitted to a Gaussian function as described in equation 11.2:

$$\rho(T) = \alpha \cdot exp(-((T - \beta)/\gamma)^2), \tag{11.2}$$

where ρ stands for the density and T represents the temperature [°C]. α, β, and γ are the fitting parameters. This guarantees a high accuracy of pipetting for liquid class optimization as well as for pipetting during experiment independently of temperature deviations.

11.2.2.2 Tip2World Interface

In order to expand the portfolio of device, which are compatible to the liquid handling stations, we constructed a set of *Tip2World* interfaces to connect tips of the liquid handling stations with standard tubings via fittings for standard thread port configurations, namely 10-32 coned and 1/4-28 flat bottom. These two versions of screw ports open up a wide range of capillary tubings to be connected to the *Tip2World* interface. The construction work was done with the 2D/3D CAD software SolidEdge (Siemens PLM Software, Plano, USA). The *Tip2World* interfaces were printed by a high resolution 3D printer namely Objet Eden260V (Stratasys, Minneapolis, USA) in polypropylene-like *PolyJet* photopolymer (DurusWhite RGD430 and VeroGray RGD85). The composition and functionality of the *Tip2World* interface is explained in detail in the following and is illustrated in Figure 11.1 by means of one version with a 10-32 coned standard output thread and an internal capillary with a diameter of 0.5 mm.



(a) (b)

Figure 11.1: Schematic illustration of the *Tip2World* interface of the dimensions 20 x 9 x 25 mm (w x d x h) (a) assembled and (b) disjointed in components (1) alignment unit, (2) O-ring and (3) distribution block by means of the version with a 10-32 standard port thread and an internal capillary with a diameter of 0.5 mm.

In general, the *Tip2World* interface is composed of one distribution block, a cylindrical alignment unit and an O-ring. The distribution block has an inlet for the LHS-Tip and an outlet port for the tubing and has dimensions of 20 x 9 x 25 mm (w x d x h). The in- and outlet connected internally with either a 0.5 mm or 1.0 mm capillary. The inscription at the edge of the *Tip2World* interface refers to this diameter. Two designs of the outlet port were chosen to connect different tubings. One has a standard 10-32 coned port to connect 1/16" OD and smaller tubings by standard 10-32 coned fittings. The other one has a standard 1/4-28 flat bottom port to connect tubings (1/32" to 3/16") by standard flat bottom fittings. Because of limitations regarding the accuracy of 3D printing only the core drill hole of the outlet port thread was included in the model to be printed. The thread was cut manually using a screw tap for tapped blind holes in order to minimize the thread intake. The cylindrical alignment unit has a cone-shaped hole with a smaller diameter of 1.2 mm which equals to the diameter of the tip end. This cylinder is used to center the tips as they move into the *Tip2World* interface.Between this cylindrical part and the distribution block a precision O-ring (0.74 x 0.02 mm), purchased from HUG (Ergoldingen, Germany), is placed within a cylindrical cut-out in order to tighten the connection between the tip and the distribution block. It consists of FPM75 (Viton) which exhibits a high chemical resistance. The *Tip2World* interface

possesses a cylindrical extrusion, which enables the fast and reversible plug-in connection to a docking station. The horizontal orientation assures that the *Tip2World* interface is not removed when the tip is retracted. The docking station has a micro plate format and can be fixed to a standard Tecan microplate carrier by M3 screws. It possesses eight sockets to fix a maximum of eight *Tip2World* interfaces which can vary with respect to outlet port format, internal capillary connection and material. The docking station enables a fast replacement and exchange of different *Tip2World* interfaces in case of the need to switch between port formats. The assembling of the docking station and the *Tip2World* interface on a liquid handling station is pictured in Figure 11.2.

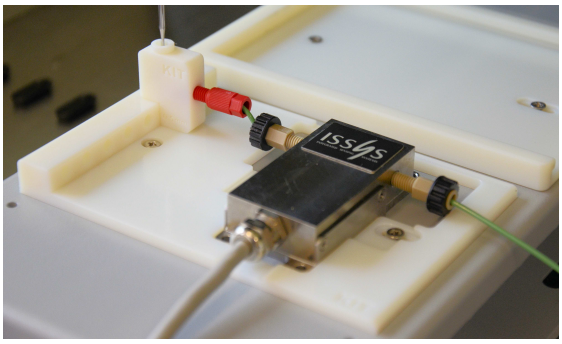


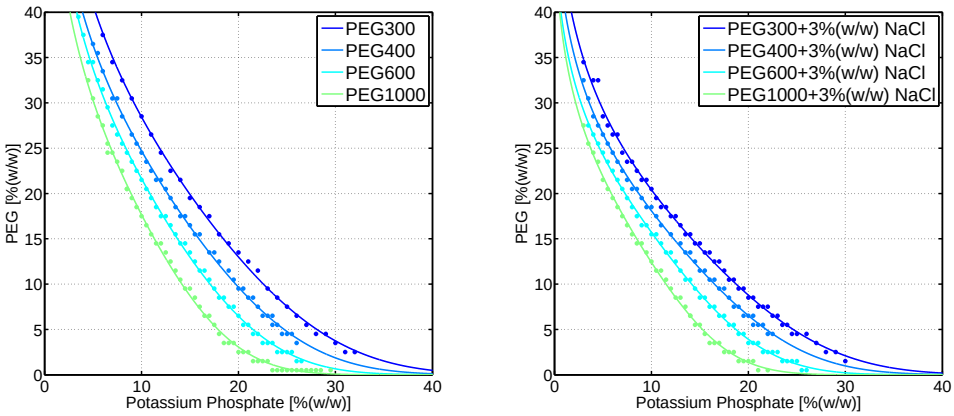
Figure 11.2: A micro Coriolis flow through sensor (MicroLDS) used as a fully automated high throughput sensor by means of the developed *Tip2World* interface and docking station. The docking station holding the MicroLDS as well as the *Tip2World* interface is fixed on a Tecan 3 position plate carrier. The stainless steel tip is immersed into the *Tip2World* interface to supply sample to the MicroLDS.

The *Tip2World* interface was tested for functionality with respect to water as a reference for aqueous solutions and a silicone oil 47 V 20 purchased from BDH Prolabo (Radnor, PA, USA) as a reference for highly viscous solutions, different flow rates and diameter of the used outlet tubing. According to vendor information the silicone oil exhibits a viscosity of 18 to 23 cSt at 25°C. Flow rates starting from 5 µL/s up to 50 µL/s were tested in combination with a high number of capillary tubings with a length of at least 50 cm. In particular, capillaries with a smallest inner diameter of 0.25 mm were used in case of silicone oil and capillaries with a smallest inner diameter of 0.13 mm were tested in case of water. The *Tip2World* interface was only operated with uncoated stainless steel tips since abrasion of tip coating caused by extensive use cannot be excluded and was not investigated in this study.

11.2.3 ATPS Characterization

11.2.3.1 Determination of Binodal Curve

The experiments for the determination of binodal curves were performed using the fully automated liquid handling station (LHS) Freedom EVO® 200 station (Tecan, Crailsheim, Germany). The binodal curves were determined by a turbidity measurement in 96 well micro titer plates (UV-Star® ,Greiner Bio-One, Frickenhausen, Germany). In particular, ultra pure water, one molecular size of PEG and potassium phosphate stock solutions were mixed to a volume of 300 µL in order to reach concentrations in a certain grid which exhibits the binodal curve. All volumes were calculated using the respective density of the stock solutions at the exact temperature during the experiment by applying the density function, which was obtained as described in section 11.2.2.1. After 180 s rigid mixing on an orbital shaker (*Te-Shake®* ,Tecan, Crailsheim, Germany) with a frequency of 1800 rpm phase separation was visually detected by turbidity. two phase systems became turbid by mixing while one-phase systems stayed clear.



(a) (b)

Figure 11.3: Compilation of derived binodal curves of potassium phosphate (pH 7) and different PEG molecular weights (a) without NaCl and (b) in the presence of 3 %(w/w) NaCl.

The individual components were varied as explained in the following. 12 different levels of potassium phosphate concentrations per plate were combined with eight different PEG

concentrations each. The PEG concentration for one potassium phosphate concentration was designed as a equally spaced linear grid varying by 1 %(w/w) steps. For each salt concentration, a binodal point was defined as the mean value of two adjacent system points, of which only one point exposed phase formation. For each system at least 30 points were used for binodal curve determination, correlating to at least three runs. The binodal curve could be fitted on base of an adequate number of points and thus with a sufficient accuracy. The binodal was exponentially fitted by least-square regression according to an empirical curve fit developed by Merchuk et al. [279], as expressed by equation 12.1. This equation was shown to be valid for PEG/phosphate systems [123, 278]:

$$x_{PEG} = f(x_{salt}) = a \cdot exp(b \cdot x^{0.5} + c \cdot x^3), \tag{11.3}$$

where x_{PEG} and x_{salt} represent the content of polymer and salt in mass percent. a, b and c are empirical fit factors.

11.2.3.2 Determination of Phase Densities

The ATPS system preparation was performed on the Freedom EVO® 200 station. The determination of the phase densities was performed on the Freedom EVO® 100. The use of a second LHS for density measurement was due to concerns regarding tip coating abrasion as mentioned above. ATPS were prepared to a total volume of 1 mL in 1.3 mL 96-well Deep Well plates (Nalgene Nunc, Rochester, NY, USA) in a 6-fold repetition. The system volume was selected due to ease sampling as well as to avoid very small pipetting volumes and thus guarantee a sufficient accuracy regarding liquid handling. The systems were prepared by pipetting water and stock solution of potassium phosphate, NaCl and polymer in these order. The system composition was selected according to a custom designed screening grid. In general, the grid design was composed one main axis with eight equidistant points and two minor axis with three eight equidistant points each covering tie lines with increasing tie line length, resulting in a grid as displayed in Figure 11.4a. The angle between major and minor axis was 30° or -30° respectively. The grid thus starts close the binodal curve and ends at the maximal feasible composition with respect to salt and polymer content, achievable with the used stock concentrations. After pipetting the stock solutions the systems were capped with an aluminum sealing film (Axygen Inc, Union City, CA, USA) and mixed by means of an overhead shaker (*Reax 2*, Heidolph, Schwabach, Germany), for at least 30 min at 40 rpm. Following, the

phase separation was accelerated by centrifugation (*Centrifuge 5810 R*, Eppendorf AG, Hamburg, Germany) at 4000 rpm for 10 min. Before density determination plates were sealed with a pre-slit well cap (Nalgene Nunc, Rochester, NY, USA) in order to minimize evaporation. The densities of top and bottom phase were derived from the MicroLDS, which is described in detail in section 11.2.2.1. The interoperability of the liquid handling station and the MicroLDS was enabled using the custom designed *Tip2World* interface, which is described in detail in section 11.2.2.2. The *Tip2World* interface enables to supply samples with the tips of the LHS to devices connected by capillary tubings and thereby transforms the standalone MicroLDS into a high throughput compatible analytic device. The assembly of the *Tip2World* interface and the MicroLDS is pictured in Figure 11.2. Top and bottom phase were sampled by the following procedure to avoid contamination by the complementary phase. 250 µL were sampled from the top phase by aspirating with a flow rate of 20 µL/s at the ATPS surface and the tip was tracked during aspiration. For sampling the bottom phase the tip moved slowly (10 mm/s) to the bottom of the deep well before slowly aspirating 250 µL with 10 µL/s. For both phases a leading airgap of 10 µL and a tailing airgap of 50 µL was chosen. The system airgap was set to 50 µL for the top phase and 80 µL for the bottom phase liquid. The airgaps are important to separate the sample from the system liquid even at increased density depending counter pressure during dispensing into the *Tip2World* interface. The MicroLDS was flushed with 200 µL of sample with a flow rate of 10 µL/s before the density was measured over a period of 15 s without a flow and at atmospheric pressure. This time range corresponds to more than 600 single measurement. The temperature was kept constant at room temperature. In order reduce measurement time, four tips were used for sampling and purging one ATPS. Tip one and three were used for handling top and bottom phase respectively, tip two and four were aspirated ultra pure water. The latter were used for rinsing the MicroLDS 900 µL ultra pure water with a flow rate of 20 µL/s between all measurements to avoid sample carryover. After density determination of one system, all tips were washed thoroughly internally and externally. The data handling was performed by automated routines written in Matlab® R2013b. Samples exposing standard deviation above 0.001 over the measurement period of 15 s were sorted out. The mean of all six replicate samples was used for the tie line determination in order to minimize effects of outliers caused by pipetting inaccuracies. The procedure of tie line determination is described in detail in the following section 11.2.3.3.

11.2.3.3 Determination of Tie Lines by Means of Phase Densities

Tie lines were determined using the lever rule with respect to phase masses of top phase (m_{TP}) and bottom phase (m_{BP}) according to equation 11.4.

$$\frac{\overline{MB}}{\overline{MT}} = \frac{m_{TP}}{m_{BP}}, \tag{11.4}$$

where $\overline{MB}$ and $\overline{MT}$ describes the distance between the mixing point M and the bottom phase BP and between M top phase TP respectively.
The phase masses were calculated by the determined densities of top phase (ϱ_{TP}) and bottom phase (ϱ_{BP}) according to the following equation system:

$$m_{total} = m_{TP} + m_{BP}, \tag{11.5}$$

$$m_{TP} = \varrho_{TP} \cdot V_{TP}, \tag{11.6}$$

$$m_{BP} = \varrho_{BP} \cdot V_{BP}, \tag{11.7}$$

$$V_{total} = V_{TP} + V_{BP}. \tag{11.8}$$

The total mass of the complete ATPS is represented by m_{total} and was calculated by summing up the pipetted masses of each stock solutions within the respective system. V_{total} refers to the total volume of the system and can be calculated by the sum of all pipetted volumes of stock solutions within the respective system. This implicates that volume contraction can be considered negligible. This assumption was validated experimentally by finding a linear correlation between one phase system composition and their densities within the salt and PEG concentration range of interest (data not shown). V_{TP} and V_{BP} refer to the volume of the top and bottom phase. This equation system results into the following expressions to determine the target factors.

$$m_{TP} = \frac{(m_{total} - V_{total} \cdot \varrho_{BP}) \cdot \varrho_{TP}}{\varrho_{TP} - \varrho_{BP}} \tag{11.9}$$

$$m_{BP} = m_{total} - m_{TP} \tag{11.10}$$

The used technical algorithm for the determination of the tie lines bases on an independently rotation of the two arms, which are spanned by the system point and the top and bottom phase respectively, around the system point. Top and bottom phase points are moved independently along the binodal curve until both arms exhibit an equal slope and

fulfill the lever rule (equation 11.4). The method has been described in detail previously [123].

11.3 Results

11.3.1 Liquid Handling

The stock solutions' density as a function of the temperature was determined in order to increase the accuracy of the system preparation procedure with respect to liquid handling used in experimental determination of binodal curve, phase density and tie line. Densities were measured over a temperature range from 8°C to 30°C over at least 4 h which correlates to more than 20000 single measurements. All stock densities were fitted to an Gaussian function as described by equation 11.2, showing an excellent agreement with coefficient of determination (R^2) above 0.99 for all liquids in the respective temperature range. The parameters of all stock solutions are listed in Table 11.1.

Table 11.1: Fitting parameters and coefficient of determination (R^2) of density functions within a temperature interval of 8 to 30°C. Percentage data correspond to weight per weight (w/w).

Stock solution	α $[g/ml]$	β [-]	γ [-]	$\mathbf{R^2}$
Water	1.0001	1.7697	422.95	0.99
KPO4 40%	1.3921	-43.127	564.44	1
PEG300 70%	1.412	-631.01	1330	1
PEG400 70%	3.2542	-2785.8	2708.7	1
PEG600 60%	1.824	-1353.9	1936.7	1
PEG1000 40%	1.1271	-166.66	816.3	1
NaCl 25%	1.4399	-907.74	2117.2	1

11.3.2 Tip2World Interface

A generic *Tip2World* interface was constructed in order to integrate a standalone device into the liquid handling station and thus expand the portfolio of devices suitable for a HTS work flow. The *Tip2World* interface was tested successfully for leak tightness with

aqueous and high viscous solutions in combination with different flow rates and connected capillary tubings. After passing the functionality tests, the MicroLDS was successfully integrated into the LHS by means of the *Tip2World* interface for fully automated sample supply.

11.3.3 Binodal Curves

The binodal curves of different PEG/phosphate ATPS systems with PEG varying in molecular weight (MW) from 300 Da, 400 Da, 600 Da,1000 Da and NaCl concentrations of 0 %(w/w) and 3 %(w/w) were determined. A minimum of 30 points derived from the screening procedure described in section 11.2.3.1 were used for exponential fitting the data by least-square regression according to an empirical curve fit (equation 12.1) as developed by Merchuk et al. [279]. All curves could be fitted with a coefficient of determination (R^2) of at least 0.99. All determined coefficients are listed in Table 11.2. The binodal curves were determined in order to calculate tie lines by means of densities of

Table 11.2: Fitting parameters and coefficient of determination (R^2) of determined binodal curves.

PEG MW [Da]	**NaCl [%(w/w)]**	**a**	**b**	**c**	**R^2**
300	0	88.8	-0.345	$-4.73\cdot 10^{-5}$	0.999
400	0	81.0	-0.354	$-6.79\cdot 10^{-5}$	0.999
600	0	81.0	-0.385	$-1.00\cdot 10^{-4}$	0.999
1000	0	79.1	-0.420	$-1.72\cdot 10^{-4}$	0.998
300	3	62.3	-0.335	$-5.77\cdot 10^{-5}$	0.997
400	3	57.2	-0.338	$-8.22\cdot 10^{-5}$	0.999
600	3	52.9	-0.341	$-1.35\cdot 10^{-4}$	0.999
1000	3	53.0	-0.382	$-2.43\cdot 10^{-4}$	0.998

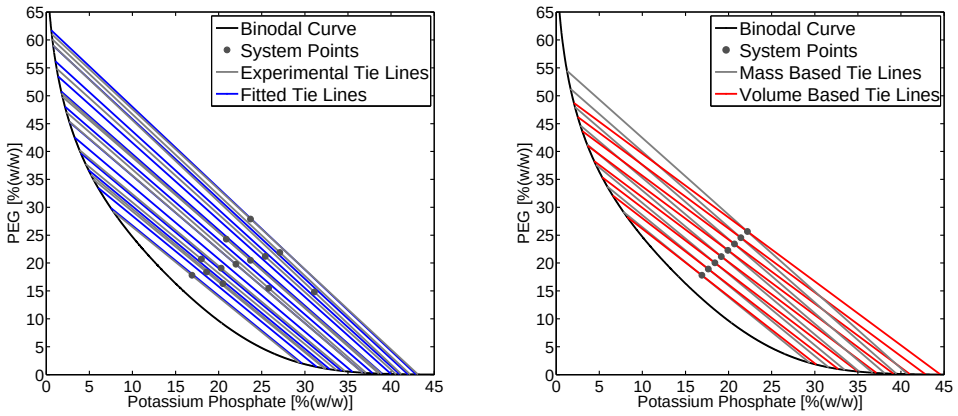
top and bottom phase. The binodal curves are shifted to lower potassium phosphate and PEG concentrations with increasing PEG molecular weight as illustrated in Figure 11.3a. The same effect was observed by adding 3 %(w/w) NaCl as shown in Figure 11.3b.

11.3.4 Phase Densities

Densities were determined by the micro fluidic Coriolis liquid density sensor in a high throughput format. The MicorLDS was established onto a liquid handling station by means of a *Tip2World* interface and thereby transformed into a fully automated HTS device as described in section 11.2.3.2. All systems were prepared six fold and the densities of top and bottom phase were measured with more than 600 single measurements. Densities of top and bottom phase of all system are illustrated color coded in Figure 11.5 and Figure 11.6. In all systems the densities of top and bottom phase increased with increasing tie line length. The differences in densities of top and bottom phase revealed a identical tendency. The addition of polymer has a minor effect on changes in density than the addition of salt. The increase of the differences in phase densities results in an accelerated phase separation by increasing tie line length.

11.3.5 Tie Line Calculation

The densities derived from the MicroLDS were used for the determination of phase masses and phase volumes according to the equation system described by equations 11.9 and 11.10. Tie lines were calculated by applying the lever rule with respect to phase masses as expressed in equation 11.4. The experimentally derived tie lines were used to generate the function F_{slope} in order to identify any tie line within the two phase region limited by the screening grid. The latter was achieved by fitting the slope of the tie lines to a second-degree polynomial surface $F_{slope}(x_{salt}, x_{poly})$. On base of $F_{slope}(x_{salt}, x_{poly})$ tie lines were recalculated in order smooth smaller deviations in experimentally tie line slopes. The measured densities were correlated to the top and bottom phases using a second-degree polynomial surface $F_{density}$. As illustrated in Figure 11.4a by means of PEG400/ phosphate ATPS tie lines could be calculated and fitted with an excellent accuracy. One of the benefits of the presented methodology is the consideration of phase densities. In order to illustrate the impact of neglecting the differences in phase densities, tie lines were evaluated for a linear grid of system points on base on phase masses and phase volumes respectively. The importance of considering differences in phase densities is demonstrated clearly by Figure 11.4b, which illustrates tie lines calculated on base of phase volumes [278] and phase mass [279]. Thereby, the influence of neglecting or considering respectively the differences in phase densities is pointed out. It is obvious that neglecting differences in phase densities by using volume based tie lines [278] results in an



(a) (b)

Figure 11.4: Comparison of (a) experimentally and fitted tie lines and (b) tie lines determined on bases of phase masses [279] and phase volumes [278] respectively by means of PEG400/ potassium phosphate systems (pH 7).

increasing error with respect to tie line slope and phase compositions. The fitted tie lines along with a selection of experimentally derived tie lines of all PEG/phosphate systems investigated in this study are illustrated in Figure 11.5 and Figure 11.6. The densities of top and bottom phase are visualized color coded along the binodal curve. The fitted tie lines were in excellent agreement with the one derived from experiment. All tie lines were almost parallel. The averaged tie line slopes are decreasing with increasing PEG molecular weight, descending from -1.19 ± 0.03 (PEG300) to -1.51 ± 0.06 (PEG1000). The addition of 3 %(w/w) NaCl led to increased tie line slopes for all investigated systems. The averaged tie line slopes within the studied region are listed in Table 11.3.

Table 11.3: Averaged tie line slopes ± standard deviation within investigated region of studied PEG/phosphate systems.

PEG MW [Da]	**300**	**400**	**600**	**1000**
0 %(w/w) NaCl	-1.19 ± 0.03	-1.33 ± 0.05	-1.42 ± 0.03	-1.51 ± 0.06
3 %(w/w) NaCl	-1.15 ± 0.06	-1.20 ± 0.07	-1.26 ± 0.02	-1.37 ± 0.03

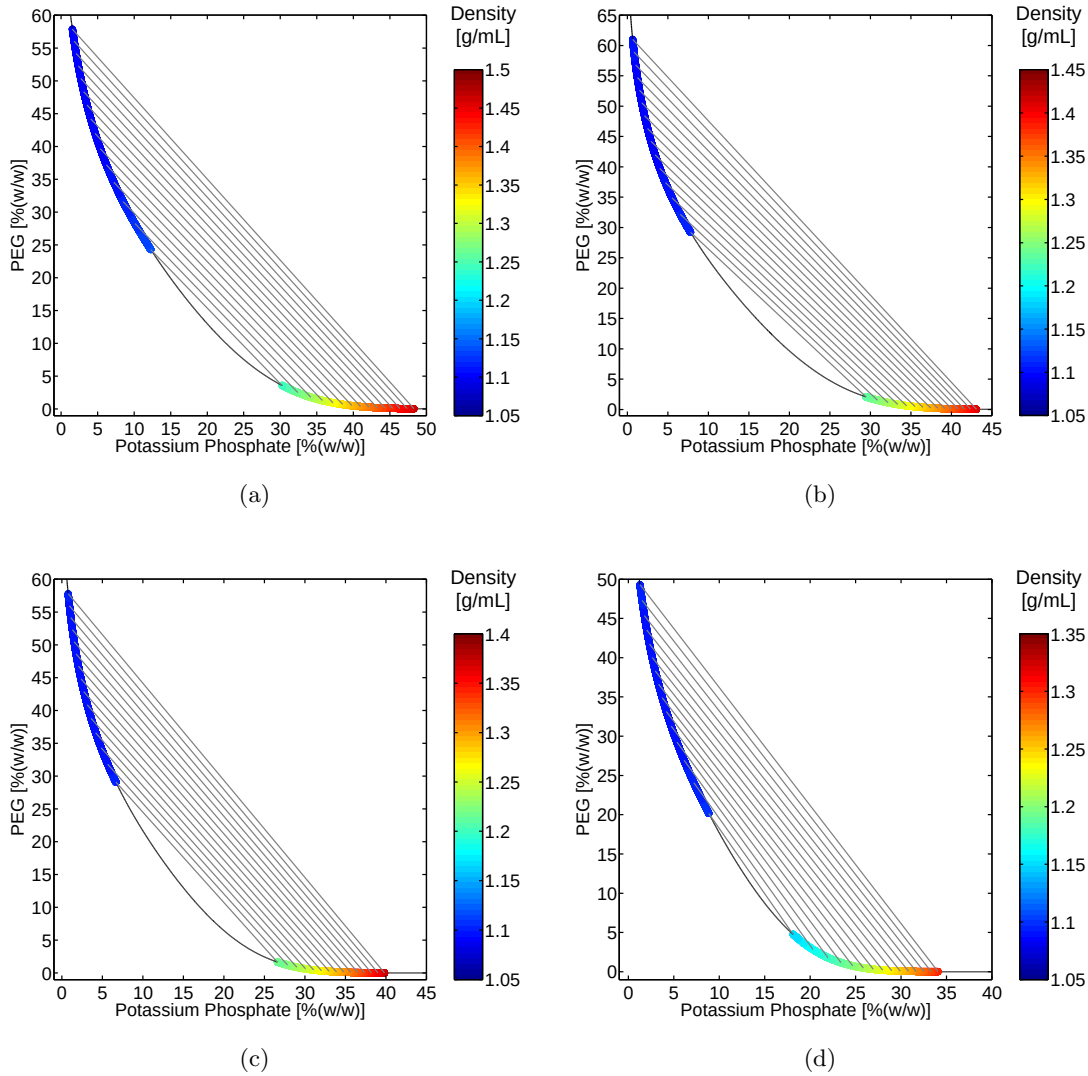


Figure 11.5: Characterization of ATPSs composed of potassium phosphate at pH 7 and (a) PEG300, (b) PEG400, (c) PEG600, and (d) PEG1000.

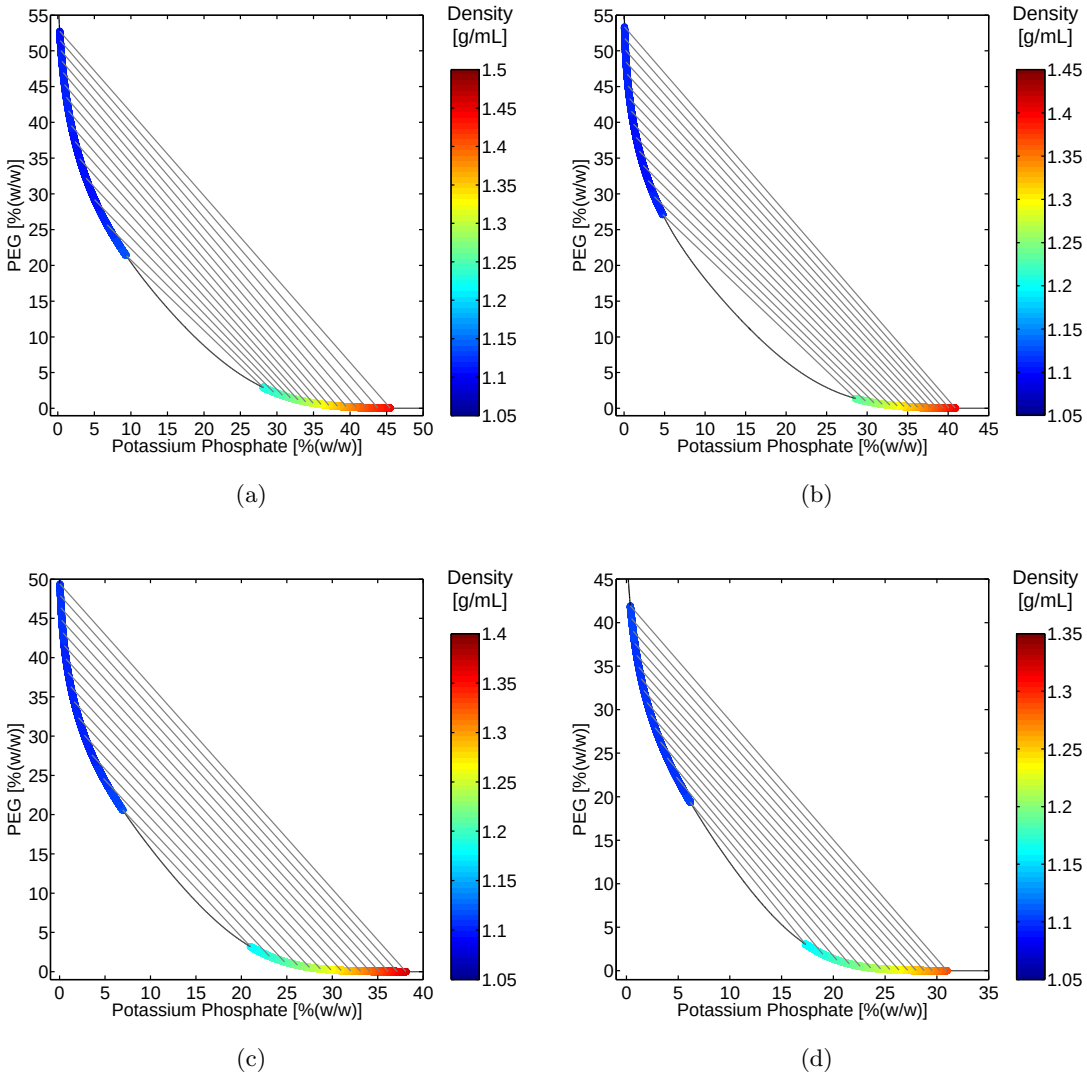


Figure 11.6: Characterization of ATPSs composed of potassium phosphate at pH 7 and (a) PEG300, (b) PEG400, (c) PEG600, and (d) PEG1000 in presence of 3%(w/w) NaCl.

11.4 Discussion

11.4.1 Tip2World Interface

The presented *Tip2World* interface was developed in order to integrate a standalone micro fluidic sensor into a liquid handling station. By means of this device we could combine two of the most promising technologies, namely HTS robotics and lab-on-chip technologies, in order to improve and accelerate process development. In the present work, we focused on ATPS and the respective density determination. However, the *Tip2World* interface can help to transform a number of standalone devices into HTS robotics. It expands the portfolio LHS compatible devices enormously. For instance analytics based on flow through cells such as flow chamber based determination of pH, conductivity, selective ion concentrations, particle analysis or any kind of spectroscopy can potentially be integrated into a LHS and performed in a daisy-chained setup to increase throughput and reduce sample consumption. The feasibility could be limited by highly increased counter pressure possibly resulting in liquid leakage at tubing connection within the LHS. However, this can be counteracted by reducing the flow rate or by increasing tubing inner diameters downstream the *Tip2World* interface. It is important to mention, that the device is not solely suited for sample supply, but even for sample drawing since it can be operated into reversed flow as well. This enables an automated sampling of systems connected to the *Tip2World* interface by capillary tubings. For instance the sample drawing of a connected reactor could be automated or movements of constructions could be controlled hydraulically. *Tip2World* interface was manufactured by a 3D printer to a reasonable price in a polyamide plastic, which is resistant against a range of aqueous solutions but not for organic solvents. However, there are a number of suppliers, which provide 3D printing with highly chemical resistance such as polyether ether ketone (PEEK), which we would recommend in case of using organic solvents.

11.4.2 ATPS Characterization

11.4.2.1 Binodal Curve

Bindodal curves were evaluated for PEG/phosphate ATPS with a sufficient coefficient of determination (R^2) by means of a least 30 points. The binodal curves positionally shifted towards the origin with increasing PEG molecular weight and NaCl content, and

thus less salt and polymer concentrations allude to phase formation. This is in good agreement with other reported systems [123, 279, 289–291]. This may be contributed to the increasing hydrophobicity of PEG with increasing molecular weight and thus the decrease in compatibility between the system components [52].

11.4.2.2 Tie Line Determination

In this work we present a new approach for tie line determination by means of the phase densities, which were derived from a HTS compatible *lab-on-a-chip* based technique. The determined tie lines are quite parallel. Although there is no analytical framework which confirm these results, the data are in good agreement with experimental observations [279, 281, 292–294]. The main advantage of these methodology is the analysis speed and thus the feasibility of HTS work flows in comparison to gravimetric analysis [279], or the use of conductivity (salt) and refractive index (PEG) analysis [272, 280, 281]. In contrast to the much faster dye method developed by Bensch et al. [278], the presented method is non-invasive and has no assumptions regarding a dye distribution or recovery. It excludes inaccuracies of top and bottom phase sampling and system dependent meniscus formation as an error source. This method enables to calculate tie lines by applying the lever arm rule with respect to phase masses and redundantize the simplification of negligible differences of phase densities. As pointed out in 11.3.5 this simplification can alter the accuracy of tie line determination enormously especially for long lie lines. Moreover, additional information and the density in particular is derived from this approach. Knowing the density of the phase liquids could be highly interesting with respect to feasibility and processability issues downstream. Moreover, the method and the presented data might be highly valuable and particularly relevant for further developments of models describing physiochemical properties of ATPS [273]. Thus the presented method is a suitable alternative approach for the characterization of ATPSs in a more accurate way and of great theoretical and practical importance.If needed the phase density determination can further be accelerated by increasing dispensing flow rates, shorten purging volume or simple numbering up of density sensors. Considering the whole ATPS characterization process, two complete ATPSs could be characterized within 24 h, including four runs per ATPS for binodal curve determination (less than 45 min/run), and tie line determination (less than 45 min/run for ATPS preparation and 8 h for density determination), which can be performed fully automated over night without requiring manpower. Therefore the presented approach can still cope with HTS

demands and can be integrated into a HTS work flow. The absorbance method [278] is highly suitable and should be favored for a fast ATPS characterization in scope of HTP work flow. In order to characterize the ATPS in a more refined way, the differences in phase densities have to be taken into account and the presented methodology should be favored. The associated additional time consumption can be compensated by appropriate time management as described above.
The presented methodology provides a cost, time and material effective approach for characterization of ATPS phase diagram based on highly accurate and comprehensive data. By this means the derived data opens the door for a more detailed description of ATPS towards generating mechanistic based models, since molecular approaches such as MD simulations or statistical descriptions along the line of QSAR heavily rely on accurate and comprehensive data. This structure based *in-silico* approaches could be uses as predictive tools for protein partitioning and thus for ATPS process development in very early stage drug development.

11.5 Conclusions

In the present work a generic interface was developed in order to enable the combination of lab-on-chip technologies and robotic liquid handling stations. This *Tip2World* interface was manufactured by high precision 3D printer to a reasonable price. It was validated for a wide range of relevant applications by testing different liquids and capillary tubings. The *Tip2World* interface provides the integration of number of analytic devices and micro fluidic applications into liquid handling station and into the HTS work flow in turn. It expands the portfolio of LHS compatible instruments enormously and can be used to transform a number of standalone analytic devices into fully automated instruments.
In this study the *Tip2World* interface was used to integrate a micro fluidic Coriolis liquid density into a fully automated robotic liquid handling station. This setup facilitates the realization of an innovative approach for the characterization of aqueous two phase systems with respect to tie lines determination by means of phase densities. The new approach can cope with demands regarding HTS work flow compatibility and accuracy.

11.6 Acknowledgment

This research work is part of the project "Molecular Interaction Engineering: From Nature's Toolbox to Hybrid Technical Systems", which is funded by the German Federal Ministry of Education and Research (BMBF), funding code 031A095B. The authors bear the complete responsibility for the content of this publication. We give special thanks to Matthias Franzreb and Jonas Wohlgemuth for performing the 3D printing.

11.7 References

27. Azevedo, A. M., Rosa, P. A., Ferreira, I. F. & Aires-Barros, M. R. Chromatography-free recovery of biopharmaceuticals through aqueous two-phase processing. *Trends Biotechnol.* **27,** 240–247 (2009).

45. Rito-Palomares, M. Practical application of aqueous two-phase partition to process development for the recovery of biological products. *J. Chromatogr. B* **807.** 12th International Conference on Biopartitioning and Purification, 3–11 (2004).

52. Harris, J. M. *Poly(Ethylene Glycol) chemistry: Biotechnical and biomedical applications* (Springer, 1992).

123. Diederich, P., Amrhein, S., Hämmerling, F. & Hubbuch, J. Evaluation of PEG/phosphate aqueous two-phase systems for the purification of the chicken egg white protein avidin by using high-throughput techniques. *Chem. Eng. J.* **104,** 945–956 (2013).

268. Wiendahl, M., Oelmeier, S. A., Dismer, F. & Hubbuch, J. High-throughput screening-based selection and scale-up of aqueous two-phase systems for pDNA purification. *J. Sep. Sci.* **35,** 3197–3207 (2012).

269. Treier, K. *et al.* High-throughput methods for miniaturization and automation of monoclonal antibody purification processes. *Biotechnol. Prog.* **28,** 723–732 (2012).

270. Hansen, S. K., Skibsted, E., Staby, A. & Hubbuch, J. A label-free methodology for selective protein quantification by means of absorption measurements. *Biotechnol. Bioeng.* **108,** 2661–2669 (2011).

271. Diederich, P. *et al.* A sub-two minutes method for monoclonal antibody-aggregate quantification using parallel interlaced size exclusion high performance liquid chromatography. *J. Chromatogr. A* **1218,** 9010–9018 (2011).

272. Hatti-Kaul, R. Aqueous two-phase systems. English. *Mol. Biotechnol.* **19,** 269–277 (2001).

273. Valavi, M., Shirazian, S., Pour, A. & Ziary, M. Calculation of the density and activity of water in ATPS systems for separation of biomolecules. English. *J. Solution Chem.* **42,** 1423–1437 (2013).

274. Valavi, M., Dehghani, M. R. & Feyzi, F. Calculation of liquid-liquid equilibrium in polymer electrolyte solutions using PHSC-electrolyte equation of state. *Fluid Phase Equilib.* **341,** 96–104 (2013).

275. Azevedo, A. M. *et al.* Downstream processing of human antibodies integrating an extraction capture step and cation exchange chromatography. *J. Chromatogr. B* **877,** 50–58 (2009).

276. Ferreira, L. *et al.* Structural features important for differences in protein partitioning in aqueous dextran-polyethylene glycol two-phase systems of different ionic compositions. *Biochim. Biophys. Acta, Proteins Proteomics* **1844,** 694–704 (2014).

277. De Barros, D. P. *et al.* Modeling the partitioning of amino acids in aqueous two phase systems. *J. Chromatogr. A* **1329,** 52–60 (2014).

278. Bensch, M., Selbach, B. & Hubbuch, J. High throughput screening techniques in downstream processing: Preparation, characterization and optimization of aqueous two-phase systems. *Chem. Eng. Sci.* **62,** 2011–2021 (2007).

279. Merchuk, J. C., Andrews, B. A. & Asenjo, J. A. Aqueous two-phase systems for protein separation: Studies on phase inversion. *J. Chromatogr. B: Biomed. Sci. Appl.* **711,** 285–293 (1998).

280. Nascimento, K. S. *et al.* Liquid-liquid equilibrium data for aqueous two-phase systems composed of ethylene oxide propylene oxide copolymers. *J. Chem. Eng. Data* **56,** 190–194 (2011).

281. Martins, J. P. *et al.* Liquid-liquid equilibria of an aqueous two-phase system containing poly(ethylene) glycol 1500 and sulfate salts at different temperatures. *J. Chem. Eng. Data* **53,** 238–241 (2008).

282. Salieb-Beugelaar, G. B. *et al.* Latest developments in microfluidic cell biology and analysis systems. *Anal. Chem.* **82,** 4848–4864 (2010).

283. Arora, A. *et al.* Latest developments in micro total analysis systems. *Anal. Chem.* **82.** PMID: 20462185, 4830–4847 (2010).

284. Weigl, B. H., Bardell, R. L. & Cabrera, C. R. Lab-on-a-chip for drug development. *Adv. Drug Delivery Rev.* **55.** Biomedical Micro- and Nano-technology, 349–377 (2003).

285. Neuži, P. *et al.* Revisiting lab-on-a-chip technology for drug discovery. *Nat. Rev. Drug Discov.* **11,** 620–632 (2012).

286. Waldbaur, A. *et al.* Microfluidics on liquid handling stations (µF-on-LHS): An industry compatible chip interface between microfluidics and automated liquid handling stations. *Lab. Chip* **13,** 2337–2343 (12 2013).

287. Sparks, D. *et al.* Measurement of density and chemical concentration using a microfluidic chip. *Lab. Chip* **3,** 19–21 (1 2003).

288. Sparks, D., Smith, R., Massoud-Ansari, S. & Najafi, N. *Coriolis mass flow, density and temperature sensing with a single vacuum sealed MEMS chip* in *Solid-State Sensor, Actuator and Microsystems Workshop, Hilton Head Island, South Carolina* (2004), 75–78.

289. Azevedo, A. M. *et al.* Partitioning of human antibodies in polyethylene glycol-sodium citrate aqueous two-phase systems. *Sep. Purif. Technol.* **65,** 14–21 (2009).

290. Foroutan, M., Heidari, N., Mohammadlou, M. & Sojahrood, A. J. (Surfactant + polymer) interaction parameter studied by (liquid + liquid) equilibrium data of quaternary aqueous solution containing surfactant, polymer, and salt. *J. Chem. Thermodyn. Thermochem.* **41,** 227–231 (2009).

291. Tubío, G., Pellegrini, L., Nerli, B. B. & Picó, G. A. Liquid-liquid equilibria of aqueous two-phase systems containing poly(ethylene glycols) of different molecular weight and sodium citrate. *J. Chem. Eng. Data* **51,** 209–212 (2006).

292. Raja, S. *et al.* Aqueous two phase systems for the recovery of biomolecules–a review. *Sci. Technol.* **1,** 7–16 (2011).

293. Simon, L. & Gautam, S. Modeling continuous aqueous two-phase systems for control purposes. *J. Chromatogr. A* **1043,** 135–147 (2004).

294. Madeira, P. P. *et al.* Correlations between distribution coefficients of various biomolecules in different polymer/polymer aqueous two-phase systems. *Fluid Phase Equilib.* **267,** 150–157 (2008).

CHAPTER 12

Molecular Dynamics Simulations on Protein Partitioning in Aqueous Two Phase Systems

Sven Amrhein and Jürgen Hubbuch

Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Science, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany

in preparation

Abstract

There is unbroken interest in alternative bioseparation operations for chromatographic separation in biopharmaceutical industry which is further boosted by the strive for continuous manufacturing. Aqueous two phase systems have proofed as highly potent separation systems, which can be performed continuously with reasonable instrumental complexity. However, since a mechanistic understanding of protein partitioning in aqueous two phase systems is still missing, the process development is still approached empirically and the systems are primary used in academia and less established in biopharmaceutical industry. In order to understand underlying mechanisms of the experimental observable partitioning phenomena, the associated free energy changes have to be investigated.
In this study extensive full-atom molecular dynamics simulations were performed in order to derive partitioning free energies of selected aqueous two phase systems composed of potassium phosphate and poylethylene glycol with molecular weights of 300, 400, 600, and 1000 Da. These systems were thoroughly characterized in terms of composition and density of the coexisting phases and were assembled by advanced packing procedures to simulate the respective solvent as good as possible. A selection of amino acids, peptides, and proteins were simulated in the respective coexisitng phases separately. The resulting trajectories were analyzed for polar (electrostatic) and non-polar (van der Waals) interactions in order to approximate the partitioning free energy by the linear interaction energy approach. The derived data gives detailed insight into the protein-solvent interaction. These data might serve for building a predictive model for partitioning within the set of investigated systems but requires experimental data for a proper parametrization of the linear interaction energy model.

Keywords: *ATPS; Free Energy Calculation; In-Silico; Linear Interaction Energy; Protein-Solvent Interactions*

12.1 Introduction

Aqueous two phase systems (ATPSs) have been demonstrated as highly promising separation technique in biotechnology offering scalability, cost-efficiency and providing mild conditions that do not harm or denature biopharmaceutical products [41, 295]. It has been shown that using ATPS can be an alternative to established purification techniques for instance for the cost intensive purification of monoclonal antibodies [296, 297].
However, after decades of research there is no adequate understanding of the partitioning of proteins within ATPSs. Since a mechanistic model is still missing, the development of purification operations based on ATPSs has to be performed experimentally with a high effort in terms of material, time, and costs. The experimental effort can be reduced by using parallelized, miniaturized, and automated high throughput screening methods. High throughput methods for ATPS screening have already been developed and established within our working group [123, 268, 298]. However, a basic mechanistic understanding for the complex partitioning phenomena requires investigations on a more detailed level of underlying protein-solvent interactions.
Molecular dynamics (MD) simulations are able to gain a profound understanding of protein partitioning mechanism and the thermodynamic and kinetic properties of partitioning solutes. MD simulations are highly appreciated tools in drug discovery and lead optimization which can be applied to understand experimental observations on an atomic level of detail. Gaining a profound understanding of experimental observables requires the evaluation of underlying associated free energy changes such as solvation [70, 71], protein-ligand binding [72–74], proton and electron transfer reactions [75, 76], transport and partitioning phenomena [79], or protein/peptide folding and stability [80, 81]. There is a number of methods to evaluate the free energy by molecular dynamics simulations. One of the most accurate ones is the thermodynamic integration (TI) and the related free energy perturbation (FEP) which are widely-used for characterization systems of small structures [299–301]. However, these methods are highly elaborate with respect to CPU-time and therefore not practical for the valuation of large systems or the aqueous two phase systems respectively, which are object of the present work. The method with the lowest effort with respect to CPU-time is probably a quantitative structure-activity relationship (QSAR) or quantitative structure-property relationship (QSPR) approach. The predictive quality highly depends on the developed model which generates a response in dependency of different descriptors. Moreover, the molecular dynamics of phase components and solutes should be taken into account. Considering

the inverse correlation of accuracy and throughput, the approximate method, namely the linear interaction energy (LIE) method, was applied to compute free energy differences within the complex ATPSs. The LIE method enables a sampling over simulation time, and thereby an estimation of error, and simultaneously is practical with respect to CPU-time. This method was pioneered by Åqvist et al. [302, 303] and was successfully applied for the determination of free energy differences for characterization binding affinities [304] and partitioning of small molecules between water and chloroform [94].
In the present work the LIE method was used to determine partitioning free energies of solvent interactions of amino acids, peptides, and proteins in ATPSs composed of phosphate and polyethylene glycol (PEG) of different molecular masses ranging from 300 to 1000 Da in order to develop a predictive model of its affinity based partitioning and thus to deepen the knowledge of the underlying partitioning mechanisms. The study thereby focuses on the true partitioning of diluted solute concentrations and thus a the partitioning which is dominated by the solute specific affinities to the coexisting phases. At high solute concentrations the partitioning is dominated by differences in solubility limitations in the respective phases and thus the partitioning coefficient does not reflect the solute specific phase affinities. This study will be experimentally substantiated by partitioning coefficients determined by means of experimental high throughput screenings.

12.2 Materials and Methods

12.2.1 Experimental Characterization of Aqueous Two Phases Systems With Respect to Binodal Curve and Tie Lines

The determination of binodal curve and tie lines of the respective systems has been described in detail previously [159] and is shorty described in the following. The system characterization was performed by means of robotic liquid handling stations. In order to prepare highly defined mixtures of relevant stock solutions, a pipetting optimization was performed for each stock solution by adjusting all relevant parameters. The accuracy was validated gravimetrically by means of an analytical balance (WXTS205DU, Mettler-Toledo, Greifensee, Switzerland). The densities as a function of the temperature of respective solutions were derived from a micro liquid density sensor, purchased from Issys (Ypsilanti, MI, USA).

12.2.1.1 Binodal Determination

The binodal curves were determined by turbidity measurements in 96 well micro titer plates. The method was performed using the liquid handling station Freedom EVO® 200, purchased from Tecan (Crailsheim, Germany), equipped with Teflon coated fixed tips and 1.0 mL dilutors. In particular, ultra pure water, PEG and phosphate stock solutions were mixed to a volume of 300 μL in order to reach concentrations in a certain grid exhibiting the binodal curve. After 60 s rigid mixing on an orbital shaker with a frequency of 1800 rpm two phase systems were visually detected. The binodal was fitted to an exponential function as stated in equation 12.1 according to Merchuk et al. [279]

$$x_{PEG} = a \cdot exp(b \cdot x_{salt}^{0.5} + c \cdot x_{salt}^{3}), \tag{12.1}$$

where x_{PEG} and x_{salt} represent the content of polymer and salt in mass percent. The coefficients a, b, and c are empirical fitting factors.

12.2.1.2 Tie Line Determination

The densities of top and bottom phase of a representative selection of aqueous two phase systems were analyzed in order to calculate the respective tie line and the phase densities along the binodal curve. These phase densities enable to calculate absolute numbers of molecules for MD simulations. Based on the phase densities the tie lines were evaluated by applying the lever arm rule on the mass ratio of top and bottom phase. The densities were measured in a high throughput way by combining a micro fluidic Coriolis sensor in combination with a liquid handler, namely Freedom EVO® 100 (Tecan), and an interface called *Tip2World* interface. The technical implementation and the algorithm for tie line evaluation is described in detail elsewhere [123, 159].

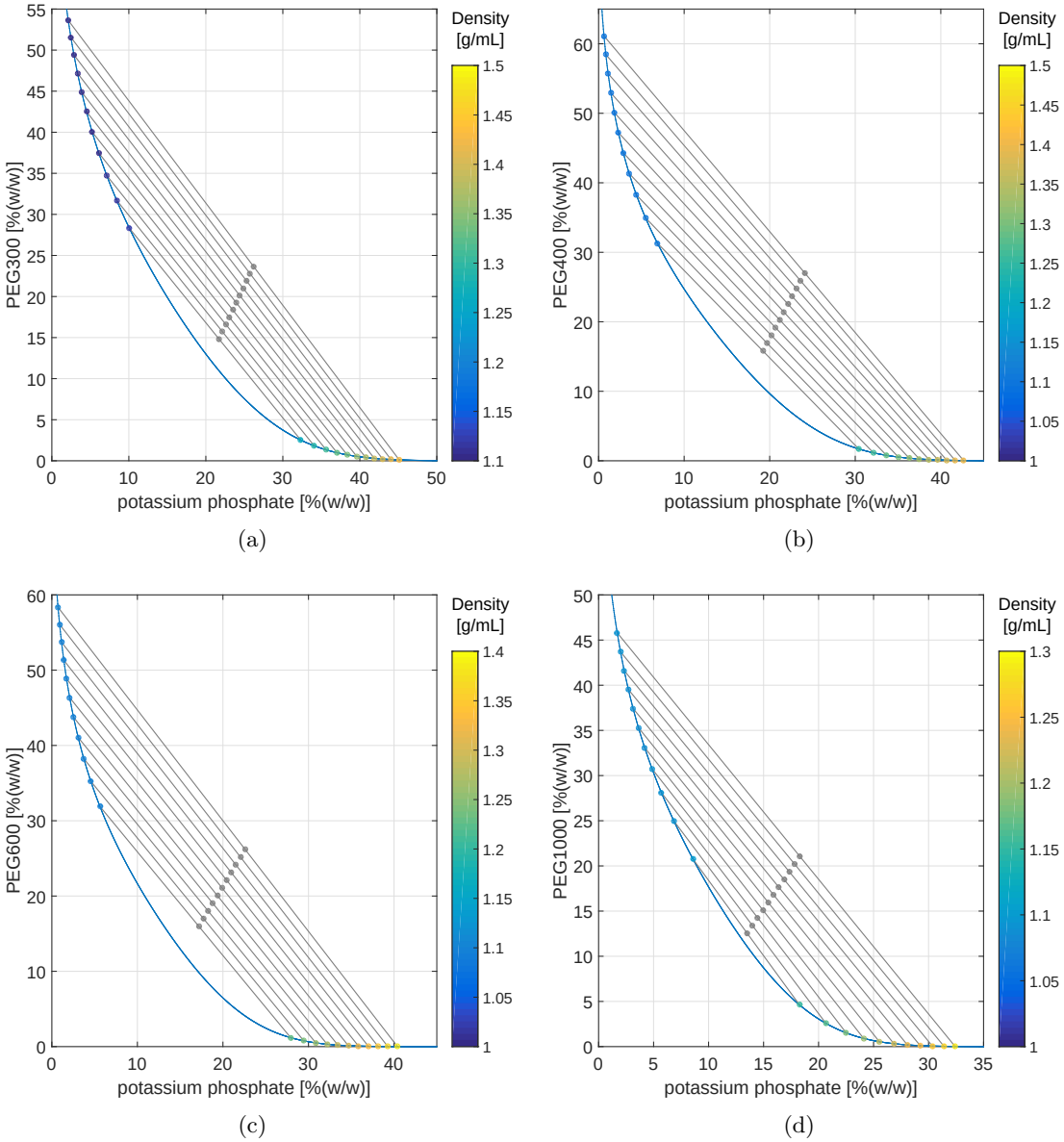


Figure 12.1: Selection of systems within the two phase region of aqueous two phase systems composed of potassium phosphate at pH 7 and (a) PEG300, (b) PEG400, (c) PEG600, and (d) PEG1000, which were investigated by molecular dynamics simulations.

12.2.2 Molecular Dynamics Simulations

The experimentally derived properties of the aqueous two phase systems were used to setup molecular dynamics simulations of selected systems, which are illustrated in Figure 12.1. The explicit MD simulations were performed on a high performance cluster computer provided by the Steinbuch Centre for Computing (SCC) at the Karlsruhe Institute of Technology (KIT). For MD simulations and evaluations the commercial software package YASARA Structure (version 14.6.23) [209] and Matlab® R2015a (The MathWorks, Natick, ME, USA) were used. The goal of the MD simulations was the determination of the free energy differences of the solute (amino acid, peptide, or protein) in top and bottom phase. These energy differences can potentially be used for building predicive models for solute partitioning. Before generating the initial scenes of top and bottom phase for the final MD simulations the solutes, phase components, and salt and PEG in particular, were prepared. These preparations were performed in order to get the most realistic model of all relevant components with respect to charges and structure. The preparation is described in the following.

12.2.2.1 Protein Preparation

The solutes subjects of this work were selected according to a number of criterions mentioned in the following. The solutes are supposed to be stable in the used aqueous two phase systems, as major conformational transformations like unfolding and denaturation processes are supposed to require much more simulation time and are not describable with this molecular dynamics simulation approach. The structural information of the concerning proteins have to be satisfying with respect to resolution and completeness of the sequence. Thereby homology modeling should be avoided as a source of error. The structures of peptides were sketched by hand. The solutes' protonation at pH 7 was calculated using the *H++* algorithm [230]. All used solutes are listed in Table 12.1. All solutes were used in their monomeric form.

12.2.2.2 PEG and Salt Preparation

The initial structures of PEG and salts were sketched by hand. These structures were simulated lonely in pure water for equilibration as described in detail elsewhere [307]. The PEG-molecules were parametrized using the AutoSMILES algorithm [214] implemented

Table 12.1: List of studied solutes.

	Solute	Abbreviation	PDB ID
Proteins	α-Lactalbumin	α-lact	1F6R [16]
	Avidin	Avi	3VGW [305]
	Duck ovotransferrin	Ovo	1DOT [306]
	Human insulin	Ins_{hum}	3W7Y
	Lysozyme from chicken egg white	Lys_{egg}	2VB1 [228]
Peptides	AlaPhe	Pep_{AlaPhe}	-
	GluValPhe	$Pep_{GluValPhe}$	-
	GlyPhe	Pep_{GlyPhe}	-
	PheGly	Pep_{PheGly}	-
	PheGlyGly	$Pep_{PheGlyGly}$	-
	TyrAla	Pep_{TyrAla}	-
	TyrPhe	Pep_{TyrPhe}	-
	TyrTyrTyr	$Pep_{TyrTyrTyr}$	-
Amino acids	Phe	Phe	-
	Trp	Trp	-
	Tyr	Tyr	-

in the MD simulation software, which has been shown to generate adequate force field parameters [307]. The basic and acid forms of phosphate ($H_2PO_4^-$ and HPO_4^{2-}) were parametrized accordingly.

12.2.2.3 Final Molecular Dynamics Simulations

The resulting energy minimized structures were used to generate the initial scene using the software *packmol* [262, 263] which performed a packing optimization for the complex phases of the respective ATPSs. To minimize the influence of the initial packaging, all systems were simulated in quintets with randomized initial configurations. The phases were packed into cubic boxes, which had dimensions with a factor 1.5 larger than the finally desired box dimensions. The boxes were subsequently compressed to reach the final dimensions and thus solvent density. After each compression step the systems were energy minimized with periodic boundaries. During the compression procedure the structure of the solute was constrained in position to avoid conformational changes due to the compression. The last compression to the desired box size was followed by a removal of all constraints. This procedure turned out to be beneficial for the packing of large structures and the PEG molecules in particular in terms of computational effort and

randomized distribution.
The top and bottom phase were simulated separately with identical box dimensions. The dimensions of the cubic simulations cell were defined by the largest dimension of the protein in addition to an extension of 15 Å at each sides. In case of peptides and amino acids, an extension of 30 Å was chosen to assure an appropriate packing. The exact numbers of all relevant molecules were calculated from the experimentally determined phase compositions in combination with the measured density as described in section 12.2.1.
The ratio of the basic and acid phosphate component was adopted from the experimental procedure of preparing the stock solution of potassium phosphate (pH 7). The alternative of using Henderson-Hasselbalch equation was not selected due to high concentrations of the stock solutions of 40 % (w/w). For all $H_2PO_4^-$ and HPO_4^{2-} molecules the corresponding number of potassium ions was added. For electronically neutralization sodium or chloride ions respectively were added to neutralize the solute charge. In order to remove bumps and correct the covalent geometry, every simulation was initially energy-minimized with the AMBER03 force field [212]. After the removal of conformational stress by a short steepest descent minimization, the procedure continued with simulated annealing until a convergence criteria was met with a time step of 2 fs and scaled down atom velocities by 0.9 every 10th step. Convergence was reached when the energy improved by less than 0.05 kJ/mol per atom during 200 steps. The temperature was kept constant at 298 K by rescaling the atom velocities using a Berendsen thermostat [211]. Simulations were performed with constant atom number, constant volume and constant temperature. The AMBER03 force field [212] was applied to describe the interactions of the proteins. The particle mesh Edwald (PME) algorithm treatment [213] was used for long-range electrostatics with a 7.86 Å cutoff for non-bonded interactions. This cutoff was selected due to concerns regarding the performance/accuracy trade-off, as the software YASARA is optimized towards a cutoff that is a multiple of 2.62 Å. The MD simulations were performed for a time of 5 ns and the trajectories were saved every 25 ps.

12.2.2.4 Free Energy Calculation

The free energy ΔG_{TB} of a compound i of each ATPS was calculated using the semi-empirical linear interaction energy (LIE) approach developed by Åqvist et al. [302, 303] as

$$\Delta G_{TB} = \alpha\Delta \left\langle V_{T-B}^{vdW} \right\rangle + \beta\Delta \left\langle V_{T-B}^{elec} \right\rangle + \gamma, \tag{12.2}$$

where $\langle\rangle$ denotes MD averages of the non-bonded van der Waals (vdW) and electrostatic (elec) interactions between target protein and the surrounding environment. The Δ denotes the difference between these averages in top and bottom phase. The coefficients α and β are used to weight the contribution of the van der Waals and electrostatic term and γ is a constant.

The LIE approach was selected over statistical mechanical approaches such as the thermodynamic integration (TI), the related free energy perturbation (FEP) method, and methods based on probability distribution functions and sampling probabilities due to the computational demand of these methods. The LIE method only considers endpoints and does not rely on any intermediate states such as in the TI or FEP formalism [59]. It was originally developed for the calculation of binding energies between ligand and protein by comparing the polar and non-polar energies in a free and a bound state. However, this scenario can be viewed as a partitioning problem of a ligand between two states and was already applied for partitioning of small structures between water and chloroform and for calculation hydration energies [94, 95].

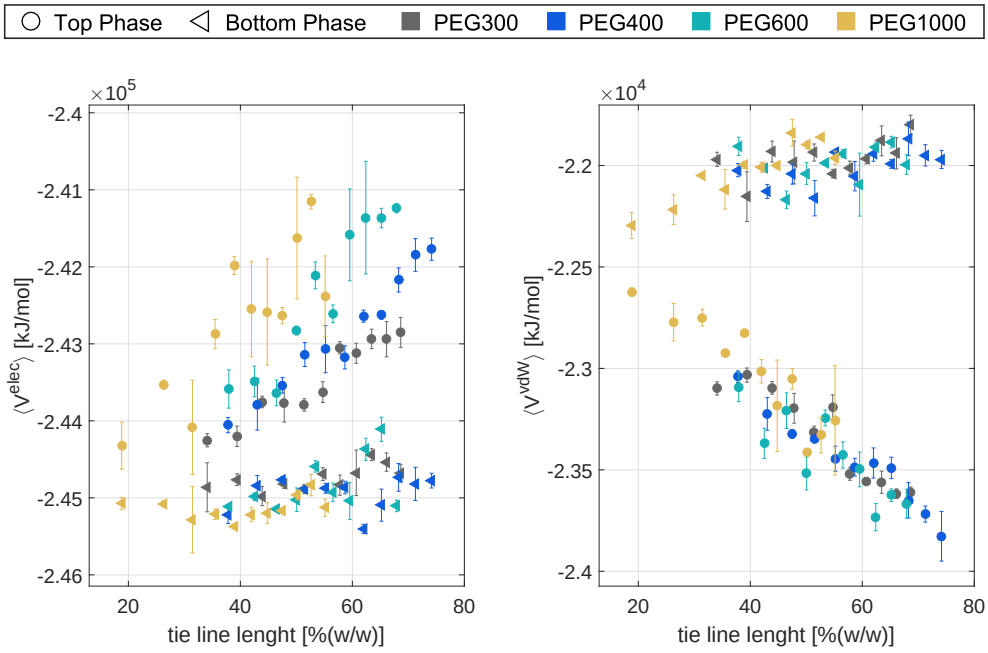
The equilibrium constant k_i for the partitioning of a given compound i between top phase T and bottom phase B is related to the corresponding free energy change of transfer $\Delta G_{i,TP}$ of compound i from bottom phase B to top phase T by equation 12.3

$$k_i = \frac{c_{i,T}}{c_{i,B}} = e^{-\Delta G_{i,TB}/RT}, \tag{12.3}$$

where R is the gas constant and T is the absolute temperature. The van der Waals and electrostatic energies were calculated over a time range starting from 10 ns into 200 ns every 25 ps. It is important to note, that the LIE approach requires a proper parametrization of the scaling factors α and β and the constant γ. These calibrating can be done by means of experimentally derived partitioning coefficients of model solutes such as peptides or amino acids.

12.3 Results

Extensive full atom MD simulations were performed with a selection of amino acids, peptides, and proteins in respective coexisting phases in quintets each. Electrostatic (V^{elec}) and van der Waals (V^{vdW}) interaction energies were evaluated and averaged over time. In scope of this manuscript the derived interaction energies of proteins only are presented and discussed due to reasons of clearness. The interaction energies of studied peptides and amino acids will be required for proper parametrization as discussed later. The derived data exposed very similar trends but differ by magnitudes according to protein size. The interaction energies are illustrated exemplarily for duck ovotransferrin which is representative for the evaluated solutes as a function of tie line length for the variety of PEG molecular weights in Figure 12.2.



(a) (b)

Figure 12.2: Interaction energies of duck ovotransferrin as a function of tie line length for a variety of systems composed of PEG of different molecular weight and potassium phosphate at pH 7 and (a) polar electrostatic interaction energy and (b) non-polar van der Waals interaction energy in particular.

The polar electrostatic interaction energies (Figure 12.2a) were more negative in the salt rich bottom phase compared to the polymer rich top phase. In addition, the electro-

static interaction energies within the bottom phase are less affected by tie line length compared to the top phase even though the variations in salt concentrations within each ATPS are quite similar for both phases according to the binodal curves. In contrast to the electrostatic interaction energy, the non-polar interaction energies (Figure 12.2a) are more negative within the top phase. The non-polar interaction energies expose a negative correlation to the tie line length within the top phase and a positive one within the bottom phase.

The variance illustrated as the standard deviation of the five fold determination is acceptable, whereas the PEG1000 systems expose the most pronounced variance of interaction energies. Based on the separate phase simulations the differences in polar (electrostatic) interaction energies ($V_T^{elec} - V_B^{elec}$) and non-polar (van der Waals) interaction energies ($V_T^{vdW} - V_B^{vdW}$) were calculated and exemplarily illustrated for duck ovotransferrin in Figure 12.3.

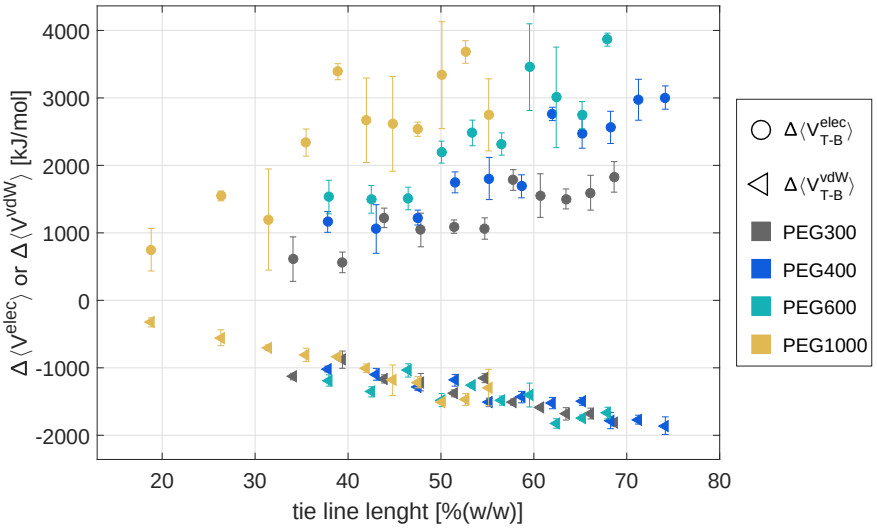
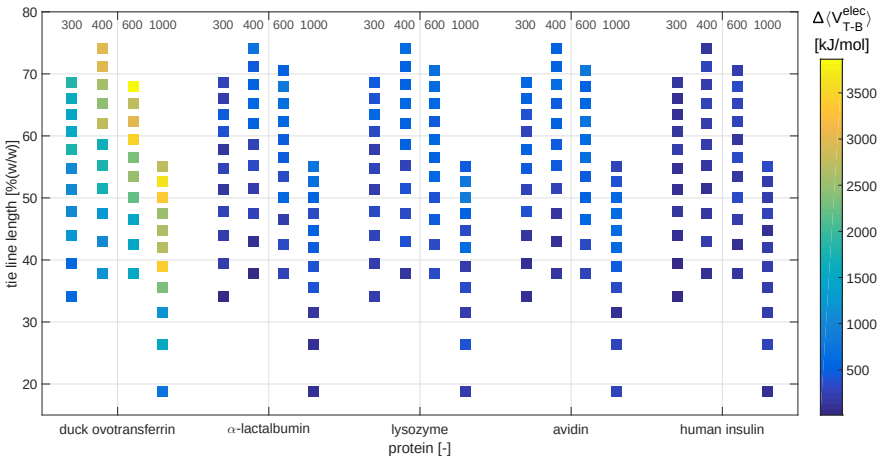
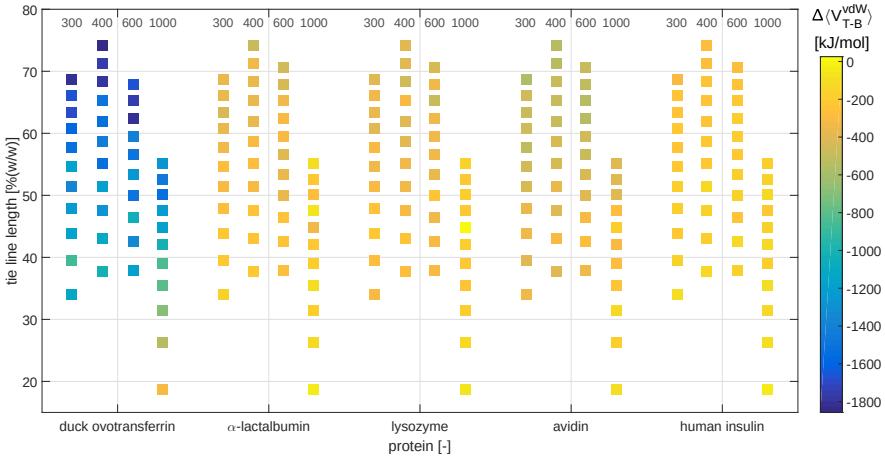


Figure 12.3: Differences in interaction energies of duck ovotransferrin as a function of tie line length within selected aqueous two phase systems composed of potassium phosphate at pH 7 and PEG300, PEG400, PEG600, and PEG1000.

The differences in electrostatic interaction energies ($V_T^{elec} - V_B^{elec}$) followed a linear trend in dependency of tie line length for each polymer molecular weight, whereas differences in van der Waals interaction energies ($V_T^{vdW} - V_B^{vdW}$) seem to correlate linearly with tie line length throughout the polymer molecular weight. The differences in electrostatic and van der Waals interaction energies of all investigated proteins are visualized color coded in Figure 12.4.



(a)

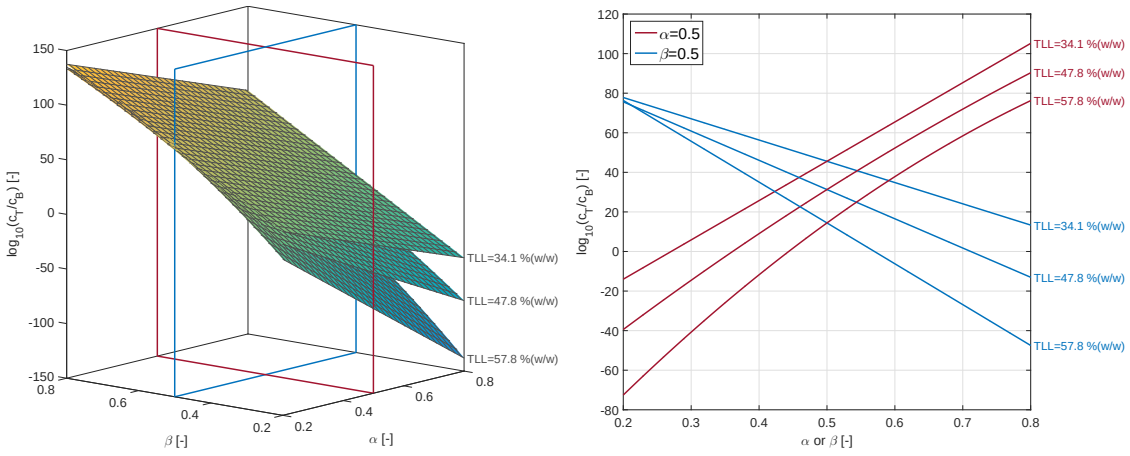


(b)

Figure 12.4: Differences of (a) polar electrostatic interaction energies (V_{elec}) and (b) non-polar van der Waals interaction energies (V_{vdW}) of investigated proteins. The numbers on the top refer to the molecular weight of the respective PEG.

The differences in polar interaction energies are positive and increased with increasing tie line length irrespective of protein species or polymer molecular weight, whereas the increase is most pronounced for duck ovotransferrin. Contrary to the polar interaction energies, differences in non-polar interaction energies are negative and decreased with increasing tie line length irrespective of protein species or polymer molecular weight, whereas the decrease is most pronounced for duck ovotransferrin again. This means that

solely according to the polar interaction energies, the protein affinity for the bottom phase is more pronounced than to the top phase. Contrary to the polar interaction energies, the affinity towards the top phase is more pronounced in terms of non-polar interaction energies. Thus, the polar and non-polar contribution factors α and β have to be properly adapted in order to apply the LIE method according to equation 12.2 and 12.3 for calculating the theoretical partitioning coefficient. Since a proper parametrization by means of experimental data is in progress, theoretical partitioning coefficients are presented in dependency of a variation of α, β, and γ values. The dependencies are illustrated by means of duck ovotransferrin in PEG300/potassium phosphate systems. The constant γ is set to zero for this purpose. The resulting response surface for theoretical partitioning coefficient in dependency of the scaling factors α and β are illustrated in Figure 12.5.



(a) (b)

Figure 12.5: Illustration of the influence of the scaling factors α and β onto the theoretical partitioning coefficient by means of a set of three systems identified by tie line length (TLL) and composed of PEG300 and potassium phosphate for the theoretical partitioning of duck ovotransferrin as (a) response surface and (b) linear correlation for fixed α (α=0.5) or β (β=0.5) respectively.

The response surfaces in Figure 12.5a as well as the exemplary theoretical partitioning coefficients for selected combinations of α and β in Figure 12.5b highlight the need of a proper parametrization of the constants α and β. The selected parameter space used in Figure 12.5b is indicated by planes in Figure 12.5a. Depending on these parameters the affinity to the top or the bottom phase varies by magnitudes and orientation. Thus, it is crucial to generate an appropriate experimental data set for training the model. Further

Table 12.2: Molecular weight (MW) and solvent accessible surface (SAS) of investigated proteins according to AMBER03 parameters and a probe radius of 1.4 Å.

Protein	**MW [g/mol]**	**SAS [Å^2]**
α-lact	14174	6343
Avi	13703	7358
Ovo	75600	27843
$\mathrm{Ins_{hum}}$	5806	3570
$\mathrm{Lys_{egg}}$	14313	6256

conceivable modifications of the LIE method in terms of weighting parameters as well as the experimental design for gaining the crucial experimental data is discussed in section 12.4.

12.4 Discussion

12.4.1 Molecular Dynamics Simulations

Extensive full-atom molecular dynamics simulations of amino acids, peptides, and proteins were performed in corresponding coexisting phases of a set of 44 aqueous two phase systems. For the purpose of clarity, amino acids, peptides, and proteins are called solutes in the following. The five fold assemblies and packing of the initial spatial configurations for the MD simulations as well as the long simulation period of 200 ns are considered to enable a sufficient sampling. This is reflected by the reasonable variance of the determined energies. The fact that the PEG1000 systems expose the most pronounced variance in interaction energies reflects the influence of the initial packing. A randomized and evenly distribution becomes more challenging with increasing polymer size and decreasing polymer molecule numbers. The magnitude of the interaction energies of the proteins in top and bottom phase vary strongly throughout the set of proteins due to the variance in protein size and surface as listed in Table 12.2. The electrostatic interactions were found to be more negative in the salt rich bottom phase, which is due to the increased number of charged molecules and salt ions in particular within the system, which lead to more pronounced electrostatic interactions onto the solute. Furthermore, the electrostatic interaction energies in the top phase seem to correlate to the molecular weight of the used polyethylene glycol. This findings are in agreement with the correlation between the PEG molecular weight and its hydrophobicity. PEG is known to

expose increasing hydrophobicity with increasing molecular weight [52]. The polarity of the polymer chain, which is majorly contributed by the polar terminal hydroxy groups, decreases with the number of repetitive sub-units. Thus, the polarity of PEG decreases with increasing PEG molecular weight. This effect cannot be identified clearly within the salt rich bottom phase due to the very low number of polymer molecules. Contrary to the polar interaction energies, the non-polar interaction energy is more negative within the polymer rich top phase. This finding is reasonable considering the physicochemical properties of the phases as well as the underlying mathematical framework of the molecular dynamics simulations. Since the presence of polyethylene glycol increased the hydrophobicity of the top phase compared to the salt rich polar bottom phase, the non-polar contribution is supposed to be more pronounced within the top phase. From a molecular dynamics' perspective, it is important to note, that the total van der Waals interactions are highly depending on the packing density of the solvent. The van der Waals interactions for a given solvent is measured by the number of Lennard-Jones interaction centers surrounding the solute. Since these interactions are attractive for all but very short distances and are very similar for all atoms, the non-polar interaction energy provides a measure of the heavy-atom number density in the surrounding. This affect can be observed in classical binding energy determinations by means of the LIE approach when the non-polar interaction energies tend to be more negative when the ligand is partly surrounded by the receptor protein than when it is free in water [308, 309]. Thus, it is reasonable that the van der Waals interaction energies are more negative within the polymer rich top phase since the coexisting phases expose significantly different heavy-atom number density in the surrounding and the determined van der Waals interaction energies reflect the local packing number density of the solvent and the van der Waals radii of the solvent components. It is important to note, that the number density is not reflected by the volumetric mass density due to the significant variation in molecular weights of solvent compounds presented in top and bottom phase.

12.4.2 Design and Conception of Further Computational and Experimental Studies on Partitioning

Further refinements of the original LIE model are possibly resulting in a more general form of equation 12.2 that also relaxes the constraint of equal scaling factors for top and bottom phase [308] according to equation 12.4.

$$\Delta G_{TB} = \alpha_A \left\langle V_T^{vdW} \right\rangle - \alpha_B \left\langle V_B^{vdW} \right\rangle + \beta_T \left\langle V_T^{elec} \right\rangle - \beta_B \left\langle V_B^{elec} \right\rangle + \gamma. \tag{12.4}$$

This refinement would allow to consider possible differences in electrostatic response properties of the solute and an improved modeling of the phases. The use of non-equal scaling factors for coexisting phases may require to subtract the intramolecular interaction energies of the solute. In theory the partitioning free energy is affected not only by interactions between solvent and protein, but is affected by the creation of curvatures of the solute in both phases to a certain extent. The energy for creating these curvatures in a solvent depends on the solvent surface tension, which can be considered by phase specific scaling factors. In addition, the constant term γ can further be modified as semiempirical intercept in order to consider entropic contributions to the free energy of partitioning. Since this study covers a large range of molecular weights of the solute ranging from single amino acids to complex proteins, an additional scaling factor will potentially be required, which depends on spatial solute properties such as surface area and volume.

In order to evaluate the options of the mathematical framework for modeling the partitioning based on derived energies, the evaluation of experimental partitioning coefficients is required. Due to the complexity of ATPSs and limitations in analytic performance a number of considerations have to be followed in experiments. The MD simulations based approach presented in this study considers ideally diluted solutes in ATPSs, which implies that interactions between the solute itself such as protein-protein interactions affecting the partitioning cannot be considered by the MD simulations based approach. The adaption of simulation cell dimensions in combination with periodic boundaries cannot overcome this shortcoming since the periodicity highly constraints the orientation of the solute to its copied neighboring ones. Thus, very low solute concentrations have to be used in experiments. The use of low concentrations is further beneficial for the precipitation issue. High salt and polymer contents in the coexisting phases are known to induce precipitation and thus the partitioning coefficient does not represent the phase affinities properly. According to precipitation models such as the Cohn [310] model stated in equation 12.5

$$log(S) = S_0 - m^* \cdot I, \tag{12.5}$$

where S is the solubility of the protein, S_0 is the idealized solubility, m^* is a salt-specific constant and I is the ionic strength of the solution attributed to the added salt, the salt tolerance of the solutes towards precipitation should be increased by decreased solute concentrations. However, the use of low solute concentrations challenges the analytics needed for quantification. It should be avoided, that the detection limit affects the determined partitioning coefficient. Especially in scope of high throughput screenings, fast quantification can be achieved by UV/Vis absorbance measurements. However, absorbance values determined by standard spectrometers bear the risk to be affected by the dependency of extinction coefficient of the respective analyte to solvent components. A dilution of the samples further emphasizes the affect of detection limit and implies further pipetting inaccuracies since the increased viscosity complicates the liquid handling. Chromatography can be used for solvent exchange to overcome the solvent effect on the extinction coefficient. A precise and accurate sample injection volume can be guaranteed by using a full loop injection procedure in combination with a sufficient loop overfill. Moreover, the presence of high salt and especially high polymer content limits the use of affinity based chromatographic analytics such as reversed phase (RP) chromatography. This limitations suggest the use of size exclusion chromatography (SEC) for quantification of undiluted samples even though the relative slow sample throughput. Since the sample loading cannot be adapted to column capacity in order to increase detection limits in SEC but is rather limited by column dimension, the detection limit might be not sufficient for an accurate quantification and further more sensitive analytical procedures might be required. Thus, the fraction of interest of the SEC elute can be collected and further analyzed by highly accurate colorimetric or fluorometric assays. Conventional assays cannot be applied to the original sample due to interference of salt and polymer. A further concentrating step can be applied for sample preparation for the colorimetric or fluorometric assays by evaporation of the sample buffer. The quantification of sample volumes can be spectrometrically controlled by liquid pathlength determination by absorbance measurement of an internal standard such as a dye within the fraction collection plate or the water absorbance in the near IR area.

All these considerations lead to the conclusion to use low solute concentrations in order to minimize precipitation effects and reduce the influence of solute-solute interactions on the partitioning behavior in combination with a SEC for buffer exchange and quantification of undiluted samples even though the relatively slow sample throughput. A potential limitation of detection in SEC can be overcome by SEC elute fractionation followed by additional assay. Potential evaporation step can increase quantification accuracy and suggest the use of volatile buffers such as ammonium bicarbonate buffer as

SEC running buffer. This experimental design should provide a useful set of experimental data to build a predictive model based on interaction energies gained by extensive molecular dynamics simulations presented in the present work.

12.5 Conclusion

In this study full-atom molecular dynamics simulations were performed of a batch of solutes ranging from single amino acids, peptides into complex proteins in coexisting phases of aqueous two phase systems composed of potassium phosphate and polyethylene glycol of different molecular weight. The comprehensive set of *in-silico* derived data covers the interaction energies within highly complex solvents which might help to gain a mechanistic understanding of mechanisms of the complex partitioning behavior. The linear interaction energy can be applied to derive free energies by means of the interaction energies, which was already shown to be sufficient for computing solvation free energies and partitioning coefficients for small compounds. The enormous amount of data might serve for developing models for partitioning within aqueous two phase systems with predictive power.

12.6 Acknowledgement

This research work is part of the project "Molecular Interaction Engineering: From Nature's Toolbox to Hybrid Technical Systems", which is funded by the German Federal Ministry of Education and Research (BMBF), funding code 031A095B.

12.7 References

16. Chrysina, E. D., Brew, K. & Acharya, K. R. Crystal structures of apo-and holobovine α-lactalbumin at 2.2-Å resolution reveal an effect of calcium on inter-lobe interactions. *J. Biol. Chem.* **275,** 37021–37029 (2000).

41. Selber, K. *et al.* Large-scale separation and production of engineered proteins, designed for facilitated recovery in detergent-based aqueous two-phase extraction systems. *Process Biochem.* **39,** 889–896 (2004).

52. Harris, J. M. *Poly(Ethylene Glycol) chemistry: Biotechnical and biomedical applications* (Springer, 1992).

59. Steinbrecher, T. & Labahn, A. Towards accurate free energy calculations in ligand protein-binding studies. *Curr. Med. Chem.* **17,** 767–785 (2010).

70. Åqvist, J. Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.* **94,** 8021–8024 (1990).

71. Barone, V., Cossi, M. & Tomasi, J. A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. *J. Chem. Phys.* **107,** 3210–3221 (1997).

72. Sousa, S. F., Fernandes, P. A. & Ramos, M. J. Protein–ligand docking: Current status and future challenges. *Proteins: Struct., Funct., Bioinf.* **65,** 15–26 (2006).

73. Woo, H.-J. & Roux, B. Calculation of absolute protein–ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 6825–6830 (2005).

74. Leach, A. R., Shoichet, B. K. & Peishoff, C. E. Prediction of protein-ligand interactions. Docking and scoring: Successes and gaps. *J. Med. Chem.* **49,** 5851–5855 (2006).

75. Yang, H. *et al.* Protein conformational dynamics probed by single-molecule electron transfer. *Science* **302,** 262–266 (2003).

76. Winget, P., Cramer, C. J. & Truhlar, D. G. Computation of equilibrium oxidation and reduction potentials for reversible and dissociative electron-transfer reactions in solution. *Theor. Chem. Acc.* **112,** 217–227 (2004).

79. Roux, B., Nina, M., Pomès, R. & Smith, J. C. Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: A molecular dynamics free energy perturbation study. *Biophys. J.* **71,** 670–681 (1996).

80. Zhou, R., Berne, B. J. & Germain, R. The free energy landscape for β hairpin folding in explicit water. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 14931–14936 (2001).

81. Wang, W. & Kollman, P. A. Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. *J. Mol. Biol.* **303,** 567–582 (2000).

94. McDonald, N. A., Carlson, H. A. & Jorgensen, W. L. Free energies of solvation in chloroform and water from a linear response approach. *J. Phys. Org. Chem.* **10,** 563–576 (1997).

95. Carlson, H. A. & Jorgensen, W. L. An extended linear response method for determining free energies of hydration. *J. Phys. Chem.* **99,** 10667–10673 (1995).

123. Diederich, P., Amrhein, S., Hämmerling, F. & Hubbuch, J. Evaluation of PEG/phosphate aqueous two-phase systems for the purification of the chicken egg white protein avidin by using high-throughput techniques. *Chem. Eng. J.* **104,** 945–956 (2013).

159. Amrhein, S., Schwab, M.-L., Hoffmann, M. & Hubbuch, J. Characterization of aqueous two phase systems by combining lab-on-a-chip technology with robotic liquid handling stations. *J. Chromatogr. A* **1367,** 68–77 (2014).

209. Krieger, E. *et al.* Making optimal use of empirical energy functions: Force-field parameterization in crystal space. *Proteins* **57,** 678–83 (2004).

211. Berendsen, H. J. C. *et al.* Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3684–3690 (1984).

212. Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24,** 1999–2012 (2003).

213. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103,** 8577–8593 (1995).

214. Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **23,** 1623–1641 (2002).

228. Wang, J. *et al.* Triclinic lysozyme at 0.65 Å resolution. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **63,** 1254–1268 (2007).

230. Gordon, J. C. *et al.* H++: A server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **33,** W368–W371 (2005).

262. Martínez, J. M. & Martínez, L. Packing optimization for automated generation of complex system's initial configurations for molecular dynamics and docking. *J. Comput. Chem.* **24,** 819–825 (2003).

263. Martínez, L., Andrade, R., Birgin, E. G. & Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **30,** 2157–2164 (2009).

268. Wiendahl, M., Oelmeier, S. A., Dismer, F. & Hubbuch, J. High-throughput screening-based selection and scale-up of aqueous two-phase systems for pDNA purification. *J. Sep. Sci.* **35,** 3197–3207 (2012).

279. Merchuk, J. C., Andrews, B. A. & Asenjo, J. A. Aqueous two-phase systems for protein separation: Studies on phase inversion. *J. Chromatogr. B: Biomed. Sci. Appl.* **711,** 285–293 (1998).

295. Rosa, P., Ferreira, I., Azevedo, A. & Aires-Barros, M. Aqueous two-phase systems: A viable platform in the manufacturing of biopharmaceuticals. *J. Chromatogr. A* **1217,** 2296–2305 (2010).

296. Rosa, P. A., Azevedo, A. M. & Aires-Barros, M. R. Application of central composite design to the optimisation of aqueous two-phase extraction of human antibodies. *J. Chromatogr. A* **1141,** 50–60 (2007).

297. Azevedo, A. M., Rosa, P. A., Ferreira, I. F. & Aires-Barros, M. R. Optimisation of aqueous two-phase extraction of human antibodies. *J. Biotechnol.* **132,** 209–217 (2007).

298. Wiendahl, M. *High throughput screening in downstream processing of biotechnological products* (Cuvillier, 2008).

299. Steinbrecher, T., Case, D. A. & Labahn, A. Free energy calculations on the binding of novel thiolactomycin derivatives to E. coli fatty acid synthase I. *Bioorg. Med. Chem.* **20,** 3446–3453 (2012).

300. Genheden, S., Nilsson, I. & Ryde, U. Binding affinities of factor Xa inhibitors estimated by thermodynamic integration and MM/GBSA. *J. Chem. Inf. Model.* **51,** 947–958 (2011).

301. Steiner, D. *et al.* Calculation of binding free energies of inhibitors to plasmepsin II. *J. Comput. Chem.* **32,** 1801–1812 (2011).

302. Åqvist, J., Medina, C. & Samuelsson, J.-E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **7,** 385–391 (1994).

303. Åqvist, J. Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions. *J. Comput. Chem.* **17,** 1587–1597 (1996).

304. Van Lipzig, M. M. H. *et al.* Prediction of ligand binding affinity and orientation of xenoestrogens to the estrogen receptor by molecular dynamics simulations and the linear interaction energy method. *J. Med. Chem.* **47,** 1018–1030 (2004).

305. Terai, T. *et al.* Rational development of caged-biotin protein-labeling agents and some applications in live cells. *Chem. Biol.* **18,** 1261–1272 (2011).

306. Rawas, A., Muirhead, H. & Williams, J. Structure of diferric duck ovotransferrin at 2.35 Å resolution. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **52,** 631–640 (1996).

307. Oelmeier, S. A., Dismer, F. & Hubbuch, J. Molecular dynamics simulations on aqueous two-phase systems - Single PEG-molecules in solution. *BMC Biophys.* **5,** 14 (2012).

308. Åqvist, J. & Marelius, J. The linear interaction energy method for predicting ligand binding free energies. *Comb. Chem. High Throughput Screen.* **4,** 613–626 (2001).

309. Marelius, J., Hansson, T. & Åqvist, J. Calculation of ligand binding free energies from molecular dynamics simulations. *Int. J. Quantum Chem.* **69,** 77–88 (1998).

310. Cohn, E. J. The physical chemistry of the proteins. *Physiol. Rev.* **5,** 349–437 (1925).

CHAPTER 13

Conclusion and Outlook

In this work protein properties and protein behavior in solution were investigated intensively applying specially developed experimental high throughput techniques and molecular dynamics (MD) simulations based approaches.
Experimental methods were developed to assess surface tension of liquids fully automatically and highly accurately. Based on surface tension profiling a non-invasive method for protein hydrophobibicity assessment was established and applied to a number of solvent conditions and proteins. In addition to experimental methods, *in-silico* approaches based on MD simulations were developed in order to characterize peptides and proteins with respect to hydrophobicity. Thus, in this thesis experimental and MD simulations based approaches were developed for molecular peptide and protein assessment and the characterization of surface hydrophobicity in particular. The achievements of the molecular protein assessment were further applied to process related issues. Protein aggregation induced by precipitants, and sodium chloride and ammonium sulfate in particular, was investigated and correlated to experimentally derived protein surface properties and conformational flexibility determined by MD simulations. The *in-silico* derived conformational flexibility in presence of precipitants in varying concentrations in combination with experimentally derived protein surface properties was found to be capable of describing protein phase behavior in solution in terms of aggregation propensity. In order to investigate aqueous two-phase systems (ATPSs), microfluidic lab-on-a-chip technology and fully automated robotic liquid handling stations could successfully be combined in order to evaluate densities and compositions of coexisting phases of ATPSs composed of potassium phosphate and polyethylene glycol with molecular weights of 300, 400, 600, and 1000 Da. The hereby developed *Tip2World* interface expands the operation range of liquid handling stations tremendously. Starting from this comprehensive experimental data set of phase densities and compositions, molecular dynamics simulations of amino acids, peptides, and proteins in selected ATPSs were performed in order to investigate

the driving forces for protein partitioning at diluted protein concentrations. In this comprehensive study a huge data base of interaction energies acting on the respective solute within coexisting salt and PEG rich phases was generated. This data serves for model building for protein partitioning within aqueous two phase systems.
Currently, the density screening platform is refined in order to determine viscosity additionally. Viscosity is of major interest in process development since it affects the processibility such as filtration, formulation, or pumpability and can be used as a measure for protein-protein interactions in solution. Viscosity determination requires a precise control and regulation of temperature, which is currently in development. The platform then will be combined with the surface tension determination developed in this work and thus multiple parameters can be assessed simultaneously, which is highly beneficial in terms of throughput and sample consumption. It has been shown that the conformational flexibility can be derived from MD simulations and has a predictive power for protein stability. At the same time, the protein partitioning in ATPSs is highly affected by differences in solubility limits within coexisting phases at high protein concentrations. These findings suggest the combination of conformational flexibility within coexisting phases and interaction energies to evaluate solubility limitations. This approach could potentially be used to evaluate protein partitioning *in-silico* for process relevant protein concentrations. Moreover, the protein stability in polymer rich solutions is of high interest for the development of printable hydrogel bioinks. The *in-silico* estimation of protein stability within these hydrogels can be highly valuable and might be a further application of the findings and approaches originating from this work. The evaluation of differences in interaction energies of solutes in different solutions, as performed in this work, could be applied to other practical partitioning phenomena such as the liberation of drug formulations.

CHAPTER 14

Comprehensive Reference List

1. Watson, J. D., Crick, F. H., *et al.* Molecular structure of nucleic acids. *Nature* **171,** 737–738 (1953).
2. Cohen, S. N., Chang, A. C. Y., Boyer, H. W. & Helling, R. B. Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **70,** 3240–3244 (1973).
3. Chames, P., Van Regenmortel, M., Weiss, E. & Baty, D. Therapeutic antibodies: Successes, limitations and hopes for the future. *Br. J. Pharmacol.* **157,** 220–233 (2009).
4. Knier, B., Hemmer, B. & Korn, T. Novel monoclonal antibodies for therapy of multiple sclerosis. *Expert Opin. Biol. Ther.* **14,** 503–513 (2014).
5. Barkhof, F. & Ciccarelli, O. Daclizumab in multiple sclerosis: A high-yield extension study. *Lancet Neurol.* **13,** 443–444 (2014).
6. Demain, A. L. & Vaishnav, P. Production of recombinant proteins by microbes and higher organisms. *Biotechnol. Adv.* **27,** 297–306 (2009).
7. FDA. *Guidance for industry process validation: General principles and practices* 2008.
8. Friend, S., Swanick, M. & Arlington, S. *Pharma 2020: Supplying the future–Which path will you take? Price Waterhouse Company* 2011.
9. Warikoo, V. *et al.* Integrated continuous production of recombinant therapeutic proteins. *Biotechnol. Bioeng.* **109,** 3018–3029 (2012).
10. Jungbauer, A. & Peng, J. Continous bioprocessing: An interview with Konstantin Konstantinov from Genzyme. *Biotechnol. J.* **6,** 1431–1433 (2011).
11. Croughan, M. S., Konstantinov, K. B. & Cooney, C. The future of industrial bioprocessing: Batch or continuous? *Biotechnol. Bioeng.* **112,** 648–651 (2015).

12. Cramer, S. M. & Holstein, M. A. Downstream bioprocessing: Recent advances and future promise. *Curr. Opin. Chem. Eng.* **1.** Open issue 1/1, 27–37 (2011).

13. Konstantinov, K., Godawat, R., Warikoo, V. & Jain, S. *Integrated continuous manufacturing of therapeutic protein drug substances* US Patent App. 14/195,481. 2014.

14. Das, R. *et al.* Structure prediction for CASP7 targets using extensive all-atom refinement with Rosetta@home. *Proteins: Struct., Funct., Bioinf.* **69,** 118–128 (2007).

15. Voet, D. & Voet, J. *Biochemistry* (Wiley, 2010).

16. Chrysina, E. D., Brew, K. & Acharya, K. R. Crystal structures of apo-and holobovine α-lactalbumin at 2.2-Å resolution reveal an effect of calcium on inter-lobe interactions. *J. Biol. Chem.* **275,** 37021–37029 (2000).

17. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16,** 404–405 (2000).

18. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* **31,** 3381–3385 (2003).

19. Kelley, L. A. & Sternberg, M. J. Protein structure prediction on the Web: A case study using the Phyre server. *Nat. Protoc.* **4,** 363–371 (2009).

20. Perrin, D., Dempsey, B. & Serjeant, E. English. in *pKa Prediction for Organic Acids and Bases* 12–20 (Springer Netherlands, 1981).

21. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157,** 105–132 (1982).

22. Sweet, R. M. & Eisenberg, D. Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure. *J. Mol. Biol.* **171,** 479–488 (1983).

23. Abraham, D. J. & Leo, A. J. Extension of the fragment method to calculate amino acid zwitterion and side chain partition coefficients. *Proteins: Struct., Funct., Bioinf.* **2,** 130–152 (1987).

24. Semisotnov, G. *et al.* Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers* **31,** 119–128 (1991).

25. Bertsch, M., Mayburd, a. L. & Kassner, R. J. The identification of hydrophobic sites on the surface of proteins using absorption difference spectroscopy of bromophenol blue. *Anal. Biochem.* **313,** 187–95 (2003).

26. Alizadeh-Pasdar, N. & Li-Chan, E. C. Comparison of protein surface hydrophobicity measured at various pH values using three different fluorescent probes. *J. Agric. Food Chem.* **48,** 328–334 (2000).

27. Azevedo, A. M., Rosa, P. A., Ferreira, I. F. & Aires-Barros, M. R. Chromatography-free recovery of biopharmaceuticals through aqueous two-phase processing. *Trends Biotechnol.* **27,** 240–247 (2009).

28. Becker, J., Thomas, O. & Franzreb, M. Protein separation with magnetic adsorbents in micellar aqueous two-phase systems. *Sep. Purif. Technol.* **65,** 46–53 (2009).

29. Rito-Palomares, M. & Lyddiatt, A. Process integration using aqueous two-phase partition for the recovery of intracellular proteins. *Chem. Eng. J.* **87,** 313–319 (2002).

30. Chen, J.-P. & Lee, M.-S. Enhanced production of Serratia marcescens chitinase in PEG/dextran aqueous two-phase systems. *Enzyme Microb. Technol.* **17,** 1021–1027 (1995).

31. Wei, D.-Z., Zhu, J.-H. & Cao, X.-J. Enzymatic synthesis of cephalexin in aqueous two-phase systems. *Biochem. Eng. J.* **11,** 95–99 (2002).

32. Albertsson, P.-Å. Chromatography and partition of cells and cell fragments. *Nature* **177,** 771–774 (1956).

33. Albertsson, P.-Å. Partition of proteins in liquid polymer-polymer two-phase dystems. *Nature* **182,** 709–711 (1958).

34. Albertsson, P.-Å. & Nyns, E. J. Counter-current distribution of proteins in aqueous polymer phase systems. *Nature* **184,** 1465–1468 (1959).

35. Philipson, L., Albertsson, P. & Frick, G. The purification and concentration of viruses by aqueous polymer phase systems. *Virology* **11,** 553–571 (1960).

36. Albertsson, P.-Å. & Philipson, L. Antigen-antibody in liquid two-phase systems: A method for studying immunological reactions. *Nature* **185,** 38–40 (1960).

37. Norrby, E. C. J. & Albertsson, P. Å. Concentration of poliovirus by an aqueous polymer two-phase system. *Nature* **188,** 1047–1048 (1960).

38. Albertsson, P.-Å., Sasakawa, S. & Walter, H. Cross partition and isoelectric points of proteins. *Nature* **228,** 1329–1330 (1970).

39. Hart, R. A. *et al.* Large scale, in situ isolation of periplasmic IGF–I from E. coli. *Nat. Biotechnol.* **12,** 1113–1117 (1994).

40. Costa, M. J. L., Cunha, M. T., Cabral, J. M. & Aires-Barros, M. R. Scale-up of recombinant cutinase recovery by whole broth extraction with PEG-phosphate aqueous two-phase. *Bioseparation* **9,** 231–238 (2000).

41. Selber, K. *et al.* Large-scale separation and production of engineered proteins, designed for facilitated recovery in detergent-based aqueous two-phase extraction systems. *Process Biochem.* **39,** 889–896 (2004).

42. Frerix, A., Müller, M., Kula, M.-R. & Hubbuch, J. Scalable recovery of plasmid DNA based on aqueous two-phase separation. *Biotechnol. Appl. Biochem.* **42,** 57–66 (2005).

43. Cunha, T. & Aires-Barros, R. Large-scale extraction of proteins. English. *Mol. Biotechnol.* **20,** 29–40 (2002).

44. Greve, A. & Kula, M.-R. Recycling of salts in partition protein extraction processes. *J. Chem. Technol. Biotechnol.* **50,** 27–42 (1991).

45. Rito-Palomares, M. Practical application of aqueous two-phase partition to process development for the recovery of biological products. *J. Chromatogr. B* **807.** 12th International Conference on Biopartitioning and Purification, 3–11 (2004).

46. Hatti-Kaul, R. *Aqueous two-phase systems: Methods and protocols* (Humana Press, 2000).

47. Harris, E. & Angal, S. *Protein purification spplications: A practical approach* (IRL Press, 1990).

48. Franco, T., Andrews, A. & Asenjo, J. Use of chemically modified proteins to study the effect of a single protein property on partitioning in aqueous two-phase systems: Effect of surface charge. *Biotechnol. Bioeng.* **49,** 309–315 (1996).

49. Hachem, F., Andrews, B. & Asenjo, J. Hydrophobic partitioning of proteins in aqueous two-phase systems. *Enzyme Microb. Technol.* **19,** 507–517 (1996).

50. Berggren, K. *et al.* The surface exposed amino acid residues of monomeric proteins determine the partitioning in aqueous two-phase systems. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1596,** 253–268 (2002).

51. Asenjo, J. A. & Andrews, B. A. Aqueous two-phase systems for protein separation: A perspective. *J. Chromatogr. A* **1218,** 8826–8835 (2011).

52. Harris, J. M. *Poly(Ethylene Glycol) chemistry: Biotechnical and biomedical applications* (Springer, 1992).

53. Andrews, B. A., Schmidt, A. S. & Asenjo, J. A. Correlation for the partition behavior of proteins in aqueous two-phase systems: Effect of surface hydrophobicity and charge. *Biotechnol. Bioeng.* **90,** 380–390 (2005).

54. Franco, T., Andrews, A. & Asenjo, J. Conservative chemical modification of proteins to study the effects of a single protein property on partitioning in aqueous two-phase systems. *Biotechnol. Bioeng.* **49,** 290–299 (1996).

55. Saravanan, S., Rao, J. R., Nair, B. U. & Ramasami, T. Aqueous two-phase poly(ethylene glycol)–poly(acrylic acid) system for protein partitioning: Influence of molecular weight, pH and temperature. *Process Biochem.* **43,** 905–911 (2008).

56. Johansson, H.-O., Karlström, G., Tjerneld, F. & Haynes, C. A. Driving forces for phase separation and partitioning in aqueous two-phase systems. *J. Chromatogr. B: Biomed. Sci. Appl.* **711,** 3–17 (1998).

57. Greer, J., Erickson, J. W., Baldwin, J. J. & Varney, M. D. Application of the three-dimensional structures of protein target molecules in structure-based drug design. *J. Med. Chem.* **37,** 1035–1054 (1994).

58. Von Itzstein, M. *et al.* Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **363,** 418–423 (1993).

59. Steinbrecher, T. & Labahn, A. Towards accurate free energy calculations in ligand protein-binding studies. *Curr. Med. Chem.* **17,** 767–785 (2010).

60. Claußen, H., Buning, C., Rarey, M. & Lengauer, T. FlexE: Efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **308,** 377–395 (2001).

61. Carlson, H. A. Protein flexibility is an important component of structure-based drug discovery. *Curr. Pharm. Des.* **8,** 1571–1578 (2002).

62. Cavasotto, N. & Narender Singh, C. Docking and high throughput docking: Successes and the challenge of protein flexibility. *Curr. Comput. Aided Drug Des.* **4,** 221–234 (2008).

63. Alonso, H., Bliznyuk, A. A. & Gready, J. E. Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.* **26,** 531–568 (2006).

64. Lindorff-Larsen, K. *et al.* Simultaneous determination of protein structure and dynamics. *Nature* **433,** 128–132 (2005).

65. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314,** 141–151 (1999).

66. Snow, C. D., Nguyen, H., Pande, V. S. & Gruebele, M. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **420,** 102–106 (2002).

67. Shaw, D. E. *et al.* Atomic-level characterization of the structural dynamics of proteins. *Science* **330,** 341–346 (2010).

68. Fan, H. & Mark, A. E. Refinement of homology-based protein structures by molecular dynamics simulation techniques. *Protein Sci.* **13,** 211–220 (2004).

69. Cornell, W. D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117,** 5179–5197 (1995).

70. Åqvist, J. Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.* **94,** 8021–8024 (1990).

71. Barone, V., Cossi, M. & Tomasi, J. A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. *J. Chem. Phys.* **107,** 3210–3221 (1997).

72. Sousa, S. F., Fernandes, P. A. & Ramos, M. J. Protein–ligand docking: Current status and future challenges. *Proteins: Struct., Funct., Bioinf.* **65,** 15–26 (2006).

73. Woo, H.-J. & Roux, B. Calculation of absolute protein–ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 6825–6830 (2005).

74. Leach, A. R., Shoichet, B. K. & Peishoff, C. E. Prediction of protein-ligand interactions. Docking and scoring: Successes and gaps. *J. Med. Chem.* **49,** 5851–5855 (2006).

75. Yang, H. *et al.* Protein conformational dynamics probed by single-molecule electron transfer. *Science* **302,** 262–266 (2003).

76. Winget, P., Cramer, C. J. & Truhlar, D. G. Computation of equilibrium oxidation and reduction potentials for reversible and dissociative electron-transfer reactions in solution. *Theor. Chem. Acc.* **112,** 217–227 (2004).

77. Hansch, C., Quinlan, J. E. & Lawrence, G. L. Linear free-energy relationship between partition coefficients and the aqueous solubility of organic liquids. *J. Org. Chem.* **33,** 347–350 (1968).

78. De Young, L. R. & Dill, K. A. Solute partitioning into lipid bilayer membranes. *Biochemistry* **27,** 5281–5289 (1988).

79. Roux, B., Nina, M., Pomès, R. & Smith, J. C. Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: A molecular dynamics free energy perturbation study. *Biophys. J.* **71,** 670–681 (1996).

80. Zhou, R., Berne, B. J. & Germain, R. The free energy landscape for β hairpin folding in explicit water. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 14931–14936 (2001).

81. Wang, W. & Kollman, P. A. Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. *J. Mol. Biol.* **303,** 567–582 (2000).

82. Chong, L. T. *et al.* Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 14330–14335 (1999).

83. Lang, K. M. *et al.* A comprehensive molecular dynamics approach to proteinretention modeling in ion exchange chromatography. *J. Chromatogr. A* **1381,** 184–193 (2015).

84. Mortier, J. *et al.* The impact of molecular dynamics on drug design: Applications for the characterization of ligand–macromolecule complexes. *Drug Discov. Today* (2015).

85. Van Gunsteren, W. F., Daura, X. & Mark, A. E. Computation of free energy. *Helv. Chim. Acta* **85,** 3113–3129 (2002).

86. McQuarrie, D. *Statistical Mechanics* (University Science Books, 2000).

87. Feig, M. & Brooks, C. L. Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr. Opin. Struct. Biol.* **14,** 217–224 (2004).

88. Marrink, S. J. *et al.* The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111,** 7812–7824 (2007).

89. Åqvist, J., Luzhkov, V. B. & Brandsdal, B. O. Ligand binding affinities from MD simulations. *Acc. Chem. Res.* **35,** 358–365 (2002).

90. Gutiérrez-de-Terán, H. & Åqvist, J. English. in *Computational Drug Discovery and Design* (ed Baron, R.) 305–323 (Springer New York, 2012).

91. De Amorim, H. L. N., Caceres, R. A. & Netz, P. A. Linear interaction energy (LIE) method in lead discovery and optimization. *Curr. Drug Targets* **9,** 1100–1105 (2008).

92. Hansson, T., Marelius, J. & Åqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput. Aided Mol. Des.* **12,** 27–35 (1998).

93. Åqvist, J. & Hansson, T. On the validity of electrostatic linear response in polar solvents. *J. Phys. Chem.* **100,** 9512–9521 (1996).

94. McDonald, N. A., Carlson, H. A. & Jorgensen, W. L. Free energies of solvation in chloroform and water from a linear response approach. *J. Phys. Org. Chem.* **10,** 563–576 (1997).

95. Carlson, H. A. & Jorgensen, W. L. An extended linear response method for determining free energies of hydration. *J. Phys. Chem.* **99,** 10667–10673 (1995).

96. Sulea, T., Corbeil, C. R. & Purisima, E. O. Rapid prediction of solvation free energy. 1. An extensive test of linear interaction energy (LIE). *J. Chem. Theory Comput.* **6,** 1608–1621 (2010).

97. Pearce, J. M. *Open-source lab: How to build your own hardware and reduce research costs* (Newnes, 2013).

98. Pearce, J. M. Building research equipment with free, open-source hardware. *Science* **337,** 1303–1304 (2012).

99. Bak, D. Rapid prototyping or rapid production? 3D printing processes move industry towards the latter. *Assembly Autom.* **23,** 340–345 (2003).

100. Jones, D. B. *et al.* Three-dimensional modeling may improve surgical education and clinical practice. *Surg. Innov.* (2015).

101. Fee, C., Nawada, S. & Dimartino, S. 3D printed porous media columns with fine control of column packing morphology. *J. Chromatogr. A* **1333,** 18–24 (2014).

102. Zhang, Y., Lyu, F., Ge, J. & Liu, Z. Ink-jet printing an optimal multi-enzyme system. *Chem. Commun.* **50,** 12919–12922 (85 2014).

103. Hung, A. M., Noh, H. & Cha, J. N. Recent advances in DNA-based directed assembly on surfaces. *Nanoscale* **2,** 2530–2537 (2010).

104. Vashist, S. K. *et al.* Immobilization of antibodies and enzymes on 3-amino"-propyl"-triethoxy"-silane-functionalized bioanalytical platforms for biosensors and diagnostics. *Chem. Rev.* **114,** 11083–11130 (2014).

105. Gdor, E., Shemesh, S., Magdassi, S. & Mandler, D. Multienzyme inkjet printed 2D arrays. *ACS Appl. Mater. Interfaces* **7,** 17985–17992 (2015).

106. Rengier, F. *et al.* 3D printing based on imaging data: Review of medical applications. English. *Int. J. Comput. Assist. Radiol. Surg.* **5,** 335–341 (2010).

107. Leukers, B. *et al.* Hydroxyapatite scaffolds for bone tissue engineering made by 3D printing. *J. Mater. Sci.: Mater. Med.* **16,** 1121–1124 (2005).

108. McGurk, M., Amis, A., Potamianos, P. & Goodger, N. Rapid prototyping techniques for anatomical modelling in medicine. *Ann. R. Coll. Surg. Engl.* **79,** 169–174 (1997).

109. Mironov, V., Kasyanov, V., Drake, C. & Markwald, R. R. Organ printing: Promises and challenges. **3,** 93–103 (2008).

110. Mironov, V. *et al.* Organ printing: Computer-aided jet-based 3D tissue engineering. *Trends Biotechnol.* **21,** 157–161 (2003).

111. Mironov, V., Kasyanov, V. & Markwald, R. R. Organ printing: From bioprinter to organ biofabrication line. *Curr. Opin. Biotechnol.* **22,** 667–673 (2011).

112. Curodeau, A., Sachs, E. & Caldarise, S. Design and fabrication of cast orthopedic implants with freeform surface textures from 3-D printed ceramic shell. *J. Biomed. Mater. Res.* **53,** 525–535 (2000).

113. Murphy, S. V. & Atala, A. 3D bioprinting of tissues and organs. *Nat. Biotechnol.* **32,** 773–785 (2014).

114. Khalyfa, A. *et al.* Development of a new calcium phosphate powder-binder system for the 3D printing of patient specific implants. *J. Mater. Sci.: Mater. Med.* **18,** 909–916 (2007).

115. Warnke, P. H. *et al.* Ceramic scaffolds produced by computer-assisted 3D printing and sintering: Characterization and biocompatibility investigations. *J. Biomed. Mater. Res. Part B Appl. Biomater* **93,** 212–217 (2010).

116. Ihalainen, P., Määttänen, A. & Sandler, N. Printing technologies for biomolecule and cell-based applications. *Int. J. Pharm.* **494.** The potential for 2D and 3D Printing to Pharmaceutical Development, 585–592 (2015).

117. Kruth, J.-P., Wang, X., Laoui, T. & Froyen, L. Lasers and materials in selective laser sintering. *Assembly Autom.* **23,** 357–371 (2003).

118. Santos, E. C., Shiomi, M., Osakada, K. & Laoui, T. Rapid manufacturing of metal components by laser forming. *Int. J. Mach. Tools Manuf.* **46,** 1459–1468 (2006).

119. Tanford, C. Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J. Am. Chem. Soc.* **84,** 4240–4247 (1962).

120. Dill, K. A. Dominant forces in protein folding. *Biochemistry* **29,** 7133–55 (1990).

121. Brems, D. N., Plaisted, S. M., Havel, H. a. & Tomich, C. S. Stabilization of an associated folding intermediate of bovine growth hormone by site-directed mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **85,** 3367–3371 (1988).

122. Nieba, L., Honegger, A., Krebber, C. & Pluckthun, A. Disrupting the hydrophobic patches at the antibody variable/constant domain interface: Improved in vivo folding and physical characterization of an engineered scFv fragment. *Protein Eng.* **10,** 435–44 (1997).

123. Diederich, P., Amrhein, S., Hämmerling, F. & Hubbuch, J. Evaluation of PEG/phosphate aqueous two-phase systems for the purification of the chicken egg white protein avidin by using high-throughput techniques. *Chem. Eng. J.* **104,** 945–956 (2013).

124. Janson, J.-C. *Protein purification: Principles, high resolution methods, and applications* (John Wiley & Sons, 2012).

125. Janin, J. Surface and inside volumes in globular proteins. *Nature* **277,** 491–492 (1979).

126. Eisenberg, D. Three-dimensional structure of membrane and surface proteins. *Annu. Rev. Biochem. Biochem.* **53,** 595–623 (1984).

127. Black, S. D. & Mould, D. R. Development of hydrophobicity parameters to analyze proteins which bear post- or cotranslational modifications. *Anal. Biochem.* **193,** 72–82 (1991).

128. Rose, G. D. & Wolfenden, R. Hydrogen bonding, hydrophobicity, packing, and protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **22,** 381–415 (1993).

129. Whitney, P. L., Tanford, C., W, P. L. & Tanford, C. Solubility of amino acids in aqueous urea solutions and its implications for the denaturation of proteins by urea. *J. Biol. Chem.* **237,** 1735–1737 (1962).

130. Nozaki, Y. & Tanford, C. The solubility of amino acids and related compounds in aqueous urea solutions. *J. Biol. Chem.* **238,** 4074–4081 (1963).

131. Nozaki, Y. & Tanford, C. The solubility of amino acids and related compounds in equeous ethylene glycol solutions. *J. Biol. Chem.* **240,** 3568–3573 (1965).

132. Nozaki, Y. & Tanford, C. The solubility of amino acids, diglycine, and triglycine in aqueous guanidine hydrochloride solutions. *J. Biol. Chem.* **245,** 1648–1652 (1970).

133. Nozaki, Y. & Tanford, C. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. *J. Biol. Chem.* **246,** 2211–2217 (1971).

134. Dooley, K. H. & Castellino, F. J. Solubility of amino acids in aqueous guanidinium thiocyanate solutions. *Biochemistry* **11,** 1870–1874 (1972).

135. Fendler, J. H., Nome, F. & Nagyvary, J. Compartmentalization of amino acids in surfactant aggregates. *J. Mol. Evol.* **6,** 215–232 (1975).

136. Radzicka, A. & Wolfenden, R. Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry* **27,** 1664–1670 (1988).

137. Shanbhag, V. P. & Axelsson, C.-G. Hydrophobic interaction determined by partition in aqueous two-phase systems. *Eur. J. Biochem.* **60,** 17–22 (1975).

138. Keshavarz, E. & Nakai, S. The relationship between hydrophobicity and interfacial tension of proteins. *Biochim. Biophys. Acta* **576,** 269–279 (1979).

139. Kato, A. & Nakai, S. Hydrophobicity determined by a fluorescence probe method and its correlation with surface properties of proteins. *Biochim. Biophys. Acta* **624,** 13–20 (1980).

140. Cardamone, M. & Puri, N. K. Spectrofluorimetric assessment of the surface hydrophobicity of proteins. *Biochem. J.* **282,** 589–93 (1992).

141. Hendriks, J. *et al.* Transient exposure of hydrophobic surface in the photoactive yellow protein monitored with Nile Red. *Biophys. J.* **82,** 1632–1643 (2002).

142. Hawe, A., Sutter, M. & Jiskoot, W. Extrinsic fluorescent dyes as tools for protein characterization. *Pharm. Res.* **25,** 1487–1499 (2008).

143. Salgado, J. C., Rapaport, I. & Asenjo, J. A. Is it possible to predict the average surface hydrophobicity of a protein using only its amino acid composition? *J. Chromatogr. A* **1075,** 133–143 (2005).

144. Miller, S., Janin, J., Lesk, A. M. & Chothia, C. Interior and surface of monomeric proteins. *J. Mol. Biol.* **196,** 641–56 (1987).

145. Lijnzaad, P., Berendsen, H. J. & Argos, P. A method for detecting hydrophobic patches on protein surfaces. *Proteins* **26,** 192–203 (1996).

146. Chennamsetty, N. *et al.* Design of therapeutic proteins with enhanced stability. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 11937–11942 (2009).

147. Salgado, J. C., Rapaport, I. & Asenjo, J. A. Predicting the behaviour of proteins in hydrophobic interaction chromatography: 1: Using the hydrophobic imbalance (HI) to describe their surface amino acid distribution. *J. Chromatogr. A* **1107,** 110–119 (2006).

148. Salgado, J. C., Rapaport, I. & Asenjo, J. A. Predicting the behaviour of proteins in hydrophobic interaction chromatography: 2. Using a statistical description of their surface amino acid distribution. *J. Chromatogr. A* **1107,** 120–129 (2006).

149. Mahn, A., Lienqueo, M. & Asenjo, J. Effect of surface hydrophobicity distribution on retention of ribonucleases in hydrophobic interaction chromatography. *J. Chromatogr. A* **1043.** 23rd International Symposium on the Separation of Proteins, Peptides and Polynucleotides, 47–55 (2004).

150. Chen, J. *et al.* Investigation of protein retention in hydrophobic interaction chromatographic (HIC) systems using the preferential interaction theory and quantitative structure property relationship models. *Reactive and Functional Polymers* **67.** Special Issue Dedicated to Professor Helfferich, 1561–1569 (2007).

151. Amrhein, S., Oelmeier, S. A., Dismer, F. & Hubbuch, J. Molecular dynamics simulations approach for the characterization of peptides with respect to hydrophobicity. *J. Phys. Chem. B* **118,** 1707–14 (2014).

152. Reißer, S., Strandberg, E., Steinbrecher, T. & Ulrich, A. S. 3D hydrophobic moment vectors as a tool to characterize the surface polarity of amphiphilic peptides. *Biophys. J.* **106,** 2385–2394 (2014).

153. Trinquier, G. & Sanejouand, Y.-H. Which effective property of amino acids is best preserved by the genetic code? *Protein Eng.* **11,** 153–169 (1998).

154. Biswas, K. M., DeVido, D. R. & Dorsey, J. G. Evaluation of methods for measuring amino acid hydrophobicities and interactions. *J. Chromatogr. A* **1000,** 637–655 (2003).

155. Bull, H. B. & Breese, K. Surface tension of amino acid solutions: A hydrophobicity scale of the amino acid residues. *Arch. Biochem. Biophys.* **161,** 665–670 (1974).

156. Absolom, D. R., Zingg, W. & Neumann, A. W. Protein adsorption to polymer particles: Role of surface properties. *J. Biomed. Mater. Res.* **21,** 161–71 (1987).

157. Genest, S. *et al.* Characterization of highly substituted, cationic amphiphilic starch derivatives: Dynamic surface tension and intrinsic viscosity. *Starch/Stärke* **65,** 999–1010 (2013).

158. Tate, T. XXX. On the magnitude of a drop of liquid formed under different circumstances. *Philos. Mag.* **27,** 176–180 (1864).

159. Amrhein, S., Schwab, M.-L., Hoffmann, M. & Hubbuch, J. Characterization of aqueous two phase systems by combining lab-on-a-chip technology with robotic liquid handling stations. *J. Chromatogr. A* **1367,** 68–77 (2014).

160. Tsierkezos, N. G. & Molinou, I. E. Thermodynamic properties of water + ethylene glycol at 283.15, 293.15, 303.15, and 313.15 K. *J. Chem. Eng. Data* **43,** 989–993 (1998).

161. Jańczuk, B., Białopiotrowicz, T. & Wójcik, W. The components of surface tension of liquids and their usefulness in determinations of surface free energy of solids. *J. Colloid Interface Sci.* **127,** 59–66 (1989).

162. Melinder, Å. *Thermophysical properties of aqueous solutions used as secondary working fluids* PhD thesis (School of Industrial Engineering and Management, Royal Institute of Technology, KTH Stockholm, 2007).

163. Pegram, L. M. & Record, M. T. Hofmeister salt effects on surface tension arise from partitioning of anions and cations between bulk water and the air-water interface. *J. Phys. Chem. B* **111,** 5411–5417 (2007).

164. Permyakov, E. A. & Berliner, L. J. alpha-Lactalbumin: Structure and function. *FEBS Lett.* **473,** 269–274 (2000).

165. Guo, D. *et al.* Prediction of peptide retention times in reversed-phase high-performance liquid chromatography I. Determination of retention coefficients of amino acid residues of model synthetic peptides. *J. Chromatogr. A* **359,** 499–518 (1986).

166. Bigelow, C. C. On the average hydrophobicity of proteins and the relation between it and protein structure. *J. Theor. Biol.* **16,** 187–211 (1967).

167. Cao, W. G. *et al.* Mechanism of the interaction between bromophenol blue and bovine serum albumin. *Spectrosc. Lett.* **36,** 197–209 (2003).

168. Wei, Y.-J., Li, K.-A. & Tong, S.-Y. The interaction of bromophenol blue with proteins in acidic solution. *Talanta* **43,** 1–10 (1996).

169. Tayyab, S. & Qasim, M. Binding of bromophenol blue to bovine serum albumin and its succinylated forms. *Int. J. Biol. Macromol.* **12,** 55–58 (1990).

170. Murakami, K., Sano, T. & Yasunaga, T. Kinetic studies of the interaction of bromophenol blue with bovine serum albumin by pressure-jump method. *Bull. Chem. Soc. Jpn.* **54,** 862–868 (1981).

171. Kragh-Hansen, U., Møller, J. V. & Lind, K. E. Relation between binding of phenolsulfophtalein dyes and other ligands with a high affinity for human serum albumin. *Biochim. Biophys. Acta* **365,** 360–371 (1974).

172. Bjerrum, O. J. Interaction of bromphenol blue and bilirubin with bovine and human serum albumin determined by gel filtration. *Scand. J. Clin. Lab. Investig.* **22,** 41–48 (1968).

173. Waldmann-Meyer, H. & Schilling, K. The interaction of bromophenol blue with serum albumin and γ-globulin in acid medium. *Arch. Biochem. Biophys.* **64,** 291–301 (1956).

174. Flores, R. A rapid and reproducible assay for quantitative estimation of proteins using bromophenol blue. *Anal. Biochem.* **88,** 605–611 (1978).

175. Mayburd, A. L., Tan, Y. & Kassner, R. J. Complex formation between chromatium vinosum ferric cytochrome c' and bromophenol blue. *Arch. Biochem. Biophys.* **378,** 40–44 (2000).

176. Heikkila, T. *et al.* 96-Well plate surface tension measurements for fast determination of drug solubility. *Lett. Drug Des. Discov.* **5,** 471–476 (2008).

177. Petereit, A. C. *et al.* Prediction of blood-brain barrier penetration of poorly soluble drug candidates using surface activity profiling. *Eur. J. Pharm. Biopharm.* **75,** 405–410 (2010).

178. Seelig, A., Gottschlich, R. & Devant, R. M. A method to determine the ability of drugs to diffuse through the blood-brain barrier. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 68–72 (1994).

179. Suomalainen, P., Johans, C., Söderlund, T. & Kinnunen, P. K. Surface activity profiling of drugs applied to the prediction of blood-brain barrier permeability. *J. Med. Chem.* **47,** 1783–1788 (2004).

180. Crandall, D. *et al.* Characterization and comparative evaluation of a structurally unique PAI-1 inhibitor exhibiting oral in-vivo efficacy. *J. Thromb. Haemost.* **2,** 1422–1428 (2004).

181. Finholt, P. & Solvang, S. Dissolution kinetics of drugs in human gastric juice - the role of surface tension. *J. Pharm. Sci.* **57,** 1322–1326 (1968).

182. Jaitely, V., Karatas, A. & Florence, A. T. Water-immiscible room temperature ionic liquids (RTILs) as drug reservoirs for controlled release. *Int. J. Pharm.* **354,** 168–173 (2008).

183. Vladisavljević, G. T., Shimizu, M. & Nakashima, T. Production of multiple emulsions for drug delivery systems by repeated SPG membrane homogenization: Influence of mean pore size, interfacial tension and continuous phase viscosity. *J. Membr. Sci.* **284,** 373–383 (2006).

184. Amrhein, S., Christin Bauer, K., Galm, L. & Hubbuch, J. Non-invasive high throughput approach for protein hydrophobicity determination based on surface tension. *Biotechnol. Bioeng.* **112,** 2485–2494 (2015).

185. Mayr, L. M. & Bojanic, D. Novel trends in high-throughput screening. *Curr. Opin. Pharmacol.* **9.** Anti-infectives/New technologies, 580–588 (2009).

186. Harkins, W. D. & Brown, F. The determination of surface tension (free surface energy), and the weight of falling drops: The surface tension of water and benzene by the capillary height method. *J. Am. Chem. Soc.* **41,** 499–524 (1919).

187. Wilkinson, M. Extended use of, and comments on, the drop-weight (drop-volume) technique for the determination of surface and interfacial tensions. *J. Colloid Interface Sci.* **40,** 14–26 (1972).

188. Richards, T. W. & Coombs, L. B. the surface tensions of water, methyl, ethyl and isobutyl alcohols, ethyl butyrate, benzene and toluene. *J. Am. Chem. Soc.* **37,** 1656–1676 (1915).

189. Vazquez, G., Alvarez, E. & Navaza, J. M. Surface tension of alcohol water+ water from 20 to 50. degree. C. *J. Chem. Eng. Data* **40,** 611–614 (1995).

190. Sell, P.-J. & Renzow, D. Bestimmung des Benetzungsverhaltens von Pigmenten. *Prog. Org. Coat.* **3,** 323–348 (1975).

191. Fowkes, F. M. Attractive forces at interfaces. *Ind. Eng. Chem.* **56,** 40–52 (1964).

192. Wolf, K. & Wolff, R. Oberflächenspannung und Oberflächenaktivität. *Houben-Weyl: Methoden der organischen Chemie. Bd. III Teil* **1,** 449–479 (1955).

193. Marwedel, G. Zusammenhänge zwischen Oberflächenspannung. *Dichte und Viskosität organischer (besonders lacktechnische) Flüssigkeiten bei übereinstimmenden Temperaturen, Farbe und Lack* **69,** 437–451 (1963).

194. Della Volpe, C. & Siboni, S. Some reflections on acid-base solid surface free energy theories. *J. Colloid Interface Sci.* **195,** 121–136 (1997).

195. Eötvös, R. Über den Zusammenhang der Oberflächenspannung der Flüssigkeiten mit ihrem Molecularvolumen. *Annalen der Physik* **263,** 448–459 (1886).

196. Yadav, J. *Advanced practical physical chemisty* (Krishna Prakashan).

197. Chattoraj, D. *Adsorption and the Gibbs Surface Excess* (Springer US, 2012).

198. Asenjo, J., Schmidt, A., Hachem, F. & Andrews, B. Model for predicting the partition behaviour of proteins in aqueous two-phase systems. *J. Chromatogr. A* **668.** 8th International Conference on Partitioning in Aqueous Two-Phase Systems, 47–54 (1994).

199. Hawe, A., Sutter, M. & Jiskoot, W. Extrinsic fluorescent dyes as tools for protein characterization. English. *Pharm. Res.* **25,** 1487–1499 (2008).

200. Hessa, T. *et al.* Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433,** 377–381 (2005).

201. Wimley, W. C. & White, S. H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Mol. Biol.* **3,** 842–848 (1996).

202. Guy, H. Amino acid side-chain partition energies and distribution of residues in soluble proteins. *Biophys. J.* **47,** 61–70 (1985).

203. Moelbert, S., Emberly, E. & Tang, C. Correlation between sequence hydrophobicity and surface-exposure pattern of database proteins. *Protein Sci.* **13,** 752–762 (2004).

204. Chothia, C. The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol.* **105,** 1–12 (1976).

205. Freed, A. S., Garde, S. & Cramer, S. M. Molecular simulations of multimodal ligand-protein binding: Elucidation of binding sites and correlation with experiments. *J. Phys. Chem. B* **115,** 13320–13327 (2011).

206. Oelmeier, S., Dismer, F. & Hubbuch, J. Gaining mechanistic understanding of aqueous two-phase systems for bioseparation. *Chem. Ing. Tech.* **84,** 1292–1292 (2012).

207. Bordoli, L. & Schwede, T. English. in *Homology Modeling* (eds Orry, A. J. W. & Abagyan, R.) 107–136 (Humana Press, New York, USA, 2012).

208. Cavasotto, C. N. & Phatak, S. S. Homology modeling in drug discovery: Current trends and applications. *Drug Discov. Today* **14,** 676–683 (2009).

209. Krieger, E. *et al.* Making optimal use of empirical energy functions: Force-field parameterization in crystal space. *Proteins* **57,** 678–83 (2004).

210. Krieger, E., Koraimann, G. & Vriend, G. Increasing the precision of comparative models with YASARA NOVA–a self-parameterizing force field. *Proteins* **47,** 393–402 (2002).

211. Berendsen, H. J. C. *et al.* Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3684–3690 (1984).

212. Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24,** 1999–2012 (2003).

213. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103,** 8577–8593 (1995).

214. Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **23,** 1623–1641 (2002).

215. Teeter, M. M. Water structure of a hydrophobic protein at atomic resolution: Pentagon rings of water molecules in crystals of crambin. *Proc. Natl. Acad. Sci. U. S. A.* **81,** 6014–6018 (1984).

216. Némethy, G. & Scheraga, H. A. The structure of water and hydrophobic bonding in proteins. III. The thermodynamic properties of hydrophobic bonds in proteins. *J. Phys. Chem.* **66,** 1773–1789 (1962).

217. Kim, J. *et al.* Phosphopeptide elution times in reversed-phase liquid chromatography. *J. Chromatogr. A* **1172,** 9–18 (2007).

218. Blonder, J. *et al.* A detergent- and cyanogen bromide-free method for integral membrane proteomics: Application to halobacterium purple membranes and the human epidermal membrane proteome. *Proteomics* **4,** 31–45 (2004).

219. Krokhin, O. *et al.* An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC: Its application to protein peptide mapping by off-line HPLC-MALDI MS. *Mol. Cell. Proteomics* **3,** 908–919 (2004).

220. Mant, C. T., Zhou, N. E. & Hodges, R. S. Correlation of protein retention times in reversed-phase chromatography with polypeptide chain length and hydrophobicity. *J. Chromatogr. A* **476,** 363–375 (1989).

221. McCue, J. T. in *Guide to Protein Purification, 2nd Edition* (eds Burgess, R. R. & Deutscher, M. P.) 405–414 (Academic Press, 2009).

222. Haskard, C. A. & Li-Chan, E. C. Hydrophobicity of bovine serum albumin and ovalbumin determined using uncharged (PRODAN) and anionic (ANS-) fluorescent probes. *J. Agric. Food Chem.* **46,** 2671–2677 (1998).

223. Chao, C.-C., Ma, Y.-S. & Stadtman, E. R. Modification of protein surface hydrophobicity and methionine oxidation by oxidative systems. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 2969–2974 (1997).

224. Sangster, J. *Octanol-water partition coefficients: Fundamentals and physical chemistry* (John Wiley & Sons, 1997).

225. Fauchere, J. & Pliska, V. Hydrophobic parameters-pi of amino-acid side-chains from the partitioning of n-acetyl-amino-acid amides. *Eur. J. Med. Chem.* **18,** 369–375 (1983).

226. Konagurthu, A. S., Whisstock, J. C., Stuckey, P. J. & Lesk, A. M. MUSTANG: A multiple structural alignment algorithm. *Proteins: Struct., Funct., Bioinf.* **64,** 559–574 (2006).

227. Kuwajima, K. The molten globule state of alpha-lactalbumin. *FASEB J.* **10,** 102–109 (1996).

228. Wang, J. *et al.* Triclinic lysozyme at 0.65 Å resolution. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **63,** 1254–1268 (2007).

229. Durek, T., Torbeev, V. Y. & Kent, S. B. Convergent chemical synthesis and high-resolution X-ray structure of human lysozyme. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 4846–4851 (2007).

230. Gordon, J. C. *et al.* H++: A server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **33,** W368–W371 (2005).

231. Cromwell, M. E., Hilario, E. & Jacobson, F. Protein aggregation and bioprocessing. *AAPS J.* **8,** E572–E579 (2006).

232. Mahler, H.-C., Friess, W., Grauschopf, U. & Kiese, S. Protein aggregation: Pathways, induction factors and analysis. *J. Pharm. Sci.* **98,** 2909–2934 (2009).

233. Philo, J. S. & Arakawa, T. Mechanisms of protein aggregation. *Curr. Pharm. Biotechnol.* **10,** 348–351 (2009).

234. Chi, E. Y., Krishnan, S., Randolph, T. W. & Carpenter, J. F. Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation. *Pharm. Res.* **20,** 1325–1336 (2003).

235. Scopes, R. K. *Protein purification: Principles and practice* (Springer Science & Business Media, 2013).

236. Goto, Y. & Fink, A. L. Conformational states in. beta.-lactamase: Molten-globule states at acidic and alkaline pH with high salt. *Biochemistry* **28,** 945–952 (1989).

237. Wang, W., Li, N. & Speaker, S. External factors affecting protein aggregation. *Aggregation of therapeutic proteins,* 119–204 (2010).

238. Hofmeister, F. Zur Lehre von der Wirkung der Salze (about the science of the effect of salts). *Archiv für experimentelle Pathologie und Pharmakologie* **25,** 1–30 (1888).

239. Wu, S.-L. & Karger, B. L. in *High Resolution Separation and Analysis of Biological Macromolecules Part A: Fundamentals* (ed Barry L. Karger, W. S. H.) 27–47 (Academic Press, 1996).

240. Scopes, R. K. in *Protein Purif. Princ. Pract.* 3rd, 71–101 (Springer-Verlag New York, Inc., 1994).

241. Shukla, A. A. *et al.* Downstream processing of monoclonal antibodies - Application of platform approaches. *J. Chromatogr. B* **848.** Polyclonal and Monoclonal Antibody Production, Purification, Process and Product Analytics, 28–39 (2007).

242. Paul, R. *et al.* Structure and function of purified monoclonal antibody dimers induced by different stress conditions. *Pharm. Res.* **29,** 2047–2059 (2012).

243. Brunsteiner, M., Flock, M. & Nidetzky, B. Structure based descriptors for the estimation of colloidal interactions and protein aggregation propensities. *PloS one* **8** (2013).

244. Lauer, T. M. *et al.* Developability index: A rapid in silico tool for the screening of antibody aggregation propensity. *J. Pharm. Sci.* **101,** 102–115 (2012).

245. Valerio, M. *et al.* Early events in protein aggregation: Molecular flexibility and hydrophobicity/charge interaction in amyloid peptides as studied by molecular dynamics simulations. *Proteins: Struct., Funct., Bioinf.* **58,** 110–118 (2005).

246. Sousa, R. Use of glycerol, polyols and other protein structure stabilizing agents in protein crystallization. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **51,** 271–277 (1995).

247. Yadav, S. *et al.* The influence of charge distribution on self-association and viscosity behavior of monoclonal antibody solutions. *Mol. Pharm.* **9,** 791–802 (2012).

248. Yadav, S., Shire, S. J. & Kalonia, D. S. Viscosity analysis of high concentration bovine serum albumin aqueous solutions. *Pharm. Res.* **28,** 1973–1983 (2011).

249. Kumar, V., Dixit, N., Zhou, L. L. & Fraunhofer, W. Impact of short range hydrophobic interactions and long range electrostatic forces on the aggregation kinetics of a monoclonal antibody and a dual-variable domain immunoglobulin at low and high concentrations. *Int. J. Pharm.* **421,** 82–93 (2011).

250. Zhang, J. & Topp, E. M. Protein G, protein A and protein A-derived peptides inhibit the agitation induced aggregation of IgG. *Mol. Pharm.* **9,** 622–628 (2012).

251. Smith, R. D. *et al.* New developments in biochemical mass spectrometry: Electrospray ionization. *Anal. Chem.* **62,** 882–899 (1990).

252. Robbins, F. M., Kronman, M. J. & Andreotti, R. E. Inter- and intramolecular interactions of α-lactalbumin V. The effect of amidination on association and aggregation. *Biochim. Biophys. Acta, Biophys.* **109,** 223–233 (1965).

253. Lüllig, H. *et al.* Isolation and characterization of human pneumocytes. *Prax. Klin. Pneumol.* **37,** 842 (1983).

254. Lundblad, G. & Lind, J. Studies on lysozyme from human leucemic urine by isoelectric focusing. *Acta Chem. Scand* **26,** 1771–1713 (1972).

255. Righetti, P. G. & Caravaggio, T. Isoelectric points and molecular weights of proteins: A table. *J. Chromatogr. A* **127,** 1–28 (1976).

256. Righetti, P. G., Tudor, G. & Ek, K. Isoelectric points and molecular weights of proteins: A new table. *J. Chromatogr. A* **220,** 115–194 (1981).

257. Baumgartner, K. *et al.* Determination of protein phase diagrams by microbatch experiments: Exploring the influence of precipitants and pH. *Int. J. Pharm.* **479,** 28–40 (2015).

258. Hunter, R. Zeta potential in colloid science academic. *New York,* 69 (1981).

259. Haynes, W. M. *CRC handbook of chemistry and physics* (CRC press, 2014).

260. Tans, A. A new type of nomogram. Aqueous ammonium sulfate solutions. *Ind. Eng. Chem.* **50,** 971–972 (1958).

261. VDI Gesellschaft and VDI-Gesellschaft Verfahrenstechnik und Chemieingenieurwesen. *VDI heat atlas* (Springer Science & Business Media, 2010).

262. Martínez, J. M. & Martínez, L. Packing optimization for automated generation of complex system's initial configurations for molecular dynamics and docking. *J. Comput. Chem.* **24,** 819–825 (2003).

263. Martínez, L., Andrade, R., Birgin, E. G. & Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **30,** 2157–2164 (2009).

264. Greenwood, R. Review of the measurement of zeta potentials in concentrated aqueous suspensions using electroacoustics. *Adv. Colloid Interface Sci.* **106,** 55–81 (2003).

265. Ziegler, G. R. & Foegeding, E. A. The gelation of proteins. *Adv. Food Nutr. Res.* **34,** 203–298 (1990).

266. Boström, M. *et al.* Why forces between proteins follow different Hofmeister series for pH above and below pI. *Biophys. Chem.* **117,** 217–224 (2005).

267. Zeelen, J. P. in *Protein Cryst.* (ed Bergfors, T. M.) 2nd, 175–194 (Internat'l University Line, 2009).

268. Wiendahl, M., Oelmeier, S. A., Dismer, F. & Hubbuch, J. High-throughput screening-based selection and scale-up of aqueous two-phase systems for pDNA purification. *J. Sep. Sci.* **35,** 3197–3207 (2012).

269. Treier, K. *et al.* High-throughput methods for miniaturization and automation of monoclonal antibody purification processes. *Biotechnol. Prog.* **28,** 723–732 (2012).

270. Hansen, S. K., Skibsted, E., Staby, A. & Hubbuch, J. A label-free methodology for selective protein quantification by means of absorption measurements. *Biotechnol. Bioeng.* **108,** 2661–2669 (2011).

271. Diederich, P. *et al.* A sub-two minutes method for monoclonal antibody-aggregate quantification using parallel interlaced size exclusion high performance liquid chromatography. *J. Chromatogr. A* **1218,** 9010–9018 (2011).

272. Hatti-Kaul, R. Aqueous two-phase systems. English. *Mol. Biotechnol.* **19,** 269–277 (2001).

273. Valavi, M., Shirazian, S., Pour, A. & Ziary, M. Calculation of the density and activity of water in ATPS systems for separation of biomolecules. English. *J. Solution Chem.* **42,** 1423–1437 (2013).

274. Valavi, M., Dehghani, M. R. & Feyzi, F. Calculation of liquid-liquid equilibrium in polymer electrolyte solutions using PHSC-electrolyte equation of state. *Fluid Phase Equilib.* **341,** 96–104 (2013).

275. Azevedo, A. M. *et al.* Downstream processing of human antibodies integrating an extraction capture step and cation exchange chromatography. *J. Chromatogr. B* **877,** 50–58 (2009).

276. Ferreira, L. *et al.* Structural features important for differences in protein partitioning in aqueous dextran-polyethylene glycol two-phase systems of different ionic compositions. *Biochim. Biophys. Acta, Proteins Proteomics* **1844,** 694–704 (2014).

277. De Barros, D. P. *et al.* Modeling the partitioning of amino acids in aqueous two phase systems. *J. Chromatogr. A* **1329,** 52–60 (2014).

278. Bensch, M., Selbach, B. & Hubbuch, J. High throughput screening techniques in downstream processing: Preparation, characterization and optimization of aqueous two-phase systems. *Chem. Eng. Sci.* **62,** 2011–2021 (2007).

279. Merchuk, J. C., Andrews, B. A. & Asenjo, J. A. Aqueous two-phase systems for protein separation: Studies on phase inversion. *J. Chromatogr. B: Biomed. Sci. Appl.* **711,** 285–293 (1998).

280. Nascimento, K. S. *et al.* Liquid-liquid equilibrium data for aqueous two-phase systems composed of ethylene oxide propylene oxide copolymers. *J. Chem. Eng. Data* **56,** 190–194 (2011).

281. Martins, J. P. *et al.* Liquid-liquid equilibria of an aqueous two-phase system containing poly(ethylene) glycol 1500 and sulfate salts at different temperatures. *J. Chem. Eng. Data* **53,** 238–241 (2008).

282. Salieb-Beugelaar, G. B. *et al.* Latest developments in microfluidic cell biology and analysis systems. *Anal. Chem.* **82,** 4848–4864 (2010).

283. Arora, A. *et al.* Latest developments in micro total analysis systems. *Anal. Chem.* **82.** PMID: 20462185, 4830–4847 (2010).

284. Weigl, B. H., Bardell, R. L. & Cabrera, C. R. Lab-on-a-chip for drug development. *Adv. Drug Delivery Rev.* **55.** Biomedical Micro- and Nano-technology, 349–377 (2003).

285. Neuži, P. *et al.* Revisiting lab-on-a-chip technology for drug discovery. *Nat. Rev. Drug Discov.* **11,** 620–632 (2012).

286. Waldbaur, A. *et al.* Microfluidics on liquid handling stations (µF-on-LHS): An industry compatible chip interface between microfluidics and automated liquid handling stations. *Lab. Chip* **13,** 2337–2343 (12 2013).

287. Sparks, D. *et al.* Measurement of density and chemical concentration using a microfluidic chip. *Lab. Chip* **3,** 19–21 (1 2003).

288. Sparks, D., Smith, R., Massoud-Ansari, S. & Najafi, N. *Coriolis mass flow, density and temperature sensing with a single vacuum sealed MEMS chip* in *Solid-State Sensor, Actuator and Microsystems Workshop, Hilton Head Island, South Carolina* (2004), 75–78.

289. Azevedo, A. M. *et al.* Partitioning of human antibodies in polyethylene glycol-sodium citrate aqueous two-phase systems. *Sep. Purif. Technol.* **65,** 14–21 (2009).

290. Foroutan, M., Heidari, N., Mohammadlou, M. & Sojahrood, A. J. (Surfactant + polymer) interaction parameter studied by (liquid + liquid) equilibrium data of quaternary aqueous solution containing surfactant, polymer, and salt. *J. Chem. Thermodyn. Thermochem.* **41,** 227–231 (2009).

291. Tubío, G., Pellegrini, L., Nerli, B. B. & Picó, G. A. Liquid-liquid equilibria of aqueous two-phase systems containing poly(ethylene glycols) of different molecular weight and sodium citrate. *J. Chem. Eng. Data* **51,** 209–212 (2006).

292. Raja, S. *et al.* Aqueous two phase systems for the recovery of biomolecules–a review. *Sci. Technol.* **1,** 7–16 (2011).

293. Simon, L. & Gautam, S. Modeling continuous aqueous two-phase systems for control purposes. *J. Chromatogr. A* **1043,** 135–147 (2004).

294. Madeira, P. P. *et al.* Correlations between distribution coefficients of various biomolecules in different polymer/polymer aqueous two-phase systems. *Fluid Phase Equilib.* **267,** 150–157 (2008).

295. Rosa, P., Ferreira, I., Azevedo, A. & Aires-Barros, M. Aqueous two-phase systems: A viable platform in the manufacturing of biopharmaceuticals. *J. Chromatogr. A* **1217,** 2296–2305 (2010).

296. Rosa, P. A., Azevedo, A. M. & Aires-Barros, M. R. Application of central composite design to the optimisation of aqueous two-phase extraction of human antibodies. *J. Chromatogr. A* **1141,** 50–60 (2007).

297. Azevedo, A. M., Rosa, P. A., Ferreira, I. F. & Aires-Barros, M. R. Optimisation of aqueous two-phase extraction of human antibodies. *J. Biotechnol.* **132,** 209–217 (2007).

298. Wiendahl, M. *High throughput screening in downstream processing of biotechnological products* (Cuvillier, 2008).

299. Steinbrecher, T., Case, D. A. & Labahn, A. Free energy calculations on the binding of novel thiolactomycin derivatives to E. coli fatty acid synthase I. *Bioorg. Med. Chem.* **20,** 3446–3453 (2012).

300. Genheden, S., Nilsson, I. & Ryde, U. Binding affinities of factor Xa inhibitors estimated by thermodynamic integration and MM/GBSA. *J. Chem. Inf. Model.* **51,** 947–958 (2011).

301. Steiner, D. *et al.* Calculation of binding free energies of inhibitors to plasmepsin II. *J. Comput. Chem.* **32,** 1801–1812 (2011).

302. Åqvist, J., Medina, C. & Samuelsson, J.-E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **7,** 385–391 (1994).

303. Åqvist, J. Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions. *J. Comput. Chem.* **17,** 1587–1597 (1996).

304. Van Lipzig, M. M. H. *et al.* Prediction of ligand binding affinity and orientation of xenoestrogens to the estrogen receptor by molecular dynamics simulations and the linear interaction energy method. *J. Med. Chem.* **47,** 1018–1030 (2004).

305. Terai, T. *et al.* Rational development of caged-biotin protein-labeling agents and some applications in live cells. *Chem. Biol.* **18,** 1261–1272 (2011).

306. Rawas, A., Muirhead, H. & Williams, J. Structure of diferric duck ovotransferrin at 2.35 Å resolution. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **52,** 631–640 (1996).

307. Oelmeier, S. A., Dismer, F. & Hubbuch, J. Molecular dynamics simulations on aqueous two-phase systems - Single PEG-molecules in solution. *BMC Biophys.* **5,** 14 (2012).

308. Åqvist, J. & Marelius, J. The linear interaction energy method for predicting ligand binding free energies. *Comb. Chem. High Throughput Screen.* **4,** 613–626 (2001).

309. Marelius, J., Hansson, T. & Åqvist, J. Calculation of ligand binding free energies from molecular dynamics simulations. *Int. J. Quantum Chem.* **69,** 77–88 (1998).

310. Cohn, E. J. The physical chemistry of the proteins. *Physiol. Rev.* **5,** 349–437 (1925).