Predictive Tools in Protein Chromatography: Extension and Optimization of a Molecular Modeling Approach

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. Katharina Maria Helena Lang
aus Berlin

Referent:	Prof. Dr. Jürgen Hubbuch
Korreferent:	Prof. Dr. Matthias Franzreb
Tag der mündlichen Prüfung:	17. Juli 2015

Für Paul

[. . .] it ain't about how hard you hit.
It's about how hard you can get hit, and keep moving forward.
How much you can take, and keep moving forward.

Rocky Balboa

Danksagung

Diese Ehre hätte mir nicht zuteilwerden können, wenn mich nicht eine Vielzahl von Menschen unterstützt hätten. Daher möchte ich diese Gelegenheit nutzen, um meinen Dank auszudrücken.

Ich danke meinem Doktorvater Herr Professor Jürgen Hubbuch für die Betreuung und Unterstützung. Außerdem möchte ich Herrn Professor Matthias Franzreb für die Übernahme des Korreferats danken.

Ich danke meinen Kollegen und Studenten, die mir viele Aufgaben abnahmen und mich auch in angespannter und gestresster Laune ertrugen. Insbesondere möchte ich „meiner“ -129 danken. Natalie und Lara, ihr habt die schönsten Zeiten maßgeblich geprägt. Ihr werdet mir fehlen.

Sven, vielen lieben Dank für alles was Du für mich getan hast. Du bist mir in der Zeit ein sehr guter Freund geworden und hast mir immer geholfen, egal wie viel Du selber zu tun hattest. Das werde ich Dir nie vergessen.

Marion, vielen Dank für Deine wundervolle, offene und herzliche Art! Du hattest immer ein offenes Ohr und hast mir damit oft geholfen. Bitte bleib wie Du bist.

Michèle und Katha, vielen Dank für die gemütlichen Abende!

Ben, Sigrid und Patrick möchte ich für die nette Zeit danken.

Mein Dank geht auch an meine Betreuer Florian, Katrin und Anna, von denen ich so viel gelernt habe - fachlich und persönlich.

Außerdem danke ich Jörg, der mir stets Mut zugesprochen und mich bestärkt hat. Hättest Du mich nicht unterstützt, wäre meine Arbeit nicht möglich gewesen. Mit allem Pipapo.

Und nicht zuletzt danke ich meinen Eltern, die in jeglicher Hinsicht die Grundsteine für meinen Weg gelegt haben.

Contents

Zusammenfassung

Die vorliegende Arbeit „Predictive Tools in Protein Chromatography: Extension and Optimization of a Molecular Modeling Approach“befasst sich mit der Erweiterung und Optimierung eines prädiktiven Werkzeugs in der Proteinchromatographie. Sie wurde am Institut für Bio- und Lebensmitteltechnik, im Bereich IV: Molekulare Aufarbeitung von Bioprodukten unter der Leitung von Prof. Dr. Jürgen Hubbuch angefertigt.

Ziel der Arbeit war ein besseres Verständnis des Bindungsverhaltens von Biomolekülen auf verschiedenen Adsorbern zu schaffen. Dieses Verständnis soll zur Vorhersage und Problemanalyse genutzt werden können. Die Herangehensweise war die Entwicklung eines Moleküldynamiksimulations-Tools, welches sowohl das Biomolekül, als auch die Ligandenoberfläche auf atomarer Ebene realistisch darstellt und Interaktionsvorgänge zwischen beiden abbilden kann. Simulationszeiten sollten ein Kompromiss zwischen Genauigkeit und Geschwindigkeit sein. Hierbei wurde mit verschiedenen Software-Programmen gearbeitet und die meisten Simulationen liefen auf dem Hochleistungscomputer des Karlsruher Instituts für Technologie. Die Simulationsergebnisse wurden mittels entsprechenden Laborexperimenten validiert.

Das Tool wurde nach der Entwicklung auf drei verschiedene Fragestellungen angewendet: In der ersten Studie wurde die Vorhersagefähigkeit von Retentionszeiten mehrerer Proteine auf einem Anionenaustauscher und einem Kationenaustauscher geprüft und mit experimentellen Ergebnissen abgeglichen. Das Retentionsverhalten wurde entsprechend den experimentellen Ergebnissen vorhergesagt. Außerdem konnten die bindungsbeeinflussenden Aminosäuren identifiziert werden. Auf dem Kationentaustauscher handelt es sich dabei größtenteils um Arginine, während auf dem Anionenaustauscher Asparaginsäure die treibende Kraft war. Dies deckt sich mit früheren experimentellen Ergebnissen aus der Literatur. Damit hat sich das Tool als fähig erwiesen, die Protein-Ionenaustauscher-Interaktionen darzustellen und Retentionsverhalten vorherzusagen. Außerdem konnte ein Einblick in das Bindungsverhalten auf molekularer Ebene gewonnen werden.

Ein weiterer wichtiger Einflussfaktor auf das Bindungsverhalten sind die Charakteristika des Adsorbers. Um das Potenzial des entwickelten Tools zu ermitteln, wurde die wahrscheinlich wichtigste Einflussgröße, die ionische Kapazität, untersucht. Die entsprechende Simulationsgröße ist die Ligandendichte auf dem Adsorber, die variiert wurde. Analog wurden die entsprechenden Versuche im Labor durchgeführt. Die Herausforderung war die Aufreinigung eines Proteingemischs, bestehend aus 3 Modellproteinen auf Adsorbern des gleichen Typs mit verschiedenen ionischen Kapazitäten. Die Simulationen ergaben einen starken Einfluss der Ligandendichte auf das Retentionsverhalten, was auch in Experimenten beobachtet wurde. Die Simulationen haben das unterschiedliche Verhalten auf unterschiedlichen Adsorbern beschreiben können und eignen sich daher zum Adsorberdesign welches spezifisch für ein Aufreinigungsproblem ausgelegt ist. Das Tool bereitet den Weg für die schnelle Entwicklung von maßgeschneiderten Adsorbern.

In der letzten Studie wurde eine Fallstudie mit speziellen Antikörpern, sogenannten Nanobodies, durchgeführt, bei denen unterschiedliche Prozessschritte untersucht wurden: Zum einen ein Kationionenaustauscherschritt in der Aufreinigung und zum anderen die Couplingreaktion in der anschließenden Immobilisierung. Für die Simulationen wurden die Proteinstrukturen von einem Industriepartner zur Verfügung gestellt. Entsprechende Experimente wurden beim Industriepartner durchgeführt. Die Simulationen mit dem Affinitätscouplingmedium ergab ein besseres Verständnis von potentiellen Couplingstellen unter der Annahme, dass energetisch favorisierte Orientierungen wahrscheinlicher die Couplingreaktion eingehen werden. Aus den Kationenaustauschersimulationen konnten die "Hot-Spots"der Bindung identifiziert und die Elutionsreihenfolge vorhergesagt werden. Dieses Ergebnis zeigt einen weiteren Vorteil der *in silico* basierten Untersuchung auf: Strukturvarianten beziehungsweise deren Aufreinigung können vorerst auf molekularer Ebene untersucht werden. Ist die Ursache identifiziert, beispielsweise eine besonders stark bindende Aminosäure, können verschiedene Szenarien am Rechner untersucht werden bevor aufwändige Experimente durchgeführt werden.

Die Entwicklung des Simulationstools konnte erfolgreich abgeschlossen werden. Die Funktionalität wurde in mehreren Untersuchungen nachgewiesen.

Abstract

This thesis titled “Predictive tools in protein chromatography: Extension and Optimization of a Molecular Modeling Approach” deals with the extension and optimization of a predictive tool in protein chromatography. It was carried out at the Institute of Process Engineering in Life Sciences, in Section IV: Biomolecular Separation Engineering under the supervision of Prof. Dr. Jürgen Hubbuch.

The aim of this study was to increase the understanding of the binding behavior of biomolecules on different adsorbents. This understanding ought to be used for prediction and problem analysis. The approach was to develop a molecular dynamics simulations tool, which represents both: the biomolecule and the ligand surface at the atomic level and realistically maps the interaction processes between them. Simulation times should be a compromise between accuracy and speed. This was achieved with various software programs and most simulations ran on the high-performance computing cluster at Karlsruhe Institute of Technology. Simulation results were validated by corresponding laboratory experiments.

After development, the tool was applied on three different issues: In the first study, the predictive capability of retention times of several proteins was tested on an anion and a cation exchanger and compared with experimental results. The retention behavior was predicted correctly. In addition, the binding influencing residues could be identified. On the cation exchanger these are mostly arginine, while on the anion exchanger aspartic acid was the driving force. This is consistent with previous experimental results from literature. The tool has proven to be able to show the protein ion exchange interactions and predict retention behavior. Furthermore, insight on the binding behavior at the molecular level was gained.

Other important factors influencing the binding behavior are the characteristics of the adsorbent. To determine the potential of the developed tools, the ionic strength, which is probably the most important factor, was investigated. The corresponding simulation parameter is the ligand density on the adsorber, which was varied. Analogously, corresponding experiments were run in the laboratory. The challenge was the purification of a protein mixture consisting

of 3 model proteins on adsorbers of the same type with different ionic capacities. The simulations showed a strong influence of ligand density on the retention behavior, which was also observed in experiments. The simulations were able to describe the different behavior on different adsorbents and are therefore suitable for adsorber design, which is specific for a purification problem. This tool is paving the way for the design of adsorbers for specific purification problems.

A wide scale application of the developed tool was shown in a case study with special antibodies called nanobodies, in which different process steps were investigated: once in the purification of a cation exchanger step and in the subsequent immobilization of the coupling reaction. For the simulations, the protein structures were provided by an industrial partner. The simulations with the affinity coupling medium resulted in a better understanding of potential coupling sites under the assumption that energetically favored orientations are more likely to undergo the coupling reaction. From the cation exchanger simulations the binding "hot spots" were identified and elution could be predicted. This result indicates a further advantage of the *in silico* analysis: problematic biomolecules or their purification can initially be examined at the molecular level. Is the culprit identified, for example a particularly strong binding residue, different scenarios could be investigated exclusively on the computer before expensive experiments are conducted.

The development of the simulation tool was completed successfully. The functionality was demonstrated in several studies.

Chapter 1

Introduction

1.1. Proteins: Structures, Properties, and Functions

Proteins play a major role in all life processes since they translate the information contained in DNA into cellular processes and structures. They regulate important functions such as gene expression and metabolism, serve as transporter for small molecules, and act as receptors or messengers. Structural proteins make up the framework of cells and stabilize entire organisms. Also, proteins support the muscle contraction and movement of cells [1]. One of the main tasks of proteins is to speed up biochemical reactions or run them: Enzymes act as biocatalysts that engage in a reaction [2–4].

Despite this wide variety of tasks, all proteins are built on a simple principle. They consist of up to 22 different units, the proteinogenic amino acids. The individual amino acids are arranged like beads in a row and linked by a peptide bond (Figure 1.1) between the carboxyl group of the first and the amino group of the next amino acid.

Proteins of the immune system, antibodies, are soluble glycoproteins in blood or extracellular fluid of the tissues. They are built by B cells and recognize specific epitopes. Through the bonding of an antibody to its specific antigen, it is marked for phagocytosis by the scavenger cells; simultaneously, an immune response is induced, leading to the formation of more antibodies of the same type. Antibodies can be divided into five classes (IgG, IgA, IgM, IgD, IgE), which differ in structural features.

1.1.1. Protein Structure

The building blocks in a protein are the amino acids, which are covalently linked in a long polypeptide chain (compare Figure 1.1). The protein structure divides into four levels: Primary, secondary, tertiary, and quaternary structure.

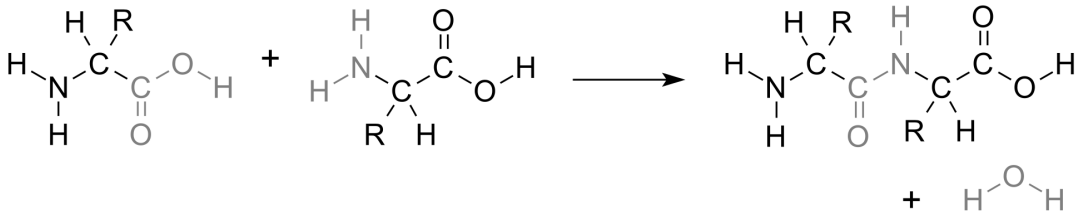


Figure 1.1: Peptide bond formation between two amino acids

The primary structure refers to a sequence of the 22 proteinogenic amino acids.

The secondary structure describes the spatial arrangement of adjacent amino acids. The formation of hydrogen bonds between the carbonyl oxygen and the nitrogen of the amino group produces two secondary structures: the α-helix and β-sheet.

The tertiary structure refers to the spatial arrangement of the entire polypeptide chain. Helical sections may alternate with β-sheet structures. Also, sections in which the polypeptide chain runs in loops occur.

Some proteins are made up of several polypeptide chains. The quaternary structure indicates, which and how many of these subunits make up a protein, as these are oriented spatially to one another and whether they are linked covalently or via hydrogen bridges.

1.1.2. PHYSICOCHEMICAL PROPERTIES OF BIOMOLECULES

Biomolecules are proteins, carbohydrates, lipids, and nucleic acids. These biomolecules differ in various physicochemical characteristics:

- molecular weight
- size and shape
- net charge
- hydrophobicity

The molecular weight of proteins ranges from 2 (Trp-Cage) to 3.000 kDa (Titin, a muscle protein) [5]. The most relevant proteins in biopharmaceutical production range from insulin to antibodies, in terms of size. Another relevant property is the charge of proteins. Amino acids, which form the protein can be positive, negative, or neutral in charge and together form a protein's total charge. Charge differences among proteins are considerable, which is a strong parameter for use in separation processes. Hydrophobic amino acids, such as alanine,

isoleucine, leucine, methionine, phenylalanine, tryptophan, and valine cluster in the core of the protein and therewith not only influence the protein structure but also result in different hydrophobicity values for different proteins.

All these properties are used in separation processes to separate the target molecule from the contaminants, which ideally have completely different characteristics.

There is the possibility to change the properties of proteins in terms of molecular weight, size, shape, charge, and hydrophobicity in some degree by PEGylation. This also extends the possibilities for separation. In PEGylation, polyethylene glycol (PEG) polymer chains are covalently bound to the protein. Through this change in characteristics, the purification of the biomolecule can be eased.

CHARGE

The amino acids of proteins have different pK-values that describe the pH-value at which the titratable groups are protonated and deprotonated in equal shares. The pK-value depends on various factors: immediate surrounding amino acids, ionic strength of the surrounding medium, and temperature, among others. The charge distribution and net charge of the protein depends on the protonation and deprotonation of the amino acids (Figure 1.2).

The isoelectric point (pI) is the pH of an aqueous solution, in which case amino acids and proteins balance positive and negative charges so that the net charge equals zero.

In practice, the pI can be used for precipitation from solution. The solubility of a protein is strongly influenced by the surrounding pH and reaches a minimum at the isoelectric point. Above or below the pI, all molecules have the same charge (positive or negative) and therefore repel each other. An aggregation into insoluble aggregates is prevented by the repulsion of the molecules and the protein remains in solution.

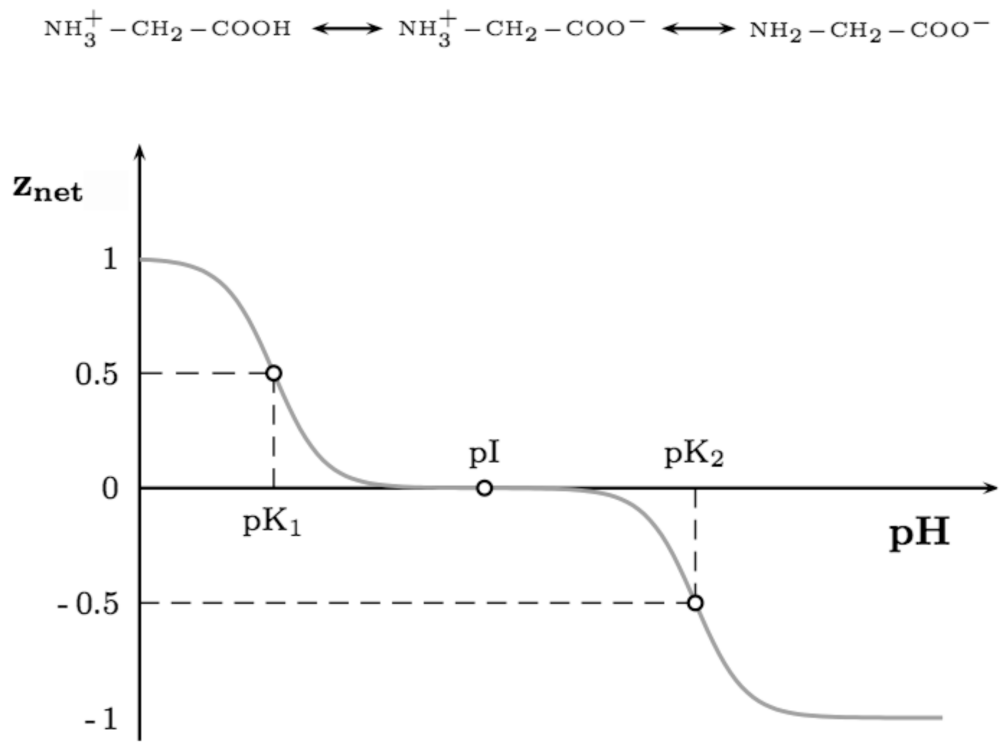


Figure 1.2: Coherence of the average net charge z_{net} and pH [6]

1.2. INDUSTRIAL APPLICATIONS AND SIGNIFICANCE OF BIOMOLECULES

Biomolecules are valuable tools in biotechnical and pharmaceutical applications: Enzymes are used in food manufacturing (e.g., cheese), enzymology, and genetic engineering. Proteins, such as insulin, are used as drug substances against multiple diseases, such as diabetes. Monoclonal antibodies are used as a therapeutic agent for a wide range of purposes, e.g., as vaccines. Also, they are used in biotechnology and pharmacology for analysis and purification of therapeutics due to their high specificity towards a single antigen.

The separation of proteins and especially antibodies are of most interest in industrial downstream processing. Proteins operate as: catalysts in terms of enzymes, antibodies for specific binding to pathogenic cells, signaling hormones such as insulin for the regulation of the sugar level in the blood, among others.

1.3. CHROMATOGRAPHY

Chromatography is a technique for the separation of mixtures. A distinction is made depending on the layout of the stationary phase: In planar chromatography, the stationary phase is in or on a plane, e.g., paper. In column chromatography, the stationary phase is located in a cylindrical tube. The packing material may fill the entire cavity of the tube (packed column), or is applied merely as a thin layer on the inner surface (capillary column).

1.3.1. A BRIEF HISTORY

The initial mentioning of the word "chromatography" was by Mikhail Semenovich Tswett, who first presented his work on the chromatographic technique at the meeting of the Biological Section of the Warsaw Society of Natural Sciences in 1903 [7]. (The conference proceedings were published in 1905 [8].) Tswett demonstrated the technique of chromatographic separation of chlorophyll in his twin papers in 1906 [9, 10]. Since Tswett investigated over one hundred inorganic and organic adsorbents, he derived a general law of adsorption, which later proved to be universally valid [11].

Not until 35 years later, in 1941, Martin and Synge [12] presented the first chromatography technique on partition chromatography. Ten years later, the first gas chromatograph was demonstrated by James and Martin [13] on the 290th Meeting of the Biochemical Society in London. Stahl and Vollmann [14] established thin-layer chromatography with their publication in 1965. Since the 1970s, a rapid development of multiple other techniques proceeded, which continues to this day.

1.3.2. COLUMN CHROMATOGRAPHY

The basic structure divides into a mobile phase, which is a fluid (gas or liquid) and an immobile stationary phase. The fluid flows through the column containing the stationary phase and the single components of the (fluid) mixture interact differently with the immobile phase depending on their physicochemical properties. Usually agarose beads or finely powdered silica gel is used as stationary phase. The components elute according to their strength of interaction. Separation can be accomplished by two mechanisms: either the biomolecule is withheld because it is retained spatially (size exclusion chromatography) or the biomolecule interacts through various mechanisms, such as electrostatics (ion exchange chromatography), hydrophobicity (hydrophobic interaction chromatography), or affinity (affinity chromatography). Elution of the biomolecules is accomplished by weakening the interaction

strength. This mode is called bind-elute. Another approach is the flow-through mode, in which the biomolecule of interest does not interact with the immobile phase, but the contaminants do.

1.3.3. CHROMATOGRAPHIC PROCESS

In liquid chromatography, the stationary bed is made up of multiple (small) particles packed tightly into a column. The liquid phase flows through the packed column (mostly under pressure = positive flow). Since the liquid phase contains a mixture of components (see Figure 1.3), the interaction of the several components with the stationary phase depend on the respective affinity of the component particles to the adsorbent particle. A higher affinity to the stationary phase leads to slower movement through the column, and therefore to later arrival at the column end. This delayed arrival results in a separation, when the eluent is collected in fractions. This process is pictured in Figure 1.3 where component B has a higher affinity towards the adsorbent and therefore elutes later from the column.

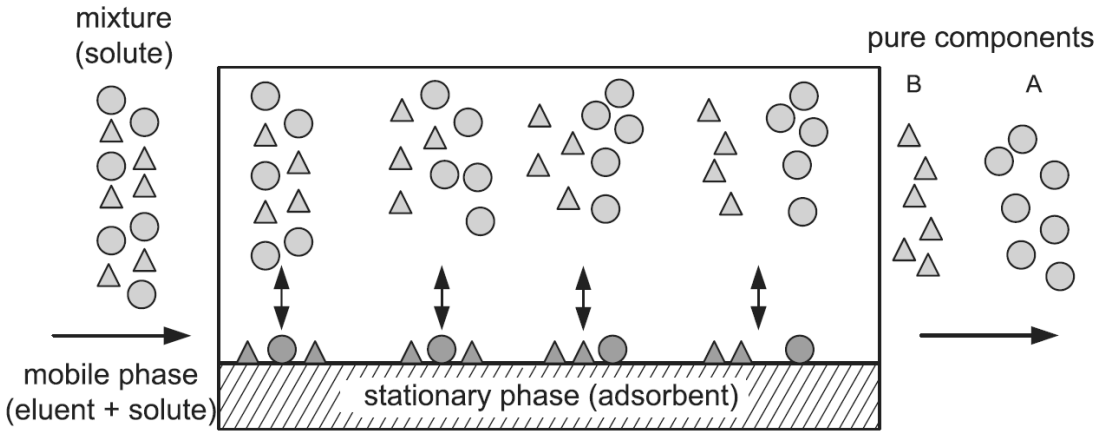


Figure 1.3: Adsorption chromatography from H. Schmidt-Traub: *Preparative Chromatography of Fine Chemicals and Pharmaceutical Agents*. p. 12. 2005. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

1.3.4. MASS TRANSFER MECHANISMS

The functional principle of chromatography is based on the mechanism of mass transfer of the molecules within the bulk phase and inside the adsorbent. The mass transfer phenomena can be classified in four categories, which are depicted in Figure 1.4. The molecular movement within the bulk phase is called convection and diffusion (1). Around the adsorbent par-

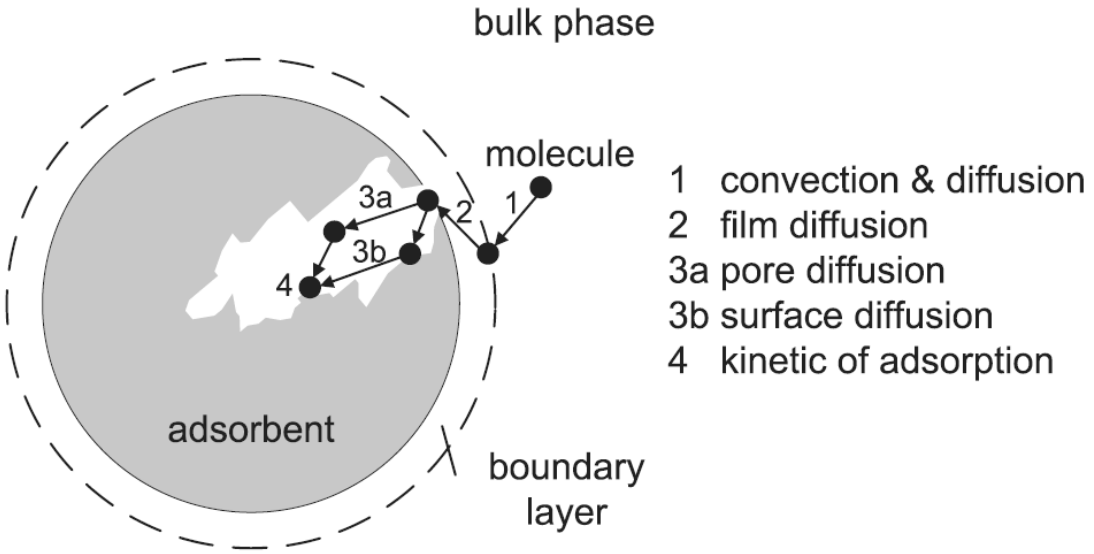


Figure 1.4: Mass transfer mechanisms from H. Schmidt-Traub: *Preparative Chromatography of Fine Chemicals and Pharmaceutical Agents*. p. 24. 2005. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

ticle is a thin boundary layer (film) that forms at the phase boundary between solid and liquid phase. The diffusive mass transport through this film is called film diffusion (2). Movement within the pore can take place by two different motions: the pore diffusion is the movement in the free spaces inside the pores (3a). The second transfer mechanism inside the pore is surface diffusion, which is defined as the movement along the pore surface (3b). The final step is the adsorption onto a free adsorption site of the adsorbent (4). Bottlenecks of the transport mechanism with regards to speed are the film diffusion (2) and transport inside the pores (3a, 3b), whereas convection in the bulk phase and adsorption are comparatively fast [15].

1.3.5. ADSORPTION

The adsorption process is based on the gathering of molecules on the immobile phase. The immobile phase in liquid chromatography is defined as the adsorbent, whereas the molecule is defined as the solute. The adsorption process is the arrangement of binding between the surface of the adsorbent and the solute. These binding forces may differ in their nature and therefore in their strength depending on surface and solute properties.

1.3.6. VISUALIZATION OF SEPARATION: CHROMATOGRAM

The chronological chromatography process can be visualized by detecting a protein signal at the column outlet via light absorption. The quantification of proteins based upon absorbance is a common method in biological science and clinical research. Proteins absorb ultraviolet (UV) light at 280 nm. This characteristic can be used as a basis for the photometric quantification. The recording over time can be represented visually as a chromatogram. An example of a chromatogram with three components is depicted in Figure 1.5. The peak represents the concentration of the component as a function of time. The retention time t_R is directly proportional to the affinity of the components to the adsorbent and is calculated from the peak maximum.

CHROMATOGRAM PARAMETERS

Retention Time $t_{R,i}$ The total retention time is the time that passes from the sample injection to reach the peak maximum of a component. The retention time of a compound is characteristic when the chromatographic conditions are kept constant.

Total Dead Time t_{total} The total dead time is the time required by a non-delayed component from sample injection until signal appearance in the detector. It demonstrates how long a substance remains in a chromatographic system, if it has no interactions with the stationary phase. In practice, the aim is to keep this time as small as possible since a lengthy stay results in back-mixing of the components and thus leads to peak broadening.

Net Retention Time $t_{R,i,net}$ The net retention time is the difference between the retention and total dead time. It represents the residence time of the components, which exclusively stay within the stationary phase (adsorbed).

Capacity Factor (Retention Factor) k' The capacity factor or retention factor k' is a measure for the rate of migration of a component in the chromatographic system. It is a more appropriate parameter than the net retention time, since it is independent from column length and flow rate.

$$k'_i = \frac{t_{R,i} - t_0}{t_0} \tag{1.1}$$

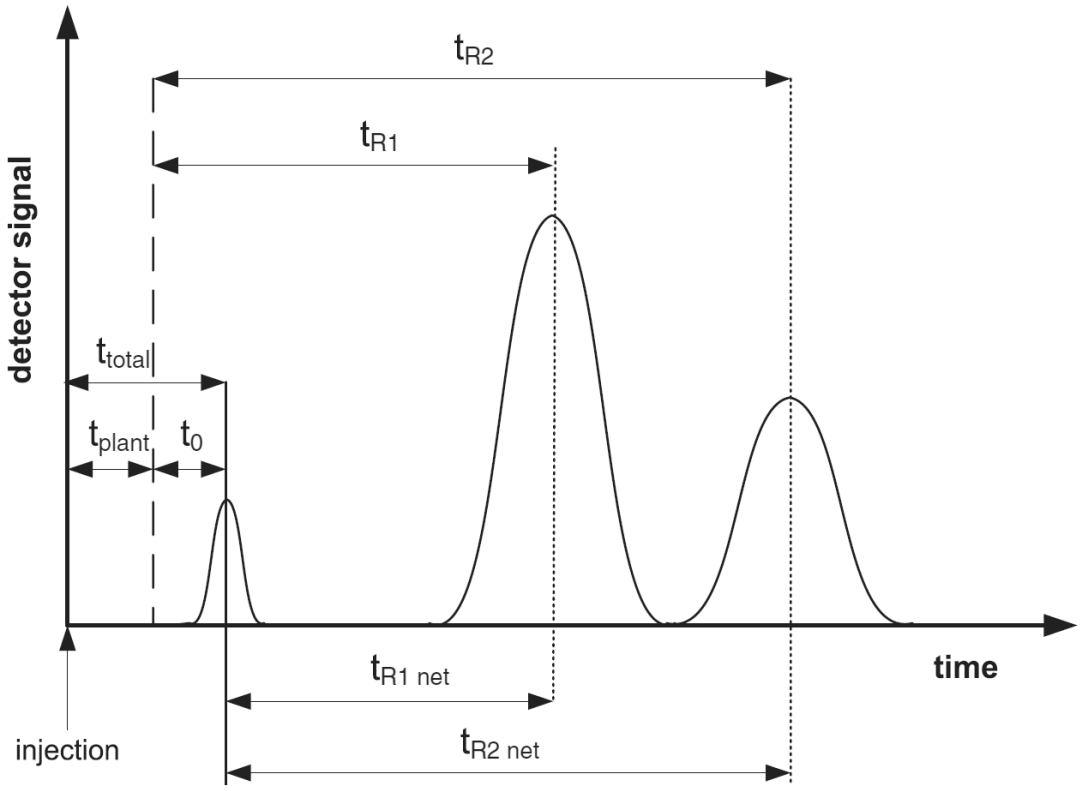


Figure 1.5: Chromatogram of three components from H. Schmidt-Traub: *Preparative Chromatography of Fine Chemicals and Pharmaceutical Agents.* p. 13. 2005. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

1.4. MODELING OF PROTEIN RETENTION

Chromatography is a widely used technique for the purification of biomolecules [16]. A profound understanding of the basic mechanism of retention behavior and therewith the opportunity to model this mechanism would be of great advantage. Models range from very basic charged surface approaches to highly detailed simulations of biomolecule and adsorbent.

1.4.1. MATHEMATICAL MODELS FOR ION-EXCHANGE CHROMATOGRAPHY

Ion exchange chromatography is one of, if not, the most widely used chromatography method due to its advantage of keeping proteins in their native state, since it is based on electrostatic interactions between protein and the charged stationary phase. Therefore, multiple mathematical models have been developed to describe and predict retention behavior. A distinc-

tion of different approaches can be made: some models are semi-empirical while others are knowledge-based.

SEMI-EMPIRICAL MODELS

The stoichiometric displacement model implies that the retention of a protein is related to the counter ion concentration under isocratic, linear conditions [17, 18]. An extension to linear gradient elution conditions was published by Yamamoto et al. [19] to describe protein retention in a gradient. In 1992, Brooks and Cramer [20] published the theory on a steric mass-action (SMA) model, that also takes the steric hindrance of counter ions into account. The SMA model can consider non-linear conditions in isocratic and gradient chromatography. Another approach is the available area model [21, 22] that uses the protein's radius to consider the geometric blocking on the adsorbent. This approach simplifies the protein shape to a symmetrical sphere, which is a simplification that was overcome by Ladiwala et al. [23] in 2005 who used the three-dimensional crystal structure of proteins and a mathematical algorithm to derive properties and behavior based on the structure of the solute. This type of modeling is called Quantitative Structure Property Relationship (QSPR), which initially was developed for small molecules. Ladiwala et al. [23] used the developed model to predict adsorption isotherm parameters. One year later, in 2006, Malmquist et al. [24] published a paper on QSPR of gradient elution conditions.

KNOWLEDGE-BASED MODELS

Ståhlberg et al. [25, 26] presented in two sequential papers a model based on Gibbs free energy calculations of two oppositely charged planar surfaces with a salt solution in-between. In their first publication, the authors took into account electrostatic interactions and in the second paper additionally van der Waals forces. On knowledge-based approaches there is a computational model that calculates electrostatic and van der Waals energies of interaction between a protein molecule and a planar, charged surface [27]. A more complex approach is the modeling of adsorbent and solute as discrete molecules. This can be realized by molecular dynamics simulations (Section 1.5.2): Noinville et al. [28] positioned a protein on a charged surface and used the AMBER force field parameters to derive interaction energy values from the system. Ravichandran et al. [29] simulate a protein on an array of charges on a planar surface at low and high salt concentrations. A more recent approach is the simulation of a protein on a ligand surface with charged active groups [30].

1.4.2. MATHEMATICAL MODELS FOR AFFINITY CHROMATOGRAPHY

Kavoosi et al. [31] presented a two-zone model (TZM), which describes solute binding to an affinity adsorber. The model is divided into two sections: an inner protein-free center and an outer zone where proteins bind. Kermani et al. [32] used support vector machine learning algorithms to model retention time of a peptide on a column with nickel based adsorbents based on the amino acid composition. Du et al. [33] also used a QSPR approach to predict retention behavior of peptides on a nickel column. The most computationally intense approach were molecular dynamics simulations of monoclonal antibodies on affinity ligands [34].

1.5. MOLECULAR MODELING

1.5.1. A BRIEF HISTORY

In the history of molecular modeling, flat, two-dimensional molecular structures were at the beginning, which then developed to precise calculations of molecular structure and intermolecular interactions on representations of the stereochemistry and 3D representations. Physical models allow only for very limited number of statements about a structure, which are mostly limited to size, outline, and stereochemistry. In addition, it can be difficult to create models for larger molecules. Computer models, however, can easily generate large structures and allow the calculation of conformational energies or the exact position of atoms in a molecule.

The first approaches in stereochemistry were made in the 19th century, when structural formulas were introduced to represent the conformation of chemical compounds. In 1874, van't Hoff discovered the tetrahedral structure of carbon [35]. In 1953, Barton introduced the conformation analysis. James Watson and Francis Crick presented the first 3D model of DNA [36]. For this purpose, they used X-ray diffraction data from Rosalind Franklin and Maurice Wilkins. In 1957, John Kendrew created the first 3D structure of a protein (myoglobin) with X-ray crystallography at a resolution of 6 Å [37]. John Kendrew and Max Perutz received the Nobel Prize in Chemistry for the 3D structure of myoglobin and hemoglobin in 1962. In 1959, the Dreiding stereomodel was developed [38]. In 1965, the Corey-Pauling-Koltun (CPK) space-filling model was developed [39, 40]. Starting in the 1970s, computer models were used.

```
SEQRES   1    129  LYS VAL PHE GLY ARG CYS GLU LEU ALA ALA ALA MET LYS
SEQRES   2    129  ARG HIS GLY LEU ASP ASN TYR ARG GLY TYR SER LEU GLY
SEQRES   3    129  ASN TRP VAL CYS ALA ALA LYS PHE GLU SER ASN PHE ASN
SEQRES   4    129  THR GLN ALA THR ASN ARG ASN THR ASP GLY SER THR ASP
SEQRES   5    129  TYR GLY ILE LEU GLN ILE ASN SER ARG TRP TRP CYS ASN
SEQRES   6    129  ASP GLY ARG THR PRO GLY SER ARG ASN LEU CYS ASN ILE
SEQRES   7    129  PRO CYS SER ALA LEU LEU SER SER ASP ILE THR ALA SER
SEQRES   8    129  VAL ASN CYS ALA LYS LYS ILE VAL SER ASP GLY ASN GLY
SEQRES   9    129  MET ASN ALA TRP VAL ALA TRP ARG ASN ARG CYS LYS GLY
SEQRES  10    129  THR ASP VAL GLN ALA TRP ILE ARG GLY CYS ARG LEU
SSBOND   1 CYS     24    CYS    145                                    2.03
SSBOND   3 CYS     48    CYS    133                                    2.04
SSBOND   5 CYS     82    CYS     98                                    2.03
SSBOND   7 CYS     94    CYS    112                                    2.04
ATOM      1  N   LYS    19      -0.270   5.015  15.269  0.09 66.00           N
ATOM      2 1H   LYS    19      -0.728   5.652  15.915  0.21 65.00           H
ATOM      3 2H   LYS    19       0.558   4.679  15.749  0.21 65.00           H
ATOM      4 3H   LYS    19      -0.881   4.224  15.105  0.21 65.00           H
ATOM      5  CA  LYS    19       0.114   5.687  14.002 -0.00 15.00           C
ATOM      6  HA  LYS    19       0.700   4.975  13.423  0.11 80.00           H
ATOM      7  CB  LYS    19      -1.146   6.053  13.183  0.02 12.00           C
ATOM      8 1HB  LYS    19      -1.597   5.130  12.819  0.02 83.00           H
ATOM      9 2HB  LYS    19      -1.860   6.550  13.839  0.02 83.00           H
```

Figure 1.6: Sample PDB file of Lysozyme. The ATOM entries contain information on every atom: atom number, atom name, residue, residue number, Cartesian coordinates, etc.

1.5.2. PROTEIN MODELING

3D COORDINATES

For simulations with peptides, the three-dimensional structure in Cartesian coordinates is mandatory. There are multiple methods to obtain this 3D structure. Structure determination is based on experimental methods like X-ray crystallography [41] or nuclear magnetic resonance (NMR) spectroscopy [42]. A merely theoretical approach is the structure prediction in bioinformatics, known as homology modeling. The structure information can then be saved as a textual file, which is in human-readable format (see Figure 1.6). The most common file format is protein data bank (PDB), which provides information on the atom coordinates, secondary structure sections and data about connectivity. To enable an open exchange between research groups, a databank system, the Protein Data Bank was established and is freely accessible on the internet [43]. Other, less widely used file formats are macromolecular Crystallographic Information File (mmCIF), protein structure file (PSF), and Kinemage file (KIN).

FORCE FIELDS

By calculating force fields in molecular mechanics, all bonds, angles, torsions and the contribution of the non-covalent interactions are calculated and their energies are summed up according to Hooke's law (Equation 1.2). The interactions between atoms that are not covalently bound, are described using Coulomb potential for electrostatic interactions and Lennard-Jones potential for van der Waals interactions. An example of a force field file is depicted in Figure 1.7.

First, each individual atom is classified in the molecule, since the atomic type is an important parameter for the calculation. In a second step, the potential energy terms for the individual atom bonds are calculated [44].

$$E_{total} = E_{bond} + E_{angle} + E_{dihedral} + E_{vdw} + E_{elec} \tag{1.2}$$

with the energy of the bond length

$$E_{bond} = \frac{1}{2} k_b (b - b_0)^2 \tag{1.3}$$

the energy of the bond angle

$$E_{angle} = \frac{1}{2} k_\theta (\theta - \theta_0)^2 \tag{1.4}$$

the energy of the torsion angles

$$E_{dihedral} = \frac{1}{2} k_\varphi \left(1 + \cos\left(n\varphi - \varphi_0\right)\right) \tag{1.5}$$

the energy of van der Waals interactions

$$E_{vdw} = \sum \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \tag{1.6}$$

and the energy of the electrostatic interactions

$$E_{elec} = \frac{1}{\varepsilon} \frac{Q_1 Q_2}{r} \tag{1.7}$$

```
# ATOM TYPES, MASSES AND POLARIZABILITIES
# =======================================
# <Atom symbol [2 letters]> <mass [g/mol]>
Br 79.90         2.880               bromine  (Applequist)
C  12.01         0.616  !            sp2 C carbonyl group
C* 12.01         0.360               sp2 arom. 5 memb.ring w/1 subst. (TRP)
C0 40.08                             calcium
CA 12.01         0.360               sp2 C pure aromatic (benzene)
CB 12.01         0.360               sp2 aromatic C, 5&6 membered ring junction
CC 12.01         0.360               sp2 aromatic C, 5 memb. ring HIS
CD 12.01         0.360               sp2 C atom in the middle of: C=CD-CD=C
CK 12.01         0.360               sp2 C 5 memb.ring in purines
CM 12.01         0.360               sp2 C  pyrimidines in pos. 5 & 6
CN 12.01         0.360               sp2 C aromatic 5&6 memb.ring junct.(TRP)
CQ 12.01         0.360               sp2 C in 5 mem.ring of purines between 2 N
CR 12.01         0.360               sp2 arom as CQ but in HIS
CT 12.01         0.878               sp3 aliphatic C
CU 63.55                             copper
```

Figure 1.7: Excerpt from the Amber03 force field. The force field file is in human readable format. The parameters are specific for each atom type and are derived from fits, which are experimentally determined.

The constants b, θ, φ, and R_{ij} describe the molecular structure: bond lengths, bond angles, dihedral angles and interatomic distances. The remaining terms are force field parameters [45]. Common force fields in molecular mechanics are AMBER [46, 47] and GROMOS [48].

SEQUENCE AND STRUCTURE ALIGNMENT

A sequence alignment bears the possibility to arrange sequences of proteins to identify regions of similarity that are a result of functional, structural, and evolutionary relationships between the sequences [49]. Gaps are inserted so that identical or similar amino acids are consecutive. There is pairwise and multiple sequence alignment. Pairwise alignment is used to find the best agreeing regions of two sequences. Multiple sequence alignment integrates multiple sequences to reveal similarities and differences of the sequences. The graphical representation is usually conducted in phylogenetic trees, that visualize the similarity between sequences and conclusions of evolutionary relationships can be drawn.

Structure alignment is the arrangement of two or more 3D structures based on their geometry and three-dimensional conformation to evaluate the significant homology between them.

ENERGY MINIMIZATION

Energy minimization is based on the assumption that the calculated structure of a molecule with the lowest potential energy corresponds to the native state.

Therefore, an optimization of a molecular model can be achieved by calculating its potential energy. The geometry of the molecule is modified iteratively so that the energy of the molecule is decreased until an energy minimum is reached. To solve this complex task, various mathematical algorithms are used in which the atomic positions can be varied either simultaneously or sequentially.

MOLECULAR DYNAMICS SIMULATION

Molecules are dynamic structures and therefore rigid conformations are inaccurate replica of the system. This is especially true for enzymes in which the dynamic characteristics are often essential for the enzymatic reaction (e.g., induced-fit enzyme mechanism). The motion ranges from picoseconds (pss) for amino acids up to ms range for domains. Molecular dynamic simulation methods enable the examination of these conformational changes over a period of picoseconds up to nanoseconds.

In molecular dynamics, the Newtonian equations of motion are calculated based on a previously selected force field (cf. Section 1.5.2). During the molecular dynamics simulation, the atomic positions and their velocity are calculated as a function of time to create a trajectory.

The disadvantage of this method is that the number of atoms and the simulation time is directly dependent on the available computing power. The larger the system, the longer the simulation must run to observe the scrutinized behavior [44].

Charge Assignment An important aspect in simulations mainly influenced by strong electrostatic interactions is the correct atom charge assignment prior to simulation. To calculate the partial charges of the atoms, the pK-values can be calculated using several techniques. Most software packages are freely available in the internet: Some approaches are based on the Poisson-Boltzmann equation (PBE). The PBE is an alteration of Poisson's equation that integrates a description of the effect of solvent ions on the electrostatic field in the surrounding of a molecule. The web servers H++ [50–52], MCCE [53–55], and Karlsberg+ [56, 57] use PBE to calculate the pK-values of the amino acids. PBE based methods calculate the difference of the pK-values of amino acids, when this side chain moves from a hypothetical fully solvated state in solution to its actual position in the protein with surrounding amino acids. To perform such a calculation, a theoretical method is needed that calculates the effects of the protein's inside on a pK-value, and the knowledge of the pK-value of amino acid side chains in its fully solvated conditions [58–61]. Other pK-value calculation methods are based on empirical methods, molecular dynamics, or calculations of free energy.

1.5.3. MOLECULAR MODELS OF PROTEIN CHROMATOGRAPHY

Multiple molecular models of ion exchange chromatography have been published in literature: Noinville et al. [28] simulated alpha-lactalbumin and lysozyme on an anion exchanger surface and studied their retention behavior. They found electrostatic patches, which dominated the binding. Ravichandran et al. [29] investigated the initial stages of hen egg-white lysozyme adsorption onto a charged solid interface using Brownian Dynamics simulation. They found that, although lysozyme has a positive net charge, it adsorbs on positively charged surfaces. Dismer and Hubbuch [30] developed a mechanistic model for the adsorption of lysozyme onto a cation exchange surface based on molecular dynamic simulations and were able to predict the retention of ribonuclease A.

Yarovsky et al. [62] simulated the interaction of peptides with reversed phase sorbents and demonstrated the potential of molecular dynamics to support the interpretation of peptide interactions with hydrophobic ligands.

Zamolo et al. [34] computationally investigated the design of affinity ligands through the study of their interactions with monoclonal antibodies and a model support material. The authors propose that the interaction between ligand and carrier material is an important parameter that should be taken into account in experimental design of ligands and that molecular dynamics can assist in the selection of suitable candidates.

The growing possibilities for the simulation of larger biomolecules through the increase in computational power makes it an interesting alternative to experimentally intense screenings. Continually increasing computational power enables ever faster and more accurate depictions of biomolecule behavior on the molecular level. As a result, molecular models not only enhance knowledge of the behavior on atomistic level but also have the potential to replace experimentally intense screenings in downstream process development in the future. This requires a versatile, easy to use, and expandable tool that enables quick gain in process understanding and thus the possibility to design it.

REFERENCES

[1] C. K. Mathews and K. E. van Holde. *Biochemistry*. Benjamin/Cummings series in life sciences and chemistry. Benjamin-Cummings Publishing Company, Subs of Addison Wesley Longman, Inc, 1996.

[2] K. Faber. *Biotransformations in Organic Chemistry: A Textbook*. Springer, 2011.

[3] L. Y. Jayasinghe, A. J. Smallridge, and M. A. Trewhella. The yeast mediated reduction of ethyl acetoacetate in petroleum ether. *Tetrahedron Lett.*, 34(24):3949–3950, 1993.

[4] A. J. J. Straathof and P. Adlercreutz. *Applied Biocatalysis*. Taylor & Francis, 2000.

[5] A. B. Fulton and W. B. Isaacs. Titin, a huge, elastic sarcomeric protein with a probable role in morphogenesis. *Bioessays*, 13(4):157–61, Apr. 1991.

[6] Wikimedia. Aminosäuren — Wikipedia, the free encyclopedia, 2015. [Online; accessed 19-February-2015].

[7] M. S. Tswett. O novoi kategorii adsorbtsionnykh yavlenii io primenenii ikh k biokhimicheskomu analizu (On a new category of adsorption phenomena and their application to biochemical analysis). In *Lecture at the March 8 (21), 1903, meeting of the Biological Section of the Warsaw Society of Natural Sciences*, volume 8, pages 20–39, 1903.

[8] *On a New Category of Adsorption Phenomena and their Application to Biochemical Analysis (translation from Russian)*, 1905. The Proceedings of the Warsaw Society of Natural Sciences, Sect XIV (6).

[9] M. S. Tswett. *Physikalisch-chemische Studien über das Chlorophyll. Die Adsorption.*, volume 24 of *Berichte der Deutschen Botanischen Gesellschaft*, chapter 24, pages 316–323. Gebrüder Borntræger, 1906.

[10] M. S. Tswett. *Adsorptionsanalyse und chromatographische Methode. Anwendung auf die Chemie des Chlorophylls.*, volume 24 of *Berichte der Deutschen Botanischen Gesellschaft*, chapter 24, pages 384–393. Gebrüder Borntræger, 1906.

[11] L. S. Ettre and K. I. Sakodynskii. M. S. Tswett and the discovery of chromatography I: Early work (1899–1903). *Chromatographia*, 35(3-4):223–231, 1993.

[12] A. J. P. Martin and R. L. M. Synge. A new form of chromatogram employing two liquid phases. *Biochem. J.*, 35(12):1358–1368, 1941.

[13] A. T. James and A. J. P. Martin. Liquid-gas partition chromatography. In *Biochem. J.*, volume 48, 1951.

[14] E. Stahl and H. Vollmann. Dünnschicht-chromatographie—XV: Trennung von terpen- und sesquiterpenalkoholen auf Silbernitrat-Kieselgel-Schichten, 1965.

[15] D. M. Ruthven. *Principles of Adsorption and Adsorption Processes*, volume 19. John Wiley & Sons, 1984.

[16] J. Ståhlberg. Retention models for ions in chromatography. *J. Chromatogr. A*, 855(1): 3–55, Sept. 1999.

[17] W. Kopaciewicz, M. A. Rounds, J. Fausnaugh, and F. E. Regnier. Retention model for high-performance ion-exchange chromatography. *J. Chromatogr. A*, 266(C):3–21, Aug. 1983.

[18] M. A. Rounds and F. E. Regnier. Evaluation of a retention model for high-performance ion-exchange chromatography using two different displacing salts. *J. Chromatogr. A*, 283:37–45, Jan. 1984.

[19] S. Yamamoto, K. Nakanishi, and R. Matsuno. *Ion-Exchange Chromatography of Proteins*. Chromatographic Science Series. CRC Press, 1988.

[20] C. A. Brooks and S. M. Cramer. Steric mass-action ion exchange: Displacement profiles and induced salt gradients. *AIChE J.*, 38(12):1969–1978, Dec. 1992.

[21] J. C. Bosma and J. A. Wesselingh. pH dependence of ion-exchange equilibrium of proteins. *AIChE J.*, 44(11):2399–2409, Nov. 1998.

[22] J. C. Bosma and J. A. Wesselingh. Available area isotherm. *AIChE J.*, 50(4):848–853, Apr. 2004.

[23] A. Ladiwala, K. Rege, C. M. Breneman, and S. M. Cramer. A priori prediction of adsorption isotherm parameters and chromatographic behavior in ion-exchange systems. *Proc. Natl. Acad. Sci. U. S. A.*, 102(33):11710–5, Aug. 2005.

[24] G. Malmquist, U. H. Nilsson, M. Norrman, U. Skarp, M. Strömgren, and E. Carredano. Electrostatic calculations and quantitative protein retention models for ion exchange chromatography. *J. Chromatogr. A*, 1115(1-2):164–86, May 2006.

[25] J. Ståhlberg, B. Joensson, and C. Horvath. Theory for electrostatic interaction chromatography of proteins. *Anal. Chem.*, 63(17):1867–1874, Sept. 1991.

[26] J. Ståhlberg, B. Joensson, and C. Horvath. Combined effect of coulombic and van der Waals interactions in the chromatography of proteins. *Anal. Chem.*, 64(24):3118–3124, Dec. 1992.

[27] C. M. Roth and A. M. Lenhoff. Electrostatic and van der Waals contributions to protein adsorption: computation of equilibrium constants. *Langmuir*, 9(3):962–972, Apr. 1993.

[28] V. Noinville, C. Vidal-Madjar, and B. Sebille. Modeling of Protein Adsorption on Polymer Surfaces. Computation of Adsorption Potential. *J. Phys. Chem.*, 99(5):1516–1522, Feb. 1995.

[29] S. Ravichandran, J. D. Madura, and J. Talbot. A Brownian Dynamics Study of the Initial Stages of Hen Egg-White Lysozyme Adsorption at a Solid Interface. *J. Phys. Chem. B*, 105 (17):3610–3613, May 2001.

[30] F. Dismer and J. Hubbuch. 3D structure-based protein retention prediction for ion-exchange chromatography. *J. Chromatogr. A*, 1217(8):1343–53, Feb. 2010.

[31] M. Kavoosi, N. Sanaie, F. Dismer, J. Hubbuch, D. G. Kilburn, and C. A. Haynes. A novel two-zone protein uptake model for affinity chromatography and its application to the description of elution band profiles of proteins fused to a family 9 cellulose binding module affinity tag. *J. Chromatogr. A*, 1160(1-2):137–49, Aug. 2007.

[32] B. G. Kermani, I. Kozlov, P. Melnyk, C. Zhao, J. Hachmann, D. Barker, and M. Lebl. Using support vector machine regression to model the retention of peptides in immobilized metal-affinity chromatography. *Sensors Actuators, B Chem.*, 125(1):149–157, 2007.

[33] H. Du, X. Zhang, J. Wang, X. Yao, and Z. Hu. Novel approaches to predict the retention of histidine-containing peptides in immobilized metal-affinity chromatography. *Proteomics*, 8:2185–2195, 2008.

[34] L. Zamolo, V. Busini, D. Moiani, D. Moscatelli, and C. Cavallotti. Molecular dynamic investigation of the interaction of supported affinity ligands with monoclonal antibodies. *Biotechnol. Prog.*, 24(3):527–39, 2008.

[35] H. A. M. Snelders. De reactie van buys ballot op van 't hoffs «voorstel tot uitbreiding der tegenwoordig in de scheikunde gebruikte structuur-formules in de ruimte». *Scientiarum Historia*, 15(1):27–38, 1973.

[36] J. D. Watson and F. H. C. Crick. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, 171(4356):737–738, Apr. 1953.

[37] J. C. Kendrew and R. G. Parrish. The crystal structure of myoglobin. III. Sperm-whale myoglobin. *Proc. R. Soc. A*, 238(1214):305–324, Jan. 1957.

[38] A. S. Dreiding. Einfache Molekularmodelle. *Helv. Chim. Acta*, 42(4):1339–1344, 1959.

[39] R. B. Corey and L. Pauling. Molecular models of amino acids, peptides, and proteins. *Rev. Sci. Instrum.*, 24(8):621–627, 1953.

[40] W. L. Koltun. Space filling atomic units and connectors for molecular models, Feb. 1965. US Patent 3,170,246.

[41] K. Müller. *Frontiers in drug research: crystallographic and computational methods: Proceedings of the Alfred Benzon Symposium 28 held at the premises of the Royal Danish Academy of Sciences and Letters, Copenhagen, June 11-15, 1989*, chapter Molecular Modelling and Structural Data Bases in Pharmaceutical, pages 210–221. Munksgaard, 1990.

[42] A. E. Torda and W. van Gunsteren. *Reviews in Computational Chemistry*, chapter Molecular Modeling Using Nuclear Magnetic Resonance Data, pages 143–172. John Wiley & Sons, 1992.

[43] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Res.*, 28(1):235–42, Jan. 2000.

[44] H.-D. Höltje and G. Folkers. *Molecular Modeling*. Wiley-VCH Verlag GmbH, 2008.

[45] K. Kuczera. *Molecular Modeling of Peptides*, volume 1268 of *Methods in Molecular Biology*, chapter 2, pages 15–41. Springer Science Business Media, 2015.

[46] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.*, 106(3):765–784, Feb. 1984.

[47] S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.*, 7(2):230–252, 1986.

[48] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: The gromos force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.*, 25(13):1656–1676, 2004.

[49] D. W. Mount. *Bioinformatics: Sequence and Genome Analysis*. Cold Spring Harbor Laboratory Press, 2004.

[50] R. Anandakrishnan, B. Aguilar, and A. V. Onufriev. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.*, 40(Web Server issue):537–41, July 2012.

[51] J. C. Gordon, J. B. Myers, T. Folta, V. Shoja, L. S. Heath, and A. Onufriev. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.*, 33(Web Server issue):W368–71, July 2005.

[52] J. Myers, G. Grothaus, S. Narayanan, and A. Onufriev. A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins*, 63(4): 928–38, June 2006.

[53] E. G. Alexov and M. R. Gunner. Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophys. J.*, 72(5):2075, 1997.

[54] R. E. Georgescu, E. G. Alexov, and M. R. Gunner. Combining conformational flexibility and continuum electrostatics for calculating pKas in proteins. *Biophys. J.*, 83(4):1731–1748, 2002.

[55] Y. Song, J. Mao, and M. R. Gunner. MCCE2: Improving protein pKa calculations with extensive side chain rotamer sampling. *J. Comput. Chem.*, 30(14):2231–2247, 2009.

[56] G. Kieseritzky and E.-W. Knapp. Optimizing pKa computation in proteins with pH adapted conformations. *Proteins: Struct., Funct., Bioinf.*, 71(3):1335–1348, 2008.

[57] B. Rabenstein and E.-W. Knapp. Calculated pH-dependent population and protonation of carbon-monoxy-myoglobin conformers. *Biophys. J.*, 80(3):1141–1150, 2001.

[58] M. R. Gunner, J. Mao, Y. Song, and J. Kim. Factors influencing the energetics of electron and proton transfers in proteins. what can be learned from calculations. *Biochim. Biophys. Acta, Bioenerg.*, 1757(8):942–968, 2006.

[59] G. M. Ullmann, E. Kloppmann, T. Essigke, E.-M. Krammer, A. R. Klingen, T. Becker, and E. Bombarda. Investigating the mechanisms of photosynthetic proteins using continuum electrostatics. *Photosynth. Res.*, 97(1):33–53, 2008.

[60] J. M. Antosiewicz and D. Shugar. Poisson-Boltzmann continuum-solvation models: applications to pH-dependent properties of biomolecules. *Mol. Biosyst.*, 7:2923–2949, 2011.

[61] D. Bashford. Macroscopic electrostatic models for protonation states in proteins. *Front. Biosci.*, 9:1082–1099, 2004.

[62] I. Yarovsky, M. T. W. Hearn, and M. I. Aguilar. Molecular Simulation of Peptide Interactions with an RP-HPLC Sorbent. *J. Phys. Chem. B*, 101(50):10962–10970, Dec. 1997.

Chapter 2

Research Proposal

Biopharmaceutical products require purification after production to achieve a high degree of purity. These high purities are mandatory for the approval in medical use. To purify the product, the downstreaming process contains multiple purification stages, usually containing several chromatographic steps. In chromatography, various techniques allow for separation based on different physicochemical properties of the biomolecule and its contaminants (cf. Section 1.1.2). The rising interest in understanding the mechanisms of attraction between a biomolecule and adsorbent material in chromatography led to several models (Section 1.4).

In this study, the approach was to extend and optimize a previously developed tool, which uses molecular dynamics simulations to model protein behavior on adsorbers. In the previous study, the model was solely applied to one protein and one cation exchanger. The tool was extended to other chromatography techniques and used to optimize a system in this approach. The simulations were underpinned by corresponding laboratory experiments.

The aim was to increase understanding of adsorption behavior of biomolecules on adsorber surfaces of various adsorbents. First, the transfer from a cation exchanger to an anion exchanger needed to be accomplished, followed by a detailed investigation of binding contributors. It was expected that negatively charged residue clusters contribute to binding behavior. Next, the tool was applied to a cation exchanger with varying ionic capacities to explore the influence on binding. By this approach, it was expected to determine an adsorber with a specific capacity of being the optimal choice for a given separation challenge. Lastly, a completely different chromatography technique was studied with the tool: affinity coupling chromatography. This was the most challenging, since it was not clear if affinity coupling can be simulated reliably. It was expected that the results indicate energetically favorable orientations, on which basis a thesis may be proposed on which residues are likely to involve in a coupling reaction.

Chapter 3

A Comprehensive Molecular Dynamics Approach to Protein Retention Modeling in Ion Exchange Chromatography

Katharina M. H. Lang, Jörg Kittelmann, Cathrin Dürr, Anna Osberghaus, Jürgen Hubbuch

In downstream processing, the underlying adsorption mechanism of biomolecules to adsorbent material are still subject of extensive research. One approach to more mechanistic understanding is simulating this adsorption process and hereby the possibility to identify the parameters with the strongest impact. So far this method was applied with all-atom molecular dynamics simulations of two model proteins on one cation exchanger. In this work we developed a molecular dynamics tool to simulate protein-adsorber-interaction for various proteins on an anion exchanger and ran gradient elution experiments to relate the simulation results to experimental data. We were able to show that simulation results yield similar results as experimental data regarding retention behavior as well as binding orientation. We could identify Arginines in case of cation exchangers and Aspartic acids in case of anion exchangers as major contributors to binding.

published in Journal of Chromatography A, Volume 1381, 13 February 2015, Pages 184-193

3.1. INTRODUCTION

Biopharmaceutical products need a purification process after product retrieval to achieve a good degree of purity [1]. These high levels (usually 99.9 % or higher) are mandatory for medical use. The downstreaming process contains many purification stages, usually containing chromatographic steps [2]. In chromatography, various techniques allow for separation based on the different physicochemical properties of the biomolecule and its contaminants. A separation method based on charge differences is ion exchange chromatography. The interaction between charged molecule and immobile phase leads to separation.

The growing interest in understanding the basic mechanisms of attraction between a biomolecule and adsorbent material in the various chromatographic techniques led to many mathematical models. Various semi-empirical and mechanistic models for the description of protein interaction behavior on ion- exchangers were developed in the past. In 1992, Brooks and Cramer [3] published a paper in which they described the SMA model, which accounts for the steric hindrance of counter ions. Bosma and Wesselingh [4, 5] have developed the available area model, which uses the protein's radius to consider the geometric hindrance on the adsorber. QSPR/Quantitative Structure Activity Relationship (QSAR) is another approach in which descriptors of molecule characteristics correlate with experiments [6, 7]. The major drawback of this approach is the need for a large data set for model training. Mechanistic models are based on electrostatic calculations and van der Waals forces [8] or calculations of Gibbs free energy [9, 10]. The downside of these approaches is that they simplify proteins to rigid bodies. Yet, proteins are not static but rather dynamic molecules, that can have different native conformations. This drawback was overcome by molecular dynamics (MD) simulations. With molecular dynamics, the adsorber and proteins can be represented either by coarse-grained models, which is a simplification of the system and simultaneous reduction of simulation time or by all-atom models, with higher precision at the cost of longer simulation time. Molecular dynamic studies of adsorption on surfaces can be classified into three levels: modeling and visualization, adsorption of small molecules, and protein adsorption [11]. The protein-surface interaction influences adsorption behavior the most and therefore is subject to diverse studies: interactions between affinity ligands and monoclonal antibodies [12], with two monoclonal antibodies and two different ligands, which resulted in comparable binding behavior as on protein A. Also, interactions between proteins and self-assembled monolayers (SAMs) [13–15] or a MgO surface [16], respectively, were studied. In all studies the obtained adsorption energies correlated well with experimental results, though mostly qualitative rather than quantitative [17–19].

As mentioned previously, not only the protein-surface interaction is matter of investigation but also the protein orientation. The orientation has a major influence on the adsorption behavior since the protein has heterogeneous characteristics, such as hydrophobic and hydrophilic patches, but also charged areas. Depending on the characteristics of ligand and matrix, the protein interaction is dominated by hydrophobicity or electrostatics. Therefore, multiple studies examine protein orientation with MD tools: Noinville et al. [20] simulated alpha-lactalbumin and lysozyme on a macromolecular surface and studied their retention behavior. They found electrostatic patches, which dominated the binding. Ravichandran et al. [21] studied lysozyme adsorbtion on a positively charged plane utilizing Brownian dynamics simulations and found that the orientation of the proteins was nonuniform, which confirms the findings of Noinville et al. [20]. Agashe et al. [22] simulated a Fibrinogen γ-Chain Fragment on SAMs and found that, if not restrained, the protein undergoes multiple translations and rotations until reaching stable orientations.

A more recent approach using MD simulations to investigate protein-surface interaction and simultaneously protein orientation was by Dismer and Hubbuch [23]. The authors showed a correlation between simulated and experimental retention factors for two model proteins on a cation exchanger surface. However, this approach is limited to one adsorber and biomolecules without non-standard residues, presenting only a small fraction of potential target molecules.

In this research project we developed a MD tool, which is capable to run simulations with proteins also containing non-standard residues. With this tool we established a mechanistic model for an anion exchange adsorber type. The tool is capable to predict retention behavior as well as gain insights into the binding mechanism on molecular level. This work follows a case-study design, with in-depth analysis of interaction behavior of various proteins with two ion exchanger surfaces. With our tool, we simulated lysozyme on a SP Sepharose FF surface to compare these results to previous experimental findings. We included all positively charged residues, lysine, arginine, and histidine, in the analysis to determine their influence on interaction. Also, various proteins were simulated on an anion exchanger surface. The analysis of these simulations focuses on the negatively charged residues, glutamic acid and aspartic acid, and their contribution to binding.

This study addresses the following research questions:

1. Do simulations yield similar results as experimental data?
2. Which residues are the major factors for the protein-adsorber interaction in ion exchange chromatography?

Note: Throughout this paper, the terms SPFF and QFF will refer to SP Sepharose FF and Q Sepharose FF, respectively.

3.2. MATERIAL AND METHODS

The proteins chosen for simulations and experiments were selected on the basis of protein size and isoelectric point. Protein size has a big influence on simulation duration, thus smaller proteins were preferred for this MD tool development and case study. In the anion exchanger experiments the protein was supposed to bind and so the isoelectric point of the protein was crucial. Another requirement was that the proteins were in monomeric native state, since the simulations were also run with monomers. Table 3.1 lists the chosen proteins, the corresponding PDB ID, concentrations in experiments, and article number from Sigma-Aldrich.

The first part of this section focuses on the experiments that were conducted and the second part focuses on the setup, the actual simulations, and the data analysis.

Table 3.1: Proteins used in this study.

Protein	PDB ID	No. of residues	Conc. [mM]	Sigma no.
Alpha-lactalbumin	1F6R	123	0.6	L5385
Beta-lactoglobulin	4TLJ	162	0.6	L3908
Phospholipase A2	1BP2	123	0.6	P8913
Ribonuclease A	1KF8	124	1.2	83833

3.2.1. PROTEIN SIZE ANALYSIS

The protein size is a good indicator for the protein's native state and aggregates. If it is too large, either the protein is denatured or aggregated. The protein size was analyzed by dynamic light scattering (DLS) in a Zetasizer Nano ZSP from Malvern Instruments (Malvern, Worcestershire, UK). The diffusion barrier method was used to prevent protein aggregation during the measurement on the cell electrodes. The U-shaped capillary is filled with sample buffer. Next, the protein sample is isolated from the electrodes by injecting it into the measurement zone of the capillary. Alpha-lactalbumin and beta-lactoglobulin were measured in 50 mM 1-Methylpiperazine titrated to pH 9.8, 50 mM 2-(N-morpholino)ethanesulfonic acid (MES) titrated to pH 6.0, and 50 mM phosphate buffer titrated to pH 6.8. Phospholipase A2 and Ribonuclease A were only measured in the pH 9.8 buffer because of their higher

theoretical isoelectric point (6.1 for Phospholipase A2 and 8.64 for Ribonuclease A, as calculated from pK values). The experimental pH values of 1-Methylpiperazine and MES were too close-by for Phospholipase A2 and thus no protein size analysis was run. The gradient elution runs were conducted with proteins whose pI is below the pH value of experiments. Since Ribonuclease A has a theoretical pI of 8.64, it was not run with the two lower pH buffers and correspondingly the protein size analysis was not run either. All buffer ingredients were purchased from Sigma-Aldrich, St. Louis, MO, USA. Sample solutions were filtered through a 0.2 µm cellulose-acetate membrane prior to measurement to ensure no aggregates were present.

3.2.2. GRADIENT ELUTION EXPERIMENTS

The proteins that were used for the gradient elution experiments are enlisted in Table 3.1. All proteins were purchased from Sigma-Aldrich, St. Louis, MO, USA. The proteins were dissolved in the buffers described in Section 3.2.1. The starting buffer was the same as the solvate buffer. The elution buffer consisted of starting buffer plus 1 M NaCl purchased from Merck (Darmstadt, Germany). All buffers were filtered and deaerated in an ultrasound bath for 15 minutes. The protein solution was filtered through 0.2 µm cellulose-acetate membrane. An ÄKTApurifier system purchased from GE Healthcare (Little Chalfont, Buckinghamshire, United Kingdom) was used for gradient elution experiments with a 1 mL prepacked column HiTrap Q Sepharose Fast Flow (material and packing by GE Healthcare). The column was equilibrated after storage by removing storage buffer with water for one column volume (CV), followed by 5 CVs of starting buffer, 5 CVs of elution buffer, and henceforth 5 CVs of starting buffer at 100 cm/h. The gradient elution procedure was carried out by 2 CVs equilibration with starting buffer, sample loading (25 µL sample), 5 CVs wash with starting buffer, 15 CVs gradient up to 0.5 mM NaCl, 5 CVs high salt at 200 cm/h to speed up the run, and 10 CVs regeneration also at higher flow rate. Unless stated otherwise, the flow rate was 100 cm/h for all steps in the gradient run until the end of the gradient. The runs were conducted in triplicates using the scouting function of Äkta's UNICORN system control software (version 5.3.1). After the scouting runs, the columns were flushed with water and stored in 20% Ethanol/water at 4 °C. Retention volume and conductivity at maximum point of elution was determined in UNICORN. Retention factor k' was calculated in accordance with Schmidt-Traub [24] by

$$k_i' = \frac{t_{R,i} - t_0}{t_0} \tag{3.1}$$

where $t_{R,i}$ is the retention time of component i and t_0 is the dead time of the column, determined with beta-lactoglobulin in high-salt buffer to hinder protein adsorption.

3.2.3. LIGAND DISTANCE CALCULATION

To position the ligands in the correct distance from each other, the assumption was made that ionic capacity correlates with ligand density. We developed a script with which we can calculate the ligand distance with

$$A_{Total} = \frac{4(1-\varepsilon)\varepsilon_P}{d_{Pore}} + \frac{6(1-\varepsilon)}{d_{Particle}} \tag{3.2}$$

where A_{Total} is the total area available for binding, ε and ε_P the extra-particle and intra-particle porosity, d_{Pore} the pore diameter, and $d_{Particle}$ is the particle diameter. Equation 3.2 is adapted from Carta and Jungbauer [25]. The total number of ligands per surface volume is defined by

$$N_{Surface} = \frac{IC \times N_A}{A_{Total}} \tag{3.3}$$

where $N_{Surface}$ is the number of ligands per surface unit, IC is the ionic capacity in mol/nm^3 and N_A is the Avogadro constant. The ligand distance is calculated by

$$d_{Ligand} = \frac{\sqrt{A_{unit}}}{\sqrt{N_{Surface}}} \tag{3.4}$$

where d_{Ligand} is the ligand distance and A_{unit} is the surface unit. Equation 3.2 follows the single pore model [26], which simplifies the pore network to cylindrical pores. The resulting pore size distribution can be described as one large pore, that is the sum of the cylindrical pores (Figure 3.1). Since adsorber characteristics are subject to fluctuations due to manufacturing processes, we used Monte Carlo simulations with 10.000 repetitions applying the fluctuation values from Table 3.2. All physical properties of the adsorber were assumed to be Gaussian distributed. The range of variation for the ionic capacity specification were assumed to be the twofold standard deviation. When not the range of variation but multiple values were given in literature, the mean and standard deviation were used for Monte Carlo simulations. The calculated result was taken as ligand distance in the simulation setup. The Monte Carlo simulations were performed in Matlab version 8.1.0.604 (R2013a).

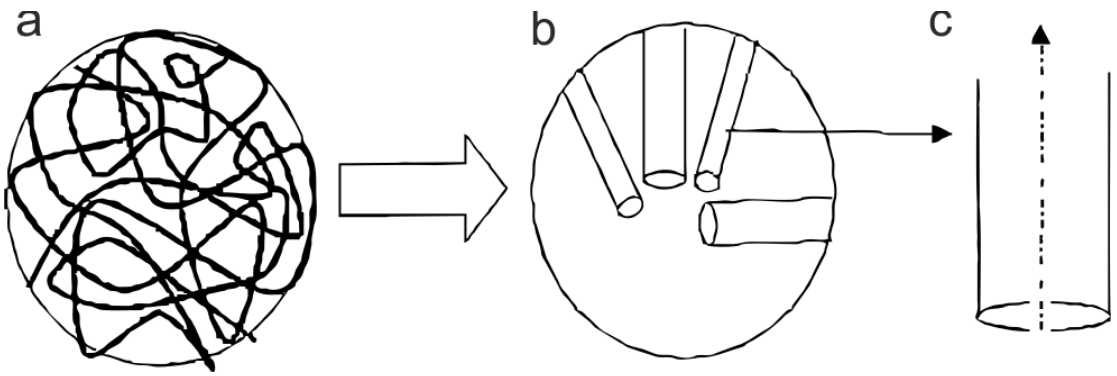


Figure 3.1: Simplification of the complex agarose network (a): to a capillary pore distribution model (b); to a single pore model (c). Adapted from [26]. Copyright 2004 American Institute of Chemical Engineers (AIChE). Reprinted with permission.

Table 3.2: Values on adsorber characteristics for SP Sepharose FF and Q Sepharose FF used for Monte Carlo experiments.

Physical quantity	Value and reference	Unit
Ionic capacity IC_{total}	180-250[a] 180-240[b] [27]	$\mu mol \times mL^{-1}$
Pore diameter d_{Pore}	30, 33.4, 38.6 [25], 32 [28]	nm
Particle diameter $d_{Particle}$	45-165 [27]	μm
Intraparticle porosity ϵ_p	0.76, 0.81, 0.85 [25], 0.9 [28]	-
Extraparticle porosity ϵ	0.35 [28]	-

[a] SP Sepharose FF
[b] Q Sepharose FF

3.2.4. MOLECULAR DYNAMICS SIMULATIONS

The molecular dynamics tool is a set of macros for Yasara and a post-simulation Matlab script for graphical output. All simulations were run fully automated from energy minimization to post-simulation analysis with simulation time ranging from two up to five days depending on the number of atoms. If not stated otherwise, all simulations were run with a cutoff at 7.86 Å. Long-range Coulomb interactions were calculated using the Particle Mesh Ewald (PME) algorithm [29] and boundaries were set to periodic. All MD simulations were run with the software Yasara Structure 13.6.16 [30] on 8 processors on a distributed memory parallel computer, running SUSE Linux Enterprise 11.

PROTEIN STRUCTURE PREPROCESSING

Most proteins contain heteroatoms or non-standard residues. Any residue of protein or nucleic acid that is not included in the list of standard residues (20 standard amino acids, 12 stan-

dard nucleotides) is considered non-standard. These residues must be parameterized prior to simulation. We used the simulation software Yasara [30], which is capable of automatically assign parameters to molecules, which are not described in the force field (in this case Calcium and its bonds in alpha-lactalbumin) through an internal algorithm, AutoSMILES [31–34]. This makes it very convenient to simulate many proteins in rather short time.

In a first step, we identified the correct UniProt ID [35] by protein name and organism to find an appropriate PDB structure. The RCSB Protein Data Bank [36] was then searched for the selected UniProt ID. Appropriate protein structure files were picked by two selection criteria: closest number of residues and best resolution. Structures derived from X-ray diffraction measurements were favored over NMR spectroscopy derived structures, since atomic information provided by X-ray crystallography is more detailed. The PDB IDs for the proteins that were used in lab experiments can be found in Table 3.1. Additionally, to this table, lysozyme (PDB ID: 2VB1) was used for simulations to compare with experimental result from a previous study [37]. Once a suitable PDB ID was selected, the 3D coordinate file of the protein was downloaded from the RCSB Protein Data Bank and saved. The atomic bonds and hydrogen atoms were checked for integrity in Yasara [30] by comparing them to the UniProt entry. The hydrogen bonds were corrected and optimized via an internal Yasara algorithm (for further details see reference [38]).

To avoid atom collisions in the protein and correct the covalent geometry, the protein structure was energy minimized with the AMBER99 force field [39], using a 7.68 Å force cutoff. Long-range electrostatic interactions were calculated with the Particle Mesh Ewald algorithm [29] . After removal of conformational stress by a steepest descent minimization, the energy minimization continued by simulated annealing (time step 2 fs, atom velocities scaled down by 0.9 every 10th step) until convergence was reached, i.e., the energy improved by less than 0.25 kJ/mol per atom within 200 steps.

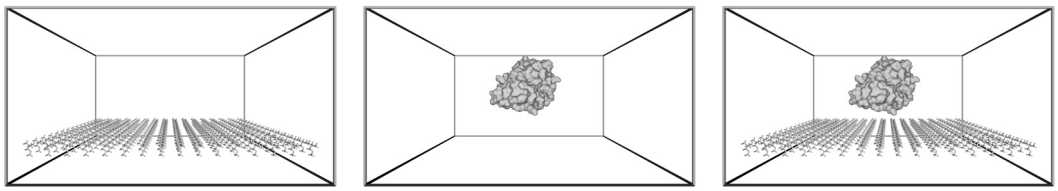
The energy minimized structure was saved in PDB format and uploaded to the H++-Tool from Virginia Tech [40] to calculate a protonated protein structure depending on the selected pH value. Confidential structures can be processed locally when installing the source code, which is available upon request. The protonated PDB file was downloaded and checked again for missing bonds (e.g., bonds to heteroatoms, such as Ca^{2+}) and, if necessary, corrected. The completed structure was then energy minimized by the same procedure as mentioned before. If the protein structure contained non-standard residues such as a heme group or metal bonds (alpha-lactalbumin) the procedure was slightly different: after the first energy minimization, the non-standard residues were depleted from the protein structure in Yasara

and saved off as an extra PDB file. After the protonation of the depleted protein structure in H++, both PDB files were loaded into Yasara and bonds between the depleted protein and the non-standard residues were rebuilt. The energy minimization was the same as described above.

COMPLEX CONSTRUCTION

The 3D ligand structure was built in Yasara by using the information on chemical composition from the manufacturer's website. The functional group of SP Sepharose FF is
$CH_2CH_2CH_2SO_3-$
and the functional group of Q Sepharose FF is
$CH_2N^+(CH_3)_3$ [41].
The ligands were replicated and placed equidistantly depending on the result of the calculated ligand distance (see Section 3.2.3) in XZ-plane spanning a quadratic surface. The adsorber matrix onto which the ligands are bound was simplified by using a dummy C-atom per ligand. The size of the plane was dependent on protein size to reduce computational cost and give all proteins equal conditions for interaction.

The complex consists of the whole setup of protein on ligand surface and the surrounding solvent molecules. The protein was rotated in 50 equidistantly distributed orientations to assure impartiality. In the next step, the protein was positioned 5 Å above the ligand surface. The distance is the minimum distance between the van der Waals radii of the ligands and the protein. A distance of 5 Å was chosen because lower distances lead to unrealistic high energies and consequently an explosion of the simulation cell. Also, parts of the protein were supposed to be within the Debye screening length of the ligands, which in our case is approx. 7 Å. Therefore, 5 Å was chosen as protein-ligand distance. After complex construction, the simulation box was defined depending on the ligand surface dimension. The cell was extended by 10 Å on each side of the protein along the plane of the ligands.



(a) Ligand surface (b) Protein (c) Complex (here without solvent)

Figure 3.2: Lysozyme on SPFF

ENERGY MINIMIZATION

The simulation cell was filled with water molecules and counter ions (i.e., sodium and chloride ions) were placed at locations of extreme electrostatic potential until the cell was neutralized. The electrostatic potential is naturally highest in regions close to charged molecules. A short solvent simulation was run and afterwards water molecules were successively deleted until the density matched 0.9966 g/cm^3. The total system was then energy minimized following the same EM protocol as for the protein alone, mentioned in Section 3.2.4.

EQUILIBRATION

In equilibration, the system is heated from 0 to 300 Kelvin, whereas the protein backbone and bond length of hydrogen atoms are restrained to reduce computational cost. The adsorber backbone is modeled by fixing the first three respectively two atoms of ligands in space for SPFF respectively QFF. The QFF ligand is shorter, fixing the first three atoms would decrease the flexibility disproportionately compared to SPFF. The used force field was AMBER03 [42].

MOLECULAR DYNAMICS SIMULATION

The data production simulation was run for 50 ps at 300 K, utilizing a Berendsen thermostat, which rescales atom velocities after calculating an average macroscopic temperature and the corresponding atom velocities. Simulations were of NVT type, meaning the number of moles (N), volume (V) and temperature (T) are kept constant. A snapshot of the system was saved each 5 ps. Time steps for calculation of intra- and intermolecular forces were 1.33 fs and 4 fs, respectively. The simulation was kept quite short since the backbone is restrained and larger conformational changes are therefore suppressed. Since no significant energy leaps could be observed beyond the equilibration phase neither in a previous study with longer simulation times [23] nor in this study, the simulations in this work were reduced to a time span in which an energetical equilibrium was reached.

DATA ANALYSIS

Post-simulation data analysis was run in two steps: previously saved simulation snapshots were opened in Yasara, energies (total, electrostatic, van der Waals, potential) were calculated and tabulated over simulation time and saved for further processing for each orientation. The interaction energy is the energy of the complex minus the energy of the separated compounds and was calculated for all energy terms. The more negative the energy, the stronger

is the interaction between protein and ligand surface. Also, an average protein structure over each simulation run was calculated and energy minimized (analogous to Section 3.2.4).

In a second step, the result tables were loaded into Matlab and various calculations were run. For each orientation average energies over time were calculated and Boltzmann weighted. The Boltzmann statistic indicates the probability of a state of a system (in this case one state represents one orientation), which is coupled to a temperature T in thermodynamic equilibrium. We make two assumptions: 1. each orientation simulation is in thermodynamic equilibrium, and 2. that the states (numbered by j = 1,2, ... N) have the associated simulation energy E_j. The probability of state j then is

$$p_j = \frac{e^{-\beta E_j}}{Z} \tag{3.5}$$

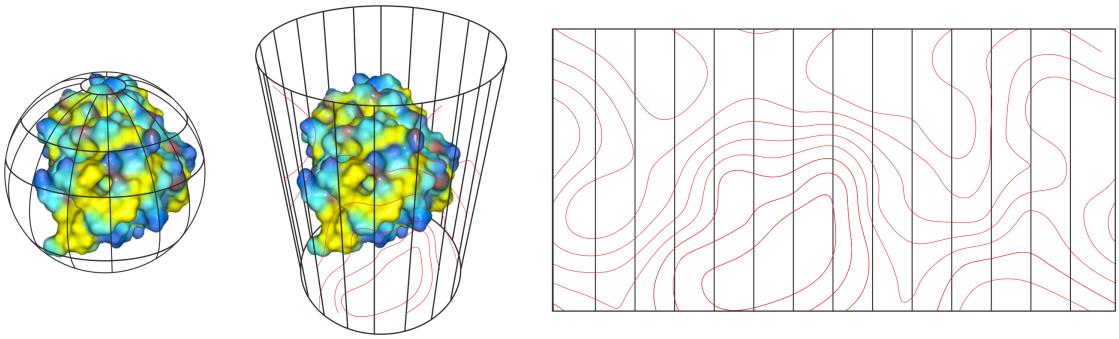
where Z is the partition function and is obtained by

$$Z = \sum_{j=1}^{N} e^{-\beta E_j} \tag{3.6}$$

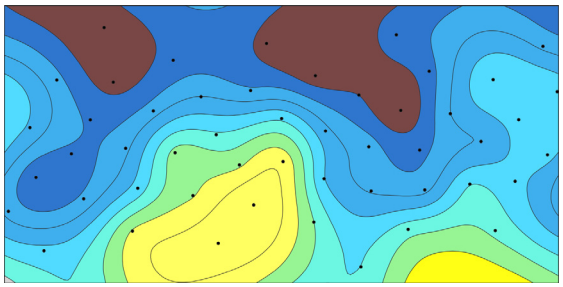
and β is a constant, which is obtained by

$$\beta = \frac{1}{k_B T} \tag{3.7}$$

where k_B is the Boltzmann constant and T the temperature. An interaction map (see Figure 3.3) for each protein is calculated by plotting the Boltzmann averaged energy values for each orientation and interpolating with Sibson's natural neighbor interpolation algorithm [43]. Protein orientations, which yield negative electrostatic interaction energy imply attraction between protein and adsorber. Only the negative interaction energies were Boltzmann weighted and averaged, since positive energies (i.e., repulsion of the protein) do not contribute to binding and would lead to a rotation to a favorable orientation. The final average is the sum of all 50 weighted electrostatic energies, and thus a distinct quantity of interaction for each protein. This protein-specific average interaction energy is then used for the correlation with experimental results.



(a)



(b)

Figure 3.3: (a) Cylindrical projection of points from a protein; unwrapping and flattening out the cylinder results in the (b) interaction map

3.3. RESULTS AND DISCUSSION

3.3.1. PROTEIN SIZE ANALYSIS BY DYNAMIC LIGHT SCATTERING

DLS measurements were conducted as described in Section 3.2.1 to determine the protein size. The results from this analysis are presented in Table 3.3. It is apparent that all proteins except Phospholipase A2 have radii of around 2 nm. Phospholipase A2 in 50 mM 1-Methylpiperazine had a radius of 44 nm.

An extraordinary large protein radius can have several causes: denaturation or aggregation. Phospholipase A2 has a theoretical radius of gyration of 1.45 nm, calculated in Yasara. The instability index was calculated as 29.77 according to Guruprasad et al. [44], with values below 40 indicating stable proteins. Since it can withstand strong reagents according to the manufacturer, denaturation can be ruled out. This implies that Phospholipase A2 aggregates. Since the experimental conditions result in aggregated protein, but MD simulations were run with monomeric Phospholipase A2, the protein had to be rejected in the further course of this work.

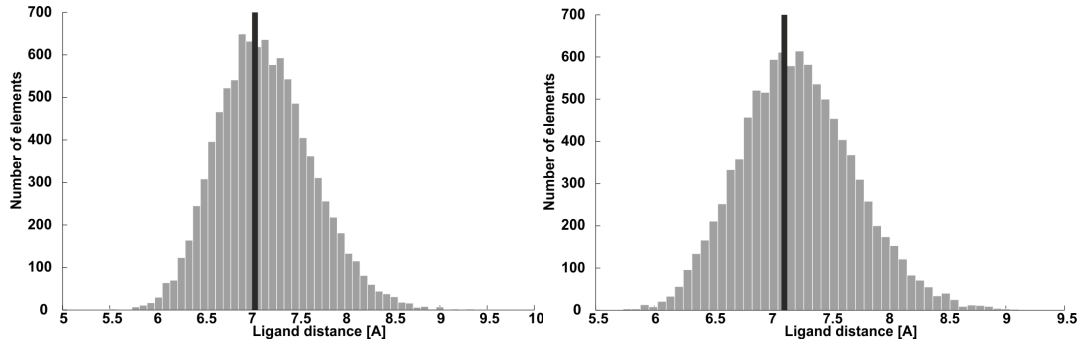
The results suggest that analyzing protein stability is a crucial step, since aggregation cannot be observed in retention experiments. In this way deviant on -column behavior can be explained.

Table 3.3: Median of protein radii of proteins at multiple pH values, measurement in triplicates

	Radius [nm]		
Protein	pH 6	pH 6.8	pH 9.8
Alpha-lactalbumin	2.21	2.17	1.89
Beta-lactoglobulin	3.44	3.15	2.04
Phospholipase A2	-	-	44.06
Ribonuclease A	-	-	2.13

3.3.2. LIGAND DISTANCES FROM MONTE CARLO SIMULATIONS

Monte Carlo simulations were run as described in Section 3.2.3 to calculate ligand distances that are valid equivalents to given ionic capacities. The one-sample Kolmogorov-Smirnov test revealed that the ligand distance distributions (Figure 3.4) were not normally distributed. Both distributions are slightly positively skewed and thus the median was calculated. The ligand distances are 7.04 Å for SPFF and 7.11 Å for QFF. The findings of the Monte Carlo sim-



(a) (b)

Figure 3.4: Ligand distance histograms for (a) SP Sepharose FF and (b) Q Sepharose FF; the median is depicted as a gray vertical line

ulations confirm the trend shown by Dismer and Hubbuch [23] who simulated SPFF with ligand distances between 10 and 20 Å and empirically confirmed the best correlation with experimental results at a ligand distance of 10 Å.

Therefore, the Monte Carlo approach is an efficient tool for computing ligand distances of adsorbent materials with known physical properties, which can be described by the single pore model.

3.3.3. SP SEPHAROSE FF

MD simulations with lysozyme on SPFF at various pH values were run, as described in Section 3.2.4. In the analysis step, all positively charged residues (arginines, lysines, and histidines) were plotted into the interaction maps to determine their influence on the electrostatic interaction. The electrostatic interaction energy decreases with increasing pH with a sharp decrease beyond pH values of 9. The simulation results with pH values of 5, 9, and 12 are presented in Figure 3.5. The net charge for these pH values are 9, 8, and 3, respectively, which corresponds well with the electrostatic interaction energy course. The protein as seen from the ligand's point of view is depicted on the right-hand side. The lowest energy reflects the highest interaction between protein and adsorber surface, this defines a hot spot. It is apparent from Figure 3.5 that there are two sets of orientations that are favorable for binding: orientations A and B. Orientation A in 3D depiction shows that the protein faces the ligands with several arginines and lysines (Lys1, Lys33, and potentially Lys13). Only arginines (Arg45, Arg61, Arg68, Arg73, Arg112, and Arg114) are in close range to orientation B. Orientation B fades for the benefit of orientations C and A with increasing pH. Lys96 and Lys97 are both close to the ligands in orientation C. Also, in orientation C Lys1 has no interaction with the adsorber surface. With rising pH, the interaction of Lys116 with the ligands gains influence.

The results of this study will now be compared to the experimental findings of Dismer et al. [37]. The authors conducted a study in which lysozyme was bound onto a cation exchanger in batch binding experiments and then labeled the lysines. A difference in labeling efficiency of lysines in lysozyme depending upon pH value was found by the authors. When looking at orientation A, the orientation sterically hinders Lys1, Lys33 and potentially Lys13. Lys13 especially is hindered if high protein concentration conditions are run, which was the case in the author's experimental set-up. At pH 12, orientation C gains importance in which Lys96/97 blocked from labeling, which is again in good agreement with experimental data from Dismer et al. [37].

Since the blocking of lysines could be shown in experiments as well as in simulations, we suggest that our depiction of simulation results is capable of identifying residues that hinder labeling. Since our interaction maps integrate all charged residues, binding contributing

residues can easily be identified. We determine not only lysine as a strong contributor but especially arginines having an important role in binding. A possible explanation for the strong influence of arginines is their side-chain flexibility. arginines are flexible and hence can turn towards the negatively charged ligand surface. arginines have a strong influence on the binding mechanism and with that comes a steric hindrance of residues that are close to the ligand surface. This hypothesis is in good agreement with Fromm et al. [45] who showed that the arginine-sulfate interaction is stronger than lysine-sulfate interaction.

3.3.4. Q SEPHAROSE FF

MD MODEL

This study related protein dynamics results with the biological adsorption mechanism. To relate the simulation results to the adsorption mechanism, the Boltzmann weighted, averaged negative electrostatic interaction energy was the measure for attraction, which corresponds to the retention volume in experiments.

The Boltzmann averaged electrostatic interaction energy, calculated as described in Section 3.2.4, was correlated with the retention factors for the experiments on QFF. Alpha-lactalbumin at pH 6.8 showed a split peak in the gradient elution experiments. Since it was impossible to determine the peaks without further assays, alpha-lactalbumin for pH 6.8 was removed from linear regression.

The model for QFF obtained from the linear regression is presented in Figure 3.6. The coefficient of determination R^2 is 0.96, which suggests a highly significant relationship. The mean error of prediction is 1.26 mL as calculated via cross validation.

A mean error of prediction of 1.26 mL with a gradient length of 15 mL and 1 mL column volume provides a high level of predictability for an unknown protein structure. Through the cross validation the model is built with a training set and then testing against a set, which is unknown to the system. This simulates the scenario of a new structure being passed into the MD tool for which gradient experiments have never been conducted.

The slope in retention for beta-lactoglobulin indicates a slight overestimation of ligand distances in the simulation. Dismer and Hubbuch [23] showed decreasing slopes in a correlation of retention factor to interaction energy with decreasing ligand distances. This is most probably due to resin characterization based on values from different sources in literature and can be avoided by characterizing the adsorbent batch in use. The negative correlation

between retention factor and binding energy was expected, as a weaker interaction energy (higher energy values) results in earlier desorption.

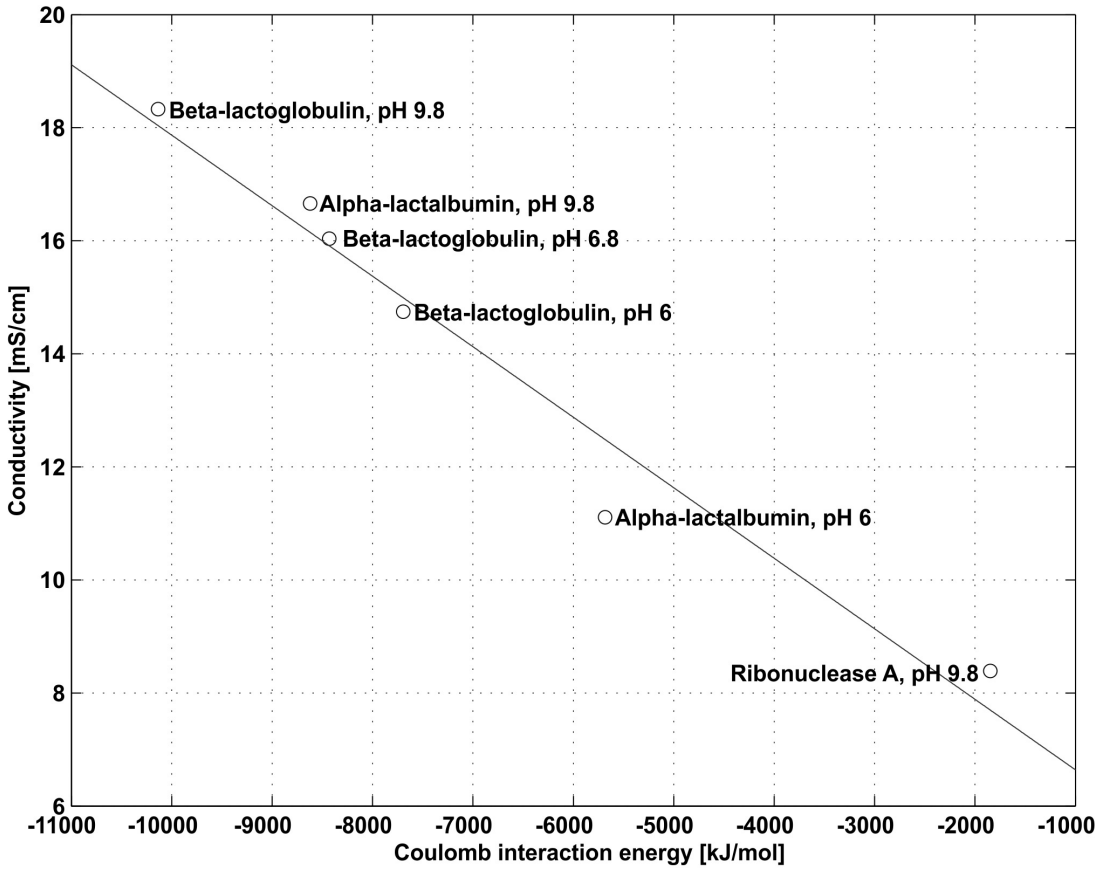


Figure 3.6: Retention model for Q Sepharose FF, linear fit, R^2 = 0.96

INTERACTION MAPS

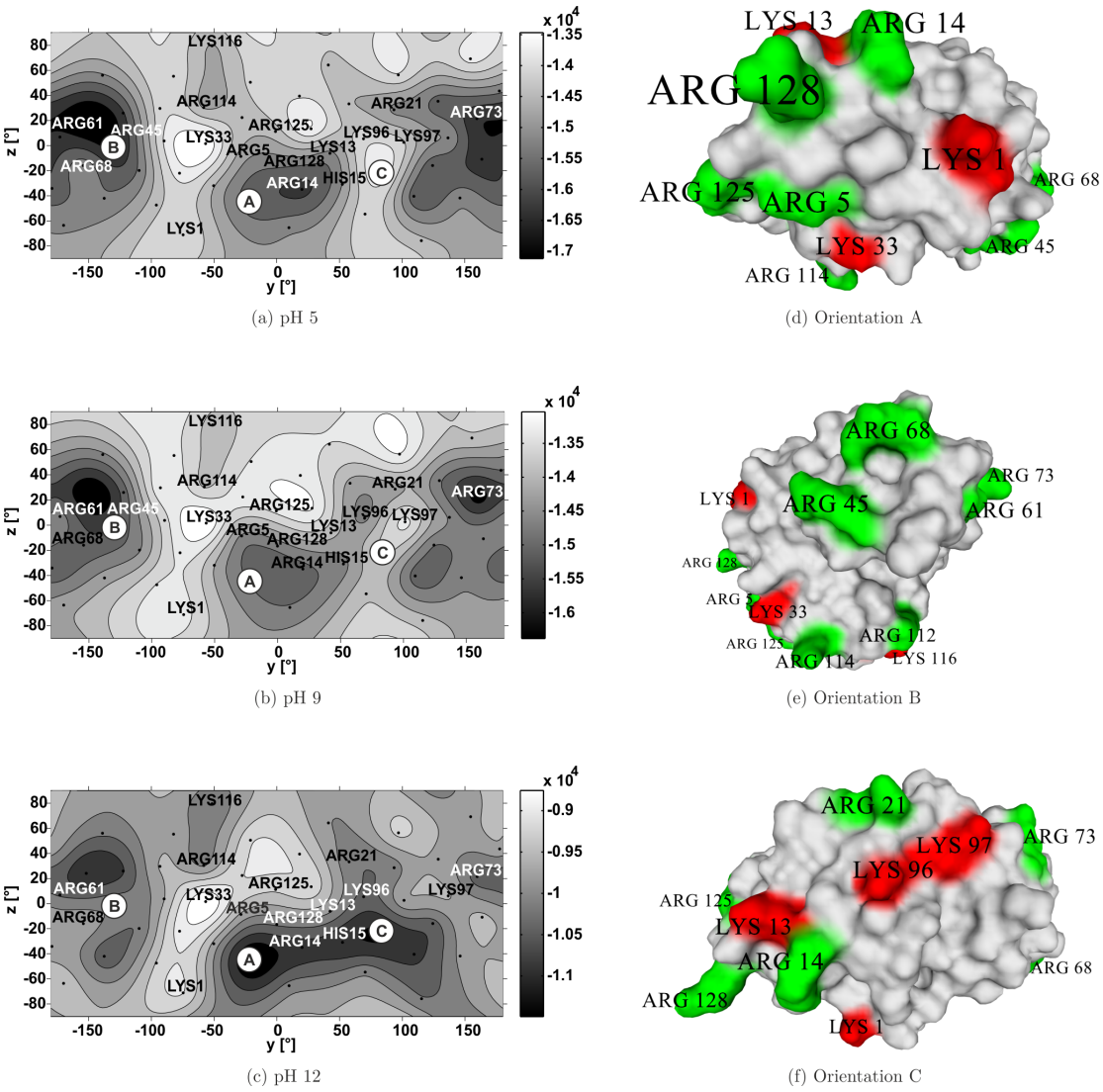
Aspartic acid and glutamic acid are both negatively charged at pH values above 4.3 and thus interact with anion exchangers. At ph 9.8, free cysteines are also negatively charged and would have been considered in the interaction maps, but all cysteines in the proteins observed (alpha-lactalbumin and beta-lactoglobulin) formed disulfide bridges and therefore not charged.

Alpha-lactalbumin Figure 3.7 compares the resulting interaction maps from alpha-lactalbumin on QFF for pH values 6.0, 6.8, and 9.8. Two orientations, A and B, are subject to

major changes with changing pH and were therefore chosen for the 3D analysis to the right-hand side of the interaction maps. The two apparent hot spots that hit the eye (y = [-50, 0], z = [-60, -80] and y = [100, 150], z = [-60, -80]) are in fact identical since the unwrapping of the sphere from 3D to 2D causes the lower part of the map to represent the same orientation (compare Figure 3.3a). The same is true for the upper part of the map. It is apparent from Figure 3.7 that a slight change in pH from 6.0 to 6.8 leads to a small shift in orientation and a detachment towards A. The pH change yields a net charge difference of 4 units from -1 to -5. The shift from pH 6.8 to 9.8 only results in a difference of 3 units. This result is mirrored in the weak change of interaction energy maxima from pH 6.8 to 9.8. The maximum interaction energy was -9047.3 kJ/mol at pH 6.8, whereas -9876.8 kJ/mol at pH 9.8. Orientation B is also subject to major changes, as the strongest interaction region at lower pH value moves towards A for higher pH values.

Not the absolute shift in net charge is decisive, but the shift in the preferred orientation. Thus, the net charge is only a rough indication to assess the interaction. The orientations A and B are both influenced by aspartic acid, Asp46, Asp97, and Asp14.

Beta-lactoglobulin In the interaction map depicted in Figure 3.8, one binding orientation (A) can be identified. The strongest interaction orientation hardly changes with increasing pH. At pH 6.0 and 6.8 the net charge of beta-lactoglobulin was -1 and -2, respectively. The net charge decreases to -4 at pH 9.8. When looking at the strongest interaction from ligand point of view both charged residue types, aspartic acid and glutamic acid, are present but the protein is tilted towards aspartic acids Asp96 and Asp98. Ding et al. [46] showed in salt gradient elution runs on an anion-exchange column with the same functional ligand group as QFF, quaternary ammonium, that glutamic acid elutes shortly before aspartic acid. This finding confirms our suggestion of a stronger interaction between aspartic acid and the ligands than glutamic acid. This means that proteins with a high fraction of aspartic acid will bind stronger onto anion exchanger ligands. Not only the fraction but almost certainly also the charged residue density has a strong impact on binding strength and binding orientation. One aspartic acid on one side versus two glutamic acid residues on the other side will most probably result in a preferential orientation of glutamic acids closest to the ligand surface. Please note that all negative interaction energies mean binding behavior and only the ratio of energies to one another give a sense of general binding preference. In case of high protein concentration and low flow rate, the preferential binding orientation(s) will give way for more proteins, potentially with less favorable binding orientations.



(a) pH 5

(b) pH 9

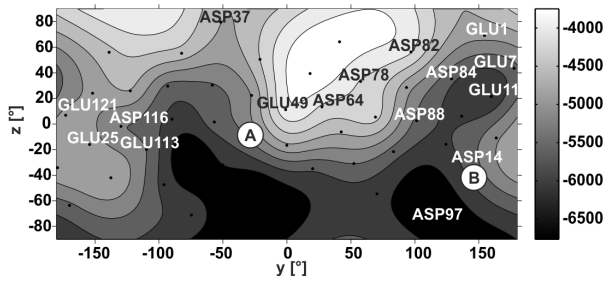
(c) pH 12

(d) Orientation A

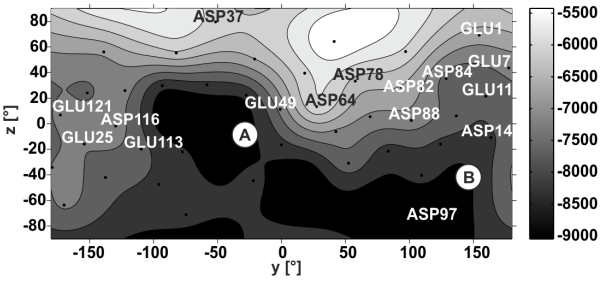
(e) Orientation B

(f) Orientation C

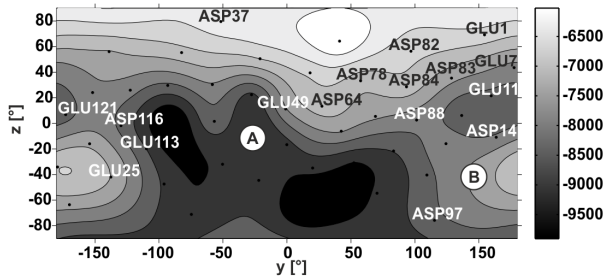
Figure 3.5: Lysozyme on SPFF: Coulomb interaction maps (energy in kJ/mol) on the left-hand side; Orientations A, B, and C on the right-hand side. Net charge for the pH values: 9, 8, and 3, respectively. Please note that the energy scales are different for the different interaction maps (left-hand side), due to the energetical differences within each interaction map. Positively charged residues were plotted into the interaction maps by labeling the according orientation in which the residue is closest to the ligands. The color of the labels only changes to achieve maximum readability and have no other indication. The color differences of the labeled residues of the three orientations on the right-hand side are from the ligand's point of view. The residues colored in red are lysines and green are arginines. The depiction is in perspective and therefore the labeling of the charged residues, which are further away from the ligands is smaller.



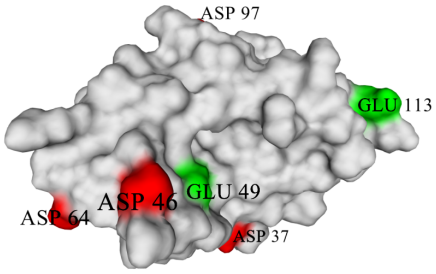
(a) pH 6.0



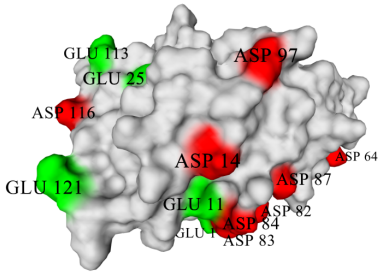
(b) pH 6.8



(c) pH 9.8

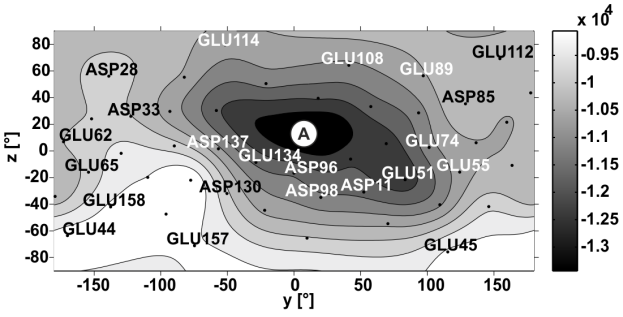


(d) Orientation A

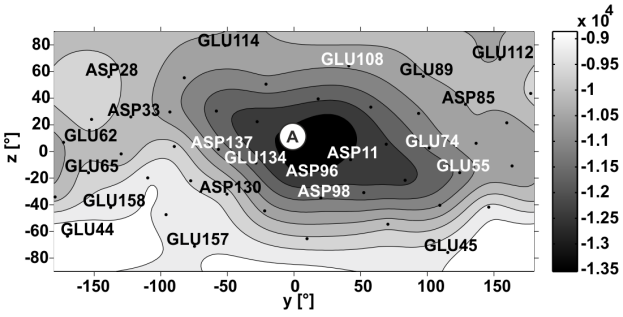


(e) Orientation B

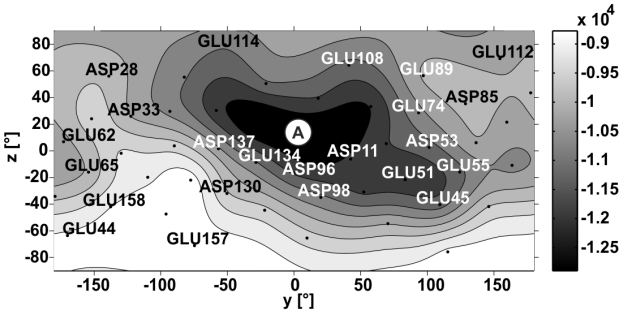
Figure 3.7: Alpha-lactalbumin on QFF: Coulomb interaction maps (energy in kJ/mol) on the left-hand side; Orientations A and B on the right-hand side. Net charge for the pH values: -1, -5, and -8, respectively. Please note that the energy scales are different for the different interaction maps (left-hand side), due to the energetical differences within each interaction map. Negatively charged residues were plotted into the interaction maps by labeling the according orientation in which the residue is closest to the ligands. The color of the labels only changes to achieve maximum readability and have no other indication. The color differences of the labeled residues of the two orientations on the right-hand side are from the ligand's point of view. The residues colored in red are aspartic acid and green are glutamic acid. The depiction is in perspective and therefore the labeling of the charged residues, which are further away from the ligands is smaller.



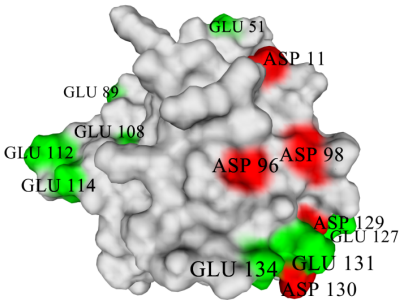
(a) pH 6.0



(b) pH 6.8



(c) pH 9.8



(d) Orientation A

Figure 3.8: Beta-lactoglobulin on QFF: Coulomb interaction maps (energy in kJ/mol) on the left-hand side; Orientation A on the right-hand side. Net charge for the pH values: -1, -2, and -4, respectively. Please note that the energy scales are different for the different interaction maps (left-hand side), due to the energetical differences within each interaction map. Negatively charged residues were plotted into the interaction maps by labeling the according orientation in which the residue is closest to the ligands. The color of the labels only changes to achieve maximum readability and have no other indication. The color differences of the labeled residues of the orientation on the right-hand side are from the ligand's point of view. The residues colored in red are aspartic acid and green are glutamic acid. The depiction is in perspective and therefore the labeling of the charged residues, which are further away from the ligands is smaller.

3.4. CONCLUSIONS

We have developed a comprehensive MD tool, which is able to simulate a wide range of biomolecules on chromatographic surfaces. The main goal of the current study was to determine if the MD tool us is capable to represent on-column behavior and to explore the many opportunities that come with this detailed view on molecular level: what influences binding. The tool was applied to anion and cation exchange adsorbers, namely Q Sepharose FF and SP Sepharose FF. Found interactions between proteins and Q Sepharose FF had a high correlation to experimental retention data. Preferred binding orientations for SP Sepharose FF were in good agreement with experimental results for binding orientation analysis previously published. It can be concluded that simulations yield similar results with regard to binding orientation and retention behavior as experimental results. Investigations on molecular level revealed interaction areas mostly defined by arginines in case of cation exchangers and aspartic acids in case of anion exchangers, which is also in good agreement with previous experimental results. Molecular dynamics reflects all influences on the binding, since the simulations are at atomic level. The tool is capable of predicting interactions and therewith retention behavior, as well as helping the user to gain insight into binding mechanisms on molecular level.

ACKNOWLEDGMENTS

We would like to thank Sven Amrhein, Thiemo Huuk, and Florian Pilgram from our work group for their helpful contribution in various sections of this work.

REFERENCES

[1] J. A. Asenjo and B. A. Andrews. Is there a rational method to purify proteins? From expert systems to proteomics. *J. Mol. Recognit. JMR*, 17(3):236–47, 2004.

[2] M. E. Lienqueo, A. Mahn, J. C. Salgado, and C. Shene. Mathematical Modeling of Protein Chromatograms. *Chem. Eng. Technol.*, 35(1):46–57, Jan. 2012.

[3] C. A. Brooks and S. M. Cramer. Steric mass-action ion exchange: Displacement profiles and induced salt gradients. *AIChE J.*, 38(12):1969–1978, Dec. 1992.

[4] J. C. Bosma and J. A. Wesselingh. pH dependence of ion-exchange equilibrium of proteins. *AIChE J.*, 44(11):2399–2409, Nov. 1998.

[5] J. C. Bosma and J. A. Wesselingh. Available area isotherm. *AIChE J.*, 50(4):848–853, Apr. 2004.

[6] A. Ladiwala, K. Rege, C. M. Breneman, and S. M. Cramer. A priori prediction of adsorption isotherm parameters and chromatographic behavior in ion-exchange systems. *Proc. Natl. Acad. Sci. U. S. A.*, 102(33):11710–5, Aug. 2005.

[7] G. Malmquist, U. H. Nilsson, M. Norrman, U. Skarp, M. Strömgren, and E. Carredano. Electrostatic calculations and quantitative protein retention models for ion exchange chromatography. *J. Chromatogr. A*, 1115(1-2):164–86, May 2006.

[8] C. M. Roth and A. M. Lenhoff. Electrostatic and van der Waals contributions to protein adsorption: computation of equilibrium constants. *Langmuir*, 9(3):962–972, Apr. 1993.

[9] J. Ståhlberg, B. Joensson, and C. Horvath. Theory for electrostatic interaction chromatography of proteins. *Anal. Chem.*, 63(17):1867–1874, Sept. 1991.

[10] J. Ståhlberg, B. Joensson, and C. Horvath. Combined effect of coulombic and van der Waals interactions in the chromatography of proteins. *Anal. Chem.*, 64(24):3118–3124, Dec. 1992.

[11] L. Zhang and Y. Sun. Molecular simulation of adsorption and its implications to protein chromatography: A review. *Biochem. Eng. J.*, 48(3):408–415, Feb. 2010.

[12] L. Zamolo, V. Busini, D. Moiani, D. Moscatelli, and C. Cavallotti. Molecular dynamic investigation of the interaction of supported affinity ligands with monoclonal antibodies. *Biotechnol. Prog.*, 24(3):527–39, 2008.

[13] J. Zheng, L. Li, H.-K. Tsao, Y.-J. Sheng, S. Chen, and S. Jiang. Strong repulsive forces between protein and oligo (ethylene glycol) self-assembled monolayers: a molecular simulation study. *Biophys. J.*, 89(1):158–66, July 2005.

[14] Y. He, J. Hower, S. Chen, M. T. Bernards, Y. Chang, and S. Jiang. Molecular simulation studies of protein interactions with zwitterionic phosphorylcholine self-assembled monolayers in the presence of water. *Langmuir*, 24(18):10358–64, Sept. 2008.

[15] J. C. Hower, Y. He, and S. Jiang. A molecular simulation study of methylated and hydroxyl sugar-based self-assembled monolayers: Surface hydration and resistance to protein adsorption. *J. Chem. Phys.*, 129(21):215101, Dec. 2008.

[16] A. N. Cormack, R. J. Lewis, and A. H. Goldstein. Computer Simulation of Protein Adsorption to a Material Surface in Aqueous Solution: Biomaterials Modeling of a Ternary System. *J. Phys. Chem. B*, 108(52):20408–20418, Dec. 2004.

[17] I. Yarovsky, M. T. W. Hearn, and M. I. Aguilar. Molecular Simulation of Peptide Interactions with an RP-HPLC Sorbent. *J. Phys. Chem. B*, 101(50):10962–10970, Dec. 1997.

[18] A. I. Liapis and B. A. Grimes. The coupling of the electrostatic potential with the transport and adsorption mechanisms in ion-exchange chromatography systems: Theory and experiments. *J. Sep. Sci.*, 28(15):1909–1926, Oct. 2005.

[19] K. Makrodimitris, D. L. Masica, E. T. Kim, and J. J. Gray. Structure prediction of protein-solid surface interactions reveals a molecular recognition motif of statherin for hydroxyapatite. *J. Am. Chem. Soc.*, 129(44):13713–22, Nov. 2007.

[20] V. Noinville, C. Vidal-Madjar, and B. Sebille. Modeling of Protein Adsorption on Polymer Surfaces. Computation of Adsorption Potential. *J. Phys. Chem.*, 99(5):1516–1522, Feb. 1995.

[21] S. Ravichandran, J. D. Madura, and J. Talbot. A Brownian Dynamics Study of the Initial Stages of Hen Egg-White Lysozyme Adsorption at a Solid Interface. *J. Phys. Chem. B*, 105 (17):3610–3613, May 2001.

[22] M. Agashe, V. Raut, S. J. Stuart, and R. A. Latour. Molecular Simulation To Characterize the Adsorption Behavior of a Fibrinogen γ -Chain Fragment. *Langmuir*, 9(9):1103–1117, 2005.

[23] F. Dismer and J. Hubbuch. 3D structure-based protein retention prediction for ion-exchange chromatography. *J. Chromatogr. A*, 1217(8):1343–53, Feb. 2010.

[24] H. Schmidt-Traub. *Preparative Chromatography of Fine Chemicals and Pharmaceutical Agents*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 1 edition, 2005.

[25] G. Carta and A. Jungbauer. *Protein chromatography: process development and scale-up*. Wiley-VCH-Verl., Weinheim, 2010. ISBN 9783527323012.

[26] J. Gutenwik, B. Nilsson, and A. Axelsson. Coupled diffusion and adsorption effects for multiple proteins in agarose gel. *AIChE J.*, 50(12):3006–3018, Dec. 2004.

[27] GE Healthcare. *Sepharose Fast Flow ion exchange media and prepacked column formats*, 2013. Product data sheet.

[28] A. R. Ubiera and G. Carta. Radiotracer measurements of protein mass transfer: kinetics in ion exchange media. *Biotechnol. J.*, 1(6):665–74, June 2006.

[29] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh Ewald method. *J. Chem. Phys.*, 103(19):8577, 1995.

[30] E. Krieger and G. Vriend. Models@Home: distributed computing in bioinformatics using a screensaver based approach. *Bioinformatics*, 18(2):315–8, Feb. 2002.

[31] A. Jakalian, D. B. Jack, and C. I. Bayly. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.*, 23 (16):1623–41, Dec. 2002.

[32] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and testing of a general amber force field. *J. Comput. Chem.*, 25(9):1157–74, July 2004.

[33] J. J. Stewart. MOPAC: A semiempirical molecular orbital program. *J. Comput.-Aided Mol. Des.*, 4(1):1–105, Mar. 1990.

[34] A. Klamt. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.*, 99(7):2224–2235, Feb. 1995.

[35] R. Apweiler, M. J. Martin, C. O'Donovan, et al. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.*, 40(D1):71–75, 2012.

[36] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Res.*, 28(1):235–42, Jan. 2000.

[37] F. Dismer, M. Petzold, and J. Hubbuch. Effects of ionic strength and mobile phase pH on the binding orientation of lysozyme on different ion-exchange adsorbents. *J. Chromatogr. A*, 1194(1):11–21, June 2008.

[38] A. Aguiar, G. H. de Haas, E. H. Jansen, A. J. Slotboom, and R. J. Williams. Proton-nuclear-magnetic-resonance/pH-titration studies of the histidines of pancreatic phospholipase A2. *Eur. J. Biochem.*, 100(2):511–8, Oct. 1979.

[39] J. Wang, P. Cieplak, and P. A. Kollman. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.*, 21(12):1049–1074, Sept. 2000.

[40] R. Anandakrishnan, B. Aguilar, and A. V. Onufriev. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.*, 40(Web Server issue):537–41, July 2012.

[41] GE Healthcare. *Ion Exchange Chromatography*, 2010. Product data sheet.

[42] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.*, 24(16):1999–2012, Dec. 2003.

[43] R. Sibson. A vector identity for the Dirichlet tessellation. *Math. Proc. Cambridge Philos. Soc.*, 87(01):151, Oct. 1980.

[44] K. Guruprasad, B. V. Reddy, and M. W. Pandit. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng.*, 4(2):155–61, Dec. 1990.

[45] J. R. Fromm, R. E. Hileman, E. E. Caldwell, J. M. Weiler, and R. J. Linhardt. Differences in the interaction of heparin with arginine and lysine and the importance of these basic amino acids in the binding of heparin to acidic fibroblast growth factor. *Arch. Biochem. Biophys.*, 323(2):279–87, Nov. 1995.

[46] Y. Ding, H. Yu, and S. Mou. Direct determination of free amino acids and sugars in green tea by anion-exchange chromatography with integrated pulsed amperometric detection. *J. Chromatogr. A*, 982(2):237–44, Dec. 2002.

Chapter 4

Custom-tailored Adsorbers: A Molecular Dynamics Study on Optimal Design of Ion Exchange Chromatography Material

Katharina M. H. Lang, Jörg Kittelmann, Florian Pilgram, Anna Osberghaus, Jürgen Hubbuch

The performance of functionalized materials, e.g., ion exchange resins, depends on multiple resin characteristics, such as type of ligand, ligand density, the pore accessibility for a molecule, and backbone characteristics. Therefore, the screening and identification process for optimal resin characteristics for separation is very time and material consuming. Previous studies on the influence of resin characteristics have focused on an experimental approach and to a lesser extent on the mechanistic understanding of the adsorption mechanism. In this in silico study, a previously developed MD tool is used, which simulates any given biomolecule on resins with varying ligand densities. We describe a set of simulations and experiments with four proteins and six resins varying in ligand density, and show that simulations and experiments correlate well in a wide range of ligand density. With this new approach simulations can be used as pre-experimental screening for optimal adsorber characteristics, reducing the actual number of screening experiments, which results in a faster and more knowledge-based development of custom-tailored adsorbers.

published in Journal of Chromatography A, Volume 1413, 25 September 2015, Pages 60-67

4.1. INTRODUCTION

Ion exchange chromatography of proteins is a standard technique in biotechnological processes, e.g., antibody purification and analytical biochemistry. One of the main advantages of this unit operation is that it maintains the biological activity of the bioproduct [1]. The underlying principle of ion exchange chromatography is the electrostatic interaction of biomolecules to a stationary phase on which charged ligands are grafted on. These ligands can vary in their density on the surface, strongly influencing the selectivity of the chromatographic material.

Various studies on the influence of ligand density on binding behavior have been conducted in the past. Wu and Walters [2] found in a study on ion exchange isocratic elution experiments with two proteins that the Z number, which describes the number of interaction binding sites of the protein with the ligand surface, increased as the ligand density increased. This behavior is comprehensible from a molecular point of view, since much more binding sites of the protein are within the reach of (electrostatic) interaction with the ligands. Hardin et al. [3] conducted experiments with a monoclonal antibody, which was eluted from an agarose based strong cation exchanger with varying ligand densities. The authors showed that the variation of the ligand densities only had subtle effects on the dynamic binding capacity, however, affected the critical conductivity. In contrast, Franke et al. [4] showed in experiments with a strong cation exchanger resin with varying ligand densities that the ligand density has a significant effect on the dynamic binding capacity. The dynamic binding capacity increased with increasing ligand density until a certain critical ligand density was reached and decreased from there on. The decrease above the critical density is explained by a high number of ligands that result in smaller pores. Similar results were obtained by Fogle et al. [5] who showed that the dynamic binding capacity for three tested monoclonal antibodies increased with ligand density. It has been shown that the relationship between ligand density, retention, and resolution are affected by the protein's characteristic charge and surface charge distribution [5]. In a consecutive work by Fogle and Persson [6], the authors claimed that the ligand density has the potential to affect protein elution profiles, which lead to changes in purification efficiency.

Overall, these studies outline that the ligand density is a factor that strongly influences the interaction between protein and adsorber surface. While some research has been carried out experimentally, no studies are available, which investigate the influence of ligand density on protein retention on a molecular level. To gain this understanding, we have conducted several experiments - in a laboratory and simulational manner.

In this study, a MD simulation tool, developed by our work group [7], was used to calculate interaction energy profiles for model proteins on an adsorber surface with different ionic capacities, corresponding to the various adsorber prototypes. The ionic capacity corresponds to ligand density. For the simulation setup, the ligands were placed in a certain distance towards each other. The ligand distances were determined based on ionic capacity values in Monte Carlo calculations.

To correlate the MD data to experimental data, the laboratory experiments were carried out as standard gradient elution experiments with adsorber prototypes with varying ionic capacities. Retention factors have been determined for model proteins on sulfopropyl based adsorbers with agarose backbone featuring different ionic capacities.

Experimental results from column experiments could be correlated to MD simulation interaction energy profiles. It is shown that the developed MD tool can describe binding behavior sensitive to the ligand distance. When using this method as a pre-experimental screening, it provides the means to reduce experimental screening efforts and early stage sample consumption in process design. Also, the *in silico* screening for optimal ligand densities for specific purification tasks makes way for tailor-made adsorbers, specific for the purification task at hand. Please note that the ionic capacity of the adsorbent is proportional to the density of the charged ligands and therefore will be handled as synonyms in this work.

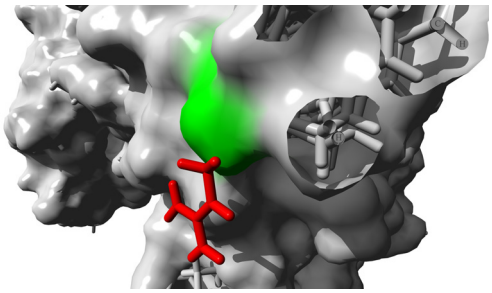
4.2. MATERIAL AND METHODS

4.2.1. PROTEINS, RESINS, AND CHEMICALS

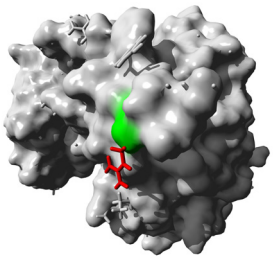
Proteins The proteins chosen for simulations and experiments were Cytochrome C, Lysozyme, Ribonuclease A, and Thaumatin. All proteins were purchased from Sigma-Aldrich (St. Louis, MO, USA). Table 4.1 lists the corresponding 3D structures in PDB format and Sigma-Aldrich article numbers.

Table 4.1: Proteins with PDB ID

Protein	PDB ID	Sigma-Aldrich no.	Net charge at pH 7
Lysozyme C	2VB1 [8]	L6876	8
Cytochrome C	1HRC [9]	30398	8
Ribonuclease A	1KF5 [10]	83833	6
Thaumatin-1	2VHK [11]	T7638	5
Thaumatin-2	3AOK [12]		5



(a) Zoom in: The surface corresponds to Thaumatin-1, Arginine from Thaumatin-2 clearly exceeds the surface of Thaumatin-1



(b) Zoom Out: the residues are on a pole of the protein and might have an influence on binding

Figure 4.1: Residue 63 for aligned Thaumatin-1 and Thaumatin-2: Arginine of Thaumatin- 2 is depicted in red; Serine of Thaumatin-1 is depicted in green

Adsorber material Six prototypes with varying ionic capacities of sulfopropyl ligands on a cross-linked agarose backbone were used (see Table 4.2). The adsorber prototypes were variants of GE Healthcare's SP Sepharose^{TM} Fast Flow (FF) and therefore consisted of polymer sepharose beads on which sulfopropyl ligands were grafted on. All adsorbent prototypes were packed in Tricorn 5/20 columns purchased from GE Healthcare (Uppsala, Sweden).

Buffers Loading buffer for gradient elution experiments was 20 mM phosphate buffer titrated to pH 7 and elution buffer consisted of loading buffer with 0.5 M NaCl added (all purchased from Merck KGaA, Darmstadt, Germany). Buffer preparation was conducted at room temperature.

4.2.2. COLUMN CHARACTERIZATION

DETERMINATION OF BED HEIGHTS

The bed heights of the packed columns were determined with a vernier caliper. The diameter of the columns is specified by the manufacturer with 5 mm.

DETERMINATION OF POROSITIES AND DEAD VOLUME

Pulse experiments were conducted to determine the total porosity, the interstitial porosity, and the particle porosity of the columns. As a non-binding but pore penetrating tracer, 1% acetone solution was used (purchased from Merck KGaA). The retention volume of acetone corresponds to the total available volume. Dextran with a chain length of 2,000 kDa

with a concentration of 1 mg/mL in loading buffer was used as a non-pore penetrating and non-binding tracer (purchased from Sigma-Aldrich). The retention volume of Dextran corresponds to the interstitial volume.

All experiments were performed on an ÄKTApurifier^{TM} from GE Healthcare. The column was equilibrated with loading buffer. Subsequently, 25 µL of acetone solution was injected using an autosampler. The volume until the acetone peak reached its maximum was recorded as the retention volume. A Dextran peak was recorded as well in the same manner. The dead volume detection was run with the same experimental setup without a column. All measurements were repeated in triplicates to assure reproducibility.

4.2.3. GRADIENT ELUTION EXPERIMENTS

The proteins were dissolved in loading buffer to a concentration of 0.2 mM.

The packed columns were equilibrated with loading buffer for 5 CVs. 25 µL of protein was injected with an autosampler onto the column. Unadsorbed protein was washed off the resin with 2 CVs of loading buffer. Adsorbed protein was eluted using a gradient length of 10 CVs with elution buffer, followed by 5 CVs of 100 % elution buffer for regeneration of the resin.

Protein sample injection was performed with an autosampler A-905. The measurements of absorbance and conductivity were carried out with an ÄKTA ^{TM} Monitor UV-900 and Monitor pH/C-900. The post-experiment calculations were run with the UNICORN^{TM} software (all purchased from GE Healthcare).

Each experiment was run in triplicate to assure reproducibility and averages of the retention values are reported. All experiments were performed at room temperature.

4.2.4. LIGAND DISTANCE CALCULATION FROM IONIC CAPACITY VALUES

The method to calculate the ligand distance for the MD simulations has been introduced in a previous paper [7]. For the calculation, a 10,000 sample Monte Carlo (MC) analysis was performed in MATLAB [13]. All physical properties of the adsorber beads were assumed to be Gaussian distributed. Resulting ligand distances are shown in Table 4.2.

4.2.5. MOLECULAR DYNAMICS SIMULATIONS

PROTEIN AND ADSORBER PREPARATION

The proteins' UniProt ID were selected in the UniProtKB [14] by organism and revision status. The protein length was looked up in the sequence annotation section of the UniProtKB entry and subsequently a corresponding PDB ID was selected by (protein) completeness and best resolution. The file with the highest Z-score was imported from the databases Research Collaboratory for Structural Bioinformatics (RCSB) [15] and PDB_Redo [16] directly into YASARA [17]. The Z-score describes the statistical significance relative to an alignment of random structures, a widely established parameter for structure quality assessment.

The protein's 3D structure was then cleaned (adding missing bonds, rebuilding side chains with missing atoms) and crystal water was removed. The hydrogen bonding network was optimized and pH was set to 7. An energy minimization with a cutoff of 7.86 Å and the force field AMBER99 [18] was run until the energy changed by less than 0.05 kJ/mol per atom during 200 steps. Afterwards, a molecular dynamics simulation was run for 50 ps to equilibrate the protein structure and all protons were deleted. The pH dependent protonation was calculated online with the H++ tool [19] from VirginiaTech. For Cytochrome C, the heme group was removed before processing in H++ and manually added afterwards because H++ is unable to process non- standard residues like a heme group.

Thaumatin from Sigma-Aldrich is a mixture of Thaumatin-1 and Thaumatin-2, which corresponds to two 3D structures (2VHK and 3AOK) for the simulations. Both proteins are 207 residues in length and are positively charged at pH 7 with a net charge of +5. The sequence alignment program Clustal X version 2.0 [20] was used to identify sequence differences. The two proteins differ in five residues at positions 46, 63, 68, 87, and 113 and therewith 97.58 % of them is identical. The root mean square deviation is 0.335 Å, which is negligible. The strongest deviation is residue 63 (depicted in Figure 4.1), where Thaumatin-1 has a Serine whereas Thaumatin-2 has an Arginine, which is positively charged as opposed to Serine and additionally a very flexible residue. As shown in a previous work [7], Arginine has a strong influence on the binding behavior of the protein in anion exchange adsorption. This may result in a different interaction with the adsorber and therefore needs particular attention in the evaluation of the simulations.

The 3D ligand structure was built in YASARA based on the chemical composition of the ligand, taken from the GE Healthcare website. The ligands were replicated and placed at equal intervals on the plane in accordance to the result of the calculated distance (see Section 4.3.1 and Table 4.2). The adsorbent matrix, to which the ligands are bound, has been simplified by

using a dummy C atom per ligand. This has been shown to be a viable method to stabilize ligands on constructed surfaces [7]. The size of the area depends on the largest dimension of the protein to reduce computational effort and to give all proteins equal starting conditions.

EQUILIBRATION AND MOLECULAR DYNAMICS SIMULATIONS

Molecular dynamic simulations were carried out on 8 processors on the InstitutsCluster II (IC2) system at Karlsruhe Institute of Technology, Germany, using the molecular dynamics software YASARA version 13.2.2 [17]. The force fields used were GAFF [21] (for ligands), Amber99 [18] (for protein minimization), and Amber03 [22] (for protein simulation). Cytochrome C, Lysozyme C, Ribonuclease A, and both Thaumatin structures were simulated on an adsorber surface, which was twice the size of the protein in its largest dimension. The MD simulations were computed as described in a previous paper [7], by simulating 50 different orientations of the protein on the ligand surface with a distance of 5 Å between protein and ligands. Each orientation was heated up from 0 K to 300 K in increments of 1 K to reach a stable state (equilibration step) and subsequently simulated at 300 K for 50 ps.

The calculated energies (total, bond, angle, dihedral, planarity, Coulomb, van der Waals) derived from the simulation were each averaged and Boltzmann weighted over the entire simulation, starting at 5 ps to 50 ps. The median of the Boltzmann weighted average of all orientations was calculated and used as the total energy of the ensemble (= protein + surface + solvent). Post-simulation calculations were performed with a YASARA analysis macro and MATLAB.

4.2.6. REGRESSION MODEL

A linear regression model was developed to attain a correlation between simulation and experimental retention values using MATLAB. The input factors were the binding influencing variables ionic capacity, protein net charge, and Coulomb interaction energy, which is the simulation model response variable. The input variables' values were centered to have a mean of zero and scaled to have a standard deviation of one, since they deviate strongly in magnitude. Analysis of coefficients of the regression model enables the analysis for the strongest influencing variables, which leads to more mechanistic understanding of the binding influencing factors.

Table 4.2: Ionic capacities of sulfopropyl prototypes

Adsorbers	IC [µmol/ml]	Mean ligand distance [Å]	Particle porosity
PT1	254	6.51	0.79
PT2	169	7.98	0.82
PT3	131	9.06	0.86
PT4	107	10.00	0.86
PT5	84	11.30	0.89
PT6	73	12.14	0.85

4.3. RESULTS AND DISCUSSION

4.3.1. MONTE CARLO CALCULATIONS OF LIGAND DISTANCE FROM IONIC CAPACITY

To obtain ligand distances corresponding to the respective ionic capacity of the adsorbers, MC simulations have been conducted, as described in Section 4.2.4. The Monte Carlo simulations resulted in a distribution of ligand distances for each adsorber variant. An exemplary histogram is depicted in Figure 4.2. The mean value of each adsorber variant was taken as ligand distance for the simulations (see Table 4.2).
To illustrate the substantial difference in ligand density, Figure 4.3 shows Lysozyme C on two different prototypes: on PT1 (Figure 4.3a) with a very high ionic capacity of 254 µmol/mL and on PT6 with the lowest ionic capacity (73 µmol/mL) and therewith the lowest ligand density. As a comparison: commercially available SP Sepharose FF from GE Healthcare has a specified ionic capacity range of 180–250 µmol/mL. It can be seen in Figure 4.3 that the protein has much more ligands to interact with on an adsorber with a high ionic capacity and high ligand density than on an adsorber with a lower ligand density. This leads to a stronger interaction between protein and ligands and consequently to more negative Coulomb interaction energies, indicating stronger binding.

4.3.2. EXPERIMENTAL RESULTS

DETERMINATION OF POROSITY

To investigate the influence of ionic capacity on the adsorber's porosity, the packed columns were characterized as described in Section 4.2.2. The porosity was calculated with acetone and Dextran retention values and the measured bed height of the individual columns. A correlation of decreasing particle porosity with rising ionic capacity could be established (see Figure4.4). More detailed interpretation of results should be conducted with care. All

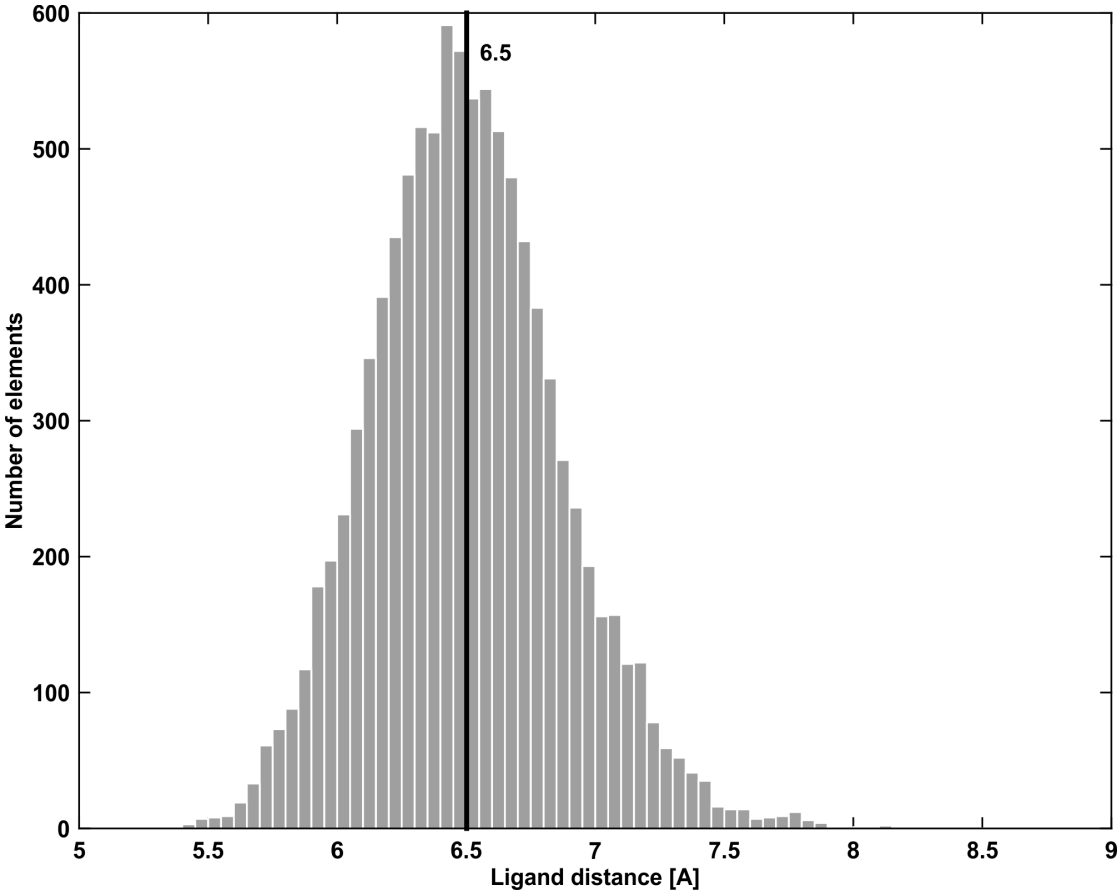


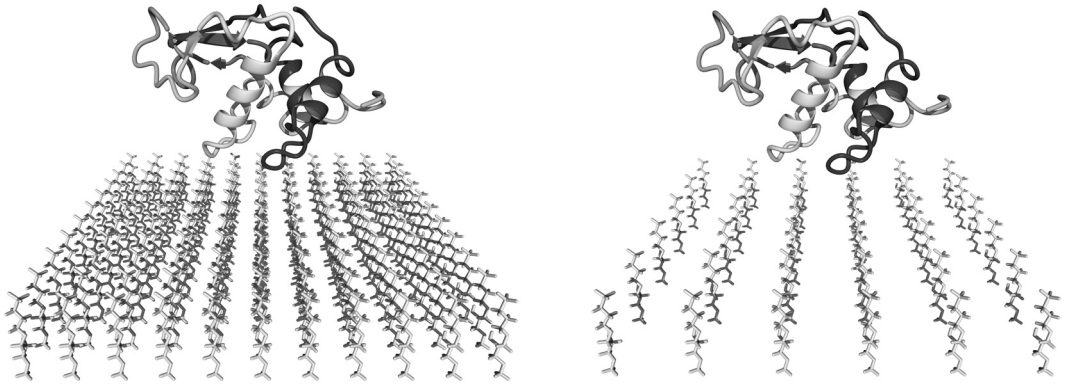
Figure 4.2: Histogram of Monte Carlo experiments to calculate the ligand distance for prototype PT1, vertical mark: mean value in Å

columns were packed and compressed manually, thus column-to-column fluctuations are self-evident.

These findings are consistent with those by Hardin et al. [3] and Franke et al. [4], who showed that increasing the ionic capacity results in steric exclusion of molecules, which corresponds to a lower porosity.

GRADIENT ELUTION EXPERIMENTS

To correlate experimental and simulation data, gradient elution experiments were conducted as described in Section 4.2.3. A correlation between the retention of proteins and the ionic capacity is plotted in Figure 4.5. The vertical lines in this figure represent the optimal ionic capacity for separation of the proteins. The values next to the lines are the exact ionic capacity, which is interpolated in-between points. Its position in y-dimension is in-between two



(a) Ligand distance 6.51 Å, the total length in one dimension is 65.1 Å

(b) Ligand distance 12.14 Å, the total length in one dimension is 60.7 Å

Figure 4.3: Lysozyme on two different adsorber variants

neighboring eluting proteins, which are therefore critical for separation. A positive correlation was found between the retention factor and ionic capacity for all proteins in the ionic capacity range from 73 to 169 µmol/mL. After this point, a flattening of the retention factor curve can be seen, suggesting that an increase of the ionic capacity (and ligand density) after this point has no beneficial effect on the separation result. This is an advantageous outcome for manufacturing, since a reduction in material consumption can be achieved while maintaining the same separation effect. As presented in Section 4.3.2, the intraparticle porosity decreases with increasing ionic capacity. This is a result of steric hindrance due to a high ligand density in the pores and clogging of the diffusive pores by ligands (compare References [3] and [4]). Since Lysozyme C is the largest protein in our set with a radius of gyration of 29.75 Å (measured in YASARA), this could be an explanation for the decrease of its retention factor above an ionic capacity of 169 µmol/mL. If the protein is too large for the clogged pore, it cannot diffuse into the pore and therefore elutes early in the gradient. This suggests that the curves of the other proteins show similar behavior at higher ionic capacities. An ionic capacity of 287 µmol/mL (PT7) corresponds to a ligand distance of 6.12 Å and previous simulation setups resulted in steric clashes for ligand distances below 5 Å, therefore a much higher ionic capacity than 287 µmol/mL is impossible to reach theoretically. This needs to be subject for further investigations. The results of Hardin et al. [3] suggested that increasing the ligand density (i.e., ionic capacity) above a critical value will not improve resin binding capacity, as the constant retention values for ionic capacities of 169 µmol/mL and above confirm.

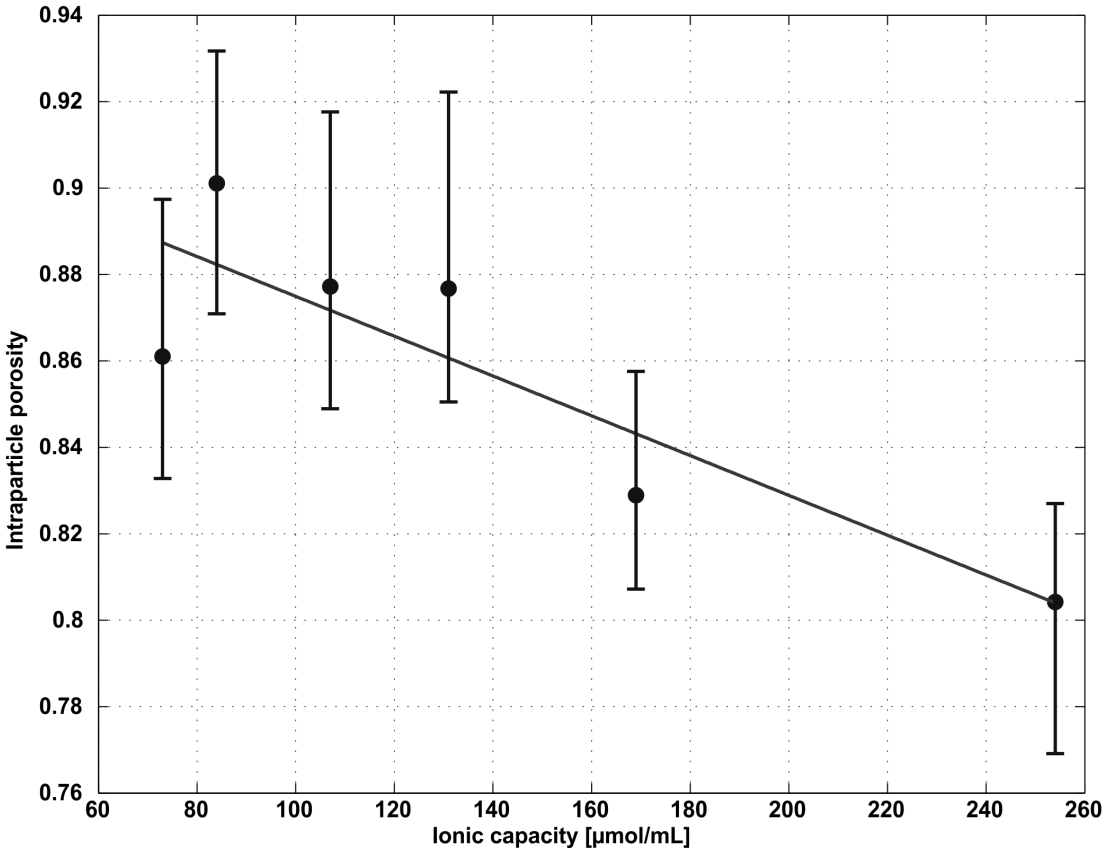


Figure 4.4: Intraparticle porosity vs. ionic capacity; bed height measurements of the columns were conducted in 10- fold replicates; the bars represent the measurement range

4.3.3. CORRELATION OF SIMULATION WITH EXPERIMENTS

The correlation was conducted to generate a model that describes protein retention based upon simulation data. Figure 4.6a shows sloping curves of the retention factor as a function of the Coulomb interaction energy for all proteins. The experimentally determined retention behavior above 169 µmol/mL (compare Section 4.3.2) corresponds to the constant region between the first and second data point for each curve in this figure. The constant region is most probable a result of inaccessible pore volume above a critical ionic capacity, as discussed in Section 4.3.2. The region less than 169 µmol/mL shows increasing retention factors with increasing ionic capacity and can be modeled well (see Figure 4.6b).

The linear regression of data for all proteins and adsorbers in the linear region produces a regression coefficient of 0.97. The strongest influencing factors on the regression model are the protein net charge with a coefficient weight value of 0.92 and Coulomb interaction en-

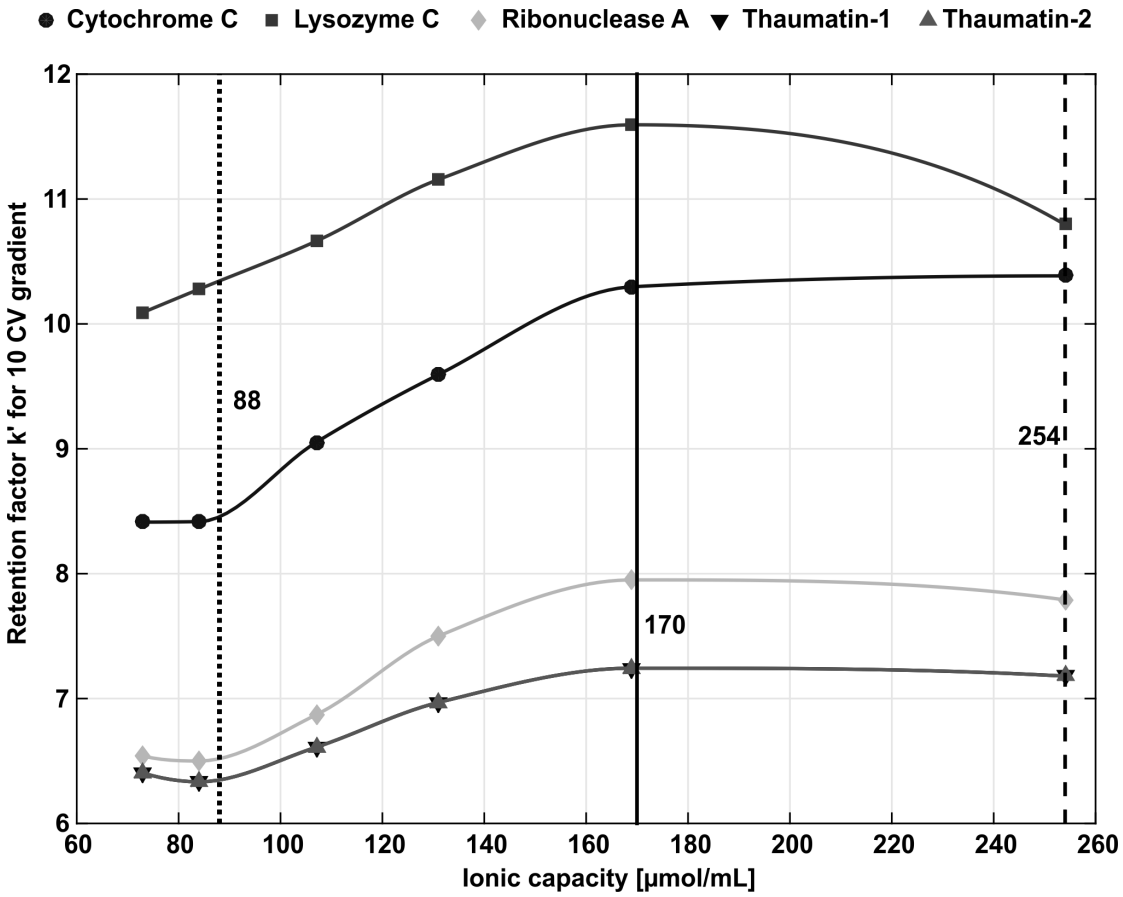
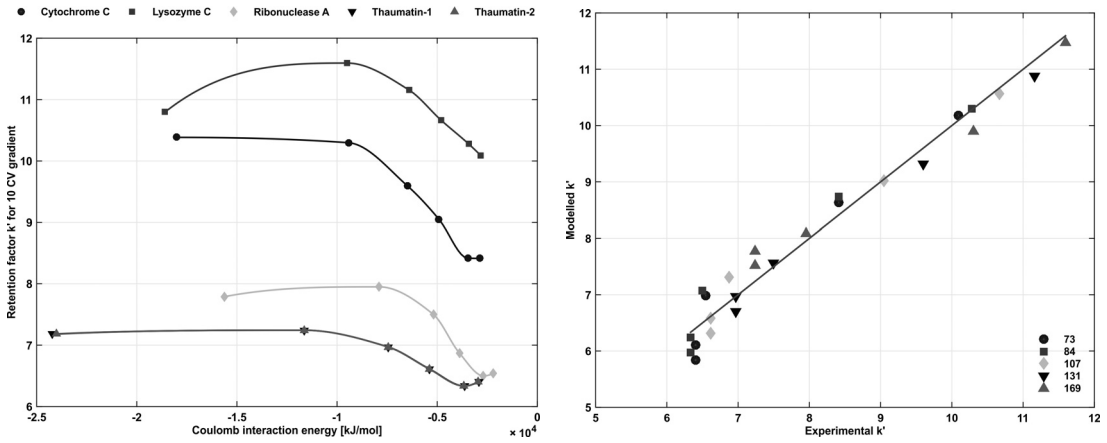


Figure 4.5: Retention factor k' vs. ionic capacity; the vertical dotted line represents the optimal ionic capacity for the separation of Lysozyme C and Cytochrome C; the vertical solid line represents the optimal ionic capacity for the separation of Ribonuclease A and Thau matin; the vertical dashed line represents the optimal ionic capacity for the separation of Cytochrome C and Ribonuclease A

ergy with a coefficient weight value of -0.29 on the retention factor k' as model response (Table 4.3).

The coefficient weight quantities are comprehensible, since the net charge contributes strongly to the electrostatic interaction and therefore influences the Coulomb energy. This result is corroborated by the findings of Dismer and Hubbuch [23] who showed in a regression model of protein retention that the net charge alone allows modeling of protein retention in the limitation to one adsorber. MD simulation based refinement, which includes the electrostatic interaction energy results in a better modeling result, but more importantly the simulations allow for the evaluation of different ligand densities.



(a) Retention factor k' for the proteins plotted over the Coulomb interaction energy

(b) Regression model of retention factor k' grouped by ionic capacity in µmol/mL; Regression coefficient $R^2 = 0.97$

Figure 4.6: Correlation of simulation with experiments

Table 4.3: Coefficient Weights of Regression Model Variables

Variable	Coefficient weight
Ionic capacity	0.02
Protein net charge	0.92
Coulomb interaction energy	−0.29

4.3.4. EXCURSUS: DIFFERENCES IN SIMULATION OF THAUMATIN VARIANTS

The differences between the two Thaumatin structures were elaborated in Section 4.2.1. The 3D structures of Thaumatin-1 and Thauma tin-2 differed in five residues and therefore were simulated separately. Even though the net charge of both Thaumatins are identical, Arginine 63 of Thaumatin-2 has a strong influence on the binding interaction [7]. Consequently, the results of the MD simulations of both proteins were investigated particularly intensive as presented in Figure 4.7. The adsorber on which the MD simulations of the two protein structures generated the highest energy deviation from each other was investigated in detail. This was PT5 with an ionic capacity of 84 µmol/mL. The Coulomb interaction energy difference was 33 kJ/mol. The differences were modeled with the previously presented regression model and plotted into the experimental chromatogram (Figure 4.7). It can be seen that both, simulated and modeled, k' deviate slightly from the actual peak maximum. The peak is a sum signal of both Thaumatins. The modeled retention factors would lead to a peak maximum, which is positioned 0.13 mL to the right of the actual experimental peak. These

results highlight that the structural differences of the Thaumatin variants cannot be resolved in experiments on this adsorber, as no peak separation can be detected under applied conditions, while simulations can depict the slightest structural changes. Further investigations with experiments on an analytical column with a higher resolution should be conducted.

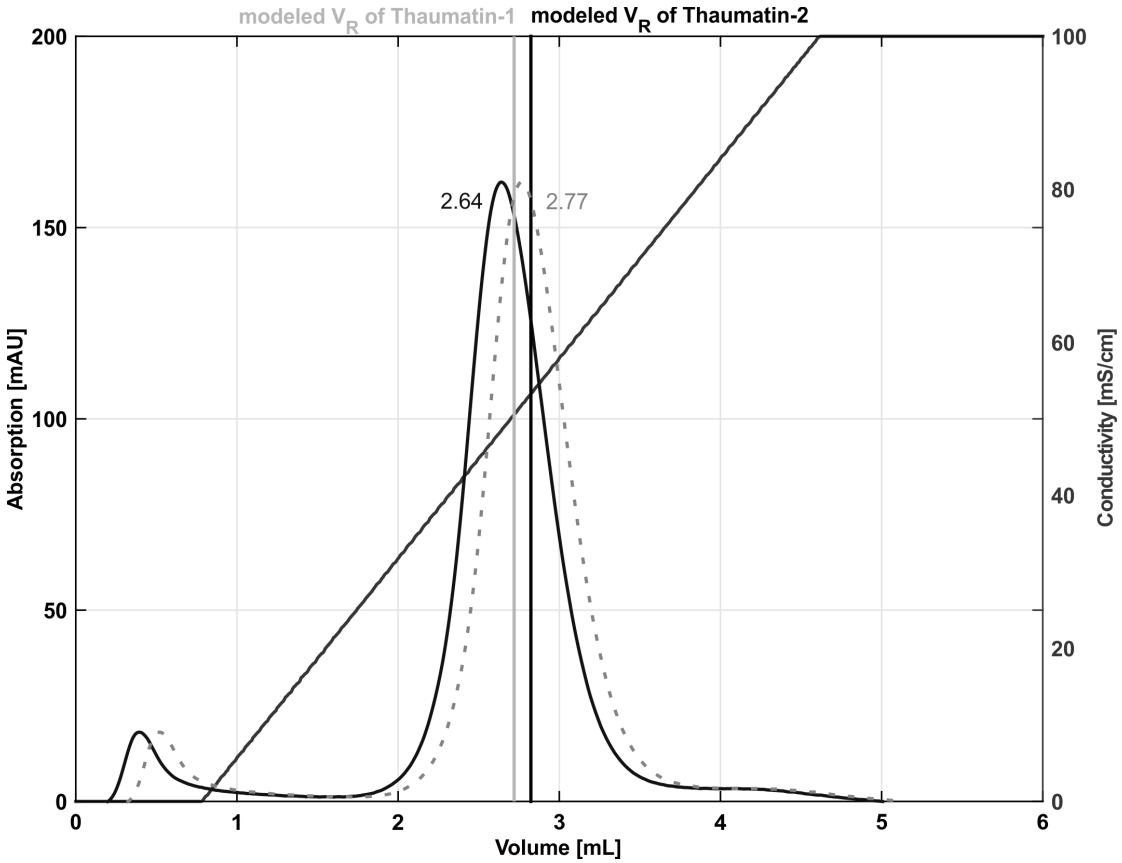


Figure 4.7: Chromatogram of Thaumatin on PT5. The vertical lines represent the modeled retention volume of Thauma tin-1 and Thaumatin-2 respectively. The dotted chromatogram is the modeled response data.

4.4. CONCLUSIONS

The main goal of the current study was to evaluate the potential of a previously developed MD tool in modeling protein binding behavior on adsorbers with varying ligand densities. This challenge was tackled with a twin-tracked strategy: laboratory gradient elution experiments of proteins on adsorbers with varying ligand density and molecular dynamic simulations thereof. This study has confirmed that the ligand density has a strong influence on the binding behavior, which can be seen in MD simulations as well as in experiments. The

following conclusions can be drawn from the present study: MD simulations are capable of describing differing retention behavior due to varying ligand densities and can therefore be used in customizing adsorbent material and troubleshooting challenging purification tasks. The results of this research support the idea that other physical characteristics of the adsorber apart from the ligand density (i.e., particle diameter, pore diameter, and available adsorber area) also influence the retention behavior for high ligand densities. To achieve optimal results with the presented method a twin tracked strategy must be run of selected small scale experiments to calibrate MD model, which subsequently identifies optimal adsorbent characteristics in terms of ionic capacity.

ACKNOWLEDGEMENTS

We would like to thank GE Healthcare for kindly providing samples of cation exchange prototype adsorber material used in this study. Also, we would like to thank Jörg Kittelmann from our work group for his helpful contribution to various sections of this work.

REFERENCES

[1] J. Ståhlberg. Retention models for ions in chromatography. *J. Chromatogr. A*, 855(1): 3–55, Sept. 1999.

[2] D. Wu and R. R. Walters. Effects of stationary phase ligand density on high-performance ion-exchange chromatography of proteins. *J. Chromatogr. A*, 598(1):7–13, May 1992.

[3] A. M. Hardin, C. Harinarayan, G. Malmquist, A. Axén, and R. van Reis. Ion exchange chromatography of monoclonal antibodies: effect of resin ligand density on dynamic binding capacity. *J. Chromatogr. A*, 1216(20):4366–71, May 2009.

[4] A. Franke, N. Forrer, A. Butté, B. Cvijetić, M. Morbidelli, M. Jöhnck, and M. Schulte. Role of the ligand density in cation exchange materials for the purification of proteins. *J. Chromatogr. A*, 1217(15):2216–25, Apr. 2010.

[5] J. Fogle, N. Mohan, E. Cheung, and J. Persson. Effects of resin ligand density on yield and impurity clearance in preparative cation exchange chromatography. i. mechanistic evaluation. *J. Chromatogr. A*, 1225:62–9, Feb. 2012.

[6] J. Fogle and J. Persson. Effects of resin ligand density on yield and impurity clearance in preparative cation exchange chromatography. ii. process characterization. *J. Chromatogr. A*, 1225:70–8, Feb. 2012.

[7] K. M. H. Lang, J. Kittelmann, C. Dürr, A. Osberghaus, and J. Hubbuch. A comprehensive molecular dynamics approach to protein retention modeling in ion exchange chromatography. *J. Chromatogr. A*, 1381:184–193, Feb. 2015.

[8] J. Wang, M. Dauter, R. Alkire, A. Joachimiak, and Z. Dauter. Triclinic lysozyme at 0.65 å resolution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 63(12):1254–1268, 2007.

[9] G. W. Bushnell, G. V. Louie, and G. D. Brayer. High-resolution three-dimensional structure of horse heart cytochrome c. *J. Mol. Biol.*, 214(2):585–95, July 1990.

[10] R. Berisio, F. Sica, V. S. Lamzin, K. S. Wilson, A. Zagari, and L. Mazzarella. Atomic resolution structures of ribonuclease a at six pH values. *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 58(3):441–450, Feb. 2002.

[11] N. Asherie, J. Jakoncic, C. Ginsberg, A. Greenbaum, V. Stojanoff, B. J. Hrnjez, S. Blass, and J. Berger. Tartrate chirality determines thaumatin crystal habit. *Cryst. Growth Des.*, 9(9): 4189–4198, Sept. 2009.

[12] T. Masuda, K. Ohta, F. Tani, B. Mikami, and N. Kitabatake. Crystal structure of the sweet-tasting protein thaumatin II at 1.27 å. *Biochem. Biophys. Res. Commun.*, 410(3):457–60, July 2011.

[13] MATLAB. version 8.0.0.783 (R2012b), 2012.

[14] R. Apweiler, M. J. Martin, C. O'Donovan, et al. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.*, 40(D1):71–75, 2012.

[15] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Res.*, 28(1):235–42, Jan. 2000.

[16] R. P. Joosten, K. Joosten, G. N. Murshudov, and A. Perrakis. PDB_REDO: constructive validation, more than just looking for errors. *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 68(Pt 4):484–96, Apr. 2012.

[17] E. Krieger, T. Darden, S. B. Nabuurs, A. Finkelstein, and G. Vriend. Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins*, 57 (4):678–83, Dec. 2004.

[18] J. Wang, P. Cieplak, and P. A. Kollman. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.*, 21(12):1049–1074, Sept. 2000.

[19] R. Anandakrishnan, B. Aguilar, and A. V. Onufriev. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.*, 40(Web Server issue):537–41, July 2012.

[20] M. Larkin, G. Blackshields, N. Brown, R. Chenna, P. McGettigan, H. McWilliam, F. Valentin, I. Wallace, A. Wilm, R. Lopez, J. Thompson, T. Gibson, and D. Higgins. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21):2947–2948, 2007.

[21] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and testing of a general amber force field. *J. Comput. Chem.*, 25(9):1157–74, July 2004.

[22] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.*, 24(16):1999–2012, Dec. 2003.

[23] F. Dismer and J. Hubbuch. 3D structure-based protein retention prediction for ion-exchange chromatography. *J. Chromatogr. A*, 1217(8):1343–53, Feb. 2010.

Chapter 5

A Theoretical Study on Binding Behavior of Single Domain Antibodies to Affinity Chromatography Media using Molecular Dynamics Simulations

Katharina M. H. Lang, Esther M. Volz, Anna Osberghaus, Jürgen Hubbuch

Single-domain antibodies (sdAbs) are promising candidates for biopharmaceutical and research applications due to their advantageous physicochemical properties. An important application in research is the usage as affinity ligands in chromatography and therefore the immobilization of the single-domain antibodies on chromatography media is crucial. To gain a better understanding and possibly even influence the coupling, we present a theoretical study on potential coupling sites via an in silico approach. Four single-domain antibody structures were simulated by MD simulations on a surface made up of N-hydroxysuccinimide (NHS) ligands. Through the evaluation of energy profiles by means of interaction maps, energetically most favored orientation states could be identified and a presumption on coupling orientation could be made. In one case, a problematic coupling reaction observed in experiments was reproduced and the issue was identified. Our tool was capable to simulate the interaction between single-domain antibodies and affinity coupling ligands through the use of MD simulations and therefore is a powerful tool to gain more mechanistic understanding on molecular level.

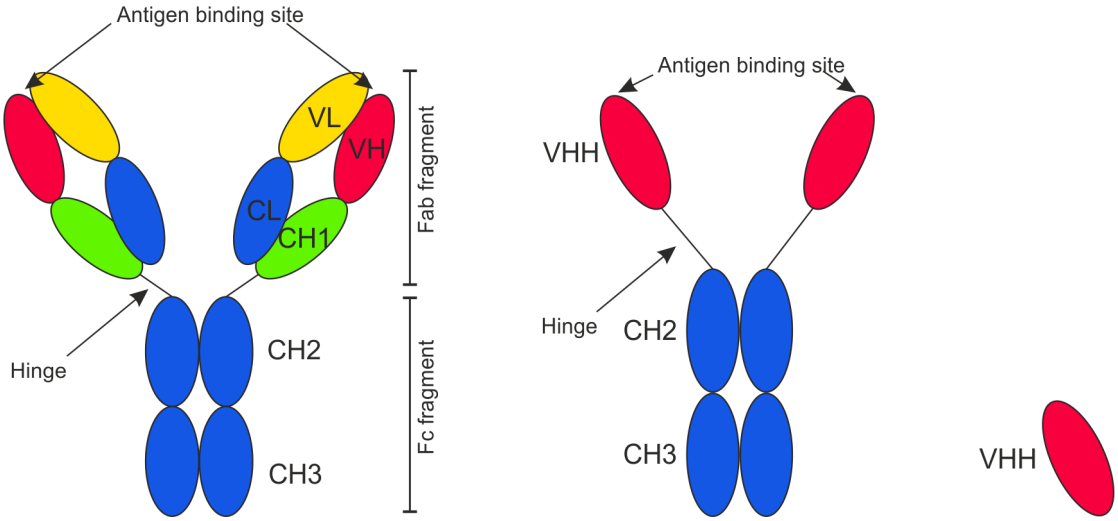
in preparation

5.1. INTRODUCTION

sdAbs, also referred to as Nanobodies® (term introduced by Ablynx) are antibody fragments that are comprised of a single monomeric variable domain of an antibody (Figure 5.1a). They are produced by specimen of the Camelid family [1] and cartilaginous fish, such as sharks [2]. sdAbs are the smallest antibody fragments with a molecular mass of about 12-15 kDa that are capable of antigen recognition. Due to their low molecular weight and their special physicochemical properties, it is expected that sdAbs can be used as novel drugs for the treatment of multiple diseases such as cancer [3], rheumatoid arthritis, inflammatory bowel disease [4], and perhaps Alzheimer's disease [5] in the future. The pharmacological and biophysical properties are: heat- resistance up to 90 °C [6], stable towards detergents [7] and generally towards pH extremes [3], highly soluble [3, 8], and easily clonable [3]. sdAbs are produced in *Escherichia coli* [9, 10], *Saccharomyces cerevisiae* [11–13], or *Pichia pastoris* [14].

Due to their special properties, sdAbs are applicable in several domains: in research, therapeutics, diagnostics, and even in cosmetics (as shampoo additive). In therapeutics, the sdAbs could be used as medication (e.g., against cancer [7]) and as vaccines for passive immunization [4]. In research and diagnostics, the sdAbs are used as affinity ligands for the purification of biomolecules [15]. For the immobilization of the sdAbs, they must be purified after cultivation. Purification of regular monoclonal antibodies (mAbs) is usually conducted with a protein A step (affinity chromatography) and a subsequent polishing step, commonly ion exchange chromatography (IEX) and/or hydrophobic interaction chromatography (HIC). The key for the purification of mAbs is the high affinity of protein A for the Fc region of monoclonal antibodies. Since the Fc region is missing in sdAbs (Figure 5.1c), the purification strategy is different: Purification with immobilized metal affinity chromatography (IMAC) and a subsequent gel filtration step yields highly pure sdAbs [4, 16]. However, in some cases the purification was conducted successfully with a protein A chromatography step [11].

In large-scale downstreaming, purification steps are usually conducted as platform processes for a given class of biomolecules (e.g., antibodies), meaning a standard set of unit operations are run successively. In early stage process development, even promising candidates that cannot be purified to a satisfactory level in this platform process get rejected and are not further pursued. We have developed and presented in previous studies [17, 18] an *in silico* tool to evaluate (critical) purification steps on molecular level and make appropriate recommendations to improve purification efficiency (e.g., by adjusting the ionic capacity of the adsorber). In a previous study we showed the application of the developed MD tool to an anion exchanger, which we extend to affinity coupling chromatography media in this study.



(a) Regular monoclonal antibody

(b) Heavy chain antibody

(c) Single domain antibody

Figure 5.1: Schematic depiction of a regular monoclonal antibody, a heavy chain antibody, and a single domain antibody.

In previous laboratory experiments, problems were observed in a cation exchange step and during affinity coupling step. To investigate this behavior, we simulate the 3D structure of four similar nanobodies supplied by BAC B.V. (Naarden, Netherlands) with MD on an adsorber surface of NHS ligands. The calculated interaction energies allow the thesis that potential NHS-coupling sites of the nanobodies can be predicted with MD simulations. Additionally, simulations with a cation exchanger were run.

5.2. MATERIAL AND METHODS

5.2.1. PROTEIN STRUCTURES

The structures of the four nanobodies (VHH3a, VHH3b, VHH5, VHH10), which were determined through homology modeling, were provided by BAC B.V. (Netherlands) in PDB file format. VHH3a has 124 residues, VHH3b 123, VHH5 127, and VHH10 118. The PDB preprocessing was performed as described in [18]. A structure analysis was performed to show the differing sequence segments and their resulting tertiary structure that might lead to different interaction behavior. The sequence alignment was run with ClustalX 2.1 [19]. The graphical preparation was conducted in Jalview 2.8.2 [20] with a BLOSUM62 matrix [21] to generate a phylogenetic tree that shows the similarities and differences based on their residue sequence.

A subsequent structural alignment was performed in Yasara 14.12.2 [22] with a MUSTANG algorithm [23].

5.2.2. LIGAND STRUCTURES

MD simulations were run with affinity coupling ligands equally distributed on a quadratic surface. MC simulations to calculate the ligand distances were run with the input data from Table 5.1 analogous to [18]. The ligand structure was obtained from the GE Healthcare manual on affinity chromatography [24]. The chemical formula of GE Healthcare's NHS-activated Sepharose 4 Fast Flow (NHS) ligand is
$-OCH_2CH(OH)CNCH_2CH_2CH_2CH_2CH_2C(=O)ONC_4O_2$.
Additionally, simulations were run on a surface made up of Capto S ligands to determine problematic interactions, which were observed in experiments. The chemical formula of GE Healthcare's Capto S ligand was obtained from [25] and is
$-OCH_2CHOHCH_2OCH_2CHOHCH_2SO_3^-$.

Table 5.1: Adsorbent properties, data from [24–27]

Adsorber	Ionic capacity [µmol/mL]	Extraparticle porosity	Intraparticle porosity	Pore diameter
NHS	16-23	0.35	0.76; 0.81; 0.85; 0.9	30; 32; 33.4; 38.6
Capto S	110-140	0.35	0.76; 0.81; 0.85; 0.9	30; 32; 33.4; 38.6

5.2.3. SIMULATION PARAMETERS AND SETUP

The simulation setup was analogous to [18].

The simulation parameters are listed in Table 5.2.

The nanobodies were structurally aligned prior to simulation to facilitate the post-simulation evaluation in the interaction maps. With this approach, the interaction maps of the individual nanobodies can be compared directly.

5.2.4. MOLECULAR DYNAMICS SIMULATIONS

The MD simulations were run automatically with a set of macros in Yasara. The procedure started with an energy minimization of the system, followed by a gradual heat-up from 0 to 300 K to slowly equilibrate it. Subsequently, the actual data producing main step was run:

Table 5.2: Simulation parameters

Parameter	Setting
Protein distance	5 Å
No. of orientations (states)	50
pH	8.3[a], 4[b], 5.5[b]
Salt concentration	2.92 % [a], 0.90 % [b], 2.87 % [b]
Temperature	300 K
Duration	100 ps
Interactions	Bond, Angle, Dihedral, Planarity, Coulomb, VdW
Cutoff	7.86 Å
Longrange	Coulomb

[a] NHS-activated Sepharose 4 Fast Flow
[b] Capto S

the MD simulation for 100 ps for every single one of the 50 states (orientations). Finally, the analysis macro loads the previously saved simulation snapshots, calculates the interaction energy terms and saves them in a text file in human- readable format.

The post-simulation evaluation was run in Matlab R2014b [28]. In this evaluation script, the energies are imported from the text file, averaged and Boltzmann weighted over time from 5 to 100 ps, and the value of each state is plotted into the interaction map.

ION EXCHANGE CHROMATOGRAPHY SIMULATIONS

Simulations with a cation exchange adsorber ligand (Capto S) were run to identify strong interaction sites and differences between the nanobodies. Simulations were run in two settings analogous to the corresponding laboratory experiments: one with low salt concentration (0.9 % sodium chloride) and a pH of 4 and the other with a higher salt concentration of 2.87 % and a pH of 5.5 to account for a point in the gradient and a high salt wash step, respectively.

AFFINITY COUPLING SIMULATIONS

Simulations with an affinity coupling ligand (NHS) were run to identify strong interaction sites and make a presumption on possible coupling sites. Simulations were run analogous to the corresponding laboratory experiments: at pH 8.3 and a salt concentration of 2.92 % sodium chloride.

Figure 5.2: Phylogenetic tree of the sequence alignment with the VHHs. The branch length is proportional to evolutionary change. It is apparent that VHH10 differs stronger from the other two VHHs than these among themselves. The tree was calculated with a BLOSUM62 matrix [21] in Jalview 2.8.2 [20].

5.3. RESULTS AND DISCUSSION

In the following sections, the results of the structure analysis and MD simulations are presented. The structure analysis was conducted to detect deviations between structures, which helps in the succeeding interpretation of MD simulation results.

5.3.1. SEQUENCE ALIGNMENT

The sequence alignment yielded a phylogenetic tree (Figure 5.2) that shows the similarities and differences of the VHHs. VHH3a and VHH3b are very similar, VHH5 is a little further away and VHH10 differentiates most to the others. From this sequential alignment no definite behavior predictions can be made, since the behavior strongly depends on the spatial arrangement of the residues.

5.3.2. STRUCTURAL ALIGNMENT

The structural alignment between VHH3a and VHH3b had an RMSD of 1.399 Å over 100 aligned residues with 100 % sequence identity. The sequence identity is merely calculated over closely matching residues and can therefore be higher even though the sequence alignment resulted in higher differences. This is an interesting result, since the two sequences only differ in one residue, namely the first. However, the homology modeling yielded different structural arrangements. The structural alignment between VHH3b and VHH5 has an RMSD of 1.108 Å over 114 aligned residues with 75.44 % sequence identity. The alignment between VHH3b and VHH10 has an RMSD of 1.297 Å over 102 aligned residues with 77.45 % sequence identity. The structural alignment between VHH5 and VHH10 has an RMSD of 1.335 Å over 104 aligned residues with 75.96 % sequence identity.

Interestingly, the 3D structures of VHH3a and VHH3b differ stronger than VHH3a and VHH5, although the sequence identity is much higher. The other structures differ mostly in the lower left part of the nanobodies (see Figure 5.3). Even though the sequences differentiate strongly, the resulting tertiary structure (beta-sheet and loop) is maintained. However, it cannot be concluded, that the ligand interaction is therefore the same. The lower left part of the nanobodies differ most in their structure and are more likely to lead to different interaction behavior. We therefore expect the most different interaction behavior in this region.

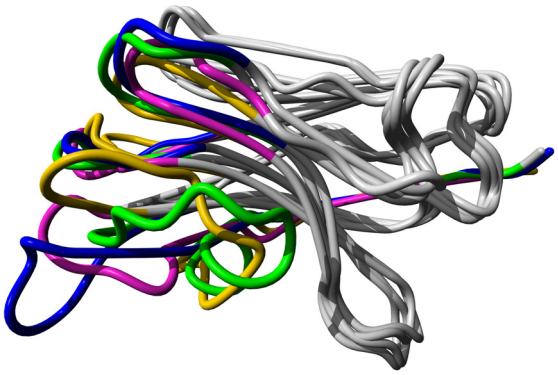


Figure 5.3: Structural alignment of the VHHs. The colored regions indicate the strongest differences between the sequences. VHH3a is colored blue, VHH3b is colored yellow, VHH5 is colored green, and VHH10 is colored magenta. The colored region in the upper part of the nanobodies nicely depicts the difference between sequential and structural alignment.

5.3.3. ANTIGEN BINDING SITES

We propose the thesis that the antigen binding site (ABS) is the side on the nanobody that contains a cluster of charged residues, since Absolom and van Oss [29] and van Oss et al. [30] suggest that antigens bind to antibodies through noncovalent bonds driven by long-range electrostatic interactions. The charged residues cluster for all nanobodies on the lower right side (Figure 5.4, the ABS of VHH10 is highlighted in magenta). Since the nanobodies all have the same ABS region, only VHH10 is depicted. The specific charged residues that make up the ABS region are displayed in Table 5.3.

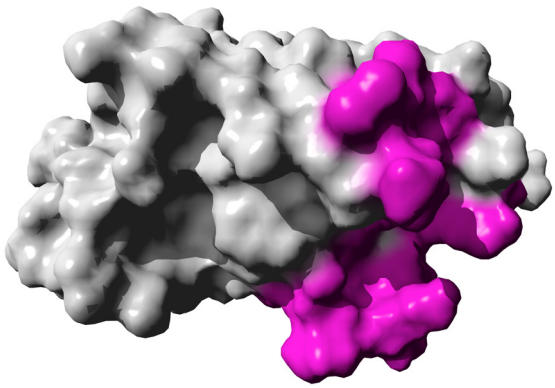


Figure 5.4: Antigen binding site of VHH10, determined by a cluster of charged residues.

Table 5.3: Antigen binding sites for the nanobodies. The binding site in each nanobody consists of three parts that are spatially close together and form a gap (cf. Figure 5.4).

Nanobody	Antigen binding site
VHH3a	ARG38–GLU46 and ASP62–ARG67 and LYS87–ASP90
VHH3b	ARG37–GLU45 and ASP61–ARG66 and LYS86–ASP89
VHH5	ARG38–ARG45 and ASP61–ARG66 and ASP84–ASP89
VHH10	ARG38–GLU46 and GLU61–ARG67 and LYS87–ASP90

5.3.4. ION EXCHANGE CHROMATOGRAPHY SIMULATIONS ANALYSIS

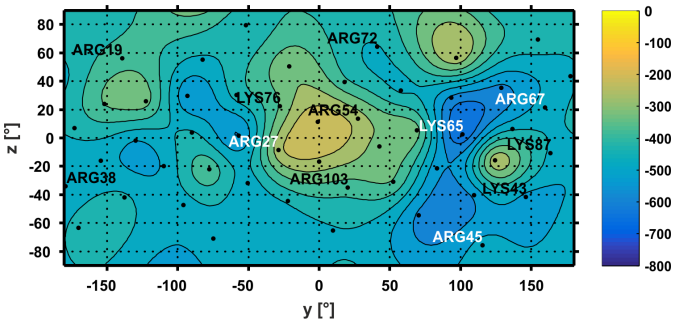
Simulations with Capto S ligands were run to identify strong interaction sites and differences between the different nanobodies as this step caused problems in experiments. Two setups were run to account for a point in the gradient and a high salt step, as described in Section 5.2.4. The resulting interaction maps for each setup are pictured in Figures 5.5 and 5.6.

ION EXCHANGE CHROMATOGRAPHY SIMULATIONS AT PH 4 AND 0.9 % SALT CONCENTRATION

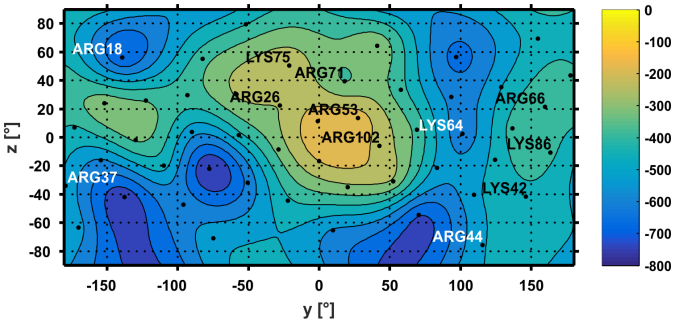
Although VHH3a and VHH3b sequentially only differ in one residue, the interaction maps differ severely. Considering the different 3D structure (cf. Section 5.3.2), this makes perfect sense and was expected. VHH3b shows stronger interaction states and the most present charged residues in these states are LYS42 and ARG44, whereas they are LYS65 and ARG67 for VHH3a. We assume that VHH3a will elute slightly earlier in the gradient. Since the elution time probably does not differ significantly, the peaks may overlap partly and therefore may be impossible to distinguish.

VHH5 shows the strongest interaction states compared to the other VHHs judging from the interaction map (Figure 5.6c). It has multiple residues that show up in at least two of the strong sites: ARG19, LYS43, ARG45, ARG57, LYS65, and ARG67.

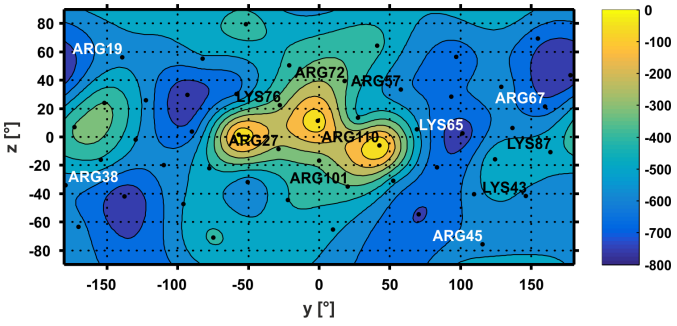
VHH10 (Figure 5.6d) shows the least strongest interaction, but same preferential regions as VHH3a. Most contributing residues are: ARG66, ARG71, and LYS86. The elution order due to the averaging and Boltzmann weighting of all 50 states would be VHH3a, VHH3b, VHH10, and at last VHH5, when optically differentiable.



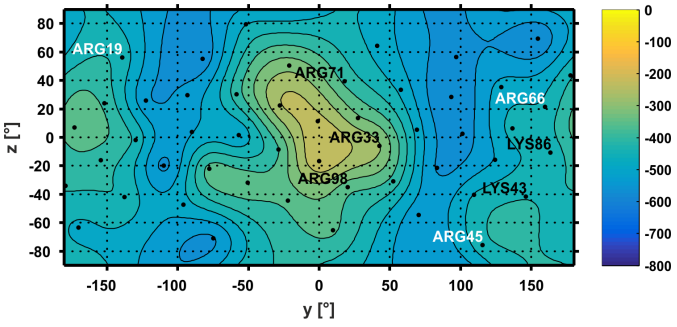
(a) VHH3a



(b) VHH3b



(c) VHH5



(d) VHH10

Figure 5.5: Interaction maps of the nanobody simulations on Capto S ligands at pH 4 and 0.9 % sodium chloride concentration: Coulomb interaction energy [kJ/mol]. Residues that are close to the ligand surface and are potential binding partners are labeled with residue name and number.

ION EXCHANGE CHROMATOGRAPHY SIMULATIONS AT PH 5.5 AND 2.87 % SALT CONCENTRATION

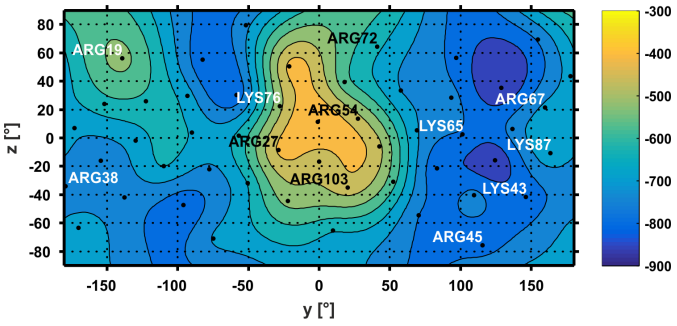
Again, VHH3a and VHH3b show different interaction behavior: VHH3a has its hot spots more on the right side of the map, mostly influenced by ARG67, LYS65, and LYS87. VHH3b has its hot spot in the lower part of the map, mostly influenced by ARG44 and LYS42. The averages of the Boltzmann weighted energies of all 50 states differ by 60 kJ/mol, with VHH3a interacting stronger.

From the averaged energy, VHH5 lies in between VHH3a and VHH3b. Most influencing residues are ARG45, LYS65, and ARG101.

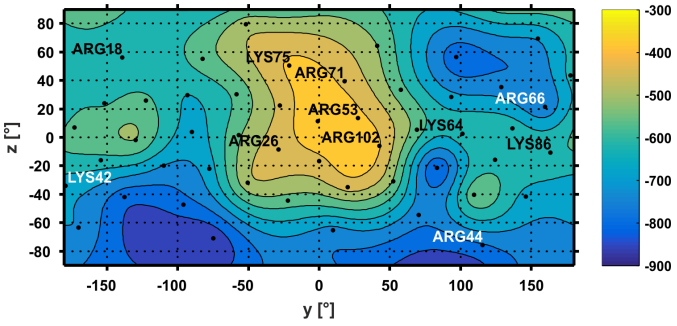
VHH10 shows the lightest interaction map with the same regions of interaction as VHH3a. Most influencing are ARG33, ARG66, and ARG71.

Elution order would be VHH10, VHH3b, VHH5, and at last VHH3a.

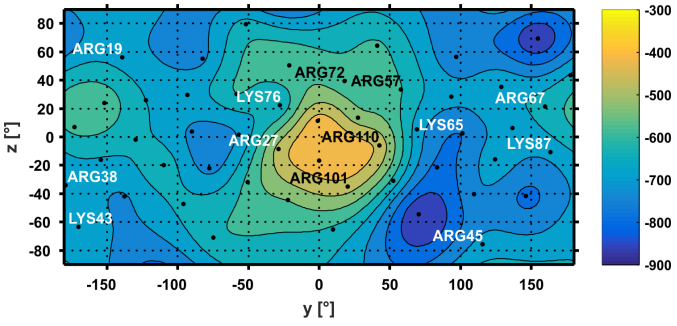
Through the representation of the interaction maps in the same scaling, one can easily compare the interaction behavior and since alignment also the changes in interaction hot spots. Most binding contributing are arginines, which also has been published previously [18]. To influence binding behavior, a possible approach could be to point mutate strongly contributing residue into uncharged residues but otherwise having similar properties.



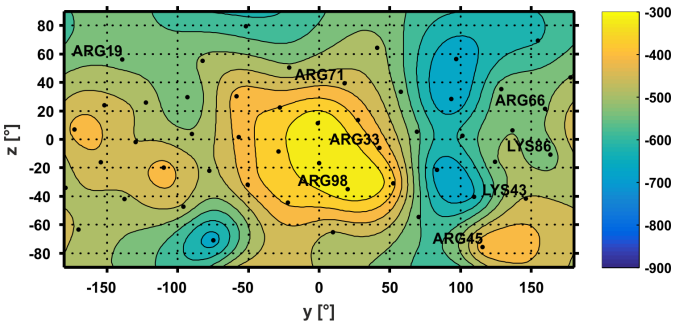
(a) VHH3a



(b) VHH3b



(c) VHH5



(d) VHH10

Figure 5.6: Interaction maps of the nanobody simulations on Capto S ligands at pH 5.5 and 2.87 % sodium chloride concentration: Coulomb interaction energy [kJ/mol]. Residues that are close to the ligand surface and are potential binding partners are labeled with residue name and number.

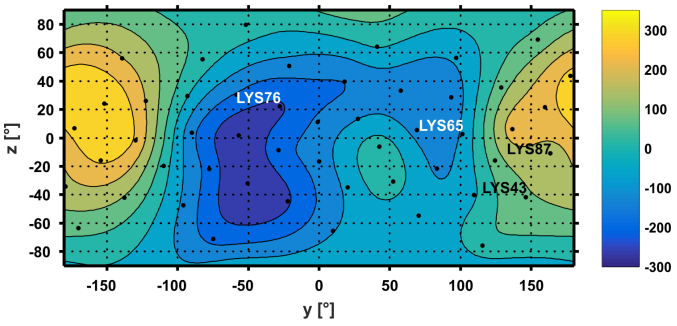
5.3.5. AFFINITY COUPLING SIMULATIONS ANALYSIS

Total interaction energies derived from simulations were Boltzmann weighted and averaged over a period of 5 to 100 ps and plotted into interaction maps to visualize states with strong total interaction. Values between the points were interpolated.

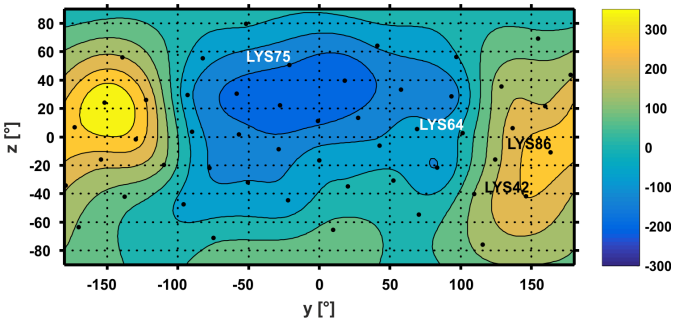
Although VHH3a and VHH3b differ in only one residue, the interaction maps (Figure 5.7) differ significantly. VHH3a shows an interaction area with energies up to -300 kJ/mol and four orientations that are located inside this area. When looking at the orientations in 3D, all four have GLN1 close to the ligand surface, in three of four cases LYS76 is also nearby.

Following the theory of Boltzmann distribution, the probability of orientation in one of these four orientations is higher compared to the other states. We postulate that the NHS coupling will more likely happen in states that are energetically more probable due to their higher likelihood of occurrence. In this study, we did not investigate the actual affinity coupling mechanism and therefore cannot make a statement to the coupling mechanism itself or its degree.

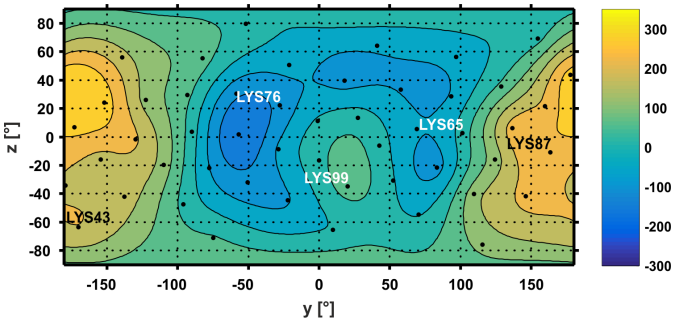
For VHH3b, LYS75 was closest to the ligands in 3 of 5 states and VAL1 in 2 of 5 states. For VHH5, there was one distinct state with the lowest interaction energy and in this state GLN1 and LYS76 were closest to the ligands. For VHH10, two states were in the region of the highest interaction and in both states GLN1 and LYS43 were closest to the ligands. In all four cases, the N-terminus was close to the ligands in the favorable states. Additionally, for VHH3a, VHH3b, and VHH5, lysine at 75th or 76th place, respectively, was also close to the ligands. In case of VHH10, LYS43 was close to the ligands. Based on our assumption that these states are more probable and therefore coupling will more likely happen with these residues, we hypothesize that the N-terminus and LYS75 or LYS76, respectively, couple with the NHS ligand. In case of VHH10, this would be LYS43.



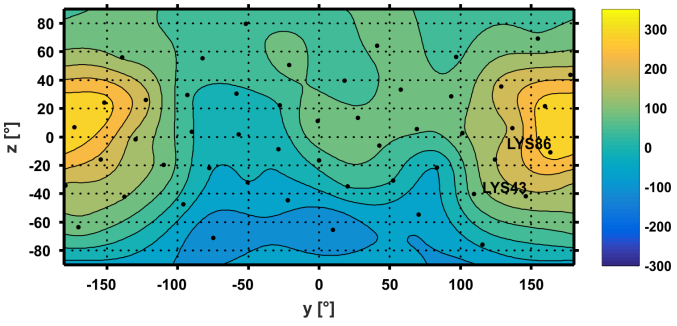
(a) VHH3a



(b) VHH3b



(c) VHH5



(d) VHH10

Figure 5.7: The nanobodies on NHS-activated Sepharose 4 FF: Total interaction energy maps (energy in kJ/mol). Residues that are close to the ligand surface and are potential coupling partners are labeled with residue name and number. The N-terminus, which can also be coupled to the NHS ligand, is not labeled due to technical reasons.

During certain coupling procedures an inactivation of VHH10 affinity was observed (personal communication BAC B.V.). When looking at the states with the highest interactions, GLN1 is at the opposite end of the ABS, however LYS43 is part of it (Figure 5.8 compared to Figure 5.4). This explains why VHH10 is inactivated. The detailed analysis of interaction hot spots demonstrates how our tool is capable of identifying difficulties in process results. Consequently, the next step would be to work out a solution strategy, which at first could also be done *in silico* by running multiple scenarios (e.g., point mutation, not part of this study).

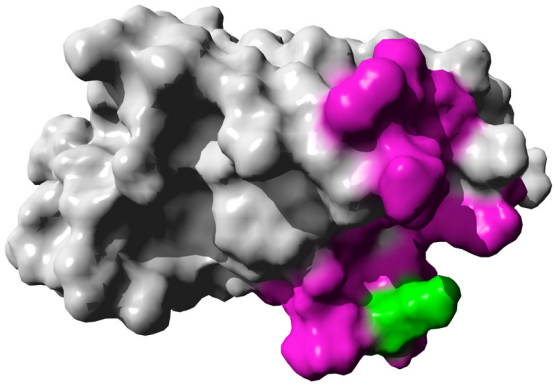


Figure 5.8: VHH10 with the ABS colored magenta and LYS43, one residue that contributes two both states with high interaction energy, colored green. This nicely depicts why VHH10 is inactivated after coupling. The coupling reaction very likely takes place with GLN1 (on the opposite side of the nanobody) and with LYS43. The ABS is sterically blocked. (cf. Figure 5.4)

5.4. CONCLUSIONS

We simulated four similar nanobodies on a cation exchanger and on an affinity coupling ligand surface. With the results of the simulations, we were able to investigate the binding contributing residues in case of the cation exchanger and forecast the expected elution order. The simulations on the NHS surface yielded a greater understanding of potential coupling sites under the assumption that energetically more favorable orientations will more likely undergo coupling reaction.

This makes way for different promising uses: if certain nanobodies cause problems during processing, e.g., are inactivated after coupling, as was the case with one nanobody in this work, this can be investigated on molecular level. Potential possibilities, e.g., point mutations, could first be tried *in silico* without even running one single experiment. However, if the simulation results look promising, they inevitably must be verified by experiments. Another potential application could be the preliminary analysis of potential coupling sites for a

whole library of nanobody variants prior to experimental screening. The computational cost can be minimized by running docking simulations prior to MD simulations and subsequently screening only the most promising candidates in more detail.

ACKNOWLEDGEMENTS

We would like to thank BAC B.V. for kindly providing the 3D structures of the VHHs used in this study. Also, we would like to thank Jörg Kittelmann and Sven Amrhein from our work group for their helpful contribution to various sections of this work.

REFERENCES

[1] C. Hamers-Casterman, T. Atarhouch, S. Muyldermans, G. Robinson, C. Hamers, E. B. Songa, N. Bendahman, and R. Hamers. Naturally occurring antibodies devoid of light chains. *Nature*, 363:446–448, 1993.

[2] A. S. Greenberg, D. Avila, M. Hughes, A. Hughes, E. C. McKinney, and M. F. Flajnik. A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. *Nature*, 374(6518):168–173, 1995.

[3] H. Revets, P. de Baetselier, and S. Muyldermans. Nanobodies as novel agents for cancer therapy. *Expert Opin. Biol. Ther.*, 5(1):111–24, Jan. 2005.

[4] S. Muyldermans. Nanobodies: natural single-domain antibodies. *Annu. Rev. Biochem.*, 82:775–97, Jan. 2013.

[5] A. R. A. Ladiwala, M. Bhattacharya, J. M. Perchiacca, P. Cao, D. P. Raleigh, A. Abedini, A. M. Schmidt, J. Varkey, R. Langen, and P. M. Tessier. Rational design of potent domain antibody inhibitors of amyloid fibril assembly. *Proc. Natl. Acad. Sci. U. S. A.*, 109(49): 19965–70, Dec. 2012.

[6] R. H. J. van der Linden, L. G. J. Frenken, B. de Geus, M. M. Harmsen, R. C. Ruuls, W. Stok, L. de Ron, S. Wilson, P. Davis, and C. T. Verrips. Comparison of physical chemical properties of llama VHH antibody fragments and mouse monoclonal antibodies. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.*, 1431(1):37–46, Apr. 1999.

[7] S. Muyldermans. Single domain camel antibodies: current status. *Rev. Mol. Biotechnol.*, 74(4):277–302, June 2001.

[8] M. M. Harmsen and H. J. de Haard. Properties, production, and applications of camelid single-domain antibody fragments. *Appl. Microbiol. Biotechnol.*, 77:13–22, 2007.

[9] M. Arbabi-Ghahroudi, A. Desmyter, L. Wyns, R. Hamers, and S. Muyldermans. Selection and identification of single domain antibody fragments from camel heavy-chain antibodies. *FEBS Lett.*, 414(3):521–526, Sept. 1997.

[10] F. Rahbarizadeh, M. J. Rasaee, M. Forouzandeh-Moghadam, and A.-A. Allameh. High expression and purification of the recombinant camelid anti-MUC1 single domain antibodies in Escherichia coli. *Protein Expr. Purif.*, 44(1):32–38, Nov. 2005.

[11] L. G. J. Frenken, R. H. J. van der Linden, P. W. Hermans, J. W. Bos, R. C. Ruuls, B. de Geus, and C. T. Verrips. Isolation of antigen specific Llama V(HH) antibody fragments and their high level secretion by Saccharomyces cerevisiae. *J. Biotechnol.*, 78(1):11–21, Feb. 2000.

[12] Y. E. Thomassen, W. Meijer, L. Sierkstra, and C. Verrips. Large-scale production of VHH antibody fragments by Saccharomyces cerevisiae. *Enzyme Microb. Technol.*, 30(3):273–278, Mar. 2002.

[13] J. M. van der Vaart. Expression of VHH antibody fragments in saccharomyces cerevisiae. In P. M. O'Brien and R. Aitken, editors, *Antibody Phage Display*, volume 178 of *Methods in Molecular Biology*, pages 359–366. Humana Press, 2002. ISBN 978-0-89603-906-3.

[14] F. Rahbarizadeh, M. J. Rasaee, M. Forouzandeh, and A.-A. Allameh. Over expression of anti-MUC1 single-domain antibody fragments in the yeast Pichia pastoris. *Mol. Immunol.*, 43(5):426–35, Feb. 2006.

[15] S. V. Tillib, M. E. Privezentseva, T. I. Ivanova, L. F. Vasilev, G. A. Efimov, Y. G. Gursky, G. P. Georgiev, I. L. Goldman, and E. R. Sadchikova. Single-domain antibody-based ligands for immunoaffinity separation of recombinant human lactoferrin from the goat lactoferrin of transgenic goat milk. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.*, 949-950: 48–57, Feb. 2014.

[16] P. J. Doyle, M. Arbabi-Ghahroudi, N. Gaudette, G. Furzer, M. E. Savard, S. Gleddie, M. D. McLean, C. R. Mackenzie, and J. C. Hall. Cloning, expression, and characterization of a single-domain antibody fragment with affinity for 15-acetyl-deoxynivalenol. *Mol. Immunol.*, 45(14):3703–13, Aug. 2008.

[17] F. Dismer and J. Hubbuch. 3D structure-based protein retention prediction for ion-exchange chromatography. *J. Chromatogr. A*, 1217(8):1343–53, Feb. 2010.

[18] K. M. H. Lang, J. Kittelmann, C. Dürr, A. Osberghaus, and J. Hubbuch. A comprehensive molecular dynamics approach to protein retention modeling in ion exchange chromatography. *J. Chromatogr. A*, 1381:184–193, Feb. 2015.

[19] M. Larkin, G. Blackshields, N. Brown, R. Chenna, P. McGettigan, H. McWilliam, F. Valentin, I. Wallace, A. Wilm, R. Lopez, J. Thompson, T. Gibson, and D. Higgins. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21):2947–2948, 2007.

[20] A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, and G. J. Barton. Jalview Version 2–A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9):1189–91, May 2009.

[21] S. Henikoff and J. G. Henikoff. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U. S. A.*, 89(22):10915–9, Nov. 1992.

[22] E. Krieger, T. Darden, S. B. Nabuurs, A. Finkelstein, and G. Vriend. Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins*, 57 (4):678–83, Dec. 2004.

[23] A. S. Konagurthu, J. C. Whisstock, P. J. Stuckey, and A. M. Lesk. MUSTANG: a multiple structural alignment algorithm. *Proteins*, 64(3):559–74, Aug. 2006.

[24] GE Healthcare. IgSelect affinity medium. Product data sheet.

[25] GE Healthcare Bio-Sciences. *Affinity Chromatography: Principles and Methods*. GE Healthcare Bio-Sciences, 2007.

[26] G. Carta and A. Jungbauer. *Protein chromatography: process development and scale-up*. Wiley-VCH-Verl., Weinheim, 2010. ISBN 9783527323012.

[27] A. R. Ubiera and G. Carta. Radiotracer measurements of protein mass transfer: kinetics in ion exchange media. *Biotechnol. J.*, 1(6):665–74, June 2006.

[28] MATLAB. R2014b (8.4.0.150421), Sept. 2014.

[29] D. R. Absolom and C. J. van Oss. The nature of the antigen-antibody bond and the factors affecting its association and dissociation. *CRC Crit. Rev. Immunol.*, 6(1):1–46, 1985.

[30] C. J. van Oss, R. J. Good, and M. K. Chaudhury. Nature of the antigen-antibody interaction: Primary and secondary bonds: Optimal conditions for association and dissociation. *J. Chromatogr. B: Biomed. Sci. Appl.*, 376(0):111–119, 1986. 6th International Symposium on Bioaffinity Chromatography.

Chapter 6

Conclusion and Outlook

The major objective of this study was the extension and optimization of an MD tool to increase the understanding of adsorption behavior of biomolecules on adsorber surfaces of various adsorbents. For this purpose, an MD tool was developed and applied to various tasks. The result is a versatile, easy to use, and expandable tool that quickly and automatically generates results and thus the user gains a first impression on the behavior of the biomolecule on the adsorbent. The approach was a gradual development of the tool: In a first step the tool was ported to a different software package and extended to include anion exchanger simulations. Next, the influence of adsorber variables on the example of varying adsorber capacities was investigated. Finally, a case study was carried out to investigate multiple issues in the downstream and immobilization process of single-domain antibodies.

The MD tool was able to simulate a wide range of biomolecules on chromatographic surfaces. The tool was applied to anion and cation exchange adsorbers. It can be concluded that simulations yield similar results with regard to binding orientation and retention behavior as experimental results. Investigations on molecular level revealed interaction areas mostly defined by arginines in case of cation exchangers and aspartic acids in case of anion exchangers, which is also in good agreement with previous experimental results. The tool was capable of predicting ion exchanger interactions and therewith retention behavior, as well as helping the user to gain insight into binding mechanisms on molecular level.

The simulations with varying ionic capacities confirmed that the ligand density has a strong influence on the binding behavior, which could be seen in MD simulations as well as in experiments. MD simulations were capable of describing differing retention behavior due to varying ligand densities and can therefore be used in customizing adsorbent material and troubleshooting challenging purification tasks. To achieve optimal results with the presented method, a twin tracked strategy must be run of selected small scale experiments to calibrate

the model. Subsequently, it is able to identify optimal adsorbent characteristics in terms of ionic capacity and therefore gains predictive power.

The simulations of four nanobodies on an affinity coupling ligand (NHS) surface yielded a greater understanding of potential coupling sites under the assumption that energetically more favorable orientations will more likely undergo coupling reaction. With the results of the simulations of four nanobodies on a cation exchanger surface, the binding contributing residues could be identified and an expected elution order was forecast. This makes way for different promising uses: problematic biomolecules in terms of processing can be investigated on molecular level. Potential possibilities, e.g., point mutations, could first be tried *in silico* without even running one single experiment. Another potential application could be the preliminary analysis of potential coupling sites for a whole library of nanobody variants prior to experimental screening.

The development of a versatile MD tool was completed successfully. It was proved of being functional in different studies. The user friendliness is ensured by a fully automatically running script in which the parameters are defined in the beginning. The developed tool is a good starting point but could be extended and optimized. Functional extensions to other chromatographic techniques must be implemented. To make it aesthetically pleasing and more intuitive, a graphical user interface ought to be designed and programmed. Multiple other refinements could be: automatically more in-detail investigation in interaction regions of interest, implementation of pH simulation preparations within the script, and plotting of interaction energies directly onto the protein's 3D surface additionally to the interaction maps, among others. In general terms it can be stated, that *in silico* calculations such as molecular dynamics simulations, Monte Carlo simulations, and docking simulations are highly promising pre-experimental techniques (screening and process design) due to their increasing accuracy, decreasing computational cost, and savings potential of material and manpower. However, not only as pre-experimental technique, they also serve as a powerful tool in the fields of failure analysis. The successful implementation of MD simulations in downstream processing strengthens the general development towards the use of computational techniques in process design application.

List of Figures

List of Tables