Accelerated screening and orientation sensitive chromatographic modeling of biopharmaceuticals

Kittelmann, Jörg

DOI
10.4233/uuid:7c2aa066-02aa-40fe-939c-343e14599de0

Publication date
2019

Document Version
Final published version

Citation (APA)
Kittelmann, J. (2019). *Accelerated screening and orientation sensitive chromatographic modeling of biopharmaceuticals*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:7c2aa066-02aa-40fe-939c-343e14599de0

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Accelerated screening and orientation sensitive chromatographic modeling of biopharmaceuticals

Accelerated screening and orientation sensitive chromatographic modeling of biopharmaceuticals

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus prof.dr.ir. T.H.J.J. van der Hagen
chair of the Board for Doctorates
to be defended publicly on
Thursday 14 November 2019 at 15:00 o'clock

by

Jörg Kittelmann

Diplom-Ingenieur (FH), Technische Fachhochschule Berlin, Germany
born in Jena, Germany.

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Assoc. Prof. dr. ir. M. Ottens	Delft University of Technology, promotor
Prof. dr. ing. J. Hubbuch	Karlsruhe Institute of Technology, promotor

Independent members:

Prof. dr. ing. M.H.M. Eppink	Wageningen University
Prof. dr. ir. H.J. Noorman	Delft University of Technology
Prof. dr. J.H. van Esch	Delft University of Technology
Prof. dr. W.R. Hagen	Delft University of Technology

ISBN 978-94-6384-075-0

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

To Paul and Oskar.

Contents

Summary

The downstream process development for biopharmaceuticals is faced with increasing challenges. A growing market of drug candidates and new molecule families, as well as a rising trend to personalized medicine lead to an increase in market diversity. At the same time more purification techniques and materials become available, resulting in an exponential growth in potential parameter combinations and conditions to be considered and screened for.
The establishment of high throughput screening (HTS) technologies and automated liquid handling stations (LHS) have driven standardization in experiments, data handling and data quality assessment in the last decade. Despite, the establishment of automation technologies for almost all purification process steps throughout the field of DSP development, a miniaturization beyond the scale of 96-well plates has not been reached, as sample handling and pipetting accuracy fell short with established LHS.

Nevertheless, with HTS technology, more and reproducible data becomes available, providing the experimental foundation for mechanistic and empirical modeling of purification processes, as molecular dynamic (MD) simulations and quantitative structure activity relationship (QSAR) models. In contrast to MD simulations, which are limited in their application range due to high computational costs, QSAR combines mechanistically derived descriptors, capturing molecular features in numerical entities, and an empirical combination of these descriptors to predictive models. Despite the wide spread application of QSAR models in other fields of research, QSAR yet lacks suitable sets of descriptors for purification process modeling, as published descriptors for proteins are primarily targeted on structure stability and folding. In downstream processing of biomolecules, interactions are often orientation sensitive between molecule and a surface as in chromatography or between identical molecules (e.g. diffusion, aggregation and crystallization processes). The lack of suitable descriptors, modeling standards, and easy to use tools hamper a wide application of QSAR techniques in process development.

In this work, a new approach to high throughput screening is presented, using accurate measurements of liquid and resin volumes to overcome the limitation of 96-well plates. The applicability is presented in a 384-well based batch isotherm process on an automated liquid handling station. Single handling and pipetting step accuracies are measured and their impact on experimental results is calculated by Monte Carlo simulations. Higher or comparable accuracy to similar processes established in 96-well format is shown. The application of spectrometric methods to resin quantification and qualification is demonstrated on SP Sepharose FF resin, using UV/Vis spectra information to measure volumes, ligand concentration and pro-

tein load based on light scattered within porous adsorbent particles without sample interference. The further potential of miniaturized high throughput screening is demonstrated by introduction of the µF-on-LHS interface, combining microfluidic structures and automated LHS systems. The potential in automated handling of microfluidic chips and fluid control by the LHS pipetting unit is shown by different microfluidic chip designs and the realization of automated droplet microfluidics in the volume range down to 0.7 µL.

For the utilization of HTS data in predictive modeling, sets of QSAR descriptors were developed. These descriptors reflect the distribution of characteristics as electrostatic potential and hydrophilic/hydrophobic constants on the molecular surface, as well as different orientations of interaction, and screening effects by the surrounding solvent. The applicability of these descriptor sets is demonstrated by the predictive modeling of the Langmuir isotherm affinity parameter, a key parameter in chromatography modeling. Found preferred orientations of the model protein Lysozyme showed good comparability to experimentally and MD simulation derived orientations in literature. The developed modeling approach, was applied to a set of monoclonal antibodies in combination with model proteins to demonstrate the use in process development for high value biopharmaceuticals. The successful prediction of mAb isotherm parameters by QSAR models without training of the model on said structure is displayed. Further, the influence of the ionic strength on the binding orientation of mAb is investigated, showing a change from a ”head-on” orientation, with the Fab fragment facing the adsorbent at low ionic strength to a ”flat” orientation at higher ionic strength. This result is consistent with literature and experimental results conducted.

To make the developed descriptor calculation approach available for other applications in research and industry, the program mantoQSAR was written, which guides the user via an graphic user interface (GUI) through the steps of predictive QSAR modeling, including structure selection, descriptor calculation, modeling of response values and response prediction for new observations. Emphasise was put on the visualization of molecule features and descriptors, to allow the user to investigate preferred orientations of interaction and the corresponding projection of descriptor values. The software was designed to be usable on all conventional computer operating systems and to be run on user computers, to ensure confidentiality of structural and experimental information, in contrast to distributed web services.

This work demonstrates the unused potential of optical methods in HTS applications for process improvement and resin sample analysis. Further, advances in predictive modeling are demonstrated by utilising mechanistic understanding in descriptor calculation. The developed methodologies have inspired ongoing research in the fields of microfluidics, lab automation and QSAR modeling and will have a significant impact on future downstream process development.

Samenvatting

De ontwikkeling van zuiveringsprocessen voor biofarmaceutische producten, wordt geconfronteerd met toenemende uitdagingen. Een groeiende markt van kandidaatgenees- middelen en nieuwe molecule families, alsook een stijgende trend voor gepersonaliseerde medicijnen, leiden tot een toename van de diversiteit van de markt.

Tegelijkertijd worden meer zuiveringstechnieken en materialen verkrijgbaar; dit resulteert in een exponentiële groei van de potentiële parameter combinaties en voorwaarden, waarnaar gescreend en beschouwd moet worden. De oprichting van high throughput screening (HTS) technologieën en geautomatiseerde vloeibare handling stations (LHS), hebben de standaardisatie van experimenten, data handling en de beoordeling van data kwaliteit in het laatste decennium bevorderd. Ondanks de oprichting van automatiserings technologieën voor bijna alle zuiveringsstappen van het gehele gebied van DSP ontwikkeling, wordt een miniaturisatie van deze niet verder dan de 96-wells platen bereikt. Dit komt doordat het monster handling en het pipetteren met de gevestigde LHS, niet nauwkeurig genoeg was.

Desondanks worden door de HTS meer en reproduceerbare gegevens beschikbaar. Deze vormen de experimentele basis voor de mechanistische en empirische modellering van zuiveringsprocessen, in de vorm van moleculaire dynamica (MD) simulaties en kwantitatieve structuuractiviteitsrelatie (QSAR) modellen. In tegenstelling tot MD-simulaties, die vanwege de hoge computationele kosten beperkt zijn in hun toepassingsgebied, combineren QSAR modellenmechanistisch afgeleidde descriptoren, die de moleculaire eigenschappen in numerieke entiteiten vastleggen met een empirische combinatie van deze omschrijvingen in voorspellende modellen. Ondanks de wijdverspreidde toepassing van QSAR modellen op andere gebieden van het onderzoek, mist QSAR geschikte descriptoren voor het modelleren van zuiveringsprocessen. Dit komt doordat de meest gepubliceerde descriptoren voor proteïnen vooral gericht zijn op de stabiliteit en de vouwstructuur. Tijdens het zuiveringsproces van biomoleculen, zijn de moleculaire interacties vaak gevoelig voor de orientatie tussen identieke moleculen (bijvoorbeeld bij diffusie, aggregatie en kristallisatieprocessen) of tussen de moleculen en een oppervlak, zoals in chromatografie. Het ontbreken van geschikte descriptoren, modellering normen, en gebruiksvriendelijk gereedschap, bemoeilijken een ruime toepassing van QSAR technieken voor het gebruik in de procesontwikkeling.

In dit werk wordt een nieuwe aanpak voor high throughput screening gepresenteerd, die met behulp van nauwkeurige metingen van vloeistof- en harsvolumes de beperking van 96-well platen overwint. De toepasbaarheid wordt gepresenteerd in een 384-wells gebaseerde batch isotherm proces op en geautomatiseerde liquid

handling station. De nauwkeurigheden van Single handling en pipetterstapen worden gemeten en hun invloed op de experimentele resultaten worden berekend met behulp van Monte Carlo simulaties. Een gelijkwaardige of betere nauwkeurigheid dan soortgelijke in 96-well formaat gevestigde processen, wordt getoond. De toepassing van spectrometrische methodes voor de kwantificatie en kwalificatie van hars, wordt gedemonstreerd voor SP Sepharose FF-hars, waardoor gebruik van UV / Vis spectra de volumes, ligand concentratie en eiwit belasting op basis van het licht verstrooid in poreuze adsorbensdeeltjes zonder monster interferentie gemeten worden. Het verdere potentieel van geminiaturiseerde high throughput screening blijkt uit de invoering van uF-on-LHS interfaces, die microfluïdische structuren en geautomatiseerde LHS systemen integreren.

Het verdere potentieel van geautomatiseerde verwerking van microfluïdische chips en vloeistofcontrole door de LHS pipetteren unit, wordt getoond door verschillende microfluïdische chip ontwerpen en de realisatie van geautomatiseerde druppel microfluidics in een volume tot een maximum van 0,7 µl.

Voor het gebruik van de HTS data in voorspellende modellen, worden sets van QSAR descriptoren ontwikkeld. Deze descriptoren spiegelen de verdeling van eigenschappen zoals elektrostatisch potentieel en hydrofiele / hydrofobe constanten op de moleculaire oppervlakte weer, evenals verschillende oriëntaties van interactie en afschermingseffect door het omringende oplosmiddel. De toepasbaarheid van deze descriptor sets blijkt uit de voorspelling van de Langmuir isotherm affiniteit parameter, een belangrijke parameter in de chromatografie modellering. De gevonden geprefereerde oriëntaties van het model eiwit lysozym, toonden een goede vergelijkbaarheid met experimentele en van MD simulaties afgeleidde oriëntaties in de literatuur. De ontwikkelde modelleringsbenadering werd op een reeks monoklonale antilichamen in combinatie met modeleiwitten gebruikt, om het gebruik ervan voor procesontwikkeling voor biofarmaceutica van hoge waarde te demonstreren. De succesvolle voorspelling van mAb isotherm parameters door QSAR modellen wordt zonder training van het model op de genoemde constructie weergegeven. Verder wordt de invloed van de ionische sterkte op de adsorptie van de mAb oriëntatie onderzocht. Dit toont de verandering van een ”head-onöriëntatie, van en Fab-fragment met uitzicht op de adsorbens bij lage ion sterkte naar een ”platte-oriëntatie bij hogere ionensterkte. Het resultaat is in overeenstemming met de literatuur en de resultaten van de uitgevoerde experimenten.

Om de ontwikkelde benadering voor de berekening van de descriptoren beschikbaar te maken voor andere toepassingen in onderzoek en industrie, wordt het programma mantoQSAR geschreven. In dit programma, dat de gebruiker via een grafische user interface (GUI) door de stappen van het voorspellende QSAR modellering beleidt, zijn de structuur selectie, desciptor berekening, modellering van de respons waarden en respons voorspelling voor nieuwe waarnemingen, inbegrepen. Nadruk lag op de visualisatie van de moleculaire functies en beschrijvingen, om de gebruiker het onderzoek van geprefereerde oriëntatie van de interactie en

de overeenkomstige projectie van descriptor waarden toe te staan. De software was ontworpen om op alle gangbare besturingssystemen toepasbaar te zijn en kan op gewone computers uitgevoerd worden, om, in tegenstelling tot gedistribueerde web services, de vertrouwelijkheid van structurele en experimentele informatie te waarborgen.

Dit werk toont het ongebruikte potentieel van optische methodes in HTS-toepassingen voor procesverbetering en harsmonster analyse. Verder wordt de vooruitgang in de voorspellende modellering door gebruik van mechanistische inzicht in het berekenen van descriptoren aangetoond. De ontwikkelde methodieken hebben lopend onderzoek op het gebied van microfluidics, laboratorium automatisering en QSAR modellering geïnspireerd en zullen een aanzienlijke impact hebben over de toekomstige downstream procesontwikkeling.

1

Introduction

1

1.1. Background and motivation

The development of purification processes for biomolecules in pharmaceutical production is challenged by the requirement for high purities and short development cycles [1–3]. Pharmaceutical products for therapeutic use need to meet high purity standards, which are to be reached by downstream processing of the target molecules. Therefore, stringent safety requirements are set by regulatory agencies as the US Food and Drug Administration (FDA) and European Medicines Agency (EMA) to ensure consistent product quality and patient safety. Short times to market, are required due to patent regulations limiting the time frame in which a new pharmaceutical product can be marketed, without competing with generic products [2]. Given the anticipated competitive market, reducing production costs has a high priority already in early stage process development. This is even more crucial, as changing approved processes is burdened with extensive regulatory requirements and additional filing [1, 4].

In recent years the development of biopharmaceutical upstream processes has reached an increase in product titer from milligrams to several grams per liter [5]. This development is not met with comparable capacity improvements in purification processes, thereby shifting the production bottleneck to downstream processes [2, 6].

This focus on and growth in the market of biomolecule purification prompted a high degree of diversification in purification process technologies in the recent decade. Liquid chromatography, which is considered the most versatile process step in high value biomolecule separation, provides a vast variety in solid phase materials, interaction characteristics and process parameters to choose from. This ongoing development increases the parameter space which needs to be screened to conclude on a viable and efficient purification process. The design space to be investigated grows thereby exponentially given the multitude of parameter interactions present.

The resulting need for strategies in bioseparation bioprocess development lead to different approaches, varying in complexity, required mechanistic understanding of the process rationale, and experimental efforts needed (see Figure 1.1). The importance of strategic and rational process design, as well as the beneficial combination of process development approaches is extensively investigated in academic and industrial research [7–13].

Theoretical knowledge and experience from other projects (often referred to as ”prior knowledge”), preliminary experimentation and ”one parameter at a time” or ”one factor at a time” constitute the traditional approach to product and process understanding, while multivariate methods such as design of experiments (DoE), principal component analytics (PCA), partial least squares (PLS) methods and various empiric, semi-empiric, and mechanistic mathematical models institute the quality by design (QbD) approach. The importance of process understanding for process control is also highlighted by initiatives from FDA and ICH for QbD driven approval

of biopharmaceutical drug substances and related processes [14, 15].

Increased process understanding and increased experimental throughput can be established by *in silico* process design based on mechanistic and empirical models. By today, this technique is widely established for single process steps as chromatographic separation and for multi-step optimization of purification tasks [16–19]. The description and modeling of biomolecule interactions at the level of the structural features of a single molecule is based in the research field of bioinformatics. The correlation of such molecular structure features to processing or interaction properties by means of molecular descriptors and empirical modeling is termed Quantitative-Structure Activity Relationships (QSAR) and is described in more detail in Section 1.3.

In addition to traditional in-process and specification tests, process analytical technologies (PAT) also play an essential role in generating product and process understanding and setting up a suitable control strategy. PAT is a term used for a system for designing, analysing, and controlling manufacturing through timely measurements of critical quality and performance attributes of raw and in-process materials and processes, with the goal of ensuring final product quality. Here it is of importance to establish a mechanistic understanding of these analytical methods to develop fast, accurate, precise and robust analytics. The approach and techniques used to develop this understanding are the same as in process development and go hand in hand with these.

All process development concepts are, to varying degrees, carried by significant advances in experimental technologies. Miniaturization and automation in experimental determination of process parameters have driven the field of process development by continuously increasing efficiency of screening experiments and providing the data foundation for growing insights in process fundamentals and their modeling.

1.2. High throughput screening technology

Production process designs are often set in an early developmental stage to reduce the time to market of a potential drug lead and to avoid costly design changes in a later stage. To get to efficient and economic processes a broad data basis is needed. This data provides, in an optimal case, process relevant parameters and a good process understanding. The development of this data is subject to the process inherent restriction of sparse sample materials in early development stages. A solution to this is the use of automated, parallelized, and strongly miniaturized screening processes. These three characteristics are combined in high throughput screening (HTS) using liquid handling stations (LHS), which provide a wide range of sample handling, storage and analytic capabilities [20, 21].

While in DSP development the use of 96 well microtiter plates is the standard

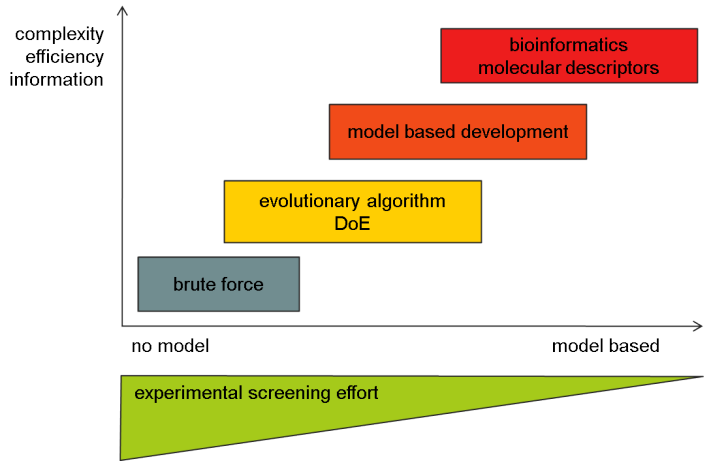

Figure 1.1: Approaches in purification process development plotted over complexity, the need of process understanding, and associated screening efforts.

format in HTS batch applications, 384 well and even 1536 well microplates, which are often referred to as ultra-high throughput screening (uHTS), are available and commonly used in forensic screening, gene databases, and blood analysis for medical purposes [22].

The first downstream processes to be established on automated HTS platforms were resin batch screening applications for chromatographic parameter estimation [23, 24]. Accurate resin volume handling proved to be the main challenge. This inspired, next to different approaches in pipetting resin slurry [24, 25], the development of commercial solutions including a resin distribution system [23] and 96 well microplates with pre-distributed amounts of resin [26]. These efforts were followed by the transfer of purification process steps into the HTS domain with solubility, precipitation, and aqueous two-phase system (ATPS) processes being described in literature [20, 27–29] among others. Further, an automated miniaturized chromatographic column system for the use on LHS was introduced [21]. The combination of packed columns with the pipetting system of the LHS allowed for parallel chromatographic runs of up to 8 columns in parallel. This technology made a fully automatic integration of chromatographic runs and analytics possible.

With the adaptation of almost all purification processes to the HTS scale and the creation of standards, HTS methods became a tool in establishing computational methods for improved screening efficiency. Design of Experiment (DoE) approaches were used in miniaturized batch and column screening [28], as well as learning algorithms as genetic algorithms [30, 31]. HTS scale experiments gained further importance with the development of computational models for large scale chromatography runs, relying on protein-adsorbent interaction parameters [9, 32, 33]. With further optimization of miniaturized processes, complex purification tasks with

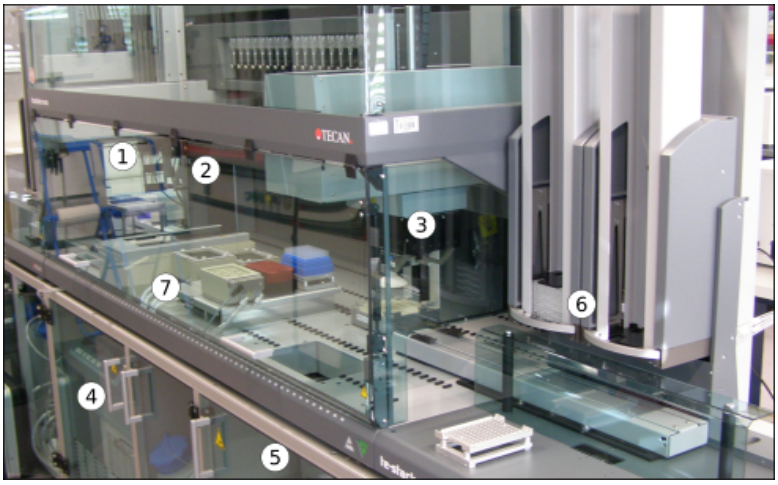

Figure 1.2: Liquid handling station (LHS) Tecan Freedom EVO with 1) a 8 channel pipetting arm, 2) a microtiter plate manipulator (RoMa) for transfer of plates on and to the workstation and 3) a 96 channel pipetting unit. A centrifuge (4) is integrated and can be run automatically with up to 4 microtiter plates. A spectrophotometer is situated beneath the workstation (5) and can be accessed by the RoMa through a recess in the workspace. Microplates can be stored prior and after processing in the TeStack units (6). Additional devices can be placed on the workstation as required (7), including magnetic separation (TeMags, left) and filtration (TeVacs, center) units.

multiple chromatographic steps can be displayed on a LHS and run iteratively for a target parameter to be optimized [34].

Reliable HTS processes moved the bottleneck to the analytical methods connected, which often limited a further increase in throughput. The optimization of established analytics, as size exclusion chromatography, towards reduced processing times and interlaced sample processing [35] as well as the use of newly developed optical methods for biomolecule quantification in complex mixtures [36], allowed for high throughput throughout the developmental process and removed this limitation.

In recent years, the data quality gained from HTS experiments, potential tradeoffs between accuracy and processing time, and transferability to production scale came into focus. This development underlines a maturation in the use of LHS applications and HTS technology. A strong argument is made towards balancing the need for fast results, data quality and number of experiments to reach a statistical confidence in the data and to adapt screening approaches accordingly [37–39].

While the increase in throughput by HTS applications in the field of downstream process development is significant compared to manual experimentation, and further increases in sample throughput are now often limited by preparatory steps, sample consumption can still be significant which hinders more comprehensive

screens. The resulting task of further miniaturization experiments beyond the 96-well format is not comparable to existing uHTS applications. These are focused on qualitative applications or sample storage for µL analytical applications. In contrast, for process development quantitative answers are needed. Recent research has shown that the accuracy in sample handling achievable with established LHS technologies is limited [37–39].

This work states that the change of the theorem that accurate sample handling and distribution leads to small experimental error can be replaced by an approach based on the measurement of intermediate sample volumes and concentrations. With analytical techniques excelling small volume handling in accuracy, an increase in data quality can be achieved. This in consequence can be used to design processes utilizing less sample volumes.

A second approach is the combination of complementary technologies to achieve further miniaturization in HTS experiments while remaining or increasing throughput and ease of use. Microfluidic systems achieve lowest volumes and a high degree of volume control [40, 41]. Today's manufacturing technology uses glass and PDMS materials to facilitate complex structures within microfluidic chips. Parallel and sequential combinations of structures for sample modification, mixing or separation, and analysis can be realized within single chips. The used materials are cheap and easy to be processed, which allows fast prototyping and low operational costs. The drawback of this fast developing field is the current lack of standards in microfluidic devices, integrated analytics, as well as pump and control systems [42, 43]. This often not only limits throughput due to manual sample preparation, but also prevents technology transfer between research facilities and large scale applicability within industry.

The integration of LHS and microfluidic devices, as also proposed in this work [44, 45], is part of a strengthened focus on integrating biological processes and biomolecule analytics in microfluidic structures. This work extends from biomolecule characterisation [46], HTS compatible microfluidic chips for spectrophotometric measurements [47] to enzymatic reaction monitoring [48, 49].

1.3. Quantitative Structure Activity Relationships

The term Quantitative Structure Activity Relationship (QSAR) was coined in 1964 and has become a well established tool in the field of chemometrics. The QSAR domain is based on three assumptions: first, biological properties are related to the molecular structure; second, the features of molecular structures can be converted to numerical entities, termed descriptors; third, a statistical relationship between the biological property and molecular descriptors can be established. [50]

The development of QSAR started with the spread of computers in research facilities and is rooted in small molecule drug research [50, 51]. Over the years,

multiple strategies to build QSAR models have been developed. With the development of 3D descriptors and especially comparative molecular field analysis (COMFA) the three-dimensional features of molecular structures were captured and applied in predicting binding modes of protein ligands [52]. Various algorithms were established and used in descriptor selection, (e.g. genetic algorithms, regression methods [53]) and in reducing the dimensionality of the data (e.g. principal component analysis, PCA, or partial least square regression, PLSR).

Today, different modeling algorithms are well established, applied to various datasets, and compared against each other with detailed descriptions of strengths and shortcomings associated with each individual approach [54, 55]. In the development of QSAR, good models characterized by a high correlation coefficient between experimental and modeled parameter were often mistaken for predictive models. This discrepancy gained increasing interest in the last decades, which led to the establishment of parameters describing model characteristics, ensuring predictive and stable models are established [52, 56]. The ongoing research in QSAR has formed a well established technology for application to small molecules, with thousands of descriptors being described and available online [54, 56, 57]. An increasing standardization in tools, research and application has lead to good comparability between models and high confidence in predictions from these [58, 59].

In contrast, applications of QSAR techniques to large molecules as proteins are still sparse. This is based in the fact that most small molecule descriptors are not applicable, as approaches describing rotation angles are not suitable for structures containing thousands of atoms, as antibodies. Further, the questions to be answered by QSAR models differ. While in drug component research, potential leads need to be identified from thousands of potential structures, in protein investigation only dozens or hundreds of molecules are available. This is often coincident with better curated data from a single source and the need for most accurate prediction of molecule properties. In DSP development qualitative answers are required to predict retention factors, diffusion coefficients, or aggregation behavior. Application of QSAR models in DSP started with the development of descriptors, specific for large molecules. Therein, the mapping of electrostatic and lipophilicity parameters to molecule surfaces became a key technique in descriptor calculation, capturing the amorph property distribution within large molecules [60, 61]. Especially the importance of molecular property distribution for describing the interactions in chromatographic systems has been of larger focus in recent years, leading to more sensitive models and larger insight in mechanistic interactions on a molecular level [62–65]. The formulation of descriptors capturing electrostatic potential as a pH dependent parameter allowed for models to be pH sensitive [66]. QSAR models applied mainly to chromatographic separation in predicting retention volumes or mechanistic model parameters in ion exchange chromatography, hydrophobic interaction chromatography, and in chromatographic displacer screening [60, 61, 66–70].

In parallel to the advancements in QSAR applications, MD simulations in DSP

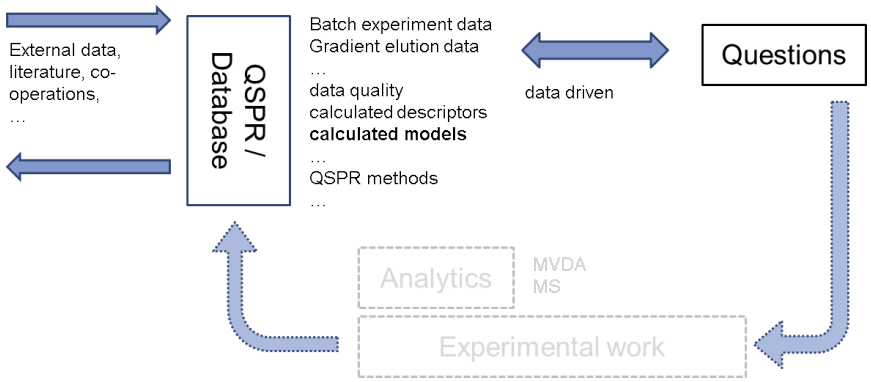

Figure 1.3: Depiction of a knowledge based approach to purification process development based on parameter databases and predictive modeling (QSAR). The structured documentation of previous screenings and research results allows for modeling of process parameters in new purification tasks (labeled as questions), reducing the screening space in need for experimental validation. Experimental results are stored in the database increasing the data pool for upcoming tasks.

development have had a strong development within the last years. Simulation tools have become more reliable and easier to apply as guidelines and graphic user interfaces were more and more established. Full atom simulations of biomolecule conditions account implicitly for all forces influencing molecule behavior in the process step investigated, given correct force field parameters; whereas QSAR models must rely on the applicability of the calculated descriptor to the property investigated. Nevertheless, despite large scale computation power being easily available via high performance computation centers at universities and companies or cloud computing vendors, simulations of large amounts of molecular structures, as they are processed in QSAR modeling, remains inapplicable, especially for large molecules as antibodies.

The current challenge in QSAR modeling approaches in the DSP domain is the lack in standards in descriptor calculation and model evaluation, which prohibits any comparison between applications. Given the complexity of descriptor calculation from scratch on a programming level and the lack of comprehensive software tools also hinders a wider application.

1.4. Knowledge databases

All *in silico* modeling of processes, independent of the nature of the model, e.g. mechanistic or empirical, relies on experimental data. Predictive models not necessarily need experimental data for the target molecule, as the knowledge for model generation can be gained from other structures and data sources. This leads to the possibility to gain insights into new target structures based on the prior knowledge from similar experiments as shown in Figure 1.3. The aim is to omit or reduce

experimental work to investigate the design space of the target molecule by predicting the properties from the accumulated knowledge from prior work.
In chemometrics, this context has led to the generation of extensive databases of structures and experimental data of different source and nature. This is due to a higher screening throughput available and necessary in drug compound screening and the fact that most investigated structures and properties are not restricted from publication. Research efforts have been targeted to provide tools for an easy access to and sharing of structures and experimental data in small molecule search [71]. Large publicly available databases contain millions of chemical compounds tested in various biological assays such as PubChem [72] have been made available. Extensive overviews are provided to guide the user to find suitable knowledge databases and modeling tools [73, 74].

Similar efforts have been started for protein characteristics applied in DSP development. Databases of protein parameters and key host cell proteins are being established [9, 75]. Also, the need for data standards in DSP are discussed and made aware of in academia and industrial research [75–77].

1.5. Research objectives

The primary objective of this research is to integrate the potentials of high-throughput experimentation and molecular structure based modeling into a unified workflow for the generation of predictive models, which are to be used for purification process parameter estimation of new molecules. This approach uses the knowledge gained by screening of previous biopharmaceutical molecules and reference structures.
In the HTS development it is to be shown, that the established 96 well microplate format is not the limit of miniaturization in HTS screening for downstream process parameters. Major focus lays on developing technologies, which overcome the problem of decreasing signal-to-noise ratios with increasing miniaturization. The expected outcome is the establishment of a batch isotherm screening process as technology platform showcasing the work in <100 µL volumes. Further, a platform for efficient parameter screening in the <10 µL volume range is to be proposed as guideline for future developments.
Gained molecular property data is to be used in predictive QSAR modeling as a tool in efficient process development. Therefore, descriptors capturing molecular properties in context of purification processing conditions are to be developed. A successful modeling approach is to be showcased on model proteins and monoclonal antibodies as an example of biopharmaceutical purification challenges.

The aim is to provide a platform for knowledge driven process development which spans from HTS screening to a predictive modeling approach targeted on biomolecule purification.

1

1.6. Outline of thesis

Chapter 1 as introductory chapter, provides background and motivation for this work. A definition and state of the art overview are given on HTS application in purification process development as well as a brief definition and overview on QSAR modeling. This provides the context for used technologies and connects the topics theoretically. Further, the research approach followed is presented in overview.

The main part of the thesis is divided into 6 chapters, each of which is centered on a key aspect of the proposed screening and modeling framework. While chapters 2 to 4 cover all aspects of the research undertaken in miniaturized molecule property screening, chapters 5 to 7 focus on the development of an integrated QSAR modeling framework.
Chapter 2 presents a newly developed batch isotherm screening method in 384 well microplate format, thereby accessing <300 µL scale in HTS screening in downstream process development on LHS. Incentive and results for the use of a developed optical resin volume quantification in favor over handling of solids, are given.
Chapter 3 extends the use of the measurements of light scattering by chromatographic adsorbent volumes presented in Chapter 2 to resin characterization and non interruptive protein load analysis. In **Chapter 4** the concept of process miniaturization below <1 µL scale is presented by the establishment of a microfluidic chip on a liquid handling station. A standardized interface between microfluidic chips and automated HTS robotic system is proposed. A successful implementation and potential of the technology are shown with a droplet generation in chip by positive and negative pressure control via the LHS pipetting device.
In the second part of the thesis a new QSAR modeling approach is presented. **Chapter 5** is introducing an orientation sensitive approach to descriptor calculation for biomolecules and a proof of concept QSAR model for langmuir batch parameter in ion-exchange chromatography. In **Chapter 6** this approach is applied to monoclonal antibodies and proves the applicability of orientation sensitive QSAR technology to large molecules distinguishing minor molecular differences in antibody F_{ab} domain and their influence on binding orientations. The developed modeling technologies are provided in a software program for use in research and process development, featuring a graphical user interface for molecule inspection, model generation, and evaluation, which is presented in **Chapter 7**.
Finally, **Chapter 8** gives an overview on the presented technologies and an outlook on future development in the field of QSAR and HTS application in biomolecule purification development.

Acknowledgement

This work was financially supported by Lonza Biologics PLC and BE-Basic, a public private program by the Netherlands Ministry of Economic Affairs and the BE-Basic partner organizations (www.be-basic.nl).

References

[1] B. D. Kelley, M. Switzer, P. Bastek, J. F. Kramarczyk, K. Molnar, J. L. Coffman, and T. Yu, *High-throughput screening of chromatographic separations: IV. Ion-exchange,* Biotechnol. Bioeng. **100**, 950 (2008).

[2] D. Low, R. O'Leary, and N. S. Pujar, *Future of antibody purification,* J. Chromatogr. B **848**, 48 (2007).

[3] S. Sommerfeld and J. Strube, *Challenges in biotechnology production—generic processes and process optimization for monoclonal antibodies,* Chem. Eng. Process. Process Intensif. **44**, 1123 (2005).

[4] G. Hodge, *Investing in process development,* Bioprocessing J. , 31 (2004).

[5] F. M. Wurm, *Production of recombinant protein therapeutics in cultivated mammalian cells,* Nat. Biotech. , 1393 (2004).

[6] B. Kelley, *Downstream processing of monoclonal antibodies: Current practices and future opportunities,* in *Process Scale Purification of Antibodies* (John Wiley & Sons, Ltd, 2017) Chap. 1, pp. 1–21.

[7] J. A. Asenjo and B. A. Andrews, *Is there a rational method to purify proteins? from expert systems to proteomics,* J. Mol. Recogn. **17**, 236 (2004).

[8] B. Nfor, T. Ahamed, L. v. Dedem, G. v.d. Wielen, E. v.d. Sandt, M. Eppink, and M. Ottens, *Design strategies for integrated protein purification processes: Challenges, progress and outlook,* J. Chem. Technol. Biotechnol. **83**, 124 (2008).

[9] B. K. B. Nfor, M. Noverraz, S. Chilamkurthi, P. D. E. M. Verhaert, L. a. M. van der Wielen, and M. Ottens, *High-throughput isotherm determination and thermodynamic modeling of protein adsorption on mixed mode adsorbents,* J. Chromatogr. A **1217**, 6829 (2010).

[10] A. T. Hanke and M. Ottens, *Purifying biopharmaceuticals: knowledge-based chromatographic process development,* Trends Biotechnol **32**, 210 (2014).

[11] S. M. Pirrung, L. A. M. van der Wielen, R. F. W. C. van Beckhoven, E. J. A. X. van de Sandt, M. H. M. Eppink, and M. Ottens, *Optimization of biopharmaceutical downstream processes supported by mechanistic models and artificial neural networks,* Biotechnology Progress , 696 (2017).

[12] T. C. Huuk, T. Hahn, and H. J. Osberghaus, Anna, *Model-based integrated optimization and evaluation of a multi-step ion exchange chromatography,* Sep. Purif. Technol. **136**, 207 (2014).

[13] T. Hahn, T. Huuk, V. Heuveline, and J. Hubbuch, *Simulating and optimizing preparative protein chromatography with chromx,* J. Chem. Educ. **92**, 1497 (2015).

[14] FDA, *Guidance for Industry PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance* (2004).

[15] ICH, *International Committee on Harmonization. Q8: Pharmaceutical Development* (2009).

[16] G. Guiochon, *Preparative liquid chromatography,* J. Chromatogr. A **965**, 129 (2002).

[17] O. Kaltenbrunner, O. Giaverini, D. Woehle, and J. Asenjo, *Application of chromatographic theory for process characterization towards validation of an ion-exchange operation,* Biotechnol. Bioeng. **98**, 201 (2007).

[18] B. Nilsson, *Aspects of modeling a preparative ion-exchange step for antibody purification,* Chem. Eng. Technol. **28**, 1367 (2005).

[19] C. Shene, A. Lucero, B. Andrews, and J. Asenjo, *Mathematical modeling of elution curves for a protein mixture in ion exchange chromatography and for the optimal selection of operational conditions,* Biotechnol. Bioeng. **95**, 704 (2006).

[20] M. Bensch, P. Schulze Wierling, E. von Lieres, and J. Hubbuch, *High Throughput Screening of Chromatographic Phases for Rapid Process Development,* Chem. Eng. Technol. **28**, 1274 (2005).

[21] M. Wiendahl, P. Schulze Wierling, J. Nielsen, D. Fomsgaard Christensen, J. Krarup, A. Staby, J. Hubbuch, and P. S. Wierling, *High-throughput screening for the design and optimization of chromatographic processes – Miniaturization, automation and parallelization of breakthrough and elution studies,* Chem. Eng. Technol. **31**, 893 (2008).

[22] J. Wölcke and D. Ullmann, *Miniaturized HTS technologies - uHTS.* Drug discovery today **6**, 637 (2001).

[23] T. Herrmann, M. Schröder, and J. Hubbuch, *Generation of equally sized particle plaques using solid-liquid suspensions,* Biotechnol. Prog. **22**, 914 (2006).

[24] J. Thiemann, J. Jankowski, J. Rykl, S. Kurzawski, T. Pohl, B. Wittmannliebold, and H. Schluter, *Principle and applications of the protein-purification-parameter screening system,* J. Chromatogr. A **1043**, 73 (2004).

[25] J. L. Coffman, J. F. Kramarczyk, and B. D. Kelley, *High-throughput screening of chromatographic separations: I. method development and column modeling,* Biotechnol. Bioeng. **100**, 605 (2008).

[26] T. Bergander, K. Nilsson-Välimaa, K. Oberg, and K. M. Lacki, *High-throughput process development: determination of dynamic binding capacity using microtiter filter plates filled with chromatography resin,* Biotechnol. Prog. **24**, 632 (2008).

[27] K. Rege, M. Pepsin, B. Falcon, L. Steele, and M. Heng, *High-throughput process development for recombinant protein purification,* Biotechnol. Bioeng. **93**, 618 (2006).

[28] M. Wiendahl, C. Völker, I. Husemann, J. Krarup, A. Staby, S. Scholl, and J. Hubbuch, *A novel method to evaluate protein solubility using a high throughput screening approach,* Chemical Engineering Science **64**, 3778 (2009).

[29] P. Diederich, S. Amrhein, F. Hämmerling, and J. Hubbuch, *Evaluation of peg/phosphate aqueous two-phase systems for the purification of the chicken egg white protein avidin by using high-throughput techniques,* Chem. Eng. Sci. **104**, 945 (2013).

[30] A. Susanto, K. Treier, E. Knieps-Grünhagen, E. von Lieres, and J. Hubbuch, *High throughput screening for the design and optimization of chromatographic processes: Automated optimization of chromatographic phase systems,* Chem. Eng. Technol. **32**, 140 (2009).

[31] K. Treier, S. Hansen, C. Richter, P. Diederich, J. Hubbuch, and P. Lester, *High-throughput methods for miniaturization and automation of monoclonal antibody purification processes,* Biotechnol. Prog. **28**, 723 (2012).

[32] K. Rege, N. Tugcu, and S. M. S. Cramer, *Predicting column performance in displacement chromatography from high throughput screening batch experiments,* Sep. Sci. Technol. **38**, 1499 (2003).

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

[34] K. Treier, P. Lester, and J. Hubbuch, *Application of genetic algorithms and response surface analysis for the optimization of batch chromatographic systems,* Biochem. Eng. J. **63**, 66 (2012).

[35] P. Diederich, S. K. Hansen, S. A. Oelmeier, B. Stolzenberger, and J. Hubbuch, *A sub-two minutes method for monoclonal antibody-aggregate quantification using parallel interlaced size exclusion high performance liquid chromatography,* J. Chromatogr. A **1218**, 9010 (2011).

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

[37] J. Kittelmann, M. Ottens, and J. Hubbuch, *Robust high-throughput batch screening method in 384-well format with optical in-line resin quantification,* J. Chromatogr. B **988**, 98 (2015).

[38] P. Diederich, M. Hoffmann, and J. Hubbuch, *High-throughput process development of purification alternatives for the protein avidin,* Biotechnology Progress **31**, 957 (2015).

[39] W. R. Keller, S. T. Evans, G. Ferreira, D. Robbins, and S. M. Cramer, *Use of minicolumns for linear isotherm parameter estimation and prediction of benchtop column performance,* J. Chromatogr. A **1418**, 94 (2015).

[40] P. F. O'Neill, A. Ben Azouz, M. Vázquez, J. Liu, S. Marczak, Z. Slouka, H. C. Chang, D. Diamond, and D. Brabazon, *Advances in three-dimensional rapid prototyping of microfluidic devices for biological applications,* Biomicrofluidics **8**, 052112 (2014).

[41] B. E. Rapp, *Microfluidics: modeling, mechanics and mathematics.* (Amsterdam: Elsevier, 2017) pp. lxiii, 766.

[42] J. El-Ali, P. K. Sorger, and K. F. Jensen, *Cells on chips,* Nature **442**, 403 (2006).

[43] H. Becker, *Mind the gap!* Lab Chip **10**, 271 (2010).

[44] A. Waldbaur, J. Kittelmann, C. P. Radtke, J. Hubbuch, and B. E. Rapp, *Microfluidics on liquid handling stations (f-on-lhs): an industry compatible chip interface between microfluidics and automated liquid handling stations,* Lab Chip **13**, 2337 (2013).

[45] J. Kittelmann, C. P. Radtke, A. Waldbaur, C. Neumann, J. Hubbuch, and B. E. Rapp, *Microfluidics on liquid handling stations (µf-on-lhs): a new industry-compatible microfluidic platform,* Proc. SPIE **8976**, 89760G (2014).

[46] E. Häusler, P. Domagalski, M. Ottens, and Bardow, *Microfluidic diffusion measurements: The optimal H-cell,* Chem. Eng. Sci. **72**, 45 (2012).

[47] C. P. Radtke, M.-T. Schermezer, Y. C. Zhai, J. Göpper, and J. Hubbuch, *Implementation of an analytical microfluidic device for the quantification of protein concentrations in high□throughput format,* Eng. Life Sci. **16** (2016).

[48] S. H. Hansen, T. Kabbeck, C. P. Radtke, S. Krause, E. Krolitzki, T. Peschke, J. Gasmi, K. S. Rabe, M. Wagner, H. Horn, J. Hubbuch, J. Gescher, and C. M. Niemeyer, *Machine-assisted cultivation and analysis of biofilms,* bioRxiv (2017).

[49] C. P. Radtke, N. Hillebrandt, and J. Hubbuch, *The biomaker: an entry□level bioprinting device for biotechnological applications,* J. Chem. Technol. Biotechnol. **93** (2017).

[50] C. Hansch and T. Fujita, *ρ-σ-π analysis. a method for the correlation of biological activity and chemical structure,* J. Am. Chem. Soc. **86**, 1616 (1964).

[51] C. Hansch, P. Maloney, and T. Fujita, *Correlation of biological activity of phenoxyacetic acids with hammett substituent constants and partition coefficients,* Nature **194**, 178 (1962).

[52] H. Kubinyi, *QSAR and 3D QSAR in drug design Part 1: methodology,* Drug Discov. Today **2**, 457 (1997).

[53] A. G. Mercader, P. R. Duchowicz, F. Fernandez, E. a. Castro, and F. M. Fernández, *Modified and enhanced replacement method for the selection of molecular descriptors in QSAR and QSPR theories,* Chemometr. Intell. Lab. **92**, 138 (2008).

[54] A. Tropsha, *Predictive quantitative structure–activity relationship modeling,* in *Comprehensive Medicinal Chemistry II*, Vol. 4, edited by Y. C. Martin (Elsevier, Amsterdam, 2006) pp. 149–165.

[55] R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors* (Wiley-VCH, 2008).

[56] A. Tropsha, *Best practices for qsar model development, validation, and exploitation,* Mol. Inf. **29**, 476 (2010).

[57] A. Tropsha and A. Golbraikh, *Predictive QSAR modeling workflow, model applicability domains, and virtual screening,* Curr. Pharm. Des. **13**, 3494 (2007).

[58] T. U. Consortium, *Reorganizing the protein space at the Universal Protein Resource (UniProt),* Nucleic Acids Res. **40**, D71 (2012).

[59] T. U. Consortium, *Activities at the Universal Protein Resource (UniProt),* Nucleic Acids Res. **42**, D191 (2014).

[60] C. B. Mazza, N. Sukumar, C. M. Breneman, and S. M. Cramer, *Prediction of protein retention in ion-exchange systems using molecular descriptors obtained from crystal structure,* Anal. Chem. **73**, 5457 (2001).

[61] 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**, 164 (2006).

[62] F. Dismer and J. Hubbuch, *A novel approach to characterize the binding orientation of lysozyme on ion-exchange resins,* J. Chromatogr. A **1149**, 312 (2007).

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

[64] K. M. 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 (2015).

[65] J. Robinson, D. Roush, and S. Cramer, *Domain contributions to antibody retention in multimodal chromatography systems,* J. Chromatogr. A , 89 (2018).

[66] T. Yang, M. C. Sundling, A. S. Freed, C. M. Breneman, and S. M. Cramer, *Prediction of pH-dependent chromatographic behavior in ion-exchange systems,* Anal. Chem. **79**, 8927 (2007).

[67] C. B. Mazza, K. Rege, C. M. C. Breneman, N. Sukumar, J. S. Dordick, and S. M. S. Cramer, *High-throughput screening and quantitative structure-efficacy relationship models of potential displacer molecules for ion-exchange systems,* Biotechnol. Bioeng. **80**, 60 (2002).

[68] L. Xu and C. E. Glatz, *Predicting protein retention time in ion-exchange chromatography based on three-dimensional protein characterization,* J. Chromatogr. A **1216**, 274 (2009).

[69] A. Ladiwala, F. Xia, Q. Luo, C. M. Breneman, and S. M. Cramer, *Investigation of protein retention and selectivity in HIC systems using quantitative structure retention relationship models,* Biotechnol. Bioeng. **93**, 836 (2006).

[70] J. Liu, T. Yang, A. Ladiwala, S. M. Cramer, and C. M. Breneman, *High throughput determination and QSER modeling of displacer DC-50 values for ion exchange systems,* Sep. Sci. Technol. **41**, 3079 (2006).

[71] I. Sushko, S. Novotarskyi, R. Körner, A. Pandey, M. Rupp, W. Teetz, S. Brandmaier, A. Abdelaziz, V. Prokopenko, V. Tanchuk, R. Todeschini, A. Varnek, G. Marcou, P. Ertl, V. Potemkin, M. Grishina, J. Gasteiger, C. Schwab, I. Baskin, V. Palyulin, E. Radchenko, W. Welsh, V. Kholodovych, D. Chekmarev, A. Cherkasov, J. Aires-de Sousa, Q.-Y. Zhang, A. Bender, F. Nigsch, L. Patiny, A. Williams, V. Tkachenko, and I. Tetko, *Online chemical modeling environment (ochem): web platform for data storage, model development and publishing of chemical information,* J. Comput. Aided Mol. Des. **25**, 533 (2011).

[72] E. E. Bolton, Y. Wang, P. A. Thiessen, and S. H. Bryant, *Chapter 12 - PubChem: Integrated Platform of Small Molecules and Biological Activities,* Annu. Rep. Comp. Chem., **4**, 217 (2008).

[73] T. I. Oprea and A. Tropsha, *Target, chemical and bioactivity databases – integration is key,* Drug Discov. Today **3**, 357 (2006).

[74] A. J. Williams, V. Tkachenko, C. Lipinski, A. Tropsha, and S. Ekins, *Free online resources enabling crowd-sourced drug discovery,* Drug Discovery World , 33 (2000).

[75] B. P. Hudson, B. Narayanan, C. L. Lawson, C. Shao, D. Dimitropoulos, E. Peisach, G. Gao, H. M. Berman, H. Yang, I. Persikova, J. D. Westbrook, L. Chen, L. Tan, L. Di Costanzo, M. A. Zhuravleva, M. R. Sekharan, R. Sala, S. Ghosh, V. Guranovic, Y. Liang, Z. Feng, J. Y. Young, S. K. Burley, A. Mukhopadhyay, A. Gutmanas, A. R. Clark, A. Patwardhan, D. R. Armstrong, E. Sanz-García, G. J. Swaminathan, G. Sahni, G. J. Kleywegt, G. van Ginkel, J. M. Berrisford, L. Mak, O. S. Smart, P. M. S. Hendrickx, S. Velankar, S. Sen, S. Gore, T. J. Oldfield, G. Kurisu, H. Nakamura, J. Sato, K. Nishiyama, M. Chen, R. Igarashi, Y. Ikegawa, Y. Kengaku, J. L. Markley, and K. Baskaran, *Worldwide Protein Data Bank biocuration supporting open access to high-quality 3D structural biology data,* Database **2018** (2018).

[76] J. Inglese, C. E. Shamu, and R. K. Guy, *Reporting data from high-throughput screening of small-molecule libraries,* Nat. Chem. Biol. , 438 (2007).

[77] F. Boukouvala, F. J. Muzzio, and M. G. Ierapetritou, *Design space of pharmaceutical processes using data-driven-based methods,* J. Pharm. Innov. **5**, 119 (2010).

2

Robust high-throughput batch screening method in 384-well format with optical in-line resin quantification

High-throughput batch screening technologies have become an important tool in downstream process development. Although continuative miniaturization saves time and sample consumption, there is yet no screening process described in the 384-well microplate format. Several processes are established in the 96-well dimension to investigate protein-adsorbent interactions, utilizing between 6.8 and 50 µL resin per well. However, as sample consumption scales with resin volumes and throughput scales with experiments per microplate, they are limited in costs and saved time. In this work, a new method for in-well resin quantification by optical means, applicable in the 384-well format, and resin volumes as small as 0.1 µL is introduced. A HTS batch isotherm process is described, utilizing this new method in combination with optical sample volume quantification for screening of isotherm parameters in 384-well microplates. Results are qualified by confidence bounds determined by bootstrap analysis and an comprehensive Monte Carlo study of error propagation. This new approach opens the door to a variety of screening processes in the 384-well format on HTS stations, higher quality screening data and an increase in throughput.

Parts of this chapter have been published in J. Chromatogr. B [1].

2.1. Introduction

High-throughput screening (HTS) techniques in downstream process development have established new approaches and lead to significant advances in productivity and time to optimized processes. High-throughput batch screening methods are an important tool in early stage development of chromatography based purification strategies, due to their advantage in throughput. Batch chromatography experiments in HTS scale gain further importance with the development of computational models for large scale chromatography column runs, relying on protein-adsorbent interaction parameters [2–4]. Different batch screening technologies have been developed in recent years. Resin volumes distributed to filter plates have been used to screen preferable binding and elution conditions for biomolecules [5–7]. Hermann et al. developed a method to prepare equally sized resin plaques [8] which have been used in batch screening experiments [9]. Commercially available filter plates with pre-packed chromatographic materials were used to determine dynamic binding capacities [10]. Wenger et al. used pipetting tips filled with resin to purify virus-like particles in a HTS application [11].
Despite different approaches and ongoing advances in the development of HTS hardware, the 96-well format has remained the limit in throughput and sample size.

Data quality remains the challenge in HTS process development, as processing uncertainties gain significance with decreasing volumes. With increasing process complexity, single step error propagation gains impact on data quality. Process downscale is limited by the volumes of sample solutions and resins which can be handled reproducible. At the same time, sample consumption per experiment often scales with the resin volume utilized, resulting in high sample consumption for complex design spaces to be screened, despite the HTS approach.
If it is possible to overcome the limitation of erroneous, miniature volume by quantification, rather than accurate volume handling, the range of 384-well would be made accessible for applications in process development. Therefore in this work we solved the quantification problem via light extinction measurements in the sub-100 µL volume range. Optical measurements can be used to quantify volumes as well as suspensions of particles. Methods to quantify volumes in microtiter plates by vertical beam photometers utilizing the absorbance of the solvent water have been described and utilized in extent, confirming reliability [12, 13].

Particles can also be quantified, as they absorb and scatter light, based on concentration and particle size distribution. In the limitation of small concentrations, the scattering of light by particles follows the Lambert-Beer law, as light beams impinge on single particles, they will most probably be represented as single spots on a sensor plate [14]. At higher concentrations particle-particle interactions occur. The linear correlation of particle concentration and light extinction does not hold true as with higher particle concentrations radiation undergoes interaction with multiple particles. Light scattered by one particle will hit a second one and so on, leading to an increase in transmission, compared to the Lambert-Beer law. According to the hard core model, particles at high concentrations will also expose more surface to radiation as they do not interpenetrate each other. This results in a decrease in

radiation transmitted to the sensor plane. [14–16]

Despite the HT approach, experimental data is limited, therefore measurements of statistical validity for reproducibility and fits of mechanistic models should be provided for such screening processes. Re-sampling techniques, as bootstrapping and Monte Carlo analysis, can be used to assess the distribution of experimental data points, and parameters derived, as well as error propagation in complex processes. Bootstrapping, a random re-sampling method, and non-parametric statistical techniques in general, can be used to analyze data without assuming a particular probability distribution. Those distribution-free methods can be applied to a wide variety of statistical problems and do not require extensive assumptions on data distributions to validate analysis. The reasoning behind and an extensive description of this statistical method can be found in detail in [17–19].

Monte Carlo simulations allow for *in silico* calculation of process errors, given the description of the process in mechanistic equations and single process step uncertainties being quantified. This allows to analyze the influence of single process steps, and their associated uncertainties, on the overall experimental results. In-detail explanation of Monte Carlo techniques can be found in literature, e.g. [20–22]. Despite their advantages, examples of statistical validation of HTS results in literature are sparse. Kurup et al. describe a Monte Carlo error estimation in simulated moving bed chromatography [23]. Osberghaus et al. evaluated the error propagation in a HTS isotherm process, but limited the investigation to the effects on single measurement points, rather than isotherm parameters estimated [24].

In this work, signal extinction due to particle light scattering is utilized to accurately quantify volumes of adsorbent resins, distributed to microtiter plates. The applicability of this technique is shown for resins of different particle sizes and backbone composition. This new approach in chromatographic resin quantification is utilized in a newly developed automated batch isotherm HTS process in the 384-well format. The automated batch screening process presented here features an optical quantification of resin and volumes pipetted into a 384-well microtiter plate and yields 384 measurement points in a run time of approximately 4 hours. Dependent on sample layout, this equals 12 (32 data points each) to 24 (16 data points) measured isotherms in one process cycle. Isotherm parameter estimations are evaluated by a bootstrap re-sampling method. Process uncertainties are quantified and their impact on parameter estimation is assessed by Monte Carlo simulation.

2.2. Materials and Methods

2.2.1. Materials

Lysozyme from chicken egg white was purchased from Sigma-Aldrich (St. Louis, MO, USA). Sodium phosphate, sodium hydroxide and sodium chloride were purchased from Merck KGaA (Darmstadt, Germany). Strong cation-exchange adsorbent SP Sepharose FF was purchased from GE Healthcare (Buckinghamshire, United

Kingdom), adsorbents Toyopearl SP 650M and SP 650C were acquired from Tosoh Bioscience GmbH (Stuttgart, Germany). Microtiter plates UVStar 96-well and UVStar 384-well (both plane bottom F-shape) were obtained from Greiner Bio-One GmbH (Frickenhausen, Germany).

2.2.2. Equipment

Resin plaques of defined volume of 7.8 and 20.8 µL, respectively, were produced with a ResiQuot device from Atoll-Bio (Weingarten, Germany) according to instructions from [8]. The batch process was automated on a Tecan Freedom Evo 200 robotic workstation (Tecan, Maennedorf, Switzerland). Main features of this station are a liquid handling arm (LiHa) consisting of eight separately controllable pipetting channels, each equipped with a fixed teflon coated pipetting tip and driven by a 1 ml dilutor. A gripper (RoMa) was used for plate transfer on the workstation. Pipetting in 384-well plates was performed by consecutive use of a 96 channel pipetting head (MCA96) which was equipped with disposable tips of 200 µL volume. An integrated Hettich Rotanta 46RSC centrifuge (Hettich GmbH, Tuttlingen, Germany) was used for centrifugation of microtiter plates. An infinite 200M spectrophotometer (Tecan, Maennedorf, Switzerland) was utilized for optical measurements. Microtiter plates and disposable tips were stored in two storage units (Te-Stack) at the workstation and transferred to the worktable as needed. Pipetted liquid volumes were quantified with an analytical scale X S250 from Mettler-Toledo (Greifensee, Switzerland) which was integrated in the workstation.
Software EVOware 2.5 was used to program automated work-flows on the liquid-handling workstation. The spectrophotometer was controlled by the software Magellan 7.1, allowing for predefinition of measurement wavelengths and positions in well. Data evaluation and Monte Carlo simulations were performed in Matlab 8.0 (The Mathworks, Natick, ME, USA).

2.2.3. Resin quantification

Samples in microtiter plates were centrifuged for 1 min at 2000 rpm prior to optical measurement to ensure even meniscus and full sedimentation of resin particles. Light extinction due to light scattered by adsorbent particles was measured at 330 nm. A grid of 6 by 6 evenly distributed measurement points, with a distance of 50 µm to the well wall, were measured in each well as shown in figure 2.2. Measurement values were averaged to account for uneven distribution of sedimented resin beads in the well.
Extinction coefficients for adsorbent materials were determined by dilution series in 384-well plates. Resin plaques of defined volumes were prepared in 96-well plates by use of the ResiQuot device and suspended in 300 µL buffer solution. Volumes of suspended resins were transferred to 384-well plates by pipetting, ensuring even distribution of adsorbent beads in suspension.

2.2.4. Batch binding process

The initial distribution of resin to a 96-well plate was performed manually utilizing the ResiQuot device. Resin plaques of 7.8 µL volume for Toyopearl SP 650C and SP 650M and 20.8 µL for Sepharose adsorbent were distributed in a 96-well plate as described in [8]. Plaques were dissolved in 300 µL 10 mM sodium phosphate buffer at pH 7.0, with varying sodium chloride concentrations, according to subsequent binding conditions. Protein stock solutions of 7.0 $\mathrm{mg\,mL^{-1}}$ concentration were prepared in the same buffer.

The robotic work-flow, depicted in Figure 2.1 consists of the following process steps:

1. Resin slurry, provided in a 96-well plate, is transferred to a 384-well plate in 30 µL buffer volume ($V_{MCA,Resin}$) by four consecutive pipetting steps utilizing the MCA96 pipetting device with disposable tips of 200 µL volume. Slurry was pipetted up and down three times prior to transfer, to ensure even distribution of resin beads in solution.
2. Resin volume per well is quantified by optical measurement of light extinction at 330 nm wavelength.
3. Dilution series of protein concentrations are pipetted in two 96-well plates by means of 8-channel LiHa ($V_{LiHa,Buffer}$, $V_{LiHa,Protein}$). A random distribution of protein concentration within single dilution series is provided, to circumvent potential bias due to pipetting order, associated with single pipetting needles.
4. 70 µL of protein samples are pipetted from the dilution series in 96-well plates to 384-well plate ($V_{MCA,Sample}$) by consecutive use of MCA96 pipetting device.
5. Samples are sealed with adhesive aluminum foil (Sarstedt AG & Co, Nümbrecht, Germany) and incubated in over-head shaker for one hour, to allow the binding equilibrium to be reached.
6. Sample plate is centrifuged for 1 min at 2000 rpm to settle resin beads and 30 µL supernatant is transferred to a 384-well UV plate and subsequently diluted with 70 µL buffer solution.
7. Protein concentration in supernatant is determined by absorption measurement of samples in measurement plate at 280 nm.

Each pipetting step is followed by a volume quantification. For this purpose samples were centrifuged for 1 min at 2000 rpm to ensure even meniscus at volume surface. Absorption was measured at 990 and 900 nm wavelength. Volumes were calculated based on (990 - 900) nm coefficient, and an absorption coefficient gained from manual calibration in an identical microtiter plate with 10 equidistant volumes from 20 to 120 µL. Pipetting and measurement steps were fully automated, excluding manual sample preparation and sample incubation on an external over-head shaker, which required the microtiter plate to be sealed.

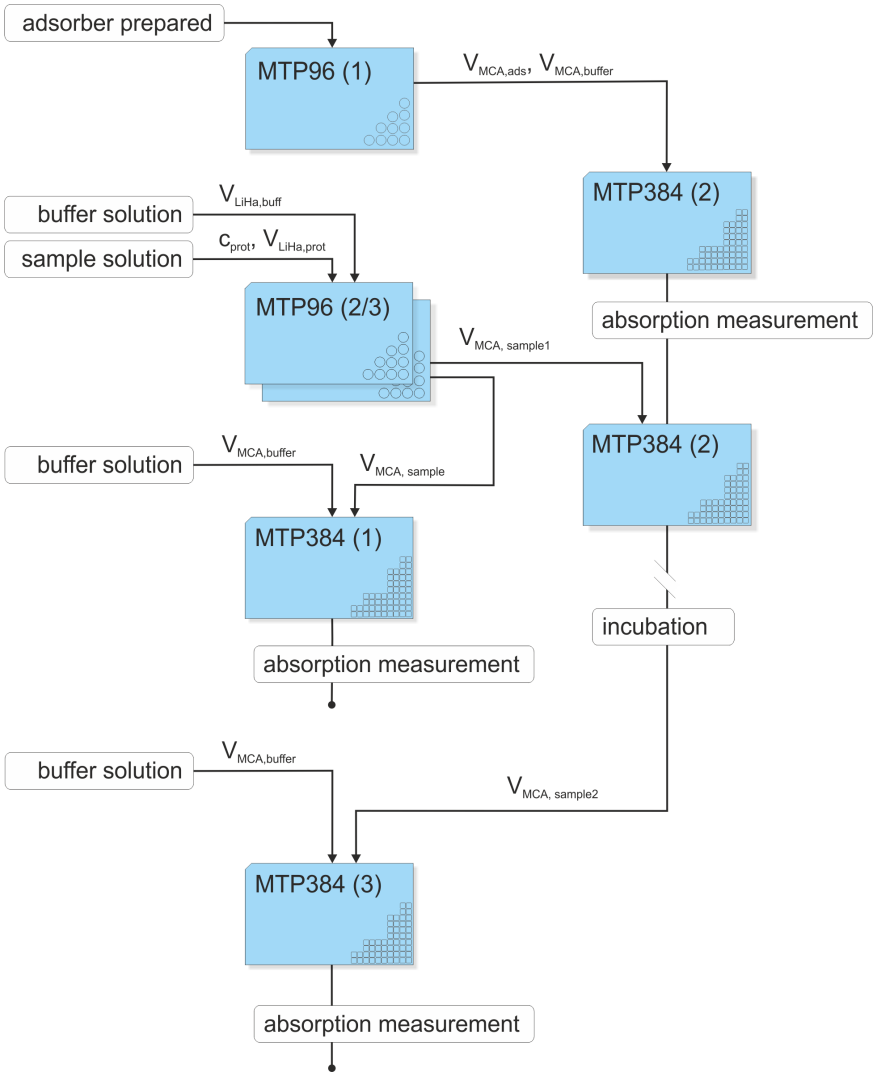

Figure 2.1: Schematic of the batch isotherm process layout for automated processing on a LHS. Left hand side solutions were manually prepared and provided at the LHS. For sample incubation the microtiter plate was manually sealed and stored in an over head shaker.

2.2.5. Langmuir parameter and confidence interval estimation

In batch binding to an adsorbent, protein molecule distribution at equilibrium follows a saturation curve which can be modelled with the Langmuir isotherm equation:

$$q = q_{max} \cdot K_a \cdot \frac{c}{1 + K_a c} \tag{2.1}$$

with q concentration of protein bound to adsorbent, q_{max} saturation coefficient

and K_a association coefficient. Langmuir equation was fitted by least square regression algorithm.

Re-sampling or bootstrapping methods can be applied were no assumptions can be drawn for the distribution of a population. A random sample from this population is used as an estimate for the distribution of the parent population. The bootstrap can further be used to calculate the value and accuracy of an estimate such as a confidence interval [25, 26]. The true estimate of a population is denoted θ. A set of randomly sampled values n from the population, yields the sample estimate $\hat{\theta}$. Sampling n values with replacement from this set provides a bootstrap sample, with a bootstrap estimate θ^*. In this study sampling is repeated 1000 times, each producing a bootstrap estimate θ^*. With the random sample being a representation of the underlying population, the Monte Carlo approximation of the sampling distribution of $\theta^* - \hat{\theta}$ was used to estimate $\hat{\theta} - \theta$. Therefore the 2.5 and 97.5 percentiles of the distribution of $\theta^* - \hat{\theta}$ estimate the $95\ \%$ confidence interval of the population based on the sample drawn.

2.2.6. Error quantification

UV reader measurement errors were assessed by 15 repeated measurements of samples in a spectrophotometer. Sample plates were removed from photometer and centrifuged for 1 min at 2000 rpm and returned to the reader in between single measurements. Measurement parameters were set according to in-process parameters. Protein concentration was measured at 280 nm wavelength. Volumes were quantified by water absorption, measured at 990 and 900 nm and calculated as outlined in Section 2.2.4. Samples containing adsorbent beads were put on an orbital shaker for 30 sec to re-suspend settled particles, ensuring random re-distribution in well. Adsorbent volume quantification by measurements at 330 nm wavelength were laid out in a grid of 6 by 6 points as described in Section 2.2.3.
Pipetting accuracy of 8-channel liquid handler was measured in a range from 10 to 300 µL by pipetting a protein solution to an analytical scale, positioned on the work table. Volume distributions for distinct sample volumes were calculated from recorded weights after correction of sample density, based on 400 measurements each.
Pipetting accuracy of 96-channel liquid handler (MCA96) for buffer and protein solutions at volumes of 30 and 70 µL, as used in this batch binding process, was quantified by measurement of the near infra-red absorbance of water in aqueous reagents. Samples pipetted into 384-well plates by MCA96 channel pipetting device were measured at 990 and 900 nm wavelength. Path length was calculated by (990 - 900) nm absorbance difference according to [13]. Corresponding volumes were calculated based on well geometry.

2.2.7. Error distribution and confidence bounds

Generalized Extreme Value (GEV) distribution was fitted to the data for error estimation and confidence bound calculation of bootstrap results and Monte Carlo

simulations. The GEV distribution combines three simpler distributions, namely the Gumbel, Fréchet and Weibull families, into a single form, allowing to fit a continuous range of possible shapes of underlying distributions. This overcomes the restriction in tail behavior of standard distributions. A random variable X is said to have a generalized extreme value distribution if its probability density function (pdf) is given by:

$$f(x|\mu,\sigma,\xi) = \frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{(-1/\xi)-1} \cdot \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \tag{2.2}$$

where σ and μ are respectively the scale and location parameters, and ξ is a shape parameter. ξ governs the tail behavior of the GEV distribution and defines the underlying distribution in use. Type I distributions are defined with $\xi = 0$ (Gumbel type), corresponding to exponentially decreasing tails. Type II refers to $\xi > 0$, the Fréchet type where tails decrease as polynomial. Type III distributions with $\xi < 0$ correspond to Weibull type distributions with finite tail.
For experimentally determined confidence bounds in liquid volumes handled by LiHa, normal distributions were assumed and standard deviations of multiple measurements were calculated according to:

$$s = \left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2\right)^{\frac{1}{2}} \tag{2.3}$$

with $\overline{x}$ the mean of samples and n the number of elements in the sample.

2.2.8. Monte Carlo simulation

The Monte Carlo algorithm was programmed to mechanistically simulate isotherm binding studies on LHS according to the process depicted in Figure 2.1. This allows to assess the effect of single process step variance and different process layouts on Langmuir isotherm parameters. In a first step the Monte Carlo algorithm calculates the initial state, sample concentration and pipetting volumina based on the assumption of a specific equilibrium behavior described by Langmuir parameters provided. In the second part of the simulation algorithm all deviations are applied to the given constants, according to the distributions experimentally determined. Errors are applied according to:

$$\widetilde{V_x} = V_x + \epsilon_x \tag{2.4}$$

with the independent random error ϵ_x, calculated corresponding to the underlying distribution.
In the third step, Langmuir isotherm parameters are calculated by non-linear least

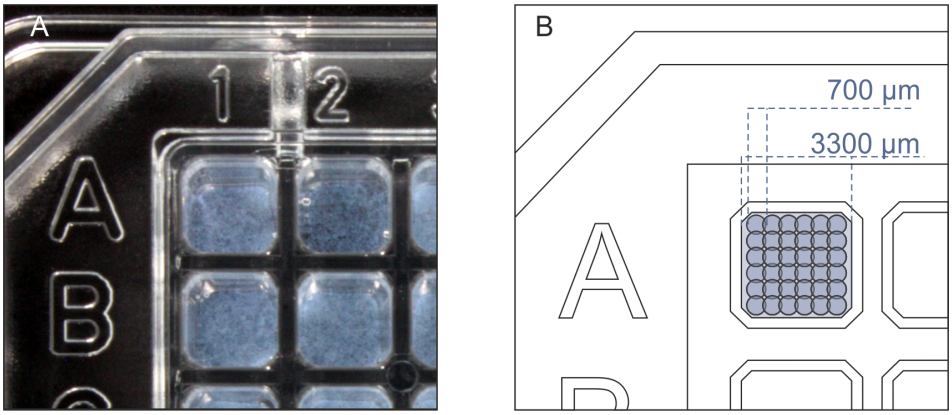

Figure 2.2: Picture of well with SP Sepharose FF resin (approx. 1.5 to 2.5 µL resin volume), schematic of a rectangular well in 384-well plate and measurement point (d = 700 µm) distribution in a 6 by 6 point grid, with 50 µm spacing to well wall.

square regression according to Equation 2.1. Steps two and three were repeated 10 000 times to determine variance of Langmuir parameters. Resulting Langmuir parameter distributions were fitted by GEV formalism (see Section 2.2.7).
For Monte Carlo simulation, variances between process steps are assumed to be independent. Error from manual sample preparation, e.g. weighing of protein and buffer preparation as well as the influences of shaking and centrifugation of sample are not in the scope of this study. For volumes pipetted by LiHa and volumes assessed by photometric measurements, standard deviations are assumed in agreement with [24].

2.3. Results and Discussion

2.3.1. Resin quantification

To quantify adsorbent bead volumes in the µL scale in microtiter plates, photometric measurements have been performed at 330 nm wavelength. Figure 2.3 shows the calibration of light extinction value over resin volumes for chromatographic adsorbent SP Sepharose FF, SP 650M and SP 650C Toyopearl. Linear dependencies could be identified for volumes up to 1.7 (SP Seph. FF), 2.1 (SP 650M), and 4.7 µL (SP 650C) per well in a 384-well plate containing 100 µL sample. The slope of light extinction over concentration differs for individual resins investigated. Analyzed resins feature different backbones and different bead sizes, with the same active group being sulfopropyl. SP Sepharose FF with a sepharose backbone shows lowest light extinction as beads are most porous, since composed of cross-linked agarose strings. Toyopearl resins feature a hydroxylated meth- acrylic polymer backbone which is more dense and has a higher opacity. Toyopearl 650C with a mean particle size of 100 µm, blocks more light from the sensor than 650M with a mean diameter of 65 µm.

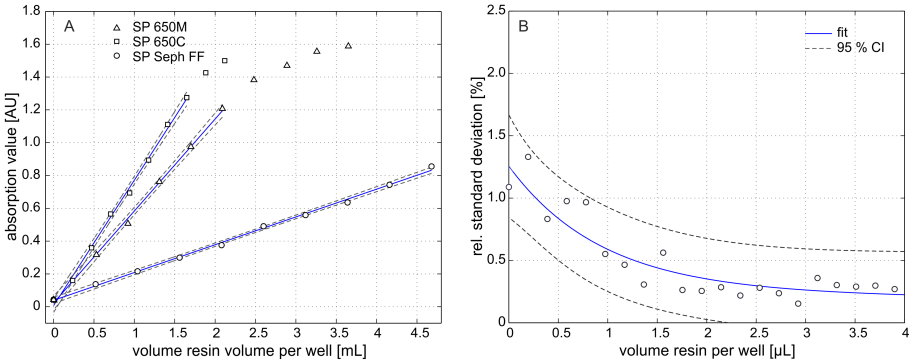

Figure 2.3: A) Calibration of resin, deviation in calibration curve due to manual pipetting, (low errors for reproducibility of quantification measurement); deviation for larger concentrations due to multiple scattering and steric hindrance of particles. B) Measurement error for SP Sepharose FF quantification by spectrophotometric measurement at 330 nm, in a 6 by 6 grid layout.

For concentrations higher than the linear range, the slope decreases, indicating an increasing contribution of multiple light scattering. This supports the assumption that large parts of light extinction are due to reflection, rather than absorption of light at 330 nm wavelength.
Repeated measurements of resin volumes, with interim redistribution of adsorbent beads in the well as described in Section 2.2.6 show high reproducibility. Figure 2.3 exemplarily shows the reproducibility for samples of SP Sepharose FF resin ranging from below 0.5 % relative standard deviation for resin volumes above 1.0 µL. Variance remains small with relative standard deviation below 1.0 % for volumes above 0.5 µL and below 1.5 % for resin volumes as small as 0.2 µL per well. The variance measured for the blank value 0 µL resin per well at 1.1 % rel. std. deviation shows the reproducibility of the photometer, indicating that the variance in measurements is due to the detection limitation of the photometer. Therefore resin redistribution in well is covered by the introduced grid layout of multiple measurements.

2.3.2. HTS batch chromatography

Isotherm experiments were performed on a LHS according to the process description in Section 2.2.4. Isotherm data for Lysozyme at pH 7.0 on SP Sepharose FF resin for varying ionic strength is shown in Figure 2.4. Confidence bounds were assessed by a bootstrapping method described previously and are depicted in Table 2.1. Isotherms show consistent data and good agreement of calculated Langmuir parameters to previously published results for similar conditions [27]. A comparison of data quality was not possible as confidence intervals are rarely described in published literature.
The process was performed in 4 hours, including manual sample preparation and 1 hour incubation of samples, thereby yielding 384 data points. Average resin volume per well was 1.8 µL, ranging between 1.6 and 2.0 µL per well. With typical

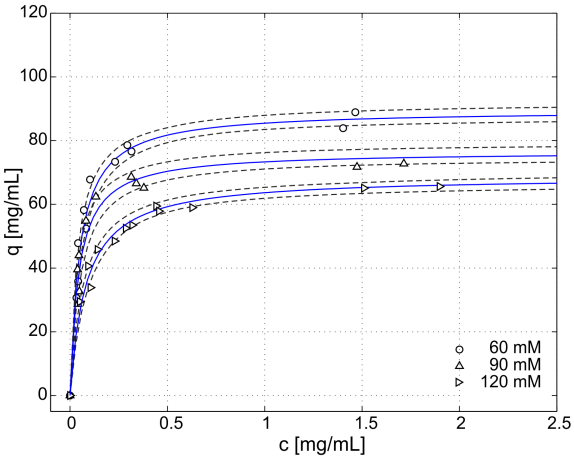

Figure 2.4: Isotherms of Lysozyme on SP Sepharose FF, pH 6.0 and ionic strength of 60, 90 and 120 mM. Grey lines indicate 95 % confidence bounds of fitted Langmuir isotherm equation.

IS (mM)	k_a coefficient			q_{max} [g L^{-1}]		
	value	confidence interval		value	confidence interval	
60	20.90	16.82	27.60	89.49	87.50	92.08
90	22.53	16.67	31.39	76.55	74.48	79.68
120	12.46	9.69	15.50	68.78	66.74	70.75

Table 2.1: Langmuir isotherm parameters for Lysozyme on SP Sepharose FF adsorbent at pH 7.0 and different ionic strengths and corresponding 95 % confidence interval.

maximum adsorber capacities in a range of 100 to 120 mg protein per milliliter adsorber, this corresponds to 0.22 mg protein per well or per data point. With 16 data points used to characterize an isotherm, 24/12 isotherms can be performed in singular/duplicate measurement with 3.46/6.91 mg protein consumed.

2.3.3. Error estimation

Overall process error results from propagated errors, introduced during single process steps. LHS operations pipetting by LiHa and MCA96, as well as photometric measurements, have been examined in terms of accuracy and error distribution. Figure 2.5 A provides the relative standard deviation of volumes pipetted over the targeted volume. Standard deviation is below 0.5 % percent for volumes above 100 µL and increases exponentially towards smaller volumes, reaching 1.0 % for 40 µL and 3.5 % for 20 µL. Figure 2.5 B shows the relative standard deviation of volumes calculated by 990-900 nm photometric method described in Section 2.2.4. The method's reproducibility is comparable to LiHa pipetting, exhibiting the exponential increase for volumes below 40 µL.

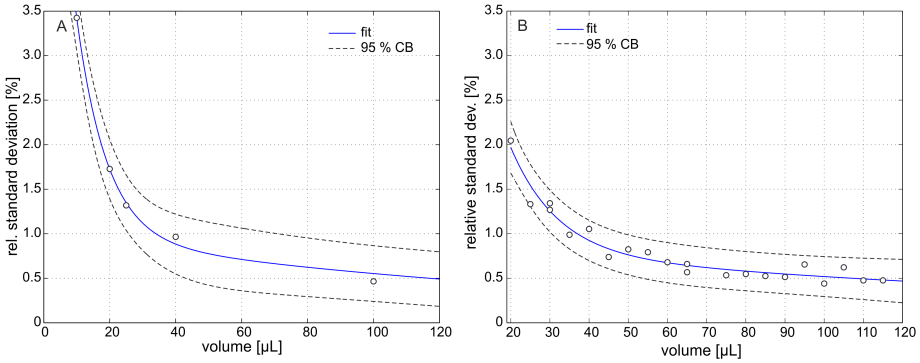

Figure 2.5: A) σ of pipetting accuracy for LiHa 8-channel pipetting head at different volumes pipetted. A sharp increase in deviation can be seen for volumes below 50 µL. B) Standard deviation of spectrophotometric measurement with n = 20.

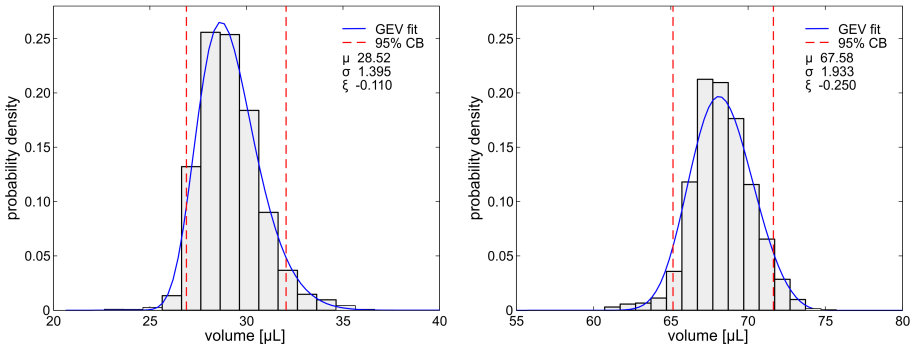

Figure 2.6: Histogram of volumes pipetted by MCA96, with n=1920. A) Distribution of dispensed volumes, with 30 µL and B) 70 µL target volume.

Pipetting by the MCA96 multi-channel device shows a higher variance in volumes pipetted. Figure 2.6 shows the distribution of volumes pipetted and 95 % confidence bounds as calculated from 97.5 percentiles. The relative standard deviation, is about 5.0 % and 3.0 % for 30 and 70 µL target volume respectively.
Pipetting errors for LiHa are in range with manufacturer specifications and published values for comparable investigations. MCA96 pipetting device reproducibility is below specifications given for small volumes. Reproducibility and accuracy could be increased by improved liquid class conditioning. Although despite extensive possibilities of liquid class adjustment, higher variations must be assumed for more viscous samples as well as samples that can not be characterized prior to processing.
For investigated conditions the error by photometric measurement is significant smaller than the one achieved by the liquid sample handling, independent of the character of sample, i.e. buffers, protein solutions, or dispensed resin particles.

Also, processes can be designed to dispense larger volumes first into a well when combining different liquids, hence avoiding more error prone quantification of small volumes. Based on these results, more accurate processes can be reached, if handled volumes are quantified by well calibrated measurements than by - even the best conditioned - pipetting steps.

2.3.4. Monte Carlo simulation

Monte Carlo simulations were performed according Section 2.2.8 to investigate the accuracy of Langmuir parameters gained from experiments, conducted with the developed experimental method. Fig. 2.7 shows parameter distributions calculated for conditions equivalent to those obtained for Lysozyme at ionic strength of 60 and 120 mM and 8/16/32 experimental data points per isotherm. It was found that affinity coefficient k_a as well as q_{max} are strongly underestimated for conditions limited to 8 data points with distribution maxima at 18.4 and 8.8 for k_a and 66.2 and 66.4 $\mathrm{mg\,mL^{-1}}$ for q_{max} values at low and high salt conditions. With 16 or more data points, the parameter distribution means converge to the set values. As expected, confidence bounds merge as the number of data points increases. Distributions of both parameters tend to be skewed towards smaller values, an effect which is reduced with increasing data points.
Experimental data points are usually set up in an equidistant manner, concerning the protein concentration in solvent. This leads to a higher density of data points in the linear range of the isotherm in case of a steep slope (= high k_a values). Subsequently, shallow slopes result in a lower data density in the linear range of the isotherm. Therefore data distribution vary between regions of the isotherm which define the linear slope and saturation area, respectively. Mathematical fits of Langmuir equation to data with experimental noise will therefore over- or underrepresent one over the other, k_a or q_{max}.

Both error analysis methods, Monte Carlo and bootstrapping, are suitable to analyze distributions since both approaches show the skewness of distributions and stronger tailing towards higher values for both parameters. Since both methods lead to similar results, the method can be selected according to the knowledge demands: Bootstrapping is the method of choice for analyzing the quality of experimental data as no prior knowledge on data inaccuracy and type of error distribution is necessary. When analyzing the influence of errors from single process steps or when analyzing error distributions under unknown, not-sampled experimental conditions, Monte Carlo simulations can be used. In MC no explicit experimental data points are necessary as long as process steps are characterized with regard to error distributions. When process steps are well characterized, Monte Carlo simulations additionally are helpful for designing future experiments as expected data quality can be assessed beforehand. Therefore screenings can be designed for most efficient sample use, running only experiments with expected most gain in knowledge. This approach increases throughput and decreases material usage which are the central idea of miniaturized high-throughput screenings.

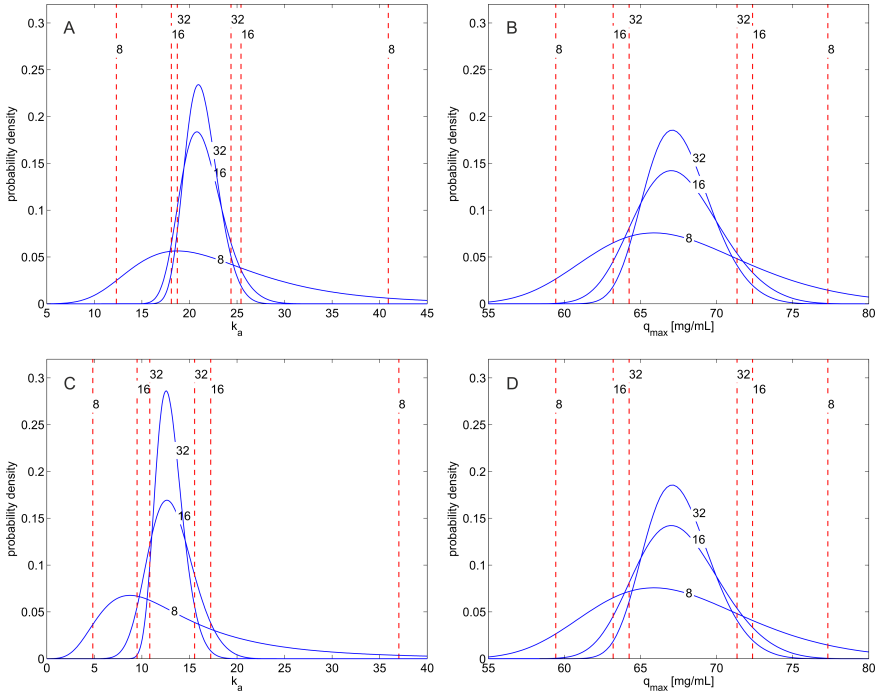

Figure 2.7: Langmuir parameter distributions calculated by Monte Carlo analysis for A) k_a and B) q_{max} parameter under strong binding conditions (equivalent to results for pH 7.0, 60 mM NaCl) and C+D) weak binding conditions (pH 7.0, 120 mM NaCl). Dotted lines indicate 95% confidence bounds for 8/16 and 32 equidistant data points per isotherm.

2.4. Conclusion and outlook

The resin quantification method by means of light extinction measurements in the 384-well microplate format, proved to be a fast, robust, and accurate method for smallest volumes of adsorbent particles dispensed in solution. A measurement grid of 6 by 6 points per well showed to give reproducible results in a volume range suitable for HTS experiments. Working volumes for resins of different structure have been established with 0.1 to 4.5 µL per well for SP Sepharose FF and 0.1 to 1.2 µL per well for Toyopearl SP 650M and SP 650C adsorbent. A new batch screening process using the 384-well format was established, thereby utilizing fewer sample volume per isotherm and increasing throughput, compared to processes described in literature. By adapting the HTS design to the premise that small volumes can be more accurately quantified than handled, reproducible screening processes in 384-well format thereby become available for a variety of applications. Monte Carlo simulation of error propagation based on single process step accuracy has shown to be an adequate tool for error propagation analysis in HTS applications.
Future work will aim to extend the method of resin quantification in microplate wells to other resins, with focus on a variety of different backbones, as it has been shown, that they strongly influence the amount of light scattered. This investigation will be

aided by the use of established *in silico* Monte Carlo methods to investigate material property influence on data quality.

References

[1] J. Kittelmann, M. Ottens, and J. Hubbuch, *Robust high-throughput batch screening method in 384-well format with optical in-line resin quantification,* J. Chromatogr. B **988**, 98 (2015).

[2] K. Rege, N. Tugcu, and S. M. S. Cramer, *Predicting column performance in displacement chromatography from high throughput screening batch experiments,* Sep. Sci. Technol. **38**, 1499 (2003).

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

[4] B. K. B. Nfor, M. Noverraz, S. Chilamkurthi, P. D. E. M. Verhaert, L. a. M. van der Wielen, and M. Ottens, *High-throughput isotherm determination and thermodynamic modeling of protein adsorption on mixed mode adsorbents,* J. Chromatogr. A **1217**, 6829 (2010).

[5] J. Thiemann, J. Jankowski, J. Rykl, S. Kurzawski, T. Pohl, B. Wittmannliebold, and H. Schluter, *Principle and applications of the protein-purification-parameter screening system,* J. Chromatogr. A **1043**, 73 (2004).

[6] J. L. Coffman, J. F. Kramarczyk, and B. D. Kelley, *High-throughput screening of chromatographic separations: I. method development and column modeling,* Biotechnol. Bioeng. **100**, 605 (2008).

[7] B. D. Kelley, M. Switzer, P. Bastek, J. F. Kramarczyk, K. Molnar, J. L. Coffman, and T. Yu, *High-throughput screening of chromatographic separations: IV. Ion-exchange,* Biotechnol. Bioeng. **100**, 950 (2008).

[8] T. Herrmann, M. Schröder, and J. Hubbuch, *Generation of equally sized particle plaques using solid-liquid suspensions,* Biotechnol. Prog. **22**, 914 (2006).

[9] A. Susanto, K. Treier, E. Knieps-Grünhagen, E. von Lieres, and J. Hubbuch, *High throughput screening for the design and optimization of chromatographic processes: Automated optimization of chromatographic phase systems,* Chem. Eng. Technol. **32**, 140 (2009).

[10] T. Bergander, K. Nilsson-Välimaa, K. Oberg, and K. M. Lacki, *High-throughput process development: determination of dynamic binding capacity using microtiter filter plates filled with chromatography resin,* Biotechnol. Prog. **24**, 632 (2008).

[11] M. Wenger, P. DePhillips, C. Price, and D. Bracewell, *An automated microscale chromatographic purification of virus-like particles as a strategy for process development,* Biotechnol. Appl. Biochem. **47**, 131 (2007).

[12] S. Kreusch, S. Schwedler, B. Tautkus, G. A. Cumme, and A. Horn, *UV measurements in microplates suitable for high-throughput protein determination,* Anal. Biochem. **313**, 208 (2003).

[13] E. L. McGown and D. G. Hafeman, *Multichannel pipettor performance verified by measuring pathlength of reagent dispensed into a microplate,* Anal. Biochem. **258**, 155 (1998).

[14] U. Riebel, *An estimate of some statistical properties of extinction signals in dilute and concentrated suspensions of monosized spherical particles,* Part. Part. Syst. Char. **8**, 95 (1991).

[15] P. Bruscaglioni, G. Zaccanti, A. Ismaelli, and P. Pili, *Comparison between measured and calculated contributions of multiply scattered radiation to the transmittance of a light beam through a turbid medium,* Radio Science **22**, 899 (1987).

[16] G. Zaccanti and P. Bruscaglioni, *Deviation from the lambert-beer law in the transmittance of a light beam through diffusing media: Experimental results,* J. Mod. Opt. **35**, 229 (1988).

[17] M. Stone, *Cross-validatory choice and assessment of statistical predictions,* J. R. Stat. Soc. Ser. B **36**, 111 (1974).

[18] B. Efron, *Bootstrap methods: Another look at the jackknife,* Ann. Stat. **7** (1979).

[19] B. Efron and G. Gong, *A leisurely look at the Bootstrap, the Jackknife, and Cross-Validation,* Am. Stat. **37**, 36 (1983).

[20] J. Ogilviet, *A Monte-Carlo approach to error propagation,* Comput. Chem. **8**, 205 (1984).

[21] O. A. Güell, J. A. Holcombe, and O. A. Guell, *Analytical applications of Monte Carlo techniques,* Anal. Chem. **62**, 529A (1990).

[22] C. E. Efstathiou and T. P. Hadjiioannou, *Monte Carlo simulation for the study of error propagation in the double known addition method with ion-selective electrodes,* Anal. Chem. **54**, 1525 (1982).

[23] A. S. Kurup, H. J. Subramani, and M. T. Harris, *A Monte Carlo-based error propagation analysis of Simulated Moving Bed systems,* Sep. Purif. Technol. **62**, 582 (2008).

[24] A. Osberghaus, P. Baumann, S. Hepbildikler, S. Nath, M. Haindl, E. von Lieres, and J. Hubbuch, *Detection, quantification, and propagation of uncertainty in high-throughput experimentation by Monte Carlo methods,* Chem. Eng. Technol. **35**, 1456 (2012).

[25] R. Wehrens, H. Putter, and L. M. Buydens, *The bootstrap: a tutorial,* Chemometr. Intell. Lab. **54**, 35 (2000).

[26] A. R. Henderson, *The bootstrap: A technique for data-driven statistics. Using computer-intensive analyses to explore experimental data,* Clin. Chim. Acta **359**, 1 (2005).

[27] 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**, 11 (2008).

3

Optical characterization of agarose based chromatographic resins by UV-Vis spectra analysis

Optimization of chromatographic processes by high-throughput screening (HTS) methodologies has become a critical part of downstream process development. But there are still no known non-invasive optical methods to characterize resin and protein-resin interaction on liquid-handling platforms. The literature describes several approaches to automated resin screening in microplates, but all those methods involve indirect measurements by removal of, and sample quantification within, supernatant volumes. In this work, we introduce light extinction by scattering to directly access resin volume and density within microplates. Methods for resin characterization are described for 96 and 384-well microplates. A sample application demonstrates ligand concentration measurement in microplates with four commercial SP Sepharose™ Fast Flow batches, a well established resin by GE Healthcare. Further, direct quantification of biomolecules bound to adsorbent is shown by example with kinetic protein-resin interaction measurement in a batch screening process. This new approach is expected to promote batch-based resin characterization and monitoring on HTS platforms and further miniaturization and intensification of chromatographic HTS processes.

This chapter has been published in J. Chromatogr. A 1397, 52 (2015) [1].

3.1. Introduction

Chromatographic separation is the the key unit operation in downstream processing of high value biomolecules. With modes of separation being available based upon molecule features from selective affinity, as binding to Protein A, to charge distribution, utilized in ion-exchange chromatography. A high degree of diversification in resin base materials, spacer and ligand characteristics are available on the market, making material selection in early stage development a challenge. The resulting need in increased screening throughput is reflected in the development of multiple HTS techniques for fast characterization of chromatography based separation. Miniaturized and automated batch binding experiments provide information on affinity constants and capacities of product-resin interactions, as well as binding kinetics.

Hermann et al. developed a vacuum device to produce resin plaques of defined volumes [2] for use in high-throughput batch screening in 96 well format [3]. Commercially available microplates with pre-packed resin material have been developed and used for dynamic binding capacity determination [4]. Pipetting tips filled with resin have been developed for parallel use within LHS and shown to purify virus-like particles within a HTS application [5]. HTS chromatography experiments gain further importance with the development of computational models for large scale chromatography column runs, relying on protein-adsorbent interaction parameters and resin batch characteristics [6, 7].

The ongoing advancements have focused in accessing molecule-resin interaction, often by quantifying biomolecules not bound to adsorbent structures. Therefore multiple limitations in HTS based chromatography process development remain. First, further miniaturization of batch screening processes is limited by the amounts of resins which can be handled reproducible. As molecule volumes bound to resin are currently assessed indirectly in supernatant, high buffer to resin volume ratios introduce error to any protein-resin interaction measurement. Furthermore, supernatant removal disrupts the system, multiple measurements e.g. over time are therefore not possible within one probe, subsequently leading to increased sample consumption as multiple probes are needed.

Second, resin characteristics, as ionic capacity and pore size, can not be analyzed by HTS methods available. These properties are only taken into account by their combined influence on biomolecule separation during screening, thereby often neglecting batch-to-batch variances. With well defined ligand structures and functionalization processes, ionic capacity scales linear with ligand concentration in resin. The relevance of ligand density of chromatographic resins on molecule retention behavior was shown throughout chromatographic modes available. Wu et al. observed a change in elution order of Cytochrome C and Lysozyme on different cation-exchange resins with different ionic capacities [8]. Lower ligand density resins showed broader bands, with a negative impact on resolution of the two proteins. Fogle et al. investigated the effect of resins with different ionic capacities

and same backbone properties on the separation of mAb variants [9]. They concluded that the effect is not straight forward and that selectivity can improve, be not affected or decline with increasing ionic capacity.

Ligand density not only defines available binding sites but also influences accessibility, as porosity and corresponding pore diffusion are impacted [10]. Which is expressed in effects on dynamic binding capacity, as shown for ion-exchange [11] and affinity binding [12] chromatography. Established methods for determination of ligand concentration, or ionic capacity in IEX chromatography, are often column based and always time consuming. Although Failla et al. introduced an optical method based on absorption, ligand conversion to chromophores, often based on amino group baring ligands and amino reaction with picrylsulfonic acid, is necessary [13].

Although, spectrophotometer are well established devices in the HTS environment, non-invasive optical assays for HTS analysis of resin batches, are not yet described. Unspecific light attenuation due to scattering as it is caused by resin beads and described as source of error in photometric measurements [14], can also be used to determine resin characteristics.

Light extinction by suspended volumes of resin in microtiter plates can be described as particles in suspension, which is defined as

$$I(\lambda_0, l) = I_0(\lambda_0)e^{-\eta_1 l}e^{-\eta_2 l} = I_0(\lambda_0)e^{-\eta_1 l}e^{-(\eta_2^a+\eta_2^b)l} \tag{3.1}$$

with the initial intensity of the incident light beam $I_0(\lambda_0)$, the propagation distance l, and the extinction coefficients of the homogenous surrounding medium η_1, mainly due to absorption, and particles η_2. η_2 can include two components, η_2^a determined by absorption of the particles and η_2^b by scattering. All terms are functions of wavelength. If particles are non-absorbing in the wavelength range of consideration, $\eta_2^a = 0$ the expression for transmittance remains as

$$T(\lambda_0) = \frac{I(\lambda_0, l)}{I_0(\lambda_0)} = e^{-\eta_1(\lambda_0)l}e^{-\eta_2^s(\lambda_0)l} = T_1(\lambda_0)e^{-\eta_2^s(\lambda_0)l} \tag{3.2}$$

where $T_1(\lambda_0)$ is the transmittance of pure homogeneous medium without particles. Suspensions of particles scatter light and thereby increase the apparent optical density. The character of scattering is influenced by particle properties. Wavelength dependent scattering by spheres of any size, smaller and larger than the wavelength of the light beam, is described by Mie scattering and is proportional to $(1/\lambda)^\eta$ with η in the range of 0 to 4 [15]. Anyway, scattering by particles smaller than the wavelength of the light beam is often termed Rayleigh scattering, with $\eta = 4$ and refers to the Rayleigh limit of Mie scattering [16]. For particles much larger than the wavelength, light scattering is due to diffuse reflection which is not wavelength dependent ($\eta = 0$) and termed Tyndall scattering.

Light extinction due to scattering is described and used within various fields of research from analysis of water, atmosphere, to biological tissues [17]. Absorption and scattering by solutions of macromolecules and suspensions have been reviewed

in detail by Heirwegh et al. [18]. Light scattering with regard to multiple scattering within aggregate structures were investigated among others by Sorensen et al. [19] and Liou et al. [20]. Turbidity and scattering measurement methods were used to determine size and quantity of particles in suspension by [21–23].

In contrast to particles used in aforementioned studies, Sepharose beads are inhomogeneous structures. Agarose is formed from dextran strings which build a matrix skeleton, occupying only a small percentage of the bead volume. The dextran strings are ordered in a double helix structure with 15 Å in diameter [24], and are bundled in a side-by-side assembly from two to a few hundreds double helix structures each [25]. Those filaments are cross-linked by junction zones, involving non-covalent binding in ordered conformations. Agarose beads are further functionalized by ligands covalently bound to the backbone. Agarose beads feature a high variance in pore shape and size, up to 0.3 µm in diameter [25], structural differences from gel to gel and within locations of gels as measured by their intrinsic birefringence [26].

In this work we introduce a high-throughput compatible optical method based on light scattering characteristics for quantification and characterization of resin volumes in microtiter plate wells. A correlation between light attenuation and particle volumes is established in 96-well and 384-well microtiter plates. The influence of ligand densities in Sepharose based resins on light extinction is investigated and used for an exemplary ligand density measurement of different commercially available batches of SP Sepharose Fast Flow (FF) (GE Healthcare), highlighting the practical use of the presented assay in batch-to-batch variance analysis. Uptake of biomolecules to resin structures is measured and accessed by the method proposed. Resulting advancements in HTS method designs are illustrated with an automated kinetic uptake experiment of Lysozyme on SP Sepharose FF, utilizing a fraction of sample volume compared to established approaches.

3.2. Materials and methods

3.2.1. Resins

Sepharose based cation-exchange resins with varying ionic capacities were kindly provided by GE Healthcare (Uppsala, Sweden). Resins consisted of sulfopropyl (SP) ligands coupled to the same type of 6 % cross-linked agarose base matrix as used in commercially available SP Sepharose FF materials. Ionic capacities of resins used were provided by GE Healthcare and were between 25 and 250 µmol mL^{-1} as listed in Table 3.1. This range of ligand densities used extends far below the commercial specification of 190 to 250 µmol mL^{-1} for SP Sepharose FF. Resins used for batch-to-batch variance analysis were purchased from GE Healthcare.

3.2.2. Liquid handling station

Dilution series of resins in microtiter plates, protein sample preparation and photometric measurements were performed automatically on a Tecan Freedom EVO 200 (Tecan, Männedorf, Switzerland) liquid handling station (LHS). The workstation was equipped with an orbital shaker (Te-Shake, Tecan) and a Rotanta RSC 46 centrifuge (Hettich, Tuttlingen, Germany) set-up to accommodate microtiter plates. The LHS was equipped with an integrated microplate spectrophotometer infinite M200 Pro (Tecan) to measure UV/Vis-spectra as well as single wavelength of suspended resin volumes and protein samples. The LHS was controlled by the software EVOware 2.5 (Tecan).

3.2.3. Resin volume preparation

Resin plaques of defined volume of 7.8 and 20.8 µL respectively were produced with the ResiQuot vacuum device from Atoll-Bio (Weingarten, Germany) according to instructions from Herrmann et al. [2]. Plaques were transferred to 96-well microplates and subsequently diluted in 200 µL buffer to prevent complete drying and to provide slurry of defined resin concentrations.

3.2.4. Spectrophotometric measurements

Experiments including photometric measurements were conducted in 96-well UV microtiter plates (MTP), polystyrene flat bottom (Greiner Bio-One, Frickenhausen, Germany), with a volumetric capacity of 300 µL or in 384-square-well MTP, polystyrene flat bottom (Greiner Bio-One), with a volumetric capacity of 110 µL. Samples in microplates were centrifuged for 1 min at 2000 rpm prior to optical measurement to ensure even meniscus and full sedimentation of resin particles. Spectra measurements were performed with varying data point density throughout the range of 230 to 1000 nm, to compromise between data resolution and measurement time. In the range of 230 to 500 nm wavelength, measurements were taken with distinct points each 2 nm, between 500 and 1000 nm measurements were performed each 4 nm. The photometer provides a bandwidth of below 5 nm and 9 nm for wavelength below and above 315 nm.

Spectra fit was performed with two equations based on [17], describing the wavelength dependent light scattering of particles.

$$\mu_s(\lambda) = a\left(f_R\left(\frac{\lambda}{280(nm)}\right)^{-4} + (1-f_R)\left(\frac{\lambda}{280(nm)}\right)^{-\eta_M}\right) + b_T \tag{3.3}$$

Equation 3.3 describes the wavelength dependence of scattering in terms of the separate contributions by Rayleigh and Mie scattering at the reference wavelength. The Rayleigh scattering is $af_R\,(\lambda/280(nm))^{-4}$ and the Mie scattering $a(1-f_R)\,(\lambda/280(nm))^{-\eta_M}$. b_T denotes the wavelength independent Tyndall scattering. The wavelength λ is normalized by a reference wavelength, in this case 280 nm,

to yield a dimensionless value, which is then raised to a power η, the scattering power. The factor a is the value $a' = \mu_s(\lambda = 280(nm))$, which scales the wavelength dependent term.

$$\mu_s(\lambda) = a\left(\frac{\lambda}{280(nm)}\right)^{-\eta} + b_T \tag{3.4}$$

Equation 3.4 simplifies the wavelength dependent scattering to one term, yielding a generalized scattering power η.

Single wavelength measurements of light extinction due to light scattered by resin particles was performed at 330 nm wavelength. The measurement thereby gains a maximized signal to noise ratio for light extinction due to scattering events, which decreases with increasing wavelength and is not effected by wavelength specific absorption from proteins or peptides. Measurements of resin volumes in 384-well microplates, a grid of 5 by 5 evenly distributed measurement points, with a distance of 50 µm to the well wall, was defined in each single well. Measurements in 96-well microplates were performed with an circular layout consisting of 48 single measurement points covering the well area, with a distance to the well wall of 50 µm, as depicted in Figure 3.1. Measurement values were averaged to account for uneven distribution of sedimented resin beads in well. Experiments with adsorbent prototypes of varying ionic capacities and batch comparison experiments, were performed in 384-well microplates to account for sparsity of prototype materials. All other experiments were performed in 96-well microplates.

3.2.5. Kinetic batch uptake measurement

Kinetic uptake experiments were also performed in 96-well microtiter plates. Samples were pipetted in duplicate and at two distinct times for increased data density. Prior to sample pipetting, 150 µL resin slurry were pipetted into wells, resulting in resin volumes of 2.0 and 3.5 µL well $^{-1}$ which were quantified by measurements at at 330 nm wavelength as described in Section 3.2.4. Solute protein samples were added to working volume of 300 µL per well. In a circular workflow samples were incubated on an orbital shaker for 40 seconds, followed by a short centrifugal step for 30 s at 1000 rpm to settle suspended resin beads, and subsequently measured at 330 nm as described. The experiment was fully automated on the LHS and required no interaction during experimentation.

3.3. Results and discussion

3.3.1. Light extinction spectra

Sepharose based resin beads are filament network structures of low density primarily formed by dextran double helices. A single bead is therefore not considered as one particle in terms of light scattering and light extinction. Immanent light beams rather penetrate the bead structure and can be scattered at inner structures of agarose and ligands. The light collides with molecules as dextran string clusters, ligands, bound biomolecules, or clusters of those and is scattered dependent on the

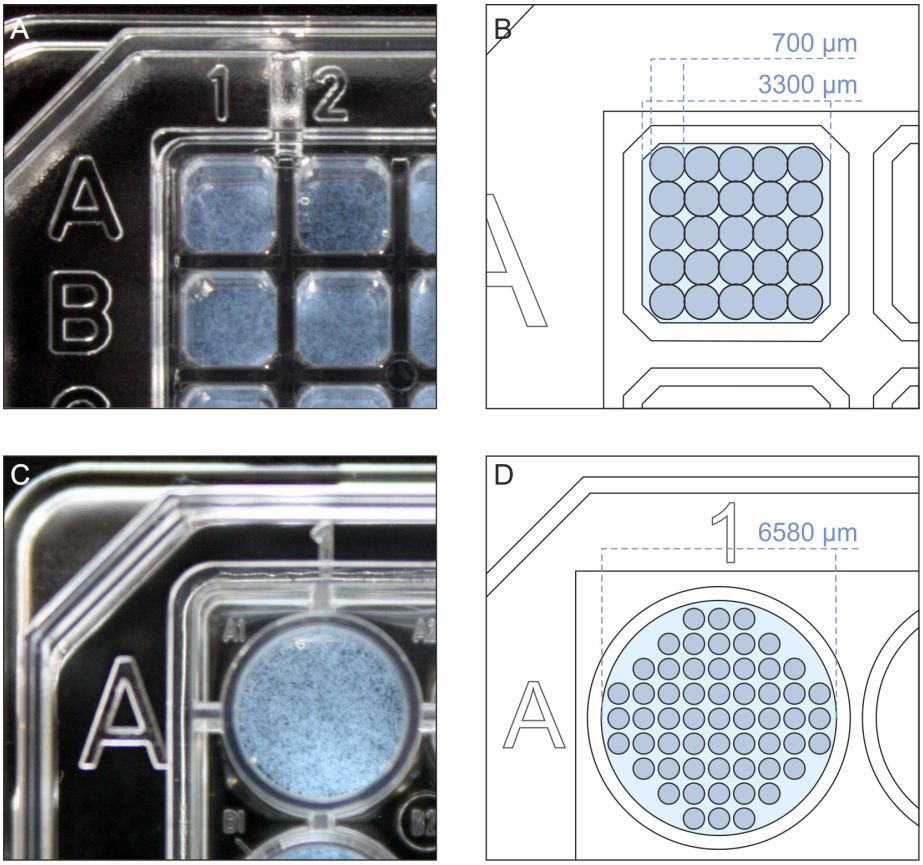

Figure 3.1: Pictures of wells with SP Sepharose FF resin (approx. 1.5 to 2.5 µL resin volume in 384-well microplate (A) and approx. 2.0 to 3.0 µL in 96-well microplate (C)). Schematic of a rectangular well in 384-well (B) and 96-well (D) microplate with measurement point (d = 700 µm) distribution in a 5 by 5 point grid (384-well) and circular 9 by 9 grid (96-well) layout, each with 50 µm spacing to well walls.

matter of collision. Given the molecular and structural diversity of resin materials, light attenuation is a result of scattering on particles of inhomogeneous structure and size. Furthermore the high density of scattering particles can result in multiple scattering, where a light beam, scattered on a particle, collides with further particles to be scattered again.

Figure 3.2A displays light extinction spectra in the UV/Vis-range of suspensions with different concentrations of SP Sepharose FF adsorbent beads. Spectra of adsorbents with different ionic capacities (Fig. 3.2B) and those of SP Sepharose FF with Lysozyme bound in varying concentrations from 10 to 65 g/L (Fig. 3.2C) are shown alongside for comparability. All spectra were measured in 96-well microplates as described in Section 3.2.4.

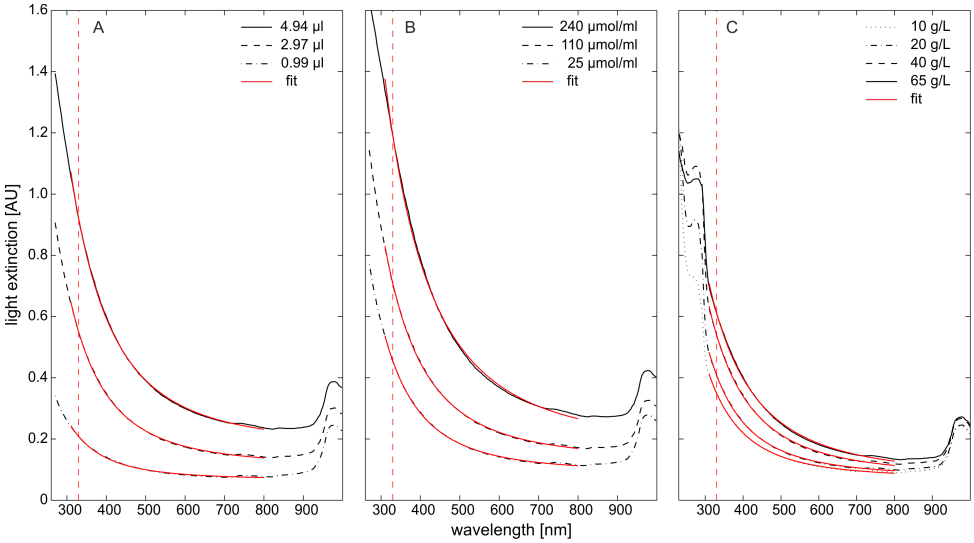

Figure 3.2: Light extinction spectra of samples of agarose based resin volumes, dispensed in microplates. Spectra were fitted with Equation 3.4 between 300 and 800 nm. Vertical line indicates single wavelength measurement at 330 nm as described in Section 3.2.4 A) Spectra of different volumes of SP Sepharose FF, suspended in 96-well microplate. B) Spectra of agarose based resins with varying ionic capacities. About 2.0 µL well $^{-1}$ resin were measured in 384-well microplate. C) SP Sepharose FF resin with varying concentrations of Ribonuclease A bound to the adsorbent. Spectra were measured in 96-well microplates.

Light extinction decreases with increasing wavelength, without wavelength specific absorption peaks in the range of 310 to 800 nm which could be related to resin material properties. Spectra in Figure 3.2C show wavelength specific absorption in the range from 230 to 300 nm with a peak maximum around 280 nm, which can be attributed to protein. In the range of 900 to 1000 nm wavelength light absorption by the water hydrogen bond is visible in all spectra. Experimental spectra in the range of 310 to 800 nm wavelength were fit to Equations 3.4 and 3.3, describing light extinction due to scattering. The quality of fit between both equations is comparable, therefore only the simplified Equation 3.4 was used for further characterization of spectral data. This reduces the wavelength dependent behavior of light scattering to one term, characterized by η, the power of light scattering wavelength dependency. The development of η for varying resin ionic capacities and protein load at resin concentrations from 50 to 150 µL well^{-1} are shown in Figure 3.3.
To shorten measurement time and to cover a potential uneven distribution of resin beads in the well cavity, a single wavelength measurement was taken at 330 nm in a grid layout as described in Section 3.2.4. The wavelength position within the spectra is also indicated in Figure 3.2.

Resin quantification The correlation between light extinction at 330 nm wavelength and resin volume in microtiter plate wells is depicted in Figure 3.4. In the range of 1.0 to 5.0 µL resin bead volume of SP Sepharose FF adsorbent a linear

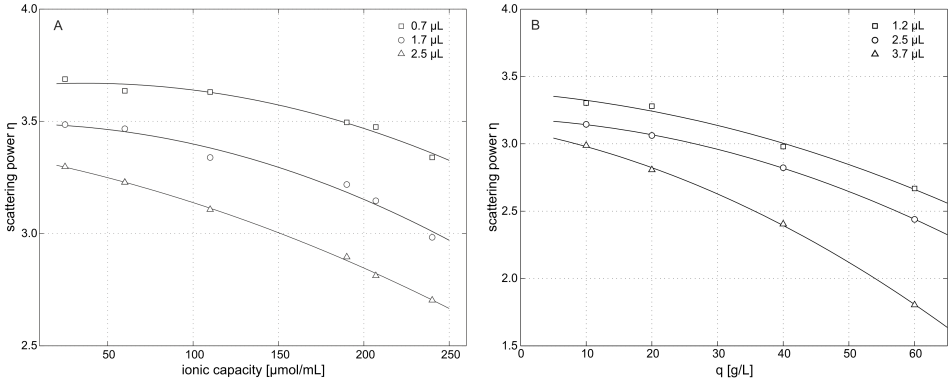

Figure 3.3: Scattering power as function of resin volume per well. A) Ionic capacity of Sepharose based resins with sulfopropyl ligands. B) Concentration of Ribonuclease A bound to SP Sepharose FF resin.

correlation could be established. This is in agreement with the assumption, that scattering media follow the Lambert-Beer law for light extinction in the limitation of low concentrations. Accuracy of calibration is limited by the manual distribution of resin by pipetting and can be further increased by a higher number of data points used. Multiple measurements of single wells show a relative standard deviation below 1.0 % (3.0 %) for concentrations above (below) 2.0 µL. The method shows to be more accurate in resin quantification in microtiter plates than described approaches relying on reproducible resin volume handling.

Ligand density The light extinction spectra of Sepharose based resins with different densities of sulfopropyl ligand are depicted in Figure 3.2B, showing an increase in scattering intensity with increasing ionic capacity. Linear correlations between resin volumes in wells and light extinction at 330 nm wavelength were established for single resins. Slope, axis intercept and quality of fit are listed in Table 3.1. Further a linear correlation between single resin calibration slope and ionic capacity was established and is shown in Figure 3.5. For comparison the range of ionic capacity given by the manufacturer for commercially available SP Sepharose FF batches is given as a gray area within the plot.

A high degree of light attenuation is caused by ligands bound within resin beads or scattering centers formed by them. Only a small fraction of scattering events can be attributed to the plain Sepharose backbone structure with a projected slope of 8.0 AU µL^{-1}, compared to 17.5 to 21.6 AU µL^{-1} for resins with ionic capacities in the range of production variance given by the manufacturer. With increasing ionic capacities, the apparent regime of scattering as determined in the factor η, changes from values of 3.5, to 3.0 for ionic capacity of 240 µmol mL^{-1}. This indicates, a change of predominant scattering centers from sizes smaller than the wavelength (η = 4.0), to comparable or larger than the wavelength within the Mie partition of scattering. The effect of Tyndall scattering increases in consonance, which is

reflected in a baseline shift to higher values throughout the spectrum.

Protein load Protein concentrations in dilution are often accessed by measurement of sample absorption between 230 to 290 nm, which can be assigned to the aromatic structures in the residues Tryptophan, Tyrosine and Phenylalanine. This approach is not possible for protein quantities bound to resin volumes, as light scattering by resin particles interferes with the absorption measurement. For that reason established methods rely on biomolecule quantification in sample supernatant and mass balance based calculation of concentrations bound to resin beads.

The introduced method of resin characterization by light scattering properties can be used for a direct quantification of protein load on Sepharose based adsorbents. Measurement of resin beads with protein, bound and unbound, are displayed in Figure 3.2C. They show a protein absorption signal superposed by light attenuation due to scattering from resin particles. While, the scattering signal throughout the spectra increases with increasing protein concentration in resin beads, the combined signal of scattering and absorption at 280 nm increases for samples up to 40 g L^{-1} protein along with the bound protein concentration, but decreases for 65 g L^{-1} protein loaded to the resin. This is due to tighter packing of protein molecules within resin beads, leading to a decrease in absorption, as single molecules are not accessible by the light beam. From Figure 3.3B it can be seen, that the generalized scattering power η decreases with increasing protein load concentration. This indicates, a decrease in scattering in the Rayleigh domain in favor of Mie scattering, caused by particles in the size of or larger as the wavelength. Tyndall scattering increases with protein load concentration but remains small, compared to the effect of volume and ligand concentration. The observations are in agreement with the assumption of molecule clusters of ligand, Sepharose backbone, and protein form scattering center of different sizes. With increasing density within resin particles, scattering center increase in number and average size, shifting the scattering pattern from a Rayleigh dominated one, towards Mie scattering. Further increases in particle density promote a decrease in the number of optical centers by affiliation, leading to wavelength independent scattering in the Tyndall domain.

The findings show, that protein load can be directly quantified within resin beads by changes in light scattering. It has to be assumed, that proteins of different size and shape lead to different packing within resin bead structures, influencing the scattering pattern yielded. A protein specific calibration, comparable to single wavelength quantification by absorption, is therefore necessary.

3.3.2. Application in process development

Batch comparison Ionic capacities of different SP Sepharose FF batches were determined using light scattering as described in Section 3.2.4 and the calibration from Sepharose batches of varying ligand concentrations shown in Table 3.1.

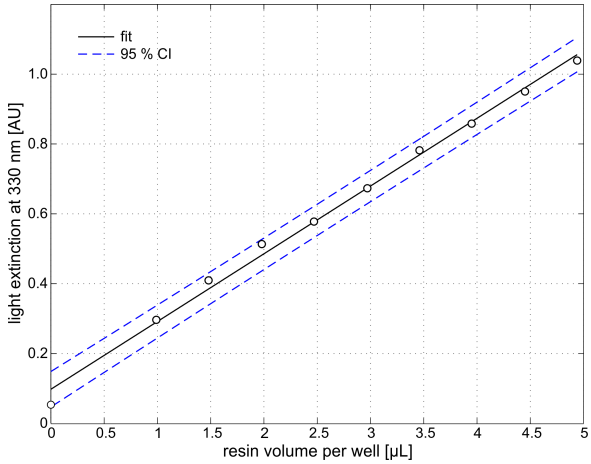

Figure 3.4: Light extinction at 330 nm wavelength as function of SP Sepharose FF resin bead volume per well in a 96-well microplate. 95 % confidence intervals (CI) for linear fit are shown.

Results are shown in Figure 3.6 and were compared against values which the manufacturer determined for each batch and provides to the customer. Ionic capacities for tested batches are in range of 210 to 250 µmol mL^{-1}, compared to 200 to 230 µmol mL^{-1}, as provided by the vendor. Confidence bounds were determined around +/- 10 µmol mL^{-1}, based on 8 measurement points per batch sample and the calibration on hand. Experimental results show high correlation towards the

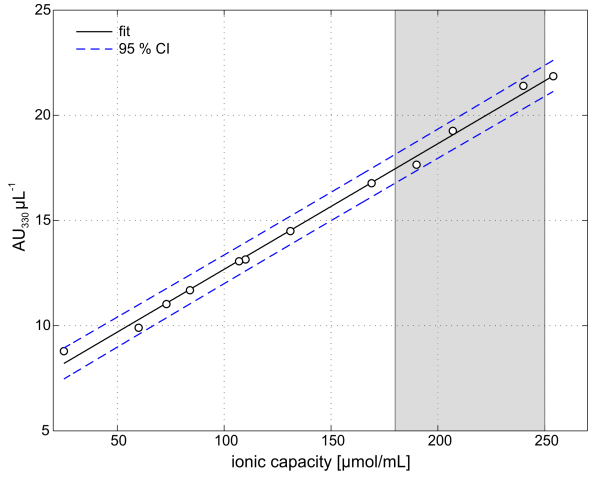

Figure 3.5: Correlation between slope of volume calibration and ionic capacity of resins with Sepharose backbone and sulfopropyl ligands. The gray area indicates potential range of ligand concentration of SP Sepharose FF resin as specified by product vendor. 95 % confidence intervals (CI) of linear fit are given, values are further listed in Table 3.1.

ionic capacity µmol mL^{-1}	slope AU_{330} µL^{-1}	stdev	intercept	stdev
25	8.79	± 0.242	3.762	± 0.730
60	9.90	± 0.206	3.262	± 0.592
73	11.02	± 0.439	3.804	± 1.160
84	11.68	± 0.137	3.532	± 0.894
107	13.06	± 0.216	3.451	± 1.694
110	13.15	± 0.530	3.396	± 1.638
131	14.50	± 0.299	3.776	± 1.566
169	16.77	± 0.491	3.998	± 1.041
190	17.68	± 0.785	4.043	± 1.897
207	19.45	± 0.430	3.844	± 1.315
240	21.86	± 0.600	4.235	± 1.897
254	22.09	± 0.426	4.036	± 1.896

Table 3.1: Slope of absorption unit over volume [µL] for Sepharose based prototype resins of varying sulfopropyl ligand concentrations.

provided values, with a consistent difference of about + 10 µmol mL^{-1}. The offset is assumed to be due to difference in method of ionic capacity measurements, between Sepharose of different ligand concentrations used in calibration and for industrially produced batches of resin material. Further, a variance in backbone structure, leading to differences in agarose content or dextran chain linkage, might influence results as the backbone, next to ligands bound, impacts the slope of light scattered over resin volume.

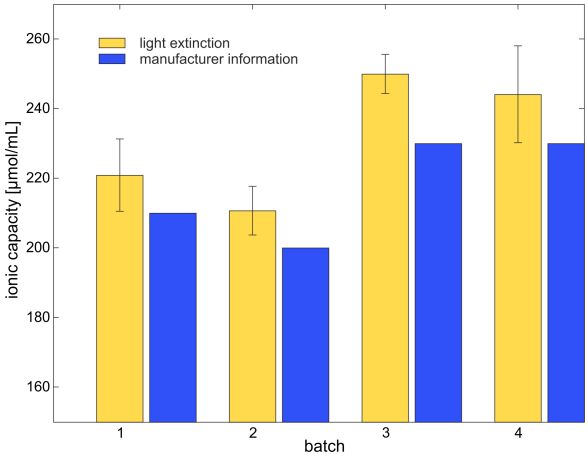

Figure 3.6: Ionic capacities of different SP Sepharose FF batches, as calculated from light extinction measurements at 330 nm wavelength in comparison to ionic capacity values provided by GE Healthcare for adsorber resin batches distributed. Error bars indicate 95 % confidence intervals.

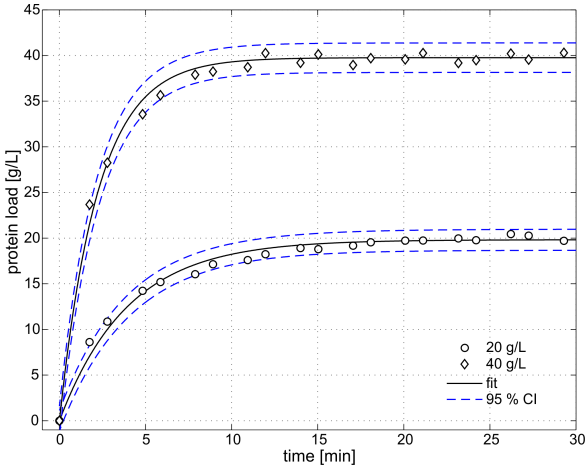

Figure 3.7: Batch uptake kinetic for Ribonuclease A on SP Sepharose FF at pH 7.0 with direct quantification of protein bound to resin by light extinction increase at 330 nm wavelength.

Kinetic uptake measurement The ability to utilize the change in light scattering characteristics due to protein bound on resin was tested in kinetic uptake measurements. Batch uptake kinetic for Ribonuclease A on SP Sepharose FF can be seen in Figure 3.7. The experimental set-up on a HTS robotic system as described in Section 3.2.5 is based on findings by Bensch et al. [27], with respect to shaking time and rate. To increase data density kinetic measurements are taken in 4 wells of a 96 well microtiter plate with equated starting times, realized by timing the initial pipetting of protein sample to resins. Average amount of resin was 1.5 µL well^{-1}, subsequently the amount of sample was about 60 µg well^{-1} for a final concentration of 40 mg mL^{-1} on resin.

Results are in good agreement with kinetic uptake measurements previously published, with regard to uptake rates on SP Sepharose FF [27, 28]. Further, data shows high quality with little variance between single data points resulting in narrow confidence intervals, confirming the high reproducibility of measurements. Sample consumption is significantly decreased, compared to established kinetic batch processes. Accurate quantification of resin volumes below 2.0 µL well^{-1} allow for less resin volumes and therefore less amount of sample to be used within single measurements. Furthermore, established processes are based on sample quantification within supernatant, and subsequent calculation of bound sample by mass balance. This, not only is prone to error, as extra pipetting steps are required to remove supernatant volumes from the well, it also disrupts the kinetic balance. Multiple measurements over time are therefore not possible within one probe, subsequently leading to increased sample consumption as multiple probes are needed. Light extinction measurements are performed on samples containing resin volumes and sample in bound and unbound conformation, rendering supernatant removal needless. Therefore, sample consumption scales with data density over time instead of total data points, as with established methods.

3.4. Conclusion and outlook

In this work an optical method for quantification and characterization of resin volumes based on light extinction measurements is introduced. For resins with sulfopropyl ligands bound on Sepharose backbones it is shown, that light beams penetrate resin beads and are scattered by inner particle structures and are sensitive to the characteristics and changes therein. Diverse scattering centers are formed by agarose backbones, ligands, and bound molecules, resulting in complex scattering patterns ranging from Rayleigh to Tyndall regime. This method is easily applicable in HTS workflow, relying on established microplate spectrophotometer, providing the potential of increased throughput and reduced sample consumption by further miniaturization of resin batch screening applications to 384 well microplates and less volumes to be used. Further, a batch specific characterization of adsorbent resins becomes available within the HTS context. This allows for fast resin characteristic screening and material lifetime monitoring. Next to ionic capacities, defined by ligand concentration, resin fouling and pore blocking are parameters which could be assessed. Above the mentioned, resin alteration by ligand leakage is a critical parameter in Protein A resin usage, which could be monitored by the method presented.

Acknowledgments

The authors thank Lonza Biologics, PLC, for financial support of the research presented within this study, GE Healthcare for providing prototype resin materials, and Katharina Lang for helpful input in preparation of the manuscript.

References

[1] J. Kittelmann, F. Hämmerling, M. Ebeler, and J. Hubbuch, *Light extinction and scattering by agarose based resin beads and applications in high-throughput screening,* J. Chromatogr. A **1397**, 52 (2015).

[2] T. Herrmann, M. Schröder, and J. Hubbuch, *Generation of equally sized particle plaques using solid-liquid suspensions,* Biotechnol. Prog. **22**, 914 (2006).

[3] A. Susanto, K. Treier, E. Knieps-Grünhagen, E. von Lieres, and J. Hubbuch, *High throughput screening for the design and optimization of chromatographic processes: Automated optimization of chromatographic phase systems,* Chem. Eng. Technol. **32**, 140 (2009).

[4] T. Bergander, K. Nilsson-Välimaa, K. Oberg, and K. M. Lacki, *High-throughput process development: determination of dynamic binding capacity using microtiter filter plates filled with chromatography resin,* Biotechnol. Prog. **24**, 632 (2008).

[5] M. Wenger, P. DePhillips, C. Price, and D. Bracewell, *An automated microscale chromatographic purification of virus-like particles as a strategy for process development,* Biotechnol. Appl. Biochem. **47**, 131 (2007).

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

[7] B. K. B. Nfor, M. Noverraz, S. Chilamkurthi, P. D. E. M. Verhaert, L. a. M. van der Wielen, and M. Ottens, *High-throughput isotherm determination and thermodynamic modeling of protein adsorption on mixed mode adsorbents,* J. Chromatogr. A **1217**, 6829 (2010).

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

[9] J. Fogle, N. Mohan, E. Cheung, 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 (2012).

[10] A. Franke, N. Forrer, A. Butté, C. Bozidar, 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**, 2216 (2010).

[11] 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**, 4366 (2009).

[12] W.-W. Zhao, Q.-H. Shi, and Y. Sun, *FYWHCLDE-based affinity chromatography of IgG: Effect of ligand density and purifications of human IgG and monoclonal antibody,* J. Chromatogr. A **1355**, 107 (2014).

[13] D. Failla and D. Santi, *A simple method for quantitating ligands covalently bound to agarose beads,* Anal. Biochem. **368**, 363 (1973).

[14] G. Geiger, J. Bernhagen, E. Wagner, H. Bisswanger, H. Brunner, and F. Vitzthum, *Standardized measurements and differential spectroscopy in microplates,* Anal. Biochem. **296**, 29 (2001).

[15] J. Q. Stewart, *Generalizations of the Rayleigh formula for molecular scattering,* J.O.S.A & R.S.I. **11**, 581 (1925).

[16] A. T. Young, *Rayleigh scattering,* Phys. Today **35**, 42 (1982).

[17] S. L. Jaques, *Optical properties of biological tissues: a review,* Phys. Med. Biol. **58**, 37 (2013).

[18] K. P. Heirwegh, J. A. Meuwissen, and R. Lontie, *Selective absorption and scattering of light by solutions of macromolecules and by particulate suspensions,* J. Biochem. Biophys. Methods **14**, 303 (1987).

[19] C. M. Sorensen, *Light scattering by fractal aggregates: A review,* Aerosol Sci. Technol. **35**, 648 (2001).

[20] K. Liou, Y. Takano, and P. Yang, *Light absorption and scattering by aggregates: Application to black carbon and snow grains,* J. Quant. Spectros. Radiat. Transfer **112**, 1581 (2011).

[21] G. Crawley, M. Cournil, and D. D. Benedetto, *Size analysis of fine particle suspensions by spectral turbidimetry: potential and limits,* Powder Technol. **91**, 197 (1997).

[22] G. S. He, H.-Y. Qin, and Q. Zheng, *Rayleigh, Mie, and Tyndall scatterings of polystyrene microspheres in water: Wavelength, size, and angle dependences,* J. Appl. Phys. **105**, 023110 (2009).

[23] S. M. Scholz, R. Vacassy, J. Dutta, H. Hofmann, and M. Akinc, *Mie scattering effects from monodispersed ZnS nanospheres,* J. Appl. Phys. **83**, 7860 (1998).

[24] S. Arnott, a. Fulmer, W. E. Scott, I. C. Dea, R. Moorhouse, and D. a. Rees, *The agarose double helix and its function in agarose gel structure,* J. Mol. Biol. **90**, 269 (1974).

[25] A. Amsterdam, Z. Er-El, and S. Shaltiel, *Ultrastructure of beaded agarose,* Arch. Biochem. Biophys. **171**, 673 (1975).

[26] J. Stellwagen and N. C. Stellwagen, *Internal structure of the agarose gel matrix,* J. Phys. Chem. **99**, 4247 (1995).

[27] M. Bensch, P. Schulze Wierling, E. von Lieres, and J. Hubbuch, *High Throughput Screening of Chromatographic Phases for Rapid Process Development,* Chem. Eng. Technol. **28**, 1274 (2005).

[28] J. L. Coffman, J. F. Kramarczyk, and B. D. Kelley, *High-throughput screening of chromatographic separations: I. method development and column modeling,* Biotechnol. Bioeng. **100**, 605 (2008).

4

Microfluidics on liquid handling stations (µF-on-LHS): an industry compatible chip interface between microfluidics and automated liquid handling stations

We describe a generic microfluidic interface design that allows the connection of microfluidic chips to established industrial liquid handling stations (LHS). A molding tool has been designed that allows fabrication of low-cost disposable polydimethylsiloxane (PDMS) chips with interfaces that provide convenient and reversible connection of the microfluidic chip to industrial LHS. The concept allows complete freedom of design for the microfluidic chip itself. In this setup all peripheral fluidic components (such as valves and pumps) usually required for microfluidic experiments are provided by the LHS. Experiments (including readout) can be carried out fully automated using the hardware and software provided by LHS manufacturer. Our approach uses a chip interface that is compatible with widely used and industrially established LHS

Parts of this chapter have been published in
Lab Chip, **13**, 2337–2343 (2013) [1] and in
Proc. SPIE 8976, Microfluidics, BioMEMS, and Medical Microsystems XII. [2]

which is a significant advancement towards near-industrial experimental design in microfluidics and will greatly facilitate the acceptance and translation of microfluidics technology in industry.

4.1. Introduction

Microfluidic devices rarely use industrially established standards [3]. Although microfluidics is considered to be a solution not only for the demands of home care diagnostics but also for industrial high throughput screening (HTS), there are still very few successfully commercialized microfluidic products in this market. Microfluidic chips are mostly designed "from scratch" with custom solutions for interfacing and fluid handling. This makes these chips very difficult to be used outside of the respective laboratory and even harder for industry to adapt. If standards were established for microfluidic chips scientific exchange as well as industrial acceptance could be dramatically increased. The issue of missing standards has been repeatedly discussed, most recently by Becker [4].

In addition to missing standards, there is also very little effort towards the implementation of automation standards [5]. If interfaces as well as peripheral components and automation were standardized, experiments could be translated from one lab to another by merely sending out digital script files. The use of standard laboratory centrifuges for operation of microfluidic chips has been suggested as one potential means to do so [6]. Centrifuges are widely available therefore allowing convenient translation of laboratory assays to an automated platform. However, liquid handling on a centrifugal format has certain limitations and fluid transfer in and out of the chips still needs to be carried out manually. In addition, readout is usually restricted to optical means due to the fact that the rotating microfluidic system is hard to interface, e.g., electronically.

We propose combining widely established industrial liquid handling stations (LHS) with microfluidics. We term this combination of technologies microfluidics on LHS (µF-on-LHS). LHS offer a high degree of freedom with respect to fluid handling and extraction. Highly developed software modules and hardware components allow for fully automated fluid handling using these systems. Using LHS, an experiment can, once a microfluidic layout is fixed, be run fully automated and, if the microfluidic chip layout is shared, be repeated in another laboratory if the same LHS is available. LHS are standard equipment in many laboratories around the globe as well as in industry. They were originally developed for fully automated handling of liquids in microwell plate (MWP) format. They generally offer features such as multichannel syringe pumps, integrated multiport valves (usually for each of the pumps individually) and numerous additional components such as detectors for readout, interfaces for MWP racks, vibrating tables, thermocyclers, etc. Although components of LHS (such as the automated syringes) have been used in combination with microfluidics [7], LHS are in general considered as being support instruments for microfluidics suitable for contactless dispensing of fluids into reservoirs [8]. Numerous examples of such systems can be found in the literature including systems for mouse genotyping using electrophoresis [9], passively pumped channels for high density cell culture [10], hydrogel microwell cavities for cell experiments [11], high-throughput drug screening 10 or parallelized immunoassays [12]. Some of these systems are adapted to the MWP format in order to make full usage of the LHS parallelization

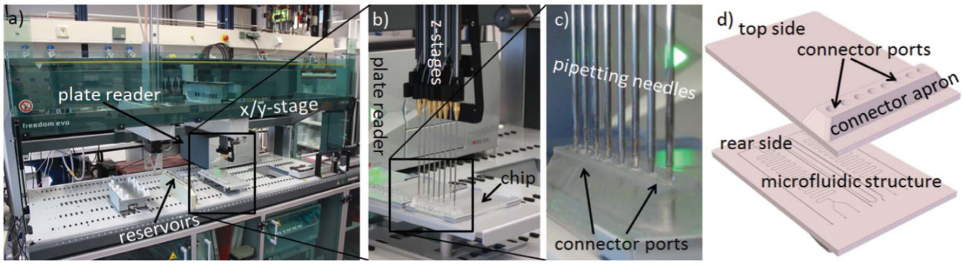

Figure 4.1: Microfluidics on a liquid handling station (µF-on-LHS) a) Photograph of a LHS type Tecan Freedom EVO 200 with chip inserted. The pipetting needles are moved by the x/y-stage of the LHS. Reservoirs provide liquid samples to be injected into the chip. b) Detailed view of the microfluidic chip on a tray of the LHS. Z-stages allow for individual insertion of the pipetting needles into the chip. The plate reader is visible in the background. c) View of pipetting needles placed in the connector ports of the microfluidic chip. Sealing is achieved by taking advantage of the soft material characteristics of polydimethylsiloxane (PDMS): the needles are slightly thicker than the channels, thus squeezing tight. d) View of a computer aided design (CAD) model of the microfluidic chip as created with the newly developed molding tool (see Fig. 4.3). The connector apron stabilizes the connector ports in which the pipetting needles of the LHS are inserted to connect with the microfluidic structure of the chip at its rear side.

capability [13, 14]. However, it has been repeatedly stated, that spotting of liquids into or onto microfluidic structures per se is not enough and that effective means for retrieving reaction products [15] or cells [16] would be required as well. The most challenging aspect of such a closed integration of LHS with microfluidics is the creation of a multi-channel reversibly sealing connection between pipetting needles (which necessarily form the interface to the LHS) and microfluidic chips [17]. Although such an integration of microfluidics with LHS has been stated as being of immense interest [18] it has, to the best of our knowledge, not been implemented in a generic way yet.

In this paper we describe a generally applicable microfluidic chip format that allows the interfacing of a (freely designable) microfluidic structure to standard LHS. This µF-on-LHS allows the creation of microfluidic assays that can be operated using standard industrial equipment thus making use of a worldwide accepted standard in industry and academia. We demonstrate the feasibility of this approach with a generic chip design, which is applicable in a wide range of chemical and biomolecular screening experiments. The design, composed of the proposed µF-on-LHS interface, a reaction channel, a subsequent T-connection for sample manipulation and a measurement channel structure for readout is described. Sample quantification is shown for a protein solution to exemplify an automated readout for biomolecular applications.

4.2. Experimental

A liquid handling station type Freedom EVO 200 (purchased from Tecan, Switzerland), controlled by EVOware Standard software package V 2.4 (provided with the

machine) was used for this work (Fig. 4.1a). It provides eight individually addressable pipetting needles connected to syringe pumps via flexible tubing. The pumps have a resolution of 3000 steps per stroke and can be equipped with syringes of 25 to 5000 µL volume. In this work we used 100 µL syringes, correlating to a volume of 0.034 µL per step and a minimum speed of 0.16 µL s^{-1}. The syringes are connected to pipetting needles which are mounted on a robotic arm, allowing free movement along the x/y-axes. The needles can be spaced flexibly in order to comply with different labware designs. In this work we use a spacing of 9 mm for the microfluidic chip, which is in agreement with the spacing of a 96 well MWP. The needles can be moved up- and downwards individually in order to enter microwell plates thus allowing fluids to be probed to and extracted from the microwells using the respective syringe provided by the LHS. We used the fact that each needle can be moved individually connect the pipetting needles to a microfluidic chip reversibly.

The microchip can be automatically transferred between functional units on the LHS such as, e.g., an integrated spectrophotometer for read out. A robotic manipulator arm (RoMa) is used to move the chip (and similar components) inside the LHS. In this work, we used a UV-VIS plate reader of type Infinite 200 M (purchased from Tecan, Switzerland) for assessing sample concentrations in microfluidic channels by absorption measurement. The plate reader is integrated in the LHS and controlled by Magellan software V 7.1 (provided by Tecan). The measurement spot size of the photometer has a diameter of 700 mm. Absorption signal processing was done with MATLAB V 8.0 (The Mathworks, Natick, USA).

4.2.1. Microfluidic chip design

For reversible connection of the pipetting needles of the LHS to the microfluidic chip connector ports matching the positioning of the needles were designed (Fig. 4.1c). The needles are guided into position via self-centering cone shapes. The material used for the chip is polydimethylsiloxane (PDMS, two component curing system type Elastosil 601, purchased from Wacker, Germany) as it is widely used and convenient for replication molding. Due to its soft properties, we designed the connectors such that the pipetting needles (once lowered) are self-sealing. These connector ports can be conveniently replicated by using a custom-design molding tool. This tool is designed as a composite tool that allows insertion of a classically structured microfluidic replication master (created, e.g., by means of lithography) that defines the shape of the microfluidic channels. This micro structured replication master can easily be exchanged if the fluidic design needs modifications while keeping the structure of the connector ports.

4.2.2. Molding tool

The developed molding tool is shown in Fig. 4.3. It is made of parts machined in stainless steel and eight uniformly shaped screws which are lathed from stainless steel studs (iron-base alloy type X5CrNi18-10, purchased from Eisenschmitt, Germany). Computer aided design (CAD) models of all parts are available in the Electronic Supplementary Information to this publication [1]. These models will

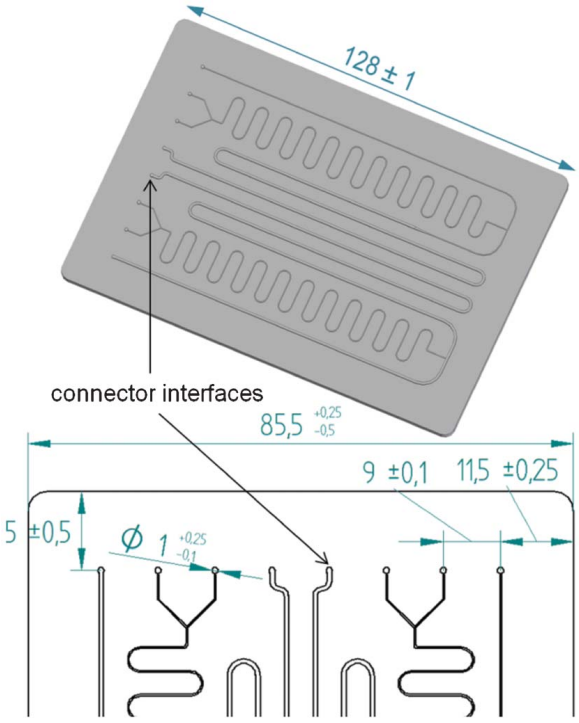

Figure 4.2: Relevant dimensions of the microfluidic replication master used in this work. This is an example of a microfluidic structure suitable for µF-on-LHS. The only important features are given with dimensions; the thickness of the bottom layer must not exceed 3 mm. All other structures (channels, intersections, etc.) can be freely adapted to the application in mind. Size and location of the connector interfaces have to match the size and pitch of the pipetting needles of the LHS. All dimensions are displayed in millimetre.

allow redoing and modifying the µF-on-LHS design described in this work. The molding tool defines the outer shape of the microfluidic chip particularly forming the connector ports of the chip. For easy dismantling the molding tool is designed of three individual parts that are screwed together. The bottom part is merely a tray to hold the replication master for the microfluidic layout. It also features a wide opening on one side serving as pour hole. Through this opening, PDMS prepolymer is poured into the closed form during replication. Onto the bottom part, the so called spacer is placed. It defines the contour of the connector apron which will be, once replicated, in the form of a heightening on top of the chip which provides access through connector ports for the pipetting needles.

The individual connector ports are created in the form of holes with a slope in order to ensure sufficient degree of self-centering when the pipetting needles are inserted. Simply providing orifices in the microfluidic chip is not sufficient: As the needles are not perfectly aligned buckling and potential breakage could occur. The

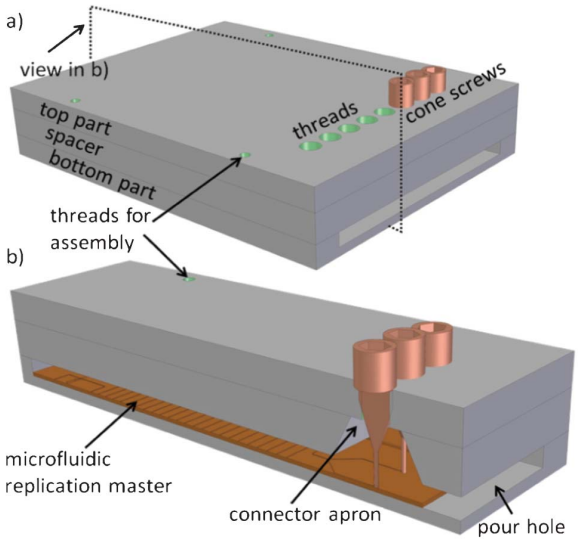

Figure 4.3: Computer aided design (CAD) view of the molding tool used to create chips as shown in Fig. 4.1c). a) Assembled molding tool with three cone screws inserted and five threads left for additional cone screws or to be sealed with stud screws. The four smaller threads are used to screw the three parts of the molding tool together. b) Cut view of a) displaying the inside of the molding tool. The bottom part serves as a tray for the replication master and has an opening through which polydimethylsiloxane (PDMS) prepolymer can be poured. The spacer shapes the connector apron. The replication master is kept in place by the cone screws which also define the cone-shaped form of the individual inlets. Once the PDMS is cured and the chip is placed on the LHS, the pipetting needles can be connected to the microfluidic network by these connector ports. Digital 3D models of these components can be downloaded from the Electronic Supplementary Information (ESI).

top part of the molding tool holds up to eight specifically designed cone screws. They are located such that they will form the described cone-shaped connector port openings during replication. These screws feature an 8 mm long cylindrical tip with a diameter of 1 mm (thus being slightly thinner than the pipetting needles of the LHS) followed by a 10 mm long cone spreading the diameter to an 8 mm thread (M8). The M8 thread on the end of the screws is 12 mm long with a hexagon socket on its end. If less than eight connector ports are required, the remaining opening threads in the top plate can be closed by stud screws (DIN 913 M8 x 8).

4.2.3. Microfluidic replication master

There is extensive work in the scientific literature describing replication masters for casting of PDMS microfluidic chips. These can be generated by, e.g., soft lithography [19] as well as numerous alternative techniques [20]. For µF-on-LHS, the only requirements in design are that the chip's footprint does not exceed the dimensions of a microwell plate (128 x 86 mm) by width and that the fluidic inlets are at locations defined by the microwell plate standard (ANSI/SBS 1-2004 to ANSI/SBS 4-2004). In this work, we used the format of a 96 microwell plate which allows a maximum of eight intake interfaces of a diameter of 1 mm being placed in a line

spaced from each other by a distance of 9 mm correlating to the minimum spacing of the LHS employed. Fig. 4.2 shows a detailed view of its structure displaying all dimensions relevant for shaping proper interconnection of the microfluidic chip to the LHS. The microfluidic replication master used in this work was created via stereo lithography. As rapid prototyping methods allow fast design changes, evaluation of different chip layouts can be carried out quickly. If round shaped silicon wafer masters are to be used instead, the sides of the wafer might have to be cut to fit the dimensions of the molding tool. If smaller wafers are used, double sided adhesive tape applied on the back of the structure ensures proper casting without unintentional moving of the replication master. Attention has to be put on the alignment of the connector interfaces and the cone screws when assembling, as silicon templates may break if the pressure applied via the cone screws is too high. A detailed description of the exemplary chip design displayed throughout this work is given in the following sections.

4.2.4. Microfluidic chip manufacturing

To create a µF-on-LHS compatible chip the microfluidic replication master is placed into the molding tool by sliding it through the pour hole. Subsequently, the required number of cone screws is screwed in until their tips connect to the connector interfaces (1 mm in height) of the replication master. This can be controlled visibly through the pour hole (see Fig. 4.4a). By this joint connection a channel through the chip will be formed during replication that connects the respective pipetting needle to the microfluidic channels formed by the replication master. PDMS is prepared by mixing its two components in a 9 : 1 mass ratio. After mixing the prepolymer can be processed for approximately one hour before the degree of crosslinking prohibits any further processing. The mixture is then poured into the molding tool. To speed up curing while also removing air trapped in the prepolymer the filled molding tool is placed in a vacuum oven (type Vacutherm VT 650, purchased from Heraeus, Germany, vacuum pressure 100 Torr) for one hour at 70 °C. Through the oven's window, the pour hole can be observed. As the pressure decreases, bubbles striving towards the surface of the PDMS can be seen. The pressure is periodically released to allow the bubbles to burst. This procedure is repeated until no more bubbles appear. After curing the molding tool is disassembled for dismantling the chip.

The microfluidic channel network can then be sealed by bonding against a planar substrate [21]. This substrate may, depending on the application, also be provided with, e.g., electrodes or similar structures for sensing or synthesis applications. At 450 kPa, the chip starts leaking at the connector port. If higher pressure is needed, the cone screws should be made thinner in order to generate a tighter interface to the pipetting needles.

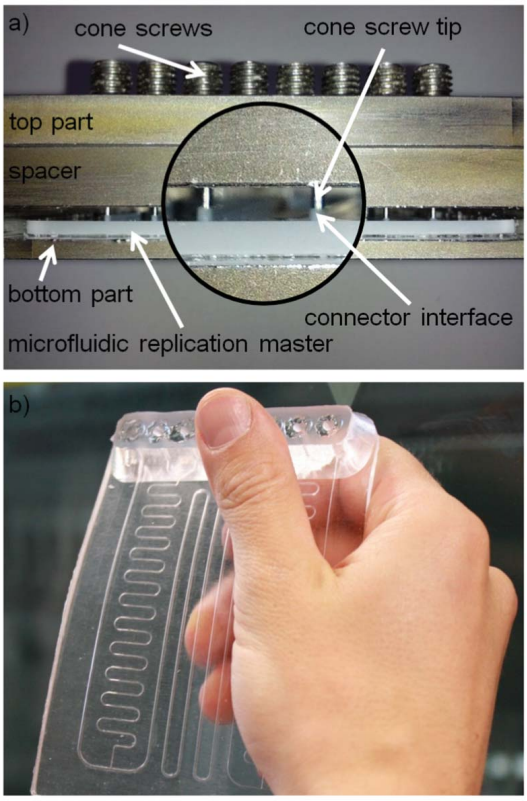

Figure 4.4: a) Photograph of the molding tool with microfluidic replication master inserted. The image shows the components when looking through the pour hole. The cone screws must form physical contact with the microfluidic replication master at the connector interfaces which can be controlled visibly (see enlarged part of the image). During replication, a microfluidic channel is created in the chip allowing interconnection of the channels with the pipetting needles of the LHS. b) Photograph of the created PDMS microfluidic chip. After bonding the chip is ready to be placed into the LHS.

4.2.5. Validation of µF-on-LHS: segmented flows

In this work, the µF-on-LHS concept was extended to allow generation of droplets. Droplet microfluidics has proven to be a suitable platform for HTS, cell-based assays as well as for analytics. Here we show that droplets can also be created using the µF-on-LHS concept and thus can be extended to industry-compatible instrumentation. To the best of our knowledge, droplet microfluidics has not been demonstrated on an industrial platform as of yet and this is the first report of doing so.

The microfluidic chip employed can be seen in Fig. 4.5. The microfluidic structure is a two-fold T-junction setup that created droplets by injection of the aqueous phase (dyed water, shown in blue) into the inert phase. In this exemplary application, we used air as an inert phase, which can be conveniently probed using the syringes supplied with the LHS. In the literature, several inert phases are being

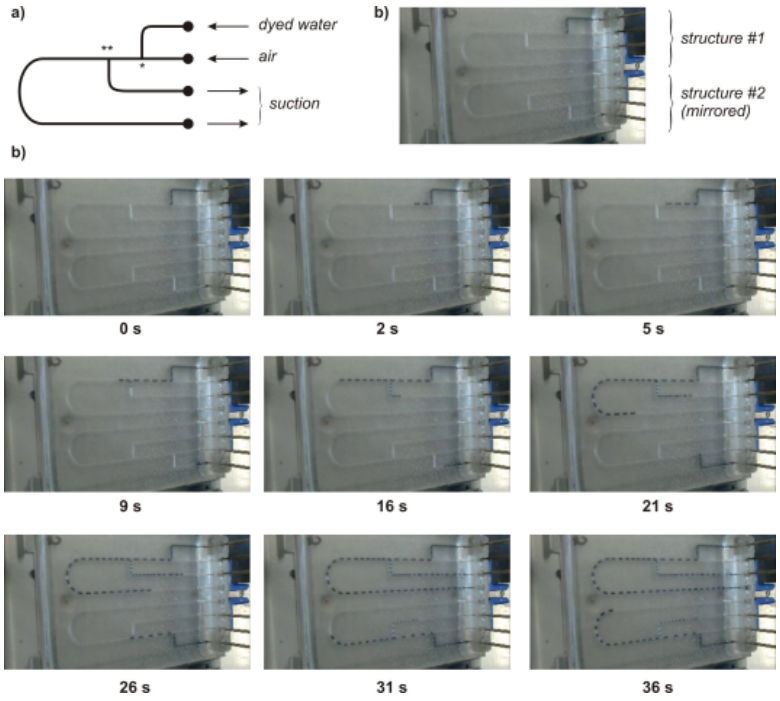

Figure 4.5: Using µF-on-LHS for droplet generation on chip. A dyed aqueous solution was used as the exemplary aqueous phase whereas air was used as an exemplary inert phase. Naturally, commonly used inert phases such as, e.g., fluorinated oils or aliphatic hydrocarbons are suitable alternatives. a) The chip features a two-stage droplet generation structure which is mirrored. Dyed water is pumped into the first injection channel where droplets are created (*) in the inert phase (air). A second channel splits the droplets again (**).
b) View of the µF-on-LHS compatible chip with the two mirrored structures and a time-lapse series of the droplet generation process. The two structures were probed with a short delay (about 20 s). At 2 s in the experiment, the upper structure already creates droplets at the first intersection. At 9 s in the experiment the second T-junction is reached which splits the droplets again by means of suction. The same process can be observed (time-delayed) in the mirrored structure.

favored, among others, fluorinated oils, such as FC-40, and aliphatic hydrocarbons, such as tetradecane. Naturally, these substances could be used as well.

The structures are designed such, that the aqueous phase is continuously injected into the inert phase at varying pump rates. These can be set conveniently on the LHS via the user control interface. The aqueous phase is thus split into droplets at the first T-junction (in Fig. 4.5, this is marked with an asterix). Given the fact that the LHS cannot only provide fluid flow due to overpressure but also fluid flow by suction, one interesting concept can be exploited. The aqueous fluid plugs created are passed along a second T-junction (in Fig. 4.5, this is marked with a double-asterix) where suction is applied via the syringes of the LHS. This suction suffices to split the droplets again by extracting portions of the liquid. In consequence,

there is a second splitting (subsplitting) and finer droplets are created.

4.2.6. Validation of µF-on-LHS: concentration determination using a MWP format UV-VIS spectrometer

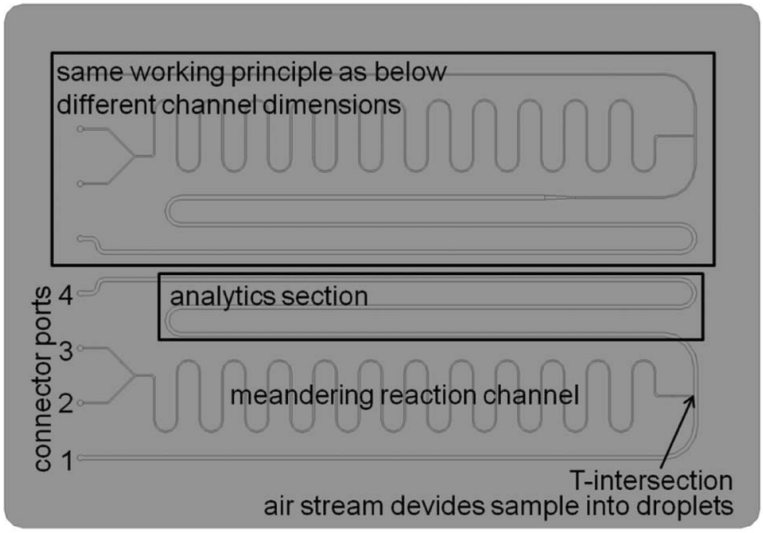

Figure 4.6: View of a computer aided design (CAD) model of the microfluidic replication master of the chip used for segmented flow analysis on the liquid handling station (LHS). Lysozyme from connector port 3 and buffer from connector port 2 are mixed diffusively in the meandering reaction channel. Droplets are formed at the T-intersection by an air stream coming from connector port 1 dividing the sample flow and pushing it towards the analytics section of the chip. Once the droplets are in the analytics section, the flows are stopped and the microfluidic chip is moved into the plate reader incorporated in the LHS for automated photometric readout. Connector port 4 is used as catch basin; also, after analysis in the plate reader individual droplets can be accessed by a pipetting needle through this port for further examination.

As a second exemplary validation of the µF-on-LHS a microfluidic chip for sample reaction, fractionation and segmented flow analysis was designed (see Fig. 4.6). The microfluidic structure used for this experiment is a channel with a set of three inlet and one outlet channels. Two independent variants of this microfluidic structure (varying in channel dimensions) were incorporated on one chip. As the LHS can operate eight pipetting needles simultaneously, this allows for two separate experiments to be carried out at the same time. This parallel processing facilitates HTS on microfluidic chips for distinct experimental tasks. Each of these two microfluidic structures features a meandering reaction channel (250 x 250 mm in cross section) in which two reactive components provided from their respective inlets channels are mixed. The channel's meandering shape allows for prolonged reaction times and enhanced mixing, which may be further augmented using more sophisticated mixing structures such as, e.g., Tesla mixers [22].

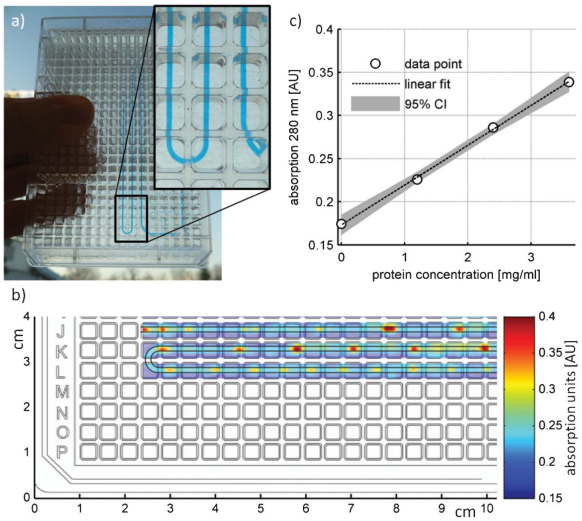

Figure 4.7: a) Photograph of a standard 384 well microwell plate (MWP) and the microfluidic chip used in this work held congruently to show microfluidic channels (for display purposes partly filled with blue liquid) match the grid of the wells of the MWP. b) False-colour representation of the absorbance (which correlates to the concentration of the protein, see c) of an aqueous lysozyme solution of 1.4 µL (row J and K) and 0.7 µL (row L) volume measured on the plate reader spectrophotometer at 280 nm wavelength. The microfluidic droplets show up as spots of high absorbance. The plate layout is displayed for representation of measurement scheme. The microfluidic channels have been highlighted with thin black lines. c) Calibration curve of lysozyme measured at 280 nm in a microfluidic channel of 700 µm width and 200 µm height, with 95 % confidence interval (CI). N = 6, error bars are smaller than data point representation and therefore not displayed.

After the reaction meander, the liquid stream is separated into droplets at a T-intersection using an air stream (liquid in air droplets). This effectively stops sample diffusion and allows for a time independent measurement. Instead of air, any fluid immiscible with aqueous solutions (such as, e.g., the fluorinated oil FC-40) could also be used for this purpose. The air stream separates the sample stream into droplets. Different volumes in the range from 0.7 µL to 1.4 µL are created by pausing the air stream. Besides pausing the air stream for intervals corresponding to the desired volume, droplets can also be induced by altering the pipetting speed, hence adjusting the flow-rate of the air stream. Thus, droplets of nearly any volume can be created. These droplets are moved into the analytics section of the microfluidic chip where they are kept in place by stopping the respective pipette. Once the droplets are placed in the analytics section of the microfluidic chip, the microfluidic chip is moved to the plate reader integrated in the LHS by the RoMa for offline measurement.

The spectrophotometer software is limited to MWP designs which are not directly applicable to microfluidic chips, as channels extend over multiple wells. To overcome this limitation, the analytics section of the microfluidic chip has been positioned in alignment with well positions in a 384-well MWP (see Fig. 4.7a). The

microfluidic channel of the analytics section features if 700 µm wide and 500 µm in height. A predefined layout for MWP with 384 squared wells was chosen for measurement in the software of the plate reader. Because the bottom of a single well in this particular MWP extends over 3.3 x 3.3 mm (which is covered by a rectangular grid of 8 x 8 measurement points) parts of the channel would not be visible to the reader. To also cover the remaining spots in the channel, a second grid layout with an offset of 2.5 mm in the x-direction to the aforementioned one was measured to cover interstitial channel volumes between single wells. Multiple measurements at overlapping area were averaged. The resulting readout of a solution of lysozyme from hen egg white (purchased from Sigma Aldrich, USA) in 0.02 M sodium phosphate buffer at pH 7 is displayed in Fig. 4.7b. The calibration curve of lysozyme dissolved in the same phosphate buffer at different concentrations for protein quantification in chip is shown in Fig. 4.7c.

For microfluidic channels with channel widths equal or smaller than the photometric measurement spot, the signal strength can vary as the beam extends beyond the microfluidic channel and light bypasses the fluidic channel (and thus the sample) resulting in higher light intensities. Analyte concentrations are hereby underestimated. This can be corrected for aqueous samples by referencing the light absorption of the analyte at a given wavelength (e.g. 280 nm for proteins) to the absorption of the solvent, determining the apparent absorption path length as described by McGown and Hafeman [23].

After screening, single droplets may be individually extracted by the pipetting needle connected to the outlet connector port of the microfluidic structure thereby recovering the liquid for further processing. Thus, interesting fractions of the sample can be selected based on photometric analysis to limit further analytical efforts. Especially time consuming analysis like, e.g., chromatography or mass spectroscopy are only applied on selected sample droplets. Of course, further analytics could be carried out fully automated on the LHS, too. Compared to experiments in MWP, µF-on-LHS is advantageous because it requires less sample volume. Additionally, by forming droplets in microfluidic channels sample quantification can be carried out in small volumes leading to accurate measurements by defined light path lengths of the photometer, which is prevented in MWP due to meniscus formation at the sample to air phase boundary.

Furthermore, in MWP high concentrations need dilution prior to measurement in the reader in order to fit in the linear range of the photometer. Lysozyme for example needs to be provided at a < 1.5 g L^{-1} dilution in a 96 well MWP, higher concentrations lead to signal saturation. As short path lengths can be achieved in microfluidic chips by defining low profile channels, no intermediate dilution of sample is necessary. Furthermore, mixing of samples can be supported by suitable microfluidic structures.

Different complex experimental set-ups can be fully automated in a short time

frame by utilization of several microfluidic chips providing standardized sub-functionality at microliter scale. Compared to classical microfluidics, µF-on-LHS offers the advantage of automation and equipment standards provided with the LHS. The use of the LHS's infrastructure leads to decreased chip complexity, as merely passive microfluidic structures are required. Active components, such as valves and pumps provided by the LHS, are not required to be integrated microfluidic components.

4.3. Conclusions

µF-on-LHS allows fast, robust and simple chip manufacturing using the developed molding tool. Our design allows almost any type of microfluidic structure to be created in a way that any LHS can be used to operate the respective microfluidic assay. Using this strategy allows the creation of microfluidic assays that are easy to migrate to a different laboratory as only passive structures and software code (in form of protocols for the respective LHS) need to be transferred. Furthermore, µF-on-LHS allows the creation of microfluidic chips relying on industry proven and well developed LHS components such as valves, pumps, detectors, etc. Development of proprietary solutions for these tasks therefore becomes unnecessary which will increase accessibility both for industry and academia.

µF-on-LHS enables microfluidic applications to be effectively deployed as routine techniques for biotechnological process development. We believe that this strategy, in combination with the molding tool developed in this work (which can be downloaded as 3D CAD compatible models from the Electronic Supplementary Information) may be a critical step in providing the microfluidics community with an industry compatible interface for HTS experiments.

References

[1] A. Waldbaur, J. Kittelmann, C. P. Radtke, J. Hubbuch, and B. E. Rapp, *Microfluidics on liquid handling stations (f-on-lhs): an industry compatible chip interface between microfluidics and automated liquid handling stations,* Lab Chip **13**, 2337 (2013).

[2] J. Kittelmann, C. P. Radtke, A. Waldbaur, C. Neumann, J. Hubbuch, and B. E. Rapp, *Microfluidics on liquid handling stations (µf-on-lhs): a new industry-compatible microfluidic platform,* Proc. SPIE **8976**, 89760G (2014).

[3] G. M. Whitesides, *The origins and the future of microfluidics,* Nature **442**, 368 (2006).

[4] H. Becker, *Mind the gap!* Lab Chip **10**, 271 (2010).

[5] J. El-Ali, P. K. Sorger, and K. F. Jensen, *Cells on chips,* Nature **442**, 403 (2006).

[6] D. Mark, M. Focke, S. Lutz, J. Burger, M. Müller, L. Riegger, M. Rombach, J. Hoffmann, G. Roth, O. Piepenburg, Y. Park, R. Zengerle, and F. Von Stetten,

Lab-on-a-chip solutions designed for being operated on standard laboratory instruments, Procedia Engineering **5**, 444 (2010).

[7] B. Godber, K. S. J. Thompson, M. Rehak, Y. Uludag, S. Kelling, A. Sleptsov, M. Frogley, K. Wiehler, C. Whalen, and M. A. Cooper, *Direct quantification of analyte concentration by resonant acoustic profiling,* Clin. Chem. **51**, 1962 (2005).

[8] S. Haeberle and R. Zengerle, *Microfluidic platforms for lab-on-a-chip applications,* Lab Chip **7**, 1094 (2007).

[9] K. L. Linask and C. W. Lo, *High-throughput mouse genotyping using robotics automation,* BioTechniques **38**, 219 (2005).

[10] I. Meyvantsson, J. W. Warrick, S. Hayes, A. Skoien, and D. J. Beebe, *Automated cell culture in high density tubeless microfluidic device arrays,* Lab Chip **8**, 717 (2008).

[11] L. Jongpaiboonkit, W. J. King, G. E. Lyons, A. L. Paguirigan, J. W. Warrick, D. J. Beebe, and W. L. Murphy, *An adaptable hydrogel array format for 3-dimensional cell culture and analysis,* Biomaterials **29**, 3346 (2008).

[12] J. S. Rossier, S. Baranek, P. Morier, C. Vollet, F. Vulliet, Y. De Chastonay, and F. Reymond, *GRAVI: Robotized microfluidics for fast and automated immunoassays in low volume,* JALA **13**, 322 (2008).

[13] A. E. Guber, M. Heckele, D. Herrmann, A. Muslija, V. Saile, L. Eichhorn, T. Gietzelt, W. Hoffmann, P. C. Hauser, J. Tanyanyiwa, A. Gerlach, N. Gottschlich, and G. Knebel, *Microfluidic lab-on-a-chip systems based on polymers—fabrication and application,* Chem. Eng. J. **101**, 447 (2004), 7th International Conference on Microreaction Technology.

[14] C. Moraes, G. Mehta, S. Lesher-Perez, and S. Takayama, *Organs-on-a-Chip: A focus on compartmentalized microdevices,* Ann. Biomed. Eng. **40**, 1211 (2012).

[15] B. W. Segelke, J. Schafer, M. A. Coleman, T. P. Lekin, D. Toppani, K. J. Skowronek, K. A. Kantardjieff, and B. Rupp, *Laboratory scale structural genomics,* J. Struct. Funct. Genomics **5**, 147 (2004).

[16] A. Y. Hsiao, Y.-C. Tung, X. Qu, L. R. Patel, K. J. Pienta, and S. Takayama, *384 hanging drop arrays give excellent Z-factors and allow versatile formation of co-culture spheroids,* Biotechnol. Bioeng. **109**, 1293 (2012).

[17] M. Pla-Roca, R. F. Leulmi, H. Djambazian, S. Sundararajan, and D. Juncker, *Addressable nanowell arrays formed using reversibly sealable hybrid elastomer-metal stencils,* Anal. Chem. **82**, 3848 (2010).

[18] C. Sauter, K. Dhouib, and B. Lorber, *From macrofluidics to microfluidics for the crystallization of biological macromolecules,* Cryst. Growth Des. **7**, 2247 (2007).

[19] D. C. Duffy, J. C. McDonald, O. J. A. Schueller, and G. M. Whitesides, *Rapid prototyping of microfluidic systems in poly(dimethylsiloxane),* Anal. Chem. **70**, 4974 (1998).

[20] A. Waldbaur, H. Rapp, K. Lange, and B. E. Rapp, *Let there be chip-towards rapid prototyping of microfluidic devices: one-step manufacturing processes,* Anal. Methods **3**, 2681 (2011).

[21] K. Haubert, T. Drier, and D. Beebe, *Pdms bonding by means of a portable, low-cost corona system,* Lab Chip **6**, 1548 (2006).

[22] C.-C. Hong, J.-W. Choi, and C. H. Ahn, *A novel in-plane passive microfluidic mixer with modified Tesla structures,* Lab Chip **4**, 109 (2004).

[23] E. L. McGown and D. G. Hafeman, *Multichannel pipettor performance verified by measuring pathlength of reagent dispensed into a microplate,* Anal. Biochem. **258**, 155 (1998).

5

QSAR modeling of orientation sensitive biomolecular binding on ion-exchange surfaces

Quantitative structure-activity relationship (QSAR) modeling for prediction of biomolecule parameters has become an established technique in chromatographic purification process design. Unfortunately available descriptor sets fail to describe the orientation of biomolecules and the effects of ionic strength in the mobile phase on the interaction with the stationary phase. The literature describes several special descriptors used for chromatographic retention modeling, all of these do not describe the screening of electrostatic potential by the mobile phase in use. In this work we introduce two new approaches of descriptor calculations, namely surface patches and plane projection, which capture an oriented binding to charged surfaces and steric hindrance of the interaction with chromatographic ligands with regard to electrostatic potential screening by mobile phase ions. We present the use of the developed descriptor sets for predictive modeling of Langmuir isotherms for proteins at different pH values between pH 5 and 10 and varying ionic strength in the range of 10 to 100 mM. The resulting model has a high correlation of calculated descriptors and experimental results, with a coefficient of determination of 0.82 and a predictive coefficient of determination of 0.92 for unknown molecular structures and conditions. The agreement of calculated molecular interaction orientations with both, experimental results as well as molecular dynamic simulations from literature is shown.

This chapter has been published in J. Chromatogr. A 1482 (2017) [1].

The developed descriptors provide the means for improved QSAR models of chromatographic processes, as they reflect the complex interactions of biomolecules with chromatographic phases.

5.1. Introduction

Chromatography is the most widely applied technique in downstream processing of biopharmaceutical products. Due to the great interest in the underlying process mechanistics and the consequential research, the fundamental understanding of chromatographic separation and adsorption mechanisms have widely advanced in the last decades. Today, a variety of modeling techniques can be applied to describe the influence of different parameters on ion-exchange chromatography, including mobile-phase composition, resin types, and protein characteristics.

Comprehensive studies on protein retention in different mobile phase conditions have lead to semi-empirical approaches of protein behavior modeling. Rounds and Regnier introduced the stoichiometric displacement model, describing the interaction between proteins and counter ions [2]. The steric mass action (SMA) model by Brooks and Cramer [3] incorporates the shielding of adsorber charges by bound macromolecules. An extension of this model was introduced by Bosma and Wesselingh, who used the charge of proteins to describe protein binding at different pH values [4]. Although, they assumed a constant electrostatic charge, neglecting the pH dependent protonation of charged residues. These models require the determination of protein and adsorbent specific parameters, and therefore can not be applied to predict the behavior of unknown molecules. Further, they are based on the assumption that the binding mechanistics remain unchanged throughout the conditional space inquired.

That this is not the case in chromatographic interactions, was shown by molecular dynamic (MD) simulations, which aim to describe the interaction of molecules on the atomic level. It has been shown that specific orientations are predominant in binding of biomolecules to charged surfaces, dependent on charge distribution of the molecule, steric factors, charge type, and capacity of the corresponding surface. The interaction energies and orientation of biomolecules in cation-exchange chromatography were investigated experimentally and *in silico* via MD simulations by Dismer et al. [5–7]. This approach was later extended to complex proteins (i.e. including non-standard residues) and anion-exchange chromatography by Lang et al. [8]. Dismer and Lang determined a pH dependent change in orientation for lysozyme on a SP Sepharose FF adsorbent. The primary binding orientation was determined by the distribution of the amino acids lysine and arginine with a change of the main binding factor from lysine 13 at low pH towards lysine 96 and 97 at high pH values [7, 8]. Despite fast computational development and detailed information to be gained, MD simulations remain inapplicable to large datasets and structures, e.g. monoclonal antibodies, due to the computational costs required.

Quantitative structure-activity relationship (QSAR) techniques aim at developing predictive models based on descriptors derived from molecular structural information. A critical part are the descriptors utilized to build a model, as they need to incorporate the structural features which lead to the observed activity. To capture the molecular information corresponding to ion-exchange chromatography,

Mazza et al. [9] introduced an approach to map electrostatic potential onto the surface of a molecule to derive pH dependent descriptors. They predicted retention times for a variety of beta-blockers and similar chemical structures published by Law and Weir [10].

Malmquist et al. [11] developed their own set of pH sensitive descriptors, based on the surface mapping of atomic charges and applied them successfully in cation and anion-exchange chromatography modeling of model proteins. Based on this work, Yang et al. [12] developed a set of electrostatic potential descriptors, and applied them in SMA parameter modeling, wherein different pH conditions of the same molecule were presented as separate observations to the model. All approaches have in common that they do not reflect the orientation of the molecule and characteristics of the mobile phase. This is especially the case, if predicted parameters are subsequently applied in aforementioned semi-empirical models.

In this approach a new set of QSAR descriptors is introduced, describing the physical proximity of biomolecules to a surface in 3D space. Thereby describing steric hindrance and interaction as well as interaction shielding by the mobile phase. This allows for orientation sensitive QSAR modeling over different ionic strengths and pH values, reflecting the parameter range in ion-exchange chromatography. The predictive capabilities are demonstrated in Langmuir isotherm parameter modeling for model proteins in cation-exchange chromatography. A detailed comparison of interaction orientations of Lysozyme on SP Sepharose FF to experimental and *in silico* data from MD simulations, shows consistent results without the need of comprehensive and time-consuming simulations.

5.2. Theory

5.2.1. Parameter projection

A molecular property on a given point in space is generated as the result of the intermolecular interactions between all fragments of the molecule and the mobile phase at the given point.

$$MP_k = \sum_{i}^{N} f_i fct(d_{ik}) \tag{5.1}$$

With MP being the molecular property at a given point k in space, i denotes the fragment and N the total number of fragments in the molecule. f_i represents the parameter constant of the fragment and fct the distance function with d_{ik} being the distance between the fragment i and the point in space k. This applies to calculation of both molecular surface properties as well as intermolecular forces as the projection towards the representation of adsorbent surface structures as described in this work. Three types of projections are used, namely the complete molecular surface, a projection towards an adjunct plane, and patches of the surface.

Molecular surface To represent the distribution and values of molecule electrostatic potential and hydrophobicity characteristics on the molecule surface, these properties are mapped to the solvent accessible surface area (SASA) of the molecule. Single values for atoms (electrostatic potential) and amino acids (hydrophobicity coefficient) are mapped over distance towards the surface represented by distinct points with a density of 1 point per $Å^2$. Descriptors based on the molecular surface of the molecule are then calculated for all points on the surface. The attenuation of electrostatic and hydrophobicity characteristics distance are described in section 5.2.2 and 5.2.3, respectively.

Plane projection To represent different orientations of a molecule towards the adsorbent surface, a representation of said surface was constructed as a grid of points in three dimensional space, which is depicted in Fig. 5.1. This representation can be repositioned around the molecule to capture the effects of reorientation of molecules. To represent all possible orientations of the molecule towards a surface, the molecule is represented by a sphere with the geometric center of the molecule in the sphere center. Based on the sphere 120 equidistant orientations were calculated. This compares to 62 grid based orientations used by Dismer et al. [7], which lead to an overrepresentation of the sphere poles and 50 equidistantly distributed orientations used by Lang et al. [8]. Orientation vectors were calculated based on the Thomson equilibrium of a sphere [13] with the vector origin placed in the geometric center of the molecule, yielding equidistantly distributed orientations. For each vector a plane of 120 Å edge length and a grid point density of 5 Å was calculated and positioned with the vector as normal vector in the plane center. The plane was shifted along the normal vector to fit the distance to the molecular surface, which was set to 5.0 Å, in accordance to the set-up of orientation sensitive MD simulations performed by Dismer [7] and Lang [8].

Surface patch To capture properties on the molecular surface sensitive to the orientation of the molecule towards an interacting surface, surface patches were calculated. Based on the calculated planes, surface patches were defined to include the SASA with a distance of below 20 Å towards the corresponding plane representation.

5.2.2. Electrostatic potential

It is known that charge and electrostatic potential properties govern the interactions of molecules in ion-exchange chromatography. In fact, the adjustment of counter ion concentration in the mobile phase to alter the long range electrostatic forces in biomolecule-adsorbent interaction is one of the most used parameters to direct protein separation (i.e. salt gradient). Coulomb interactions between molecules are reduced due to the polarization of the particles in a dielectric medium as water. Particles, depending on their character, form induced and permanent dipoles which will be oriented around a free charge so to terminate some of the field lines originating from the charge. This effect increases with increasing concentration of ions

in the medium, not only reducing the strength of the charge but also changing the shape of the potential energy, approaching zero exponentially beyond a characteristic distance. This dielectric screening of charge is one of the key parameters in ion-exchange chromatography, which is realized by adding salt to induce desorption and therewith elution.

A detailed calculation of energies and dipole formation around biomolecules can only be performed in MD simulations, with the known drawback of computational power increasing exponentially with the size of the molecular system investigated. An approximation is the calculation of the electrostatic potential EP, which for small distances d where there are no further charges considered, is written as

$$EP(d) = \frac{q}{\varepsilon d} \tag{5.2}$$

with the charge q and relative permittivity $\varepsilon = \varepsilon_R \cdot \varepsilon_0$ given as the absolute permittivity of the material as a ratio to the permittivity of vacuum. For long distances the screening of charges can be described by the addition of the Debye screening length λ_D

$$EP(d) = \frac{q}{\varepsilon d} e^{-d/\lambda_D} \tag{5.3}$$

λ_D writes for a monovalent electrolyte (e.g. NaCl),

$$\lambda_D = \left(\frac{\varepsilon k T}{2 \rho e^2} \right)^{1/2} \tag{5.4}$$

with the Boltzmann constant k, the number density of electrons ρ, and the charge of electron e. With the electron charge is entered squared, the ionic strength has a strong influence on the screening length which decreases drastically with increasing ionic strength. The Debye screening length, is for example 96 Å at 1 mM, 9.6 Å at 100 mM and 3.0 Å at 1 M NaCl in water at 22 °C.

5.2.3. Molecular hydrophobicity

The hydrophobic character of amino acids is of great interest in protein biochemistry, as it defines protein 3D structure and folding. Hydrophobicity is also known to influence the interaction with adsorbent structures as especially adsorbent backbone structures can show hydrophilic or hydrophobic characteristics, thereby influencing protein interaction.

Despite the strong interest and extensive research in the hydrophobic character of amino acids, no consensus on hydrophobicity parameters is reached as of today. This is primarily due to a wide range of measurement techniques used, including partitioning, RPLC chromatographic techniques, accessible surface area calculations, site-directed mutagenesis, and physical property measurements. Comprehensive reviews on techniques, amino acid hydrophobicity scales and their limitations were presented by Karplus [14] and Biswas [15], among others.

The parameters by Kyte and Doolittle are one of the most frequently cited amino acid hydrophobicity scales. They combine accessible surface area measurements with water-vapor partitioning values as well as an extensive library of proteins for property calculation [16].

Next to the hydrophobicity value, a distance function is needed to describe the decrease of the hydrophilic parameter in space. At the origin of the amino acid the value is at maximum, and approaches zero over increasing distances. Audry introduced the concept of molecular lipophilicity potentials and used a hyperbolic distance function [17], while Fauchère used an exponential function to study hydrophobicity at molecular solvent accessible surfaces [18]. Gaillard introduced a modified Fauchère function [19] with $e^{-d/2}$ and a cut-off at 4 Å.

In this work, the hydrophobicity parameters from Kyte and Doolittle [16] were assigned to the corresponding amino acids and projected to points on the solvent accessible surface, which were calculated as described in 5.2.1. The original distance function by Fauchère without cut-off was used to calculate the lipophilicity potential LP defined by the lipophilicity parameter l within distance d [18].

$$LP(d) = l * e^{-d} \tag{5.5}$$

Although the terms ”lipophilic” and ”hydrophobic” are not strictly synonymous, they do describe similar tendencies with regard to dispersion forces and are referenced as hydrophobicity parameter in this work.

5.2.4. Descriptors

The descriptors calculated are listed in Table 5.1 and are based on the descriptors developed by Malmquist et al. [11] and Yang et al. [12]. Hydrophobicity based descriptors were calculated based on the corresponding surface projection and surface patches but not for plane based descriptors. This is to account for the short distanced nature of hydrophobic interactions. Descriptors based on electrostatic potential were calculated based on all three projections (molecular surface, surfaces patches, and plane projection) as described in 5.2.1. From the orientation sensitive descriptor sets, only those of the orientation with the highest sum of projected electrostatic potential were selected for model generation.

5.3. Materials and methods

5.3.1. Batch isotherm experiments

The Langmuir affinity coefficients were determined experimentally for different model proteins under varying mobile phase conditions to provide data for QSAR model generation and evaluation. Adsorption isotherms were generated in batch mode based on an automated high-throughput screening process in 384 well plates as described in [20]. In these experiments protein solutions of different concentration are incubated with defined amounts of resin until an equilibrium is reached. Then, the resin is sedimented by centrifugation and the protein concentration in the supernatant is measured. Based on concentration change in the supernatant

descriptor	definition
totalSurf	surface area of the projection (Å)
nSurfP	number of surface points
sum	sum of mapped property
mean	mean value of mapped property
median	median value of mapped property
meanRes	mean value corrected for resolution of surface points
max	maximum value of mapped property
min	minimum value of mapped property
dev	value deviation
var	variance of property
nPos	number of points with positive value
nNeg	number of points with negative value
relPos	fraction of points with positive value
relNeg	fraction of points with negative value
averPos	average value of points with positive value
averNeg	average value of points with negative value
bin0 – bin9	number of points with mapped property value in the range of the bin

Table 5.1: Projection descriptor set with names and description of calculation. Descriptors are calculated based on plane projection, surface patch and the full molecular surface for electrostatic potential (EP) and hydrophobicity characteristics (hyd).

and volume ratios, the protein load on the resin is calculated.
Proteins used are listed in Table 5.2 and were all purchased from Sigma Aldrich (St. Louis, MO, USA). Experiments were performed in the range of pH 5.0-11.0. Mobile phase was prepared with 10 mM buffer systems of sodium acetate at pH < 6.0, sodium phosphate in the range of pH 6 to 7.5, and bicine at pH > 7.5. Solvent ionic capacities were adjusted with sodium chloride to values in the range of 30 to 120 mM. All buffer components and salts were purchased from Merck (Darmstadt, Germany). The adsorbent SP Sepharose FF, a cation exchange resin, was obtained from GE Healthcare (Uppsala, Sweden). Resulting isotherms were fitted with the Langmuir isotherm equation:

$$q = q_{max} \cdot k_a \frac{c}{1 + k_a \cdot c} \tag{5.6}$$

where q and c are equilibrium protein concentrations on the stationary, respectively mobile phase, q_{max} the saturation concentration in the stationary phase and k_a the equilibrium coefficient.

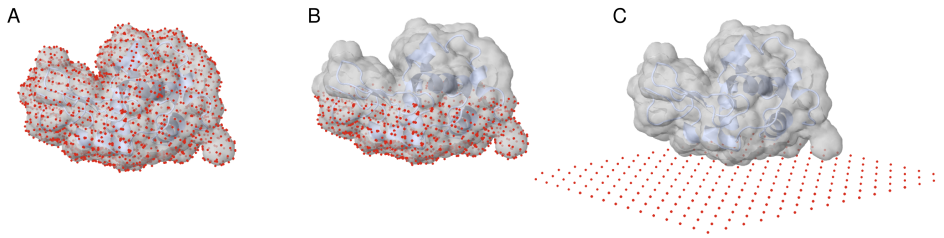

Figure 5.1: Parameter projection types applied. A) Molecular properties are projected to the solvent accessible surface of the molecule, which is represented by distinct points with a density of 1 point per $Å^2$, B) In patch orientation only a part of the surface is considered in descriptor calculation to represent an oriented interaction between molecule and solid phase during adsorption, and C) in plane orientation molecule electrostatic properties are projected towards a theoretical plane which is placed in a fixed distance to the molecule, thereby reflecting the steric hindrance of charges by the 3 dimensional structure of the molecule. Orientation sensitive projection types are calculated for equally distributed orientations and the orientation with the strongest interaction is included in the QSAR model generation.

5.3.2. Modeling

The modeling work consisted of three major steps, namely protein structure identification and preparation, descriptor calculation, and generation of the actual QSAR model, including its evaluation.

Protein Structure Preparation Proteins were selected based on isoelectric point, to ensure binding on cation-exchange resin at chosen conditions, and availability of structural information. Corresponding RCSB entries were selected based on most complete structure and high resolution. Table 5.2 lists the chosen proteins, corresponding UniProt IDs and PDB IDs for structural information used for descriptor calculation.

Structure preparation was done with the simulation software YASARA [21], which provides a graphic user interface and is capable of parameterizing non-standard residues (e.g. bound calcium or a heme-group) via the integrated AutoSmiles algorithm [22–24]. Loaded structures were cleaned from water and substitutes, a simulation box around the molecular structure was defined and the corresponding pH value was set. The structure was protonated according to set pH value.

To correct the covalent geometry, an energy minimization was performed with the AMBER03 force field [25], using a 7.86 Å cut-off. After an initial steepest descent minimization, a simulated annealing was performed to obtain an optimized structure. Partial charges of molecule atoms were derived based on the AMBER03 force field and saved along with molecule information in pqr file format, which captures atomic charges and radii information. Solvent accessible surface area with probing radius of 1.4 Å and a point density of 1 point per $Å^2$ surface was calculated to represent the molecular surface. The mapping of electrostatic potential on the SASA of proteins was based on equation 5.2, while the projection to a plane as described in 5.2.1 was based on equation 5.3 under consideration of the electrolyte concentration in the medium.

name, origin	UNIProt ID	PDB
lysozyme C,hen egg	P00698	1LYZ
chymotrypsinogen A, bovine	P00766	2CGA
α-chymotrypsin, bovine	P00766	1YPH
cytochrome C, equine	P00004	1HRC
cytochrome C, bovine	P00125	2B4Z
ribonuclease A, bovine	P61823	1FS3
phospholipase A2, bovine	P00593	1BP2
myoglobin, equine	P68082	2V1F

Table 5.2: Proteins used in QSAR model generation and evaluation with their corresponding UniProt ID and PDB ID selected for descriptor calculation.

Descriptors Calculated descriptor values were prepared by applying a unity-based normalization to allow for comparison of descriptor weights in the final model. To reflect the nonlinear character of the affinity coefficient in the Langmuir isotherm equation, k_a values were modeled in logarithmic scale.

Model The QSAR model was built as an ensemble of single regression models, with an internal cross validation. Thereby, the dataset was repeatedly split randomly into a training set (containing 80 % of protein data) and an internal test set (20 %). 100 regression models were generated and those with an r^2 above 0.8 and an predictive r^2 above 0.6 for the internal test set were subsequently pooled within the ensemble. This approach proved to be robust against over-fitting of single models and the influence of outliers, especially in small datasets [26].
The enhanced replacement method (ERM) published by Mercader et al. [27] was used for single model generation. This algorithm follows a simulated annealing approach, which prevents sub-optimal results by local optima. It has been shown, that the method is performing better than regression and partial-least-square (PLS) models, and comparable to genetic algorithms (GA), without the need to set parameters beforehand [28].
Additionally, different external test sets were split from the dataset prior to model development to demonstrate the predictive capabilities of the model.

5.4. Results and discussion

5.4.1. Model response

A successful QSAR model correlates experimentally determined molecular properties to descriptors. A model which proves to be predictive, can be used to obtain molecular activities for experimental conditions the model was not trained on.

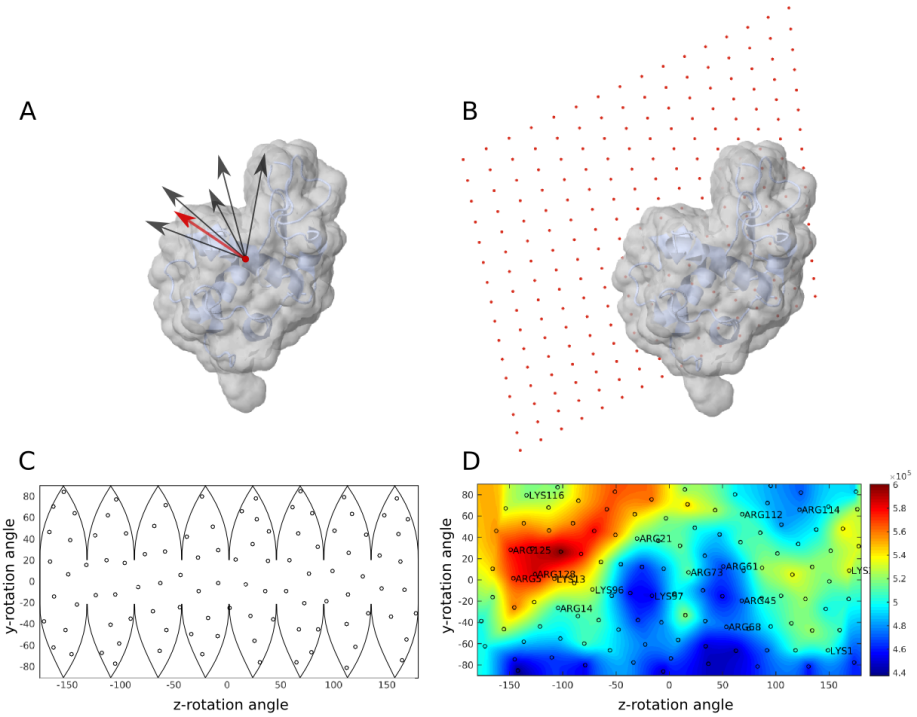

Figure 5.2: Representation of surface plot generation based on electrostatic potential mapping in plane projection approach. Originating from the molecule center 120 equidistant vectors are calculated (A). A surface plane is constructed with the reversed vector as the normal vector. The plane is positioned along the vector to a distance of 5 Å to the molecular surface (B). Molecular properties are projected onto the surface and descriptors are calculated based this on projection. Descriptors from the different orientations are mapped towards a two dimensional representation (C) of the molecule surface and descriptor values are interpolated (D).

A model correlating the affinity coefficient k_a of the Langmuir equation was calculated for the interaction of model proteins with SP Sepharose FF at 31 different conditions. 5 sets of descriptors were calculated, characterizing electrostatic potential and hydrophilicity based on the presented projections. Hydrophobicity parameter based on surface projection where not included, as a full molecule representation of short range hydrophobicity characteristics was not considered representative for the oriented interaction mechanistic in investigation. The model is shown in Fig. 5.3. Experimentally determined k_a values of 31 samples of proteins at different conditions of pH value and ionic strength were used for model generation.

As proof of concept, 5 conditions not used in training of the model were used as an external test to test model predictive ability. The conditions predicted reflect different molecular structures as well as combinations of high and low pH conditions and ionic strength. In model calculation and representation experimental values

were applied in logarithmic scale to reflect the non-linear character of k_a value in the Langmuir isotherm equation. The model correlation, described by the r^2 coefficient was 0.816, the relative standard deviation for observations was 0.206. The correlation for the external test set, the predictive r^2 was 0.922. The accuracy of the model is within the means of the experimental data variation, which is discussed in detail in [20]. In the forementioned study it is described, that low k_a values are easily underestimated in the fit of the Langmuir isotherm equation to experimental data, which is in agreement with the deviation between experimental and predicted results for k_a values below 1.0, where higher values were predicted. It is therefore concluded, that the deviation between QSAR model and experimental data fit is due to the limitation in the experimental method for small k_a values.

Predictions for external test sets are in good agreement with experimental data and model quality is in range of experimental data accuracy [20]. The agreement of experimental results and predicted parameters is exemplarily shown in the resulting iostherms for ribonuclease A at high pH and low ionic strength and cytochrome C at low pH and high ionic strength (see Fig. 5.4).

In the representation of biomolecules by descriptors, QSAR modeling relies on the accuracy of the initial molecule structure these descriptors are based on. Therefore the approach does not account for structural changes due to the interaction with the solid phase or the formation of complexes as dimers. Molecular modeling could reflect these conditions, but seldomly is conducted over sufficient time frames. In a preparative environment, these conditions are considered unfavorable for chromatography.

5.4.2. Selected descriptors

Descriptors used in the calculation of QSAR models can provide insight in the mechanistic characteristics of the molecular property modeled. Although, this insight can not be conclusive, as different combinations of descriptors can potentially result in predictive models. Nevertheless, the type of descriptors utilized can confirm mechanistic understanding and thereby qualify the model or direct further research. The eight main descriptors, which describe 90.9 % of the model variance based on descriptor weight are listed in Table 5.1. Six out of eight descriptors are related to electrostatic potential (EP) and two depict hydrophobicity characteristics (hyd). This reflects the driving forces of electrostatic binding in ion-exchange chromatography. Positive electrostatic potential have a high weight in the model with descriptors bin-Abs 7 and 9 (descriptors 3-6). Negative electrostatic potentials (descriptors 1 and 2) are accounted negatively, which is in agreement with a positively charged molecule binding onto a negatively charged surface as in cation-exchange chromatography. Hydrophobic molecule character is weighted positively in the model (descriptor 8) and low hydrophobicity values are being weighted negatively (descriptor 7), therefore favoring interaction of hydrophobic areas of the molecular surface with the stationary phase.

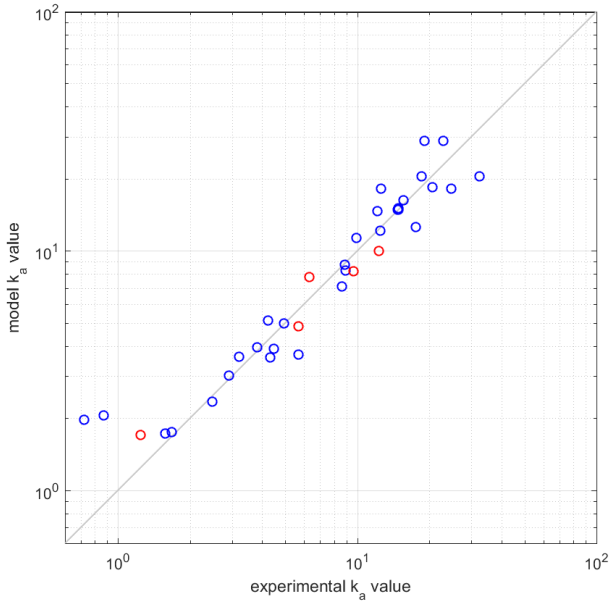

Figure 5.3: QSAR model for k_a affinity coefficient of Langmuir binding isotherm parameter with depicton of training set (blue) and external test set (red) with molecular structures which were not part of model training.

All three property projection types are represented in the model descriptors. Plane based descriptors are most used with 6 out of 8 main model descriptors (1-4 and 7-8). Patch and surface based projections are represented by one descriptor each (5 and 6). This is in agreement with different ionic strength conditions being implemented in the model. The steric hindrance in the interaction of molecule areas and the resin and the increasing shielding of electrostatic potential by increasing ionic strength of the medium, i.e. the buffer system, are represented only in the plane descriptors.

5.4.3. Molecular orientation

The position of charged amino acids in proteins define the electrostatic potential projected to a surface close to the molecule. To represent this three dimensional information surface plots were used, which are similar to those used by Lang et al. [8]. The concept of the surface plots is shown in Figure 5.2.

The electrostatic potential, as projected to different plane orientations around lysozyme is presented in Figure 5.5. It is apparent from the surface plots for pH values from 5 to 12, that descriptor values vary dependent on the protein orientation, showing preferred binding orientations. Herein, high descriptor values indicate preferred binding to negatively charged cation-exchange adsorbers. Positively

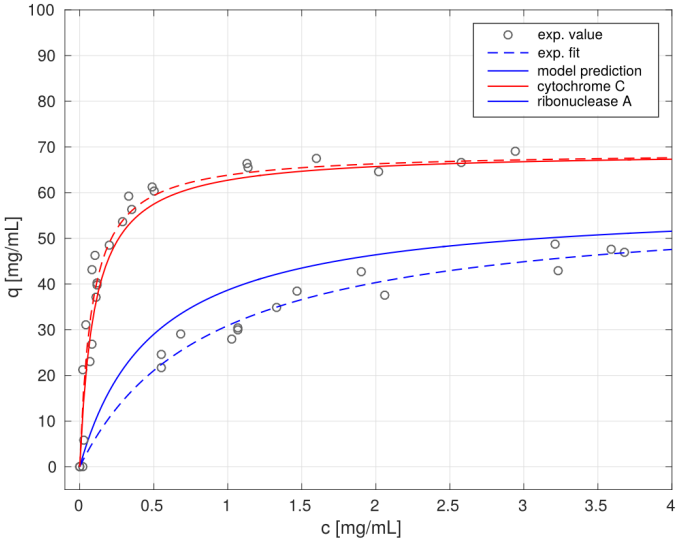

Figure 5.4: Batch isotherms of ribonuclease A (pH 9.0, 10 mM ionic strength) and cytochrome C (pH 5.0, 90 mM ionic strength), with fit of experimental data and predictions from QSAR model. Both conditions were part of the external test set.

charged amino acids (arginines, histidines, lysines) are plotted with their index and highlighted in corresponding structure representations to visualize their influence on electrostatic potential projected to the plane.

Four distinct preferred binding orientations (A-D) are identified, with a change in preference throughout the pH parameter range investigated. At pH values of 5 and 9 orientation A shows the highest descriptor value, with Lys 1, 33 and Arg 14 and 128 facing the surface plane (Figure 5.5 A). The preferred binding patch at low pH values extends to orientation B, which is defined by the proximity of Lys 116 and Arg 21, and 123 to the surface plane. At high pH values of 11 and 12, orientations A and B are not longer preferred in favor of patches C and D, with C being more distinct at pH 11. Orientation C is composed by the charged amino acids Lys 33 and 116, as well as Arg 112 and 114. Lys 13 and 96/97 and Arg 14, 21, and 73 contribute to orientation D.

The found electrostatic potential maps are in good agreement with experimental data from Dismer et al. [6], who investigated the binding orientation of lysozyme on SP Sepharose FF experimentally, by comparing labeling efficiency of lysines in bound and unbound protein state at different pH values. They found steric hindrance of Lys 1 and 33, at low pH values with Lys 13 being especially hindered if high saturation of resin is reached. This indicates Lys 13 being in close range to the surface, without being hindered by the surface, rather than by other proteins in close range. This is consistent with the found preferred orientation A. Both orienta-

	descriptor	group	weight
1	binAbs 4	plane, EP	-0.1259
2	min	plane, EP	-0.0735
3	binAbs 9	plane, EP	0.0636
4	binAbs 7	plane, EP	0.0609
5	binAbs 7	patch, EP	0.1656
6	binAbs 9	surface, EP	0.1113
7	min	plane, hyd	-0.0972
8	sum	plane, hyd	0.2109

Table 5.3: Most influential descriptors and their weight in the QSAR model. The properties mapped are indicated with EP (electrostatic potential) and hyd (hydrophobicity). The 8 descriptors listed account for 90.9 % of model variance based on descriptor weight.

tions, A and B correspond to the previously by MD simulations identified dominating binding orientations at low pH values [7]. Dismer et al. further described hindrance of Lys 96/97 at pH 12, which is consistent with found orientation D, where the binding site forms around these two neighboring lysines [6] and Arg 14, 21, and 73.

The mapped electrostatic potential descriptor value expectedly decreases as the the pH gets closer to the pI, as positively charged amino acids lysine and arginine are deprotonated. The descriptor values are in the negative range for pH 11 and 12 with maximum values of -4×10^4 (pH 11) and -5.9×10^7 (pH 12). Nevertheless, protein retention on SP Sepharose FF is documented experimentally [6] and *in silico* via MD simulations [7, 8] at pH 11 and 12 for selected molecule orientations. This is due to the descriptor calculation approach, which does not account for relocation of amino acids at the molecule surface and ligands, when in close contact. MD simulations have shown that amino acids reorient to energetically more favorable positions away from or towards the ligand surface, depending on the charge. Lang et al. [8] describe the strong contribution of arginines to the binding of lysozyme to SP Sepharose FF, as they are positively charged over a wide pH range and provide a high side-chain flexibility.

A critical parameter in protein retention on ion-exchange resins is ionic strength, which is controlled by salt concentration. Figure 5.6 displays the influence of the mobile phase ionic strength on electrostatic potential mapping to a surface plane for lysozyme at pH 5 and ionic strength of 60, 90, and 120 mM, conditions applicable in chromatographic separation. Electrostatic potential descriptor values, representing binding strength, decrease with increasing ionic strength. The strongest potential with 6.0×10^5 at 60 mM for orientation A, is reduced to 3.95×10^5 at 90 mM, and 2.96×10^5 at 120 mM ionic strength, exhibiting an exponential decay of electrostatic potential mapped concurrent to the screening function applied. Further, a difference in screening between orientations A and B can be observed, with orientation B being less favorable for binding at higher ionic strength (120 mM), compared to low ionic strength (90 mM). This is consistent with the observation of a more flat

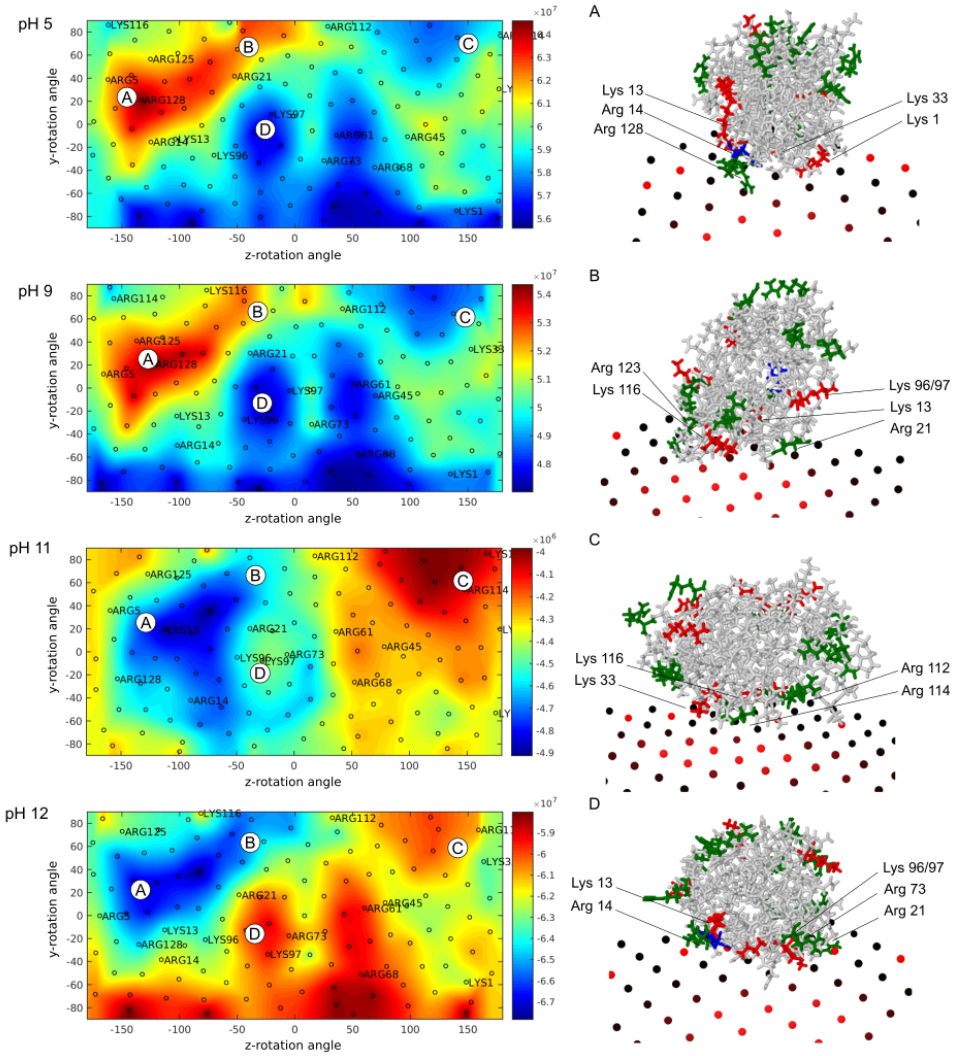

Figure 5.5: Mapped representation of electrostatic potential as calculated by plane projection with 120 orientations. Orientations of interest are labeled (A-D). Charged amino acids close to the angle of the projection plane are referenced, for better orientation. Labels do not indicate binding contributing amino acids. Right hand side structures visualize selected orientations A to D. Charged amino acids lysine (red), arginine (green) and terminal histidine (blue) are color coded.

position of lysozyme in orientation A, which brings more charged amino acids in proximity to the surface representation. In contrast, orientation B is more upright, where the binding strength is defined by fewer amino acids, which are closer to the surface, but are therefore screened stronger. The screening of electrostatics

also emphasize the impact of van der Waals forces and hydrophilic properties on molecule binding.

Mapping results and comparison to literature show that descriptor value mapping provides reliable information on preferred protein orientation and contributing amino acids. Found binding orientations are throughout consistent with computational more expensive MD simulations and experimental results. The non-linear screening of electrostatic forces and steric hindrance of contact between charged amino acids and ligands can result in different preferential orientations.

5.5. Conclusion and outlook

The implementation of interaction characteristics as preferred orientation, steric hindrance and influences by the surrounding liquid phase in QSAR descriptor calculation enables a detailed representation of biomolecule retention in ion-exchange chromatography. A predictive QSAR model to describe protein affinity on SP Sepharose FF was calculated for eight model proteins over a wide pH range and varying ionic strength conditions. Plane projection descriptors were further used to investigate preferred binding orientation of lysozyme, which was found to be in agreement with experimental and MD results. This QSAR approach is therefore robust against changes in preferential orientations in binding of biomolecules to stationary phases in chromatography due to mobile phase conditions.

The computational costs of descriptor calculation and modeling are low, compared to MD simulations. This allows for larger datasets and large molecular structures, as monoclonal antibodies. Further, the application is not limited to ion-exchange chromatography, as an orientation sensitive interaction of biomolecules towards a counterpart is intrinsic to many process unit operations, e.g. other modes of chromatography, aggregation or solubility. The presented calculation methods are therefore a powerful tool for the prediction of molecular properties in downstream process development.

Acknowledgments

Financial support for this work was provided by Lonza Biologics, PLC. The authors wish to thank Frank Hämmerling for his suggestions on preparing the manuscript.

References

[1] J. Kittelmann, K. M. Lang, M. Ottens, and J. Hubbuch, *An orientation sensitive approach in biomolecule interaction QSAR and application in ion-exchange chromatography,* J. Chromatogr. A **1482**, 48 (2017).

[2] 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. **283**, 37 (1984).

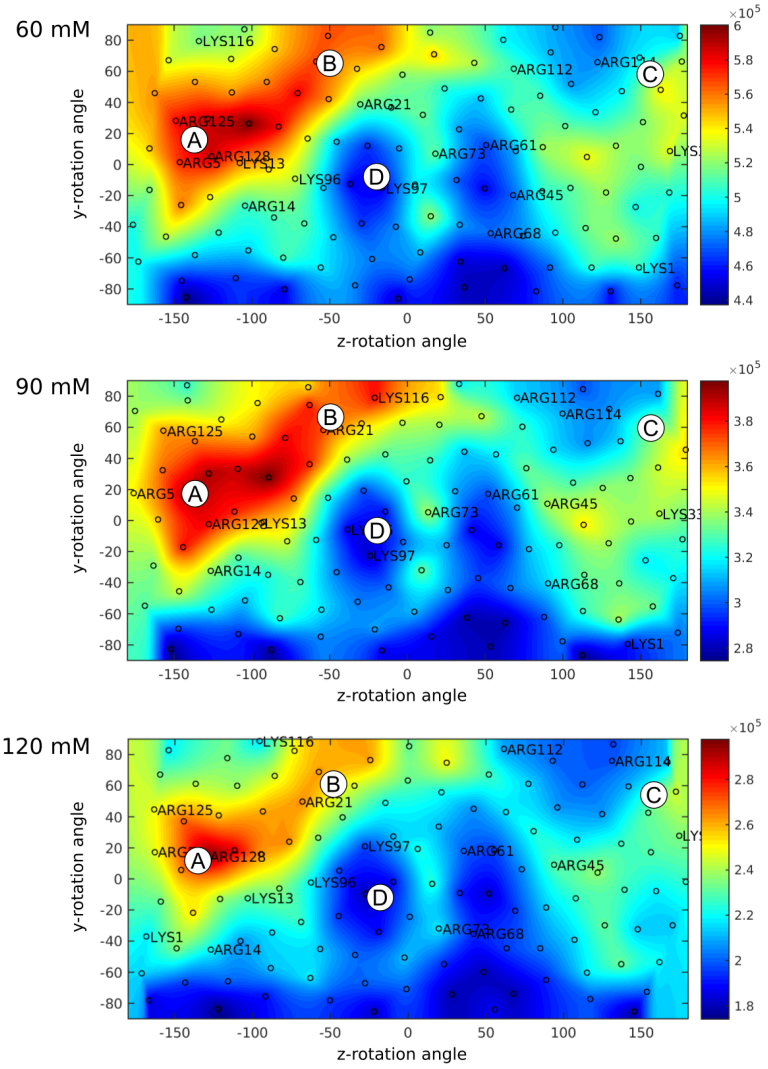

Figure 5.6: Mapped representation of electrostatic potential as calculated by plane projection with 120 orientations for varying ionic strength values of 60, 90 and 120 mM. Charged amino acids close to projection plane are referenced.

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

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

[5] F. Dismer and J. Hubbuch, *A novel approach to characterize the binding orientation of lysozyme on ion-exchange resins,* J. Chromatogr. A **1149**, 312 (2007).

[6] 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**, 11 (2008).

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

[8] K. M. 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 (2015).

[9] C. B. Mazza, N. Sukumar, C. M. Breneman, and S. M. Cramer, *Prediction of protein retention in ion-exchange systems using molecular descriptors obtained from crystal structure,* Anal. Chem. **73**, 5457 (2001).

[10] B. Law and S. Weir, *Quantitative structure-retention relationships for secondary interactions in cation-exchange liquid chromatography,* J. Chromatogr. A **657**, 17 (1993).

[11] 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**, 164 (2006).

[12] T. Yang, M. C. Sundling, A. S. Freed, C. M. Breneman, and S. M. Cramer, *Prediction of pH-dependent chromatographic behavior in ion-exchange systems,* Anal. Chem. **79**, 8927 (2007).

[13] J. Thomson and F. Cavendish, *On the structure of the atom,* Phil. Mag. **7**, 237 (1904).

[14] P. A. Karplus, *Hydrophobicity regained,* Protein Sci. **6**, 1302 (1997).

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

[16] J. Kyte and R. F. Doolittle, *A simple method for displaying the hydropathic character of a protein,* J. Mol. Biol. **157**, 105 (1982).

[17] E. Audry, J. Dubost, J. Colleter, and P. Dallet, *A new approach to structure-activity relations: the molecular lipophilicity potential,* Eur. J. Med. Chem **21**, 71 (1986).

[18] J.-L. Fauchére, P. Quarendon, and L. Kaetterer, *Estimating and representing hydrophobicity potential,* J. Mol. Graphics **6**, 203 (1988).

[19] P. Gaillard, P. A. Carrupt, B. Testa, and A. Boudon, *Molecular lipophilicity potential, a tool in 3D QSAR: Method and applications,* J. Comput. Aided Mol. Des. **8**, 83 (1994).

[20] J. Kittelmann, M. Ottens, and J. Hubbuch, *Robust high-throughput batch screening method in 384-well format with optical in-line resin quantification,* J. Chromatogr. B **988**, 98 (2015).

[21] E. Krieger, G. Koraimann, and G. Vriend, *Increasing the precision of comparative models with YASARA NOVA — a self-parameterizing force field,* Proteins **47**, 393 (2002).

[22] 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**, 1623 (2002).

[23] 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**, 1157 (2004).

[24] J. P. Stewart, *MOPAC: A semiempirical molecular orbital program,* J. Comput. Aided Mol. Des. **4**, 1 (1990).

[25] 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**, 1999 (2003).

[26] D. M. Hawkins, S. C. Basak, and D. Mills, *Assessing model fit by cross-validation,* J. Chem. Inf. Comput. Sci. **43**, 579 (2003).

[27] A. G. Mercader, P. R. Duchowicz, F. Fernandez, E. a. Castro, and F. M. Fernández, *Modified and enhanced replacement method for the selection of molecular descriptors in QSAR and QSPR theories,* Chemometr. Intell. Lab. **92**, 138 (2008).

[28] A. G. Mercader, P. R. Duchowicz, F. M. Fernández, and E. a. Castro, *Replacement method and enhanced replacement method versus the genetic algorithm approach for the selection of molecular descriptors in QSPR/QSAR theories,* J. Chem. Inf. Model. **50**, 1542 (2010).

6

Orientation of monoclonal antibodies in ion-exchange chromatography: A predictive quantitative structure–activity relationship modeling approach

Chromatographic separation of biopharmaceuticals in general and monoclonal antibodies (mAbs) specifically is the bottleneck in terms of cost and throughput in preparative purification. Still, generalized platform processes are used, neglecting molecule specific characteristics, defining protein-resin interaction terms. Currently used in silico modeling approaches do not consider the orientation of the molecule towards the chromatographic resins as a result of the structural features on an atomic level. This paper describes a quantitative structure–activity relationship (QSAR) approach to model the orientation of mAbs on ion exchange chromatographic matrices as a function of property distribution and mobile phase characteristics. 6 mAbs were used to build a predictive QSAR model and to investigate the preferred binding orientations and resulting surface shielding on resins.

This chapter has been published in J. Chromatogr. A 1510 (2017) [1].

Thereby different dominating orientations, caused by composition of F_{ab} fragments of the mAbs, could be identified. The presented methodology is suitable to gain extended insight in molecule orientation on chromatographic resins and to tailor purification strategies based on molecule structure.

6.1. Introduction

Ion-exchange chromatography is a major unit operation in biomolecule separation processes, with anion and cation-exchange chromatography used in monoclonal antibody (mAb) purification. mAb purification processes are often based on platform processes, with the promise of less development efforts, therefore reducing the time-to-market and early stage material consumption [2]. Nevertheless, standardized processes compromise on efficiency as they restrict the design space and rely on unchanged molecule properties within the application space. With the growing number of mAb in development for biopharmaceutical and bioanalytical applications, optimized and scalable processes, which work outside the box of predefined process sequences and materials, are needed [3]. This can be achieved by the use of *in silico* approaches in process development. The application of semi-empirical models to describe chromatographic separation spread wide in the last decades due to increasing fundamental understanding of process mechanistics and adsorption mechanisms. Today, a variety of modeling techniques can be applied to describe the influence of different parameters on ion-exchange chromatography, including mobile-phase composition, resin types, and protein characteristics to different extent [4–6].

These models require the determination of protein and adsorbent specific parameters, and therefore can not be applied to predict the behavior of new molecules ab initio. Physical adsorption is affected by multiple factors as van der Waals forces, electrostatic interaction, or hydrophobic effects, which are defined by molecular properties, chromatographic material and mobile phase composition as pH and ionic strength. Due to their size and structure, mAbs show different characteristics for their subunits. The isoelectric point (IEP), which defines the pH value at which the net charge of the molecule is zero, is generally higher for F_{ab} fragments than for the complete antibody. In consequence, the IEP of F_c fragments is smaller, compared to the full molecule [7–9]. At the IEP, the F_{ab} fragments will therefore carry a positive charge, while the F_c fragment is negatively charged. The dipole moment of antibodies was found to point from the F_c to the $(F_{ab})_2$ fragments which leads to an "end-on"/F_c binding orientation on positively charged surfaces. A "head-on"/F_{ab} orientation was observed for negatively charged surfaces. Further, at low electrostatic forces, where van der Waals interactions dominate, mAb showed a lying flat orientation on surfaces [10].

A shift in electrostatic potential of a mAb can therefore alter the binding orientation on surfaces as chromatographic resins. For example, a pH change from pH 4 to 8 was shown to change the preferred binding orientation of IgG from a "head-on" to an "end-on" orientation [11]. Binding orientation is of interest in development of assays, which use passive binding of IgG onto hydrophobic surfaces as used in enzyme-linked immunosorbent assays (ELISA). Here, an "end-on"/F_c binding would increase capture efficiency. In antibody purification, binding orientation defines key parameter as affinity coefficients and ligand shielding, as lying flat orientations block more space on an adsorbent surface. A higher variance in affinity is expected for

mAb binding in a "head-on"/F_{ab} orientation, compared to an interaction with the F_c fragment.

The orientation of molecules on charged surfaces is subject of investigation via different approaches, balancing the need of information with complexity of the simulation and corresponding computational efforts, which are presented by the size of IgG. Monte Carlo simulations were used to describe protein-surface interactions in various studies for smaller molecules with varying degree of model abstraction [12, 13]. Juffer et al. used explicit atom simulations to improve orientation matching for model proteins [14]. The united-residue model, in which each amino acid is represented as a sphere centered at the ¸-carbon position, allowed for analysis of the effect of amino acid composition on antibody orientation [10, 15]. Brownian dynamics simulation methods or molecular dynamic (MD) simulations, representing single atoms in the molecular structure and with mobile phase described as continuous dielectric media were established for small molecules, such as peptides [16]. This approach was used to investigate interaction energies and orientation of biomolecules in cation-exchange chromatography by Dismer et al. [17–19]. They used an atomistic representation of molecule and the interacting surface including bound ligands. This approach was later extended to complex proteins (i.e. including non-standard residues) and anion-exchange chromatography by Lang et al. [20, 21]. Due to molecule size and an exponential increase in computational cost with number of atoms in simulation, this approach is still prohibitive for application to mAbs in all-atom representation and corresponding adsorber surface. A recently introduced approach to QSAR descriptor calculation, capturing steric hindrance in protein surface interaction as well as electrostatic potential shielding by the mobile phase showed good predictivity regarding mobile phase properties and molecule structures [22]. This approach is extended to describe the binding orientation of mAb on ion-exchange chromatographic resins.

6.2. Materials and methods

Monolconal antibody structures and samples were provided by Lonza Biologics PLC. To provide a diverse data source for model generation further model proteins were included in lab and *in silico* experiments.

6.2.1. Structure preparation

Model proteins were selected based on isoelectric point to ensure binding on cation-exchange resin at chosen conditions and availability of structural information. Corresponding RCSB entries were selected based on most complete structure and high resolution. Table 6.1 lists the chosen proteins, corresponding UniProt IDs and PDB IDs for structural information used for descriptor calculation. Structure preparation was done with the simulation software YASARA [23], which is capable of parameterizing non-standard residues (e.g. bound calcium or a heme-group) via the integrated AutoSmiles algorithm [24–26].

Name, origin	UNIProt ID	PDB	pI
Lysozyme C, hen egg	P00698	1LYZ	9.32
Chymotrypsinogen A, bovine	P00766	2CGA	3.52
α-chymotrypsin, bovine	P00766	1YPH	6.09
Cytochrome C, equine	P00004	1HRC	9.59
Cytochrome C, bovine	P00125	2B4Z	6.50
Ribonuclease A, bovine	P61823	1FS3	8.64
Phospholipase A2, bovine	P00593	1BP2	6.10
Myoglobin, equine	P68082	2V1F	7.36

Table 6.1: Proteins used in QSAR model generation and evaluation with their corresponding UniProt ID, PDB ID, and theoretical pIs as calculated according to Bjellqvist et al. [28, 29] selected for descriptor calculation.

Loaded structures were cleaned from water and substitutes, a simulation box around the molecular structure was defined, and the corresponding pH value was set. The structure was protonated according to set pH value. To correct the covalent geometry, an energy minimization was performed with the AMBER99 force field [27], using a 7.86 Å cut-off. After an initial steepest descent minimization, a simulated annealing simulation was performed to obtain an optimized structure. Partial charges of molecule atoms were derived based on the AMBER99 force field and saved along with molecule information in pqr file format, which captures atomic charges and radii information.

6.2.2. Descriptor calculation

To represent the orientation of a biomolecule towards an adsorbent surface, three types of descriptors were calculated.

Projection plane.

To capture the shielding of electrostatic potential and steric hindrance, 120 theoretical planes were positioned around the molecule in an equidistant manner with a set distance of 5 Å towards the molecule surface. The plane-molecule distance was selected according to [19, 20, 22]. Electrostatic potential was then projected onto the plane. Descriptor values were calculated based on plane values.

Surface patch.

To capture orientation sensitive interaction potential of heterogeneous molecules and short range hydrophic interactions, surface patches were calculated for hydrophobic and electrostatic properties. Surface patch sizes and orientation were calculated according to [22]. From the 120 different orientations calculated, the orientation with the highest electrostatic potential density as projected on a plane was selected to be included in descriptor calculation. Further, the orientation with highest hydrophobic potential based on a solvent accessible surface area (SASA) patch were included in QSAR model calculation, as an alternative orientation driven by hydrophobic interactions for high ionic strength conditions of the mobile phase.

Descriptor	Definition
totalSurf	surface area of the projection (Å)
nSurfP	number of surface points
sum	sum of mapped property
mean	mean value of mapped property
median	median value of mapped property
meanRes	mean value corrected for resolution of surface points
max	maximum value of mapped property
min	minimum value of mapped property
dev	value deviation
var	variance of property
nPos	number of points with positive value
nNeg	number of points with negative value
relPos	fraction of points with positive value
relNeg	fraction of points with negative value
averPos	average value of points with positive value
averNeg	average value of points with negative value
bin0 - bin9	number of points with mapped property value in the range of the bin

Table 6.2: Projection descriptor set with names and description of calculation. Descriptors are calculated based on plane projection, surface patch and the full molecular surface for electrostatic potential(EP) and hydrophobicity characteristics (hyd) as described in [22].

Full surface.

The complete surface of the molecule was considered in descriptor calculation for a non-orientation sensitive approach for electrostatic and hydrophobic properties. The descriptors calculated for the three different projection types and for hydrophobicity and electrostatic potential are listed in Table 6.2. In total, 120 different descriptors were considered per experimental condition.

A detailed description of electrostatic potential and hydrophobicity property calculation and their projection towards molecule surface and projection planes including applied shielding factors is given in [22].

6.2.3. QSAR modeling

The QSAR model was built as an ensemble of single regression models with an internal cross validation. Thereby, the dataset was repeatedly split randomly into a training set (containing 80 % of protein data) and an internal test set (20 %). 100 regression models were generated and those with an r^2 above 0.8 and an predictive r^2 above 0.6 for the internal test set were subsequently pooled within the ensemble. This approach proved to be robust against over-fitting of single models and the influence of outliers, especially in small datasets [30]. An implementation of the enhanced replacement method (ERM) published by Mercader et al. [31] was used for single model generation. The algorithm follows a simulated annealing approach, which prevents sub-optimal results by local optima. It has been shown that the method is performing better than regression and partial least square (PLS) models

and is comparable to genetic algorithms (GA), without the need to set parameters beforehand [32]. Additionally, an external test set of 4 conditions (2 mAbs at 60 and 90 mM ionic strength) was split from the dataset prior to model development to demonstrate the predictive capabilities of the model.

6.2.4. Batch isotherm experiments

Batch isotherm experiments were performed for QSAR model generation and evaluation with an automated high-throughput screening process. This process is performed in a 384-well microtiter plate and based on optical quantification of resin particles, which amount to 1-3 µL volume per well. The concentration of protein bound on the resin is calculated based on protein concentration differences in the supernatant before and after a 60 min incubation phase. This process is described in more detail in [33]. Batch binding experiments of mAb were conducted at pH 5.0 and ionic strengths of 60 and 90 mM, adjusted with sodium chloride. Model proteins used are listed in Table 6.1 and were purchased from Sigma Aldrich (St. Louis, MO, USA). Experiments were performed in the range of pH 5.0-11.0 and solvent ionic capacities of 30-120 mM. All buffer components and salts were purchased from Merck (Darmstadt, Germany). The adsorbent resin SP Sepharose FF was obtained from GE Healthcare (Uppsala, Sweden).

Resulting isotherms were fitted with the Langmuir isotherm equation:

$$q = q_{max} \cdot k_a \frac{c}{1 + k_a \cdot c} \tag{6.1}$$

where q and c are equilibrium protein concentrations in the stationary, respectively mobile phase, q_{max} the maximum adsorption capacity of the solid phase and k_a the adsorption affinity coefficient. Nonbinding conditions and conditions with $k_a < 0.5$ values were omitted in model building and analysis as those conditions have shown to lead to erroneous binding parameters [33].

6.3. Results and discussion

6.3.1. Antibody orientation

Monoclonal antibodies show a variation of properties over their accessible surface, due to their complex structure and shape, as compared to small proteins and peptides. To capture these distributions and to identify interaction orientations, the electrostatic potential of mAbs were mapped as described in Section 6.2.2.

Electrostatic potentials were mapped in Fig. 6.1 and orientations of interest are displayed in Fig. 6.2. Two antibodies are compared in regard to their electrostatic potential distribution at pH 5.0, representing the diversity of mapping results found for antibodies investigated.
Two major different patterns of preferred binding orientations, as indicated by high averaged electrostatic potential, to an interaction plane were found. mAb1 shows the expected ”head-on” orientation in interaction with negatively charged surfaces.

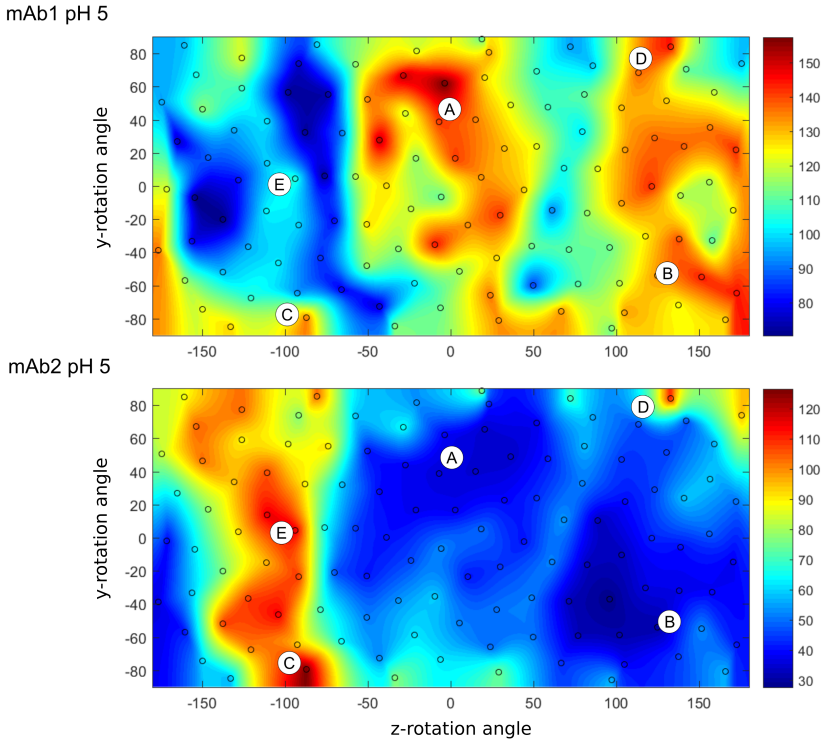

Figure 6.1: Representation of molecular electrostatic potential on adsorbent plane. Highest electrostatic potential represents preferred binding orientations of mAb on a cation exchange surface. Orientations A and B indicate "head-on" orientations preferred at conditions of high electrostatic interaction. Orientations C and D show the two opposing "flat-on" surface orientations, as found to be preferred by mAb2. Position E corresponds to the "end-on" orientation.

The highest potential values are found in an upright position (Fig. 6.1 and 6.2 positions A and B). Electrostatic potential values found for both F_{ab} fragments are comparable to each other, as expected in an identical 3D structure of the fragments. The flat-on orientation also yielded high interaction terms (postions C and D), although not reaching the values found in "head-on" conformation. The "end-on" or F_c fragment based binding orientation (E) is found to be not preferable for mAb1 at investigated conditions.

mAb2 shows a different pattern with low electrostatic potential in the F_{ab} fragment based binding orientation, which is attributed to differences in amino acid composition (see Table 6.3). The preferred orientations here are C and D, where the antibody is lying flat on the surface. Mapped parameters are comparable between mAb1 and mAb2 for F_c fragment based and flat-on orientation. This is expected as the F_c fragments are identical between antibodies and the influence of the active domain in the F_{ab} fragments is low in the flat-on orientation.

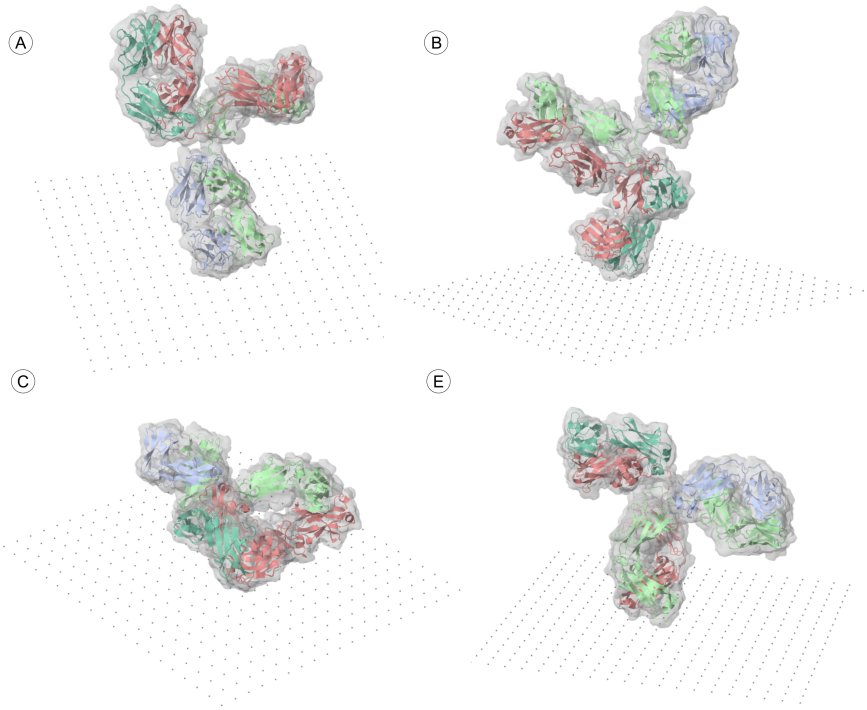

Figure 6.2: MAb orientation to surface plane. Orientations A and B show ”head-on” orientation preferred at conditions of high electrostatic interaction. Orientation C shows a ”flat-on” surface orientation, as found to be preferred by mAb2 and low electrostatic potential, indicating an increase in hydrophobic interactions.

The differences in the functional domain of F_{ab} fragments in monoclonal antibodies lead to a shift in binding strength and orientation considering the electrostatic potential mapping to a charged surface representation. The number and position of charged amino acids defines the electrostatic potential of the macromolecule in the binding domain. Herein, the addition or depletion of charges, or even a change of neutral amino acids influencing steric hindrance in interaction lead to a shift in interaction potential and molecule orientation.

It is assumed that within a population of mAbs only differencing in the active domain, a change in orientation is an on-off event occurring with the electrostatic potential presented by the active domain dropping below the interaction term of the flat-on orientation. Once a flat orientation is preferred, the composition of the F_{ab} fragment is assumed to have only small influence on the interaction. With increasing ionic concentration in the solvent and resulting electrostatic shielding effects,

structure	charge	lys/arg	asp/glu
mAb1	8.547	94/30	59/52
mAb2	6.497	87/40	73/48

Table 6.3: Net charge and charged amino acid count in mAb1 and mAb2 at pH 5.0.

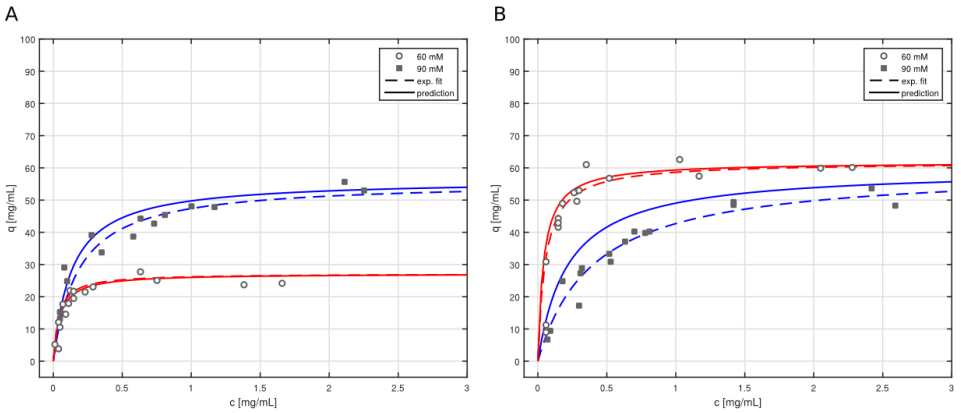

Figure 6.3: Langmuir batch binding iostherms of A) mAb1 and B) mAb2 on SP Sepharose FF. Predicted Langmuir affinity coefficients k_a (dashed lines) correlate with experimental values. Different maximum binding capacity indicates a shift in mAb1 orientation between "head-on" with q_{max} of 40 g L^{-1} and "flat-on" with q_{max} of 60 g L^{-1} at low ionic strength (60 mM) indicating a reduced resin pore clogging for mAb2 and for both mAb at higher ionic strength.

the contribution of short ranged hydrophobic interactions increases. This leads to a flat-on orientation, increasing the molecule to surface interaction area.

The QSAR approach uses a fixed molecular structure to calculate descriptors. Therefore, conformational changes induced by the interaction are not considered to full extent. A binding conformation including both variable regions of the F_{ab} fragments is possible in a head-on orientation. This increases the overall electrostatic potential involved, although has minor influence on the electrostatic potential density, as the surface area blocked by the second F_{ab} is considered as well.

The Langmuir adsorption affinity coefficient k_a, together with the maximum adsorption capacity q_{max} provides the means to model retention behavior of the molecule on different column scales at same conditions [34, 35]. Therefore, a QSAR model predicting molecule parameters for varying pH and ionic strength conditions allows for an *in silico* screening of retention and separation behavior on chromatographic columns. Fig. 6.3 shows the Langmuir isotherms for mAb1 (A) and mAb2 (B) batch binding experiments on SP Sepharose FF resin at pH 5.0 and ionic strength of 60 and 90 mM in the mobile phase.

6.3.2. Adsorption affinity coefficient k_a

At 60 mM, both mAbs show higher affinity coefficients than at 90 mM ionic strength. This is expected as increasing ionic strengths shield the electrostatic interactions between molecule and ligands and therefore results in lower affinity. Addinity coefficients at 90 mM are comparable between the investigated mAbs. This is consistent with a "flat-on" orientation being preferred at higher ionic strength (see Fig. 6.1) where the impact of the differently charged F_{ab} regions decreases due to the shielding effect of the solvent compared to the identical backbone of the mAbs. This means that the mAb differences have lesser influence on interaction behavior, with regards to affinity.

6.3.3. Maximum adsorption capacity q_{max}

A strong variance can be observed in adsorption capacity at 60 mM ionic strength between the molecules. While mAb1 reaches a q_{max} of 25 mg mL^{-1}, mAb2 yields double the amount with 56 mg mL^{-1}. The difference in adsorption capacity confirms the modeling results (see Fig. 6.1), with mAb1 showing the preferred binding orientation in a "head-on", upright position, compared to mAb2 binding in a "flat-on" orientation to the resin. The significant difference in adsorption capacity can be attributed to different molecule-resin interaction aspects:

- A change in orientation alters the coverage of the resin surface, where a "flat-on" orientation covers more interaction sites compared to a "head-on" orientation, thus leading to a higher binding capacity in a "flat-on" orientation.
- However, it needs to be assumed that in a "head-on" orientation, molecule-resin interaction is stronger sterically hindered as bound molecules reach further into the resin pore volume, hereby decreasing pore diameter. This ultimately leads to more molecule-molecule interactions.

Decreased adsorption capacity for strong interaction scenarios of mAbs in ion exchange chromatography are well described in literature and often attributed to increased pore clogging. Molecules are binding at pore entrances, leading to a decreased availability of pore volume due to electrostatic repulsion by bound molecules and steric hindrance. An increase of ionic strength leads to a stronger charge shielding, which allows for a more efficient molecule transport into the pores and therefore a higher adsorption [36, 37]. This is described to be more pronounced with increasing ligand length and density [38]. The modeling results of different orientations at 60 mM for mAb1 and mAb2, indicates that the molecule orientation has an impact on mentioned behavior. The upright orientation of mAb1 at low ionic strength favors clogging of pores by stronger binding to the resin and steric hindrance by the molecule volume. In contrast, the "flat-on" orientation and weaker binding of mAb2 allows for more molecules to reach into the pore.

6.3.4. QSAR model

The QSAR model correlates experimentally determined molecular properties to descriptors. A model which proves to be predictive, can further be used to obtain

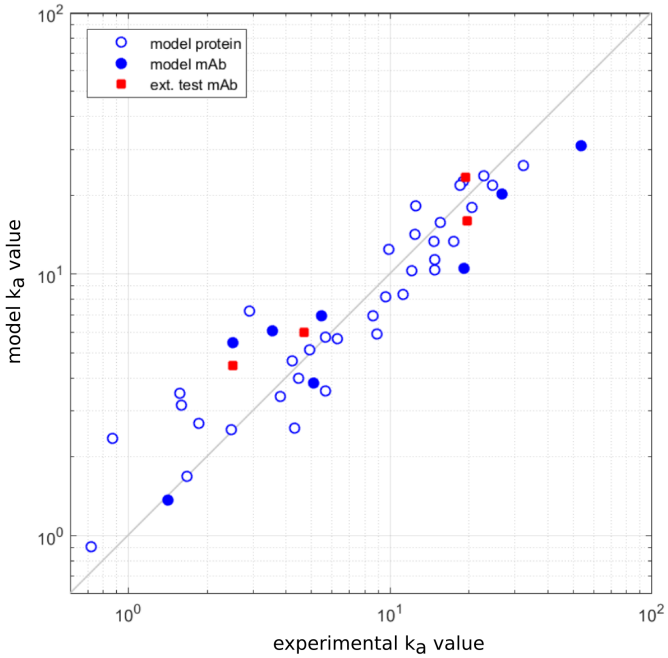

Figure 6.4: QSAR model for binding of mAb and model proteins on SP Sepharose FF, represented as Langmuir adsorption affinity coefficients k_a. The model includes varying mobile phase conditions in the pH range of 5-9 and an ionic strength of 30-120 mM for model proteins (open circle). Binding conditions for mAbs were pH 5 with mobile phase ionic strengths of 60 and 90 mM (closed circle). New structures in the external test set were predicted successfully (squares) with variation in range of experimental data accuracy.

molecular activities for experimental conditions and molecular structures the model was not trained on. A QSAR model correlating the affinity of monoclonal antibodies and model proteins to a cation exchange adsorbent under varying conditions of pH and ionic strength of the mobile phase was calculated for 42 conditions. The model was then used to predict the adsorption affinity coefficient k_a for two mAbs not included in the training set. The resulting model, shown in Fig. 6.4, shows a high correlation between experimentally determined and modeled k_a values, with a r^2 value of 0.92 and a predictive r^2 value of 0.86.

The accuracy of the model is within the means of the experimental data variation, which is discussed in detail in [33]. In that study it is described that low k_a values are underestimated in the fit, which is in agreement with the deviation between experimental and predicted results for k_a values less than 1.0, where in average higher values were predicted.

6.4. Conclusion and outlook

Mapping of molecule electrostatic properties confirm a preferred "head-on" orientation of mAb1 on negatively charged surfaces and a potential orientation change to a flat-on orientation with increasing shielding by solvent ions. It is also shown, that this does not hold true for all mAbs, as differences in the active domain of the F_{ab} fragment can lead to a flat-on orientation at high electrostatic potential conditions, as shown with mAb2 at 60 mM compared to 90 mM ionic strength conditions.

The switch of preferred orientations has a major impact on surface coverage and maximum binding capacity and therefore represents a property cliff, in experiment as well as in modeling, which needs to be considered in process design. This also shows that a generic platform approach for one mAb is not ideal for all mAbs, as chromatographic interaction properties can change significantly, based on small changes in the part of the molecule interacting with a resin surface.

The presented approach of orientation sensitive descriptors is shown to be capable of describing the change in electrostatic potential patterns and resulting change in preferred orientation on atomic level. Further, a predictive model can be built and be supported by a diverse set of model proteins. Thereby reducing experimental efforts and model training on mAb, of which availability is not only sparse in early process design but also expensive. The described methodology can already be utilized in protein design to forecast downstreaming capabilities. This work provides the grounds for further research on various resins in ion exchange chromatography as well as on mixed mode adsorbents to describe the difference and shift of orientation from electrostatic dominated binding conditions to hydrophobic interactions.

6.5. Acknowledgments

Financial support and monoclonal antibodies for this work were kindly provided by Lonza Biologics, PLC.

References

[1] J. Kittelmann, K. M. Lang, M. Ottens, and J. Hubbuch, *Orientation of monoclonal antibodies in ion-exchange chromatography: A predictive quantitative structure–activity relationship modeling approach,* J. Chromatogr. A **1510**, 33 (2017).

[2] R. Hahn, R. Schlegel, and A. Jungbauer, *Comparison of protein A affinity sorbents,* J. Chromatogr. B **790**, 35 (2003).

[3] D. Low, R. O'Leary, and N. S. Pujar, *Future of antibody purification,* J. Chromatogr. B **848**, 48 (2007).

[4] 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. **283**, 37 (1984).

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

[6] J. C. Bosma and J. A. Wesselingh, *pH dependence of ion-exchange equilibrium of proteins,* AIChE J. **44**, 2399 (1998).

[7] J. Buijs, W. Norde, and J. W. T. Lichtenbelt, *Changes in the secondary structure of adsorbed igg and f(ab')2 studied by ftir spectroscopy,* Langmuir **12**, 1605 (1996).

[8] J. Buijs, P. A. van den Berg, J. W. Lichtenbelt, W. Norde, and J. Lyklema, *Adsorption dynamics of igg and its f(ab')2and fc fragments studied by reflectometry,* J. Colloid Interf. Sci. **178**, 594 (1996).

[9] F. Fogolari, R. Ugolini, H. Molinari, P. Viglino, and G. Esposito, *Simulation of electrostatic effects in Fab–antigen complex formation,* Eur. J. Biochem. **267**, 4861 (2000).

[10] J. Zhou, H.-K. Tsao, Y.-J. Sheng, and S. Jiang, *Monte Carlo simulations of antibody adsorption and orientation on charged surfaces.* J. Chem. Phys. **121**, 1050 (2004).

[11] L. Dávalos-Pantoja, J. Ortega-Vinuesa, D. Bastos-González, and R. Hidalgo-Álvarez, *A comparative study between the adsorption of IgY and IgG on latex particles,* J. Biomater. Sci. Polym. Ed. **11**, 657 (2000).

[12] V. Zhdanov and B. Kasemo, *Monte carlo simulation of the kinetics of protein adsorption,* Proteins: Struct., Funct., Bioinf. **30**, 177 (1998).

[13] V. Castells, S. Yang, and P. R. Van Tassel, *Surface-induced conformational changes in lattice model proteins by Monte Carlo simulation,* Phys. Rev. E **65**, 031912 (2002).

[14] A. H. Juffer, P. Argos, and J. de Vlieg, *Adsorption of proteins onto charged surfaces: A Monte Carlo approach with explicit ions,* J. Comput. Chem. **17**, 1783 (1996).

[15] J. Zhou, S. Chen, and S. Jiang, *Orientation of Adsorbed Antibodies on Charged Surfaces by Computer Simulation Based on a United-Residue Model,* Langmuir **19**, 3472 (2003).

[16] A. M. Bujnowski and W. G. Pitt, *Water structure around enkephalin near a {PE} surface: A molecular dynamics study,* J. Colloid Interf. Sci. **203**, 47 (1998).

[17] F. Dismer and J. Hubbuch, *A novel approach to characterize the binding orientation of lysozyme on ion-exchange resins,* J. Chromatogr. A **1149**, 312 (2007).

[18] 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**, 11 (2008).

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

[20] K. M. 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 (2015).

[21] K. M. Lang, J. Kittelmann, F. Pilgram, A. Osberghaus, and J. Hubbuch, *Custom-tailored adsorbers: A molecular dynamics study on optimal design of ion exchange chromatography material,* J. Chromatogr. A **1413**, 60 (2015).

[22] J. Kittelmann, K. M. Lang, M. Ottens, and J. Hubbuch, *An orientation sensitive approach in biomolecule interaction QSAR and application in ion-exchange chromatography,* J. Chromatogr. A **1482**, 48 (2017).

[23] E. Krieger, G. Koraimann, and G. Vriend, *Increasing the precision of comparative models with YASARA NOVA — a self-parameterizing force field,* Proteins **47**, 393 (2002).

[24] 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**, 1623 (2002).

[25] 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**, 1157 (2004).

[26] J. P. Stewart, *MOPAC: A semiempirical molecular orbital program,* J. Comput. Aided Mol. Des. **4**, 1 (1990).

[27] J. Chen, T. Yang, Q. Luo, C. M. C. Breneman, and S. M. S. Cramer, *Investigation of protein retention in hydrophobic interaction chromatographic (HIC) systems using the preferential interaction theory and quantitative structure property relationship models,* React. Funct. Polym. **67**, 1561 (2007).

[28] B. Bjellqvist, G. J. Hughes, C. Pasquali, N. Paquet, F. Ravier, J.-C. Sanchez, S. Frutiger, and D. Hochstrasser, *The focusing positions of polypeptides in immobilized ph gradients can be predicted from their amino acid sequences,* Electrophoresis **14**, 1023 (1993).

[29] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch, *Protein identification and analysis tools on the expasy server,* in *The Proteomics Protocols Handbook*, edited by J. M. Walker (Humana Press, Totowa, NJ, 2005) pp. 571–607.

[30] D. M. Hawkins, S. C. Basak, and D. Mills, *Assessing model fit by cross-validation,* J. Chem. Inf. Comput. Sci. **43**, 579 (2003).

[31] A. G. Mercader, P. R. Duchowicz, F. Fernandez, E. a. Castro, and F. M. Fernández, *Modified and enhanced replacement method for the selection of molecular descriptors in QSAR and QSPR theories,* Chemometr. Intell. Lab. **92**, 138 (2008).

[32] A. G. Mercader, P. R. Duchowicz, F. M. Fernández, and E. a. Castro, *Replacement method and enhanced replacement method versus the genetic algorithm approach for the selection of molecular descriptors in QSPR/QSAR theories,* J. Chem. Inf. Model. **50**, 1542 (2010).

[33] J. Kittelmann, M. Ottens, and J. Hubbuch, *Robust high-throughput batch screening method in 384-well format with optical in-line resin quantification,* J. Chromatogr. B **988**, 98 (2015).

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

[35] B. K. B. Nfor, M. Noverraz, S. Chilamkurthi, P. D. E. M. Verhaert, L. a. M. van der Wielen, and M. Ottens, *High-throughput isotherm determination and thermodynamic modeling of protein adsorption on mixed mode adsorbents,* J. Chromatogr. A **1217**, 6829 (2010).

[36] C. Harinarayan, J. Mueller, A. Ljunglöf, R. Fahrner, J. Van Alstine, and R. van Reis, *An exclusion mechanism in ion exchange chromatography,* Biotechnol. Bioeng. **95**, 775 (2006).

[37] A. Ljunglöf, K. M. Lacki, J. Mueller, C. Harinarayan, R. van Reis, R. Fahrner, and J. M. Van Alstine, *Ion exchange chromatography of antibody fragments,* Biotechnol. Bioeng. **96**, 515 (2007).

[38] K. Wrzosek, M. Gramblicka, and M. Polakovic, *Influence of ligand density on antibody binding capacity of cation-exchange adsorbents,* J. Chromatogr. A **1216**, 5039 (2009).

7

mantoQSAR: A graphical user interface driven software for molecular orientation sensitive QSAR modeling in downstream process development

Quantitative Structure-Activity Relationship (QSAR) modeling of biopharmaceuticals has been and is a fast growing research topic within downstream process development. As an application in an industrial development setting QSAR is compared to alternative methods, with regard to costs and time to application. This opens the need for a standardized and easy-to-use software platform which allows users in academia and industry to focus on model results, rather then structure preparation and descriptor calculation and provides an easy way of data and model transfer between institutions and research fields. With mantoQSAR we developed an easy to use QSAR modeling software, which gives users a graphical user interface to compose QSAR models and investigate molecular structures without the need to program descriptors or modeling algorithms. mantoQSAR is a JAVA framework which can be run on computers without the need to share molecular structures with third parties. This approach will provide research and industry a mutual platform to gain comparable results and reach short development cycles in generating predictive QSAR models for biomolecule processing.

7.1. Introduction

Quantitative Structure-Activity Relationship (QSAR) modeling is gaining increasing focus in modeling of biomolecule interactions in downstream process development [1]. With the development of miniaturized screening applications, which use a high degree of parallelization, the amount of data available is increasing in an exponential fashion. This leads to faster development cycles, while reducing the amount of early stage material necessary for process development.

Parallel to the amount of experimental data, knowledge on biomolecule structure is growing rapidly with more structural information being available in the public domain. Databases as UniProt (http://www.uniprot.org/) [2], GenBank [3], and Worldwide Protein Data Bank (www.wwpdb.org) [4] are now holding over 100.000 three-dimensional structures with new information being added daily. The number is increasing with the advances in homology modeling and *ab initio* prediction of 3D molecule formations [5] and establishment of corresponding tools [6–8]. Molecular structure information provides the foundation for data mining studies, aiming at new insights in molecule characteristics and parameters in downstream processing applications [9].

Quantitative structure-activity relationship modeling aims at correlating molecule properties to descriptors which are derived from the structure of the molecule [10]. With QSAR originating from chemometrics and activity prediction for small molecules, there are well established numerical features describing structure attributes for small molecules and peptides. Over 3000 distinct descriptors are available to be calculated for small molecules, with the most established tools and descriptor sets being DRAGON [11], TOMOCOMD-CARRD [12], PADEL [13], CDK descriptor calculator [14], ADRIANA CODE [15], CODESSA-PRO and CERIUS [16, 17]. Being developed for small molecules, most of these descriptors do not take the three dimensional structure of the molecule into account, which has a large impact on the characteristics of native proteins. Descriptors considering conformation, are mostly established for applications in protein classification, similarity analysis and function prediction, with regard to active site identification and interaction with cofactors. [18]. Therefore, most of the 3D-structural descriptors published to date typically capture information relating to the protein structure and are largely used to show correlation of folding properties as folding rate constant [19–21].

With the development of QSAR models in the domain of biomolecule downstream processing, new sets of descriptors have been developed. Mazza et al. [22] introduced an approach to map electrostatic potential onto the surface of a molecule to derive pH dependent descriptors. These were used to predict retention times for a variety of betablockers and similar chemical structures previously published by Law and Weir [23]. Malmquist et al. [24] developed a set of pH sensitive descriptors, based on the surface mapping of atomic charges and applied them successfully in cation and anion-exchange chromatography modeling of proteins. Based on this work, Yang et al. [25] developed a set of electrostatic potential de-

scriptors, and applied them in SMA parameter modeling, wherein the protein in each pH condition was treated as a separate molecule. All approaches have in common that they do not reflect the interaction of molecules with the containing phase. We recently presented a new set of descriptors, which shifted the domain from describing molecule properties, to capturing molecule interaction potentials with the surrounding phase. Therein, we consider different shielding of molecule properties based on liquid phase conditions. This approach proved successful in modeling biomolecule solubility and precipitation. Further, the interaction with a solid phase as present in solid-liquid chromatography is modeled to capture steric hindrance in the binding of amorph biomolecules. Descriptor sets derived from these approaches were successfully used in a series of studies to generate predictive models in ion-exchange chromatography for a range of model proteins and antibodies, as well as in prediction of solubility parameters [26–28]. While modeling antibody binding to a cation-exchange chromatographic resin, it has been shown that binding orientation variation between different mAbs were described in the descriptor set and resulting model [27] (see Chapter 6).

Most present applications for descriptor calculation are realized as a web-service tool. This requires the user to upload a molecule structure and have necessary calculations being carried out on a remote server. The user has the advantage of ease of use, as he does not need to maintain a computational set-up while avoiding potential software dependencies or conflicts. Nevertheless, in biopharmaceutical research, target molecules are often considered intellectual property, therefore restricting the use of third party web servers. This renders these tools useless, or requires an extensive effort for integration in internal server-client structures. A software to be installed on the user's hardware circumvents these restrictions, as presented here. In this work, we present mantoQSAR, a software, which calculates projection dependent molecule descriptors. The Java based software provides management of QSAR projects and corresponding molecular structure files, descriptors, and model calculations.

7.2. Implementation

mantoQSAR provides a graphical user interface (GUI) (see Fig. 7.1) which generates descriptors and predictive models for proteins. The program accepts two different input formats for molecular structures: PQR and PDB files. Loaded structures are organized in modeling projects, where corresponding experimental data can be added and descriptor calculation parameters be set. Resulting descriptor values can be viewed within the software as well as exported in generic formats, like plain text and Excel tables. Further, descriptor value distribution and projection orientations can be visualized in a dynamic three dimensional representation. The software is implemented in JAVA (JDK 1.7) as this provides cross-platform support for any system. The Jmol software (version 14.2) is used as a 3d party library for structure and descriptor visualization.

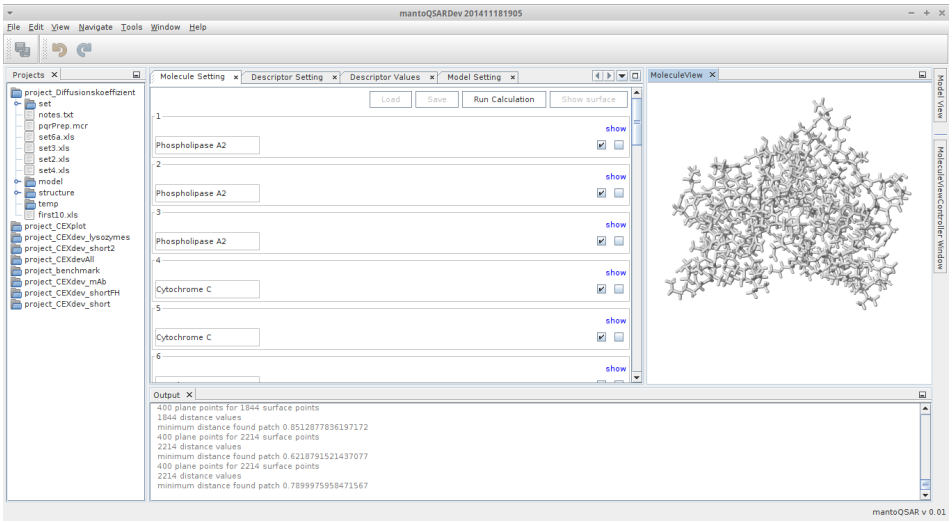

Figure 7.1: mantoQSAR window displaying structures loaded within the active project (center window), a project overview (left side) for easy change of active projects and interactive display of the currently selected biomolecule structure (right side).

7.2.1. Molecular descriptors

Molecule characteristics A successful biomolecule downstream processing strategy uses different biomolecule characteristics in the purification process. Electrostatic potential and hydrophobicity characteristics are key parameters in the processing of molecules and are therefore implemented with mantoQSAR and have been proven successful in describing molecule properties [26, 28]. The molecular electrostatic potential calculation is based on single atom charges provided with structure files to the mantoQSAR software. Hydrophobicity characteristics are based on single amino acids and the parameter values published by Kyte and Doolittle [29]. Further parameter sets can be easily added to descriptor calculation by the user.

Surface Surface based descriptors are calculated for a grid of equidistant points on the surface of the molecule as shown in Fig. 7.2.A. Surface herein is defined as solvent accessible surface area (SASA), which is defined by the center of a probe, which is rotated over the outermost atoms of the molecule. While the radius of the probe is usually set to 1.4 Å, representing a water molecule, other sizes can be set.

Surface patch To capture properties on the molecular surface sensitive to the orientation of the molecule towards an interacting surface, surface patches can be calculated. Herein surface patches are defined to include solvent accessible surface areas below a defined distance towards the corresponding plane projection. The size of the patches can be adjusted by the maximum distance in which molecule surface points are considered. Herein, the distance is defined along the normal

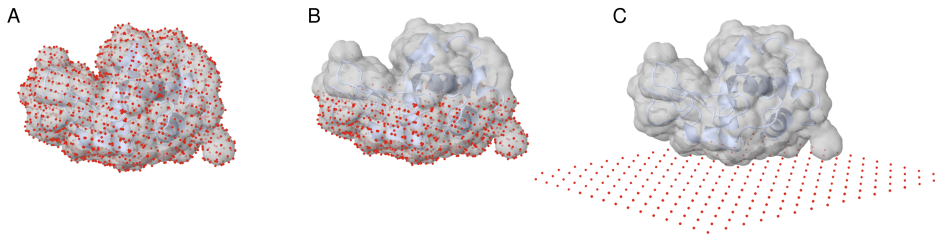

Figure 7.2: Parameter projection types applied. A) Molecular properties are projected to the solvent accessible surface of the molecule, which is represented by distinct points with a density of 1 point per $Å^2$, B) In patch orientation only a part of the surface is considered in descriptor calculation to represent an oriented interaction between molecule and solid phase during adsorption, and C) in plane orientation molecule electrostatic properties are projected towards a theoretical plane which is placed in a fixed distance to the molecule, thereby reflecting the steric hindrance of charges by the 3 dimensional structure of the molecule. Orientation sensitive projection types are calculated for equally distributed orientations and the orientation with the strongest interaction is included in the QSAR model generation.

vector of the plane as shown in Fig. 7.2.B. A variety of descriptors can be calculated based on the different projection matrices. Descriptors and their incentive in capturing molecule interaction are described in [26].

Plane projection To represent different orientations of a molecule in a directed interaction, a representation of a surface plane was constructed as a grid of points in three dimensional space, which is shown in Fig. 7.2.C. To represent all possible orientations of the molecule towards a surface, the molecule is approximated as a sphere with the sphere center positioned on the geometric molecule center. Orientation vectors are calculated based on the Thomson equilibrium [30] with the vector origin placed in the center of the molecule, yielding equidistantly distributed orientations. For each vector a plane of grid points with defined size and density is calculated and positioned with the vector as normal vector in the plane center. The plane is further shifted along the normal vector to fit the distance to the molecular surface. Molecular property values are then projected from the molecule surface or atomic/amino acid position towards the grid points of the plane.

Sphere projection The sphere projection combines the calculated descriptor values for all plane orientations, therefore giving a representation of molecule properties in a defined distance to the surface of the molecule. These descriptors have been found to be advantageous in a series of applications including diffusion coefficient [31] and protein precipitation modeling [28].

Projection algorithms While being projected from the molecule to a point of interaction, intermolecular forces decrease in strength. This is defined by the distance between the source and target of projection as well as screening by the liquid phase. mantoQSAR implements a variety of projection algorithms described in literature. Those include but are not limited to Coulomb's law distance functions and

electrostatic screening functions described by Israelachvili [32] as well as functions applicable to hydrophobicity mapping including those described by Fauchére [33] and Audry [34] as well as modifications thereof [35].

7.2.2. Predictive models

The modeling algorithm establishes a correlation between single descriptors and the molecule activity to be modeled. A wide variety of modeling approaches exist, from different regression modeling and partial least square (PLS) approaches to evolutionary algorithms in broad and genetic algorithms (GA) specifically. The mantoQSAR software provides an enhanced replacement method (ERM) algorithm which was presented by Mercader et al. [36].

This algorithm follows a simulated annealing approach, which prevents suboptimal results by local optima. It has been shown, that this method is performing better than regression and PLS models and comparable to genetic algorithms for a variety of tasks [37]. It is considered preferable to GA as it does not require algorithm parameters to be set prior to use.

Further, ensemble modeling with a variable number of submodels is supported as depicted in Figure 7.3. Hereby, the dataset is repeatedly split randomly into a training and an internal test set. The final model combines all submodels reaching a set threshold in r^2 for both data sets. All parameters regarding number of submodels, submodel complexity, data split ratios, and quality parameters can be adjusted to model needs and data quality on hand. This approach is proven to provide robustness against overfitting of single models, leading to low predictive capabilities. Also, the influence of single outliers in the data set is omitted, as poorly performing submodels are not introduced in the model ensemble. This is of special importance when small data sets are used [38].

7.2.3. Modeling workflow

A new QSAR model can be built by defining a new project or copying an existing project. Molecules are imported into the project folder via drag and drop of the corresponding structure file in PQR format from another project or folder. Hereby, structure files are copied to the project, to reduce dependencies between single projects. The molecule setting window provides an overview of all molecules in the current project and lets the user define further molecule properties and experimental conditions as pH and ionic strength of the mobile phase. Single molecules can be activated or deactivated as well as set to be included in the external testset for model validation. The molecule visualization window allows the user to confirm the identity and integrity of the imported structure.
In the descriptor setting window, all descriptor parameters can be set. Multiple descriptor sets can be defined and changed, as mantoQSAR allows detailed control of all projection parameters. This includes the selection of molecule properties, sur-

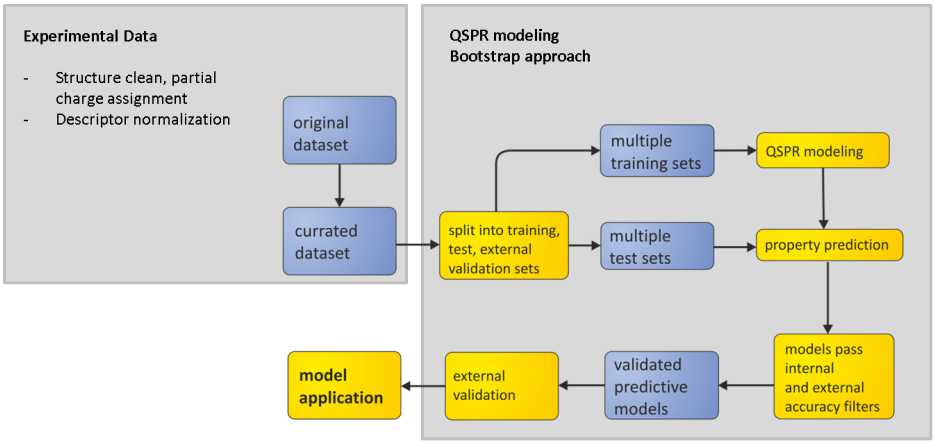

Figure 7.3: Bootstrap modeling approach in mantoQSAR software. Observations are split multiple times in internal test and training sets. Internal models are validated against internal test sets. Models meeting set requirements on r^2 and predictive r^2 are pooled to a final model for application on external test sets and subsequent clearance for predicting new structure properties.

face definitions, projection type and parameters. Predefined parameters provide a basic setup, which has been proven to be successful in QSAR modeling of different biomolecule parameters [27, 28, 31].
The descriptor values window provides an overview of calculated descriptors and progress during calculation. Descriptor values can be exported to txt and xlsx format for use in third party applications. Further, descriptor values can be selected by the user to display them on the molecule structure. Hereby, descriptor projection mode and grid are displayed together with descriptor value distribution. This allows the user to visually check preferred orientations calculated and the mapping of descriptor values.
The model setting window provides the interface to define a single ERM model and parameters as well as a model ensemble. During model calculation a dynamic feedback is given on the inclusion of single models in the final ensemble, ensemble parameters, and model responses for training and test data sets.

7.3. Conclusion and outlook

mantoQSAR is a QSAR modeling software package with an extensive graphical user interface for modeling and visually accessing biomolecule properties. The software provides an extensive set of descriptors and projection options for modeling of biomolecule interactions with focus on directed interactions. The platform can easily be extended by the user, while a common standard in QSAR project handling and exchange between different research groups and industry is maintained. Established predictive models can be extended with new observations as work progresses or tested on new molecule structures and conditions.

With mantoQSAR as a modeling framework in research, a first QSAR model can be established within hours and development can be focused on areas of interest, regarding molecule and descriptor selection, programming of new descriptors or modeling algorithms.

7.3.1. Outlook

The mantoQSAR software is an active project in development under GLP license which can be altered and further developed by its users as needed. Current development activities include the integration of further modeling algorithms and import as well as export formats for a seamless integration within existing modeling workflows.

Availability and requirements

Project name: mantoQSAR
Project home page: mantoQSAR.com
Operating system(s): Platform independent
Programming language: Java
Other requirements: JDK-7 or higher
License: GNU GPL

Acknowledgement

We would like thank Frank Hämmerling for ongoing user tests and constructive input in the development of mantoQSAR.

References

[1] A. T. Hanke and M. Ottens, *Purifying biopharmaceuticals: knowledge-based chromatographic process development,* Trends Biotechnol **32**, 210 (2014).

[2] T. U. Consortium, *Activities at the Universal Protein Resource (UniProt),* Nucleic Acids Res. **42**, D191 (2014).

[3] D. A. Benson, M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers, *GenBank,* Nucleic Acids Res. **41**, D36–D42 (2013).

[4] H. Berman, K. Henrick, H. Nakamura, and J. L. Markley, *The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data,* Nucleic Acids Res. **35**, D301 (2007).

[5] M. A. Martí-Renom, A. C. Stuart, A. Fiser, R. Sánchez, F. Melo, , and A. Šali, *Comparative protein structure modeling of genes and genomes,* Annu. Rev. Biophys. Biomol. Struct. **29**, 291 (2000).

[6] T. Schwede, J. Kopp, N. Guex, and M. C. Peitsch, *Swiss-model: an automated protein homology-modeling server,* Nucleic Acids Res. **31**, 3381 (2003).

[7] L. Bordoli, F. Kiefer, K. Arnold, P. Benkert, J. Battey, and T. Schwede, *Protein structure homology modeling using SWISS-MODEL workspace,* Nat. Protocols **4**, 1754 (2009).

[8] R. Rodriguez, G. Chinea, N. Lopez, T. Pons, and G. Vriend, *Homology modeling, model and software evaluation: three related resources.* BMC Bioinform. **14**, 523 (1998).

[9] J. Buyel, J. Woo, S. Cramer, and R. Fischer, *The use of quantitative structure–activity relationship models to develop optimized processes for the removal of tobacco host cell proteins during biopharmaceutical production,* J. Chromatogr. A **1322**, 18 (2013).

[10] C. Hansch and T. Fujita, *ρ-σ-π analysis. a method for the correlation of biological activity and chemical structure,* J. Am. Chem. Soc. **86**, 1616 (1964).

[11] A. Mauri, V. Consonni, P. Manuela, and R. Todeschini, *DRAGON software: An easy approach to molecular descriptor calculations,* MATCH-Commun. Math. Co. **56**, 237 (2006).

[12] G. Cruciani, M. Pastor, and W. Guba, *VolSurf: a new tool for the pharmacokinetic optimization of lead compounds,* Eur. J. Pharm. Sci. **11, Supplement 2**, S29 (2000).

[13] Z. Li, L. Han, Y. Xue, C. Yap, H. Li, L. Jiang, and Y. Chen, *MODEL-molecular descriptor lab: A web-based server for computing structural and physicochemical features of compounds,* Biotechnol. Bioeng. **97**, 389 (2007).

[14] H. Hong, Q. Xie, W. Ge, F. Qian, H. Fang, L. Shi, Z. Su, R. Perkins, and W. Tong, *Mold^2, Molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics,* J. Chem. Inf. Model. **48**, 1337 (2008).

[15] C. W. Yap, *PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints,* J. Comput. Chem. **32**, 1466 (2011).

[16] Y. Marrero-Ponce, J. A. Castillo-Garit, E. Olazabal, H. S. Serrano, A. Morales, N. Castañedo, F. Ibarra-Velarde, A. Huesca-Guillen, A. M. Sánchez, F. Torrens, and E. A. Castro, *Atom, atom-type and total molecular linear indices as a promising approach for bioorganic and medicinal chemistry: theoretical and experimental assessment of a novel method for virtual screening and rational design of new lead anthelmintic,* Bioorg. Med. Chem. **13**, 1005 (2005).

[17] Y. Marrero-Ponce, F. Torrens, R. García-Domenech, S. Ortega-Broche, and V. Zaldivar, *Novel 2D TOMOCOMD-CARDD molecular descriptors: atom-based stochastic and non-stochastic bilinear indices and their QSPR applications,* J. Math. Chem. **44**, 650 (2008).

[18] Y. B. Ruiz-blanco, W. Paz, J. Green, and Y. Marrero-ponce, *ProtDCal: A program to compute general-purpose-numerical descriptors for sequences and 3D-structures of proteins,* BMC Bioinform. **162**, 1 (2015).

[19] D. E. Makarov, C. A. Keller, K. W. Plaxco, and H. Metiu, *How the folding rate constant of simple, single-domain proteins depends on the number of native contacts,* Proc. Natl. Acad. Sci. USA **99**, 3535 (2002).

[20] C. Micheletti, *Prediction of folding rates and transition-state placement from native-state geometry,* Proteins **51**, 74 (2003).

[21] K. W. Plaxco, K. T. Simons, I. Ruczinski, and D. Baker, *Topology, stability, sequence, and length: Defining the determinants of two-state protein folding kinetics,* Biochem. **39**, 11177 (2000).

[22] C. B. Mazza, N. Sukumar, C. M. Breneman, and S. M. Cramer, *Prediction of protein retention in ion-exchange systems using molecular descriptors obtained from crystal structure,* Anal. Chem. **73**, 5457 (2001).

[23] B. Law and S. Weir, *Quantitative structure-retention relationships for secondary interactions in cation-exchange liquid chromatography,* J. Chromatogr. A **657**, 17 (1993).

[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**, 164 (2006).

[25] T. Yang, M. C. Sundling, A. S. Freed, C. M. Breneman, and S. M. Cramer, *Prediction of pH-dependent chromatographic behavior in ion-exchange systems,* Anal. Chem. **79**, 8927 (2007).

[26] J. Kittelmann, K. M. Lang, M. Ottens, and J. Hubbuch, *An orientation sensitive approach in biomolecule interaction QSAR and application in ion-exchange chromatography,* J. Chromatogr. A **1482**, 48 (2017).

[27] J. Kittelmann, K. M. Lang, M. Ottens, and J. Hubbuch, *Orientation of monoclonal antibodies in ion-exchange chromatography: A predictive quantitative structure–activity relationship modeling approach,* J. Chromatogr. A **1510**, 33 (2017).

[28] F. Hämmerling, C. L. Effio, S. Andris, J. Kittelmann, and J. Hubbuch, *Investigation and prediction of protein precipitation by polyethylene glycol using quantitative structure–activity relationship models,* J. Biotechnol. **241**, 87 (2017).

[29] J. Kyte and R. F. Doolittle, *A simple method for displaying the hydropathic character of a protein,* J. Mol. Biol. **157**, 105 (1982).

[30] J. Thomson and F. Cavendish, *On the structure of the atom,* Phil. Mag. **7**, 237 (1904).

[31] K. C. Bauer, F. Hämmerling, J. Kittelmann, C. Dürr, F. Görlich, and J. Hubbuch, *Influence of structure properties on protein–protein interactions—QSAR modeling of changes in diffusion coefficients,* Biotechnol. Bioeng. **114**, 821 (2017).

[32] J. N. Israelachvili, *Intermolecular and surface forces* (Elsevier, 2011).

[33] J.-L. Fauchére, P. Quarendon, and L. Kaetterer, *Estimating and representing hydrophobicity potential,* J. Mol. Graphics **6**, 203 (1988).

[34] E. Audry, J. Dubost, J. Colleter, and P. Dallet, *A new approach to structure-activity relations: the molecular lipophilicity potential,* Eur. J. Med. Chem **21**, 71 (1986).

[35] P. Gaillard, P. A. Carrupt, B. Testa, and A. Boudon, *Molecular lipophilicity potential, a tool in 3D QSAR: Method and applications,* J. Comput. Aided Mol. Des. **8**, 83 (1994).

[36] A. G. Mercader, P. R. Duchowicz, F. Fernandez, E. a. Castro, and F. M. Fernández, *Modified and enhanced replacement method for the selection of molecular descriptors in QSAR and QSPR theories,* Chemometr. Intell. Lab. **92**, 138 (2008).

[37] A. G. Mercader, P. R. Duchowicz, F. M. Fernández, and E. a. Castro, *Replacement method and enhanced replacement method versus the genetic algorithm approach for the selection of molecular descriptors in QSPR/QSAR theories,* J. Chem. Inf. Model. **50**, 1542 (2010).

[38] D. M. Hawkins, S. C. Basak, and D. Mills, *Assessing model fit by cross-validation,* J. Chem. Inf. Comput. Sci. **43**, 579 (2003).

8

Outlook

This chapter provides perspectives and views on future development directions in bioseparation process development. To date the miniaturization and parallelization of screening experiments and the use of the resulting high data density for in silico modeling of process steps and biomolecule properties are major drivers in purification research.
These developments are occurring at the interface of different professions in academia, raising a synergetic potential by introducing existing technologies into the field of biomolecule separation engineering. This includes microfluidic engineering, computational sciences, and bioinformatics. Initial steps have been taken in these fields, as outlined in this thesis. Future efforts will focus on bringing these approaches further, increasing the potential of screening methodologies and data driven design of purification strategies.

The development of high-throughput screening (HTS) applications and mechanistic modeling approaches in chromatographic column performances have decoupled the scale of development and technical applications. Biomolecule characteristics and interactions with solvents, static phases as chromatographic resins, or other biomolecules can be theoretically miniaturized to a single molecule or resin particle.
Liquid handling stations (LHS), the state of the art in miniaturized downstream process development, have not yet reached their limitation, as shown in chapter 2 and 3. Nevertheless, beyond the 384 well microplate format and <10 µL volume, the challenges of accurate liquid handling have not yet been solved for conventional LHS [citation]. Therefore, microfluidic applications which handle < 1 µL volumes with high accuracy are the consequential step in further miniaturization [citation]. They not only provide liquid handling and conditioning, but also a wide range of analytical methods on-chip [citation].
Currently, the lack of standardization in microfluidic technologies hinders a further establishment of this technology. This problem has been identified within the microfluidic research community and standardized platforms are in development [citation]. With the formation of an universal development platform, a boost in request for and development of microfluidic technologies is expected.

The increasing throughput in screening provides the data basis for investigation in mechanistic and molecular properties, driving the development of *in silico* approaches to process design. A comparison to the field of chemometrics shows the direction and potential of this development. Here, public databases for molecule structures and properties are in place and are actively expanded and curated by research and industry. Structure driven modeling is based on these large and diverse datasets and benchmarks and quality standards are widely agreed upon [citation]. This extent and quality in quantitative modeling will be seen in biomolecule process development, as computational power is becoming accessible by new computer generations and distributed computing. The lack of suitable descriptors to describe large scale molecule interactions has been identified and is topic of ongoing research. Also the lack of suitable software programs for fast development cycles extending from descriptor to prediction and deployment is addressed. These developments will indubitably lead to a higher degree in model standardization and comparability between research projects.

These developments are not limited to a single purification technique, but are rather paving the change from empirical and purely experimental approaches to a model and data focused process design in biomolecule purification.

Acknowledgements

First, I would like to express my sincere gratitude to my promoter Dr.ir. Marcel Ottens for his continuous support of my PhD and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

My sincere gratitude also goes to Prof.dr. Jürgen Hubbuch, who provided me an opportunity to join his team, and who gave me access to the laboratory and research facilities. Without his support it would not have been possible to conduct my research.

I sincerely thank Prof.dr. Eppink, Prof.dr. Noorman, Prof.dr. van Esch, and Prof.dr. Hagen for being part of my PhD defence committee.

Dr. Florian Dismer is gratefully acknowledged for stimulating discussions around protein simulation and modeling. Florian always took time to listen and share his views.

Kathy, I enjoyed (and still do) our interesting and long-lasting discussions about science and technology, that often end in rather philosophical discussions. I would also like to thank you for critically reviewing my work. Your contribution not only improved the quality but your ongoing support was key to my thesis.

Sebastian, Mathias, Jan-Hendrik, and Carsten which I had the pleasure to supervise during their study projects: Your work contributed to this thesis and your engagement and vivid interest and curiosity was highly motivational.

Ansgar and Bastian for our great collaboration. It was enlightening to work with you across scientific domains.

I thank my fellow mates at MAB, Katrin, Patrick, Sigrid, Ben, Anna, Natalie, Katha, Michèle, Lara, and Sven: For the fruitful discussions and for all the fun we have had.

Last but not the least, I would like to thank my family, my parents and my brother for supporting me throughout everything I do.

Curriculum Vitæ

Jörg Kittelmann was born in Jena, Germany on March 31st, 1982. After finishing his Abitur at Kienberg Gymnasium, Berlin in 2001, he studied biotechnology at Beuth University of Applied Sciences. There, he specialised in bioprocess technology, including study visits at Commonwealth Scientific and Industrial Research Organisation in Canberra, Australia and at Forschungszentrum Jülich, Germany. He finished his studies in 2007 with his diploma thesis in the field of protein refolding and high-throughput liquid handling. His interest in automation technology led him to a technical position at the University of Karlsruhe, Germany, where he focused on lab automation. After two years he started his doctoral thesis at Delft University of Technology in collaboration with Karlsruhe Institute of Technology under supervision of Dr. Marcel Ottens and Prof. Jürgen Hubbuch working on automated screening processes and protein modeling resulting in this thesis and further publications.

List of Publications and Conferences

Publications

A. Berg, S. A. Oelmeier, **J. Kittelmann**, F. Dismer, J. Hubbuch, *Development and characterization of an automated high throughput screening method for optimization of protein refolding processes*, J. Sep. Sci., 35 (2012).

J. Kittelmann, C. P. Radtke, A. Waldbaur, C. Neumann, J. Hubbuch, B. E. Rapp, *Microfluidics on liquid handling stations (μF-on-LHS): a new industry-compatible microfluidic platform*, Proc. SPIE 8976, Microfluidics, BioMEMS, and Medical Microsystems XII (2014).

A. Waldbaur, **J. Kittelmann**, C. P. Radtke, J. Hubbuch, B. E. Rapp, *Microfluidics on liquid handling stations (μF-on-LHS): an industry compatible chip interface between microfluidics and automated liquid handling stations*, Lab Chip, 2337 (2013).

K. M. Lang, **J. Kittelmann**, C. Dürr, A. Osberghaus, J. Hubbuch, *A comprehensive molecular dynamics approach to protein retention modeling in ion exchange chromatography*, J. Chromatogr. A 1381 (2015).

K. M. Lang, **J. Kittelmann**, F. Pilgram, A. Osberghaus, J. Hubbuch, *Custom-tailored adsorbers: A molecular dynamics study on optimal design of ion exchange chromatography material*, J. Chromatogr. A 1413 (2015).

J. Kittelmann, M. Ottens, J. Hubbuch, *Robust high-throughput batch screening method in 384-well format with optical in-line resin quantification*, J. Chromatogr. B 988 (2015).

J. Kittelmann, F. Hämmerling, M. Ebeler, J. Hubbuch, *Light extinction and scattering by agarose based resin beads and applications in high-throughput screening*, J. Chromatogr. A 1397 (2015).

K. C. Bauer, F. Hämmerling, **J. Kittelmann**, C. Dürr, F. Görlich, J. Hubbuch, *Influence of structure properties on protein-protein interactions - QSAR modeling of changes in diffusion coefficients*, Biotechnol. Bioeng. 114 (2016)

F. Hämmerling, C. L. Effio, S. Andris, **J. Kittelmann**, J. Hubbuch, *Investigation and prediction of protein precipitation by polyethylene glycol using quantitative structure–activity relationship models*, J. Biotechnol. 241 (2017)

J. Kittelmann, K. M. Lang, M. Ottens, J. Hubbuch, *An orientation sensitive approach in biomolecule interaction Quantitative Structure-Activity Relationship modeling and its application in ion-exchange chromatography*, J. Chromatogr. 1482 (2017)

J. Kittelmann, M. Ottens, J. Hubbuch, *Orientation of monoclonal antibodies in ion-exchange chromatography: A predictive quantitative structure–activity relationship modeling approach*, J. Chromatogr. 1510 (2017)

Conferences

J. Kittelmann, E. Willimann, J. Hubbuch, *High throughput process development and the implication for liquid handling*, Oral presentation at the ELRIG Conference „Liquid Handling and Label Free Detection Technologies", March 5-6 2010, Whittlebury, England.

J. Kittelmann, K. Treier, J. Hubbuch, *Advances in chromatographic high-throughput screening techniques*, Poster presentation at the 1st High-Throughput Process Development (HTPD) Conference, October 4-7 2010, Krakau, Poland.

J. Kittelmann, F. Dismer, J. Hubbuch, *A combined approach: Predictive modeling based on miniaturized high-throughput experiments*, Poster presentation at the GVC/Dechema Conference "Bioverfahrenstechnik an Grenzflächen", May 29-30 2011, Potsdam, Germany.

J. Kittelmann, F. Dismer, J. Hubbuch, *A combined approach: Predictive modeling based on miniaturized high-throughput experiments*, Oral presentation at the European Congress of Chemical Engineering / European Congress on Applied Biotechnology (ECCE/ECAB), September 25-29 2011, Berlin, Germany.

J. Kittelmann, F. Dismer, J. Hubbuch, *Ion-exchange chromatography parameter prediction by protein binding orientation sensitive QSPR modeling*, Oral presentation at the GVC/Dechema Vortrags- und Diskussionstagung "Biopharmazeutische Produktion", May 14-16 2012, Freiburg, Germany.

J. Kittelmann, F. Dismer, J. Hubbuch, *384-well based HTS batch chromatography and orientation sensitive QSPR modeling of protein adsorption on ion-exchange resins*, Oral presentation at the 2nd High-Throughput Process Development (HTPD) Conference, June 4-7 2012, Avignon, France.

J. Kittelmann, J. Hubbuch, *QSPR descriptor development for protein binding orientation sensitive ion-exchange chromatography parameter prediction*, Poster presentation at 19th EuroQSAR – Knowledge Enabled Ligand Design, August 26-30 2012, Vienna, Austria.

J. Kittelmann, C. P. Radtke, J. Hubbuch, *Microtiter plate reader based spectrophotometric analytics in microfluidic chips*, Poster presentation at Lab-on-a-Chip European Congress, March 5-6 2013, Barcelona, Spain.

J. Kittelmann, J. Hubbuch, *3D ligand-protein interaction: A new set of descriptors for ion-exchange chromatography parameter prediction by QSPR modeling*, Oral presentation at the 2nd European Congress of Applied Biotechnology (ECAB), April 21-25 2013, The Hague, The Netherlands.