

Overcoming Hurdles in Downstream Processing of Tobacco-derived Biopharmaceutical Proteins

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

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

genehmigte
DISSERTATION

von
Dr. rer. nat. Johannes Felix Buyel, M. Sc.
geboren in Erkelenz

Referent: Prof. Dr. Jürgen Hubbuch
Korreferent: Prof. Dr. Rainer Fischer
Tag der mündlichen Prüfung: 23.11.2017



This document is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0): <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>

Acknowledgements

I would first like to thank my supervisor Prof. Dr. Jürgen Hubbuch for giving me the opportunity to conduct my PhD studies in such an excellent research environment as the chair of Biomolecular Separation Engineering (MAB) as part of the Institute of Engineering in Life Sciences at the Karlsruhe Institute of Technology. I am also very grateful for his constant support and scientific input, which I could count on whenever needed.

I thank Professor Dr. Rainer Fischer for being open for the collaboration between KIT and RWTH from which this thesis emerged and giving me the opportunity to conduct parts of this thesis at his institute at the RWTH Aachen University. In this context I would like to acknowledge the funding by the European Research Council Advanced Grant Future-Pharma, proposal number 269110.

I would also like to thank Prof. Dr. Steven M. Cramer for allowing me to visit his laboratory and gain insight into his fascinating research.

Another “Thanks-a-lot” goes out to Drs. Tanja Holland, Thomas Rademacher and Markus Sack for intense scientific discussions, useful comments, setting up hydroponic plant culture together and for simply being great people to work and BBQ with. Stay as you are folks...I know you will =).

Also many thanks to Drs. Sven Amrhein and Pascal Baumann as well as Ms. Margret Meixner for being the most helpful colleagues I could think of.

Thank you Dr. Richard M. Twyman for checking my Englisch ☺.

I would also like to acknowledge Carsten Erdmann from 3M and Drs. Christian Flogaus, Gergor Kalinowski and Knut Thiele from Pall for their support during initial depth filter testing and for granting me access to pilot scale STAX-filter system equipment and HyperCel chromatography media.

Finally, and most importantly, I would like to thank my parents. The only words that can catch what I owe you are: I love you!

“Prediction is very difficult, especially about the future.”

Niels Bohr

A. Table of contents

I.	Zusammenfassung	1
II.	Summary.....	5
III.	Introduction	9
III.1	Diseases and healthcare	9
III.2	Biopharmaceuticals	9
III.3	Production of biopharmaceutical proteins	10
III.4	Protein expression strategies in plants.....	10
III.5	Plant-made biopharmaceuticals	11
III.6	Downstream processing.....	12
III.6.1	Extraction and clarification	13
III.6.1.1	Pre-processing and extraction.....	13
III.6.1.2	Flocculation	13
III.6.1.3	Clarification methods	16
III.6.1.4	Filtration	16
III.6.2	Recovery and purification	21
III.6.3	Packed-bed chromatography	22
III.6.4	Models describing chromatographic separations	27
III.6.5	Prediction of chromatographic separation.....	31
III.6.5.1	Motivation and isoelectric point.....	31
III.6.5.2	Homology modeling of protein structures.....	32
III.6.5.3	Molecular dynamics simulation	34
III.6.5.4	Quantitative structure activity relationship modeling	36
III.6.6	Economic and regulatory considerations for downstream process design	38
III.7	Design of experiments.....	40
IV.	Materials and methods.....	45
IV.1	Equipment and chemicals.....	45
IV.2	Expression vectors and cloning	45
IV.2.1	<i>Agrobacterium tumefaciens</i> infiltration of plants.....	45
IV.2.2	pGFD vector	45
IV.3	Plant growth.....	46
IV.3.1	Plant species	46
IV.3.2	Greenhouse	46
IV.3.3	Post-infiltration incubation.....	46

IV.4	Sample preparation and analysis	46
IV.4.1	Sampling from transgenic plants and process steps.....	46
IV.4.2	Sampling from infiltrated plants	46
IV.4.3	Protein quantitation	47
IV.4.4	SDS-PAGE analysis.....	47
IV.4.5	Western blot and immunodetection	47
IV.5	Homogenization of leaf material.....	48
IV.6	Bag filtration	49
IV.7	Depth filtration and screening.....	49
IV.8	Flocculant and additive screening and evaluation	50
IV.9	Pretreatment of leaves or bulk plant extracts	51
IV.9.1	Blanching of plant leaves	51
IV.9.2	Heat treatment of bulk extract.....	51
IV.10	QSAR modeling.....	52
IV.10.1	HCP detection by mass spectrometry	52
IV.10.2	Homology modeling	52
IV.10.3	Descriptor calculation	53
IV.10.4	QSAR model building and prediction.....	53
IV.11	Chromatography.....	53
IV.12	Design of experiments	54
IV.12.1	Flocculant screening and model.....	54
V.	Research proposal	56
VI.	Results and discussion	58
VI.1	Scale-down of process equipment.....	58
VI.1.1	Scale-down of unit operations.....	58
VI.1.2	Verification of operational quality	60
VI.2	Screening bag filters for initial clarification	60
VI.3	Development of a disposable bag filtration step.....	61
VI.4	Screening depth filters	65
VI.5	Reducing extract turbidity by flocculation.....	71
VI.5.1	Screening of flocculants.....	71
VI.5.2	Flocculant selection and optimization.....	76
VI.5.3	Robustness of flocculation	79
VI.5.4	Process scale-up	84

VI.6	Screening of filter additives.....	87
VI.7	Economic impact of process improvements.....	89
VI.8	Pre-treatment of plants or bulk extract.....	94
VI.9	Separation of HCPs by chromatography.....	101
VI.9.1	AEX.....	101
VI.9.1.1	DEAE Sepharose FF.....	101
VI.9.1.2	ANX Sepharose FF.....	105
VI.9.1.3	Q Sepharose FF and Q HyperCel.....	106
VI.9.2	CEX – SP Sepharose.....	109
VI.9.3	HIC.....	111
VI.9.4	MMC and salt-tolerant resins.....	112
VI.9.4.1	STAR AX.....	113
VI.9.4.2	CaptoAdhere.....	115
VI.9.4.3	HEA.....	116
VI.9.4.4	MEP.....	117
VI.9.4.5	PPA.....	117
VI.9.5	pH gradients.....	118
VI.9.6	Implications for a purification strategy.....	122
VI.10	Predicting chromatographic separation.....	128
VI.10.1	Identification and homology modeling of tobacco HCPs.....	129
VI.10.2	QSAR model generation.....	132
VI.10.3	Generation of “synthetic” chromatograms.....	137
VI.10.4	Comparison with experimental separation.....	138
VI.10.5	Future improvements of prediction quality.....	141
VII.	Conclusion and scope.....	143
VII.1	Process design and control for plant-derived bio-pharmaceuticals.....	143
VII.2	The basis for a Quality-by-Design approach.....	143
VIII.	References.....	145
IX.	Appendix.....	165
IX.1	List of publications.....	165
IX.2	Register of equipment.....	166
IX.3	List of chemicals.....	167
IX.4	List of buffers.....	168
IX.5	List of identified tobacco HCPs.....	170

IX.6	List of proteins included in QSAR training sets	174
IX.7	pH gradient buffers for IEX	176

B. Abbreviations

(v/v)	volume per volume
(w/v)	weight per volume
3D	three-dimensional
aa.....	amino acid
ADP	adenosine diphosphate
AEX	anion exchange
AFM	atomic force microscopy
Amp	ampicillin
API.....	active pharmaceutical ingredient
ASA	accessible surface area
ATPS	aqueous two phase system
BCIP	5-bromo-4-chloro-3-indolyl phosphate
BET.....	Brunauer-Emmett-Teller
BLAST	basic local alignment search tool
BSA	bovine serum albumin
cAMP.....	cyclic adenosine monophosphate
CaMV	<i>Cauliflower mosaic virus</i>
CCD.....	central composite design
CEX	cation exchange chromatography
cGCP.....	current good clinical practice
cGLP.....	current good laboratory practice
cGMP.....	current good manufacturing practice
CFT.....	colloid filtration theory
CHO.....	Chinese hamster ovary
CHS	chalcone synthase
CI	confidence interval
C.V.....	coefficient of variation
Da	Dalton
df.....	degree of freedom
DLVO	Derjaguin-Landau-Verwey-Overbeek
DNA	deoxyribonucleic acid
DoE.....	design of experiments
dpi	days post injection
dpi	dots per inch
dps.....	days post seeding
DSP.....	downstream processing
DSSP.....	define secondary structure of proteins
EB	extraction buffer
EBA	expanded bed adsorption
EDC	1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide
EDTA	ethylenediaminetetraacetic acid
ELISA.....	enzyme-linked immunosorbent assay
EP	polysorbate-20 (Tween-20)
EP	electrostatic potential
ER.....	endoplasmic reticulum
FDA	Food and Drug Administration
FF.....	fast flow
FMEA.....	failure mode and effects analysis
FPLC.....	fast protein liquid chromatography

FT	flow through
GF	gel filtration
GOI	gene of interest
HBS	HEPES-buffered saline
HCP	host cell protein
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HETP	height equivalent to one theoretical plate
HIC	hydrophobic interaction chromatography
HSV	<i>Herpes simplex virus</i>
ICH	The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
IEC	ion exchange chromatography
IEP	isoelectric point
IgA	immunoglobulin A
IgG	immunoglobulin G
Kan ^r	kanamycin (resistant)
LB	left border (vector description)
LB	lysogeny broth
LDS	lithium dodecylsulfate
LMO	leave-many-out
LOF	lack-of-fit
LRF	logarithmic reduction factor
MALDI	matrix-assisted laser desorption/ionization
MALLS	multi-angle laser light scattering
MD	molecular dynamics
MES	2-(N-morpholino)ethanesulfonic acid
MLP	molecular lipophilicity potential
MLR	multiple linear regression
MMC	mixed mode chromatography
MOE	molecular operating environment
MS	mass spectrometry
MWCO	molecular weight cut off
NB-LRR	nucleotide binding–leucine-rich repeat
NBT	nitroblue tetrazolium
NEB	New England Biolabs
NHS	N-hydroxysuccinimide
NMR	nuclear magnetic resonance
NTU	nephelometric turbidity units
OD _{600nm}	optical density at 600 nm
OFAT	one factor at a time
PAA	polyacrylamide
PAGE	polyacrylamide gel electrophoresis
PAMP	pathogen-associated molecular patterns
PAT	process analytical technology
PBS	phosphate buffered saline
PBS-T	PBS with Tween-20
PDE	partial differential equation
PEST	property-encoded surface translator
pH	negative decadic logarithm of H ⁺ concentration
pI	isoelectric point
pK _a	negative decadic logarithm of the acid constant

PID.....	pipng and instrumentation diagram
PLS	partial least squares
PMF	peptide mass fingerprint
POI.....	protein of interest
polyA	polyadenylate
ppp	price per piece
PRESS	predicted residual sum of squares
PTM.....	post-translational modification
PVPP.....	polyvinylpolypyrrolidone
Q	quaternary ammonium
QA	quality assurance
QbD	quality by design
QC.....	quality control
QSAR	quantitative structure activity relationship
RB.....	right border (vector description)
RBCL.....	RuBisCO large subunit
RBCS	RuBisCO small subunit
RECCR.....	Rensselaer Exploratory Center for Cheminformatics
RMSE	root-mean-square error
RN.....	retention number
rpm.....	revolutions per minute
RPN	risk priority number
RSM.....	response surface methodology
RuBisCO.....	ribulose-1,5-bisphosphate carboxylase
SAR	scaffold attachment region
SDS.....	sodium dodecylsulfate
SEC.....	size exclusion chromatography
SMA	steric mass action
SPR.....	surface plasmon resonance
SVM	support vector machine
TAE	transferable atom equivalent
T-DNA.....	transfer DNA
TLF.....	two-level factorial
TOC	total organic carbon
Tris.....	tris (hydroxymethyl)aminomethane
TSP	total soluble protein
USP.....	upstream production
UTR	untranslated region
UV	ultraviolet
VDW.....	Van der Waals
YEB	yeast extract broth

I. Zusammenfassung

Monoklonale Antikörper (mAKs) haben im vergangenen Jahrzehnt die Landschaft der biopharmazeutischen Industrie dominiert und werden diese Vormachtstellung vermutlich auch noch einige Zeit behaupten. Dies liegt an der Vielseitigkeit ihrer Einsatzmöglichkeiten, z.B. in der Krebstherapie. Typischerweise werden mAKs in Säugerzellen, z.B. Chinese hamster ovary (CHO) Zellen, hergestellt. In der Folge haben sich biopharmazeutische Unternehmen auf dieses Expressionssystem fokussiert um das mittlerweile eine ganze Industrie entstanden ist. Trotzdem ist in den letzten Jahren festzustellen, dass das Interesse an Biopharmazeutika, die in Pflanzen produziert wurden, steigt. Der Grund dafür ist die wachsende Anzahl und Diversität von biopharmazeutischen Proteinen bei denen es sich neben mAKs auch um proteinbasierte Toxine (z.B. Visumin aus der Mistel (*Viscum album*)) oder Enzyme wie Glycocerebrosidase handelt. Diese Substanzen können u.a. in der Krebstherapie oder zur Behandlung von Stoffwechselerkrankungen wie Morbus Gaucher eingesetzt werden. Im Mai 2012 erhielt eine in Karottenzellen von Protalix Biotherapeutics hergestellte Glycocerebrosidase als erstes Biopharmakum aus Pflanzen die Zulassung für die Anwendung im Menschen durch die FDA und konkurriert seither das erfolgreich mit dem in Säugerzellen produzierten Gegenstück. Zusammen mit der Veröffentlichung der ersten Richtlinien für die Herstellung von Biopharmazeutika in Pflanzen durch die regulatorischen Behörden der USA und der EU ist es wahrscheinlich, dass eine steigende Anzahl pflanzlich hergestellter Proteine in klinischen Phasen getestet und letztlich auch in den Markt eintreten werden. Entsprechend ist mit einem Bedarf an Produktionskapazitäten für diese Proteine zu rechnen, die aktuell jedoch vor allem nur in den USA vorhanden sind. Seit 2009 ist das Fraunhofer-Institut für Molekularbiologie und Angewandte Oekologie IME die einzige Einrichtung in der EU, die eine behördliche Genehmigung für die pflanzliche Herstellung von rekombinanten Proteinen für klinische Studien besitzt. Mehr als sieben Chargen des gegen HIV gerichteten mAKs 2G12 wurden bereits erfolgreich hergestellt.

Generell können Proteine auf zwei verschiedene Arten in Pflanzen hergestellt werden. Transiente Expression beruht auf dem Transfer der für das Produkt kodierenden DNA Sequenz in die Pflanzenzellen durch Viren oder im Prozessmaßstab häufiger verwendet durch Infiltration mit *Agrobacterium tumefaciens*. Diese Methode ermöglicht die Herstellung von Proteinen innerhalb von zwei Wochen nachdem ihre DNA-Sequenz bekannt ist. Außerdem sind die Produkttiter bei dieser Methode oft höher als bei transgenen Pflanzen. Letztere haben allerdings die Vorteile, dass ihr genetischer Status klar definiert ist und sogenannte Master und

Working Seed Banks ähnlich den Master Cell Banks bei Zellkulturen erstellt werden können. Außerdem zeigen transgene Pflanzen sehr reproduzierbare Produktausbeuten, was günstig für die Beurteilung des Prozesses durch die Behörden sein kann, da diese Konsistenz zwischen Produktionschargen als Qualitätskriterium der Herstellungsprozesse auffassen. Weiterhin sind Prozesse die auf transgenen Pflanzen basieren einfacher zu skalieren, da im Gegensatz zur transienten Expression nicht jede Pflanze einer speziellen Behandlung unterzogen werden muss, durch die eine Produktbildung initiiert wird, sondern die Pflanzen *per se* das Produkt herstellen können. Entsprechend sind transgene Pflanzen gut für die Herstellung von biopharmazeutischen Proteinen im Großmaßstab geeignet. Vor allem Tabak (*Nicotiana tabacum*) hat sich zu einer Standardproduktionsplattform entwickelt, wobei die Zielproteine typischerweise aus der Blattmasse extrahiert werden.

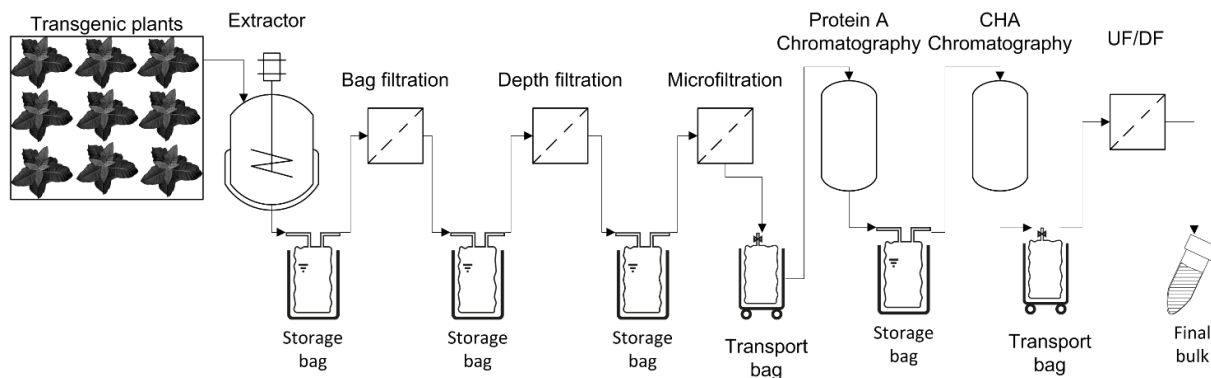


Figure I-1: Typischer Downstreamprozess für in Pflanzen hergestellte Biopharmazeutika, startend mit der Extraktion aus Pflanzen.

Bei dem resultierenden Rohextrakt (Figure I-1) handelt es sich um ein problematisches Prozessintermediat, da er nicht nur eine große Anzahl von Partikeln im Mikro- und Millimeterbereich enthält, sondern auch eine hohe Konzentration pflanzlicher Metabolite und Proteine aufweist. Während die Partikel die Kapazität von zur Klärung eingesetzten Filtern herabsetzen, können Metabolite und vor allem pflanzliche Proteine einen negativen Einfluss auf nachfolgende chromatographische Reinigungsschritte oder das Produkt haben, z.B. durch proteolytische Aktivität. Als Folge können bei pflanzenbasierten Prozessen die Produktreinigungskosten mehr als 80% der gesamten Prozesskosten ausmachen. Diese hohen Kosten haben zu Diskussionen geführt, ob pflanzlich hergestellte Biopharmazeutika überhaupt wirtschaftlich konkurrenzfähig hergestellt werden können. Daher beschäftigen sich die in dieser Dissertation beschriebenen Arbeiten mit Strategien welche die Kosten der Produktreinigung aus Pflanzen reduzieren und somit die Konkurrenzfähigkeit dieses Expressionssystems verbessern sollen. Auf der einen Seite wurde ein etablierter Filtrationsprozess optimiert, d.h. die Anzahl der Tiefenfiltrationsschritte wurde von drei auf einen reduziert. Dadurch konnten

die entsprechenden Verbrauchsmittelkosten um 50% gesenkt werden während sich die Handhabung des Systems und seine Abgeschlossenheit verbesserten. Der letzte Punkt wird vor allem auch für Prozesse die auf transiente Expression zurückgreifen wichtig sein, da dort mit S1-Bakterien gearbeitet werden muss. Außerdem wurden Hitzefällungsmethoden zur Entfernung von pflanzlichen Proteinen aus Rohextrakt standardisiert und für eine Maßstabsvergrößerung vorbereitet (Figure I-2).

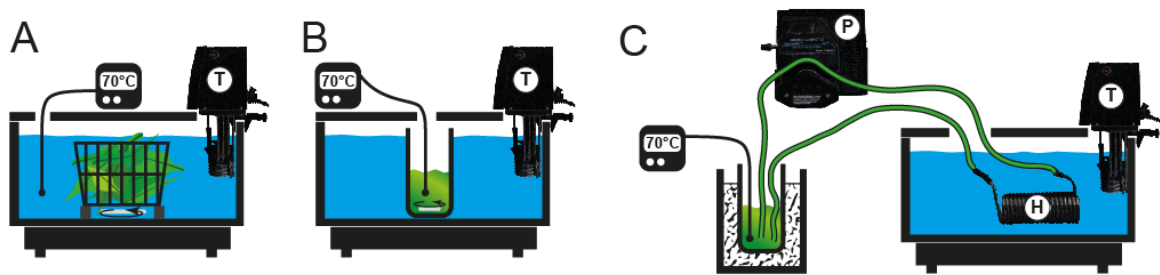


Figure I-2: Experimenteller Aufbau für die Hitzebehandlung von Pflanzenmaterial. A. Blanchieren von Blättern in heißem Puffer. B. Hitzebehandlung in einem Rührkessel. C. Erhitzen mittels Wärmetauscher.

Auf der anderen Seite wurde ein kombinierter Ansatz aus experimentellen und modellgestützten Untersuchungen verfolgt mit dem eine Datenbank der häufigsten Tabakproteine erstellt wurde welche wiederum zur wissensbasierten Vorhersage des chromatographischen Trennverhaltens dieser Proteine genutzt wurde. Dieser Ansatz wird in Zukunft eine vereinfachte und beschleunigte Prozessentwicklung für die Reinigung von biopharmazeutischen Proteinen aus Tabakextrakt ermöglichen, was wiederum die Wettbewerbsfähigkeit dieses Expressionssystems verbessern wird. Die erzeugten Modelle können darüber hinaus auch mit solchen der Proteinexpression kombiniert werden, wie sie für Tabak bereits etabliert wurden und so zu einer ganzheitlichen Prozessbeschreibung im Sinne eines QbD Ansatzes dienen (Figure I-3).

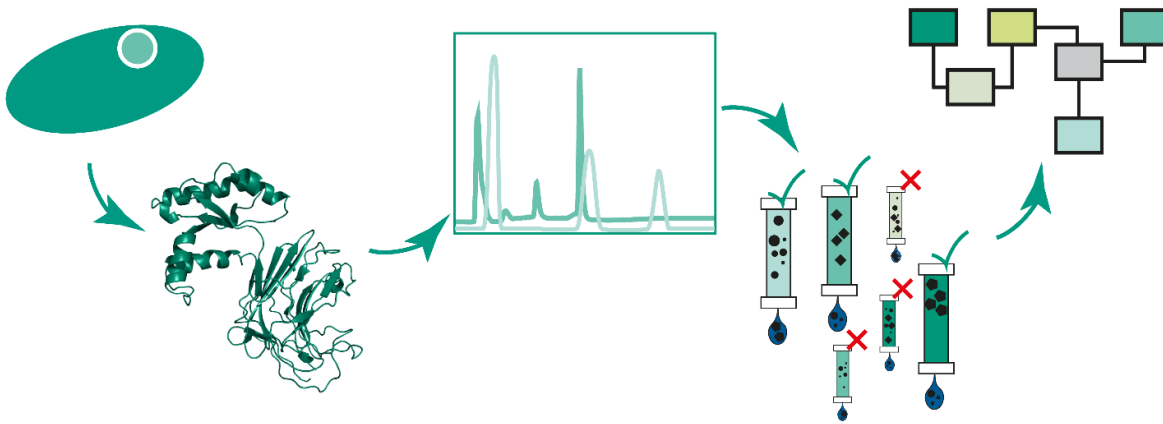


Figure I-3: Die Auswahl von Prozessstrategien kann durch die Vorauswahl von vielversprechenden Reinigungsbedingungen beschleunigt werden.

II. Summary

Monoclonal antibodies (mAbs) have dominated the biopharmaceutical landscape for the last decade and are likely to maintain their prevalence due to the versatility of their use, e.g. in cancer therapy. Such mAbs are typically produced in mammalian cell cultures, most often Chinese hamster ovary (CHO) cells. This has caused biopharmaceutical companies to focus on this expression system around which an entire industry has emerged. However, interest in plant derived biopharmaceuticals has increased over the last years, with an increasing diversity in active pharmaceutical ingredients (API) being one of the major drivers for this development. Apart from mAbs, such APIs can include protein-based toxins, e.g. viscumin originally found in mistletoe (*Viscum album*), or enzymes like glucocerebrosidase, which can be used to treat different types of cancer or Gaucher's disease respectively. Additionally, subunit vaccines directed against diseases like Malaria or HPV are now being developed in different plant-based expression systems. In May 2012, recombinant glucocerebrosidase produced in carrot cell culture by Protalix Biotherapeutics has received the full regulatory approval as the first plant-derived biopharmaceutical protein and since then successfully competed with the mammalian cell culture-derived counter part of the API. In concert with the regulatory guidelines that have been drafted by the according authorities in the US and EU for the production of biopharmaceutical proteins in plants, this development is likely to fuel the pipeline with plant-derived APIs to be tested in clinical phase trials ultimately entering the biopharmaceutical market. Therefore, production capacity is required to manufacture the corresponding plant-derived proteins in the milligram to gram quantities necessary for clinical phase trials. Currently, these capacities are mostly located in the US with only one site in the EU, i.e. the Fraunhofer Institute for Molecular Biology and Applied Ecology IME in Aachen, being operational and a second site being currently commissioned in the UK. Since 2009 the Fraunhofer IME has the permission to produce recombinant proteins in plants for clinical phase trials and has already manufactured more than 12 batches (seven under GMP conditions) of clinical-grade mAbs such as 2G12, a mAb with the potential to block the transmission of Human immunodeficiency virus (HIV).

Protein expression in plants can be achieved via two major techniques. Transient expression relies on the transfer of the DNA sequence encoding the gene of interest into the plant cell which is typically achieved by viruses or more frequently in a scalable process by infiltration with recombinant *Agrobacterium tumefaciens*. This method facilitates rapid protein production within less than two weeks after obtaining the DNA sequence of the protein of interest.

Additionally, protein expression levels are often higher compared to the other method, transgenic plants. However, transgenic plants offer the advantage of a more defined genetic status and the creation of master and working seed banks which is similar to well-known cell culture and microorganism-based approaches where master cell banks are established. Furthermore, the expression levels are highly reproducible which can be beneficial for regulatory approval where inter-batch consistency can be regarded as a quality attribute of the manufacturing process. Additionally, processes using transgenic plants are simpler compared to transient expression because the target protein is per se expressed by the plants, whereas transient expression requires that each plant of every batch is treated with the bacteria to induce target protein synthesis, which is associated with an increased handling effort and process complexity as well as additional labor and costs. Therefore, transgenic plants are the more suitable method for large scale production of biopharmaceutical proteins such as mAbs. *Nicotiana tabacum* (tobacco) has emerged as a standard production platform due to its short life cycle and high biomass output with target proteins being typically extracted from the leaves by blade-based homogenizers (Figure I-1).

The resulting raw extract is a challenging process intermediate because it contains a large number of micro- to millimeter particles, a high concentration of plant metabolites and host cell proteins. The former can induce premature blocking of the filters used to clarify the extract before chromatographic purification, which can substantially increase the consumables costs for the according filters, whereas the latter two can interfere with this purification or even cause product degradation, e.g. due to protease activity. As a consequence, more than 80% of the total production costs for plant-derived biopharmaceutical can be associated with downstream processing. These potentially excessive costs have raised questions about the economic viability of plant-based biopharmaceuticals in general. The work in this thesis has therefore focused on strategies aiming to reduce downstream processing costs and thus contribute to the competitiveness of plants as a production platform for biopharmaceutical proteins. Therefore, an established but costly filtration train was optimized by reducing the number of depth filtration steps from three to one, thereby reducing the consumables costs by more than 50% and simultaneously cutting operator handling and improving process containment. This simplification of the process was achieved by a three stage approach consisting of i) screening for more efficient filters, ii) testing of flocculants to promote aggregation of dispersed particles and thus facilitate their separation from the liquid process stream and iii) the use of filters aids to delay the clogging of depth filters. For the latter two stages a statistical experimental approach (design of experiments) was selected in order to obtain predictive models for the

flocculation and filter coating processes, which are currently too complex to be simulated by mechanistic models. The initial filter capacity of $\sim 25 \text{ L m}^{-2}$ was increased to more than 1000 L m^{-2} when flocculants and filter aids were used in combination. Polyethylene imine (Polymine P) was found to be the most effective flocculant and was best complemented with the cellulose-based filter aid LuvoZell C200. The effectiveness of the flocculants was confirmed in a 100-L pilot scale process. In total, these improvements have the potential to reduce the filtration associated costs by more than 75%.

In the context of filtration, a novel design for a single-use bag filter was developed which can replace current multi-use stainless-steel solutions. This will be a major improvement for upcoming processes using *Agrobacterium*-mediated transient expression, which will require cautious handling of plant extracts due to the safety class 1 (or higher) rating of the bacteria. The new bag filter will not only increase the process containment, i.e. prevent bacteria from contaminating surfaces of the production equipment, e.g. during cleaning, but will render cleaning and cleaning validation obsolete for this process step, reducing costs for labor, validation and monitoring accordingly.

Additionally, previously reported heat treatment methods (Figure I-2) that reduce the host cell protein concentration in tobacco extracts before purification have been standardized for easy scalability. For this purpose heat treatment by blanching (submersion of intact leaves into a hot buffer), a heat exchanger and a heated vessel (Figure 2) were compared in terms of host cell protein removal, product recovery and effect on filter capacity. Blanching at 65°C was found to be most effective in this respect and compatible with existing purification processes whereas all methods were able to remove more than 90% of the host cell proteins. This has the potential to significantly reduce the overall purification effort, e.g. reduce the number of chromatographic purification steps, and to increase the product stability due to the inactivation of host proteases. Of course, this method is only applicable to products that can withstand temperatures around 65°C without relevant conformational changes or even denaturation.

The second part of this thesis describes a combined experimental and modeling approach which was chosen (i) to establish a dataset of host cell proteins present in aqueous tobacco extracts, and (ii) to use these data for the knowledge-based design of chromatographic purification strategies for various target proteins. In a first step, a decision tree for protein purification from clarified tobacco extract was developed based on an empirical approach characterizing the separation profiles of various host cell proteins on different chromatography resins and under several conditions typically used during downstream processing of plant-derived biopharmaceutical proteins. The decision tree can be used as a simple guide to process design

heuristics. In a second step, a model based-approach was pursued and the three-dimensional structures of about 100 of the most abundant host cell proteins from tobacco were modelled. These models were combined with data from chromatographic separations of model proteins to build quantitative structure activity relationship models for the retention of these host cell proteins on defined chromatography resins. With an increasing size of model data, the accuracy for the predicted separation of host cell proteins increased and was best for SP Sepharose. This approach will facilitate and accelerate process development for new target proteins produced in tobacco as promising purification conditions can be identified more rapidly (Figure I-3), thereby reducing costs and again contributing to the competitiveness of plant-based production platforms. The models generated for the different clarification and chromatography techniques can also be combined with the models for protein expression that have been previously developed to extend the QbD approach for the whole process.

III. Introduction

III.1 Diseases and healthcare

Pathogens and their hosts exist in dynamic equilibrium, based on a balance between pathogenicity and resistance which can be shifted by favorable mutations in either species [1]. This has resulted in a co-evolutionary process similar to an ‘arms race’ [2], but humans have attempted to shift the balance in their favor by using potions and herbs which can be regarded as early biopharmaceuticals [3].

III.2 Biopharmaceuticals

Before the development of conventional drugs, people used crude extracts or mixtures, often derived from plants [4, 5]. In the 18th century, the concept of drugs based on specific active pharmaceutical ingredients (APIs) was developed, causing many of the earlier traditional remedies to fall into disuse [5]. Drugs subsequently contained chemically-synthesized APIs such as acetylsalicylic acid or sulfonamides [6-8], but following the pioneering works of researchers such as Emil von Behring (diphtheria antitoxin) and Alexander Fleming (antibiotics) in the 19th century, biopharmaceuticals became re-established [9-12]. The advantage of biopharmaceuticals was their complexity, which allowed more diverse disease targets to be addressed [13-18]. The production of such complex agents is facilitated by the selectivity and efficiency of enzymes [19], which may be provided by cultured microbial or mammalian cells as well as whole organisms such as plants [20]. However, this enhanced complexity also requires more complex manufacturing processes (e.g. aeration in biofermenters) thus introducing limitations in terms of scalability.

Small-molecule drugs (e.g. secondary metabolites) were the first of two major groups of biopharmaceuticals that gained attention in the pharmaceutical industry, because these products are often synthesized naturally by cells and have well-defined structures [9, 21, 22]. Additionally, metabolic engineering allows alternative hosts to produce the same molecules. For example, paclitaxel (Taxol) is used to treat breast cancer and can either be isolated from *Taxus brevifolia* (Pacific yew tree) or produced in optimized plant cell cultures [13]. The second group of biopharmaceuticals consists of proteins, but most medically-relevant proteins are human in origin which means natural source cannot be exploited. Therefore, such proteins are usually produced in genetically-engineered cells or organisms. This class of biopharmaceuticals is more delicate in terms of structure because (i) the molecules are larger and more complex than metabolites (e.g. antibodies are ~150 kDa in size compared to salicylic acid at 0.138 kDa)

[23], (ii) their structure is flexible and can adopt different conformations [24], which means (iii) they may undergo denaturation under unfavorable conditions (e.g. low pH, high temperature) [25].

III.3 Production of biopharmaceutical proteins

The first recombinant biopharmaceutical proteins were produced in the bacterium *Escherichia coli* [26, 27], but bacteria produce endotoxins that can be fatal if present in the final drug product [28]. An additional disadvantage is that bacteria are unable to produce complex biopharmaceuticals because they do not carry out post-translational modifications (PTMs) such as phosphorylation and glycosylation, which are often required for the efficient activity of human proteins [29]. Therefore, many recombinant proteins are now produced in eukaryotic expression platforms such as yeasts or mammalian cell cultures [27]. Chinese hamster ovary (CHO) cells are sometimes termed the ‘gold standard’ of expression platforms e.g. in the manufacture of monoclonal antibodies, which are the most prevalent biopharmaceutical products on the market [30-33]. However, mammalian cells present a risk of contamination with human pathogens and require expensive sterile cultivation equipment and media [34]. The up-front costs are therefore high despite the increasing use of disposable technologies [35]. This has encouraged the development of protein expression strategies based on plants.

III.4 Protein expression strategies in plants

Recombinant proteins can be expressed in plants by stable expression (transgenic plants) or transient expression [36, 37]. In transient expression, wild type plants are used as the transmission target, and these can be inoculated with viral vectors, bombarded with particles or injected/infiltrated with genetically-modified *Agrobacterium tumefaciens* [38, 39]. These strategies avoid the time-consuming principles of stable transformation, which include tissue culture, selection, regeneration and self-pollination, but the target proteins are rapidly expressed at high levels [38, 40-42]. *Agrobacterium*-mediated transient expression can be achieved using standard transfer DNA (T-DNA) expression constructs but hybrids containing elements from plant viruses can also be used [39]. The T-DNA, containing the gene of interest (GOI), is exported from the bacterium to nearby plant cells via a type IV secretion system [43], and imported into the nucleus of recipient cells by the combined activity of plant and bacterial proteins, when it is expressed [39]. The resulting protein of interest (POI) can be recovered from whole plants or plant tissues after a typical incubation period of 3–8 days [44-47]. Using this method, it is possible to obtain the recombinant protein less than two months after the DNA

construct is provided, and within 2–3 weeks if wild-type tobacco plants of a suitable age are maintained ready for use [48]. The ability to scale up production rapidly is a significant advantage over both transgenic plants and conventional production platforms, allowing for examples of production of vaccines in response to sudden epidemic outbreaks of influenza. Transient expression also allows the production of proteins that are toxic to plants if they accumulate during vegetative growth [49].

Transgenic plants are generated by the stable integration of a transgene into the host genome, which can be achieved by particle bombardment or other direct transfer methods, or biological delivery using *A. tumefaciens* [50]. The selection of stable integration events is facilitated by the simultaneous introduction of marker genes such as neomycin phosphotransferase (*nptII*), which confers resistance to aminoglycoside antibiotics [51]. Selection is usually implemented in callus culture, followed by regeneration into fertile plants [52] which are self-pollinated for several generations to produce homozygous transgenic plants in which the transgene is stable [53, 54]. The duration of this process is determined by the generation time of the plant species, e.g. ~24 months is sufficient for the Petit Havana SR1 cultivar of tobacco, assuming five rounds of self-pollination (T_5) [37]. The advantages of stable transgenic plants are the reproducible expression levels of the transgene, the defined master seed bank and the ease of scale-up [37, 38, 55].

III.5 Plant-made biopharmaceuticals

The first transgenic plants expressing recombinant proteins were developed in the early 1980s but the first biopharmaceutical proteins were not produced in plants until 1989 [56-58]. Different plant species and tissues offer unique advantages but also suffer from certain limitations that must be taken into account when a species is selected for production [37, 59]. Cereal crops, such as wheat (*Triticum* spp.), barley (*Hordeum vulgare*) and maize (*Zea mays*), allow the long-term storage of recombinant proteins in seeds but pharmaceutical crops could contaminate the human food chain [60, 61]. *Nicotiana* species, such as *N. tabacum* (common tobacco) or *N. benthamiana*, offer a high biomass yield and recombinant proteins are generally produced in the leaves [62]. However, the removal of toxic secondary metabolites such as nicotine must be demonstrated during the purification process, although this can be facilitated by selecting a variety with low basal levels of nicotine. The first plant-derived pharmaceutical protein (a monoclonal antibody) was expressed in tobacco [58] although since then many other pharmaceutical proteins have been produced in a diverse range of plants [63].

The major economic benefits of biopharmaceutical protein production in plants include the low cost of upstream production, the potential for large-scale cultivation, and the inability of human pathogens to replicate in plants improving the safety profile of this production platform [34, 37, 59, 60, 64]. Additionally, oligomeric proteins are correctly assembled in plant cells and PTMs such as glycosylation are introduced correctly [65-67]. The glycans can be modified further to improve API efficacy and reduce immunogenicity in humans by additional genetic engineering steps [68, 69].

Despite the advantages of plants, the first plant-derived biopharmaceutical product for human use was only granted full FDA approval in May 2012 (Elelyso, produced by Protalix Biotherapeutics, Carmiel, Israel; FDA application number (NDA) 022458). Many technological and economic hurdles caused this delay [55]. One technological issue is the burden of downstream processing costs, which account for a large proportion of overall production expenses [63]. Measures to reduce downstream processing costs will significantly improve the competitiveness of plant-based expression systems compared to traditional platforms for the production of biopharmaceutical proteins such as CHO cells.

III.6 Downstream processing

The production of biopharmaceutical proteins can be divided into two major work packages: upstream production and downstream processing. Upstream production (USP) involves the generation and cultivation of all the biological components of an expression system such as cells for fermentation, tobacco plants for cultivation, and modified *A. tumefaciens* for infiltration. It also includes all stages of the production cycle in which the recombinant protein is synthesized and accumulated. Downstream processing (DSP) involves the separation of the product from the other biological components and water, thus ensuring it is purified to a pre-defined degree. DSP can be sub-divided into three steps as shown in Table III.1 [70].

Table III.1: Common steps during the downstream processing of biopharmaceuticals.

Step	Aim
Pre-processing, extraction and clarification	Condition source material for processing; extract product from bulk biomass; remove particulate matter to facilitate further processing
Recovery ¹	Capture target protein from bulk extract; remove bulk water and some major impurities
Purification and polishing	Remove impurities and aggregates to achieve desired final purity

¹ The transition from recovery to purification and polishing is a continuum as both steps remove impurities

III.6.1 Extraction and clarification

III.6.1.1 Pre-processing and extraction

Pre-processing is the conditioning of bulk biomass into a form that can easily be handled in later processing stages. For a plant-based system, this can involve harvesting (removing the leaves from plant stems) and cleansing them to reduce bioburden after extracting the protein [71].

Protein extraction from leaves can be achieved by pressing, milling or blending, using rotating blades [63, 72-74]. Homogenization by blending often involves the addition of an extraction buffer containing a buffering agent, such as phosphate, to stabilize the pH during extraction, and salt, such as sodium chloride, to improve protein solubilization and prevent precipitation [63]. For tobacco, a pH of 7.5 achieves the most effective protein extraction [75]. Detergents can increase the efficiency of protein extraction but may also solubilize larger amounts of host cell proteins (HCP) [73]. Antioxidants are often included in extraction buffers to prevent protein modifications that can affect product quality and the oxidation of polyphenols that can interfere with filtration or chromatography [73, 76-78]. Further substances may be added to the buffer to facilitate subsequent process operations, e.g. polyvinylpolypyrrolidone (PVPP) is used to bind polyphenols and tannins [63, 78].

III.6.1.2 Flocculation

Flocculants are additives, based on macromolecular polymers with varying cationic or anionic charges, charge densities and molecular masses [79-81]. They are used to remove dispersed particles from solutions. Flocculation is achieved by bridging (cross-linking) particles, like cell debris, with a polymer of opposing charge, thereby forming large flocks that will sediment more rapidly than smaller individual particles (Figure III-1 C) [79, 82]. Bridging requires the polymer to extend beyond the diffuse double layer (the Debye length κ^{-1} , Equation 1 [83]) surrounding a dispersed particle and attach to a second particle (Figure III-1 A). Polymers with a low charge density often function in this manner, and increasing the flocculant concentration above an optimal value will result in the re-stabilization of dispersed particles. Another mode of action is charge neutralization, where polymers bind to and shield charged surface patches on dispersed particles thereby rendering them insoluble [79, 84] (Figure III-1 B). This is often observed for highly-charged flocculants, where flocculation is independent of the polymer concentration, but for complete aggregation of dispersed particles quantitative amounts of charged flocculant must be added in order to neutralize all particle charges.

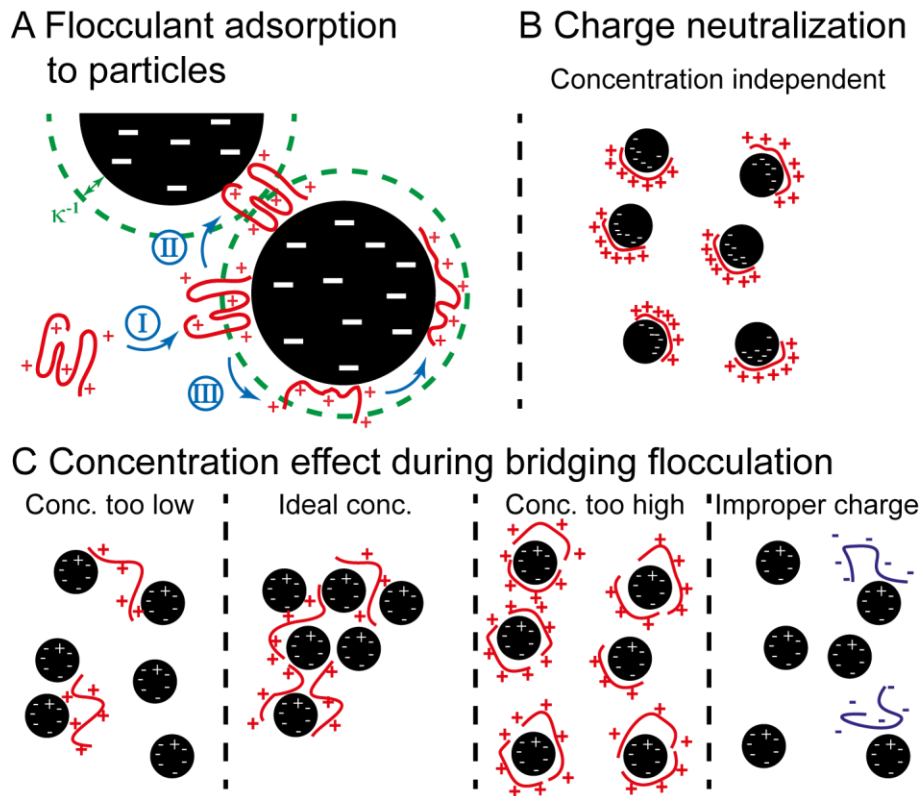


Figure III-1: Adsorption mechanisms of flocculants.

A. Charged polymers (red) initially bind to oppositely-charged particles (black) in a coiled state (I). Depending on the relaxation time t_R , adsorption time t_A and collision time t_C , the polymer can then either interact with a second particle to form bridges (II) if the polymer extends beyond the inverse Debye length (green dashed line, κ^{-1}) or adopt an expanded conformation (III) and neutralize surface charges. B. For charge neutralization, flocculation is concentration-independent but its quantity depends on the amount and charge density of the added polymer. C. Bridging flocculation is concentration-dependent where low concentrations of polymer fail to induce aggregation and excessive concentrations prevent effective bridging due to surface saturation.

$$\kappa^{-1} = \sqrt{\frac{\varepsilon_r \varepsilon_0 k_B T}{2 N_A e^2 I}}$$

Equation 1: Debye length κ^{-1} in a colloidal dispersion, with the ionic strength of the electrolyte I [mol L⁻¹], permittivity of free space ε_0 , the dielectric constant ε_r , the Boltzmann constant k_B , the absolute temperature T , Avogadro's number N_A and the elementary charge e .

The forces driving the adsorption of flocculants to dispersed particles are electrostatic and hydrophobic interactions as well as hydrogen bonding and ion binding [79]. The adsorption kinetics can be described by a second-order reaction (Equation 2) and the rate constant can be calculated for non-stirred conditions (only diffusion) and stirred conditions involving shear forces (Equation 3 and Equation 4, respectively) [79, 80]. The rate constant will also determine the time required for a certain fraction of polymer to adsorb to the dispersed particles (t_A) and the characteristic collision time (t_C) as shown in Equation 5 and Equation 6, respectively [80]. These equations also show that the process of flocculation can be accelerated by increasing the shear rates, resulting in shorter adsorption and collision times. However, higher shear rates can

also lead to the breakage of flocks and thus reduce particle aggregation [85]. The equilibrium state of flocculation can be represented by a Langmuir isotherm (Equation 30) [80].

$$J_{ij} = k_{ij}n_i n_j$$

Equation 2: Collision rate J_{ij} between dispersed particles i and polymer j , with particle flocculation rate constant k_{ij} as well as number concentrations of particles n_i and polymer n_j .

$$k_{ij} = \frac{2k_B T (r_i + r_j)^2}{3\eta r_i r_j}$$

Equation 3: Particle flocculation rate constant k_{ij} calculated for diffusion, with Boltzmann constant k_B , absolute temperature T , viscosity η and effective radii of particle r_i and polymer r_j .

$$k_{ij} = \frac{4}{3} G (r_i + r_j)^3$$

Equation 4: Particle flocculation rate constant k_{ij} calculated for shear, with shear rate G and effective radii of particle r_i and polymer r_j .

$$t_A = -\frac{\ln(1-f)}{k_{ij}n_i}$$

Equation 5: Adsorption time t_A for a specific fraction f of the added polymer, with flocculation rate constant k_{ij} and number concentration of dispersed particles n_i .

$$t_C = \frac{1}{k_{ij}n_i}$$

Equation 6: Collision time t_C with flocculation rate constant k_{ij} and number concentration of dispersed particles n_i .

Polymer adsorption is a sequential process involving the rearrangement of polymers from a coiled to a flat conformation after initial binding to the particle surface (Figure III-1 A III) [79, 86]. This process is called relaxation or inactivation, and is described by the relaxation time t_R [80, 86]. Thus, t_R needs to be larger than t_A and t_C for effective bridging flocculation to occur because this mechanism requires a polymer to extend beyond the diffuse double layer of a particle as stated above [80]. Under these conditions ($t_R > t_A$ and $t_R > t_C$) the reaction kinetics are biased towards higher-order rather than second-order reactions because more than two molecules are usually involved per interaction [87].

The efficacy of flocculants therefore depends on their concentration, charge density and length, the shear rate and particle concentration in the feed stream as well as the solution pH and ionic strength [79-81, 88]. Higher ionic strength, or background salt concentration, can affect flocculation by (i) screening the particle surface charge, (ii) reducing the thickness of the diffuse double layer surrounding the charged particle, (iii) reducing the flatness of the adsorbed polymer, (iv) reducing polymer binding affinity, and (v) reducing the polymer effective charge [79, 80]. A higher salt concentration can therefore facilitate bridging flocculation as often observed for polymers with a low charge density, because the coiled polymer structure is

avored allowing simultaneous binding to several dispersed particles. In contrast, high salt concentrations can hamper flocculation based on charge neutralization driven by highly-charged polymers, by reducing their affinity towards the dispersed particles [80]. Depending on the selected conditions, flocculants may also precipitate proteins and therefore their impact on a target protein must be carefully assessed [78, 89]. Several models have been developed to describe the settling behavior of the resulting flocks [90].

III.6.1.3 Clarification methods

The bulk plant extract contains dissolved molecules such as proteins, DNA, carbohydrates and secondary metabolites as well as colloid particles. These particles must be removed because many subsequent unit operations are not compatible with suspensions, e.g. packed-bed chromatography [63, 73]. Heat treatment can be used to precipitate HCPs but the target protein may also be affected [70]. Aqueous two-phase systems (ATPS) can remove particles from bulk extract inexpensively, simultaneously enriching the target protein [91, 92]. In principle, a solution containing two phase-forming substances (salt-salt, salt-polymer or polymer-polymer) is mixed with the extract and incubated until two phases form. Ideally, the target protein will accumulate in the top phase, corresponding to a distribution coefficient above unity, whereas particles will settle in the bottom phase under gravity [92]. The ATPS buffer can also be used as the extraction buffer, thereby reducing the number of unit operations. Drawbacks of ATPS include the high concentrations of salt and/or polymer required for phase formation and the long time that may be needed for phase separation [93]. Phase formation is also sensitive to salt concentration and pH, two factors that may differ according to the properties of the target protein. It is therefore difficult to develop a generic platform process with defined parameters, but this may become feasible with new automated screening procedures [94]. Particle removal can also be achieved using (semi)continuous centrifuges [63, 82]. These accelerate the natural sedimentation of particles by applying multiples of the gravitational force. Although this method is well known, it may be difficult to implement in a large-scale process due to the high investment costs for this equipment, although this may not be the case for multi-product facilities [95]. However, cleaning procedures between batches and campaigns can be laborious and time-consuming without completely excluding the risk of cross-contamination.

III.6.1.4 Filtration

Cross-contamination and high investment costs are less of a concern when disposable filters are used for particle removal, but consumables costs may be higher because high-quality filter material must be used [96, 97]. Filters featuring different materials and progressively smaller

pore sizes can be combined into an optimal filter train that will guarantee a defined, low-turbidity processed extract (Figure III-2 A) [98, 99]. The first filter has a large pore size to retain large particles that build up a filter cake, which will in turn provide progressively higher filter efficiency and increasing backpressure assuming a constant flow rate. Bag filters comprising needle felt or nylon filament are suitable for this step. Depth filters, with a finer retention rating, are used in the subsequent steps to remove particles with an average size of 1 μm [95, 98]. These filters remove particles by surface retention, like bag filters, but also through depth straining, settling zones and particle adsorption (Figure III-3 A) [82, 100, 101]. Depth filters can therefore remove particles larger and smaller than the nominal retention rating.

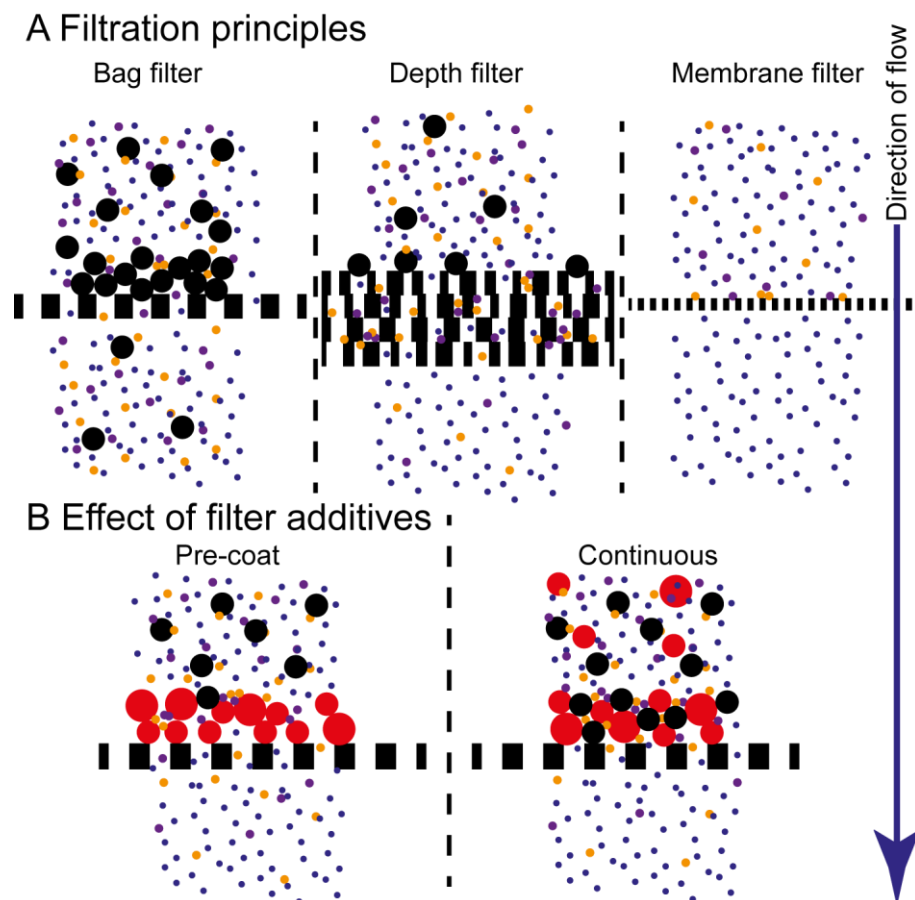


Figure III-2: Schematic representation of filtration methods.

A. Bag filters use retained particles from the filter stream to gradually build up a finer filter cake for more effective particle retention. Size exclusion, depth straining in narrow flow channels and adsorption are the key mechanisms in depth filtration, whereas membrane filters use an absolute pore size, defined by their microstructure, to retain dispersed particles. B. Filter additives (red) can prevent the blocking of filter pores, and provide additional surfaces for particle adsorption thereby prolonging filter capacity.

The retention achieved with depth filters is not absolute, i.e. particles larger than the nominal retention rate may pass through the filter with a certain probability. A visible filter cake is not always formed due to the low abundance of large particles in the feed stream. Filter additives such as cellulose or diatomaceous earth may be added to the feed stream to artificially

build-up a filter cake over time that can increase the filter capacity (Figure III-2 B) [95]. Depth filters can also bind dissolved molecules such as host cell proteins (HCP) and DNA on charged surfaces, thereby simplifying subsequent purification steps [102].

An initial, macroscopic description of depth filtration was provided in the late 1930s [103] and experimentally verified in the late 1960s [104]. First order kinetics (Equation 7) and the filter coefficient λ were used to describe the declining dispersed particle concentration along the length of a depth filter, as recently reviewed [105]. In the early 1970s the classical colloid filtration theory (CFT) was developed, splitting the adsorption of dispersed particles (colloids) to a collector (filter grain, solid phase) into two steps [106]. First, the single collector efficiency (collision rate) η was used to express particle transport towards a collector as the proportion of dispersed particles in a feed stream colliding with a collector r_s relative to the total number of particles approaching the collector r_f (Equation 8). Hence, η always adopts a value between zero and unity. Second, the collision efficiency α was used to determine the fraction of particle–collector collisions r_s that resulted in particle attachment r_a (Equation 9). As for η , values for α are in the range 0 to 1. A relationship between the filter coefficient λ and α and η can be established using Equation 10 [107, 108].

$$-\frac{\partial C}{\partial L} = \lambda C$$

Equation 7: Change of particle concentration ∂C due to adsorption to a filter material with filter depth L , particle concentration C at a given depth L of the filter and the filter coefficient λ .

$$\eta = \frac{r_s}{r_f} = \frac{r_s}{\pi b^2 u C_0} = -\frac{2}{3} \frac{d_c}{(1 - \epsilon)L} \ln\left(\frac{C}{C_0}\right)$$

Equation 8: Theoretical and experimental calculation of single collector efficiency η (modified from [109] according to [110] and [111]) with the number of particle-collector collisions r_s (the letter I is also used in the literature), the total number of particles streaming towards the collector r_f , the radius of the fluid envelope in Happel's model b , the approach velocity of the fluid u and the number concentration of particles in the fluid approaching the collector C_0 as well as collector diameter d_c , the bed porosity ϵ , the filter length (bed height) L and number concentration of particles at the outlet C .

$$\alpha = \frac{r_a}{r_s} = \frac{\eta}{\eta_0} = -\frac{2}{3} \frac{d_c}{(1 - \epsilon)L\eta_0} \ln\left(\frac{C}{C_0}\right)$$

Equation 9: Theoretical and experimental calculation of collision efficiency α (according to [108]) with the number of collisions resulting in particle attachment to the collector r_a , the number of particle-collector collisions r_s (the letter I is also used in the literature), the single collector *removal* efficiency η , the single collector *contact* efficiency η_0 (η in the absence of repulsive DLVO forces), as well as collector diameter d_c , the bed porosity ϵ , the filter length (bed height) L and number concentration of particles at the outlet C .

$$\lambda = \frac{3(1 - \epsilon)}{2} \frac{\alpha \eta_0}{d_c}$$

Equation 10: Correlation between filter coefficient λ and single collector efficiency η , as well as collision efficiency α with the bed porosity ϵ , the collector grain diameter d_c (compare with definition of k below).

Happel's sphere-in-cell porous media model [112] is often used to calculate trajectories for colloids approaching a collector and their collision with its surface, even though alternative models have been proposed including hemispheres-in-cell [113], array-of-spheres [114], and simple cubic and dense cubic packing geometries [101].

In the initial model, η was defined by the physicochemical mechanisms of diffusion (Brownian motion, Figure III-3 A I), interception (collision of a particle and a filter grain reflecting their corresponding diameters and relative positions in the flow pattern, Figure III-3 A II) and sedimentation (action of gravity, Figure III-3 A IV). Subsequent models introduced additional mechanisms such as inertia (Figure III-3 A III) [109], hydrodynamic action (particles shifting between streamlines due to their irregular shape, Figure III-3 A V) [105, 110], surface straining (retention of particles in pores smaller than the particle diameter, Figure III-3 A VI) [107] and settling in stagnant flow zones without attachment (Figure III-3 A VII) [115, 116]. Several equations have been presented to calculate η based on dimensionless numbers such as the Peclet number [109, 110] or filtration experiments [111].

The collision efficiency α is often calculated based on colloid-surface interactions according to the Derjaguin-Landau-Verwey-Overbeek (DLVO) theory [117, 118] including London-van-der-Waals interactions and electrostatic double-layer repulsion (Figure III-3 B) [119-121]. Under ideal, destabilized conditions, the attachment of dispersed particles to a collector is termed *favorable* (no repulsive forces or energy barriers present [122]) and the value of α approaches unity [106]. The concentrations of dispersed particles (C) and attached particles (S) under these conditions can be calculated according to the CFT using Equation 11 and Equation 12, respectively [122, 123], but note that different variants of these equations have been described and discussed in the literature [124].

$$C(x) = C_0 e^{-\frac{k}{v}x}; k = \frac{3(1-\epsilon)v}{2d_c} \eta$$

Equation 11: Dispersed particle concentration C , dependent on filter depth x assuming a clean filter (t_0) according to CFT with the number concentration of particles in the fluid approaching the collector C_0 , the particle deposition coefficient k and the interstitial particle velocity v . A constant value of k is assumed with the bed porosity ϵ , the collector grain diameter d_c and the single collector *removal* efficiency η .

$$S(x) = \frac{t_0 \epsilon k C_0}{\rho_b} e^{-\frac{k}{v}x}$$

Equation 12: Attached particle concentration S , dependent on filter depth x assuming a clean filter (t_0) according to CFT with the bed porosity ϵ , the number concentration of particles in the fluid approaching the collector C_0 , the porous medium bulk density ρ_b , the particle deposition coefficient k and the interstitial particle velocity v . A constant value of k is assumed.

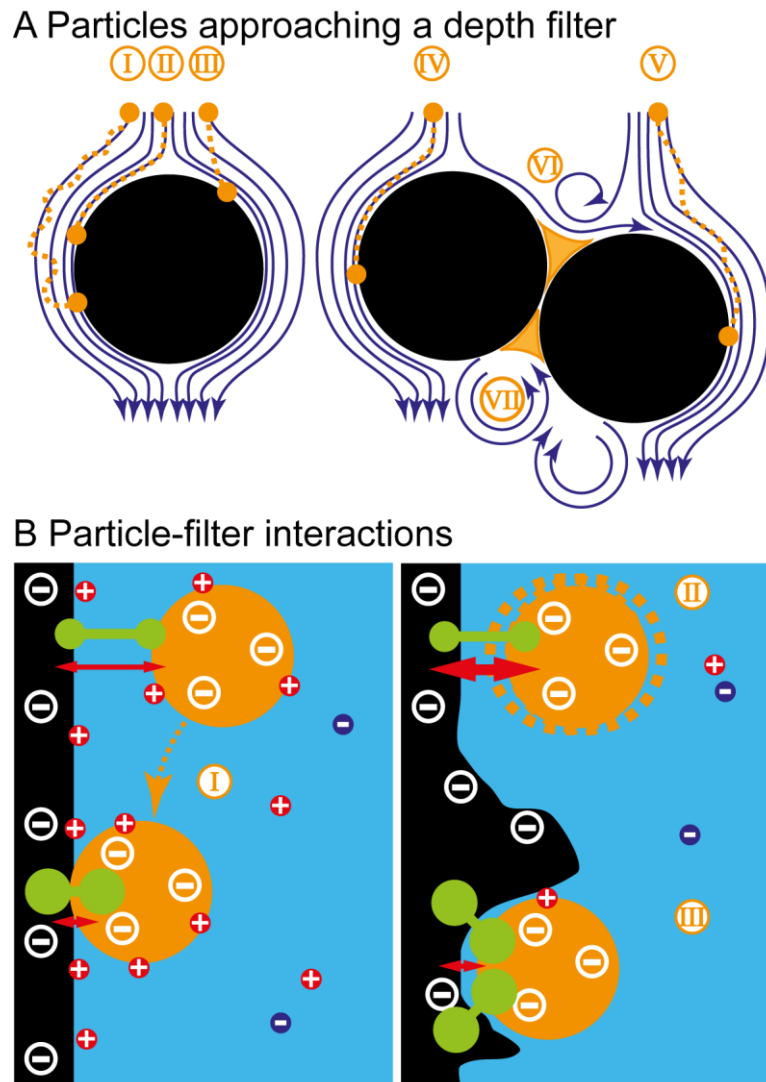


Figure III-3: Schematic representation of depth filtration mechanisms.

A. Particles (orange) in the feed stream (blue arrows) can approach depth filter material due to different mechanisms such as diffusion (I), interception (II), inertia (III), sedimentation (IV) and hydrodynamic action (V) as well as straining (VI) and deposition in stagnation zones (VII) (see text for references and detailed explanation). B. Under favorable conditions (I) electrostatic repulsion (red double arrows) is shielded by dissolved counter-ions (red) whereas London-van-der-Waals forces (green dumbbells) result in particle attachment to the filter surface. Unfavorable conditions (II) can result in a second, local energy minimum at a certain distance from the filter layer within which a particle can become trapped. Derivations from the model predictions (III) occur for example due to surface roughness or uneven charge distribution (see text for detailed discussion). Section A of this figure was modified from [125].

Conditions deemed *unfavorable* for particle adsorption exhibit significant repulsive DLVO interactions [122]. These can arise from electrostatic repulsion between particles and collectors sharing the same charge, which can be measured in terms of ζ -potentials [126]. Unfavorable conditions are often generated by low ionic strength buffers, such as <100 mM salt. Accordingly, the solution pH, ionic strength and approach velocity will affect particle adsorption [101, 111, 122, 127]. Particles may therefore be trapped in a secondary energy minimum (in addition to the first minimum, resembling particle attachment to the collector surface) preventing permanent adsorption to the collector surface [101, 127, 128]. However,

significant differences between CFT predictions and experimental data have been observed under such unfavorable conditions [120, 129, 130]. Several mechanisms have been proposed to account for these observed deviations, including (i) surface roughness on the collector and the dispersed particles (as well as their size and shape) resulting in a locally diminishing/disappearing energy barrier towards the first minimum [111, 131-133]; (ii) particles wedged between two energy barriers at sites of filter grain-to-grain contact, which may overcome this barrier by diffusion or hydrodynamic action [101]; (iii) colloid retention in stagnant flow zones without attachment [134]; and (iv) the effect of shear forces on loosely-attached particles [116]. In this context, a bimodal model accounting for fast (favorable) and slow (unfavorable) adsorption has been proposed [119].

Membrane filtration (Figure III-2 A) is often used as a final step before chromatography to remove bacteria and any remaining colloidal particles $<1 \mu\text{m}$ in diameter [95]. These filters have a fixed pore diameter which will guarantee the removal of all particles above this size. Particle removal is achieved by surface filtration and pore sizes are typically 0.45 or 0.22 μm . Some of the concepts described above for depth filtration also apply to membrane filtration [135].

III.6.2 Recovery and purification

The objective of recovery is to separate a POI from major HCP impurities and water with minimal losses of the product while maintaining its integrity and functionality [63]. In subsequent purification steps, a pre-defined product purity must be achieved by the removal of any trace HCP [136]. A final polishing step is often included for recombinant proteins with biopharmaceutical applications in order to remove product aggregates, such as dimers in the case of a monoclonal antibody [137].

Various techniques can be used for recovery and purification, each with specific benefits and drawbacks, e.g. precipitation, ultrafiltration, crystallization and various chromatographic methods, the latter introduced in more detail in section III.6.3 [70, 73, 138]. Precipitation can easily be applied on a large scale by adding kosmotropic salts such as ammonium sulfate, by adding charged polymers or by applying heat, but the costs to dispose of salt-containing waste may be high [70, 139] and the potential exists to alter the conformation of the target protein which might affect its functionality [140]. Ultrafiltration using membranes with certain molecular weight cut-off (MWCO) values can also be applied on a large scale but the resolution may be low compared to other techniques. High purities can be achieved by crystallization but finding conditions suitable for crystal formation can be difficult and time-consuming [141,

142]. Expanded bed adsorption (EBA) is a variant of chromatography that can tolerate non-clarified and partially clarified feed-streams due to the wide spaces between the floating resin particles [143]. However, EBA-specific problems include axial mixing, aggregative fluidization, channeling, bed collapse or dead water zones within the bed, and conditions must be adjusted to individual feed stream properties [144, 145]. High flow rates can be applied using membrane chromatography cassettes thus reducing overall process times, but the binding capacity of membranes can be low making them less economical than traditional packed-bed chromatography [146].

III.6.3 Packed-bed chromatography

Packed-bed chromatography is the most common chromatography technique in bioprocessing and it is the central component of many recovery and purification operations [147]. The packed bed consists of spherical resin particles, which are beads conforming to a particular diameter range that can feature certain functionalities (Table III.2). These particles often have a porous structure and are densely packed into a column (Figure III-4 as well as Equation 13 and Equation 14).

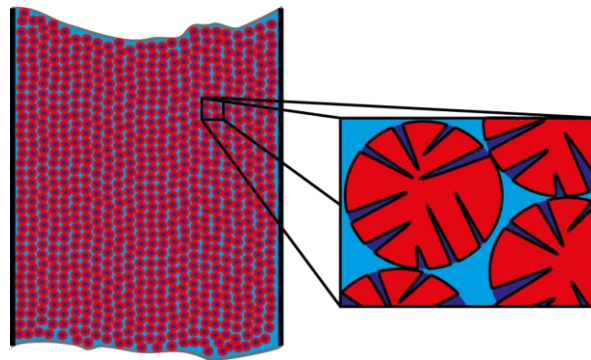


Figure III-4: Packed resin particles (beads) in a column section.

The liquid mobile phase (blue, V_M) can be subdivided into a bulk interstitial part (light blue, V_I) outside the porous particles (red, V_S) and a part inside the pores (dark blue, V_P) where different diffusion, mass transfer and size exclusion limits apply (see Equation 26).

Most resin particles are constructed from a cross-linked polymer, such as agarose, which is functionalized by covalent modification with specific ligands, e.g. quaternary ammonium ions [148]. These ligands interact with proteins passing through the column and define the type of chromatography: (i) hydrogen-bonding for some mixed-mode chromatography (MMC) resins; (ii) electrostatic interactions for anion exchange (AEX) or cation exchange (CEX) and MMC resins; and (iii) hydrophobic and Van-der-Waals (VDW) interactions in hydrophobic interaction chromatography (HIC) and MMC resins (Table III.2).

$$V_C = r^2 \cdot \pi \cdot L = V_M + V_S = (V_I + V_P) + V_S$$

Equation 13: Column volume V_C , with column radius r , column length L , mobile phase volume V_M , solid phase volume V_S , interstitial volume V_I (liquid volume outside the particles) and pore volume V_P (liquid volume inside the particle pores).

$$\beta = \frac{V_M}{V_S}$$

Equation 14: Phase ratio β , with mobile phase volume V_M and solid phase volume V_S .

Table III.2: Functional groups of ligands used in liquid chromatography including their mode of action, general interaction strength and capacity.

Ligand ¹	Functional group	Mode of action	Capacity/ligand density ²
Q	Quaternary amine	AEX, strong	0.18-0.25 mmol [Cl ⁻ mL ⁻¹]
ANX	Diethylaminopropyl	AEX, weak	0.13-0.17 mmol [Cl ⁻ mL ⁻¹]
DEAE	Diethylaminoethyl	AEX, weak	0.11-0.16 mmol [Cl ⁻ mL ⁻¹]
SP	Sulfopropyl	CEX, strong	0.18-0.25 mmol [H ⁺ mL ⁻¹]
CM	Carboxymethyl	CEX, weak	0.09-0.13 mmol [H ⁺ mL ⁻¹]
Phenyl (high)	Phenyl	HIC, aromatic	40 [μmol mL ⁻¹]
Phenyl (low)	Phenyl	HIC, aromatic	25 [μmol mL ⁻¹]
Octyl	Octyl	HIC, aliphatic	5 [μmol mL ⁻¹]
Butyl	Butyl	HIC, aliphatic	40 [μmol mL ⁻¹]
Butyl-S	Butyl-S	HIC, aliphatic	10 [μmol mL ⁻¹]
MMC	N-benzoyl-homocysteine	CEX (weak) and aromatic	0.07-0.09 mmol [H ⁺ mL ⁻¹]
MMA	N-benzyl-N-methyl ethanolamine	AEX (strong) and aromatic	0.09-0.12 mmol [Cl ⁻ mL ⁻¹]
<i>MEP</i>	4-mercapto-ethyl-pyridine	Charge induction, aromatic	70-125 [μmol mL ⁻¹]
<i>HEA</i>	Hexylamine	AEX and aliphatic	n.d.
<i>PPA</i>	Phenylpropylamine	AEX and aromatic	n.d.
<i>STAR AX</i>	Primary amine	AEX	n.d.
<i>(Hy)Q</i>	Quaternary amine	AEX, strong	0.09-0.14 mmol [Cl ⁻ mL ⁻¹]

¹ Media from Pall are italicized; ² Values provided for Sepharose fast flow (FF) (GE) or HyperCel (Pall). Note that weak exchangers will alter their capacity with changing pH while strong do not.

The strength of any protein–ligand interaction is dependent on the properties of the protein (e.g. charge density distribution), the ligand (e.g. cationic charge), the buffer conditions (e.g. pH and conductivity), and the resin (e.g. ligand density and pore size) [148]. The interaction strength in turn determines whether a protein will pass through the column (no interaction or weak interaction), elute after a delay compared to the sample front (medium interaction) or bind to the resin (strong interaction) and hence determines the retention volume V_R (Equation 15) [149]. The interaction strength also determines the retention factor k of a protein (Equation 16) as well as the selectivity coefficient $k_{A/B}$ and separation factor α for each pair of protein species in a chromatography setup (Equation 17 and Equation 18 respectively)

[150, 151]. If only one protein species binds to the resin, it is defined as specific, e.g. some affinity resins [152].

$$V'_R = V_R - V_M$$

Equation 15: Adjusted retention volume V'_R , with retention volume V_R and mobile phase volume (hold-up volume) V_M .

$$k = \frac{V'_R}{V_M} = \frac{t'_R}{t_M} = \beta \frac{Q}{C} = \frac{1 - R}{R}$$

Equation 16: Retention factor k , with the adjusted retention volume V'_R , the mobile phase volume V_M , the adjusted retention time t'_R , the hold-up time t_M , the phase ratio β , the amount of component in the stationary phase Q , the amount of component in the mobile phase C and the fraction of component in the mobile phase R . Note that in the literature, the retention factor k may also appear as k' .

$$k_{A/B} = \frac{[A]_S/[B]_S}{[A]_M/[B]_M}$$

Equation 17: Selectivity coefficient $k_{A/B}$, with the amounts of components A and B in the stationary $[A]_S$, $[B]_S$ and mobile phases $[A]_M$, $[B]_M$ respectively.

$$\alpha = \frac{V'_{R2}}{V'_{R1}} = \frac{t'_{R2}}{t'_{R1}} = \frac{k_2}{k_1}$$

Equation 18: Separation factor α , with the adjusted retention volume, retention time and retention coefficient of components 1 and 2, respectively: V'_{R1} , V'_{R2} , t'_{R1} , t'_{R2} , k_1 , and k_2 .

Even resins without ligands can affect the retention of proteins on a column: if the pore diameter allows a protein to enter, it will have a longer journey through the branched pores of the particles compared to a larger protein, unable to enter the pore, which can migrate around the particle in the buffer flow [153]. Therefore, small proteins will take more time to pass through the column. This effect is called gel filtration (GF) or size exclusion chromatography (SEC) (Equation 19) [149, 153].

$$V'_R = k_{SEC} V_I; k_{SEC} = 1 - \frac{2}{\sqrt{6\pi}} \frac{R}{d}; R \approx 0.81 M_r^{1/3}$$

Equation 19: Adjusted retention volume in SEC V'_R with SEC retention factor k_{SEC} , interstitial volume V_I (liquid volume outside the particles), protein diameter R , particle pore diameter $2d$ and protein molecular mass M_r .

Changing one of the factors influencing the strength of protein–ligand interactions will alter the protein binding/elution behavior and thus result in different selectivity coefficients and separation factors during chromatography. Typical changes include linear or stepwise salt or pH elution gradients, selecting a different pH for the running buffer, switching the buffering agent, adding mobile phase modifiers or choosing a different resin [149, 154, 155]. Even ligands with the same functional group can have different selectivities if the base matrix is distinct, e.g. Q Sepharose FF (GE, Uppsala, Sweden) and Q HyperCel (Pall, Port Washington, NY, USA) [156]. The purification of a target protein can therefore be improved by carefully adjusting the

chromatography conditions. Fewer impurities will bind when there is higher selectivity for the target protein, resulting in higher levels of purity.

Separative performance can be evaluated by comparing the separation factors of a target protein and major impurities under different chromatography conditions [157]. However, separation factors alone can be misleading because a pair of proteins with α values unequal to unity can still extensively overlap during elution if they have broad peaks (Figure III-5). Resolution and column efficiency are additional indicators that can help to overcome such ambiguities. The resolution R_S describes the separation of two peaks in terms of their average peak width at the base (Figure III-5 and Equation 20) [149].

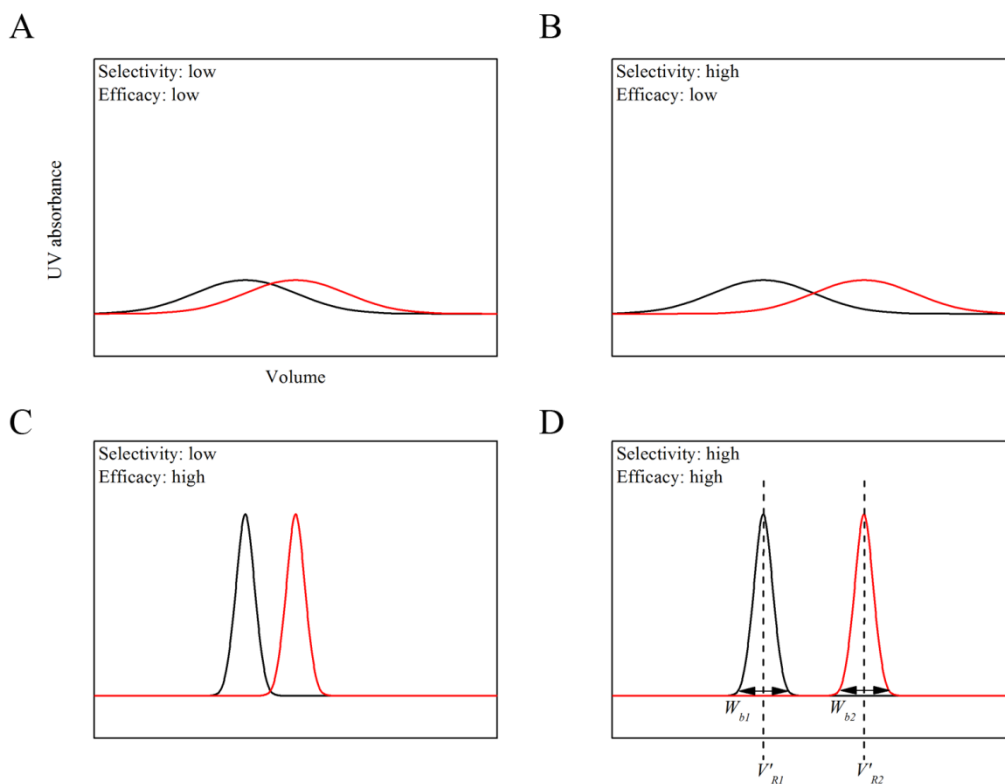


Figure III-5: Effect of ligand selectivity and column efficacy on protein separation by liquid chromatography.

A. At low selectivity and efficacy, the peaks are poorly resolved. B. Higher selectivity increases the difference in the (adjusted) elution volumes of two peaks $\Delta V'_R$ (Equation 15). C. Higher efficacy results in a narrower peak width W_b (Equation 20). D. High selectivity and efficacy results in good baseline separation.

$$R_S = \frac{V'_{R2} - V'_{R1}}{\frac{W_{b1} + W_{b2}}{2}} = \frac{2(V'_{R2} - V'_{R1})}{W_{b1} + W_{b2}} = \frac{\alpha - 1}{\alpha + 1} \cdot \frac{\bar{k}}{1 + \bar{k}} \cdot \frac{\sqrt{N}}{2}$$

Equation 20: Resolution R_S of two components as a function of peak shape and retention volume as well as separation factor α , with adjusted retention volumes of components 1 and 2 (V'_{R1} , V'_{R2} respectively), baseline peak width of components 1 and 2 (W_{b1} and W_{b2} respectively), mean retention factor \bar{k} and plate number N .

$$H = A + \frac{B}{u} + C \cdot u$$

Equation 21: Plate height H , with Eddy-diffusion A , longitudinal diffusion B , mass transfer kinetics for a solute between stationary and mobile phase C and the linear velocity u . Note that H is also referred to as height equivalent to one theoretical plate (*HETP*) (Figure III-6 C).

$$A_S = \frac{b}{a}$$

Equation 22: Peak asymmetry factor A_S , with a and b being the distance from the peak center towards the peak front and back respectively at 10% peak height.

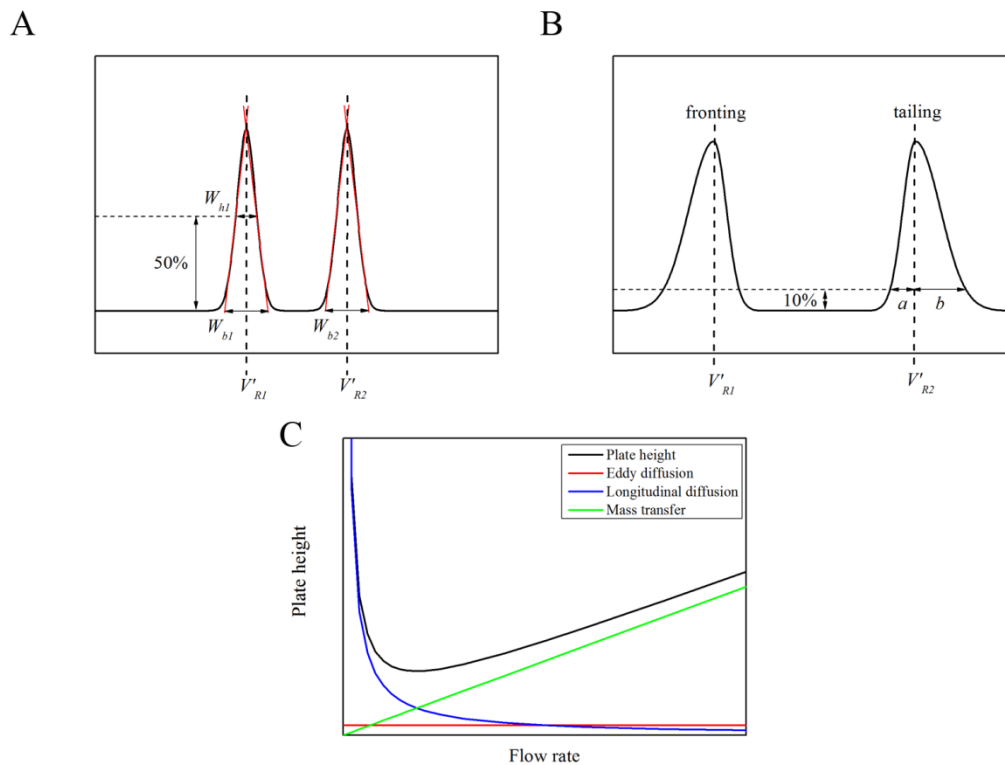


Figure III-6: Geometric indices used for peak characterization and the van Deemter equation.

A. The (adjusted) retention volume V'_R describes the volume at which an elution peak reaches its maximum. The peak width at base W_b and peak width at half height W_{hl} are a measure for peak broadening. B. The asymmetry A_S is described by the ratio of the width left and right of the peak center at 10% peak height (Equation 22). C. The impact of increasing the flow rate on plate height H , the output of the van Deemter equation (Equation 21) and its terms. The Eddy diffusion term A (red line) is independent of the linear velocity (flow rate) u and thus constant. The longitudinal diffusion B (band broadening, blue curve) decreases with increasing u while the mass transfer C (green line) is constrained by higher linear velocity.

$$N = \frac{L}{H} = \left(\frac{V'_R}{\sigma} \right)^2 = 16 \cdot \left(\frac{V'_R}{W_b} \right)^2 = 8 \cdot \ln(2) \cdot \left(\frac{V'_R}{W_{hl}} \right)^2$$

Equation 23: Number of theoretical plates N of a column, with column length L , plate height H , adjusted retention volume V'_R , standard deviation of the Gaussian peak σ , baseline peak width W_b , and peak width at half height W_{hl} . Note that $W_b = 4\sigma$.

$$\bar{k} = \frac{k_1 + k_2}{2}$$

Equation 24: Mean retention factor \bar{k} , with retention factors for components 1 and 2 shown as k_1 and k_2 respectively.

III.6.4 Models describing chromatographic separations

The plate model describes column efficiency in terms of the number of theoretical plates N [158-160]. This concept dates back to rectifying columns where equilibria were formed at each actual plate during the separation of liquid mixtures [161]. N can be calculated using Equation 23 if an elution peak has a Gaussian shape (Figure III-6 A) [157]. Any deviation from the Gaussian shape is expressed using the asymmetry factor A_S (Equation 22 and Figure III-6) [149]. An A_S larger than unity indicates peak “tailing” while a value below unity indicates “fronting”. Given a specific column (including resin), the plate number N can only be changed by altering the linear velocity u and thus the flow rate, because plate height H is calculated according to Equation 21, the Van Deemter equation [162]. Factor A in this equation is dependent on particle shape and diameter, and factor B is proportional to the diffusion coefficient of the mobile phase. Factor C also depends on this coefficient, and on the diffusion coefficient of the stationary phase and particle diameter. There is an optimal linear velocity u for which H adopts a minimal value as shown in Figure III-6 C.

The relation between resolution R_S and plate number N is given in Equation 20 [163]. The mean retention factor \bar{k} is defined in Equation 24. Resolution can be improved by increasing the values of α , \bar{k} or N . However, small changes in the separation factor will have a stronger effect on the resolution than changes in the plate number, thus α should be targeted when improving R_S . Protein elution profiles can be calculated according to the plate model using Equation 25 [158].

$$\frac{dC_{(i)}}{d\tau} = \frac{N_p [C_{(i-1)} - C_{(i)}] - C_{(i)} H \frac{dK[C_{(i)}, I]}{dI} \frac{dI}{d\tau}}{1 + H \left(K[C_{(i)}, I] + C_{(i)} \frac{dK[C_{(i)}, I]}{dC_{(i)}} \right)}$$

$$H = \frac{(V_c - V_M)}{V_M} = \frac{1}{\beta} ; \tau = \frac{t \cdot u}{V_M} ; X_0 = \frac{V_L}{V_M}$$

Equation 25: Calculation of the change in mobile phase protein concentration $dC_{(i)}/d\tau$ at plate i with the concentration difference dependent plate number $N_p[C_{(i-1)}-C_{(i)}]$, the mobile phase protein concentration $C_{(i)}$, the inverse phase ratio H , the protein concentration and ionic strength-dependent equilibrium constant $K[C_{(i)}, I]$ and the ionic strength gradient $dI/d\tau$. The initial and boundary conditions are: $C_{(1)}=C_{(2)}=\dots=C_{(N_p)}=0$ for $\tau=X_0$ and $C_{(0)}=C_0$ for $-X_0<\tau\leq 0$ and $C_{(0)}=0$ for $\tau>0$ respectively. The injected sample volume is V_L and the protein concentration in the sample is C_0 .

The rate model is a more fundamental alternative to the plate model, which calculates protein concentrations and elution profiles in the stationary and mobile phases over the whole column length and separation time by solving three partial differential equations (PDEs) [164]. The model is more costly to establish than the plate model and does not contain key figures that would facilitate the comparison of two chromatography setups, such as the number of theoretical plates N . Rate models differing in complexity and applications have been published over the last 60 years [164-171]. The first two PDEs of the rate model describe mass balances for the mobile and stationary phases (Equation 26) [164, 171]. The third equation describes adsorption and desorption kinetics (Equation 27).

$$-D_{bi} \frac{\delta^2 C_{bi}}{\delta Z^2} + v \frac{\delta C_{bi}}{\delta Z} + \frac{\delta C_{bi}}{\delta t} + \frac{3k_i(1 - \varepsilon_b)}{\varepsilon_b R_p} (C_{bi} - C_{pi,R=R_p}) = 0$$

$$(1 - \varepsilon_p) \frac{\delta C_{pi}^*}{\delta t} + \varepsilon_p \frac{\delta C_{pi}}{\delta t} - \varepsilon_p D_{pi} \left[\frac{1}{R^2} \frac{\delta}{\delta R} \left(R^2 \frac{\delta C_{pi}}{\delta R} \right) \right] = 0$$

Equation 26: Bulk mobile phase and stationary phase mass transfer equations with axial/radial dispersion coefficient D_{bi} , bulk mobile phase concentration C_{bi} , concentration of i in the stagnant fluid phase inside particles C_{pi} , adsorption saturation capacity C^∞ , concentration of i in the particle solid phase C_{pi}^* , effective diffusivity D_{pi} , film mass transfer coefficient k_i , adsorption and desorption rate constant k_{ai} and k_{di} , dimensional time t , radial coordinate for particle R , particle radius R_p , interstitial velocity v , axial coordinate Z , bed void volume fraction ε_b and particle porosity ε_p .

$$\frac{\delta C_{pi}^*}{\delta t} = k_{ai} C_{pi} \left(C^\infty - \sum_{j=1}^{N_s} C_{pi}^* \right) - k_{di} C_{pi}^*$$

Equation 27: Adsorption and desorption kinetics, same parameters as in Equation 26, with the number of components N_s . Note that if adsorption/desorption rates are faster than mass transfer, both sides of this equation adopt a zero value resulting in a Langmuir isotherm. The following initial and boundary conditions apply to these equations:

$$t = 0: C_{bi} = C_{bi}(0, Z) \text{ and } C_{pi} = C_{pi}(0, R, Z); Z = 0: \frac{\delta C_{bi}}{\delta Z} = \frac{v}{D_{bi}} [C_{bi} - C_{fi}(t)]$$

$$Z = L: \frac{\delta C_{bi}}{\delta Z} = 0; R = 0: \frac{\delta C_{pi}}{\delta R} = 0; R = R_p: \frac{\delta C_{pi}}{\delta R} = \frac{k_i}{\varepsilon_p D_{pi}} (C_{bi} - C_{pi,R=R_p})$$

Where the feed concentration of i is C_{fi} .

More equations have been developed since the late 19th century specifically to describe the binding and elution behavior of proteins on columns and thus predict their resolution. The behavior is commonly represented by sorption isotherms which characterize equilibriums at a constant temperature [172]. The simplest model assumes a linear relationship between equilibrium bulk protein concentration C_{eq} and the amount of protein bound to the resin q with the proportionality constant known as the linear coefficient, K_d (Equation 28 and Figure III-7).

$$q = K_d \cdot C_{eq}$$

Equation 28: Linear isotherm equation to calculate the amount of protein on the stationary phase q depending on the equilibrium mobile phase protein concentration C_{eq} with the linear coefficient K_d .

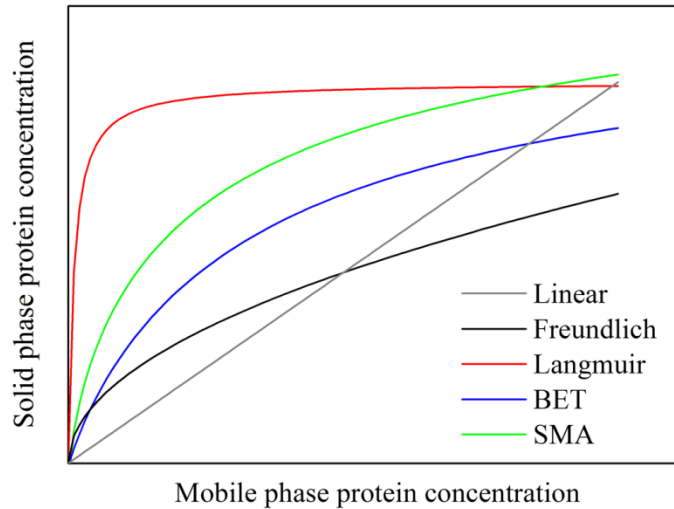


Figure III-7: Protein sorption isotherms calculated according to different models (Equation 28 to Equation 32).

The linear (gray) and Freundlich (black) isotherms do not account for saturation effects at high protein concentrations, so the predicted capacity for the resin is unlimited. Langmuir (red) and BET (blue) isotherms converge to a maximum, q_{max} , but differ in the proposed binding mechanism and thus in the steepness of the isotherm slope at low protein concentrations. BET and SMA (green) isotherms have a similar shape but the SMA isotherm is advantageous because it links maximum binding capacity to actual physical properties and can account for varying salt concentrations.

$$q = K \cdot C_{eq}^n$$

Equation 29: Freundlich isotherm to calculate the amount of protein on the stationary phase q depending on the equilibrium mobile phase protein concentration C_{eq} with the Freundlich coefficient K_f and the Freundlich exponent n . The exponent can also be written as $1/n$ if values of n above unity are preferred for convenience.

$$q = \frac{K_L q_{max} C_{eq}}{1 + K_L C_{eq}}$$

Equation 30: Langmuir isotherm to calculate the amount of protein on the stationary phase q depending on the equilibrium mobile phase protein concentration C_{eq} with the Langmuir sorption coefficient K_L and the maximum value for the mass of adsorbed protein q_{max} .

$$q = \frac{K q_{max} C_{eq}}{(C_{sat} - C_{eq}) \cdot \left(1 + \frac{(K-1)C_{eq}}{C_{sat}}\right)}$$

Equation 31: BET model to calculate the amount of protein on the stationary phase q depending on the equilibrium mobile phase protein concentration C_{eq} with the sorption coefficient K , the maximum value for the mass of adsorbed protein q_{max} and the protein solubility C_{sat} .

$$q = C_{a,eq} K_a \left(\frac{\Lambda - (v_a + \sigma_a)q}{C_s} \right)^{v_a} ; v_a = \frac{z_p}{z_s}$$

Equation 32: SMA formalism to calculate the amount of protein on the stationary phase q depending on the equilibrium mobile phase protein concentration $C_{a,eq}$ with the mobile phase salt concentration C_s , the equilibrium constant K_a , the resin total capacity Λ , the characteristic charge v_a and the steric factor of the protein σ_a . The characteristic charge v_a is the ratio of protein binding charge z_p and counter-ion charge z_s .

The obvious error in this model is the predicted infinite capacity of the resin with increasing protein concentration. This error is addressed in the empirical model by Herbert Freundlich using an additional exponent for C_{eq} and an adjusted proportionality constant K_f (Equation 29) [173]. However, introducing this modification does not correct the error, it only reduces its magnitude. Instead, the Langmuir isotherm has an explicit maximum value for the mass of adsorbed protein q_{max} and is thus able to describe the saturation of resin binding sites with protein (Equation 30) [174]. The Langmuir equation is similar to the Monod and Michaelis-Menten kinetics for microbial growth and enzyme kinetics, respectively [175, 176]. The Langmuir model is frequently used for chromatography but several assumptions are made that limit its predictive power, e.g. that bound molecules will not affect each other. Therefore, several modifications exist to adapt the model for multi-component [177-180] and cooperative adsorption [181] (Equation 31). An important expansion of the Langmuir model is the steric mass action (SMA) formalism for ion-exchange chromatography (IEC), which incorporates the dependence of protein adsorption on the mobile phase salt concentration [182] (Equation 32). SMA parameters determined for a protein are therefore valid for different salt concentrations whereas parameters for each salt concentration must be determined in the “standard” Langmuir model. The key concept of the SMA formalism is the introduction of a steric factor σ_a accounting for potential resin binding sites shielded by bound proteins in addition to the resin sites involved in protein binding described by the characteristic charge v_a (Equation 32). The formalism can also be adapted for multi-component adsorption and its parameters can be determined by the chromatographic analysis of single, pure proteins (Table III.3). Additional approaches for parameter determination have been published recently [183].

Table III.3: Methods to determine SMA parameters for proteins of interest according to [182].

Parameter	Method	Reference
Bed capacity Λ	Frontal analysis	[184]
Characteristic charge v_a	Multiple linear gradients	[185]
Equilibrium constant K_a	Multiple linear gradients	[186]
Steric factor σ_a	Breakthrough volume	[182]

$$q = C_{a,eq} K_a \left(\frac{\Lambda}{c} \right)^n \left(1 - \frac{(n - \sigma_a)q}{\Lambda} \right)^n \gamma_a ; \gamma_a = \frac{\alpha_a \cdot \alpha_L^n}{\alpha_{aL}} ; a + nL \rightleftharpoons aL_n$$

Equation 33: SMA formalism adapted for HIC to calculate the amount of protein on the stationary phase q depending on the equilibrium mobile phase protein concentration $C_{a,eq}$ with the overall molar concentration c , the equilibrium constant K_a , the total ligand concentration Λ , the number of ligands involved per bound protein molecule n , the steric factor of the protein σ_a , and the activity coefficient γ_a . Activity coefficient γ_a is defined as the ratio of protein activity α_a and ligand activity α_L to the protein-ligand complex activity α_{aL} .

$$q = C_{a,eq} K_a \left(\frac{\Lambda_{IEC} - (z_p + \sigma_a)q}{z_s C_s} \right)^{v_a} \left(\frac{\Lambda_{HIC} - (n + \delta_a)q}{c} \right)^n \gamma_a ; v_a = \frac{z_p}{z_s}$$

Equation 34: Mixed-mode isotherm to calculate the amount of protein on the stationary phase q depending on the equilibrium mobile phase protein concentration $C_{a,eq}$, the equilibrium constant K_a , the mobile phase salt concentration C_s , the overall molar concentration c , the ionic capacity Λ_{IEC} , the ligand density Λ_{HIC} , the characteristic charge v_a , the stoichiometric coefficient of the ligand n , the steric hindrance factor for charged ligands of the protein σ_a , the steric hindrance factor for hydrophobic ligands δ_a and the activity coefficient γ_a as defined above. The characteristic charge v_a is the ratio of protein binding charge z_p and counter-ion charge z_s

Similar models have been established for HIC [187-189] (Equation 33) and MMC [190] (Equation 34) or generalized for all types of adsorptive chromatography [191] (Equation 35). However, determining all parameters precisely, such as the number of ligands n involved in the binding of each protein molecule, is tedious using these systems. Therefore, non-mechanistic equations have also been developed and fitted to experimental data [192].

$$q = K_a \Gamma_a C_{a,eq} \left(\frac{c}{C_s} \right)^{v_a \beta_a \sigma} \left(\frac{q_L}{c} \right)^{v_a} ; \sigma = \frac{z_L}{z_s} ;$$

$$P_a^{z_a} + v_a L^{z_L} S_{\sigma}^{z_s} \rightleftharpoons P_a^{z_a} L_{v_a}^{z_L} S_{\sigma v_a (1 - \beta_a)}^{z_s} + v_a \beta_a \sigma S^{z_s}$$

Equation 35: Generalized formula to calculate the amount q of protein (P) bound to ligands (L) on the stationary phase displacing salt (S) depending on the equilibrium mobile phase protein concentration $C_{a,eq}$ with the mobile phase salt concentration C_s , the equilibrium constant K_a , the equilibrium excess energies Γ_a , the overall molar (protein) concentration of the mobile phase c , the concentration of free ligands q_L , stoichiometric coefficient v_a (number of ligands involved in protein binding), interaction mode determining factor β_a ($\beta = 0$ for HIC, $\beta = 1$ for IEC) and the charge ratio σ , with the charge of each protein species z_a , ligand charge z_L and the charge of the salt counter ion z_s .

III.6.5 Prediction of chromatographic separation

III.6.5.1 Motivation and isoelectric point

A major drawback of all the model approaches discussed above is that the determination of chromatography parameters, such as SMA parameters, minimally requires several separation runs using pure protein [182, 193]. Pure protein samples are rarely available at an early stage of process development for a given POI and are even more difficult to obtain for all HCPs. Additionally, performing the separation runs is laborious and time consuming. Therefore, several *in silico* methods have been developed to predict protein adsorption behavior on different chromatography resins while avoiding the need to determine chromatography

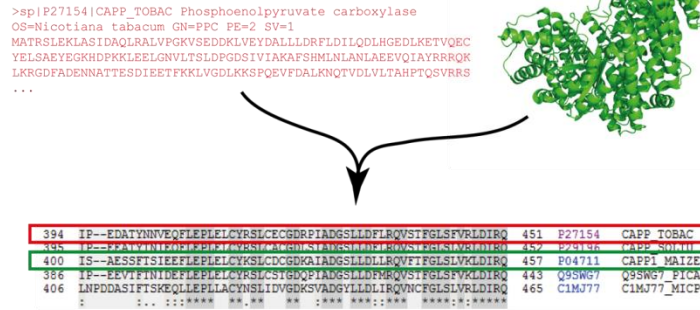
parameters for each protein. One of the simplest approaches uses a theoretical isoelectric point (IEP, pI) in IEC [194]. The pI is calculated for a defined mobile phase pH based on the pK_a values of all amino acids of a protein. However, this method does not account for the different leverage of amino acids buried within a protein compared to those on the protein surface.

III.6.5.2 Homology modeling of protein structures

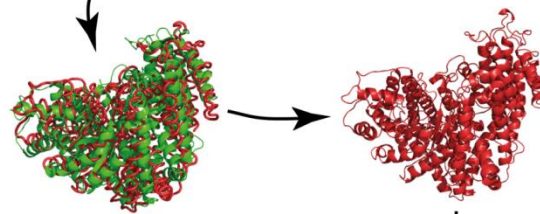
In order to take the steric effects of amino acid positions into account, the three-dimensional (3D) structure of the POI must be known. Experimental methods to determine protein structures include nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography [195, 196]. However, both methods require pure preparations of the POI and HCPs, raising the same problems discussed above. Homology modeling is used to circumvent this limitation by generating 3D models for uncharacterized proteins based on the structures of homologous molecules. This method involves three major steps: (i) template selection and alignment, (ii) model building and (iii) model evaluation (Figure III-8) [197] (note that all following statements in this paragraph are based on this reference). In the first step, algorithms such as BLAST are used to identify proteins with a sequence identity $\geq 30\%$ to the target protein. Sequence(s) with the highest identity to the target and with solved 3D structures are then selected as a template for modeling. Three major approaches exist for the second, model-building step. First, rigid body assembly can be used to gradually increase the level of ‘detail’ in the model starting with an alignment of the c_α atoms of core folds and adding the remaining backbone atoms of these regions, then including loops and finally side chain atoms before energy minimization. Second, segment matching algorithms select all-atom peptide sequences of about six residues based on guiding positions, usually c_α , to reconstruct a homology model iteratively through the sequence of the template. Third, a broad set of structural constraints and restraints can be derived from the template structure, such as angles and bond lengths, and applied to each corresponding atom in the target sequence. For the final model, the sum of all constraint/restraint violations is then minimized. The advantage of this last approach is that different data sources can be exploited to build up the set of constraints/restraints (e.g. NMR, atomic force microscopy (AFM) and cross-linking experiments). Once a model has been built, several structural parameters are hierarchically assessed to determine the model quality. First, the overall fold is confirmed by evaluating the sequence identity of target and template over the whole sequence range, to identify low-quality regions. Any identity below 30% can be challenging because a relationship between sequence and structure is difficult to establish. Then the stereochemistry of the model is checked by investigating bond lengths, angles and correct

folding indicators such as the formation of a hydrophobic core and accessible surface area (ASA). These indicators are evaluated using different scoring functions (Table III.4).

A Select templates and align



B Build homology models



C Evaluate models

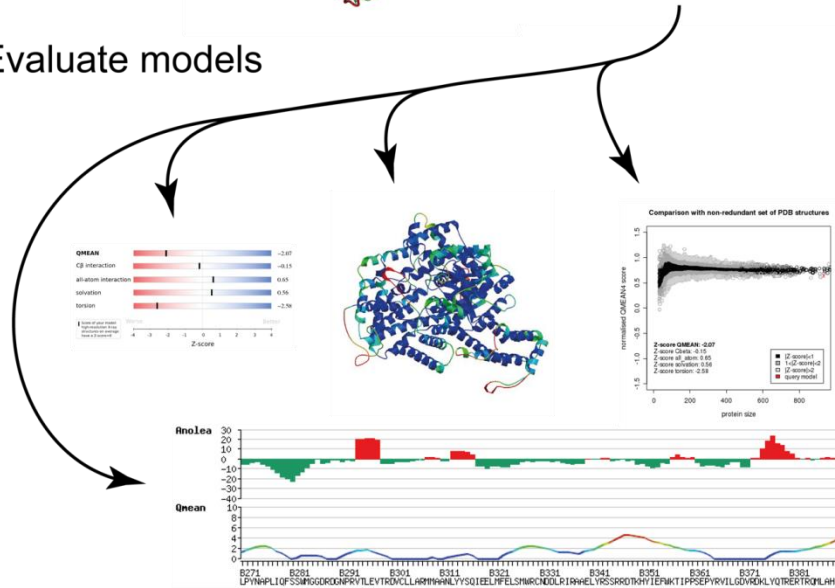


Figure III-8: Common workflow for the homology modeling of protein structures.

A. Potential templates (green protein) are identified by alignment with the target protein sequence (red). Available template structures are evaluated based on sequence identity (red vs. green, boxed sequences). B. Homology models are built by aligning the target protein sequence with the structure of the most suitable template (red string vs. green cartoon) and subsequent refinement of the homology model (red cartoon). C. Finally, homology models are evaluated based on scoring functions such as QMEAN-Z (Table III.4) that compare structural properties of the model, such as binding angles, with those of structures solved by X-ray diffraction or NMR spectroscopy (Figures in panel C taken from SWISS-Model evaluation [198]).

Homology models generated and evaluated in this way can then be used as the input data for other *in silico* methods for the prediction of protein retention and separation during chromatography, which are more sophisticated than *pI* prediction, e.g. molecular dynamics

(MD) simulations or quantitative structure-activity relationship (QSAR) models. Several online tools have been developed recently that facilitate the process of model building and evaluation, including the SWISS-MODEL and QMEAN servers [199, 200].

Table III.4: Scoring functions used to evaluate homology models according to the QMEAN server [199].

Scoring function	Description
Torsion ¹	Extended torsion potential over three consecutive residues. Bin sizes: 45° for the center residue, 90° for the two adjacent residues.
Pairwise ¹	Residue-level, secondary structure specific interaction potential using C β atoms as interaction centers. Range 3...25 Å, step size: 1 Å
Solvation ¹	Potential reflecting the propensity of a certain amino acid for a certain degree of solvent exposure approximated by the number of C β atoms within a sphere of 9 Å around the center C β .
All_atom ¹	All-atom, secondary structure specific interaction potential using all 167 atom types. Range 3...20 Å, step size: 0.5 Å
SSE_agree	Agreement between the predicted secondary structure of the target sequence (using PSIPRED) and the calculated secondary structure of the model (using DSSP).
ACC_agree	Agreement between the predicted relative solvent accessibility using ACCpro (buried/exposed) and the relative solvent accessibility derived from DSSP (>25% accessibility → exposed)
Z-score	Estimate of the absolute quality of a model by relating it to reference structures solved by X-ray crystallography. The QMEAN Z-score is an estimate of the "degree of nativeness" of the structural features observed in a model by describing the likelihood that a model is of comparable quality to high-resolution experimental structures

¹ A linear combination of these scores is the QMEANscore4 used in SWISS-MODEL. DSSP - *Define Secondary Structure of Proteins* algorithm [201].

III.6.5.3 Molecular dynamics simulation

MD simulations numerically solve the motion of molecules, like proteins and ligands, in space over time using a predefined environment and force field [202]. A force field, such as AMBER or CHARMM, provides information about the forces that act upon the atoms/molecules present in the simulated environment (Equation 36) as well as parameter sets for the different atoms/atom groups such as VDW radius, partial charge or mass [203, 204]. According to these forces, the changes in velocity and position are calculated for each atom/molecule in the simulation. For this purpose the 3D structure of each molecule has to be known, e.g. as a pdb-file. The predefined environment comprises the 3D size (the "box", typically 10 x 10 x 10 nm), the number of each molecule species, the duration and time increment as well as the solvent properties (explicit or implicit) of the simulated system [205]. An implicit solvent is represented by a fixed parameter such as the relative permittivity (dielectric constant), which is 80.10 [-] at 20°C for water [206]. In contrast, each solvent molecule is simulated individually for explicit solvents increasing both simulation accuracy, especially for molecule interactions, and the required computational power.

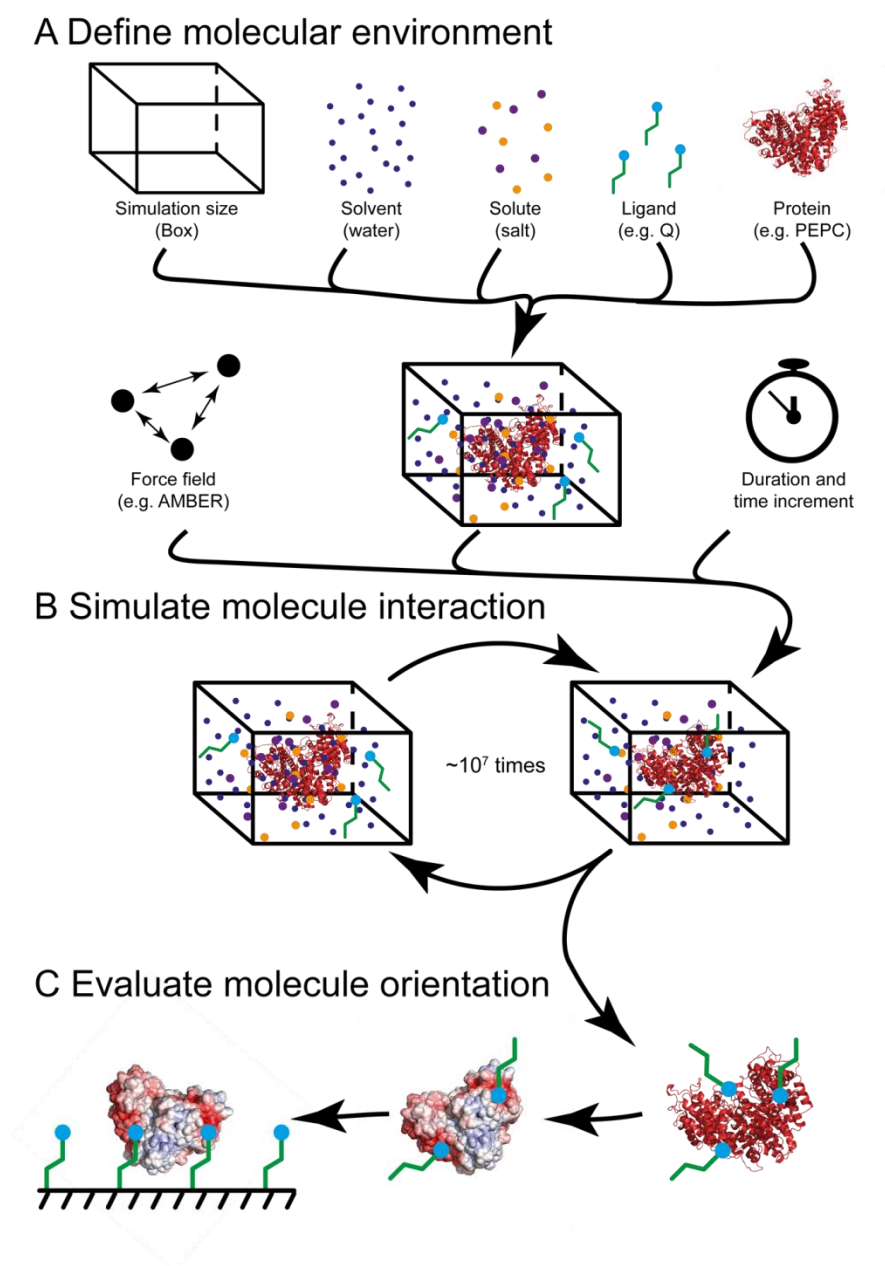


Figure III-9: Workflow of a MD simulation.

A. First, the size of the molecular environment is set, the components to be modeled are defined and the force fields, time increment and duration are selected. B. The forces are then calculated using an iterative process based on a force field, and the molecules are moved accordingly. C. Finally, the relative positions of the protein and ligand are evaluated to localize interaction spots and predict the binding orientation of the protein on a ligand-coated surface.

The simulation itself is an iterative process as shown in Figure III-9. A small time increment, typically about one femtosecond (10^{-15} s), and a total simulation time of 10–100 nanoseconds (10^{-9} s), matching the kinetics of the reaction/interaction to be studied, are crucial for reliable calculations [207, 208]. As a consequence, forces and coordinates must be calculated ~ 10 million times for each atom/molecule in the simulation. In combination with an explicit solvent this adds up to a significant demand on computational power for each simulation [209, 210]. This is especially true for the evaluation of chromatographic separations

because all conditions such as salt concentration and pH must be calculated separately. The parameter values in current force fields are also based on experimental data and can thus be flawed, which is challenging because of potential error progression in the numerous simulation steps [211]. Therefore, MD simulations are currently inadequate for the prediction of competitive protein binding to ligands from crude solutions given available computational power. Instead this method is currently limited to single protein-ligand interactions [212].

$$E_{total} = E_{bonded} + E_{nonbonded} = (E_{bond} + E_{angle} + E_{dihedral}) + (E_{elec} + E_{VDW})$$

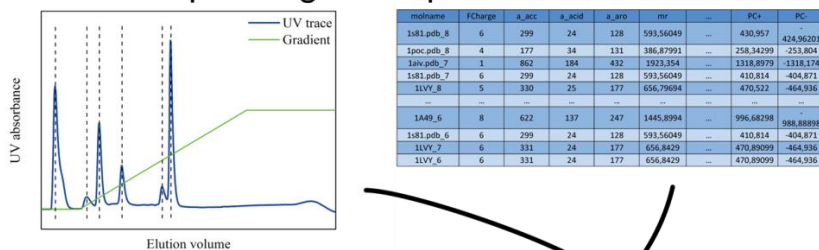
Equation 36: Composition of the total potential energy of a system E_{total} described by a force field including the potential energies for covalent bonds E_{bond} , bond angles E_{angle} , dihedral angles $E_{dihedral}$, electrostatic interactions E_{elec} , and VDW interactions E_{VDW} .

III.6.5.4 Quantitative structure activity relationship modeling

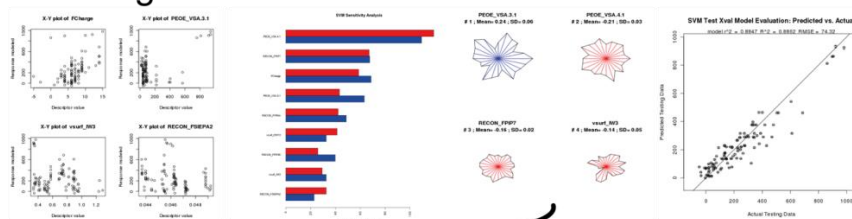
QSAR modeling is an alternative approach to predict protein retention during column chromatography that requires less computational power than molecular dynamics simulation. The method can be divided into three major parts (Figure III-10) [213]. First, a training data set is established, which consists of the retention factors for a set of model proteins under different chromatographic conditions and resin types, and descriptors for this set of model proteins. Descriptors are protein properties, such as ASA, charge and diameter, which are calculated based on the 3D structure of each individual model protein [213]. Second, a set of regression models is built that correlates retention factors with descriptor values. Algorithms commonly used to select descriptors with significant correlation to protein retention include multiple linear regression (MLR), partial least squares (PLS) and support vector machines (SVM) [214, 215]. The individual models are then critically evaluated to ensure reliable predictions for unknown proteins [214]. Over-fitting, including non-significant descriptors in the model, is a major concern because the number of descriptors is often much larger than the number of proteins in the dataset [216]. Evaluation methods include y-randomization, leave-many-out (LMO) analysis, bootstrapping and r^2 analysis [214, 217, 218]. Additionally, QASR models should only be used within their applicability domain to perform predictions and prevent unjustified extrapolations [213]. For example, if all proteins in a training data set had a small mass (<50 kDa) no attempt should be made to use the established model for predictions of proteins with a high mass (>150 kDa). This also holds true for other chromatography conditions such as pH and resin type. In a third step, the regression model with the best performance is selected to predict retention factors for proteins that have not been included in the training data set, but for which 3D data are available and descriptors have been calculated, allowing the outcome of a chromatographic separation to be predicted. However, retention factors may vary in practice

when crude mixtures are used as the feed stream because competition in binding is not considered during modeling, and training data are often generated using pure proteins.

A Experimental retention data and corresponding descriptors



B Descriptor selection, model building and evaluation



C Prediction of retention data for unknown proteins

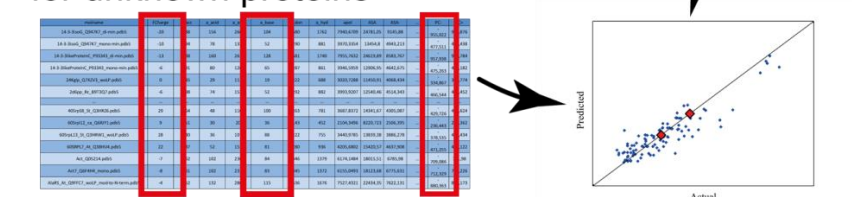


Figure III-10: Workflow for building a chromatography retention model by QSAR.

A. Descriptors are calculated for proteins from an experimental retention data set. B. Descriptors correlated to the observed chromatographic retention are selected by algorithms such as SVM and the quality of the resulting model is evaluated, e.g. by y-randomization. C. The final model is used to predict chromatographic retention of other proteins for which the same set of descriptors has been calculated based on their known/modeled structure. Graphs in panel B were generated with YAMS webserver [219].

Molecular docking is another method to predict molecular interactions that combines aspects of MD simulation and QSAR modeling. Either force field calculations or surface shape descriptors are used to “dock” two molecules to each other, e.g. protein and ligand [220]. A scoring function is then used to evaluate the most likely interaction sites. However, the main purpose of this method is to identify preferred binding orientations. It may therefore fail to

predict multi-site and weak interactions that are essential for protein retention during chromatography.

In summary, the most promising approaches to generate models for protein chromatography are (i) SMA formalism to determine Langmuir isotherms, and (ii) QSAR modeling. These use experimental and *in silico* analysis, respectively, and can help to accelerate the development of downstream processes through guided experiment design. They can also be used as part of a QbD strategy for the cGMP-compliant production of biopharmaceuticals.

III.6.6 Economic and regulatory considerations for downstream process design

Recent data suggest that DSP accounts for a major part of the production costs when plants are used to produce biopharmaceutical proteins [63]. On one hand this is because upstream costs are lower for plant-based production compared to other expression systems and DSP costs thus account for a larger share of the total costs [34, 37]. On the other hand plant secondary metabolites and polymers may cause problems during processing by clogging equipment due to oxidization and precipitation, so actual DSP costs may be higher [73]. Several additives have been used to deal with these problems as discussed above (III.6.1.2 and III.6.1.4). Using new additives or combining them can help to reduce DSP costs further. Recombinant proteins are often extracted from within the plant cell, releasing abundant HCPs into the extract. This is not the case if the culture supernatant is processed without cell lysis (e.g. when monoclonal antibodies are produced using CHO cells) [95, 221]. This issue can be addressed if inexpensive chromatography resins can be used in flow-through mode to bind protein impurities, which lowers the HCP burden on the more expensive capture resin and allows the use of smaller columns.

The use of disposable equipment can also reduce the initial investment costs for a production process although this increases the costs of consumables [96]. Disposable technology is therefore advantageous for low-margin, high-value products and multi-purpose production sites [96]. DSP costs can also be reduced by using as many generic unit operations as possible for every process, because consumables can be ordered in bulk to reduce the unit price. The inclusion of generic operations also accelerates process development, leaving more time to optimize critical process steps hence reducing costs by improving product yield and/or purity.

In addition to actual DSP, there are also costs associated with aseptic filling and pharmaceutical formulation [222], the latter aiming to provide an optimal delivery vehicle for the API, such as tablets, capsules, creams or solutions [223]. Optimization criteria for

formulation include shelf life, bioavailability, activity and dosage [224]. Several excipients are used for this purpose, including fillers, binders and disintegrators [224]. Adjuvants are a special kind of additive relevant in vaccine formulations [225]. The final formulation is known as the medicinal product. If possible, the final step in DSP should be adapted to the first step in formulation to avoid superfluous handling operations. This might be difficult because the final formulation is often developed during phase II/III clinical trials by which time the DSP is largely irreversible.

Regulatory principles must also be taken into account during process development, reflecting the demands of current good laboratory practice (cGLP), good manufacturing practice (cGMP) and good clinical practice (cGCP) as set by the International Conference on Harmonisation of Technical requirements for Registration of Pharmaceuticals for Human Use (ICH) to improve safety standards for medicinal products [226]. Mandatory aspects of these guidelines include the need for comprehensive documentation, quality assurance (QA) and quality control (QC) at each step of the biopharmaceutical manufacturing process. DSP must address process- and product-related impurities, the latter defined as API precursors or degradation products lacking the necessary activity, efficacy or safety [227]. High-resolution techniques such as SEC may be needed to remove such impurities, e.g. multimers in the case of monoclonal antibodies [137].

Process-related impurities have diverse origins. Some arise from the production platform. For example, plant cell DNA and the enzyme ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO) are among the most abundant process-related impurities in plant-based systems [226]. Such impurities are present in the bulk extract and can be depleted by precipitation and removed by polishing chromatography steps. Their presence can be monitored by immunochemical methods including the enzyme-linked immunosorbent assay (ELISA) [228]. Other process-related impurities originate from the processing equipment and chemicals, e.g. sulfate residues from buffers, leached Protein A from chromatography resins, and plasticizers from tubing. The latter are characterized as ‘extractables’ if they are only released under harsh conditions and ‘leachables’ if they are also released under normal process/storage conditions [229]. The detection of extractables/leachables can be challenging because they may be diverse, unknown, present at low concentrations and quantitative assays may be unavailable. Total organic carbon (TOC) is often used as an indicator for the presence of extractables/leachables and reversed-phase high-performance liquid chromatography (RP-HPLC) may be suitable for detection and analysis [230]. However, extractables and leachables are often separated from the product by chromatography due to their low molecular mass and

charge. High-quality materials with a low extractables/leachables burden can improve process safety but this will increase costs. Finally, process-related impurities can also include cross-contamination from previous batches (in the case of reusable equipment such as homogenizers and all stainless steel equipment that has not been cleaned sufficiently). The easiest and safest way to avoid cross-contamination is to use disposable technologies wherever possible (compare III.6.1.4) [97].

The removal of human pathogens from the product is a key regulatory requirement. Spiking experiments are used to determine the ability of a process to remove certain indicator pathogens, such as *Canine parvovirus* and *Herpes simplex virus* (HSV), because the abundance of pathogens is usually low from the beginning [231-233]. The ability to remove these pathogens is expressed as a logarithmic reduction factor (LRF), and LRFs of up to 16 may be necessary for processes connected to upstream production using mammalian cells [232]. For plant expression systems, lower LRFs may be sufficient because plants do not support the growth of human pathogens [64]. If a process is not capable of effective pathogen removal, specific inactivation steps must be introduced including low-pH treatment steps [231].

In summary, the purity of a biopharmaceutical product is not solely a technical objective but economic and regulatory constraints also apply.

III.7 Design of experiments

The outcome of a process can be predicted if the parameters with the most significant impact are identified and kept within defined ranges. These significant parameters are said to span the so-called design space of a process [234, 235]. Such a process can be modeled if the information from different data points (experiments) within the design space is combined. The classical approach for the selection of data points is to vary one factor at a time (OFAT). However, this concentrates the data points along lines within the design space leaving the remainder beyond consideration (Figure III-11 A) [236]. Accordingly, the resulting models can be flawed and interactions between factors are unlikely to be found, resulting in poor predictions and suboptimal processes.

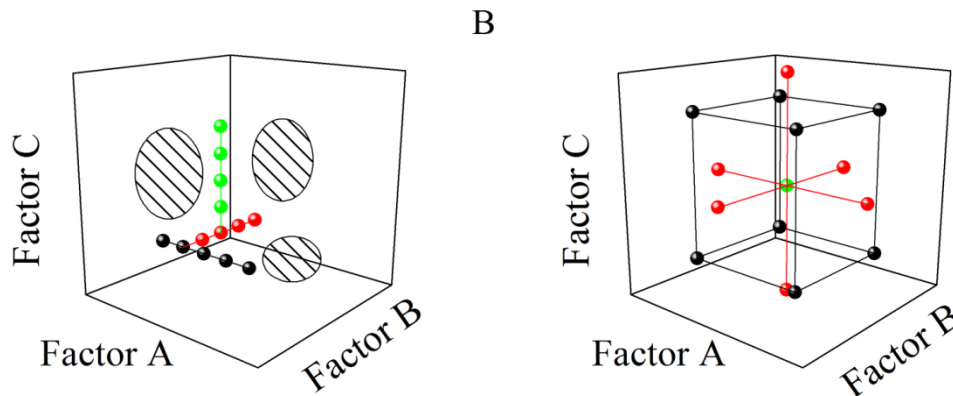


Figure III-11: Comparison of OFAT and DoE designs.

A. Experiments of an OFAT design resemble pearls-on-a-string if projected into the design space, leaving wide areas (hatched ovals) for which no information is available. B. In a DoE approach, experiments are distributed evenly throughout the design space increasing the obtained information about the system and likelihood to find a desired operation point.

In contrast, a statistical design of experiments (DoE) provides information about the entire design space. This is achieved most effectively using specialized software packages that allow the user to define the problem as well as constraints affecting the factors that are tested. The idea behind a statistical DoE is to scatter the data points (experiments) evenly throughout the design space under investigation and thus collect the maximum amount of information with the minimum effort [236, 237]. Parameters included in the design space are called *factors* and may be varied, whereas those excluded must be kept constant at all times. There are *categoric* and *numeric* factors. Categoric factors are subdivided into *nominal* factors, such as different colors, and *ordinal* factors such as the leaves on a plant. Numeric factors are normally *continuous*, like concentration or length measurements, but practical considerations may limit them to *discrete* values, e.g. if only a limited number of temperatures can be tested. The measured variable is called a *response*, e.g. filter lifetime or extract turbidity.

Factor ranges considered in a DoE are often determined in initial experiments or known from experience or the literature. They can also be determined or narrowed down by full or fractional factorial designs. Categoric factors can be delimited by irregular factorial designs. Two-level factorial (TLF) designs are often selected for screening purposes, e.g. when many numeric factors are investigated. Each of these factors then adopts one of two possible values in every experiment so the mean (linear) effect of each factor can be assessed (Equation 37) [238]. Selecting a meaningful range for each factor is crucial because otherwise an important effect may remain concealed, as shown in Figure III-12. This problem can be avoided by augmenting a TLF with additional central and optional star point experiments, forming a so called central-composite design (CCD) [238]. A CCD can also estimate the non-linear effects

of a factor allowing the construction of a so-called response surface model (RSM). Using an optimal design algorithm, the number of experiments required to build a RSM can be reduced compared to a CCD. The three major optimality criteria are listed in Table III.5, but IV-optimal algorithms provide a model that has an evenly-distributed prediction quality throughout the design space and is thus ideal for the evaluation of process robustness. There are also several other design types for specific applications such as mixture plans [239].

$$\text{Effect of factor } A = \frac{\sum_{i=1}^k y_i}{k} - \frac{\sum_{j=1}^l y_j}{l}$$

Equation 37: Effect of factor A as the difference in average response at high factor level y_i and low factor level y_j with k and l representing the number of experiments with high and low factor levels respectively. In a full-factorial design, $k = l$.

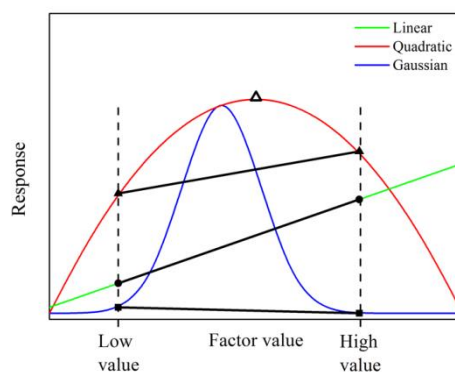


Figure III-12: The importance of meaningful factor range selection.

In a linear factor-response relationship (green line) the effect of the factor is correctly estimated with a TLF (circles). However, for a quadratic correlation (red curve) the optimal value is not captured by a TLF (triangles) and additional experiments are required (e.g. in a CCD, open triangle). In narrow dose-response relationships (blue Gaussian shaped curve) the effects can remain concealed due to the inadequate selection of factor levels (squares).

Table III.5: Criteria for the generation of optimal DoE algorithms commonly used in RSM designs.

Design optimality	Criterion
A	Minimize the average prediction variance of the model
D	Maximize accuracy of model coefficient estimates
IV	Minimize prediction variance of the model across the entire design space

An important advantage of the statistical DoE approach over the OFAT approach is that interactions between factors are more likely to be found, although data analysis is dependent on specialized software and care must be taken to avoid misinterpretations based on missing or incorrect data. Common indicators for model quality are r^2 (Equation 38), adjusted r^2 and predicted r^2 (Equation 39 and Equation 40) [240]. The inclusion of additional factors in the model is penalized in the adjusted r^2 calculation by reducing the *degree of freedom* (df) and thus the adjusted r^2 value so that only factors which substantially increase the model quality will also increase the adjusted r^2 . This indicator can therefore be used to detect and prevent model over-fitting. The predicted r^2 is similar to a leave-one-out analysis and thus measures the

predictive power of a model based on the predicted residual sum of squares (PRESS, Equation 41) [240]. Instruments to identify low-quality models (section III.6.5.4) may not be suitable to evaluate a DoE because data redundancy is low by design and thus LMO analysis is likely to result in models with artificially low quality. Alternatively, lack-of-fit (LOF) tests can be used to correlate the pure error of replicate experiments with the deviation of model and experimental data (Equation 42) [240-243]. A significant LOF indicates a low-quality model or problems with the raw data. Indicators such as normal plots of the residuals, predicted-versus-actual and residuals-versus-run plots allow the non-quantitative evaluation of a model [240].

$$r^2 = 1 - \left[\frac{SS_{residual}}{SS_{residual} + SS_{model}} \right]$$

Equation 38: R-squared (r^2) calculation, with residual sum of squares $SS_{residual}$ and model sum of squares SS_{model} .

$$adj. r^2 = 1 - \left[\frac{\left(\frac{SS_{residual}}{df_{residual}} \right)}{\left(\frac{SS_{residual} + SS_{model}}{df_{residual} + df_{model}} \right)} \right]$$

Equation 39: Adjusted r^2 calculation, with residual sum of squares $SS_{residual}$, model sum of squares SS_{model} , degrees of freedom of the residuals $df_{residuals}$ and degrees of freedom of the model df_{model} .

$$pred. r^2 = 1 - \left[\frac{PRESS}{SS_{residual} + SS_{model}} \right]$$

Equation 40: Predicted r^2 calculation, with predicted residual sum of squares (PRESS, Equation 41), residual sum of squares $SS_{residual}$ and model sum of squares SS_{model} .

$$PRESS = \sum_{i=1}^n (e_{i,-i})^2 = \sum_{i=1}^n \left(\frac{e_i}{1 - h_{ii}} \right)^2 ; e_{i,-i} = y_i - \hat{y}_{i,-i} = \frac{e_i}{1 - h_{ii}}$$

Equation 41: Predicted residual sum of squares (PRESS) as the sum of predicted residuals $e_{i,-i}$, with the residuals e_i , the diagonal element of the hat matrix (H) h_{ii} , and the actual and predicted values for each point y_i and \hat{y}_i respectively. The suffix “i,-i” indicates that the calculation was performed for the i^{th} element of a data set but excluding that specific element from the calculation [244, 245].

$$SS_{residuals} = SS_{LOF} + SS_{pure\ error} ; F = \left[\frac{\left(\frac{SS_{LOF}}{df_{LOF}} \right)}{\left(\frac{SS_{pure\ error}}{df_{pure\ error}} \right)} \right]$$

Equation 42: F-test for lack of fit with residual sum of squares $SS_{residual}$, lack of fit sum of squares SS_{LOF} , pure error sum of squares $SS_{pure\ error}$, degrees of freedom of the lack of fit df_{LOF} and degrees of freedom of the pure error $df_{pure\ error}$.

Using DoE, the robust analysis of complete datasets results in accurate predictions of optimal parameter settings within the design space covered by the model. Once a good model is established, factors (process parameters included in the model) can be adjusted to accommodate desired responses (outcomes), e.g. a shorter process time or a higher yield. By assigning suitable weighting terms to two or more different responses, the factor setting can be

trimmed as required. In a subsequent step, the process model can be combined with a cost function [246, 247]. The resulting process cost model then allows the process to be optimized for different cost positions, e.g. overall product costs, consumables costs or downstream costs. Recent reports indicate that applying a DoE modeling approach to highly complex systems such as competitive protein binding during chromatography can be challenging [248].

IV. Materials and methods

IV.1 Equipment and chemicals

All equipment is listed together with the manufacturers' information in the appendix, section IX.1. All chemicals and buffers are listed in the appendix, sections IX.3 and IX.4, respectively.

IV.2 Expression vectors and cloning

IV.2.1 *Agrobacterium tumefaciens* infiltration of plants

A. tumefaciens strain GV3101:pMP90RK was transformed with plasmids described in section IV.2.2 by electroporation [249]. Bacteria were cultivated in yeast extract broth (YEB) containing carbenicillin ($50 \mu\text{g mL}^{-1}$), kanamycin ($25 \mu\text{g mL}^{-1}$) and rifampicin ($25 \mu\text{g mL}^{-1}$) for the selection of recombinant clones at 25°C . For transient expression in *N. benthamiana*, *A. tumefaciens* was cultured to an optical density at 600 nm ($OD_{600\text{nm}}$) of ~ 5.0 and diluted with two-fold infiltration medium and tap water to the desired final $OD_{600\text{nm}}$. An $OD_{600\text{nm}}$ of 1.0 corresponded to $1.43 \pm 0.12 \times 10^9$ colony forming units per mL ($n = 6$). Bacteria were inoculated by vacuum-infiltration of whole plants at ~ 50 Pa absolute pressure for 15 min followed by sudden vacuum release.

IV.2.2 pGFD vector

The pGFD vector, a derivative of pPAM (GenBank AY027531), was kindly supplied by Dr. Thomas Rademacher, Fraunhofer IME [250]. This vector included a backbone β -lactamase gene for selection in *A. tumefaciens* and a T-DNA containing the genes for DsRed (GenBank AF168419; R2G mutant), the 2G12 κ light chain (F62 version) and the 2G12 γ heavy chain, separated by scaffold attachment regions (SARs) from the tobacco *RB7* gene (GenBank U67919). The SARs were included even though they are not expected to function during transient expression so that exactly the same constructs could be compared in transgenic plants. This holds true for all other constructs described below. The *DsRed* gene was fused to the transit peptide sequence from the barley granule-bound starch synthase I gene (*gbsSI*) allowing the recombinant protein to be imported into plastids (GenBank X07932), whereas the light and heavy antibody chain genes included their native (human) signal peptide sequences allowing secretion to the apoplast. Each gene was expressed under the control of the *Cauliflower mosaic virus* (CaMV) 35S promoter. Transgenic tobacco seeds carrying the same integrated T-DNA were also kindly supplied by Dr. Thomas Rademacher.

IV.3 Plant growth

IV.3.1 Plant species

The *N. tabacum* varieties Petit Havana SR1 (for brevity referred to as SR1 hereafter) and K326 as well as their transgenic progeny GFD (based on SR1) and pfs25, pfs25KO and E25T (all based on K326) were used for the studies described in this thesis. SR1 and *N. benthamiana* were also used for transient protein expression (IV.2.1).

IV.3.2 Greenhouse

Tobacco seeds were germinated on rockwool blocks (Cultilène, The Netherlands) and were cultivated in a greenhouse at 25/22°C day/night temperature with a 16-h photoperiod (180 $\mu\text{mol s}^{-1} \text{m}^{-2}$; $\lambda = 400\text{--}700 \text{ nm}$) and at 70% relative humidity. The plants were irrigated with a 0.1% (w/v) solution of Fertyl 2 Mega (Kammlott GmbH, Germany) and were grown for either 35 (no bud) or 42 days (developing bud) prior to infiltration with *A. tumefaciens*, or for 47 days (mature bud) prior to harvest in case of transgenic SR1 plants. Transgenic K326 plants were harvested 51 days post seeding.

IV.3.3 Post-infiltration incubation

Infiltrated plants were transferred into phytotrons and incubated for 5 d at 70% relative humidity with a 16-h photoperiod, using six Osram cool white 36 W fluorescent tubes per 0.7 m^2 (75 $\mu\text{mol s}^{-1} \text{m}^{-2}$; $\lambda = 400\text{--}700 \text{ nm}$). The incubation temperature was set to 25°C. Treated leaves were sampled 5 days post injection (dpi) when plants reached 47 days post seeding (dps).

IV.4 Sample preparation and analysis

IV.4.1 Sampling from transgenic plants and process steps

Samples were taken during extraction and subsequent DSP steps (Figure VI-1) and thus did not require maceration. Samples were centrifuged twice at 16,000 $\times g$, 20 min, 4°C and supernatants were stored at -80°C .

IV.4.2 Sampling from infiltrated plants

Samples were taken from infiltrated leaf parts (IV.2.1) using a cork borer, and 3 μL of extraction buffer (50 mM sodium phosphate buffer, 500 mM NaCl; pH 8.0) was added per 1 mg of fresh biomass. Proteins were extracted by grinding with an electric pestle. Extracts were centrifuged twice at 16,000 $\times g$, 20 min, 4°C, and supernatants were stored at -80°C .

IV.4.3 Protein quantitation

The quantity of total soluble protein (TSP) in the supernatants was determined using the Bradford method [251, 252]. Briefly, 2.5 or 5.0 μL of supernatant was mixed with 200 μL of Bradford reagent (Thermo Fisher Scientific, Rockford, Illinois) in 96-well plates and incubated for 10 min at 22°C before measuring absorbance at 595 nm using a Synergy HT plate reader (BioTek Instruments, Winooski, Vermont). Eight dilutions of bovine serum albumin (0–2000 $\mu\text{g mL}^{-1}$) were prepared in triplicate and used to build a standard curve.

DsRed fluorescence in supernatants was determined using a Synergy HT plate reader fitted with 530/25 nm (excitation) and 590/35 nm (emission) filter sets in 96-well half area plates. Reads were averaged over triplicate samples of 50 μL and a standard curve was generated with dilutions in the range 0–225 mg mL^{-1} .

The quantity of 2G12 antibody was determined by surface plasmon resonance (SPR) spectroscopy using a Biacore T100 (GE Healthcare, Uppsala, Sweden) measuring the amount of antibody binding to Protein A (Sigma-Aldrich, St. Louis, MO) immobilized on the surface of a CM5 chip by EDC/NHS coupling [253, 254]. A 585 ng mL^{-1} reference solution of 2G12 (Polymun Scientific, Klosterneuburg, Austria) was used for one-point calibration with HBS-EP+ as the running buffer.

IV.4.4 SDS-PAGE analysis

Pre-cast 4–12% (w/v) continuous Bis-Tris gels and additional equipment from Life Technologies were used for sample (IV.4.1, IV.4.2) analysis by reducing lithium dodecyl sulfate polyacrylamide gel electrophoresis (LDS-PAGE). Samples were prepared and separated according to the manufacturer's protocol. Briefly, samples were boiled in LDS-running sample buffer and reducing agent, and 10 μL was loaded per well. Samples were separated for 37 min at 200 V in MES running buffer. After washing in water, gels were stained for 1 h in SimplyBlue™ SafeStain and excess staining solution was removed by washing in water. Stained gels were scanned with a Canon 8800 (Canon, Krefeld, Germany) at a resolution of 600 dpi using Adobe Photoshop 6.0 (San Jose, CA, USA). Silver staining was carried out using the SilverQuest™ Kit. The gels were washed in water, fixed using an acidic ethanol:water mixture, sensitized, stained and developed with intervening washing steps.

IV.4.5 Western blot and immunodetection

Samples separated by SDS-PAGE (IV.4.4) were transferred at 30 V for 2 h onto nitrocellulose membranes (GE Healthcare, Waukesha, WI, USA) using a tank blotting device from Life Technologies. After blocking with 5% (w/v) milk powder in phosphate buffered saline (PBS)

containing 0.05% (v/v) Tween-20, proteins were specifically labeled with (pairs of) mAbs (Table IV.1). The last mAb in the incubation series was always conjugated with alkaline phosphatase (AP) allowing quantitative detection using a colorimetric reaction with nitroblue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP). The NBT and BCIP solutions (stock concentrations 0.3 and 0.15 mg mL⁻¹, respectively) were prepared by adding 100 µL stock per 10 mL AP buffer during development.

Table IV.1: Monoclonal antibody (pairs) used for specific protein detection as 1:5000 dilutions.

Target protein/domain	1 st antibody	2 nd antibody
2G12	Goat α -Human-H+L-AP (Dianova 109-055-003)	---
DsRed	Rabbit α -DsRed (MBL PM005)	Goat- α -rabbit-H+L-AP (Jackson 111-045-045)
His ₆	Rabbit α -his (Genescript A00174)	
E25T	Mouse α -MSP1(19) (mAb 5.2)	Goat- α -mouse-Fc-AP (Jackson 115-005-008)
Pfs25	Mouse α -pfs25 (mAb 4B7)	

IV.5 Homogenization of leaf material

Plant extracts were prepared at three scales (Figure VI-1). For bench-top scale preparations, 50–150 g of fresh leaf biomass was processed in a blender (Waring, Torrington, CT, USA) with three volumes of non-cooled extraction buffer (3 mL g⁻¹) and ground three times for 30 s with 30 s intervals.

For laboratory-scale preparations, 0.5–2.0 kg of biomass was homogenized in a Polytron PT6100 (Kinematica, Lucerne, Switzerland) with a custom blade tool scaled down according to a previously-developed large-scale process [255]. Initially, 250 g of biomass was mixed with 1500 mL non-cooled extraction buffer in a stainless steel bucket connected to the PT6100. After 2 min mixing at 8,200 rpm, the remaining 250 g biomass was added and mixing continued for another 6 min at the same speed.

For pilot-scale preparations, 12–20 kg of biomass was homogenized for 8 min at 3600 rpm in a REACTRON® RT50 (Kinematica) containing 36–60 L of non-cooled extraction buffer. All the biomass was processed in one operation.

Different extraction buffer compositions and pH values were tested (section VI.5). The turbidity, conductivity and pH of the homogenate were monitored and the pH was adjusted if required. Total soluble protein (TSP), 2G12 and DsRed concentrations were determined in homogenate samples as described above (section IV.4.3).

IV.6 Bag filtration

Different bag filters (Table VI.2) were tested alone or in combination with additives and/or flocculants to remove bulk particles from the homogenate. The working parameter settings are provided in Table IV.2. The bag filtrate was monitored for turbidity, conductivity, pH, TSP and the concentrations of 2G12 and DsRed (IV.4.3). In the bench-top and laboratory-scale processes, residual liquid was drained from the filters and filter cake by manual pressing to improve extract recovery because the filter size was slightly too large for the amount of processed biomass.

Table IV.2: Bag filter sizes and flow rates at the different process scales.

Filter	1POG04 (Pall)	BP-410 (Fuhr)	BP-420 (Fuhr)	123-D (3M)
Length [m]	0.210	0.218	0.457	0.800
Diameter [m]	0.107	0.102	0.102	0.180
Area [m ²]	0.080	0.078	0.155	0.478
Volume [L]	1.89	1.78	3.73	20.36
Volumetric flow rate [L min ⁻¹]	0.190	0.125	0.250	1.000
Volumetric areal flow rate [L m ⁻² min ⁻¹]	2.375	1.602	1.612	2.128
Linear flow rate [cm h ⁻¹]	14.25	9.61	9.67	12.77
Volume exchange [min ⁻¹]	0.1	0.07	0.07	0.05

IV.7 Depth filtration and screening

Different depth filters (Table IV.3, Table IV.4 and Table VI.3) were tested alone or in combination with additives and/or flocculants to remove fine particles from the homogenate. The volumetric loading flow rate was kept approximately constant over the different scales. Properties of the different filter types and manufacturers are listed in Table IV.3 and Table IV.4. The filtrate was monitored for turbidity, conductivity, pH, TSP and the concentrations of 2G12 and DsRed (IV.4.3).

Table IV.3: Depth filter properties at the bench-top scale.

Filter	PD series ¹ (Pall)	Sartoclear Caps (Sartorius)	BC series (3M)
Diameter [m]	0.0529	0.0564	0.0564
Area [m ²]	0.0022	0.0025	0.0025
Volumetric flow rate [L min ⁻¹]	0.006	0.006	0.006
Volumetric areal flow rate [L m ⁻² min ⁻¹]	2.733	2.400	2.400
Linear flow rate [cm h ⁻¹]	16.4	14.4	14.4

¹ effective diameter after sealing, diameter of the filter element is 0.06 m.

Table IV.4: Depth filter properties at the laboratory and pilot scales.

Filter	Supracap 100 series (Pall)	Sartoclear Maxicap (Sartorius) (20'')	Stax series (Pall)	Sartoclear L-Drum (Sartorius)
Diameter [m]	0.154	0.100	0.442	0.310
Area [m ²]	0.025	0.150	1.000	2.000
Volumetric flow rate [L min ⁻¹]	0.065 – 0.130	0.100	2.500	3.000
Volumetric areal flow rate [L m ⁻² min ⁻¹]	2.600 – 5.200	0.666	2.500	1.500
Linear flow rate [cm h ⁻¹]	15.6 – 31.2	4	15.0	9.00

IV.8 Flocculant and additive screening and evaluation

Different flocculants (Table VI.5 and Table VI.6) and additives (Table VI.9) were tested in a range of concentrations and buffer conditions for their ability to reduce extract turbidity at various points during clarification (Table IV.5 and section IV.12.1). Flocculants were prepared as 2% or 4% (w/v) stock solutions in 50 mM phosphate buffer, based on the formulation delivered by the manufacturer. The pH of the stock solution was adjusted to that of the extract with which it was mixed. In screening experiments, the amount of stock solution needed to achieve the desired final concentration of flocculant was added to 20 mL of crude extract and the volume was adjusted to 22 mL with extraction buffer. The samples were mixed thoroughly for 20 s by manual shaking and allowed to settle for varying times at different temperatures according to the DoE (Table IV.5). The samples were then filtered through two layers of Miracloth (Calbiochem/Merck-Millipore, Darmstadt, Germany) and the turbidity was determined immediately and after 24 h as a 1:10 dilution in extraction buffer using a Hach 2100P (Loveland, CO, USA). In processing experiments, the required amount of flocculant stock was added to the homogenate and mixed using the appropriate homogenization device for 10–15 s.

Table IV.5: Parameter ranges tested during flocculant screening by DoE.

Parameter	Lower boundary	Upper boundary
Extraction buffer pH [-]	4	8
Extraction buffer conductivity [mS cm ⁻¹]	15	55
Extract temperature [°C]	4	30
Flocculant concentration [g L ⁻¹]	0	4
Flocculation time [h]	0.08	1.25

For testing at the laboratory scale, appropriate volumes of flocculant stock solution were added to the extract after homogenization and mixed with the PT6100 for 15 s to 5 min. The

pH was adjusted to 7.5 if necessary and the particles were allowed to settle for 5–30 min before filtration through BP-410.

Different additives (Table VI.9) were mixed directly with the extract prior to bag or depth filtration and stirred until a homogenous dispersion was achieved, or dispersed in buffer and circulated over the filter to build up a filter cake prior to filtration. The circulation was continued until additive dispersion became clear, then the extract was filtered through filter and cake. Additives were tested in concentrations of 0–5 g L⁻¹.

IV.9 Pretreatment of leaves or bulk plant extracts

IV.9.1 Blanching of plant leaves

Blanching was always carried out with fresh leaf material before extraction (IV.5). Therefore, 150 g of plant biomass were packed into a 23 × 23 × 23 cm polypropylene basket with wide meshes and a lid preventing the leaves from floating out. The filled basket was then submerged into a blanching bath filled with ~8L of water or 50 mM phosphate buffer (pH 8.0) pre-heated with a Lauda E300 (Lauda Dr Wobser GmbH, Germany) to the required temperature in the 60–85°C range. The liquid in the blanching bath was agitated with a 4 cm magnetic stir bar with 180 rpm to reduce stagnant layers at solid-liquid interfaces. After incubation for the time required in the respective experimental setup (0.5–5.0 min) the basket was removed from the blanching bath, drained for 30s and the plant material was subjected to extraction (IV.5).

IV.9.2 Heat treatment of bulk extract

Heat treatment was conducted using either a heated vessel or a heat exchanger after protein extraction from leaf biomass (IV.5) but never combined with blanching (IV.9.1). In the vessel setup, a 2-L stainless vessel was thermally equilibrated in a water bath (setup as described under IV.9.1) for 15 min to the temperature required for heat treatment (60–85°C range). Then, 300 mL of extract were poured into the vessel and incubated as required (0.5–5.0 min). A 4-cm magnetic stir bar was also placed in the vessel to agitate the extract during heating with 150 rpm. The heat treated extract was then subjected to filtration (IV.6). As an alternative to the heated vessel, plant extract was processed in a 12-loop custom heat exchanger that was thermally equilibrated as described for the vessel. An 0.5-L bucket was connected to the heat exchanger's inlet and outlet hose using L/S 24 tubing and placed in a 1.0-L vessel containing cotton wool as insulation material. A peristaltic pump Masterflex L/S (Masterflex, Germany) was hooked up to the feed side tubing of the heat exchanger to apply a constant plant extract circulation rate of 300 mL min⁻¹. The plant extract was circled through the assembly for the

required time (0.5-5.0 min). The heat exchanger was then transferred into another water bath containing 8 L of ice-cold water and circulation was continued until the extract temperature fell below 30°C. All liquid was drained into the 0.5-L vessel using the pump. Subsequently, the heat treated extract was subjected to filtration (IV.6).

IV.10 QSAR modeling

IV.10.1 HCP detection by mass spectrometry

Samples from crude tobacco extracts were separated by SDS-PAGE, stained (IV.4.4) and the amount of protein in each band was calculated as a fraction of the total protein per lane averaged over three different dilutions using the Aida Image Analyzer software (raytest GmbH, Germany). Individual bands were isolated, the gel pieces were washed with water and acetonitrile and then dried in a concentrator 5301 (Eppendorf, Germany). The proteins were then reduced with dithiothreitol and re-oxidized under defined conditions with iodoacetamide. After washing the gel pieces in ammonium carbonate and acetonitrile, the proteins were digested in-gel with trypsin for analysis by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (MS) on an amaZon ETD (Bruker, MA, USA). Peptide mass fingerprints (PMFs) were evaluated using Mascot (Matrix Science, Boston, MA, USA). Individual proteins were quantitated by dividing the total protein fraction of a band (see above) by the number of identified proteins. These amounts were summed over all bands containing a specific protein to calculate its relative abundance in plant extract as a fraction of TSP.

IV.10.2 Homology modeling

The sequences of tobacco proteins identified by PMF (or the closest tobacco homologs when the identified proteins originated from other species) were compiled from UniProt [256]. If no tobacco homologue was found, the originally-identified protein was used instead. The 3D structures of these proteins were then obtained from PDBe or RCSB PDB [257, 258]. If no 3D structure was available the final processed amino acid sequences (without leader peptides) were used as input data for SWISS-MODEL homology modeling in automatic modeling mode [198, 200, 259]. The TargetP 1.1 webserver was used to predict leader/signal peptides in case no annotation was available [260, 261]. For each protein, the model with the highest sequence coverage was selected and its quality was ranked by QMEANscore4 [262, 263], Z-Score [264] and sequence identity. Only templates with a sequence identity greater than 30% compared to the target were used for modeling. Missing residues were added in MOE (Chemical Computing Group, Montreal, Canada) using the homology modeling algorithm with the initial homology

model as the template, whereas all water and hetero atoms were removed. Three rounds of energy minimization on increasingly-flexible atoms were carried out in protonation states corresponding to pH 5.0, 6.0, 7.0 and 8.0 of each host cell protein. For model proteins included in the QSAR training data set (IX.6) pH values corresponding to the reported experimental conditions were selected (VI.10.2). The interaction surfaces were calculated with a corresponding algorithm in MOE for all states and proteins. Final models were manually curated to prevent structural clashes.

IV.10.3 Descriptor calculation

MOE was used to calculate 2D, TAE (transferable atom equivalent), i3D and x3D descriptors for all proteins at the different pH values (IV.10.2). The PEST (Property-Encoded Surface Translator) algorithm was used to calculate EP (electrostatic potential) and MLP (molecular lipophilicity potential) for proteins at pH 7.0 [265].

IV.10.4 QSAR model building and prediction

Retention data for model proteins at different pH values on different resin types were obtained from the literature [155, 266-274]. Descriptors for these model proteins and HCPs were calculated as described above (IV.10.3). Descriptors with a significant correlation to protein retention were selected from this dataset using the YAMS (yet another modeling software) webserver by Dr. Michael Krein [219], hosted at the Rensselaer Exploratory Center for Cheminformatics Research (RECCR; <http://reccr.chem.rpi.edu/ONR-QSPR/>). Eight iterations of descriptor selection were carried out, each time removing from the dataset 30% of the descriptors with the least correlation. Descriptors were selected using a SVM algorithm. The built-in quality control tools of the web-server, such as k -fold cross-validation (here $k = 10$) [275], Y -scrambling (y -randomization) and r^2 , were used to evaluate the QSAR model quality. The method used to assemble the predicted elution salt concentrations and individual protein amounts into a “synthetic” chromatogram is described in the results (section VI.10.3).

IV.11 Chromatography

All chromatographic separations were carried out using an ÄKTA purifier 10 or 100 and ÄKTA explorer fast protein liquid chromatography (FPLC) systems from GE Healthcare. A variety of columns from different manufacturers was tested at different flow rates (Table IV.6) and details are provided in the corresponding results section (VI.9). Protein, DNA and DsRed elution from columns was monitored at 280, 260 and 495 nm, respectively.

Table IV.6: Column sizes, volumes and flow rates during chromatographic protein separation on ÄKTA FPLC systems

Column	HiTrap Series (GE)	PRC series (Pall)	XK26/10 (GE)	MiniChrom (ATOLL)
Bed height [cm]	2.5	5.0	10.0	10.0
Bed diameter [cm]	0.7	0.5	2.6	0.8
Column (bed) volume [mL]	1.0	1.0	53.1	5.0
Wash and elution flow rate [mL min ⁻¹]	1.00	1.00	8.00	5.00
Linear flow rate [cm h ⁻¹]	155.9	305.6	90.4	600.0
Loading flow rate [mL min ⁻¹]	0.125 – 1.00	0.125 – 1.00	n.a.	5.00
Linear flow rate [cm h ⁻¹]	19.4 – 155.9	38.1 – 305.6	n.a.	600.0

IV.12 Design of experiments

Design Expert 8.0 (Stat-Ease, MN, USA) was used to build and evaluate all experimental designs. Unless otherwise noted, error bars on the predicted values indicate the 95% confidence interval of the model which equals $2 \times \sigma$ and is thus twice the size of the standard deviation normally shown using error bars.

IV.12.1 Flocculant screening and model

Eighteen different flocculants (Table VI.5) and controls without flocculants were tested in an IV-optimal RSM DoE consisting of 88 conditions covering a range of pH values and flocculant concentrations (Table IV.7) but constant conductivity (30 mS cm⁻¹). A refined set of six flocculants was used in an IV-optimal second DoE with 60 different conditions based on the same parameters. For optimization of the three best flocculants, an IV-optimal RSM with 30 conditions at additional pH values of 5 and 7 was performed. Based on the results, a second set of flocculants was screened in an IV-optimal RSM of 70 conditions at different concentrations, conductivities, pH values and incubation times (Table VI.5). The efficacy of the best-performing flocculant was evaluated in a 57 run IV-optimal DoE with the same parameters. The preferred concentration range of the final flocculant was then evaluated using an IV-optimal RSM design with 70 conditions at different concentrations, conductivities, pH values, incubation times and temperatures (Table IV.5). This DoE was applied to transgenic plants harvested 44 and 49 dps, corresponding to the late budding and mature flower stages respectively. Factors showing a significant influence on turbidity were pre-selected from a quadratic or cubic model by automatic backwards selection using a p-value threshold of 0.100, and factors with p-values greater than 0.050 were removed manually. Exceptions were made to

maintain model hierarchy, a requirement for high-quality polynomial regression models [276, 277].

Additional designs were set up for combinations of flocculants (VI.5.2).

Table IV.7: Parameter ranges during flocculant screening experiments by IV-optimal DoE.

Parameter	Lower boundary	Upper boundary
Extract pH [-]	4.00	8.00
Flocculant concentration [g L ⁻¹]	0.01	3.00
Flocculant type [-]	none	18 th flocculant

V. Research proposal

As described in the introduction, plants can offer a vital alternative to cell culture-based approaches for the production of biopharmaceutical proteins. Even though a plethora of plant species has been tested for their capabilities to produce such proteins in the last decades, little attention has been paid to the challenges encountered during product extraction and purification which has been identified as the most cost-intensive part in plant-based biopharmaceutical production. As a consequence, concerns have come up as to whether an economical production of biopharmaceutical proteins can be achieved in plants at all. Therefore, there is a distinct need to improve the process performance of plant-based production approaches in terms of DSP including not only technical challenges such as clarification and product purification but also in terms of economic aspects, i.e. consumables costs.

Hence, this thesis considers different aspects of DSP for a plant-derived biopharmaceutical protein (Figure V-1). First, an existing cGMP process for the production of monoclonal antibodies was scaled down and optimized in terms of duration, labor, and the cost and performance of the homogenization and clarification operations. This involved the introduction of several new methods such as the use of flocculants (section III.6.1.2). Different model-building strategies were used to favor a QbD approach for subsequent process development in concert with a novel generic extraction and clarification process (Part I). Second, experimental and modeling methods were also evaluated for their potential to facilitate the rational development of chromatographic purification processes for clarified plant extracts (section III.6.4 and III.6.5) (Part II). The monoclonal antibody 2G12 (which binds HIV glycoprotein gp120) and the fluorescent protein DsRed were used as model proteins throughout the work described in this thesis. An antibody was chosen because it represents the dominant class of biopharmaceutical products (section III.3), whereas DsRed was selected because it is easy to detect and quantify by fluorescence analysis. Additionally, a GMP-production process for these two proteins has already been established recently providing background data for comparison.

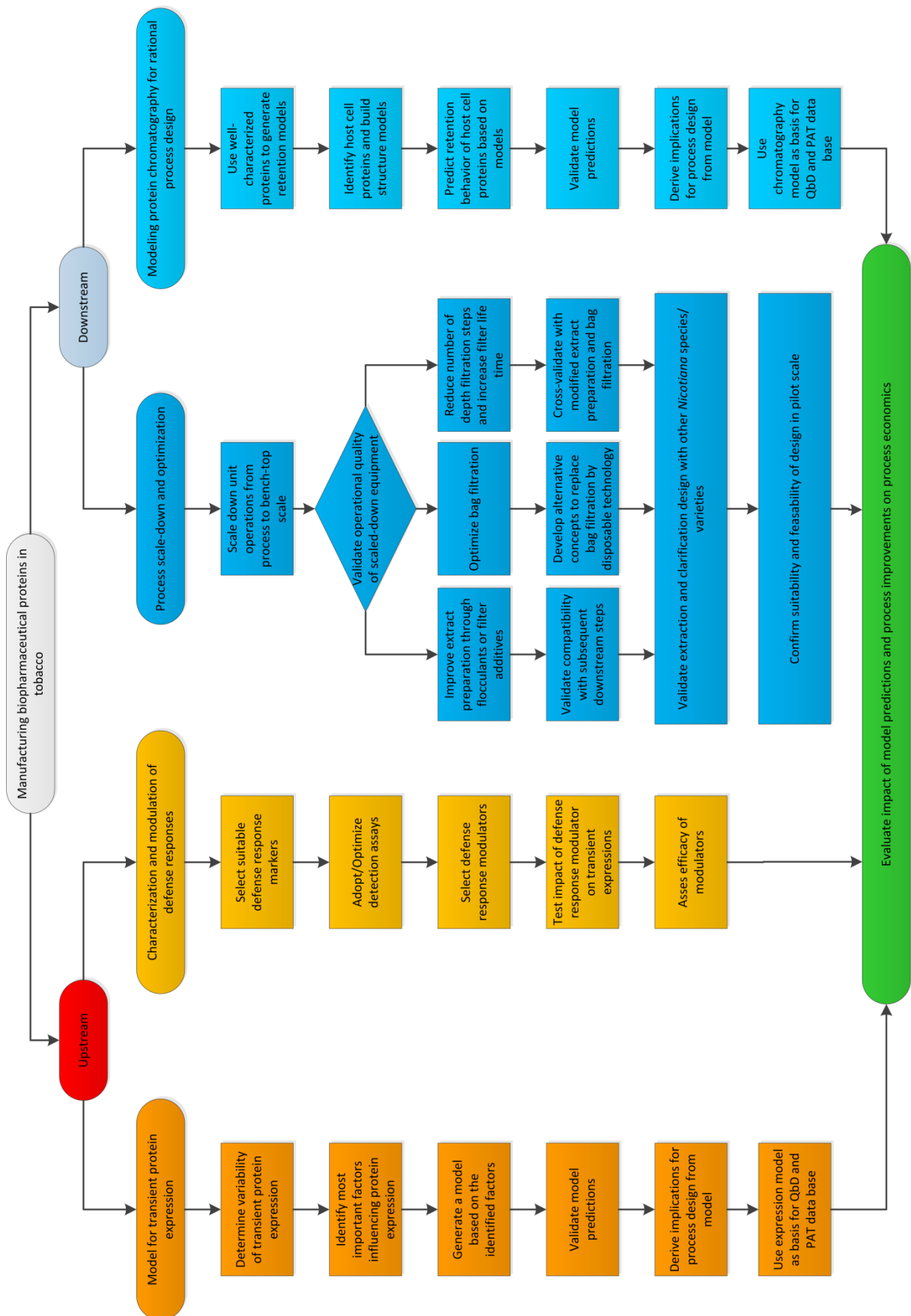


Figure V-1: Workflow during this PhD thesis, as described in section Error! Reference source not found.. The work packages in orange are part of an accompanying PhD thesis entitled “Manufacturing biopharmaceutical proteins by transient expression in *Nicotiana tabacum* (L.)” [278].

VI. Results and discussion

Part I: Extraction and clarification

VI.1 Scale-down of process equipment

The results presented in this paragraph as well as VI.3 and VI.4 have been published as the following manuscript:

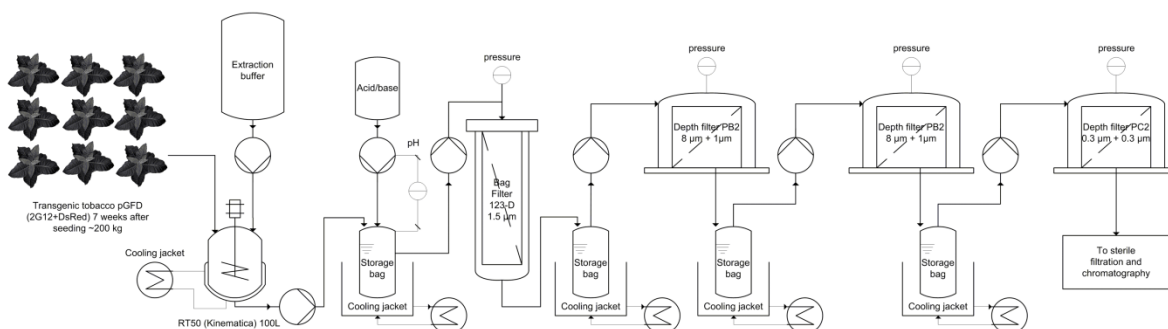
1. Buyel JF, Fischer R. 2014. Scale-down models to optimize a filter train for the downstream purification of recombinant pharmaceutical proteins produced in tobacco leaves. *Biotechnology Journal* 9(3):415-25.

VI.1.1 Scale-down of unit operations

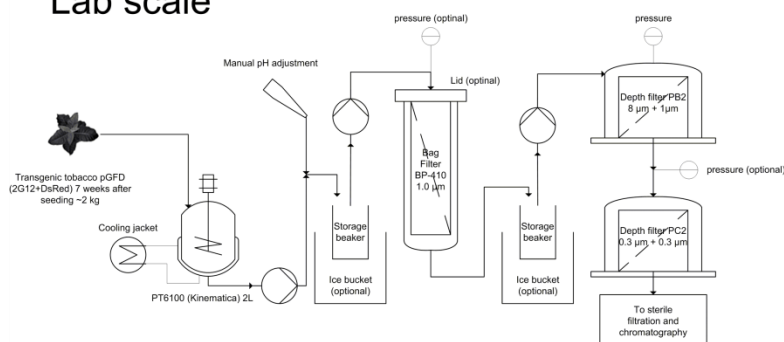
A homogenization device and filter train for processing 200 kg of transgenic tobacco leaves was developed by Dr. Martin Lobedann during his PhD project (Figure VI-1 A top) [255].

A

Process scale



Lab scale



B

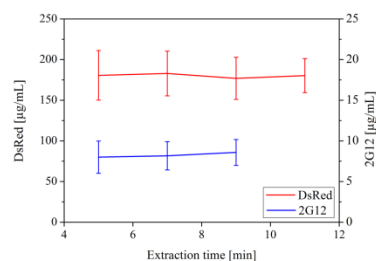


Figure VI-1: Scale-down of equipment for the production of two biopharmaceutical proteins in tobacco.

A. Homogenization and filter train in the 20 and 200 kg scale processes (top) and the 2 kg scale process (bottom). In the 2 kg scale process, some storage devices could be omitted because smaller volumes were processed. B. Verification of the operational quality of the 2 kg scaled-down process by extracting antibody 2G12 and fluorescent protein DsRed. Extraction did not vary significantly over a range of homogenization times. Error bars indicate standard deviation ($n = 16$).

Table VI.1: Scaling parameters and components in 200 kg and 2 kg extraction processes.

Parameter	Abbreviation	Scale	
		Process	Lab
Homogenizer	---	RT50	PT6100
Total vessel volume [m ³]	V_T	0.1200	0.0033
Vessel diameter [m]	d_V	0.508	0.150
Vessel radius [m]	r_V	0.254	0.075
Vessel height [m]	h_V	0.643	0.186
Blade diameter [m]	d_b	0.200	0.089
Blade radius [m]	r_b	0.100	0.045
Blade revolutions [s ⁻¹]	n	60	137
Blade top velocity [m s ⁻¹]	v_{max}	37.7	38.3
Biomass per run [kg]	m	24.0	0.5
Extraction buffer per run [m ³]	V_B	0.0720	0.0015
Working volume [m ³]	V_W	0.096	0.002
Vessel diameter/height ratio [-]	d/h	0.790	0.806
Working/Total volume ratio [-]	W/T	0.800	0.615
Blade/Vessel diameter ratio [-]	b/V	0.394	0.595
Homogenization time [min]	t	8	8
Bag filter	---	8x 123-D	1x BP-410
Nominal retention rating [μm]	p_{bag}	1.5	1
Length [m]	h_{Bag}	0.800	0.218
Diameter [m]	d_{Bag}	0.180	0.102
Area [m ²]	A_{Bag}	0.478	0.078
Volume [L]	V_{Bag}	20.36	1.78
Volumetric flow rate [L min ⁻¹]	v_{Bag}	1.000	0.125
Volumetric areal flow rate [L m ⁻² min ⁻¹]	$v_{dot{Bag}}$	2.128	1.602
Linear flow rate [cm h ⁻¹]	u_{Bag}	12.77	9.61
Volume exchange [min ⁻¹]	V_E	0.05	0.07
Depth filter 1/2	---	PB2/PC2	PB2/PC2
Nominal retention rating [μm]	p_f	8+1/0.3+0.3	8+1/0.3+0.3
Diameter [m]	d_f	0.310	0.056
Area [m ²]	a_f	2.0000	0.0025
Volumetric flow rate [L min ⁻¹]	v_f	3.000	0.006
Volumetric areal flow rate [L m ⁻² min ⁻¹]	$v_{dot{f}}$	1.5	2.4
Linear flow rate [cm h ⁻¹]	u_f	9.0	14.4

Briefly, the process consisted of a blade-based homogenization device for leaf material followed by a bag filtration step (123-D) to remove bulk particles. Subsequently, finer particles were removed in three depth filtration steps, the first two consisting of the same depth filter material (PB2) but using different operating modes: differential pressure and filtrate turbidity,

respectively. The third depth filter (PC2) was also operated in turbidity mode. This setup was scaled down to a process volume of 2 kg for the further optimization of extraction and clarification (Figure VI-1 A bottom, VI.2 to VI.6) with the parameters shown in Table VI.1.

VI.1.2 Verification of operational quality

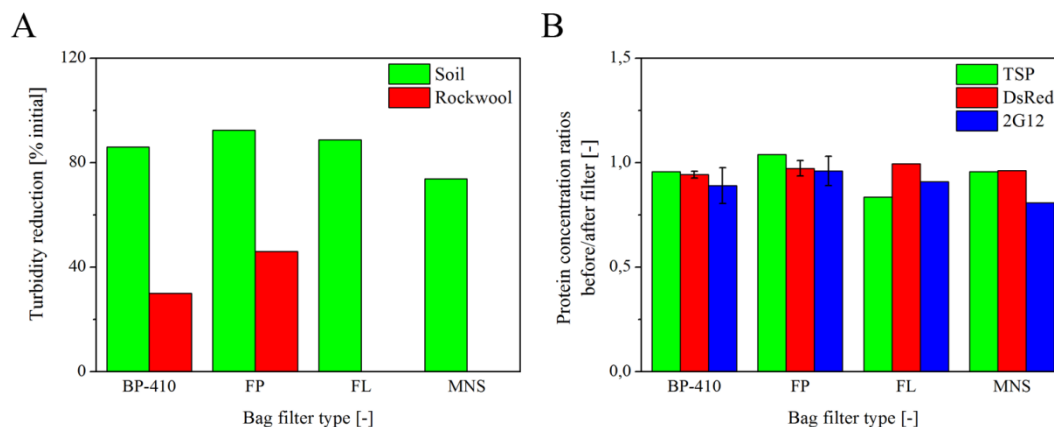
The functionality of the scaled-down homogenization device was confirmed by the efficient extraction of antibody 2G12 and the fluorescent protein DsRed from transgenic tobacco leaves, resulting in the same target protein concentrations reported for the 200 kg scale process (Figure VI-1 B) [255]. Homogenization for 8 min was sufficient for the complete extraction of both target proteins, and even shorter times did not cause a significant loss of product. The depth filtration step also scaled well, taking the different flow characteristics and reduction to a two-step filtration into account (1x PB2 + 1x PC2). The filter lifetime in the 200 kg scale process (tangential flow) was 61 L m^{-2} [255] and in the 2 kg scale process (frontal flow) it was $44.2 \pm 3.7 \text{ L m}^{-2}$ ($n = 3$).

VI.2 Screening bag filters for initial clarification

A variety of bag filter materials (Table VI.2) was tested to reduce extract turbidity more efficiently before depth filtration without affecting the yield of the target proteins compared to the initial bag filter BP-410 (Table IV.2). The FP filter performed best in terms of turbidity reduction and protein yield (Figure VI-2), the MNS filter produced the most turbid filtrate and resulted in the greatest loss of target protein and the performance of the BP-410 and FL filters was intermediate. The BP-410 was selected for the 2 kg scale process because the per piece price (ppp) of €5 is much less than the others (€20). However, the FP filter may be more beneficial in the larger-scale process because filter lifetime and product yield become more relevant cost factors compared to small-scale screening experiments. The FP filter also reduced turbidity more efficiently than the BP-410 filter when processing leaves from tobacco plants grown on rockwool rather than soil. This is relevant for the production of plant-derived biopharmaceuticals under cGMP guidelines because rockwool has a lower bioburden than soil and is a consistent raw material, helping to reduce batch-to-batch variability. Neither BP-410 nor FP reduced the total soluble protein content (TSP) or the levels of the target proteins after filtration (Figure VI-2 B).

Table VI.2: Properties of bag filters tested for initial particle removal during the processing of tobacco extracts.

Filter	BP-410	FP	FL	MNS
Filter material [-]	Polypropylene needle felt	Polypropylene needle felt	Polyester needle felt	Nylon monofilament
Nominal retention rating [μm]	1.0	1.0	1.0	1.0

**Figure VI-2: Bag filter performance in the 2 kg scale process (IV.4.1, IV.4.3 and IV.6).**

A. Turbidity reduction achieved with different filter materials and suppliers, with values closer to 100% representing superior filter performance. B. The impact of different filter materials on protein yield, with values closer to unity representing insignificant effects and thus superior filter performance. Error bars indicate standard deviation ($n = 3$). Some error bars are missing because bag filters were unavailable.

VI.3 Development of a disposable bag filtration step

The results presented in this paragraph have been filed as the following patent:

1. Housing for a single-use bag filter; European patent application 12 176 847.7; US patent application 13/943,288

Bag filtration efficiently removed coarse particles ($>10 \mu\text{m}$) from the plant homogenate in a cost effective manner (bag filter costs were $\sim 5\%$ of depth filter costs, as shown by the comparison in Table VI.4). However, the bag filters were mounted in a reusable stainless-steel housing (Figure VI-1 and Figure VI-20), introducing a potential source of contamination and requiring laborious and time-consuming cleaning in place, assembly and disassembly, which took 1-2 person hours each.

These drawbacks could be addressed using disposable filter modules [96, 97] but there were no reports of disposable bag filter modules in the literature published up to February 2012, and there were no patents covering such modules in the prior art according to a search of the Fraunhofer division of patents and licenses (FPL) performed in March 2012 (personal communication). This revealed a gap in disposable technologies within the production train for

plant-derived biopharmaceuticals. Single-use solutions are available for plant growth (e.g. by Cultilène, or Grodan, Netherlands), storage bags and connectors (e.g. GE Healthcare or Sartorius), depth and membrane filters (e.g. 3M, Pall, Sartorius) as well as chromatography (e.g. GE Healthcare or ATOLL) but not for homogenization and bag filtration. Two types of disposable bag filter housings were therefore developed to eliminate this gap and improve the liquid/product recovery at the same time (European patent application number 12 176 847.7).

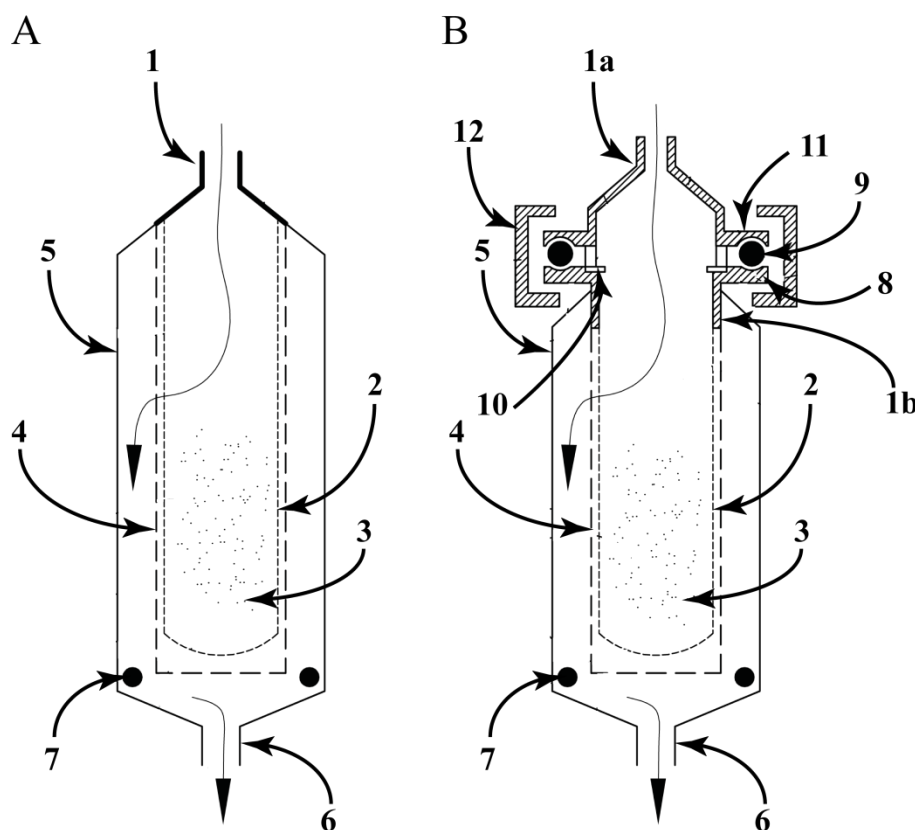


Figure VI-3: Layout for disposable bag filter housings.

A. Layout for an integrated disposable bag filter housing used to retain particles (3) dispersed in plant extract entering through the inlet (1). The filter (2) may be supported by an optional outer basket (4) to withstand elevated pressures and it is kept away from the drain (6) and the flexible envelope (5) by a spreading ring (7). B. Layout for a multi-purpose disposable bag filter housing that can accommodate different bag types by sequentially assembling the filter (2) at a specific radial groove (10) of lower part of the inlet cap (1b) which is connected to a supporting basket (4) and a flexible envelope (5). Afterwards, a gasket (9) is placed in a groove (8+11) of the upper (1a) and lower (1b) inlet cap. The whole assembly is then sealed with a clamp (12) and dispersed particles (3) can be retained while cleared filtrate leaves through the bottom drain (6).

The first solution was an integrated bag and housing resulting in a one-piece ready-to-use filter (Figure VI-3 A) comprising a rigid inlet cap (1, which may contain a standard connector such as a tri-clamp for connection to upstream equipment) welded to the bag filter material (2) retaining the solids (3). An optional stiff basket (4) was included, which may be required to support the filter material during high-pressure operation, as well as a flexible outer envelope (e.g. a foil, 5). The envelope was fitted with a lower drain (6, e.g. realized as a tri-clamp for connection to downstream equipment) that was kept away from the filter/basket by a

spreading ring (7) either contained in the envelope or mounted on the outside of the envelope by external lugs (not shown). The benefits of this solution include the ease of assembly into the process train and the minimal contamination risk due to the self-contained manufacturing and design. However, the design may be hard to integrate into an existing GMP process if the manufacturer of the disposable bag filter housing does not support the incorporation of a specific filter material which has been used in that GMP process, because a change in the filter material can attract regulatory concerns [279].

Therefore, a second design was developed that allows the disposable filter housing to be fitted with commercially-available bag filters from different manufacturers (Figure VI-3 B). For this purpose, the inlet of the bag filter housing was split into two parts (1a and 1b). The lower part of the inlet assembly (1b) was welded to an optional filter basket (4) and an envelope (5) with the same features described above (6 and 7). However, the lower part was not welded to a bag filter. Instead, the lower part was disk-shaped and featured a groove (8) for a gasket (9, e.g. a rubber O-ring) and also a radial groove (10) that attaches to the sealing ring of a commercially-available bag filter (2), which retains the solids in the feed stream (3). The base of the upper part (1a) of the inlet assembly was also disk shaped featuring a groove (11) for a gasket at a similar position to that on the lower part of the assembly. The upper part of the assembly (1a) tapered towards the inlet which was realized as a common connector type (e.g. tri-clamp). The upper (1a) and lower parts (1b) of the inlet assembly were held together by a clamp (12) and the connection was made water-tight by the gasket (9). The diameter and length of the housing could thus be adjusted to fit most common bag filters, e.g. size 2 (\varnothing 177.8 mm (7"), length 812.8 mm (32")) [280].

Both solutions featured a flexible foil as an outer envelope allowing the compression of the solids collected in the filter at the end of the process, facilitating the release of residual liquid and thus better recovery of the product dissolved in the liquid (Figure VI-4). This compression was achieved with external rolls that approach the filter housing at the top (Figure VI-4, large black circles) and squeeze the flexible envelope with a pre-defined pressure that does not cause damage to the material and thus avoids filter breakage or leakage. The rolls then move downward to drain the residual liquid (b) from the solids (a) deposited inside the bag filter, thus increasing the volume of recovered extract and reducing the mass of waste that needs to be inactivated.

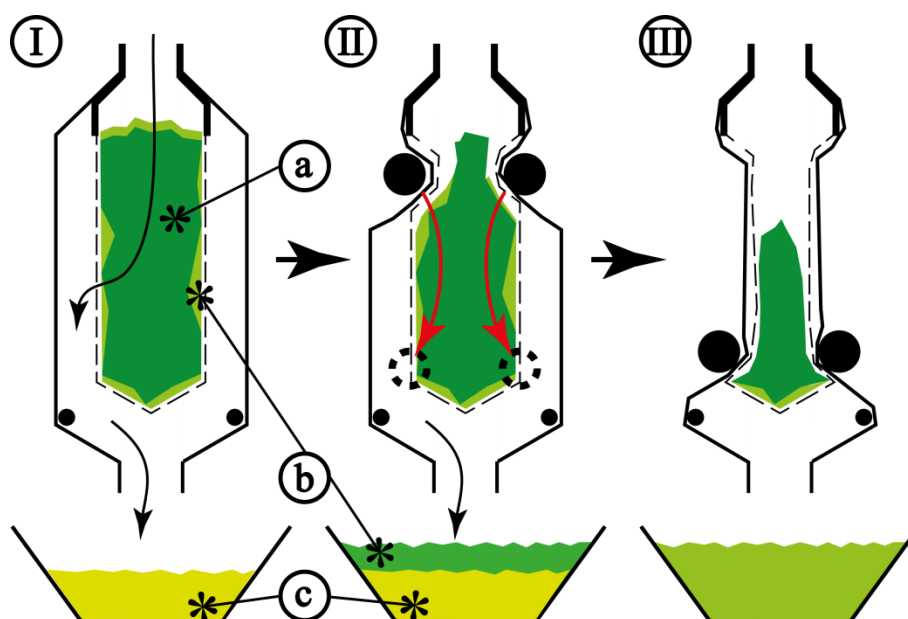


Figure VI-4: Removal of residual liquid from a bag filter after capacity has been reached.

I. Filter capacity has been reached after a certain filtrate volume (c) while not only solids (a) but also liquid (b) remains in the filter cake. II. Rolls (large black circles) pressing against the outside of the flexible bag filter envelope can remove residual liquid by moving downwards (red arrows). III. This procedure increases the recovered filtrate volume and reduces the mass of biological waste that has to be discarded together with the bag filter.

Another approach involves the use of an external rigid container (e.g. stainless-steel) that supports the filter envelope during high-pressure operation without making direct contact with the plant extract (Figure VI-5). The lid of such a container can have a circular or slit-like opening surrounded by spring-mounted rolls (Figure VI-5 c). At the end of the filtration process, the disposable housing including the bag can be removed from the external housing by lifting the filter assembly through the opening in the lid. Simultaneously, the rolls will squeeze residual liquid from the retained solids with a pre-defined contact pressure that still preserves the integrity of the filter and its envelope. The drained liquid can exit the filter through the lower drain which is also used during normal operation. The inlet and lower drain of the disposable filter housing can be mounted to the external container (e.g. by screw connection) to improve pressure resistance.

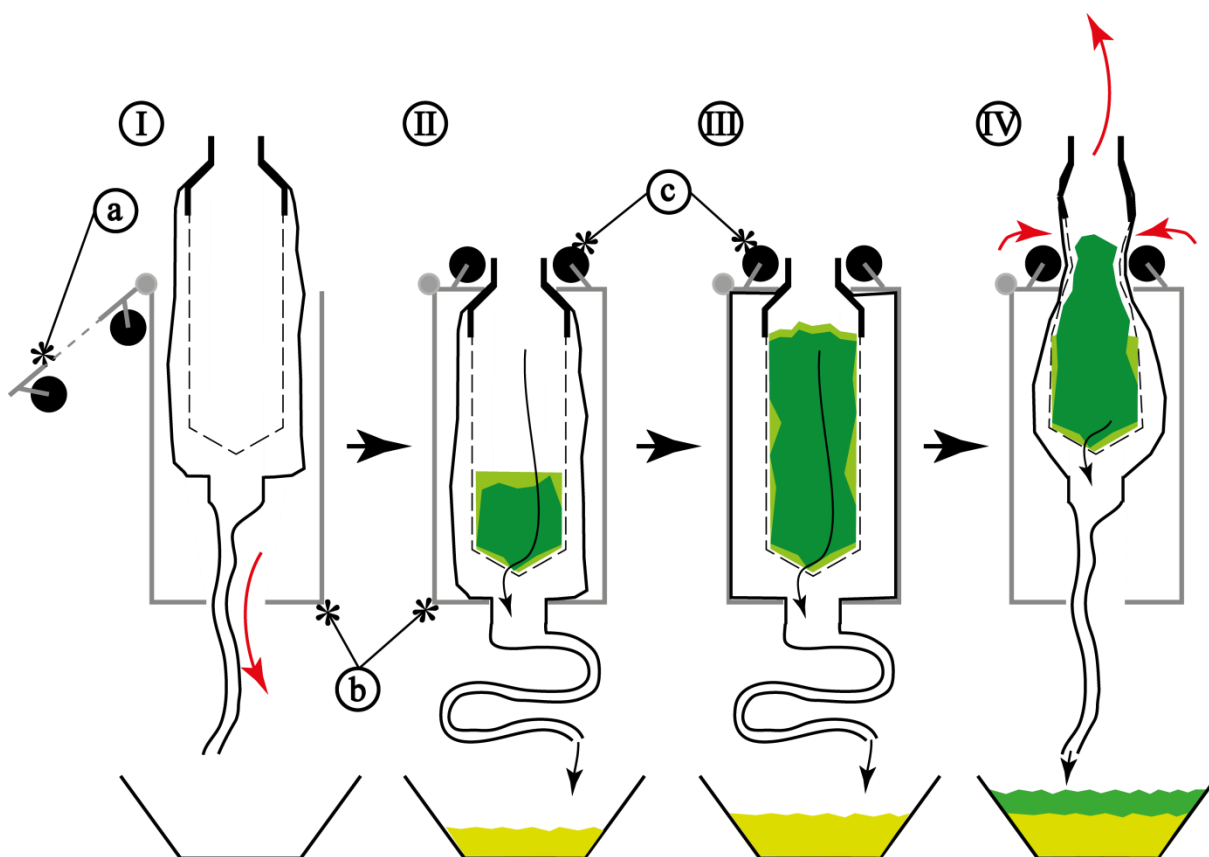


Figure VI-5: Removal of residual liquid from a bag filter contained in a rigid container after capacity has been reached.

I. A disposable bag filter with flexible housing is mounted into a rigid container (b) with a top lid (a). II and III. During operation, the filter will fill-up with retained particles (dark green) and residual liquid (light green) while the increasing pressure pushes the filter housing (e.g. a foil) against the inner wall of the container. IV. The filter can be removed from the container through a hole in the lid (a, upward red arrow) once its capacity is reached. Rolls (c) mounted on top of the lid can drain residual liquid from the filter cake by squeezing (short red arrows) the flexible bag filter housing.

VI.4 Screening depth filters

Two drawbacks were associated with the initial PB2+PC2 filter combination introduced by Dr. Martin Lobedann (Table VI.3). First, the filter lifetime was highly variable, as shown by the large standard deviation which can reduce the robustness of the final process (Figure VI-6). Second, the filter was a combination of two individual filters (three in the large-scale process) which increased the consumables costs and the labor and time required for process preparation. Therefore, several depth filters from different manufacturers were tested alone and in combination (IV.7 and Table VI.3), using extracts from plants grown in soil or rockwool so that the data could be (i) compared to previous filtration tests using extracts from plants grown in soil, and (ii) used to develop an automated process with plants grown in rockwool to meet cGMP recommendations. A final filtrate turbidity of less than 10 NTU was considered the threshold for further evaluation.

Table VI.3: Depth filters (and combinations) tested for the clarification of leaf extracts from tobacco plants grown in soil or rockwool.

Filter (combination)	Manufacturer	Nominal retention rating [μm]	
		1 st filter	2 nd filter
PB2 + PC2 ^{1,2,3}	Sartorius	8 + 1	0.3 + 0.3
PB2	Sartorius	8 + 1	---
PB1+PC1 ^{2,3}	Sartorius	11 + 4	1 + 0.3
PB1+PC2 ^{2,3}	Sartorius	11 + 4	0.3 + 0.3
PB1	Sartorius	11 + 4	---
30SP03+60ZA05 ²	3M	1-2 + 0.8	0.8 + 0.45
30SP03+90ZA05 ²	3M	1-2 + 0.8	0.8 + 0.2
30SP03	3M	1-2 + 0.8	---
60SP05	3M	0.8 + 0.45	---
60SP02	3M	3-5 + 0.45	---
90ZA05 ³	3M	0.8 + 0.2	---
PDE2	Pall	1-3 + 0.2-0.5	---
PDF4 ³	Pall	3-6 + 0.4-0.8	---
PDH4 ³	Pall	6-15 + 0.4-0.8	---
PDE4 ⁴	Pall	1-3 + 0.4-0.8	---
PDF2 ⁴	Pall	3-6 + 0.2-0.5	---
PDG2 ⁴	Pall	4-9 + 0.2-0.5	---
PDG4 ⁴	Pall	4-9 + 0.4-0.8	---

¹ Filter used in the initial process designed by Dr. Martin Lobedann, ² Combinations of two depth filters, ³ Filters used to establish a correlation between depth filter retention gradient and filter lifetime, ⁴ Filters selected for second testing according to depth filter retention gradient (Figure VI-8).

Regardless of the filter, the clarification of extracts from plants grown in rockwool reduced the filter lifetime by 50-70% compared to the clarification of extracts from plants grown in soil (Figure VI-6 A and B). This correlated with the increased turbidity of extracts from rockwool plants compared to those from plants grown in soil (Figure VI-6 C). Furthermore, the plants grown in rockwool showed significant signs of stress, including small leaves (10.1 ± 1.8 g compared to 18.5 ± 4.2 g for plants in soil) and early flowering (6–8 weeks, compared to 8–9 weeks for the plants grown in soil). The filter lifetime was therefore reduced by the processing of extracts from plants exposed to stress, perhaps reflecting the increased synthesis of polyphenols and lignins [281, 282] which have adhesive properties that can rapidly block filters [283-287]. Additionally, the stress-induced expression of proteins cross-linked to cell wall components, can increase the likelihood of biofilm formation [288] thus blocking the filter media. This was supported by the observation of a dark green varnish-like coat on the first layer of the depth filters used to clarify rockwool plant extracts (Figure VI-6 D 5).

Another stress response that could affect filter lifetime is the loss of cell wall stability, which contributes to the smaller size of plants growing in rockwool [289] and reduces the amount of cellulose in the extracts while it increases the levels of soluble pectins [290-292]. Cellulose is used as a filtration additive to increase filter lifetimes, so the presence of cellulose in the extract is beneficial, thus reducing the amount would have a negative impact on the lifetime of the filter (VI.6). Pectins help to form large aggregates that are removed in the pre-filtration steps, so the stress-induced solubilization of pectins [293] would increase the particle burden during depth filtration (due to reduced aggregation) also reducing the effective lifetime of the depth filters. A lack of cellulose and thus a reduced cell wall stability was also consistent with the observed smaller plant size on rockwool [289].

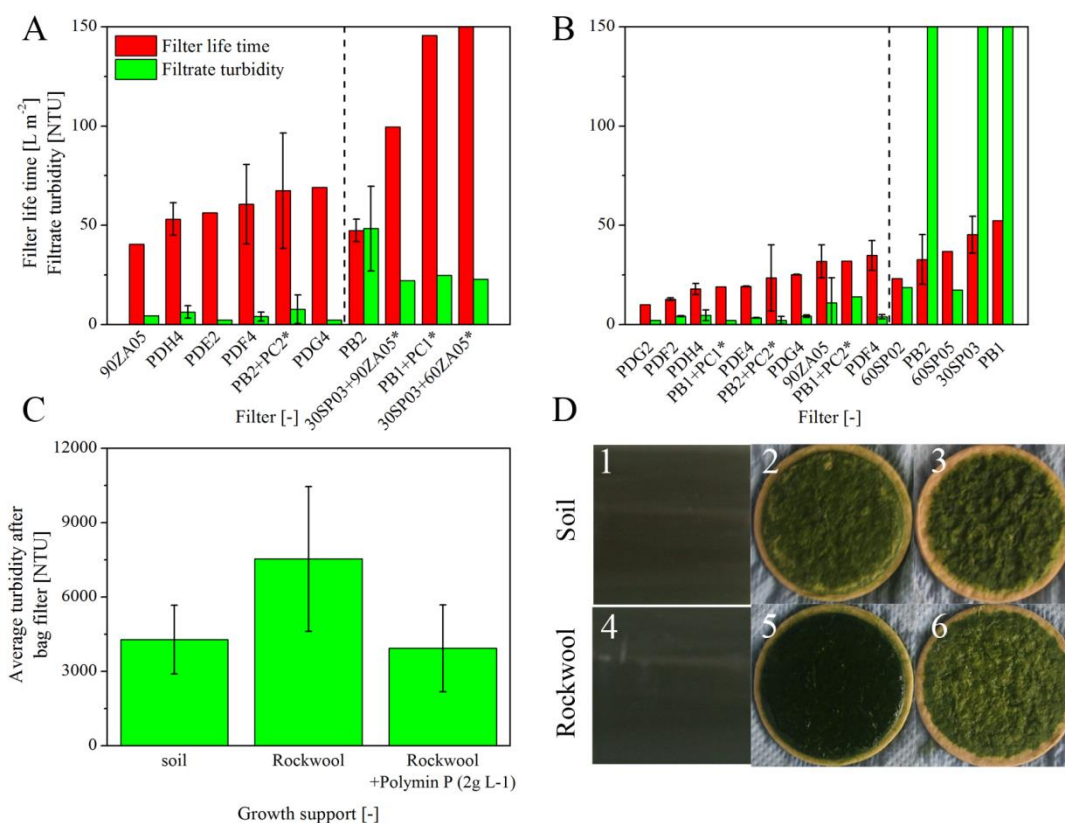


Figure VI-6: Lifetime analysis of different depth filter types used to process extracts derived from SR1 tobacco plants grown in soil or rockwool (IV.4.1 and IV.7).

The filter lifetime was higher when plants were grown in soil (A) compared to rockwool (B) correlating with the turbidity of the crude extract (C). The dashed line is the performance cutoff (10 NTU final turbidity), with only filters to the left of the line selected for further characterization. Asterisks denote tandem filter pairs, indicating that the effective filter capacity is half the value shown. Error bars indicate standard deviations ($n \geq 3$). D. Particle burden before depth filtration (1+4) as well as on first (2+5) and second (3+6) PDF4 filter layers, for extracts from plant grown in soil (top row) or rockwool (bottom row). The particle burden was higher for the rockwool plant extracts, resulting in a darker green cake with particle removal biased towards the first filter layer (compare 2 vs. 5).

Interestingly, it was found that only tobacco variety Petit Havana SR1 (and its transgenic derivative pGFD) was affected by stress, whereas other varieties such as K326 and other

Nicotiana species such as *N. benthamiana* grew normally. Accordingly, the filter lifetime was longer when extracts from these other plants grown in rockwool were processed (Figure VI-7).

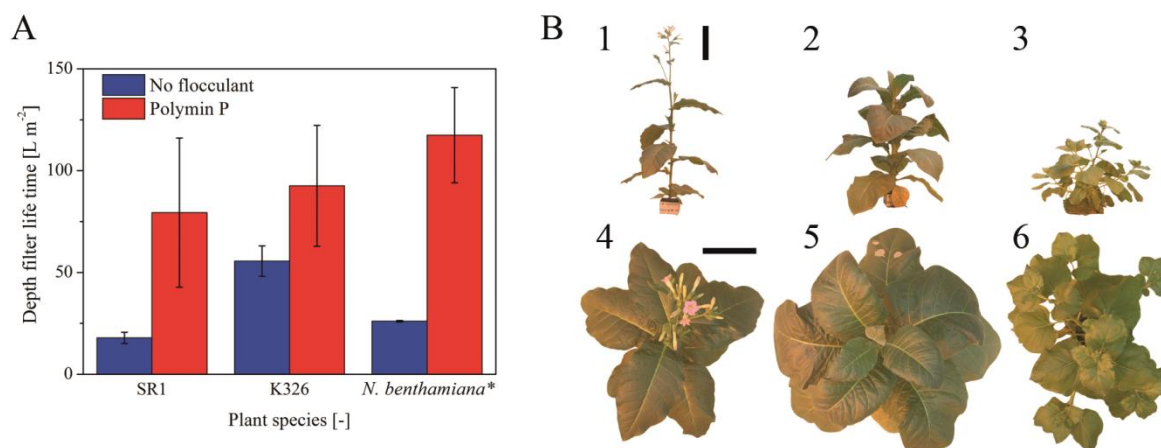


Figure VI-7: Effect of plant species and variety on filter lifetime (IV.3, IV.4.1, IV.7 and IV.8).

A. The processing of extracts from transgenic tobacco variety SR1 growing in rockwool caused a 30% reduction in filter lifetime compared to infiltrated *N. benthamiana* extracts, and a 60% reduction in lifetime compared to transgenic tobacco variety K326 (blue columns). The flocculant Polymin P increased the filter lifetime significantly for all extracts, but the difference between SR1 and *N. benthamiana* extracts remained constant at 30%, whereas the difference between SR1 and K326 shrank to 15% (red columns, VI.5). * Indicates plants infiltrated with *A. tumefaciens*. Error bars indicate standard deviation ($n \geq 3$). B. Characteristic height (top row) and diameter (bottom row) of SR1 (1+4), K326 (2+5) and *N. benthamiana* (3+6) at the time of harvest. Vertical and horizontal scale bars indicate a length 200 mm.

When the rockwool plant extract was processed, the single depth filters 90ZA05 and PDF4 as well as the initial filter combination PB2+PC2 (Table VI.3) had the longest lifetime among the filters meeting the turbidity cutoff value of 10 NTU (Figure VI-6). These three filters were investigated in more detail, to check for undesirable interactions with the target proteins. The PDF4 filter was the most suitable, reducing the levels of TSP, DsRed and 2G12 by 30%, 5% and 23%, respectively (Figure VI-8 A). Regardless of the filter type or combination, DsRed was always retained to a lesser degree than 2G12. This is unlikely to reflect size selection because under native conditions both proteins are tetramers with similar molecular masses of ~150 kDa and ~140 kDa, respectively [294, 295]. However, antibodies are prone to form aggregates, and the presence of glycans on the recombinant antibody increases its size and charge, providing a possible explanation for its retention in preference to DsRed. The charge characteristics may be important because antibodies have a basic pI (the pI of 2G12 is 8.10-8.25 [296]) which means they become positively charged at pH 7.5–8.0 and can thus interact with the negative charges of the diatomaceous earth compounds in the depth filters [297]. Increasing the pH of the extract to ~8.5 before clarification could help to reduce the loss of 2G12 during filtration, but further pH adjustment in the filtrate might be necessary before subsequent processing steps which could result in protein precipitation and the need for additional filtration steps. Omitting diatomaceous earth from the filter can be an alternative to increase product

recovery but is likely to reduce the clarification properties of the according filter [298]. This in turn would increase the number of filtration steps and thus overall costs as well as the likelihood to loose product.

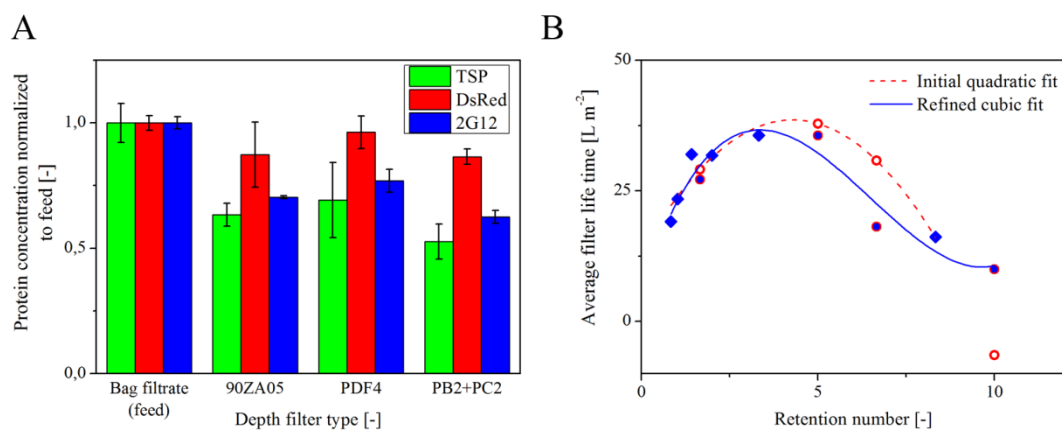


Figure VI-8: Lifetime analysis of three high-performance depth filters and their retention of target proteins (IV.4.3 and IV.7).

A. The three depth filters with the longest lifetime retained less than 20% of the DsRed but up to 30% of 2G12 and 50% of TSP. PDF4 retained the lowest levels of target protein and PB2+PC2 the highest. Error bars indicate standard deviations ($n = 3$). B. The lifetimes of six depth filters matching with the filtrate turbidity performance cutoff of <10 NTU (blue diamonds) were used to predict the lifetime of four non-tested filters (open red circles) based on the observed quadratic correlation between retention number and filter capacity (red dashed line). The actual observed values for the four filters (blue circles with red border) were then used to build a refined cubic model (blue line).

The six depth filters achieving a turbidity of less than 10 NTU in the filtrate (Figure VI-6) were selected from the initial test set (Table VI.3) and used to establish a correlation between filter lifetime and the nominal retention rating of the depth filter layers (Figure VI-8 B). A good quadratic correlation ($r^2 = 0.89$) was found between filter lifetime and the average ratio of nominal retention ratings for each pair of subsequent filter layers divided by the total number of layers. This dimensionless number was called the retention number (RN , Equation 43). The correlation was used to predict the filter lifetimes of four non-tested depth filters (Table VI.3, Figure VI-8 B). Experimental verification resulted in strong agreement with the prediction for RN values below 5.0. Predictions for larger values showed significant deviation from the quadratic model. Therefore, a new cubic model was established that provided a good correlation between filter lifetime and RN ($r^2 = 0.90$). Unlike the quadratic model, all predictions of the cubic model resulted in physically meaningful values for the filter lifetime (>0.0 L m⁻²). The physical interpretation of RN was that low filter lifetime was associated with the use of only one filter layer (or multiple layers with an identical retention rating, $RN \approx 0$) or the use of layers with large differences in their individual retention ratings ($RN > 7.5$). Long filter lifetime was related to moderate differences in the retention ratings and a limited number of filter layers ($1.5 < RN < 5.0$). Thus, a limited number of filtration experiments can be used for the rational

selection of additional filter setups with extended lifetimes. A theoretical approach to determine the optimal number and filter layer depth has been reported [299].

$$RN = \frac{\frac{r_1}{r_2} + \frac{r_2}{r_3} + \dots + \frac{r_{n-1}}{r_n}}{(n-1)n} = \frac{\sum_{i=1, j=2}^{i=n-1, j=n} \frac{r_i}{r_j}}{(n-1)n}$$

Equation 43: Dimensionless retention number RN as an indicator for depth filter lifetime depending on the nominal retention ratings of the filter layers. With n as the total number of layers, r_i as the nominal retention rating of the more porous layer in the pair, and r_j as the nominal retention rating of the finer layer in the pair.

Table VI.4: Pilot scale comparative testing of the PB2+PC and PDF4 filter assembly using plants grown in soil.

Process	Initial		Improved
Depth filter type	PB2	PC2	PDF4
Nominal retention rates [μm]	8.0/1.0	1.0/0.3	3.0-6.0/0.4-0.8
Filter capacity [L m^{-2}]	103	200	101
Loading flow rate [$\text{L min}^{-1} \text{m}^{-2}$]	3	3	2,7
Turbidity after filter [NTU] (pooled filtrate)	20	<5	3
Approximate filter area required to process 1000 L plant extract (= 250 kg biomass)	6	2	10
Filter preparation [solution; L m^{-2}]	buffer; 200		water; 50
Filter module sizes [m^2]	2	2	1.0;0.5 and 0.25
Price per m^2 filter area [€]	480 (with 30% discount)	480 (with 30% discount)	350 (still >10% discount possible)
Number of filter modules needed to process 1000 L extract	3	1	10
Approximate direct filter costs for processing 1000 L extract	2,880	980	3,500
Approximate costs for tubing, bags and connectors [€]		3,500	900
Total process costs for 1000 L extract [€] (without labor)		7,360	4,400
Work load for system setup [person-hours]		24	2
Work load for system disassembly [person-hours]		4	0,5
Contamination risk (due to number of tubing connections)		high	very low
Process complexity (due to number of pumps, valves etc.)		very high	very low
Approximate space requirement for support equipment (holder, storage tanks etc.) [m^2]		25	6
Ease of filter handling (due to component size, venting, number of connections)		poor	good

The performance of the PDF4 filter module, which had the longest filter lifetime in the scaled down test, was compared to the original PB2+PC2 filter setup in a pilot scale process. This showed that the direct costs of the filter were lower, but more importantly there were significant savings in terms of associated consumables (e.g. tubing) and the time taken to set up the filtration train (Table VI.4).

In summary, PDF4 was identified as the most suitable depth filter for the clarification of extracts from plants grown in soil or rockwool because of its extended lifetime, minimal interactions with target proteins and low cost. PDF4 was also easier to handle than the original PB2+PC2 setup because the three-step depth filtration train (PB2+PB2+PC2 cascade [255]) was replaced with a single filter, with simplified stock keeping and ordering. The consumables costs were lower because less tubing was required, and operational costs were reduced because the filter could be mounted quickly, reducing the process downtime. Filter lifetime was generally shorter when plants were grown in rockwool rather than soil, but rockwool is advantageous in terms of low bioburden and process consistency. Therefore, additional measures were investigated to increase filter lifetime when plants are grown in rockwool.

VI.5 Reducing extract turbidity by flocculation

The results presented in this section have been published as the following manuscript:

1. Buyel JF, Fischer R. 2014. Flocculation increases the efficacy of depth filtration during the downstream processing of recombinant pharmaceutical proteins produced in tobacco. *Plant Biotechnology Journal* 12(2):240-52.

VI.5.1 Screening of flocculants

Eighteen different flocculants were tested for their ability to reduce extract turbidity after bag filtration, therefore increasing depth filter lifetime and reducing consumables costs (IV.8, Table VI.5). A set of DoE approaches was used to evaluate the flocculants under different buffer pH and conductivity conditions to guarantee the identification of a flocculant that is compatible with a broad range of operational conditions (IV.12.1).

In the initial screen, the Praestol flocculants 822 BS, 855 BS, 2350 and 2610, as well as Sedipur CL 950, Sedipur CL 951 and Polymin P, were found to reduce the turbidity of the crude extracts at two or three different pH values (Figure VI-9). Sedipur CL 950 and 951 have very similar physicochemical properties (Table VI.5), therefore only Sedipur CL 950 was selected for further testing because the slightly lower charge density compared to Sedipur CL 951 reduces the risk of protein precipitation.

Table VI.5: Flocculants used to precipitate particulate matter from tobacco extracts prior to depth filtration.

Flocculant	Chemistry [-]	Charge [-]	Charge density [meq g ⁻¹]
Praestol 2350	Polyacrylamide	anionic	6.9
Praestol 2610	Polyacrylamide	anionic	1.4
Praestol 2640	Polyacrylamide	anionic	5.5
Praestol 822 BS	Polyacrylamide	cationic	1.0
Praestol 851 BC	Polyacrylamide	cationic	1.2
Praestol 855 BS	Polyacrylamide	cationic	2.8
Praestol 859 BS	Polyacrylamide	cationic	4.6
Lupamin 9095	Polyvinylformamide	cationic	12.0
Magnafloc LT 37	Polydiallyldimethylammonium chloride	cationic	4.7
Magnafloc LT 38	Polydiallyldimethylammonium chloride	cationic	4.7
Polymin P	Modified polyethylenimine	cationic	13.0
Polymin SK	Modified polyethylenimine	cationic	6.5
Polymin VT	Modified polyethylenimine	cationic	3.2
Sedipur CL 950	Polyamine	cationic	7.0
Sedipur CL 951	Polyamine	cationic	7.6
ZETAG 7587	Polyacrylamide	cationic	4.5
pDADMAC 0.8MDa	Polydiallyldimethylammonium chloride	cationic	6.1
pDADMAC 1.5MDa	Polydiallyldimethylammonium chloride	cationic	6.1
Lupasol PS ¹	Modified polyethylenimine	cationic	19.0
Catofast GM ¹	Modified polyethylenimine	cationic	11.0
Paragas ¹	Modified polyethylenimine	cationic	14.0
Catofast VSH ¹	Polyvinylformamide	cationic	7.5
ZETAG 7109 ¹	Dimethylaminoethyl acrylate methyl chloride	cationic	5.5

¹ These flocculants were selected for a second round of screening according to their chemical properties based on the initial screening results.

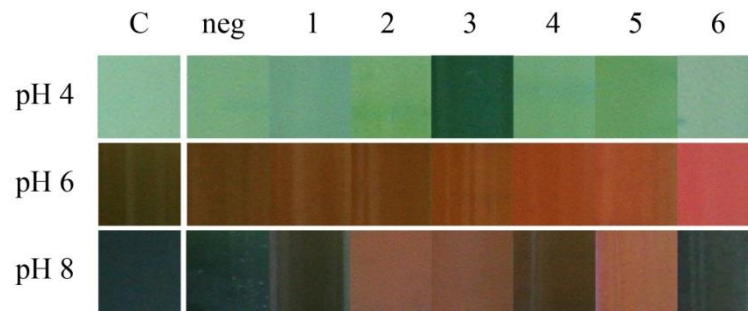


Figure VI-9: Impact of different flocculants on the turbidity of tobacco extracts at different pH values (IV.8 and IV.12.1).

Compared to an untreated extract (C), the flocculants Sedipur CL 950 (1), Praestol 2610 (2), 2350 (3), 855 BS (4) and 822 BS (5), as well as Polymin P (6), reduced the turbidity of the extract and the green color at two or all three of the tested pH values after settling for 24 h. The other flocculants had a more limited effect such as pDADMAC 1.5 MDa (neg).

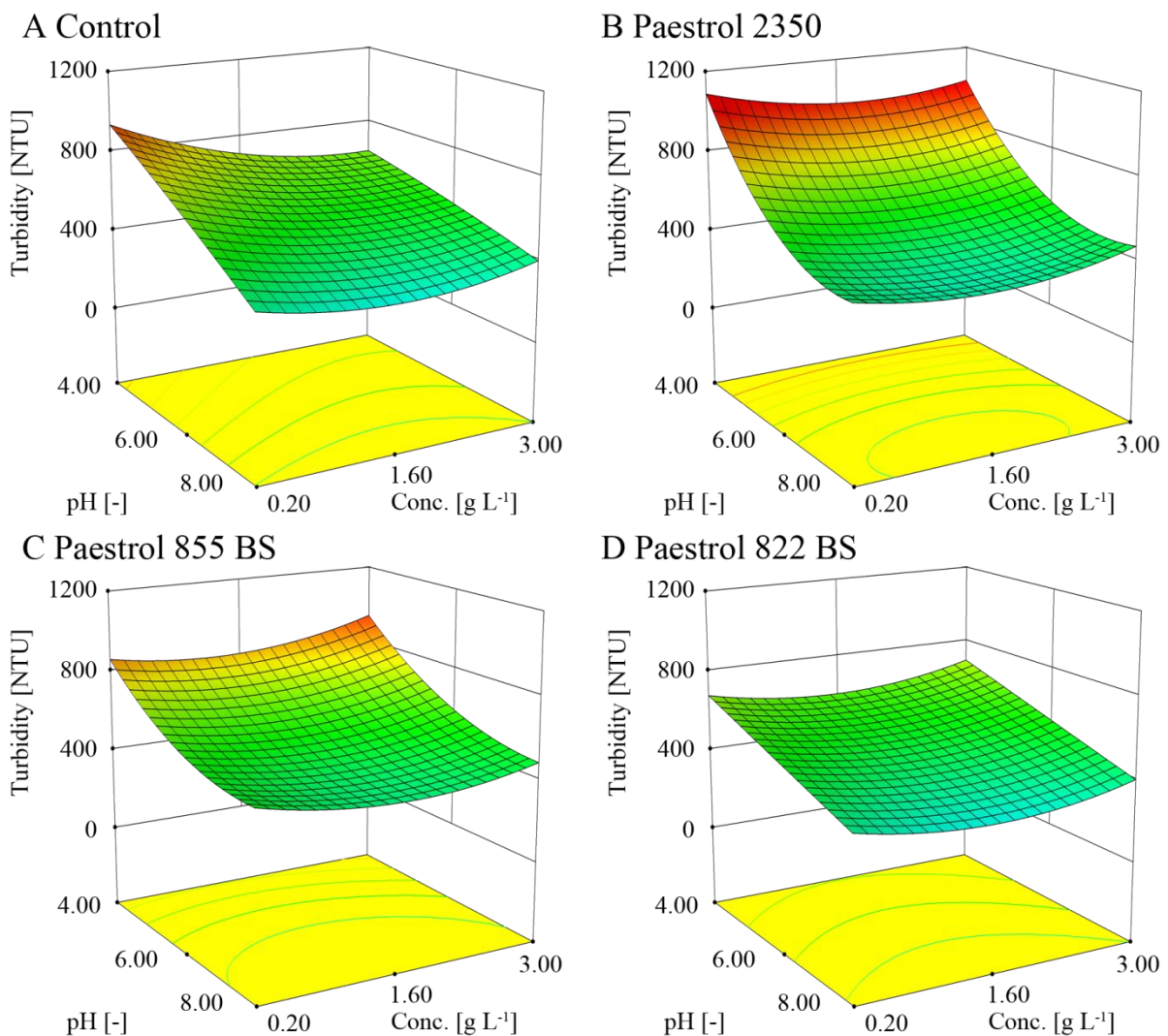


Figure VI-10: RSM for the reduction of tobacco extract turbidity achieved with different flocculants immediately after filtration, depending on concentration and pH (IV.8 and IV.12.1).

Praestol 2350 (B) and 855 BS (C) increased the extract turbidity under low pH conditions compared to flocculant-free extract (A) and had no beneficial effect otherwise, whereas Praestol 822 BS (D) reduced turbidity slightly. The turbidity values are shown as 1:10 dilutions in buffer.

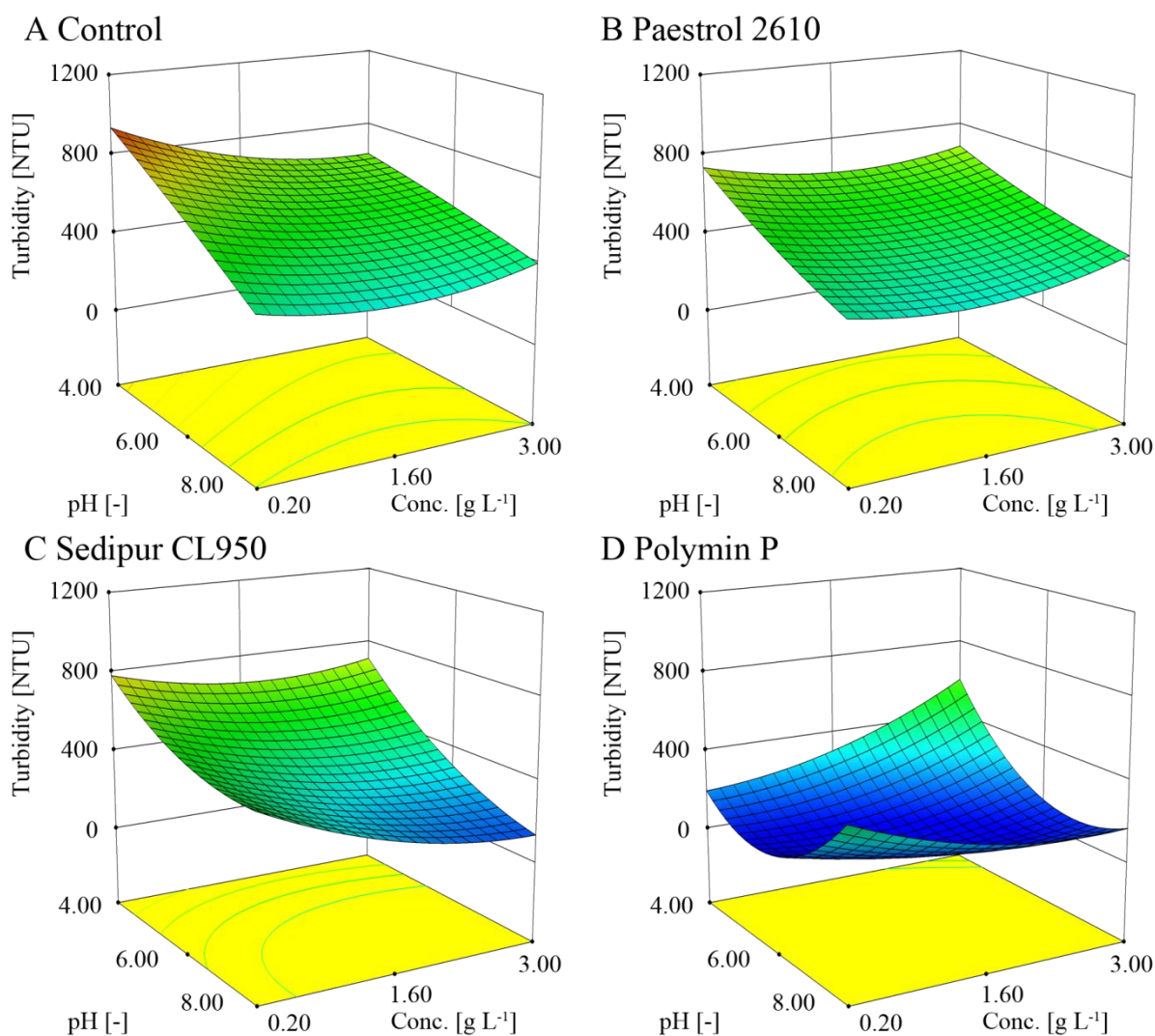


Figure VI-11: RSM for the reduction of tobacco extract turbidity achieved with different flocculants immediately after filtration, depending on concentration and pH (IV.8 and IV.12.1).

Praestol 2610 (B) reduced turbidity only slightly under low pH conditions compared to flocculant-free controls (A). Sedipur CL 950 (C) achieved a significant reduction in turbidity under high pH conditions at high concentrations, whereas Polymin P (D) reduced turbidity over a broad pH range even at low concentrations. The turbidity values are shown as 1:10 dilutions in buffer.

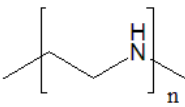
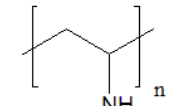
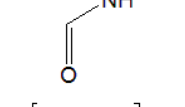
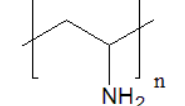
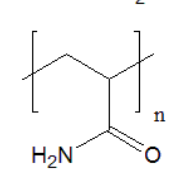
The effect of these flocculants was tested in more detail in a second set of experiments and evaluated directly after filtration. This revealed that Polymin P reduced the turbidity of the extract most efficiently, followed by Sedipur CL 950, then Praestol 822 BS and 2610 which performed equally well, and then Praestol 855 BS and 2350 which had the most limited effect (Figure VI-10 and Figure VI-11).

The efficacy of Polymin P, Sedipur CL 950 and Praestol 2610 was confirmed at pH 5 and 7. Praestol 2610 was used in preference to 822 BS because the stock solution had a lower

viscosity, which can simplify dosing during process-scale production. Polymin P was again the best flocculant, reducing extract turbidity efficiently even for pre-filtration incubation times of less than 1 h. Sedipur CL 950 had a limited effect on turbidity after 1 h and Praestol 2610 appeared to have no effect at all within this incubation time.

These results agree with the principles of flocculation presented in section III.6.1.2. The highly-charged cationic polymer Polymin P achieved the fastest and most effective flocculation of dispersed cell debris, probably reflecting charge neutralization and electrostatic interactions with negatively-charged cell fragments [79, 143]. The same mode of action can be attributed to Sedipur CL 950, but its lower charge density probably explains the slower flock formation. In contrast, the almost neutral Praestol 2610 is likely to promote bridging flocculation alone, resulting in slower and less effective flocculation than the two other polymers [79]. Cationic polymers generally caused more effective flocculation than anionic polymers, but the optimal concentration of all flocculants was in the range of 1 g L^{-1} , corresponding to 4 mg polymer per gram of suspended solids (wet mass).

Table VI.6: Physicochemical properties of flocculants representing different polymer classes.

Flocculant	Average polymer mass [kDa]	Macro structure [-]	Chemical structure [-]
Polymin P	800	branched	
Sedipur CL 950	50-200	branched	
Catofast VSH	400	linear	
Lupamin 9095	400	linear	
Praestol 2610	5000	linear	

The polymer backbone chemistry also had an important impact on flock formation induced by cationic polymers. The most efficient were those with amine or imine functional groups alone, i.e. polyamines (Sedipur CL class) and polyethyleneimines (Polymin class). In contrast, the carbonyl groups present in polyvinylformamide (Lupamin class) and polyacrylamide (Zetag and Praestol classes) slowed down flocculation and reduced the aggregation of cell debris even when the polymers were highly-charged. This could reflect the

greater hydrophilicity of polymers containing carbonyl groups, which can reduce their interactions with lipids in the cell debris and hence limit flock formation. Longer and/or branched polymers were more efficient flocculants than shorter, linear ones even if the charge density was similar (Polymin P vs. Lupamin 9095) or higher (Sedipur CL 950 vs. Lupamin 9095) (Table VI.5 and Table VI.6). Branching and high charge density also promoted rapid flock formation (<10 minutes).

VI.5.2 Flocculant selection and optimization

Based on these findings, additional cationic polymers with high charge densities were selected for a second round of screening (Table VI.5). Catiofast VSH and Lupasol PS achieved the efficient flocculation of tobacco cell debris under standard extraction conditions (pH 8.0, conductivity 50 mS cm^{-1}) and were chosen for direct comparison with Polymin P (IV.12.1).

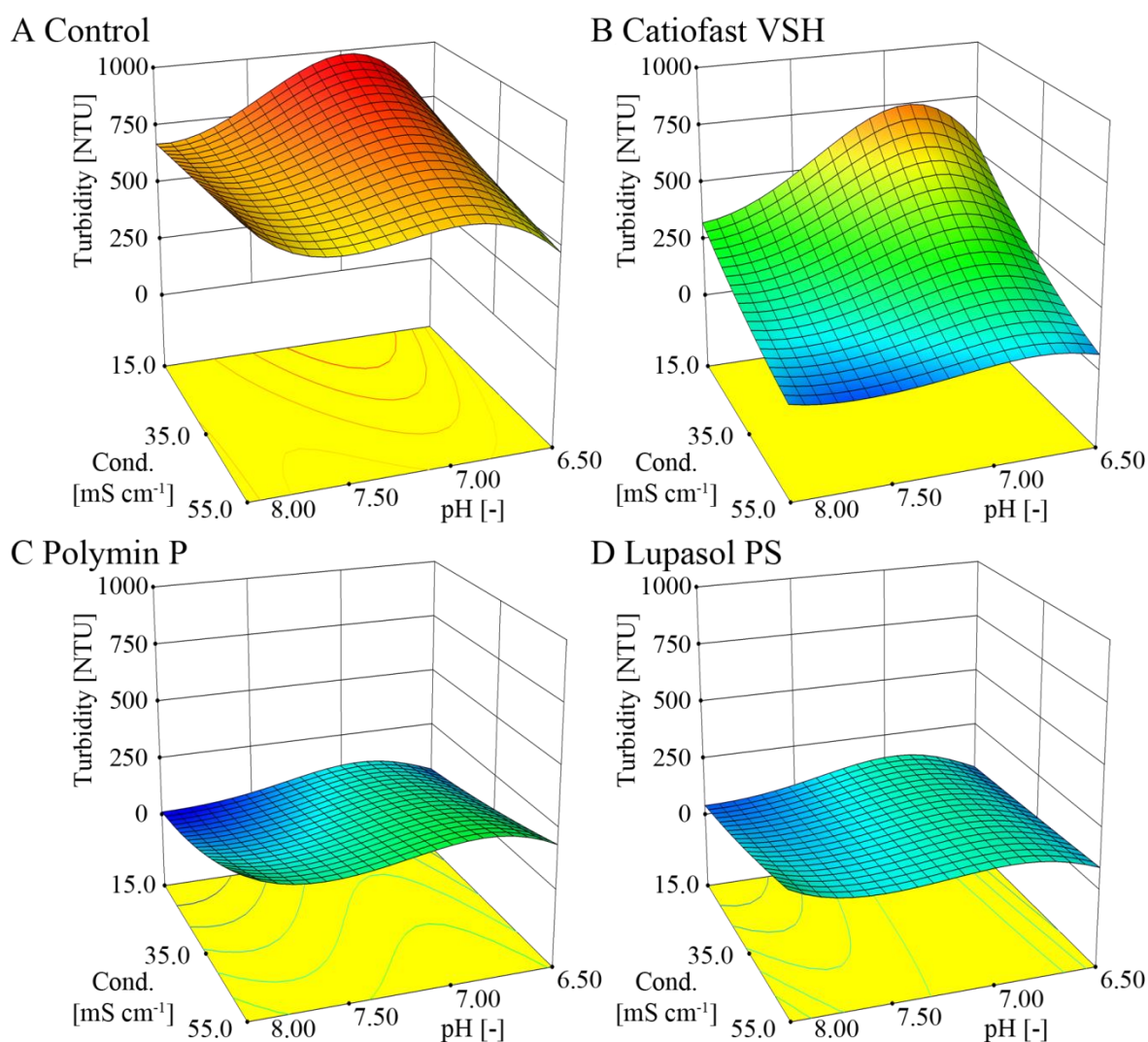
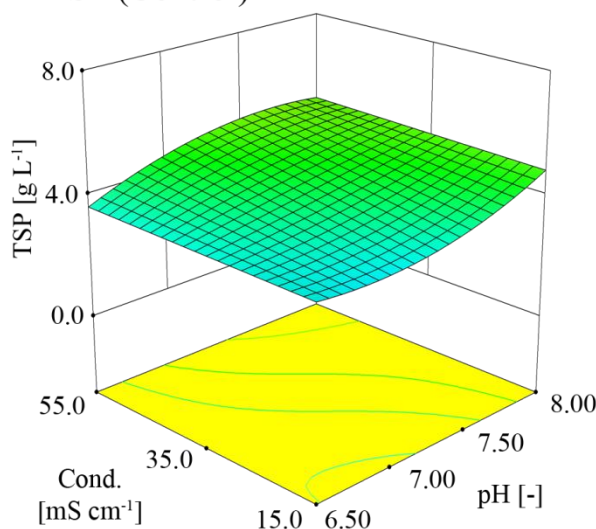


Figure VI-12: Turbidity reduction achieved with highly-charged cationic polymers (IV.8 and Table VI.5).

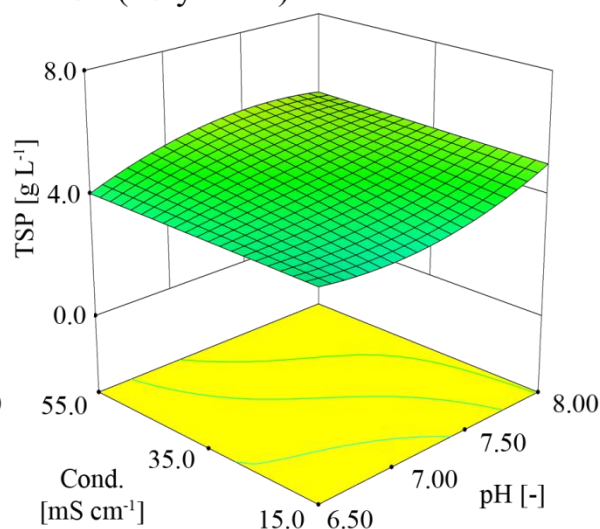
Turbidity in non-treated extract varied with pH and conductivity (A). Catiofast VSH (B) reduced extract turbidity at high conductivities, whereas Polymin P (C) and Lupasol PS (D) reduced extract turbidity at moderate and low conductivities regardless of the extraction buffer pH. All response surfaces were calculated for 4.0 g L⁻¹ of polymer and an incubation time of 15 min.

At conductivities >35 mS cm⁻¹, Catiofast VSH reduced the extract turbidity by ~90% but had little impact at lower conductivities (Figure VI-12 B). This complemented the activities of Lupasol PS and Polymin P which performed best at conductivities <45 mS cm⁻¹ and achieved similar reductions in turbidity (Figure VI-12 C and D) but flocculation was more rapid with Polymin P and required about 10% less polymer.

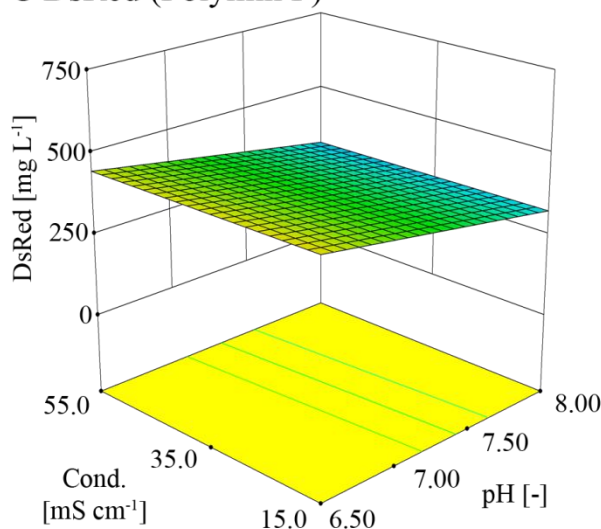
A TSP (Control)



B TSP (Polymin P)



C DsRed (Polymin P)



D 2G12 (Polymin P)

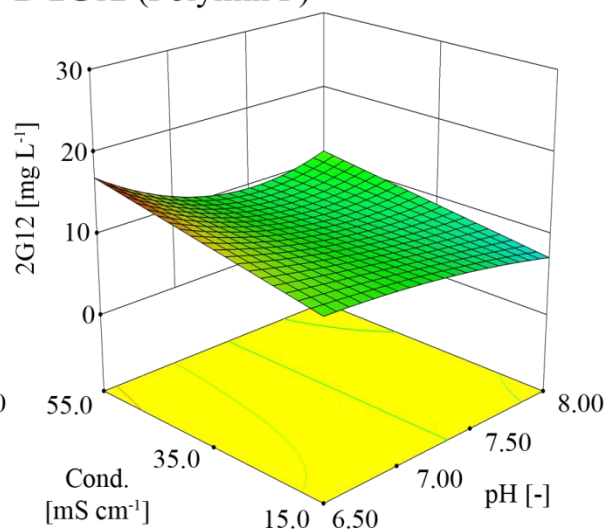


Figure VI-13: Effect of flocculants on TSP, DsRed and 2G12 concentrations in leaf extracts (IV.4.3 and IV.8).

None of the polymers affected TSP, DsRed or 2G12 concentrations significantly compared to controls. The example shown here is impact of control vs. Polymin P on TSP (A versus B). More DsRed and 2G12 were extracted at pH 6.5 (C and D respectively), whereas most TSP was extracted at higher pH and conductivity levels (A). All response surfaces were calculated for 2.0 g L⁻¹ of polymer and an incubation time of 15 min.

The concentrations of TSP, DsRed and 2G12 varied 5–10% among the polymers and controls, which is within the error ranges of the models and quantitation assays (Figure VI-13).

All three polymers therefore had an insignificant impact on protein levels within the tested concentration range ($0.5\text{--}4.0\text{ g L}^{-1}$) and did not interfere with target protein recovery. Regardless of the polymer type, more TSP was extracted at higher pH and conductivity levels, but a slightly acidic pH ($5.0\text{--}6.5$) was favorable for the extraction of DsRed and even more so for 2G12, particularly at high conductivities.

The effect of polymer combinations on extract turbidity was also tested but the simultaneous use of Polymin P and Praestol 2610, or Polymin P and Kaolin, a commonly used co-flocculant based on aluminum silicate, did not result in increased flocculation (data not shown).

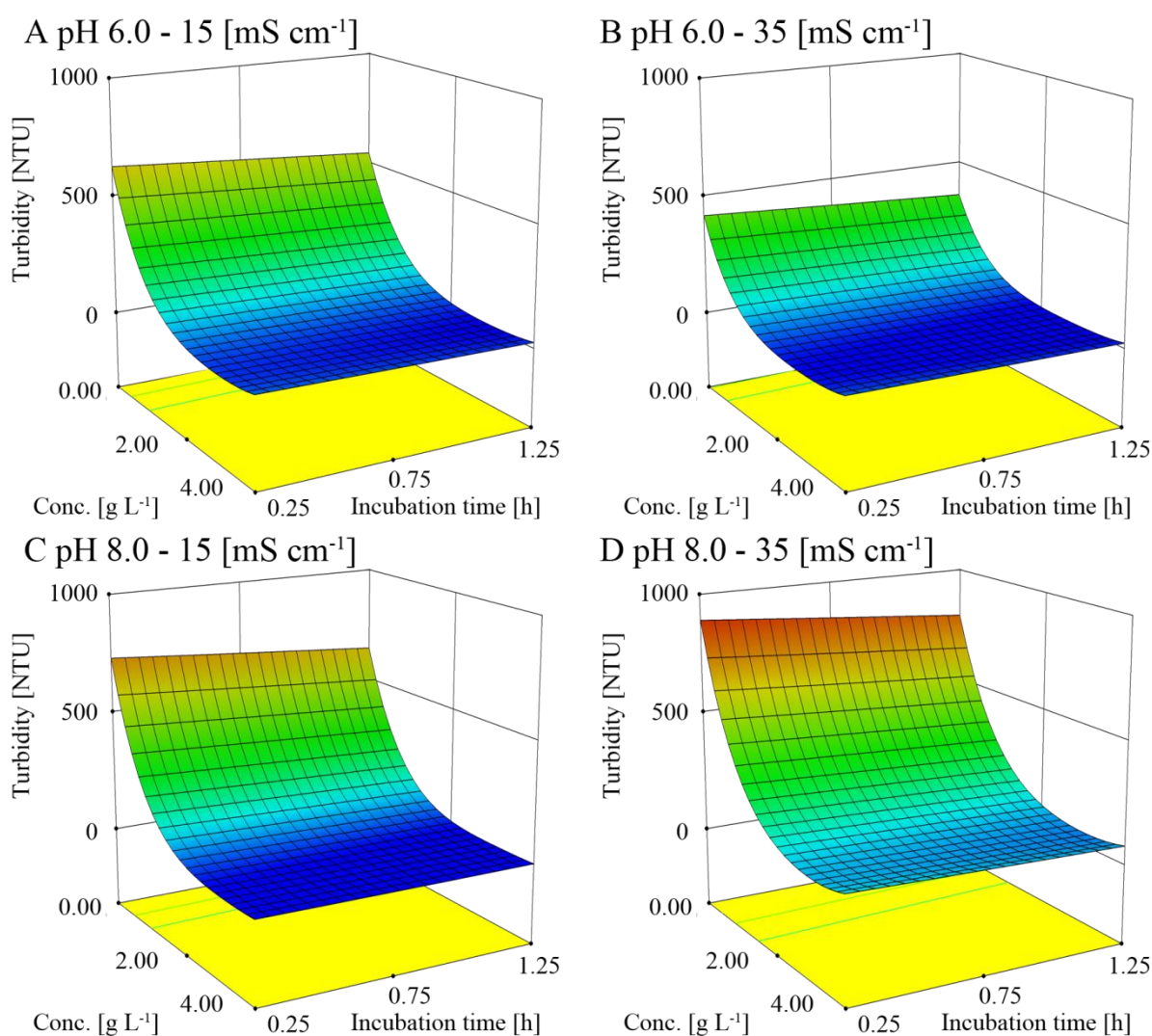


Figure VI-14: RSM for the effect of Polymin P concentration, incubation time, pH and conductivity on extract turbidity determined directly after bag filtration (IV.8 and IV.12.1).

A. Polymin P concentration of 2 g L^{-1} was sufficient for full flocculation and increasing the incubation time slightly reduced the turbidity of the extracts. The turbidity was lower at pH 6 (top row) than at pH 8 (bottom row). Increasing conductivity from 15 mS cm^{-1} (left column) to 35 mS cm^{-1} (right column) improved extract clearance at low pH (B) but increased turbidity at high pH (D). The turbidity values are shown as 1:10 dilutions in buffer.

Polymin P was selected for further experiments to test additional process parameters including incubation time and temperature (Table IV.5). Polymin P concentrations of 1.7–3.5 g L⁻¹ were optimal, reducing extract turbidity below 10% of the initial value, whereas a steep increase in extract turbidity was observed at lower concentrations (Figure VI-14). The concentrations were calculated based on the Polymin P formulation provided by BASF, which was a 50% solution in water. Therefore, the effective concentrations are only 50% of the values stated in the text. A slight increase in extract turbidity was observed at concentrations >3.5 g L⁻¹ probably reflecting an over-saturation effect as previously reported and indicating that Polymin P also acted by bridging flocculation [81] (III.6.1.2). The efficiency of Polymin P declined in the pH range 7.0–7.7 and the extract appeared milky after bag filtration indicating the presence of fine, dispersed precipitates. These were probably fragments of lignified cell walls, which are uncharged in this pH range [300] and thus unlikely to interact with the polycation, preventing formation of flocks.

High conductivities improved the performance of Polymin P in low-pH extracts but had the opposite effect at high pH values (Figure VI-14). This may reflect the ability of salt ions to shield positive particle charges at low pH, thereby preventing the repulsion of polycationic flocculants, but to compete for binding sites on the particles at high pH, thereby preventing interactions with the flocculant. Increasing the incubation time from 0.25 h to 1.25 h after the addition of Polymin P improved clarification but the effect was small compared to that of the other factors highlighting the rapid kinetics of flock formation.

VI.5.3 Robustness of flocculation

Because process economics and regulatory compliance favor the minimal use of additives, a Polymin P concentration of 2 g L⁻¹ (effective concentration 1 g L⁻¹) was selected for further tests because this was the lowest concentration that achieved complete flocculation. The robustness of flocculation was investigated because the impact of small changes in other factors that control turbidity can help to determine the potential impact of errors in buffer preparation, flocculant dosing or fluctuations in temperature on the process. However, robustness was assessed over a wider parameter range to test the compatibility of Polymin P with a broad variety of extraction conditions that may be required for different target proteins. It was possible to establish a model that reliably predicted the effects of flocculant concentration, incubation temperature, pH, conductivity, and incubation time on extract turbidity (Table VI.7 and Table VI.8).

Table VI.7: Model factors with a significant impact on the turbidity of tobacco extracts during flocculation.

Source	Sum of squares	Degrees of freedom	F-value	p-value
Model	3205.08	29	39.64	< 0.0001
Incubation temperature [°C] (A)	2.93	1	1.05	0.3114
pH [-] (B)	205.50	1	73.72	< 0.0001
Conductivity [mS cm ⁻¹] (C)	98.86	1	35.46	< 0.0001
Incubation time [h] (D)	61.29	1	21.99	< 0.0001
Concentration [g L ⁻¹] (E)	3.35	1	1.20	0.2795
AB	42.86	1	15.38	0.0003
AC	2.31	1	0.83	0.3683
AD	0.07	1	0.03	0.8732
AE	13.73	1	4.93	0.0322
BC	2.27	1	0.81	0.3724
BD	17.23	1	6.18	0.0172
BE	149.20	1	53.52	< 0.0001
CD	22.42	1	8.04	0.0071
CE	7.27	1	2.61	0.1142
A ²	31.07	1	11.14	0.0018
B ²	319.38	1	114.57	< 0.0001
C ²	13.46	1	4.83	0.0338
D ²	23.71	1	8.51	0.0058
E ²	0.15	1	0.05	0.8176
ABC	91.52	1	32.83	< 0.0001
ABD	22.12	1	7.94	0.0075
ACD	36.53	1	13.10	0.0008
A ² E	33.46	1	12.00	0.0013
AC ²	17.31	1	6.21	0.0170
B ² C	563.56	1	202.15	< 0.0001
BC ²	61.84	1	22.18	< 0.0001
BE ²	18.24	1	6.54	0.0144
C ² D	12.00	1	4.31	0.0445
B ³	408.64	1	146.59	< 0.0001
Residual	111.51	40	n.a.	n.a.
Lack of fit	100.05	32	2.18	0.1239
Pure error	11.46	8	n.a.	n.a.

Table VI.8: Parameters used to evaluate the flocculation model.

Evaluation parameter	Value
r^2	0.966
Adj. r^2	0.942
Pred. r^2	0.887
PRESS ¹	372.509

¹ Predicted residual sum of squares

It was challenging to use response surfaces to identify factor settings associated with high or low extract turbidity due to the complexity of the flocculation model. Instead, the flocculation model was used to predict conditions resulting in desirable and undesirable extract turbidities and the factor levels associated with these conditions were then analyzed (Figure VI-15). This revealed that lower extract turbidities were associated with pH values in the range 4.5–5.5 or ~8.0 as well as incubation times >15 min. No clear trend was observed for incubation temperature or conductivity indicating a broad operational window for these parameters. Interestingly, extreme Polymin P concentrations were also preferred, but this was probably an artifact reflecting the inclusion of the non-significant quadratic effect of the Polymin P concentration (E^2 in Table VI.7). This effect was included to maintain the model hierarchy [276, 277].

Polymin P concentrations <2 g L⁻¹ were associated with high turbidity, as were pH values of ~4.0 or ~7.5 and incubation times <15 min. Incubation temperature and conductivity did not exhibit a clear trend. For certain factors such as conductivity with an ambiguous influence on turbidity, extreme high or low values were over-represented, suggesting the model predictions become increasingly uncertain at the borders of the design space, resulting in larger fluctuations in the calculated turbidity values favoring extreme predictions.

Polymin P reduced TSP levels by 20–25% at all pH values when present at a concentration of 2 g L⁻¹ in a buffer with a conductivity of 15 mS cm⁻¹ (Figure VI-17 A) confirming previous studies suggesting that Polymin P can be used to precipitate proteins at low conductivities [78]. However, SDS-PAGE analysis did not reveal significant changes in the protein bands or abundance, except at pH 4.0 (Figure VI-16) when flocculant-free samples were compared with Polymin P treated ones. Additionally, Polymin P had no effect on TSP at conductivities >30 mS cm⁻¹. The concentration of DsRed was reduced 15–20% by Polymin P, regardless of the pH or conductivity of the buffer, with the exception of extractions carried out at pH 7 in buffers with conductivities below 30 mS cm⁻¹. The beneficial effect of this combination probably reflected the ability of the charged flocculant to increase the buffer conductivity, thereby promoting the release of DsRed.

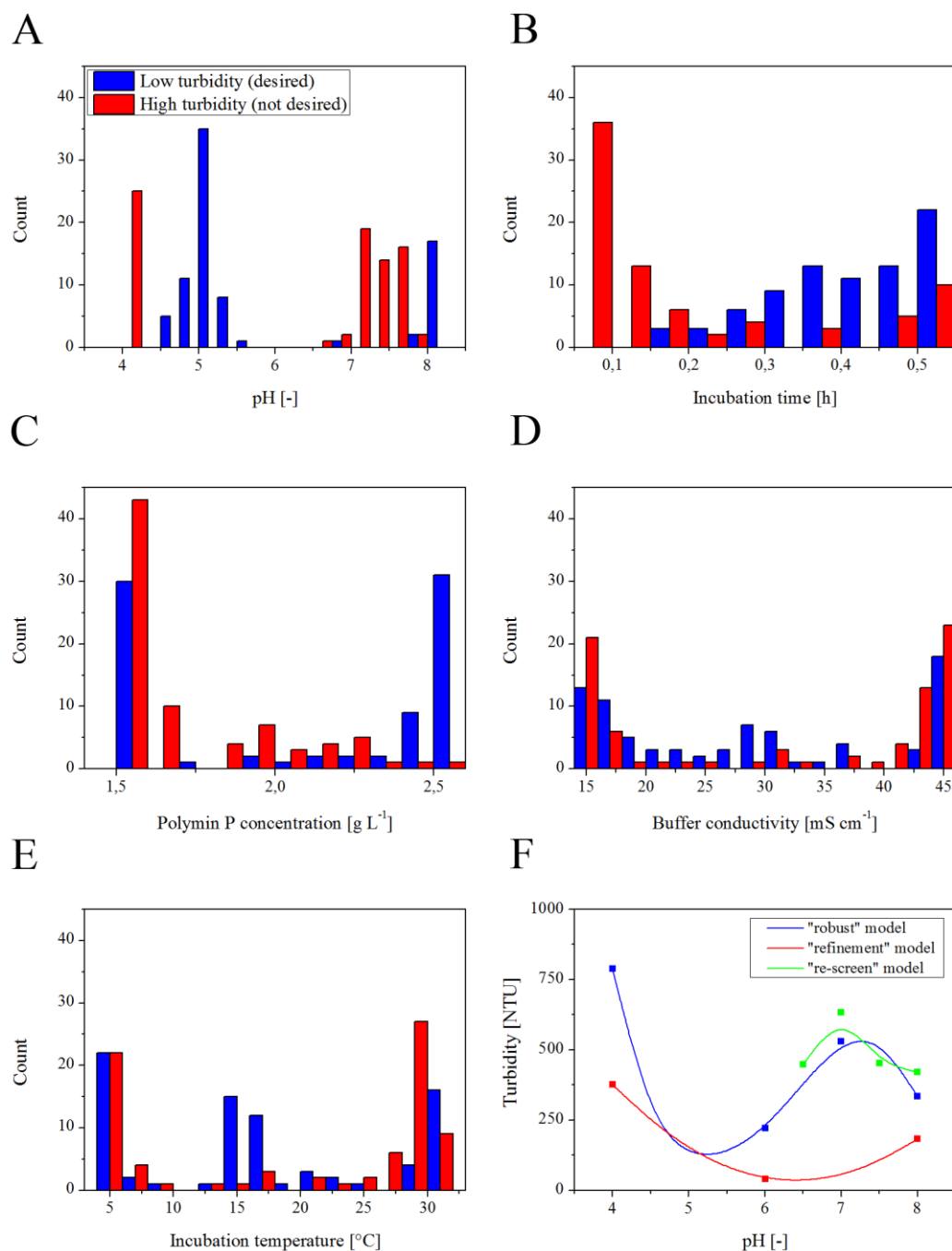


Figure VI-15: Predicted correlation between process parameters and extract turbidity (IV.8 and IV.12.1). The robustness model (Table VI.7) was used to predict conditions resulting in low (blue, desirable) or high (red, undesirable) extract turbidity. The parameter values for pH (A), incubation time (B), Polymin P concentration (C), conductivity (D) and incubation temperature (E) were analyzed separately, revealing their correlation to low or high turbidities. F. The robustness model was compared to models built during the second flocculant screening and initial optimization, revealing a large discrepancy at pH ~7 indicating that the initial model had limited predictive power over this pH range.

The presence of Polymin P in leaf extracts did not alter the concentration of 2G12 by more than 10% at pH 6.0 or above, regardless of the conductivity, but the concentration of 2G12 fell by 30% at pH 4.0.

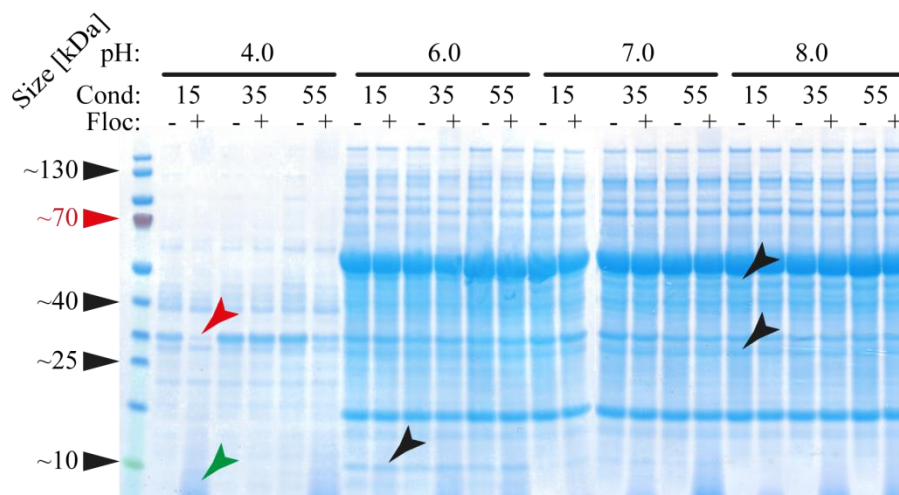


Figure VI-16: TSP composition of tobacco extract in dependence of pH, conductivity and presence of flocculant (IV.4.3, IV.4.4, IV.8 and IV.12.1).

Only at pH 4.0 and 15 mS cm⁻¹ the addition of 2.0 g L⁻¹ Polymin P (+) resulted in a visible reduction of TSP concentration (red arrow indicates diminished DsRed band) compared to polymer-free samples (-). A characteristic coloration below 10 kDa appeared in all polymer containing samples (green arrow). Depending on the pH, certain HCP were more or less extracted (black arrows): a 10 kDa protein at pH 6.0 as well as a 33 kDa and 50 kDa protein at pH 7.0 and 8.0.

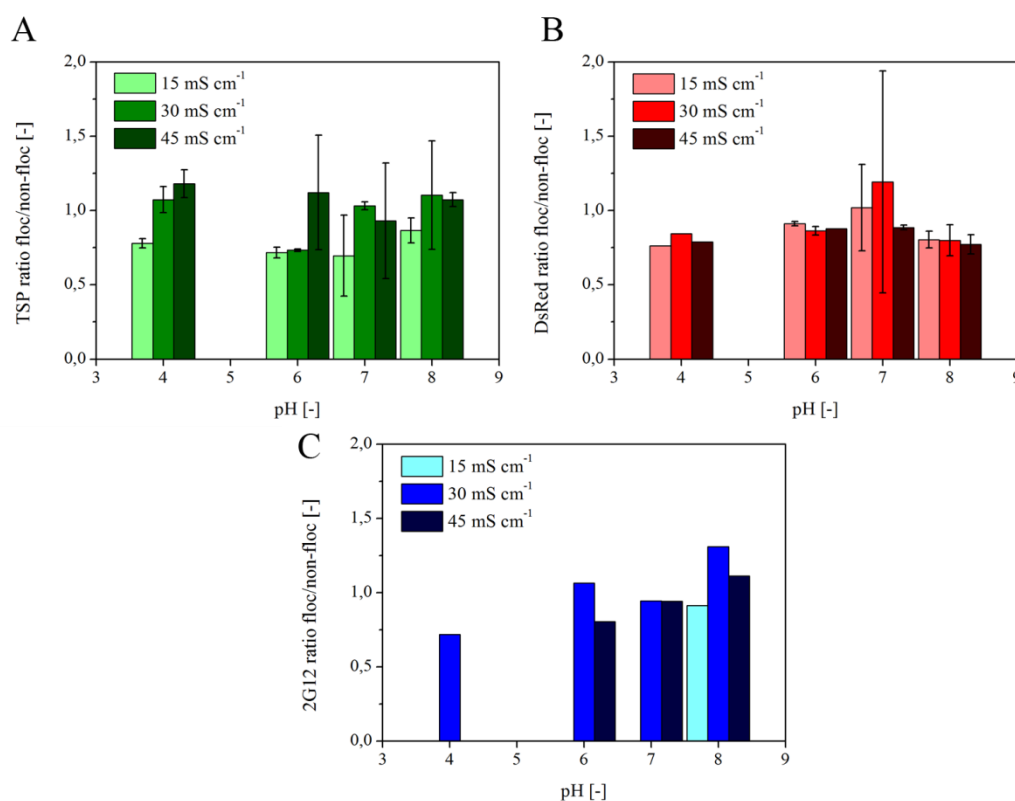


Figure VI-17: Effect of Polymin P on the concentrations of TSP, DsRed and 2G12 in leaf extracts (IV.4.3, IV.8 and IV.12.1).

A. Under low pH and low conductivity conditions, Polymin P reduced TSP levels by ~25%. B. Regardless of the conductivity, Polymin P slightly increased DsRed concentrations in extracts at pH 7.0 but otherwise reduced them by 15%. C. The level of 2G12 was reduced by <10% at pH values above 6.0, but reduced by 30% at pH 4.0. See text for comments on missing data. Error bars indicate standard deviation ($n \geq 3$). Error bars in C are missing due to the deleterious effect of the flocculant on the CM5 chip surface.

The precise quantitation of 2G12 under these conditions was difficult because the polymer generated a high background signal on the reference flow cell of the CM5 chip during SPR spectroscopy. This was expected due to the opposite charge of Polymin P (positive) and the CM5 matrix (carboxymethyl, negative) promoting electrostatic interactions especially in samples with low conductivities. However, the results agreed with those presented above indicating that flocculants have a limited impact on proteins under the high conductivity conditions tested here (VI.5.2).

VI.5.4 Process scale-up

The efficiency of Polymin P was confirmed in bench-top and pilot-scale tests. Manual mixing after the addition of flocculant was replaced with automated mixing in the homogenization device. Short mixing times of 10–15 s were optimal whereas longer mixing times reduced flocculation by disintegrating the flocks (Figure VI-18 A). In bench-top tests, Polymin P increased the lifetime of depth filters by three-fold (PB2+PC2 and PDF4) or seven-fold (PDH4) at concentrations of 1.5–2.0 g L⁻¹ (Figure VI-18 B). The flocculant probably increased the average diameter of particles in the extract by bridging flocculation, thus shifting the filter burden towards the first filter layer and favoring PDH4, with its larger nominal retention rating in this layer (Table VI.3). Filtrate turbidity was not affected by the use of Polymin P.

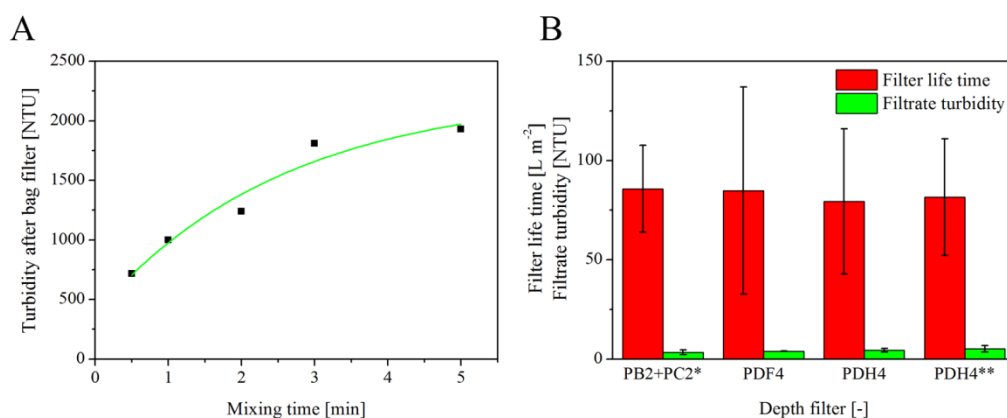


Figure VI-18: The mixing time dependence of flocculation and its effect on depth filter lifetime (IV.5, IV.7 and IV.8).

A. Increasing the mixing time reduced flocculant efficacy. B. Polymin P increased the lifetime of depth filters by 3–7-fold (depending on the configuration) compared to batches without flocculant (Figure VI-6 and Figure VI-8). The longevity increased at higher Polymin P concentrations, as observed for extract turbidity (Figure VI-14). The asterisk indicates a Polymin P concentration of 1.5 g L⁻¹, whereas 2.0 g L⁻¹ was used for the other setups. Error bars indicate standard deviation ($n \geq 3$).

Polymin P was found to increase TSP levels by 20–25% compared to a flocculant-free setup at all steps after homogenization whereas DsRed and 2G12 concentrations remained

unaffected under standard process conditions at pH 7.5–8.0 with conductivities ranging from 35 to 45 mS cm⁻¹ and at different process steps (Figure VI-19 B to D).

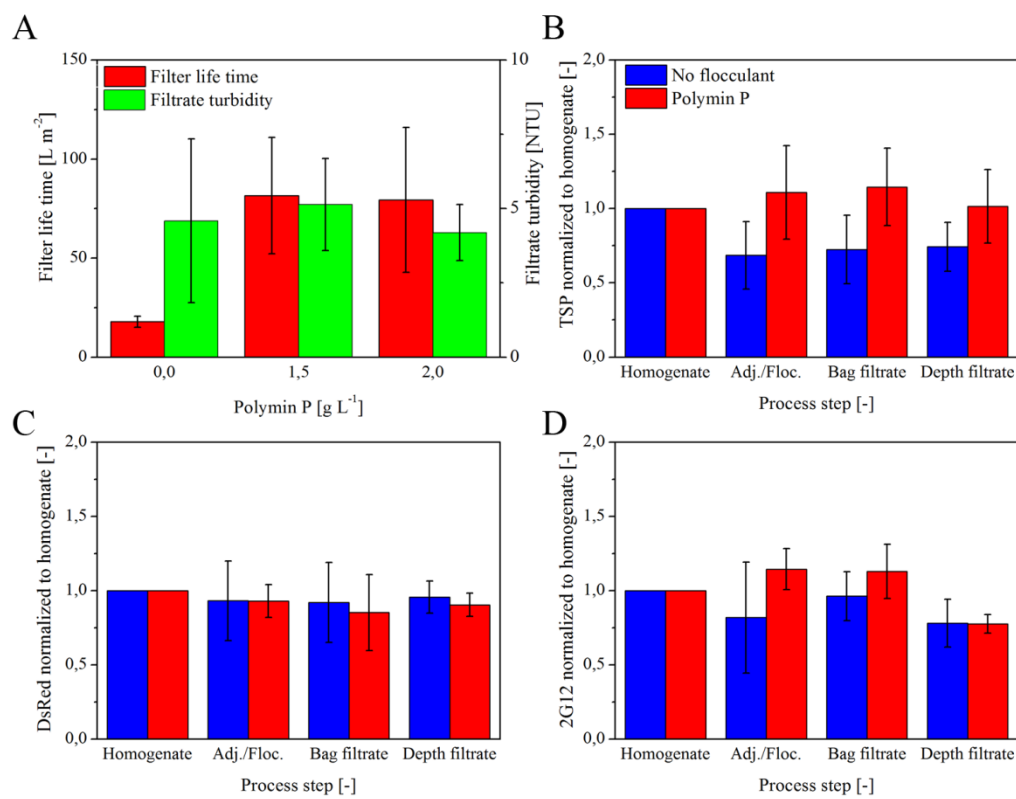


Figure VI-19: Comparison of process performance with and without Polymin P (IV.4.3, IV.5, IV.7 and IV.8).

A. Addition of Polymin P in the concentration range 1.5–2.0 g L⁻¹ increased depth filter lifetime approximately seven-fold. B – D. Polymin P did not reduce TSP or target protein levels compared to flocculant-free controls. Error bars indicate standard deviation (n ≥ 3).

When the process volume was increased to the 100 L scale, protein concentrations were not significantly affected by the addition of Polymin P (Figure VI-20). In contrast, the filter lifetime was increased substantially, to more than 280 L m⁻² compared to ~80 L m⁻² at the bench-top scale. This significant improvement can be attributed to two scale-dependent factors. First, the utilization of the upstream bag filter was higher at the 100 L scale, improving the filter performance by promoting the build-up of filter cake and reducing the particle burden to 500 NTU instead of ~5000 NTU in the feed for subsequent depth filtration. Second, the flow pattern at the filter layer surface changed from direct flow at the bench-top scale to tangential flow in the STAX modules (pilot scale), which can prevent premature clogging. Process scale-up is likely to reduce the costs further due to the optimized flow pattern and equipment utilization.

In conclusion, flocculants were shown to reduce the turbidity of the process stream after bag filtration and to increase depth filter lifetime to more than 110 L m⁻² (depending on the

scale and filter type) even when processing extracts from plants grown in rockwool. The best-performing flocculant was Polymin P.

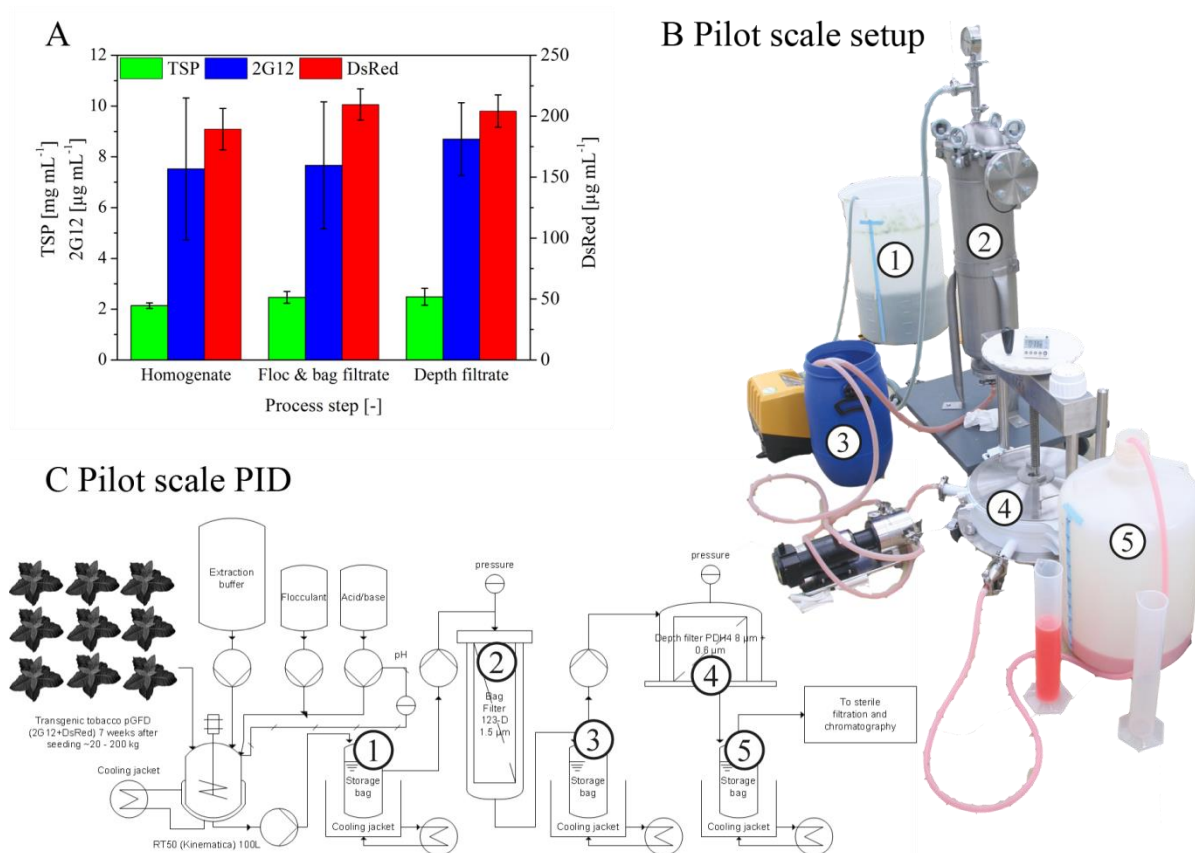


Figure VI-20: Pilot scale process validation (IV.4.3, IV.5, IV.7 and IV.8).

A. The levels of TSP, DsRed and 2G12 did not decline during the course of the filter train as also observed for the bench-top scale process (Figure VI-19 B to D). Error bars indicate the standard deviation of at least three technical replicates. B. A photograph of the pilot scale setup (without the homogenizer) indicating the position of the homogenate storage tank (1), bag filter (2), bag filtrate storage tank (3), depth filter (4) and depth filtrate storage tank (5). C. The piping and instrumentation diagram (PID) corresponding to B indicating the same devices.

The costs for filter consumables were reduced by more than 50% compared to a process without flocculant and by 20% compared to a process using plants grown in soil (without flocculant). There were no negative effects on target protein concentration. Polymin P also reduced the variance among depth filter lifetimes observed for the processing of different tobacco varieties and *Nicotiana* species (Figure VI-7 B). Natural polymers such as chitosan can be a sustainable alternative to synthetic polymers but recent studies using such flocculants in combination with plant extracts have shown that synthetic polymers were more effective in terms of turbidity reduction and thus of greater benefit for the clarification process [301].

VI.6 Screening of filter additives

Six cellulose-based filter additives were tested during bag or depth filtration alone or in combination with flocculants in order to improve particle clearance and depth filter lifetime (IV.8, Table VI.9). Additives included before bag filtration had only a moderate effect on filtrate turbidity and did not increase the lifetime of the subsequent depth filter, even if Polymin P was used at the same time (Figure VI-21 A, C200***).

Table VI.9: Cellulose-based additives tested during bag and depth filtration.

Additive	Manufacturer	Fiber size [-]
CelluFluxx F15	Erbslöh	Very fine
CelluFluxx F25	Erbslöh	Fine
LuvoZell C90	Lehmann & Voss	Mid
CelluFluxx F45	Erbslöh	Mid
LuvoZell C200	Lehmann & Voss	Coarse
CelluFluxx P30	Erbslöh	Coarse

This may have occurred because the bag filters used in this step were arranged in a vertical orientation as preferred for this filter type, which prevented the build-up of a solid filter cake by the additives and thus negated one of their two major modes of action. Second, the raw extract already contained many plant-derived cellulose particles from the homogenization process so even fine cellulose fibers such as CelluFluxx F25 did not improve particle retention. Instead, feeding these additives into the process stream before bag filtration increased the liquid hold-up, and thus reduced product recovery. When used during depth filtration, the additives alone did not significantly reduce the filtrate turbidity or enhance the depth filter lifetime (Figure VI-21 A, C200**).

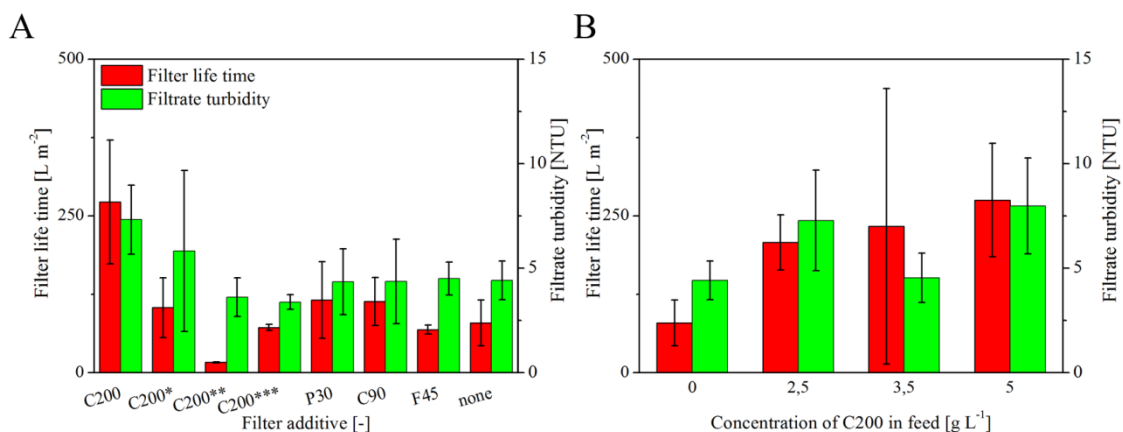


Figure VI-21: Effect of cellulose additives on depth filter lifetime (IV.7 and IV.8).

A. Adding coarse cellulose fibers (LuvoZell C200) to the feed stream increased the depth filter lifetime approximately three-fold compared to an additive-free setup. Other fibers, especially the finer ones, were less effective. The resulting filtrate turbidity was within the acceptance range of <10 NTU for all setups. * indicates a

reduced Polymin P concentration of 1.5 instead of 2.0 g L⁻¹; ** indicates the absence of Polymin P; *** indicates the addition of LuvoZell C200 during bag filtration instead of depth filtration. B. The effect of LuvoZell C200 was concentration dependent and the longest filter lifetime was achieved with 5 g L⁻¹. Higher concentrations resulted in premature blocking of the filter housing (see text for details). Error bars indicate standard deviations (n ≥ 3).

In contrast, when additives were used during depth filtration after Polymin P had been added prior to bag filtration, a significant (>3-fold) increase in the filter lifetime was achieved (Figure VI-21 A). The best performance was achieved with the coarse LuvoZell C200 fibers, and increasing their concentration in the feed stream to 5.0 g L⁻¹ was optimal in terms of filter lifetime and reproducibility (Figure VI-21 B). Higher concentrations of LuvoZell C200 produced poor results because (i) extensive stirring was required in the feed reservoir to prevent sedimentation of the fibers, potentially exposing the target proteins to oxidative stress and (ii) the resulting back pressure from the filter cake, rather than the blocked filter, became a limiting factor for the filter lifetime because all the head space in the filter housing was filled with cellulose fibers.

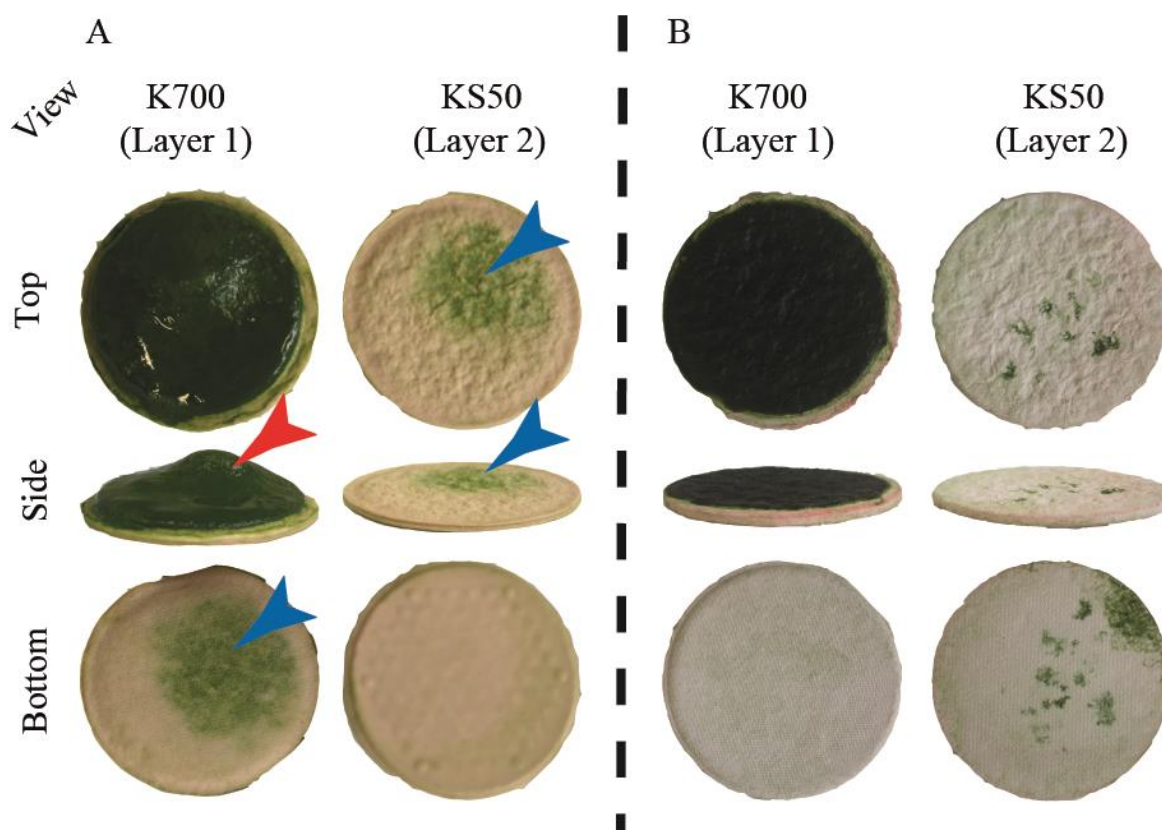


Figure VI-22: Effect of cellulose fibers on filter blockage (IV.7 and IV.8).

A. The addition of cellulose fibers resulted in the formation of a cake on the first depth filter layer (red arrow), thereby prolonging its permeability to the feed stream indicated by the particles passing onto the second filter layer (blue arrows). B. In the absence of cellulose fibers, a dark-green varnish-like coat rapidly formed on the first depth filter layer, limiting its permeability for dispersed particles and thus reducing the filter lifetime.

Other cellulose fibers, especially the finer ones, either increased the filter lifetime albeit to a lesser extent than LuvoZell C200 (e.g. LuvoZell C90) or had no impact on filter lifetime at all (Figure VI-21 A). For example, the filter cake formed by CelluFluxx F45 effectively

prevented small particles from entering and blocking the depth filter layers but at the same time formed a dense filter cake causing high back pressure and eliminating its own beneficial effect, again reducing the overall filter lifetime. This also occurred when the concentration of CelluFluxx F45 in the feed was reduced from 5.0 to 2.5 g L⁻¹. Combining fine and coarse fibers in the feed stream is another option to increase depth filter lifetime which could be tested in future experiments.

LuvoZell C200 and other cellulose fibers exhibited two simultaneous modes of action to extend filter lifetime. First, they prevented the formation of a varnish-like dark green coat on the upper depth filter layer by providing additional surfaces and binding sites for dispersed particles, resulting in a dark green filter cake (Figure VI-22). It therefore took longer to block the first filter layer and the filter lifetime was prolonged. Second, the filter cake provided a mesh-like scaffold with a loose structure to which particles could attach, reducing the back pressure and the number of particles entering the filter, and thus extending its lifetime.

In summary, cellulose fiber additives were identified as a safe and convenient strategy to increase depth filter lifetime in combination with flocculants. Their use will further reduce downstream costs during the manufacturing of plant-derived biopharmaceuticals and hence improve the competitiveness of this production platform. In the meanwhile, it was shown that the capacity can be increased about 1000-fold by an adequate combination of flocculants and filter aids compared to a setup without any additives [302].

VI.7 Economic impact of process improvements

The PDF4 depth filter was found to be superior in terms of process economics including direct costs, the costs of auxiliary equipment (e.g. tubing and connectors) as well as the labor required for operation as described in section VI.3 (Table VI.4). However, that comparison focused on the ‘original’ process setup for plants grown in soil. Future campaigns will probably use rockwool as the growth support because of its improved biosafety profile. However, this growth support may cause a significant change in filter capacity as discussed above (VI.3, e.g. Figure VI-6), and this will also affect process economy. Therefore, a direct economic comparison of the different filters (PB2+PC2, PDF4 and PDH4, see also Table VI.3) and growth supports is presented here, including the impact of the process improvements described in sections VI.5 and VI.6.

The following values and assumptions were used to calculate the different economic indicators of the depth filtration process in Table VI.10: (i) filter capacities as determined at the laboratory scale (Figure VI-6); (ii) costs for PB2+PC2 filters (including a ratio of 3:1 for

PB2:PC2) as well as auxiliary consumables (bags, tubing etc.) taken from the breakdown of costs from the Pharma-Planta consortium 2G12 GMP-production batches (Table VI.4); (iii) 60% of the auxiliary consumables costs for the PB2+PC2 filters (~2000 € in Table VI.4) were assumed to be the threshold for system setup (e.g. storage bags, tube connections between filter holders etc.) the remaining 40% were assumed to scale directly with the filter area (or number of L-Drums; 188 € per m² or 375 € per L-Drum); (iv) for the same system, a threshold of 20 h was assumed for setup with an additional hour for each L-Drum and a down time of 4 h for a system with 4 L-Drums plus 1 h for each four additional L-Drums; (v) for the PB2+PC2 system, a required area of 25 m² was assumed up to a total number of 20 L-Drums and an additional 5 m² for each new set of four L-Drums (because they require a new filter holder skid); (vi) for the PDF4 and PDH4 filters, the auxiliary consumables costs were assumed to scale linearly with the number of filter holders required (a maximum of 10 m² filler area per holder is possible); (vii) a setup time of 2 h and a disassembly time of 0.5 h was assumed per STAX holder; and (viii) a threshold area of 4 m² was assumed for the STAX system (for storage bags etc.) with an additional 2 m² required for each holder.

The calculated costs are shown in Figure VI-23 A and Table VI.10. The PB2+PC2 setup was more expensive than PDF4 or PDH4 in each scenario. This reflects three major factors: (i) the costs per unit filter area were higher for PB2+PC2 compared to the other filters (Table VI.4); (ii) the costs for auxiliary consumables (e.g. tubing, bags, etc.) were higher for the PB2+PC2 setup because more individual filter housings were connected also requiring a larger number of storage bags (Figure VI-1 A vs. Figure VI-20 C and Table VI.4); and (iii) when processing unmodified extract from plants grown on rockwool, the PB2+PC2 filter lifetime of was significantly lower than PDH4 and especially PDF4 (Figure VI-6).

The processing of unmodified extracts from rockwool-grown plants was inadvisable because this increased the consumables costs (filters and tubing etc.) by ~1.5–3-fold compared to soil-grown plants regardless of the filter type. As a consequence, the total downstream processing costs for a plant-based production process would also increase by 50–100% because the cost for filters, bags and tubing account for ~50% of DSP costs [247].

Therefore, the use of flocculants and additives is vital for the cost-effective production of biopharmaceuticals in plants if rockwool is used as a growth support. Even for soil-grown plants, these process modifications can prove useful as they improve the competitiveness of the platform compared to established counterparts such as CHO cells. Based on the process costs reported previously [247], the following assumptions were made to estimate the impact of flocculants and additives on the total production costs of a plant-derived 2G12 antibody in a

process aiming to deliver 5 g of pure product: (i) a 2G12 recovery factor (overall process yield) of 0.7; (ii) an average harvested biomass of 0.12 kg per plant; (iii) an average 2G12 concentration of 0.014 g L⁻¹ in leaf extracts when three volumes of buffer were used with fresh biomass (e.g. 3 L per 1 kg, corresponding to an expression rate of 0.0448 g kg⁻¹ fresh biomass); (iv) a liquid recovery of 80% from crude homogenate (these assumptions were based on the Pharma-Planta consortium 2G12 production batches); (v) the costs for plant growth as previously reported [247] were adjusted for the longer growth period of transgenic plants (57 days) compared to transient expression (47 days) and increased to 8.5 € per plant; and finally (vi) the costs for infiltration were neglected as a process based on transgenic plants was considered. The resulting total production costs are shown in Figure VI-23 B using PDH4 as an example.

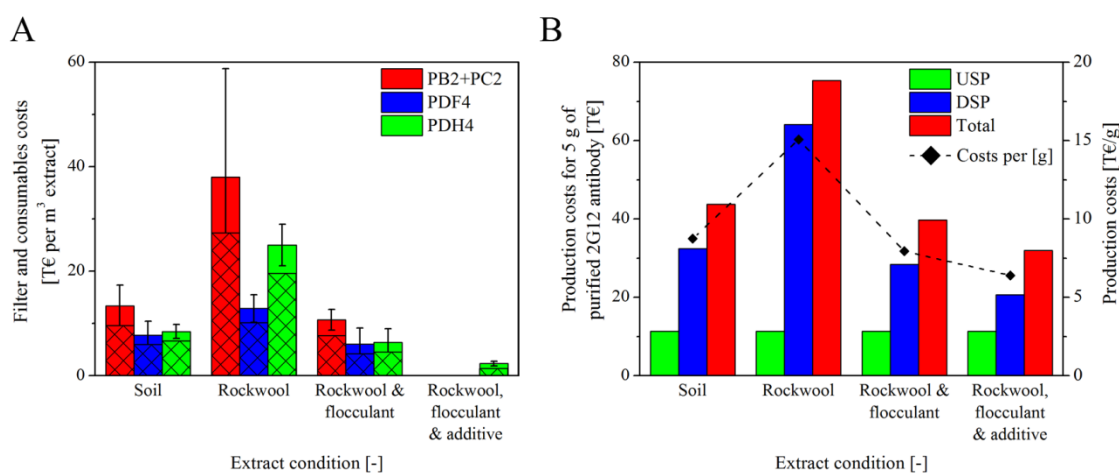


Figure VI-23: Effect of different depth filtration setups on the costs of process-scale production of 2G12 antibody (IV.7 and IV.8).

A. Costs for the filters PB2+PC2, PDF4 and PDH4 are compared for the ‘original’ setup using plants grown in soil (Table VI.4) as well as those grown on rockwool, also including flocculant (VI.5) and additives (VI.6). Filter lifetimes determined at the laboratory scale (Figure VI-6) were used for these calculations. The costs for auxiliary consumables and filters were taken from the breakdown of costs from the Pharma-Planta consortium 2G12 GMP-production batches, as discussed in the text and summarized in Table VI.10. Shaded column sections indicate the filter costs as a proportion of total consumables costs. Error bars indicate uncertainty of costs based on the filter lifetime variability. PB2+PC2 data are missing for the ‘rockwool, flocculant & additive’ condition because the small scale filters of this type do not have sufficient head space to allow cake formation by the additive. Data for PDF4 are missing because this filter type was not available at that time. B. Calculation of upstream production (USP), downstream processing (DSP) and total production costs of 2G12 under different production conditions using the PDH4 depth filter. Use of flocculants and additives reduced the production costs of rockwool-based production below the level of the soil-based process. A detailed discussion of the assumptions made for this calculation is provided in the text.

The combined use of flocculants and additives reduced the DSP costs by 35% and 66% compared to the soil-based and unmodified rockwool-based processes respectively, and the corresponding total production costs were reduced by 25% and 55%. On a per gram basis, the production costs for 2G12 were approximately €6400, which is not competitive compared to the CHO expression system. However, taking into account the low expression rate of the 2G12

antibody (0.0448 g kg^{-1}) it is likely that optimized expression cassettes/gene sequences can increase this level above 0.5 g kg^{-1} as observed for the monoclonal antibody M12 (Dr. Nicole Raven, personal communication). Assuming such an expression level, this would reduce the per gram production costs to €700-800. At this point the major cost factors in DSP will be the costs of labor and chromatography resins. The latter point can be addressed by reducing the complexity of the separation problem or using alternative separation techniques as will be discussed in section VI.9.6, whereas the first can also be approached by the use of filter systems such as STAX (PDF4 and PDH4) that require less handling time and thus help to reduce the labor costs. The small footprint of these systems will also help to reduce investment costs because they will fit within a smaller layout in the production facility.

Table VI.10: Calculated economic impact of different depth filtration setups on production-scale biopharmaceutical protein manufacturing in plants.

Extract condition [-]	Soil			Rockwool			Rockwool and flocculant			Rockwool, flocculant and additive
	PB2+ PC2	PDF4	PDH4	PB2+ PC2	PDF4	PDH4	PB2+ PC2	PDF4	PDH4	PDH4
Total filter area required to process 1 m ³ extract (~300 kg biomass)	20	17	19	57	29	56	12	12	13	4
Total number of L-Drums or STAX holders [-]	10	2	2	29	3	6	6	2	2	1
Price per m ² filter area [€]	480	350	350	480	350	350	480	350	350	350
Filter costs for 1 m ³ extract [€]	9600	5950	6650	27360	10150	19600	5760	4200	4550	1400
Approximate costs for tubing, bags and connectors [€]	3750	1800	1800	10688	2700	5400	3000	1800	1800	900
Total consumables costs for 1 m ³ extract [€]	13350	7750	8450	38048	12850	25000	8760	6000	6350	2300
Work load for system setup [person-hours]	30	4	4	49	6	12	26	4	4	2
Work load for system disassembly [person-hours]	6	1	1	11	2	3	5	1	1	1
Total labor time [person-hours]	36	5	5	60	8	15	31	5	5	3
Approximate space requirement for support equipment (holder, storage tanks etc.) [m ²]	25	8	8	40	10	16	25	8	8	6

VI.8 Pre-treatment of plants or bulk extract

The results presented in this section have been published as the following manuscript:

1. Buyel JF, Hubbuch J, Fischer R. 2016. Comparison of Tobacco Host Cell Protein Removal Methods by Blanching Intact Plants or by Heat Treatment of Extracts. *Journal of Visualized Experiments*, e54343(114):1-9.

Different options exist to remove HCPs from a feed stream or process intermediate, including precipitation at low pH [303]. However, not all target proteins are stable at low pH and thus alternative methods to remove HCPs have to be explored for plant-based expression systems in order to facilitate subsequent chromatographic product purification, especially if specific affinity ligands such as Protein A are not available. Heat precipitation can be such an alternative as previously demonstrated for several thermostable plant-derived recombinant proteins [304-306]. The heat precipitation of HCPs can be implemented in different ways, including blanching, i.e. the submersion of intact leaves into a hot liquid, stirred vessels or heat exchangers (Figure VI-24).

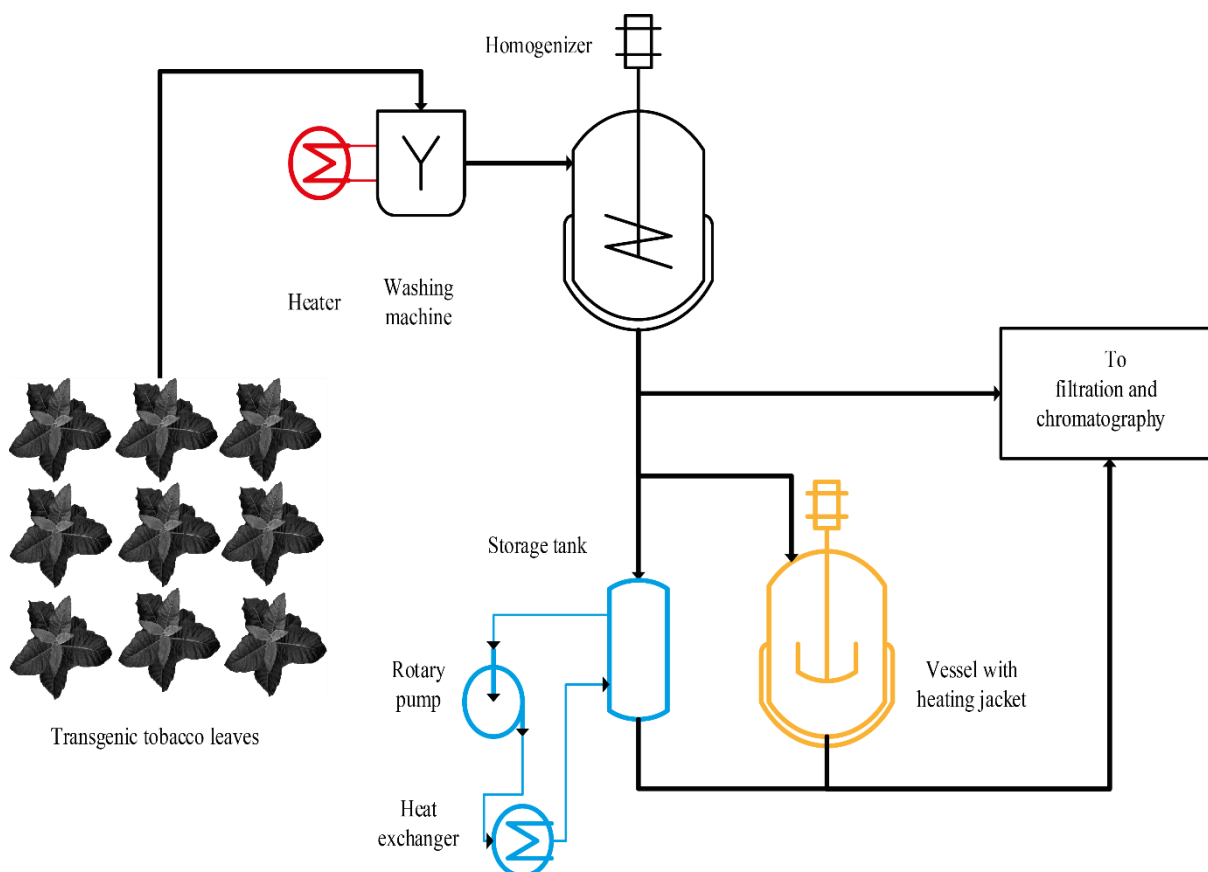


Figure VI-24: Process flow scheme illustrating the implementation of three different methods for tobacco HCP heat precipitation.

The plant material is washed and homogenized before clarification and purification. The equipment for the blanching step (red) can easily be added to the existing machinery. In contrast, using a stirred vessel (orange) and especially a heat exchanger (blue) requires one or several additional devices and tubing.

VI.8.1 Heat precipitation of tobacco host cell proteins by blanching

Blanching was successfully used to precipitate HCPs from tobacco leaves with 70°C, reducing the TSP by 96±1% (n = 3) while recovering up to 51% of the Vax8 target protein, thus increasing its purity from 0.1% to 1.2% before chromatographic separation [305]. It was also possible to recover 83±1% (n = 3) of the fluorescent protein DsRed, increasing its purity from 3.3% to 64.1%. The blanching procedure was easily integrated into a standard extraction and clarification scheme consisting of biomass washing, homogenization, bag filtration and depth filtration (Figure VI-24) [307].

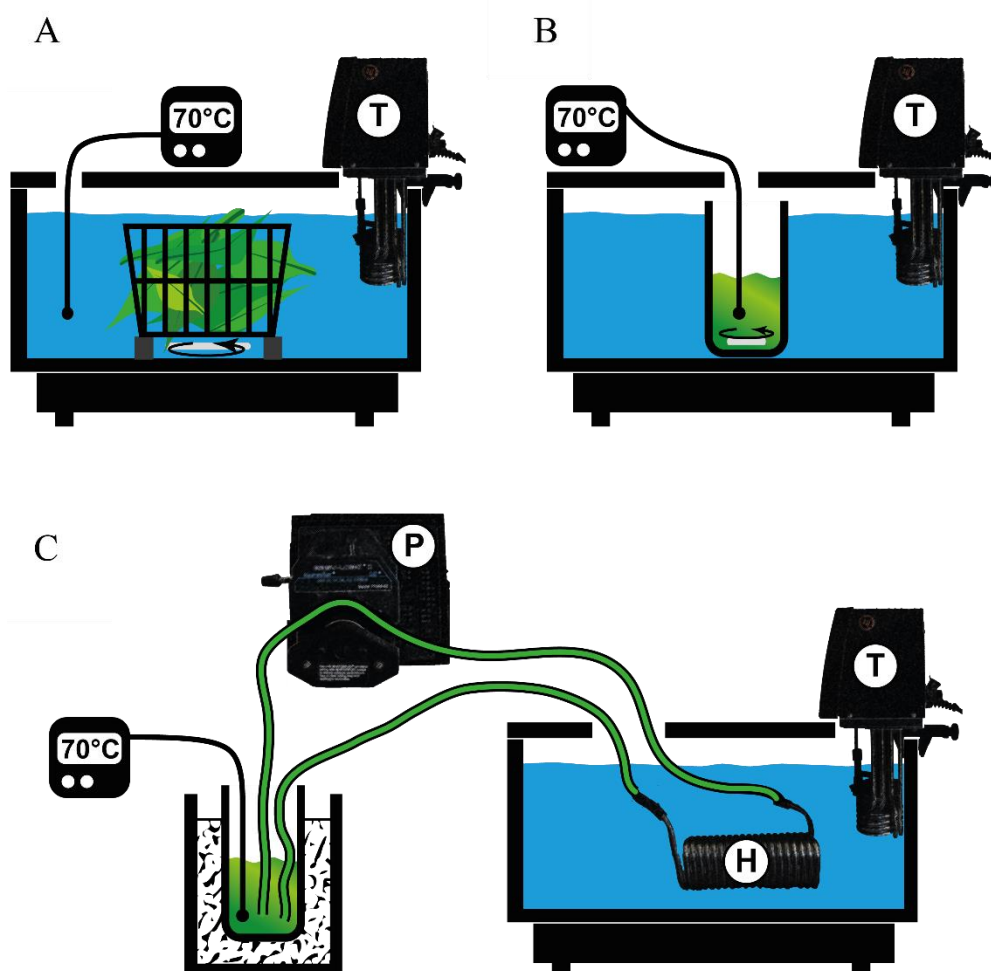


Figure VI-25: Schematic setup of three methods to precipitate tobacco HCPs in intact leaves or extracts thereof.

A. Blanching was carried out in a water bath heated with a thermostat (T) into which a basket containing intact leaves was submerged. The water bath was agitated to ensure a homogenous and constant temperature. B. A vessel containing a magnetic stir bar and leaf extract was submerged into a water bath. The temperature in the extract was monitored to ensure that the required temperature was achieved. C. A heat exchanger (H) was connected to a pump (P) and a thermally insulated storage vessel containing the plant extract. The heat exchanger was submerged in a water bath and the temperature of the heated extract was monitored.

Preparing the blanching equipment (Figure VI-25) added about 5 min to the set-up time for the clarification devices, which routinely takes 20 min. Another 7 min was required to

perform the blanching of intact leaves in addition to the typical extraction and clarification time of 45 min. However, only 2 min of the additional 7 min was actual “hands-on” time. Additionally, short incubation times of less than 1 min are possible, reducing the blanching time from 7 min to approximately 3 min. Therefore, blanching not only increased the initial purity of a product in crude plant extracts, but was also rapidly completed with no additional process equipment, thus offering the potential to replace at least an initial chromatography step. The blanching bath temperature remained constant, i.e. $<0.2^{\circ}\text{C}$ fluctuation, during all experiments even immediately after the addition of harvested leaves which were at ambient temperature. This ensured the process was repeatable, i.e. an average coefficient of variation of 17% ($n = 24$), in terms of TSP reduction, product yields and filter capacity in the subsequent clarification steps (Figure VI-26).

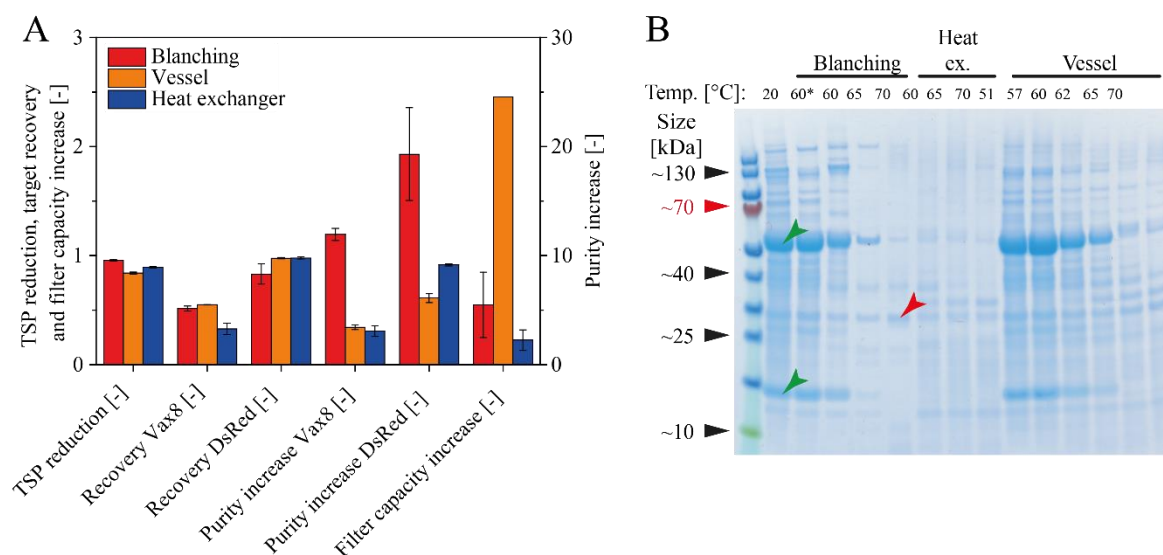


Figure VI-26: Comparison of three heat precipitation methods showing their effect on process performance and the purification of two target proteins.

A. Conditions supporting the removal of more than 90% of HCPs were identified for blanching, a stirred vessel and a heat exchanger setup, all of which increased the purity of the target proteins Vax8 and DsRed by a minimum of 2.5-fold and a maximum of 19-fold, with blanching performing best. In contrast, only the stirred vessel setup increased the capacity of the subsequent depth filtration step used for clarification of the heat treated plants or extracts. Error bars indicate the standard deviation ($n = 3$). B. The HCP content of samples after different heat treatment conditions was analyzed and compared using Coomassie-stained polyacrylamide gels (IV.4, IV.9.1). RuBisCO (green arrows) was removed along with other HCPs as the temperature during heat treatment increases, whereas DsRed (red arrow) remained in solution. * indicates a sample that was exposed to heat for 0.5 min, all other samples were treated for 3.0 min or more. Re-print with permission from [305].

However, the filter capacity declined at blanching temperatures $>63^{\circ}\text{C}$, potentially increasing the costs of filter consumables. This can be addressed by adding flocculants after protein extraction [308] and filter aids after bag filtration [302], which can restore or even increase filter capacities. A temperature of 60°C was sufficient to remove $80\pm 3\%$ ($n = 3$) of HCPs and increase Vax8 purity 2.6 ± 0.1 -fold ($n = 3$) without affecting filter capacity. Therefore, HCP removal by blanching is compatible with target proteins that have moderate heat stability,

i.e. a melting temperature below 70°C [309, 310]. However, increasing the temperature to 70°C or more may result in an HCP removal of over 95% (Figure VI-26). It was useful to conduct the corresponding set of experiments in a well-designed manner using a statistical approach [311] because this allowed the rapid identification of the most relevant process parameters, i.e. heating time and temperature. At the same time, the DoE method generated a predictive model to facilitate process optimization [305].

VI.8.2 Heat precipitation of HCPs in a stirred vessel

A stirred vessel for heat precipitation removed a maximum of $84\pm 1\%$ ($n = 3$) HCPs, achieving a purity of $0.33\pm 0.02\%$ ($n = 3$) for Vax8 and $20.2\pm 1.4\%$ ($n = 3$) for DsRed (Figure VI-26). The heat treatment was carried out in a vessel separate from the homogenizer to prevent delays reflecting device occupancy when processing multiple samples in series. The handling effort required for the laboratory-scale process was similar to that for blanching but an additional stainless steel vessel and a dedicated cooling step were required. Furthermore, heat transfer to the extract was slower than during blanching, with incubation times of at least 5 min, i.e. 10 times longer than for blanching. The delayed heating was caused by the vessel, which posed an additional barrier to heat transfer, and the ~300% greater mass that was heated in the vessel due to the presence of extraction buffer in addition to the plant biomass. Precipitating proteins also adhered to the walls of the vessel, gradually building up an additional heat transfer barrier and increasing the effort required for subsequent cleaning. Setting the water bath temperature 8°C above the temperature used for heat precipitation compensated for energy losses in the partially open system and achieved the desired extract temperature. In contrast to blanching (VI.8.1) and a heat exchanger setup (VI.8.3), HCP precipitation in a vessel increased the capacity of downstream depth filtration by 2.5-fold (Figure VI-26 A), reflecting the lower sheer forces in the vessel compared to pumping extract through the heat exchanger or homogenization after blanching, probably resulting in larger aggregates that were easier to remove in the bag filtration step.

VI.8.3 Heat precipitation of HCPs in a heat exchanger

Approximately $88.3\pm 0.7\%$ ($n = 12$) of the HCP content was consistently removed from the extract using a heat exchanger within the temperature range 60–70°C, achieving a purity of $0.31\pm 0.01\%$ ($n = 12$) for Vax8 and $27.6\pm 2.0\%$ ($n = 12$) for DsRed (Figure VI-26). The average coefficient of variation was 13% ($n = 24$) indicating that the repeatability of this procedure was even better than blanching. The desired extract temperature was achieved after ~3 min if the water bath temperature was set 4.5°C higher. As for the heated vessel, a dedicated cooling step

was required after heat treatment. The heat exchanger involved more handling effort than the other methods because the pumping apparatus and heat exchanger required intensive cleaning due to the precipitate adhering to the walls of the narrow bore stainless-steel tubing. The heat exchanger also achieved the lowest downstream depth filter capacity, clarifying only 13.5 ± 6.0 ($n = 3$) L m^{-2} before clogging.

VI.8.4 Evaluation of HCP precipitation by heat treatment methods

The three methods for heat precipitation described above can effectively remove tobacco HCPs prior to any chromatographic purification step [305, 312]. They complement other strategies that aim to increase initial product purity, e.g. guttation [313], rhizosecretion [314] or centrifugal extraction [315, 316], all of which are limited to secreted proteins. However, the heat-based methods can only be used in a meaningful way if the target protein to be purified can withstand the minimum precipitation temperature of $\sim 60^\circ\text{C}$ for more than 1 min. Therefore, the first step in any of the three methods is to design a target molecule with a sufficiently high melting temperature, which has been described for several malaria vaccine candidate proteins consisting of different domains from several *Plasmodium falciparum* antigens [304, 305, 317]. Once the thermal stability of the target protein has been demonstrated, one of the three methods can be selected based on the available equipment and media, anticipated final process scale and subsequent DSP operations [305].

Blanching was the fastest of the methods and additional equipment requirements were minimal, so it can easily be implemented into existing laboratory-scale purification protocols for plant-derived recombinant proteins. Thorough agitation of the blanching liquid is an important process parameter that affects the efficiency of HCP precipitation based on both empirical data and theoretical calculations [304, 318]. Failing to achieve good mixing can impair heat transfer and result in only partial HCP removal, which in turn can be detrimental to the product if host proteases remain active [304, 319]. Several other parameters can also affect HCP precipitation, e.g. the heating temperature and incubation time, and a DoE approach can therefore be useful to characterize the most relevant factors and provide predictive models to quantify their effects on responses such as product purity, recovery and the performance of subsequent DSP steps [311].

In the vessel setup, longer incubation times were required to achieve complete HCP precipitation and this may increase the likelihood of undesirable target protein denaturation reflecting the extended exposure to high temperatures. More thorough mixing in the vessel could improve the heat transfer and reduce the duration of heating. The long temperature ramp

in this setup can also be challenging if proteases in the extract [304, 319] become more active before final heat inactivation, causing product losses.

The increased depth filter capacity observed for the vessel setup can help to reduce consumables costs, allowing a larger number of samples to be handled in a project with a fixed budget or reducing the overall funding requirements for a given set of experiments. However, this benefit may be outweighed by the cost of the additional vessel, which is necessary in addition to the homogenizer to prevent the introduction of process hold steps if several extraction runs are required in a series of experiments, e.g. as part of a DoE approach. The positive effect on filter capacity may also diminish if a more intensive mixing regime is used to reduce heating times as suggested above.

A dedicated cooling step is necessary for both extract-based heat precipitation methods, requiring not only additional resources but also prolonging the overall processing time per sample, which can also conflict with fluent DoE procedures or experimental sequences in general. The heat exchanger setup is well characterized from an engineering perspective [320] and can easily be designed and scaled up for specific temperature differences, in contrast to the vessel, whose surface-to-volume ratio changes during scale up. However, once the heat exchanger size is defined, it can be difficult to adjust to alternative temperature differences because its length and thus the heat transfer area are fixed.

Changing other parameters, such as the residence time (or flow rate) and temperature of the heat exchanger medium, can restore flexibility to some degree, but only in small-scale experiments because these factors are typically operated in narrow windows in process scale operations due to restrictions imposed by the available equipment and media like water and steam. The demand for a combination of short incubation times of 3–5 min and a temperature difference of 40–60°C becomes increasingly difficult to solve at the device level as the process scale increases because the dimensions of the heat exchanger become larger. This is especially true for the cooling step because the temperature difference between the medium and desired extract temperature after cooling is often smaller ($\Delta T = 10\text{--}15^\circ\text{C}$) than the heating step ($\Delta T = 20\text{--}40^\circ\text{C}$) resulting in large equipment dimensions or longer cool-down times.

In the future, the methods can be adapted to biopharmaceutical proteins other than vaccine candidates which in this study were specifically designed for thermal stability. Many antibodies can withstand temperatures of $>70^\circ\text{C}$ [321, 322] which is already compatible with the current temperature regime. This natural thermal stability can be increased further by engineering the different antibody domains [323], thereby increasing the number of proteins that can be subjected to the methods presented here. The blanching method has already been

applied to transgenic tobacco plants expressing a monoclonal antibody (2G12) [312] which has not been subject to selection for thermal stability or protein engineering. Heat treatment at 65°C increased the purity of the antibody by a factor of two prior to chromatographic purification while the recovery was similar to that observed without blanching.

Additionally, characterizing the individual HCP melting temperatures of an expression system could facilitate the identification of a temperature with which the process could be conducted similar to pasteurization of milk: high temperature, short time [324]. The heat treatment (except for blanching) may also be applied to other biological starting materials to remove HCPs if the product can withstand the necessary temperatures. The latter may deviate from the ones discussed here if other expression platforms such as mammalian cell culture supernatants are being processed. In any case, the cost-benefit-ratio should be taken into account, i.e. does the benefit of reduced HCP levels outweigh the cost for implementing a heat treatment step that causes additional investment costs, increases the process time and may reduce the product yield [305]. A critical parameter in this context is the product activity. If it depends on the presence of linear epitopes as for some protein-based vaccines, then heat treatment is unlikely to have an effect [325]. In contrast, if protein structure is important, e.g. for conformational epitopes or enzymatic activity, the precise orientation of amino acid side chains in an enzyme's active site or the correct folding of antibody complementarity determining regions, a heat treatment may interfere with protein activity [326, 327]. Therefore, suitable analysis assays should be established to monitor product performance before and after heat treatment.

Part II: Chromatography

VI.9 Separation of HCPs by chromatography

The results presented in this section have been published as the following manuscript:

1. Buyel JF, Fischer R. 2014. Generic chromatography-based purification strategies accelerate the development of downstream processes for biopharmaceutical proteins produced in plants. *Biotechnology Journal*, 9(4):566-577.

Chromatography is widely used for the purification of target proteins because although the separation conditions are mild they have a high resolving power [147]. For the initial capture step, it is important to separate the target protein from the bulk HCP impurities and water, thus the resolution of the resin is of secondary importance at this stage [63]. Selection can be straightforward in the case of monoclonal antibodies secreted from mammalian cells because a specific capture step is available (using Protein A) and the target accounts for 75% or more of the TSP [328]. But in plants, the POI could represent only 0.7–7.0% of TSP [329] and resin capacity and price therefore become more important as selection criteria. This is especially true if a POI-specific capture resin is not available and the column capacity must also accommodate the binding of HCPs such as RuBisCO. It is therefore important to find conditions that separate the POI and HCPs and reduce the demand for column capacity, which is ideally achieved by directing HCPs into the flow-through (FT) fraction. Furthermore, the process will benefit from capture conditions that do not require sample conditioning (e.g. buffer exchange). The identification of such desirable capture conditions and resins was the aim of the experiments described in this chapter.

VI.9.1 AEX

VI.9.1.1 DEAE Sepharose FF

Once the number of particles is reduced below a certain threshold as described above (VI.2 to VI.6), different chromatography techniques are available for the isolation of a target protein from leaf extracts (III.6.3). The development of such an isolation procedure can be accelerated if the binding and elution behavior of the contaminating HCPs are known because this allows the selection of the most suitable separation conditions based on the target protein properties.

Initially, the reproducibility of protein elution profiles was confirmed by loading different samples onto diethylaminoethyl (DEAE) Sepharose FF resin, a weak anion exchanger. Three protein-containing elution peaks were detected in addition to the FT peak (Figure VI-27

A). The large and small RuBisCO chains were major components in the FT fraction and elution peak 2, the latter containing a large number of other proteins (Figure VI-29). The protein concentration in peak 3 was lower than in the other peaks and its UV 254 nm signal exceeded that at 280 nm, indicating the presence of DNA/RNA. Deviations for repeated separations were less than 0.30 mL (0.28 ± 0.09 mL; $n = 3$) for the retention volume of peak maxima, corresponding to a shift in salt concentration of ~ 15 mM NaCl (Figure VI-27 A). However, components eluting in peak 3 of the directly-processed samples (<36 h after filtration) shifted to the FT fraction if samples were stored for more than 40 h at 4°C before chromatography. The area of peak 3 therefore decreased by $\sim 40\%$ and that of the FT peak increased 2.5-fold without changing the total peak area significantly ($<2.0\%$). Most of these shifting components were DNA/RNA fragments as indicated by the high UV 254 nm signal in peak 3 (fresh samples) and the FT fraction (in aged samples) as well as agarose gel electrophoresis results (Figure VI-30 B). RNA and DNA were probably degraded during sample storage by RNAses/DNases released during leaf homogenization [330] reducing their binding to the positively-charged DEAE resin. Even though RNA/DNA was not the target molecule, these results imply that only freshly-prepared samples should be used to screen chromatography conditions to ensure authentic separation behavior of the extract.

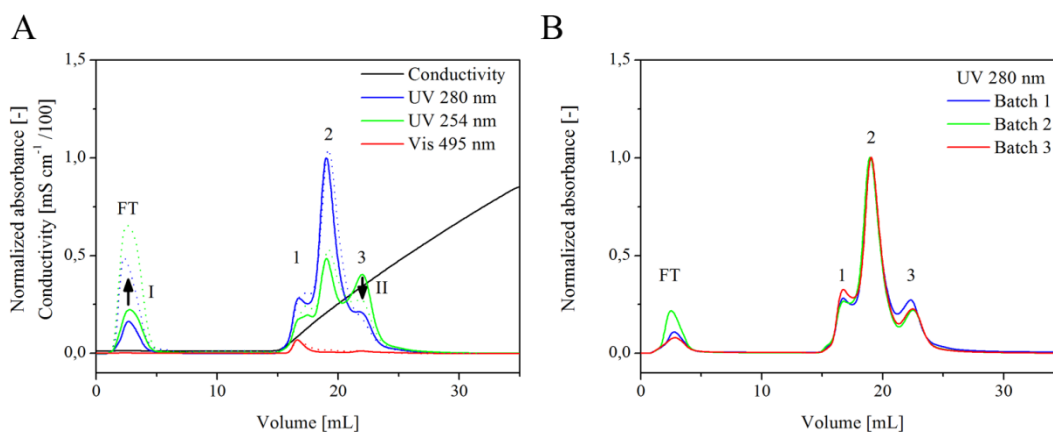


Figure VI-27: Effect of sample aging and different plant batches on tobacco HCP separation by DEAE (IV.11).

A. Samples stored for more than 40 h (dotted lines) exhibited a shift of the UV 254 nm (green line) intense peak 3 towards an increased UV signal in the FT fraction (arrows II and I respectively), whereas elution peaks 1 and 2 were hardly affected. B. Only minor differences were observed between elution peaks of samples from different batches that were subjected to chromatography within 36 h after filtration. Salt gradient as in A. Depth filter PDF4 was used for all experiments. The protein composition of the peaks is discussed in the text.

Samples that were processed immediately showed comparable elution profiles (Figure VI-27 B) even if they originated from different plant batches or underwent individual filtration or desalting procedures. The retention volumes of the FT fraction and elution peaks 1 to 3 varied

by <0.15 mL (0.09 ± 0.07 mL; $n = 4$) corresponding to ~ 8.0 mS cm^{-1} . Hence, the separation behavior of samples from different plant batches was similar, making it unnecessary to carry out separation runs under standard conditions for normalization.

The effect of different depth filters (VI.3) on separation was also investigated. Higher UV 280 nm signals were observed in the FT fraction and peak 1 after filtration with PDF4 (compared to all other filters) indicating that less-positively-charged extract components bound to the filter (Figure VI-28 A). This supported the inert behavior of this depth filter as described above (VI.3). Declining UV 254 nm and 280 nm signals in peak 3 were observed as the nominal retention ratings of the depth filters became finer, implying that more RNA/DNA fragments were binding to the finer depth filters.

The binding of HCPs to DEAE resin was measured over the pH range 7.0–8.0 in combination with low (1.5 mS cm^{-1}) or moderate (5.0 mS cm^{-1}) conductivities during protein loading (Figure VI-28 B). Lower buffer pH was associated with a higher UV 280 nm signal in the FT fraction and subsequent wash step, whereas the intensity of elution peaks 1 and 3 decreased. This is in good agreement with the binding mode of DEAE resin (Table III.2) because a lower pH reduces the negative charges on the protein surface and thus weakens binding to the positively-charged ligand. The binding strength was further weakened by the increase in conductivity, resulting in an additional shift in the UV 280 nm signal towards the FT fraction. Again, these results are in good agreement with chromatographic theory (III.6.4).

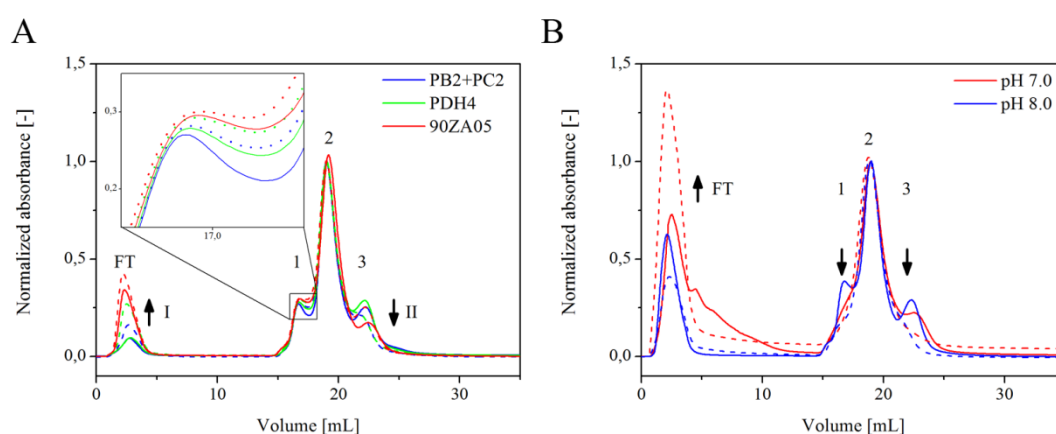


Figure VI-28: Effect of depth filters, pH and conductivity on tobacco HCP separation with DEAE (IV.11).
 A. Higher UV 280 nm signals were observed in FT (arrow I) and peak 1 (inset) after filtration with PDF4 (dashed lines) compared to all other filters (solid lines). Decreasing UV signals in peak 3 (arrow II) were observed as nominal retention ratings of the depth filters used in sample preparation got finer (Table VI.3). See text for interpretation. B. Decreasing the buffer pH from 8.0 (blue lines) over 7.5 to 7.0 (red lines) during the entire separation resulted in a decreased height of peak 1 and 3 (downward arrows), whereas more signal was observed in FT (upward arrow). Increasing the buffer conductivity from 1.5 to 5.0 mS cm^{-1} (dashed lines) enhanced this effect. Note that only UV 280 nm for runs at pH 7.0 and 8.0 are shown for clarity. Salt gradients as in Figure VI-27 A.

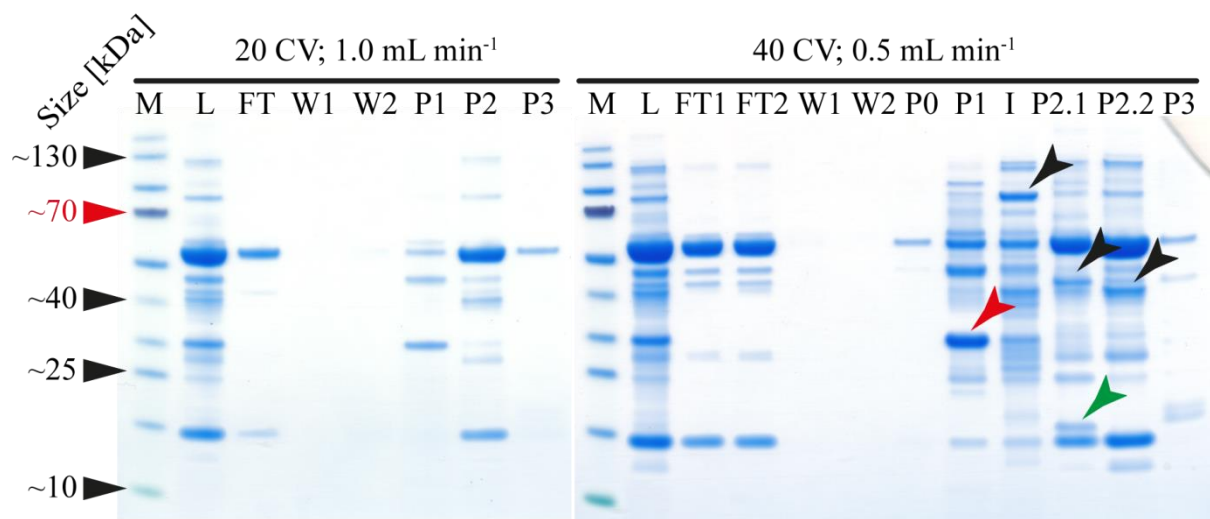
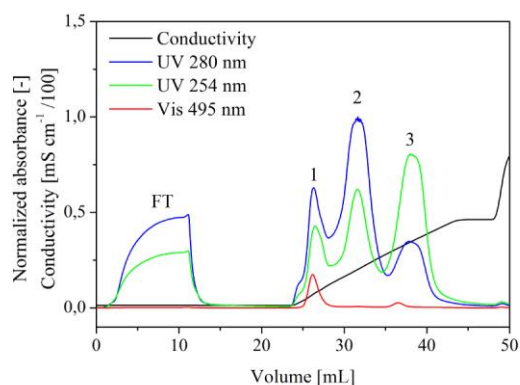


Figure VI-29: Impact of gradient steepness on resolution of HCPs by DEAE (IV.11).

Using a 20 CV gradient at pH 8.0 some RuBisCO was found in the FT while most other proteins eluted in peaks 1 and 2. Peak 3 hardly contained any protein (Figure VI-27 A). Increasing the gradient length two-fold and reducing the flow rate to 50% improved resolution of HCPs (Figure VI-30 A). DsRed (red arrow, P1) was separated from a 80 kDa major HCP (black arrow, I) and the RuBisCO containing second peak. Additionally, a 17 kDa and a 45 kDa HCP (green and black arrow, P2.1) were separated from a 42 kDa HCP (black arrow, P2.2) within elution peak 2 under these conditions. Note that 5.0 instead of 1.0 mL sample were loaded when a 40 CV gradient was used in order to improve detectability of eluted proteins. M – marker; L – load; FT – flow through; W – wash; P – peak; I – intermediate.

A



B

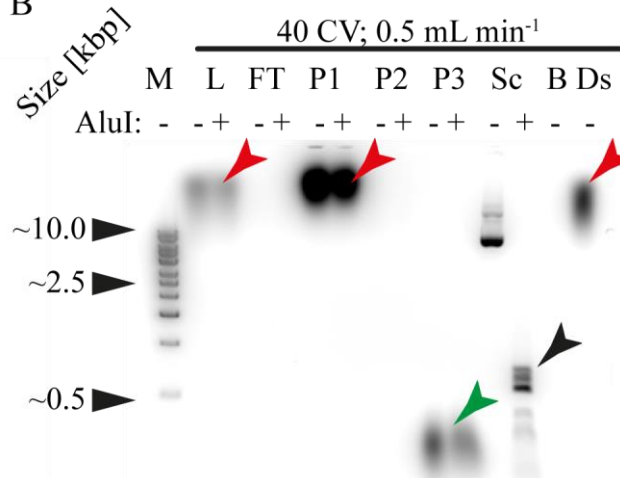


Figure VI-30: Effect of increased gradient length on HCP separation on DEAE (IV.11).

A. An increased gradient length improved separation of the three elution peaks. B. RNA/DNA was only detected in elution peak 3 (green arrow) independent of digestion with *AluI* in agarose gels. DsRed protein in load and peak 1 gave a false positive signal after ethidium bromide staining (red arrows). Note that 5.0 instead of 1.0 mL sample were loaded when a 40 CV gradient was used in order to improve detectability of eluted proteins. M – marker; L – load; FT – flow through; W – wash; P – peak; Sc – Sc plant expression vector (vector Sc as reported in [331], positive control, black arrow); B – buffer; Ds – DsRed protein. The protein composition of the peaks is described in the text and Figure VI-29.

The resolution achieved with a 20 CV gradient from 0–1000 mM NaCl was poor because only three protein peaks (including the FT fraction) and one RNA/DNA elution fraction were resolved. SDS-PAGE analysis of the peaks revealed their complex composition, which

further reduced the separation performance (Figure VI-29, 20 CV). To improve the resolution of tobacco HCPs, the gradient length was increased to 40 CV and the flow rate was reduced from 1.0 to 0.5 mL min⁻¹. These changes increased the separation factor α (III.6.3) of peaks 1 and 2 by 40% whereas the separation of peaks 2 and 3 was not affected. At the same time, the resolution of peaks 1 and 2 also increased from 0.85 to 1.20 (~40%) whereas that of peaks 2 and 3 decreased from 1.18 to 1.00 (~15%). An 80 CV gradient length did not improve the separation significantly (data not shown).

Despite the improvement achieved with a longer gradient, elution peaks still consisted of a number of proteins. Furthermore, the presence of the large and small RuBisCO subunits in all the protein-containing fractions indicated that the target protein cannot be separated from major HCP impurities using DEAE resins. The most effective use of this resin would be (i) operation in FT mode to bind ~80% of HCPs and DNA while a POI with positive/neutral charge at pH 7.0-8.0 can pass through the column, and (ii) operation in binding mode for strongly-retained target proteins eluting at conductivities >30 mS cm⁻¹ (~275 mM NaCl). The drawbacks of an initial FT step include the lack of a concentrating and volume-reducing effect as well as the need for a high-capacity column to bind the HCPs. In contrast, binding mode is likely to contaminate the target protein with tobacco HCPs in case the POI elutes at low conductivity or with RNA/DNA as this also elutes at >30 mS cm⁻¹ (Figure VI-27 and Figure VI-30).

VI.9.1.2 ANX Sepharose FF

Reproducible results were obtained when freshly-prepared samples were separated using the weak anion exchanger ANX Sepharose FF (VI.9.1.1). Again, fine depth filters reduced the amount of RNA/DNA binding to the resin. The shape of the UV 280 nm chromatograms was similar to that seen with DEAE resin (Figure VI-31 A) consisting of a FT fraction and three elution peaks. However, higher separation factors and resolutions were achieved with ANX resin because the elution of three peaks occurred in a larger volume under the same chromatographic conditions (1.0 mL sample, pH 8.0, 1.5 mS cm⁻¹, 20 CV gradient, flow rate 1.0 mL min⁻¹). A 20 CV gradient was therefore sufficient with ANX to achieve the same degree of separation as a 40 CV gradient with DEAE (Figure VI-30 A and Figure VI-32 A). The best separation between elution peaks was achieved at pH 7.5 with a resolution of 1.32 for peaks 1 and 2 and 1.11 for peaks 2 and 3. Reducing the pH to 7.0 increased the height of the FT peak and diminished peaks 1 and 3 (Figure VI-31 B). Increasing the conductivity during sample loading from 1.5 to 5.0 mS cm⁻¹ enhanced this effect. Therefore, ANX can be used for an initial

chromatography step in the same way as DEAE (VI.9.1.1) with the advantage of faster sample processing time.

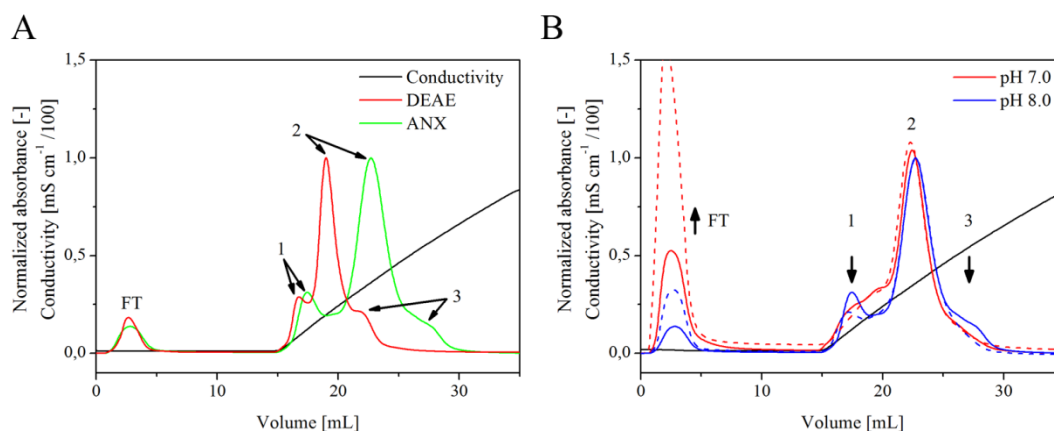


Figure VI-31: Tobacco HCP separation using ANX Sepharose FF (IV.11).

A. Three major protein containing elution peaks appeared at pH 8.0 after sample loading to ANX (green line) similar to DEAE (red line) but elution was spread over a larger volume for ANX. B. Decreasing pH and increasing conductivity during sample loading resulted in an increased intensity of the FT peak while intensity of peak 1 and 3 diminished relative to peak 2. Note that only runs at pH 7.0 and 8.0 are shown for clarity. All curves indicate UV 280 nm trace. The protein composition of the peaks is discussed in the text and Figure VI-32.

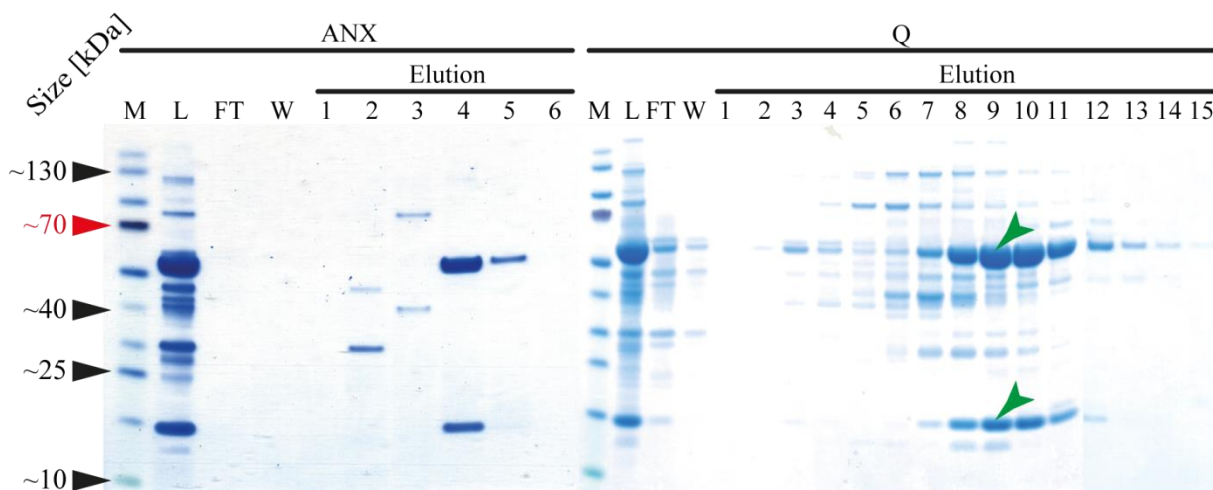


Figure VI-32: HCPs in fractions obtained from TSP separation by ANX or Q (IV.4.4, IV.11).

Using ANX with a 20 CV gradient to 1000 mM NaCl at 1.0 mL min^{-1} tobacco proteins were separated as well as on DEAE with a 40 CV gradient and a flow rate of only 0.5 mL min^{-1} (see also Figure VI-29). RuBisCO bound to the resin in the tested pH range from 7.0-8.0 (here results for pH 8.0 are shown) and made up the majority of the UV 280 nm signal during elution (peak 2 in Figure VI-31) corresponding to elution fraction 4 on the gel. The fraction 2 corresponds to ANX peak 1 in Figure VI-31 and fraction 6 to peak 3. When Q Sepharose FF was used to separate HCPs a similar sequence of the eluting proteins was observed even when desalting was performed with a 30 kDa membrane instead of a G-25 desalting resin. Again RuBisCO (green arrows) bound to the resin in a pH range from 5.5 to 8.5 (here results for pH 8.0 are shown). Interestingly, also the 15 kDa small subunit of RuBisCO was found in the load and elution fractions despite filtration through a 30 kDa membrane (see also Figure VI-33). M – marker; L – load; FT – flow through; W – wash.

VI.9.1.3 Q Sepharose FF and Q HyperCel

The ability of the strong anion exchangers Q Sepharose FF (Q) and Q HyperCel (HyQ) to bind tobacco HCP was compared over the pH range 5.5–8.5. Q exhibited a three-peak elution profile

like DEAE and ANX (Figure VI-33 A) but the resolution increased for the two pairs of subsequent peaks to 1.48 and 1.51, respectively (compare VI.9.1.1 and VI.9.1.2) even at a flow rate of 1.0 mL min^{-1} using a 20 CV gradient. Only ~15% of the UV 280 nm signal was found in the FT fractions.

Q was therefore selected for testing under conditions that resembled a production process more closely. First, the samples were desalted using a Hydrosart 30 kDa membrane (Sartorius, Germany) instead of a HiPrep 26/10 column packed with Sephadex G-25 resin (GE) which facilitated the conditioning of clarified extract for chromatography in a scaled-up process. Second, 5 mL instead of 1 mL of extract was loaded onto the 1 mL column to determine resin capacity and the effect of bound protein concentration on the resolution of separation. The 30 kDa membrane was found to permeable not only for NaCl but also for DNA/RNA from the plant extract (discarded with the permeate) as indicated by the absence of peak 3 and its formerly high UV 254 nm signal in the chromatograms (Figure VI-33 A vs. B). Interestingly, the small subunit of RuBisCO (~15 kDa) was still found in the filtration retentate and fractions collected during subsequent chromatography (Figure VI-32, green arrows) indicating its association with the large subunit under the pH conditions investigated here. This agreed with the stability of the RuBisCO complex as previously reported [332].

As expected, the large amount of protein loaded onto the column increased the UV 280 nm signal in the FT fraction (Figure VI-33 A vs. B and Figure VI-34) accounting for ~29% of the total signal. Interestingly, the UV signal in the FT fraction increased with higher pH, indicating less protein was binding to the column (Figure VI-34 A). This appears to contradict the binding principle of Q because more negative charges should be present on the protein surfaces at higher pH values, facilitating binding to the positively-charged ligand. However, the association of large and small chain of RuBisCO mentioned above and their oligomeric structure might have been affected by the change in pH (e.g. 4:4 heterooctamer at pH 5.5 but 2:2 heterotetramer at pH 7.5) thus reducing the number of binding sites per molecule and causing weaker binding [149]. This assumption was supported by the finding that RuBisCO appeared in the FT fraction at $\text{pH} \geq 7.5$ whereas it was not found in the FT fraction at $\text{pH} \leq 6.5$ (data not shown).

Under high protein loading conditions, the two protein-containing peaks 1 and 2 were still separated by Q Sepharose FF whereas only one elution peak was observed using DEAE resin (Figure VI-33 B). The resolution of these peaks increased from 1.03 at pH 5.5 to 1.63 at pH 8.5 and even up to 2.30 if phosphate was replaced with glycine as the buffering agent (Figure

VI-34 A). This exchange also resulted in the appearance of a third elution peak (1.2) between former peaks 1 and 2.

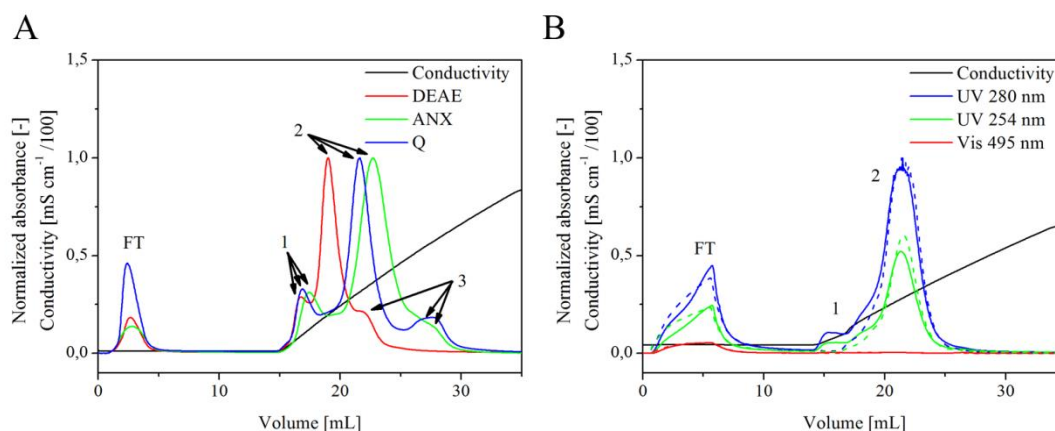


Figure VI-33: HCP separation by Q Sepharose FF compared to DEAE and ANX (IV.11).

A. The three resins exhibited a three-peak elution profile (labels 1-3) but peaks were best resolved using Q. B. Q (solid lines) and DEAE (dashed lines) showed significant breakthrough in the FT fraction after ~2.5 mL (~12 mg TSP loaded) when 5 instead of 1 mL sample was loaded to the column. Under these conditions only Q was capable of resolving peaks 1 and 2. Desalting the sample with a 30 kDa membrane instead of a G-25 resin removed most RNA/DNA molecules resulting in the disappearance of elution peak 3 (compare A and B) and its strong UV 254 nm signal. All curves in A represent a UV 280 nm signal. The protein composition of the peaks is discussed in the text and Figure VI-32.

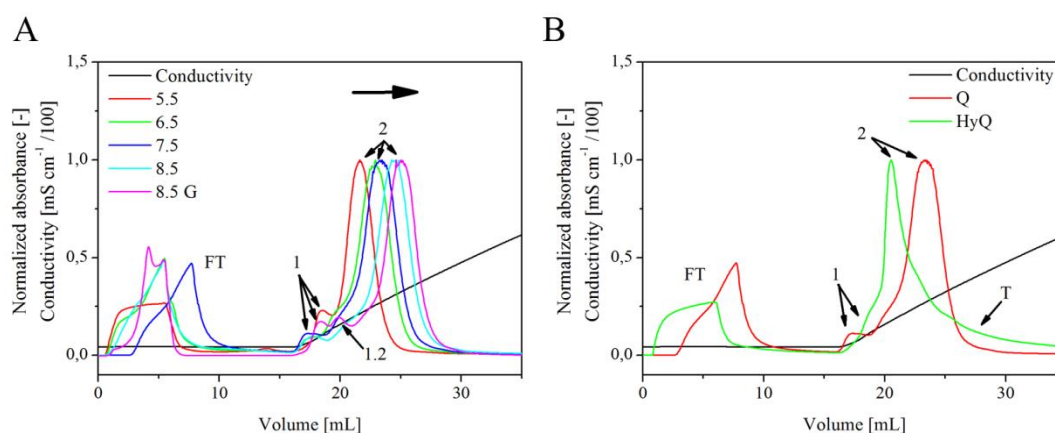


Figure VI-34: Influence of pH and resin structure on protein binding to quaternary ammonium ligands (IV.11 and III.6.3).

A. An increasing pH shifted the elution of peak 2 towards higher salt concentrations, ~1.0 mL (~50 mM NaCl) per 0.5 pH units while peak 1 was less affected. As a consequence resolution increased from 1.03 at pH 5.5 to 1.63 at pH 8.5. Changing the buffering agent from phosphate to glycerin further improved the resolution as an additional peak 1.2 appeared. B. The cellulose based resin HyQ exhibited less breakthrough at pH 7.5 than Q but did not resolve the two elution peaks well and showed extensive tailing (T). Note that the chromatogram of Q at pH 7.5 (blue line in A and red line in B) had a wash step that was 2 mL shorter than that of all other runs and thus the FT peak is shifted by this volume, however, the salt gradient starts at the same volume as for all other chromatograms.

Separation of the peaks was less efficient when a cellulose resin matrix (HyQ) was used instead of an agarose-based matrix (Q) (Figure VI-34 B) because the resolution dropped to 0.75 at pH 7.5. However, the percentage of proteins in the FT fraction remained at ~27%. The HyQ

resin showed extensive tailing of the major elution peak containing RuBisCO, preventing the effective separation of this impurity from potential target proteins. HyQ may be suitable in binding mode only if RuBisCO has been removed in a previous step.

The tailing may reflect the longer spacer arms of HyQ compared to Q Sepharose FF resulting in solid diffusion in the pores rather than pore diffusion [333-335]. The binding capacity of HyQ can therefore be higher (indicated by the shallower FT peak in Figure VI-34 B) but desorption may be inhibited.

VI.9.2 CEX – SP Sepharose

In CEX the pH of the feed stream is usually set to 4.0–6.5 to ensure the presence of positive surface charges, allowing proteins to bind the negatively-charged ligands such as sulfopropyl [336]. At pH values <6.0, tobacco HCPs including the RuBisCO large subunit started to precipitate as described above (Figure VI-16). This can be advantageous if the target protein is stable and soluble in this pH range, because its abundance relative to the HCPs increases.

Setting the pH at 4.5 resulted in the quantitative precipitation of the RuBisCO large subunit and 65% of all HCPs (Figure VI-35). Although the majority of the remaining proteins were found in the FT fractions (Figure VI-35 and Figure VI-36 A), the RuBisCO small subunit and a 40 kDa protein eluted early and late in a salt elution gradient, respectively. These conditions are favorable for a target protein binding to the SP resin because even a coarse fractionation of the elution or stepwise gradient will leave only one major impurity (RuBisCO small subunit or the 40 kDa protein) together with the POI. Resolution between the six detected elution peaks was on average only 0.53 ± 0.16 , which can be improved by increasing the bed height or reducing the flow rate if SP is considered for an intermediate purification step (III.6.3). However, for an initial capture step, the achieved separation was sufficient.

Approximately 25% of all HCPs precipitated at pH 5.5 but both RuBisCO subunits remained soluble and bound to the SP ligand as did 65% of the loaded HCPs (Figure VI-35 and Figure VI-36 A). The RuBisCO large subunit was found in the FT only during the second half of sample loading, indicating that the dynamic binding capacity of the resin had been reached. Bound proteins eluted in the first half of the subsequent salt gradient as a single peak showing extensive tailing.

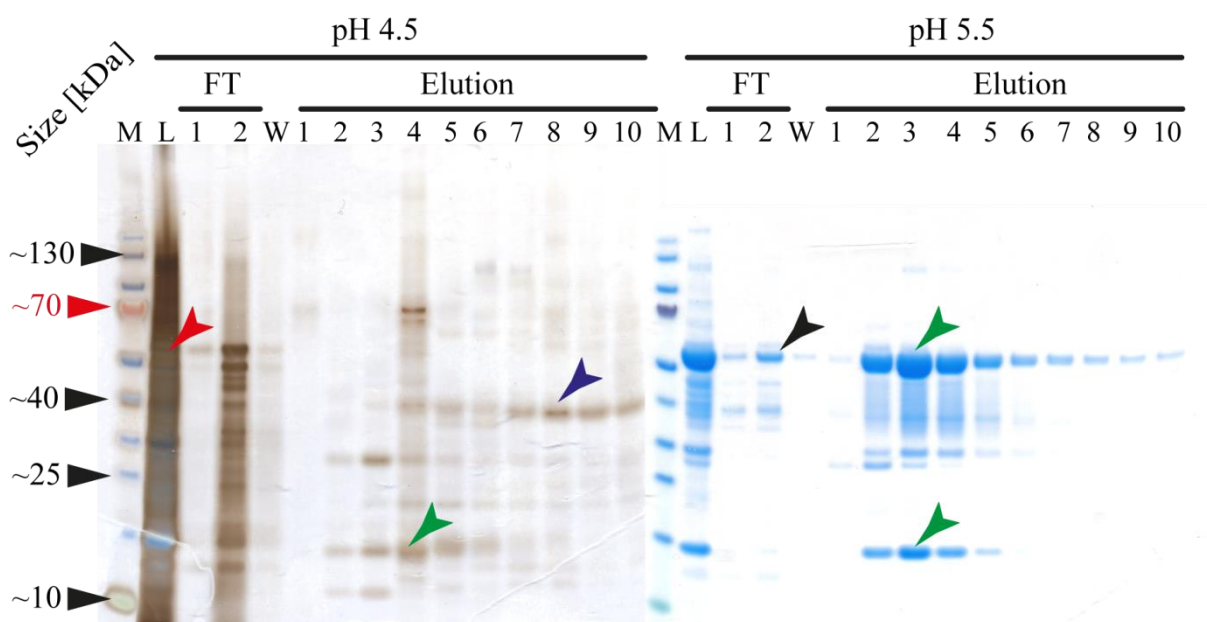


Figure VI-35: SDS-PAGE analysis of fractions collected during HCP separation by SP Sepharose FF (IV.4.4, IV.11).

HCP concentrations, e.g. that of the large subunit of RuBisCO (red arrow), had decreased by about 65% through pH adjustment to 4.5 (L). Of the remaining proteins 80% were found in the FT while the rest eluted within the first 15 CV of a 20 CV gradient to 500 mM NaCl. The most dominant species to bind to SP were a 15 kDa protein (likely RuBisCO small subunit, green arrow) and a 40 kDa protein (blue arrow). At pH 5.5 25% of TSP was removed by precipitation following pH adjustment (L). In the course of 5 mL sample loading an increasing amount of RuBisCO (black arrow) was found in the FT. Most proteins eluted in a single sharp peak (fractions 2-4) with RuBisCO as the major component (green arrows) exhibiting extensive tailing (Figure VI-36 A). For both gels elution fractions 1-10 cover the volume from 17-27 mL shown in Figure VI-36 A. M – marker; L – load; FT – flow through; W – wash.

Therefore, SP seemed inadequate for binding-mode at this pH because separation from RuBisCO can only be achieved for target proteins tightly binding to SP and eluting only at salt concentrations >500 mM NaCl. FT mode is a possible alternative but as shown by the early breakthrough of the RuBisCO large subunit, resin capacity could be a limiting factor.

Interestingly, and as described above, the RuBisCO small subunit was found in the loading and elution fractions despite sample preparation with a 30 kDa MWCO membrane, which should be permeable to the RuBisCO small subunit (~15 kDa) if present as a monomer. Therefore the RuBisCO large and small subunits also formed oligomers to a certain extent at pH 4.5 and 5.5.

Minimal HCP binding to SP was observed at pH 6.5 and ~95% of TSP was found in the FT fraction, including both RuBisCO subunits (Figure VI-36). Therefore, operating SP in binding mode can be advantageous because HCPs will not limit column capacity and bound impurities will only account for a small share of the total eluting proteins. Six elution peaks were observed (as described at pH 4.5) with a comparably low average resolution of 0.52 ± 0.40 but acceptable separation factors of 1.39 ± 0.15 (colored arrows in Figure VI-36 B) indicating high selectivity but low efficacy (Figure III-5). This low efficacy can be compensated by (i) a

longer column (Equation 23) or (ii) by a shallower gradient (Equation 25) because all proteins were eluted within the first 25% of the salt gradient (up to ~250 mM NaCl). Both measures can increase the plate count N and thus help to separate bound HCPs from the target protein during elution. The separation achieved with the current setup can be sufficient if (i) it is used as an initial capture step because only traces of HCP (~5%) bind to SP or (ii) the target protein elutes at NaCl concentrations >250 mM.

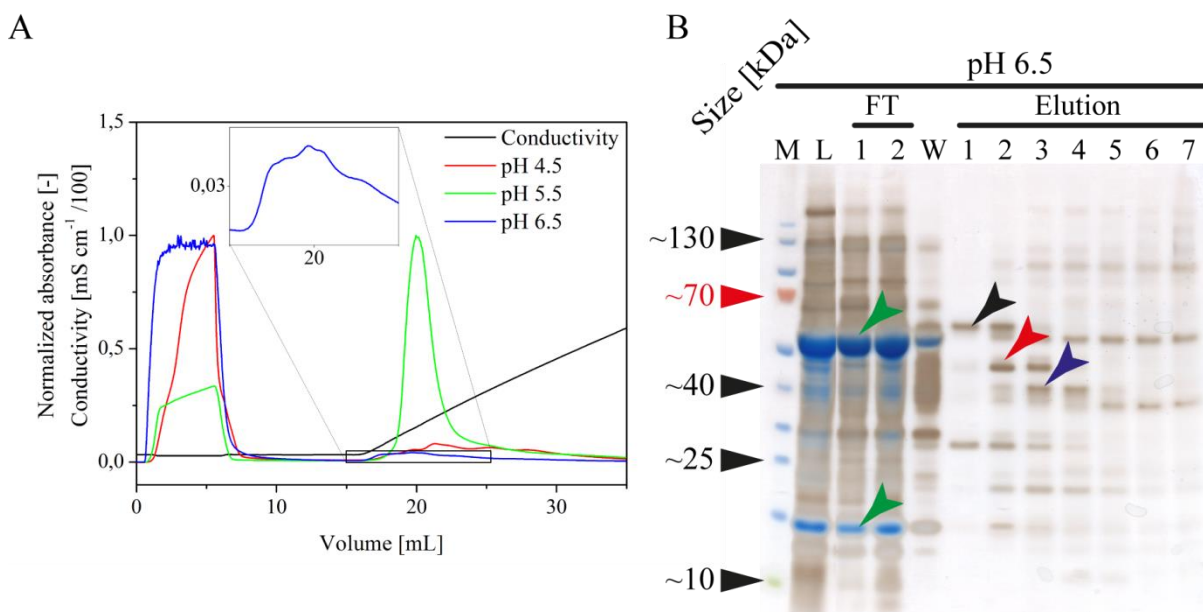


Figure VI-36: Chromatographic separation of HCPs by SP Sepharose FF (IV.4.4, IV.11).

A. About 65% of the HCPs bound to SP at pH 5.5 (green curve) whereas only 5% did at pH 6.5 (blue curve). Most of the binding protein species had been removed due to precipitation at pH 4.5 (red curve) (compare Figure VI-35 A). B. At pH 6.5 RuBisCO small and large subunits were found in the FT fractions (green arrows). The few binding protein species eluted within the first 7 mL of a 20 CV gradient to 500 mM NaCl (here: 1 CV = 1 mL). Six major protein species were identified (three of which are labeled by a black, red or blue arrow) exhibiting an average separation factor of 1.38 and an average resolution of 0.52. Note that for clarity three lanes of the gel between L and FT 1 have been removed. Elution fractions 1-7 cover the volume from 17-24 mL shown in Figure VI-36 A. M – marker; L – load; FT – flow through; W – wash.

VI.9.3 HIC

High concentrations (1000–1500 mM) of salts, e.g. NaCl or sulfate, are required for effective protein binding to ligands in hydrophobic interaction chromatography (HIC) [149]. These concentrations and salts can cause protein precipitation and/or denaturation and increase disposal costs [149]. However, it is easier to adjust the intermediate salt concentration (~500 mM) after routine tobacco protein extraction to these conditions in a process, e.g. by mixing with high salt solutions, rather than desalting the bulk extract, e.g. by membrane filtration, to be compatible with IEC (VI.9.1 and VI.9.2).

Therefore, protein binding was investigated to one aliphatic (octyl) and one aromatic (phenyl) ligand, and the latter was tested in two different ligand densities (high and low substitution, Table III.2). Only Phenyl Sepharose FF high substitution was found to bind ~56%

tobacco HCPs in the presence of 1000 mM NaCl at pH 6.5 and 7.5 (Figure VI-37). Column breakthrough was observed for this resin during the second half of 5 mL sample loading (~10 mg TSP loaded), indicated by the presence of RuBisCO in the FT fraction. A broad elution ‘peak’ was found in the subsequent gradient to a salt-free buffer. This peak contained RuBisCO as the dominant protein component in all its fractions (Figure VI-37 B) and additional proteins appeared with decreasing salt concentrations. Therefore, no effective separation of bound proteins was achieved with the elution gradient and operation should be restricted to FT mode. However, even in FT mode, the limited resin capacity in terms of RuBisCO binding can limit its usefulness in a purification process for tobacco-derived proteins. Higher salt concentrations may improve the performance of Phenyl Sepharose FF but can also promote protein precipitation. Therefore, the tested HIC resins appeared inadequate for an initial capture step.

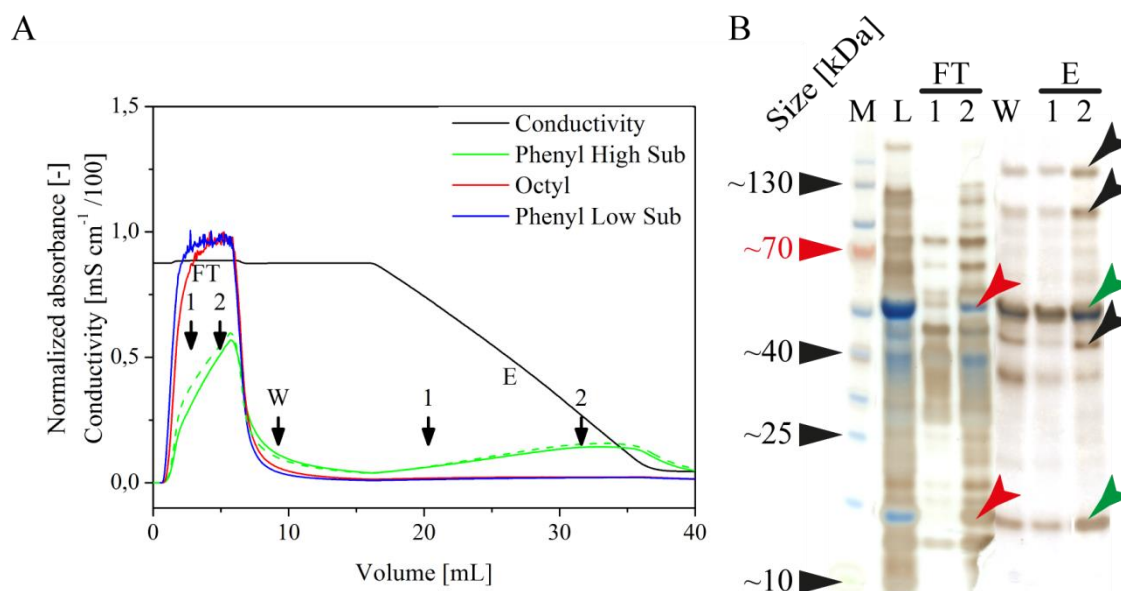


Figure VI-37: Tobacco HCP binding to hydrophobic chromatography ligands (IV.4.4, IV.11).

A. At a conductivity of 80 mS cm^{-1} (1000 mM NaCl) HCPs only bound to Phenyl Sepharose High Substitution (green lines) at both, pH 6.5 (dashed) and 7.5 (solid) with no significant difference between the pHs. Protein breakthrough occurred in the course of 5 mL sample loading (~20 mg protein). One broad elution peak was observed over the whole range of a 20 CV gradient to 4.5 mS cm^{-1} (no NaCl). B. RuBisCO (red arrows) was found in the later FT fraction from Phenyl Sepharose High Substitution when protein breakthrough occurred. Both RuBisCO subunits (green arrows) accounted for the majority of the UV 280 nm signal of the broad elution peak observed for this ligand as exemplarily shown for two fractions. With decreasing conductivity the elution peak contained more of a 50 kDa, a 120 kDa and a 150 kDa protein (black arrows). M – marker; L – load; FT – flow through; W – wash; E – elution.

VI.9.4 MMC and salt-tolerant resins

Mixed-mode and salt-tolerant resins (Table III.2) can be used as an alternative to HIC resins in terms of direct protein capture from tobacco extracts because they allow binding at intermediate conductivities of $15\text{--}35 \text{ mS cm}^{-1}$ (100–350 mM NaCl) and thus require no or only minimal sample conditioning, such as dilution with water [337].

VI.9.4.1 STAR AX

The salt tolerance of the primary-amine-based anion exchange resin STAR AX was tested with tobacco extracts adjusted to a conductivity of either 9 or 14 mS cm⁻¹ (IV.11) by dilution with water. Furthermore, phosphate concentrations of 15–25 mM were investigated because in this concentration range the buffer ion can affect the binding strength of proteins to the resin according to the manufacturer's information [338].

Regardless of the conductivity or phosphate concentration, the UV 254 nm signal exceeded the UV 280 nm signal in the FT and wash fractions, indicating the presence of RNA/DNA as well as pigments such as anthocyanins (data not shown). The presence of the latter molecules was further implied by the yellow-greenish color of the FT fractions, the typical color of anthocyanins at alkaline pH [339]. These compounds had been removed before the IEC separations described above in the course of desalting by column or membrane filtration due to their small size (<0.5 kDa) compared to proteins [339].

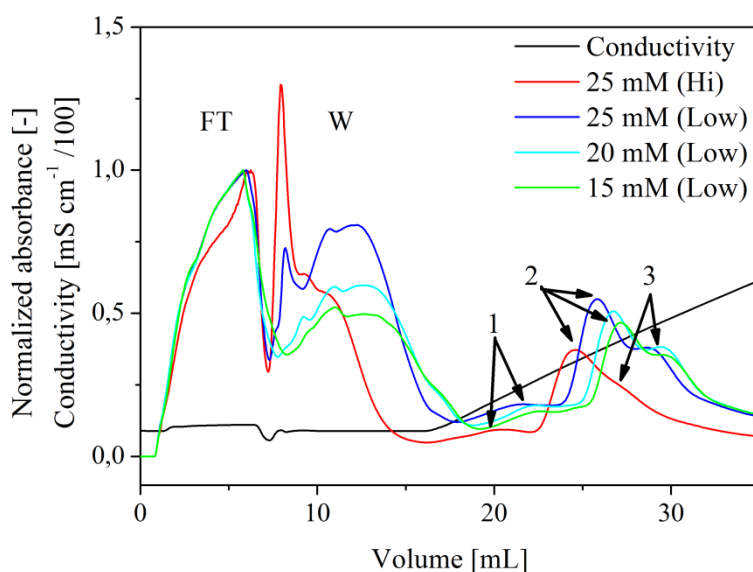


Figure VI-38: Separation of tobacco HCPs by HyperCel STAR AX at pH 7.5 (IV.11).

Depth filtered extract was diluted 1:1 with de-ionized water and a total of 10 mg TSP was loaded to STAR AX in the presence of 15–25 mM phosphate at a conductivity of 14 mS cm⁻¹ (Hi) or 9 mS cm⁻¹ (Low). The UV 254 nm signal exceeded the UV 280 nm signal in FT and W (Wash) fractions. Decreasing conductivity and phosphate concentration increased the retention volumes of the three observed elution peaks. See text for discussion.

Both RuBisCO subunits and an abundant 40 kDa protein were bound under all the conditions tested and no breakthrough of these proteins was observed during sample loading (5 mL, corresponding to ~10 mg TSP per 1 mL resin) (Figure VI-39). At a conductivity of 14 mS cm⁻¹ ~27% of the UV 280 nm signal appeared in the elution fractions and a reduction to 9 mS cm⁻¹ increased this share to 32% corresponding to >65% of the loaded protein (Figure VI-38). Accordingly, 80 kDa and 120 kDa proteins shifted from the FT fractions into the elution

fractions (Figure VI-39) and average resolution between the three observed peaks increased from 0.95 to 1.13. Lower phosphate concentrations shifted protein elution to elevated salt concentrations, indicating stronger binding to the resin, but resolution was not affected.

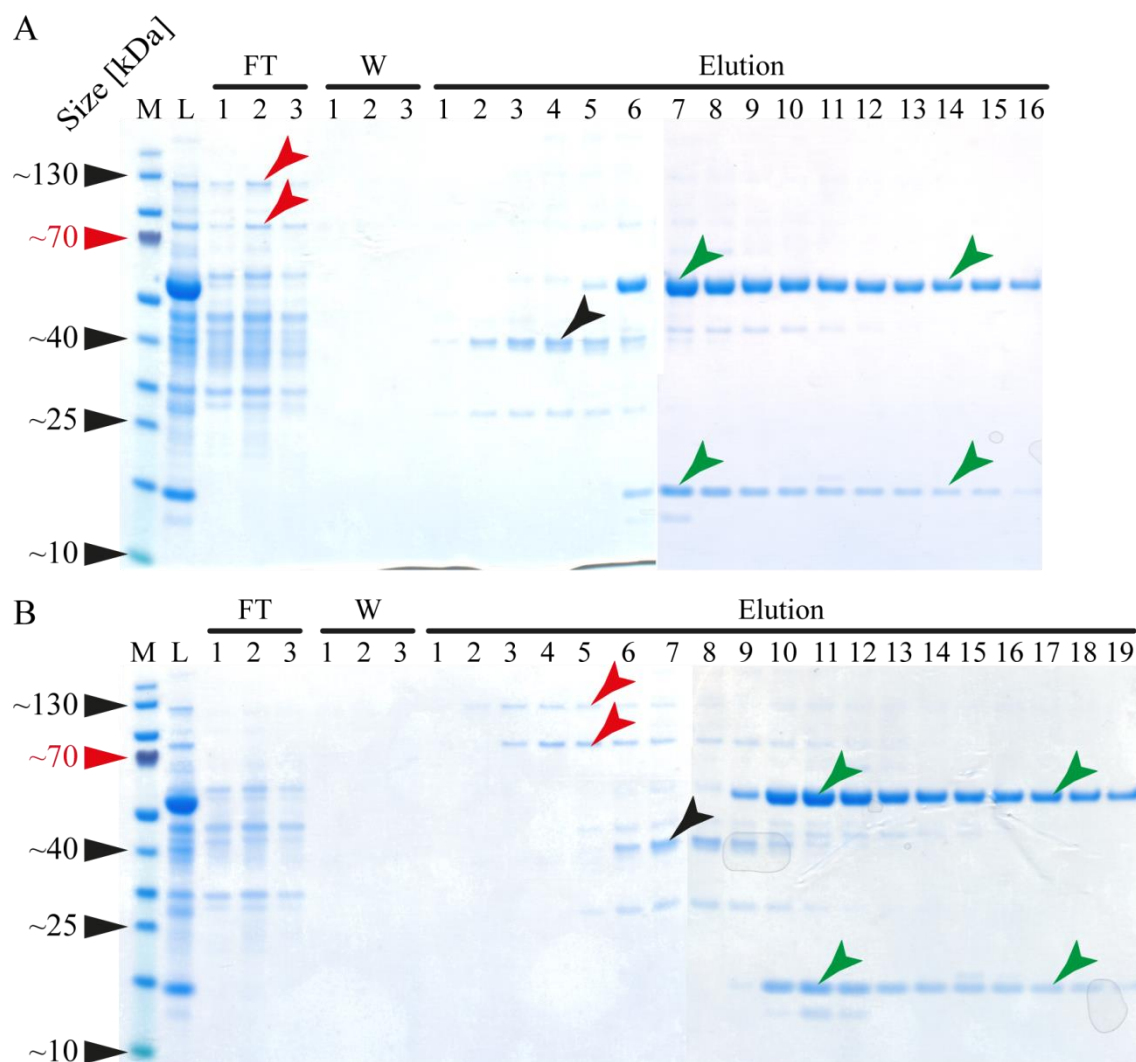


Figure VI-39: SDS-PAGE analysis of fractions collected during HCP separation from diluted tobacco extract by HyperCel STAR AX (IV.4.4, IV.11).

A. Fractions obtained during separation with 25 mM phosphate in liquid phase and 14 mS cm^{-1} conductivity during sample loading (Figure VI-38). Both RuBisCO subunits eluted in peak 2 and 3 exhibiting extensive tailing (green arrows) but were not found in the FT. The wash fractions did not contain detectable amounts of protein. A 30 kDa and a 40 kDa protein band were found in fractions of elution peak 1 (black arrow). Elution fractions cover the gradient volume from 15-31 mL. B. Fractions obtained during separation with 25 mM phosphate in liquid phase and 9 mS cm^{-1} conductivity during sample loading. The reduced conductivity resulted in a shift of a ~90 kDa and a ~125 kDa band from the FT fractions to early gradient elution fractions (red arrows, compare with A). Elution fractions cover the gradient volume from 15-34 mL. M – marker; L – load; FT – flow through; W – wash. Note that gels in A and B were each merged from two individual gels.

RuBisCO was present in several subsequent elution fractions as observed for Q HyperCel (Figure VI-34) and therefore this tailing effect was attributed to the shared cellulose-based resin matrix rather than the ligand. Accordingly, STAR AX can be operated effectively in FT mode, to separate a target protein from most HCPs, or in binding mode, if the target elutes before RuBisCO. However, in both operation modes, the resin capacity needs

further evaluation to ensure cost-effective separation of the POI and HCPs in the presence of the abundant RuBisCO proteins. The STAR AX resin can simplify the purification of biopharmaceuticals produced in tobacco because it allowed separation from HCP directly after filtration by simple 1:1 dilution with water. By adjusting the extraction buffer composition to a lower salt concentration, it can be even possible to skip this conditioning step.

VI.9.4.2 CaptoAdhere

Proteins were allowed to bind to CaptoAdhere (Table III.2) at conductivities of 30 or 75 mS cm⁻¹ and pH values of 6.0, 7.0 or 8.0, to determine whether (i) electrostatic or hydrophobic interactions were the driving forces for tobacco HCP binding to the ligand, and (ii) whether filtered tobacco extract (commonly 30–50 mS cm⁻¹) can be directly loaded onto a column packed with this resin. Elution was induced by a combined gradient to pH 3.0 at a conductivity of ~3.0 mS cm⁻¹ as recommended by the manufacturer [340].

At 75 mS cm⁻¹ >95% of the UV 280 nm signal (Figure VI-40 A) and all detectable proteins (data not shown) were found in the FT and wash fractions regardless of the pH. In contrast, at 30 mS cm⁻¹ the share of UV 280 nm found in the FT and wash fractions increased with the pH but only from 30% at pH 5.0 to 46% at pH 8.0. This indicated that proteins were bound more strongly to CaptoAdhere at low conductivities, thus electrostatic interactions were the major driving force for protein adsorption in the investigated conductivity range.

The increasing share of the total UV 280 nm signal in the FT and wash fractions with increasing pH was not due to less protein binding to the resin, which would have been inconsistent with the positive charge of the ligand, but instead marked the occurrence of irreversible protein binding to CaptoAdhere at elevated pH levels. This was consistent with the observation that not all proteins loaded onto the column were retrieved in the FT, wash or elution fractions, e.g. RuBisCO (Figure VI-40 B). Also, manufacturer's data [340] indicate that recombinant antibody yields fall significantly at pH >7.5 when CaptoAdhere is operated in FT mode, indicating progressive and irreversible protein binding with increasing pH. The irreversible binding at low conductivity also explains the reduction of the total peak area to ~70% of that observed at high conductivity.

In summary, CaptoAdhere can be used like STAR AX in FT mode to bind many HCPs including RuBisCO, but is insensitive to phosphate and compatible with conductivities of at least 30 mS cm⁻¹ avoiding the need for a conditioning step. However, harsh and time-consuming cleaning procedures may be necessary because some protein binding was not reversible using a low-salt and low-pH elution gradient. This could be overcome by a gradient

of increasing salt concentration or a combination of pH and chaotropic salts [341] but protein denaturation can still occur upon binding to the ligand.

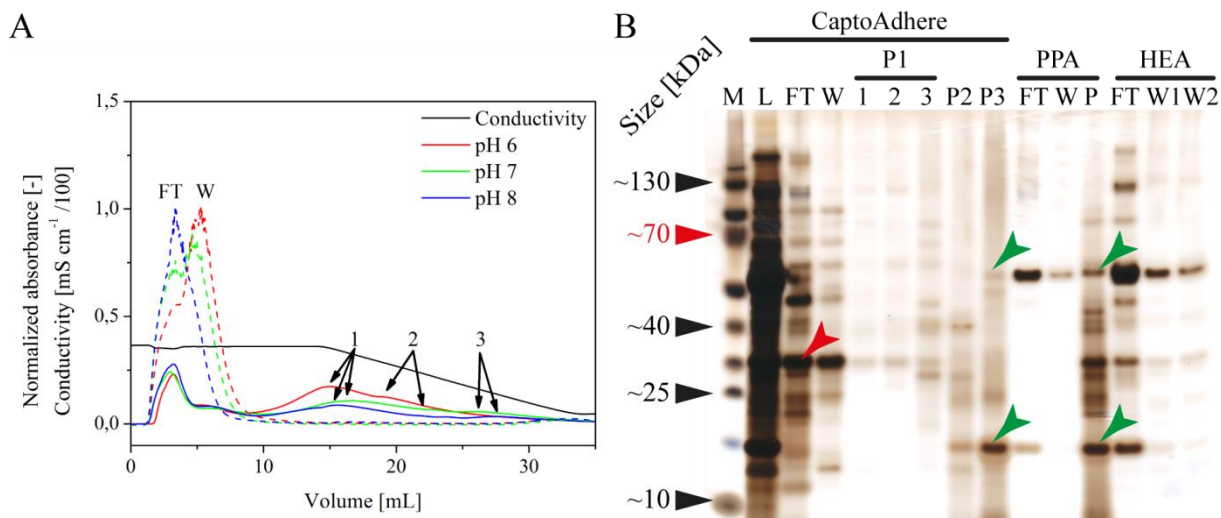


Figure VI-40: Analysis of fractions collected during HCP separation by different mixed-mode resins (IV.4.4, IV.11, Table III.2).

A. At a conductivity of 75 mS cm^{-1} (dashed lines) the entire UV 280 nm signal was observed in the FT and wash fractions for chromatograms of HCP separation by CaptoAdhere in a pH range from 6.0-8.0. Three elution peaks (numbered arrows) appeared at a conductivity of 30 mS cm^{-1} (solid lines). Their share of the total area under the curve decreased from 70% at pH 6.0 to 54% at pH 8.0. B. Silver-stained SDS-PAA gel loaded with protein samples from fractions obtained from chromatography runs at pH 7.0 and a conductivity of 30 mS cm^{-1} . Only a small fraction of the loaded amount of RuBisCO was found in the FT, wash or elution fractions of CaptoAdhere or PPA derived samples (green arrows). DsRed (red arrow) was found in the FT of CaptoAdhere and HEA, while it bound to PPA. All fractions were obtained from chromatography runs at pH 7.0 and 30 mS cm^{-1} . M – marker; L – load; FT – flow through; W – wash; P – peak.

VI.9.4.3 HEA

Proteins were allowed to bind to HEA under the same conditions as above (VI.9.4.2). At 75 mS cm^{-1} all UV 280 nm absorbance (Figure VI-41 A) and protein bands detectable on silver-stained SDS-PAA gels (data not shown) were found as a single peak in the FT and early wash fractions, indicating weak protein-ligand interactions, if any. An individual second peak was observed in the wash fractions when conductivity was reduced to 30 mS cm^{-1} during sample loading (Figure VI-40 B and Figure VI-41 A) indicating a stronger interaction between protein and ligand which was still not sufficient for complete retention. The second peak contained the RuBisCO large subunit as a major protein component and diminished at pH values >7.0 . This was consistent with the ligand pK_a value of 8.0, resulting in only 50% of ligand groups becoming charged at pH 8.0 and thus reducing resin binding capacity compared to pH 7.0 (90% charged ligand) [342].

Further reduction of the extract conductivity can increase protein-ligand interactions and thus result in the complete retention of some proteins. However, this would require sample

conditioning and thus CaptoAdhere (no sample conditioning necessary) or STAR AX (no hydrophobic ligand part that can cause protein denaturation) remain better choices.

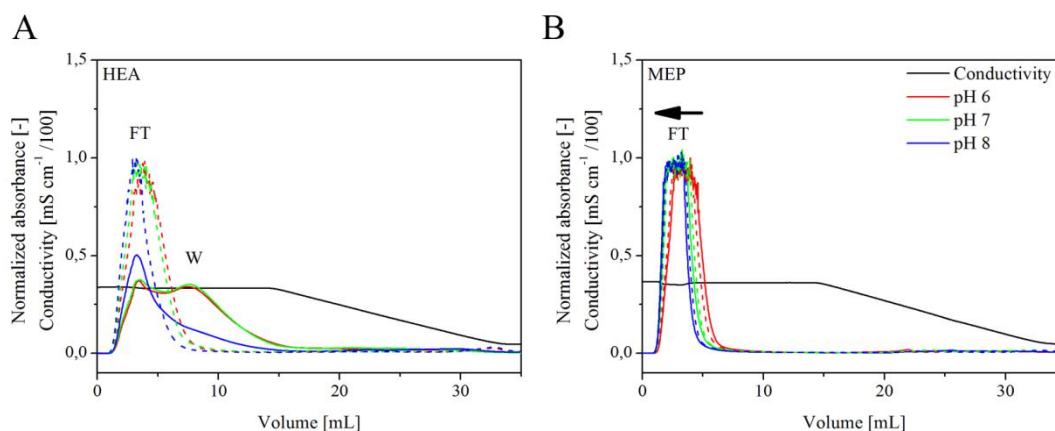


Figure VI-41: Tobacco HCP binding to HEA and MEP mixed-mode resins at pH 6.0-8.0 at 30 or 75 mS cm⁻¹ (IV.11, Table III.2).

A. For HEA, at 75 mS cm⁻¹ (dashed lines) all UV 280 nm signal was found in FT independent of pH while reducing the conductivity to 30 mS cm⁻¹ (solid lines) shifted 50% of the signal into the wash (W) fractions but still no peak was observed during the subsequent low salt/low pH gradient. B. Independent of conductivity, all UV 280 nm signal was found in the FT when proteins were loaded to MEP. An increasing pH slightly shifted the FT peak towards a smaller retention volume (arrow). FT – flow through; W – wash.

VI.9.4.4 MEP

When the same loading conditions (VI.9.4.2) were applied to MEP, UV 280 nm absorbance was found mostly in the FT fraction and partly in the wash fractions (Figure VI-41 B) regardless of conductivity, reflecting ineffective binding to the resin. Increasing the buffer pH to 8.0 prevented the slight interactions of proteins with the resin that were observed at pH 6.0. This ligand can ideally be used in binding mode if the target protein can adsorb under any of the investigated conditions, allowing separation from nearly all tobacco HCPs and avoiding capacity limitations reflecting the binding of impurities. However, target protein binding to the ligand is unlikely because MEP was designed to be immunoglobulin-selective [343].

VI.9.4.5 PPA

Proteins loaded onto PPA at high conductivity (VI.9.4.2) passed through the column as a single FT peak (Figure VI-42 A). Reducing the buffer pH from 8.0 to 6.0 shifted this peak slightly towards higher retention volumes, but no elution peak was observed either by UV 280 nm or silver-staining the fractions separated on SDS-PAA gels (data not shown). A reduced buffer conductivity increased protein ligand interactions resulting in a second peak in the wash fractions containing RuBisCO (Figure VI-40 B) as observed for HEA (VI.9.4.3). A pH of 8.0 partially reversed this shift due to the ligand pK_a as discussed for HEA [342].

Although no elution peak was observed at pH 6.0, about 16% of the total UV 280 nm signal was found in a peak at the end of the pH 7.0 elution gradient, and this contained the RuBisCO small subunit as a major protein component but only a little of the large subunit, reflecting the irreversible binding of this protein to the resin. Accordingly, PPA should only be considered for FT mode operations to avoid target protein denaturation/irreversible binding. However, RuBisCO binding may quickly exhaust the resin capacity, so large-volume columns would be required in order to remove all HCPs.

In summary, CaptoAdhere was found to be the most promising mixed-mode ligand for the direct processing of filtered plant extracts because (i) the highest percentage of proteins was released from the resin during gradient elution (limited ‘irreversible’ binding) rendering binding mode a feasible operation mode, and (ii) because this resin achieved the widest operation range in terms of pH (Figure VI-42 B).

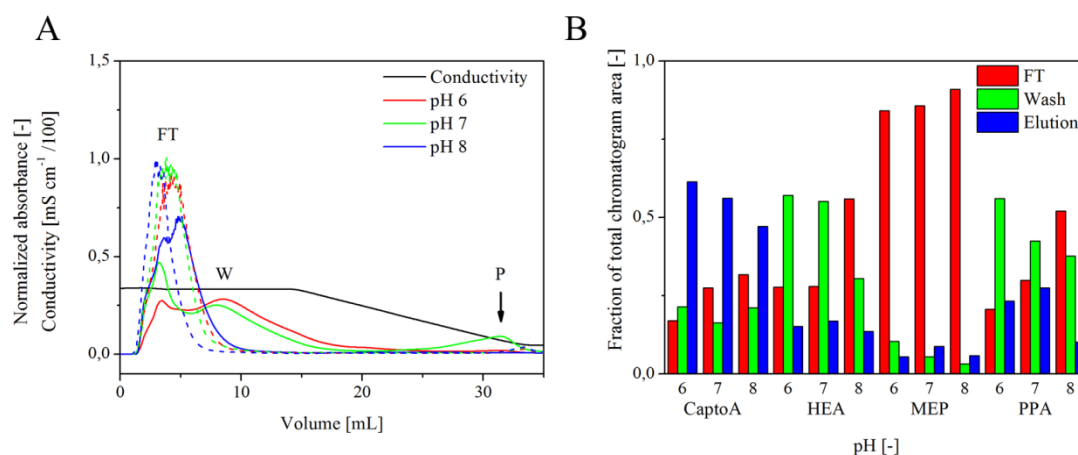


Figure VI-42: HCP binding to PPA and comparison of different mixed-mode ligands (IV.11, Table III.2).

A. For PPA, at 75 mS cm^{-1} (dashed lines) all UV 280 nm signal was found in FT and wash (W) fractions independent of pH while reducing the conductivity to 30 mS cm^{-1} (solid lines) shifted 50% of the signal into the wash fractions. Only for pH 7.0 a single peak at the end of the low salt/low pH gradient was observed (P). B. CaptoAdhere exhibited the largest share of UV 280 nm during elution (blue columns) compared to FT and wash fractions (red and green columns respectively) indicating that proteins not only bound to the ligand at 30 mS cm^{-1} but were also eluted again. Hence, binding-mode is an option for this resin and protein denaturation appeared less likely. Additionally, protein binding was achieved over a wider pH range than for the other ligands. FT – flow through; W – wash; P – peak.

VI.9.5 pH gradients

In addition to salt elution gradients, a shift in pH can also be used to elute bound proteins from a resin as discussed above (VI.9.4.2 to VI.9.4.5). The elution pH of HCPs can be used to optimize conditions for intermediate or polishing steps. This is because commonly applied salt gradients will achieve a higher protein elution selectivity if the buffer pH is close to the elution pH of one of the proteins in the mixture (either HCPs or the target protein) since this protein

will have only a low surface charge compared to the other proteins and will thus respond quickly to a change in salt concentration [344].

Linear pH gradients are required for the reliable determination of protein elution pH values but these can be tedious to establish [345]. Therefore, different buffers, resins and gradients were tested for their ability to maintain a linear pH gradient during tobacco HCP elution.

A gradient from 40 mM Tris pH 8.0 to 40 mM citrate pH 3.9 did not achieve linear pH reduction at the outlet of a DEAE column during tobacco HCP elution when a 2 mL sample (~8.0 mg TSP) was loaded onto a 1 mL column at a constant flow rate of 1 mL min⁻¹ (Figure VI-43 A). A single steep shift from pH 6.5 to 4.5 was observed within 4 CVs. This issue was not solved by using buffer mixtures of 20 mM Tris/20 mM citrate containing 10 mM NaCl at pH 8.0 and 3.9. Replacing DEAE with Q and applying a non-linear buffer gradient resulted in a near-linear pH gradient (Figure VI-43 B) but this was not reproducible because there was an occasional sharp drop from pH 6.0 to 4.0 within 4 CVs. This seemed to reflect the simultaneous elution and denaturation of tobacco HCPs bound to the resin at pH <6.5 (see also VI.5.3 and VI.9.2) because (i) the pH shift was accompanied by a UV 280 nm peak and (ii) no sharp pH drop was observed in the absence of loaded protein (data not shown). The linearity and reproducibility of the pH gradient was no issue when the buffers developed by Frieder Kröner (personal communication) was used in combination with MonoQ or MonoS columns of 10 cm bed height (GE) (Figure VI-44, IX.7).

Separation of tobacco HCPs in a pH 11.0–3.5 gradient on a MonoQ column resulted in a FT fraction and three major gradient elution peaks, the first and last of which consisted of three non-baseline separated sub-peaks (Figure VI-44 A). The FT fraction and peak 1 contained barely detectable amounts of protein (Figure VI-45). DsRed was found in elution peak 2 which was consistent with the high UV 495 nm signal (absorbance shoulder of DsRed) observed for this peak and thus had an elution pH of ~7.3. The RuBisCO small subunit was found in peak 3 with an elution pH of ~3.8-4.8 but the large subunit was not detected in any of the elution fractions. This indicated that the RuBisCO large subunit was precipitated at a pH above its elution pH (<5.5) based on the previously observed quantitative precipitation of this protein below pH 5.5 (VI.5.3 and VI.9.2). Interestingly, only ~25% of the loaded TSP was retrieved during pH elution, indicating major protein precipitation/denaturation before the individual elution pH levels were reached.

In contrast, >70% of loaded TSP was retrieved in a pH 4.0–11.0 gradient on MonoS, and 18 of the 25 elution peaks were detected in the first 40% of the gradient (pH 4.0–6.8, Figure

VI-44 B). Accordingly, the protein composition of the corresponding samples changed for each fraction (Figure VI-46). The RuBisCO small and large subunits were the major protein components of the dominant elution peaks at pH 7.0 and 10.6, respectively. DsRed eluted shortly before the small subunit at pH 6.5. Additional 130, 20 and 12 kDa proteins eluted with large subunit. Interestingly, many proteins including DsRed and the RuBisCO small subunit exhibited extensive tailing during pH elution and were detected in most of the elution fractions. This tailing made it difficult to determine a precise elution pH for the different proteins.

Some proteins exhibited similar elution pH values on MonoQ and MonoS, including DsRed (pH 7.3 and 6.5 respectively), which has also been observed for monoclonal antibodies [344]. However, these values normally differed for the two resins, e.g. the elution pH values for the RuBisCO small subunit were ~4.3 and 7.0 respectively. This may reflect differences in the binding sites and orientations involved in protein adsorption to the Q and S ligands and it may be necessary to determine the protein elution pH for AEX and CEX separately [346].

Another problem during elution pH determination was protein precipitation due to (i) sample pH adjustment prior to column loading to the column and (ii) precipitation/denaturation of proteins bound to the column. It was therefore impossible to determine the elution pH for affected proteins such as the RuBisCO large subunit on MonoQ. A closely-related issue was that the elution pH may not reflect the pH at which protein-ligand interactions were disrupted but rather the pH at which interactions between proteins of a bound protein complex were disrupted, resulting in the release of one component while the others remained bound to the resin. The release of single components from a complex can induce or facilitate protein denaturation [347, 348] and thus explain for example the lack of RuBisCO large subunit (RBCL) in any of the MonoQ elution fractions despite the presence of the small subunit (RBCS). This mechanism was also consistent with the tailing observed for DsRed (a homotetramer) and the small subunit on MonoS assuming that a partial release will initially result in bound DsRed trimers then dimers and monomers. More complex rearrangements seemed likely in the case of RuBisCO (hetero hexadecamer of 8:8 RBCL:RBCS) such as 8:4 (RBCL:RBCS) releasing four small subunits per complex, resulting in elution peak 1 (fraction 10 in Figure VI-46), followed by 8:2 causing the weak tailing of the small subunit and then the release of the remaining complex in elution peak 2.

In summary, the analysis of elution pH can help to characterize heterogeneous feedstock [328, 344, 349], but additional studies of tobacco extracts will be required in order to rule out the abovementioned issues and determine reliable values for the elution pH of different proteins.

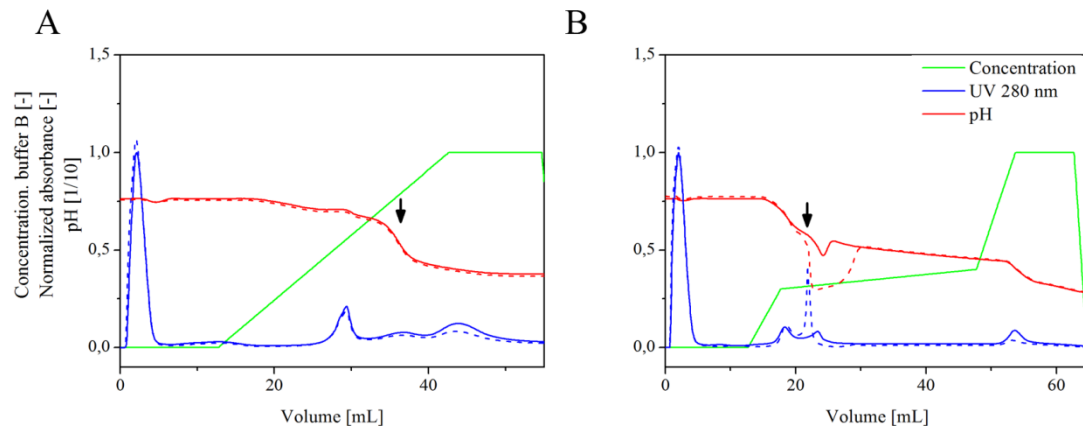


Figure VI-43: Non-optimal buffer conditions can prevent a reproducible linear pH gradient (IV.11).

A. A linear gradient from pH 8.0 (40 mM Tris) to pH 3.9 (40 mM citrate) did not result in a linear pH gradient at the outlet of a DEAE column to which tobacco HCP had bound. Instead a sharp drop from pH 6.5 to 4.5 (arrow) was observed even if Tris/citrate mixtures (20 mM each) containing 10 mM NaCl were used (dashed lines). B. Using Q instead of DEAE and a non-linear buffer gradient results in a partially linear pH gradient at the column outlet (solid lines). However, this gradient was not reproducible (dashed lines) and a sharp pH drop can still occur coinciding with protein elution/denaturation.

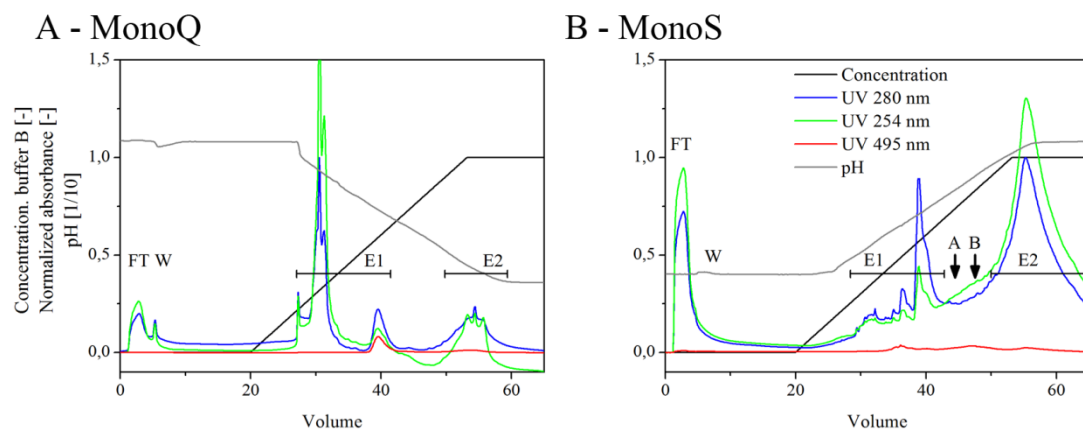


Figure VI-44: Separation of tobacco HCPs on 10 cm bed height MonoQ or MonoS columns using an optimized buffer composition (IV.11, IX.7).

A. Besides the FT, three major elution peaks were observed during pH 11.0-3.5 gradient elution on MonoQ. The first and last elution peak consisted of three non-baseline separated sub-peaks. Horizontal bars E1 and E2 indicate elution ranges of which subsequent 1 mL fractions were analyzed by silver-staining of SDS-PAA gels (Figure VI-45). B. Two dominant peaks were found among the total of 25 elution peaks during tobacco HCP separation on MonoS. Horizontal bars E1 and E2 indicate elution ranges of which subsequent 1 mL fractions were analyzed by silver-staining of SDS-PAA gels (Figure VI-45). Additional samples A and B were also analyzed. FT – flow through; W – wash; P – peak; E – elution.

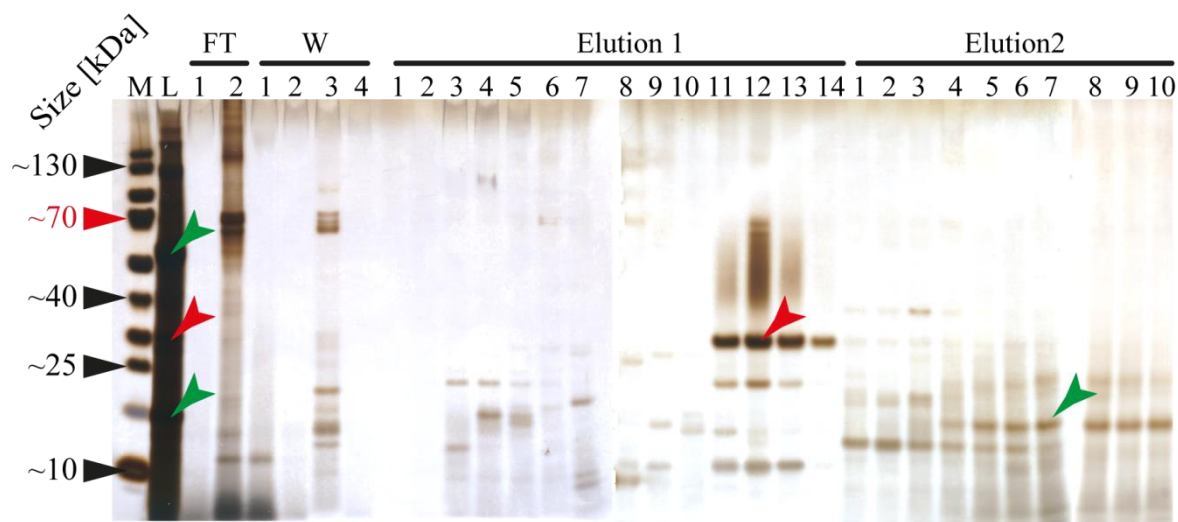


Figure VI-45: SDS-PAGE and silver-staining analysis of 1 mL fractions from MonoQ pH gradient elution (IV.4.4, IV.11 and IX.7).

The late FT fraction and the small wash peak (W3) contained several ~70 kDa proteins, whereas fractions from elution peak 1 and its sub-peaks (Elution 1, 1-8; pH 10.0-9.0) did not contain significant amounts of protein. DsRed (red arrow) was found to be the major protein component of elution peak 2 (~pH 7.3) being consistent with the UV 495 nm signal of this peak (Figure VI-44 A). In the initial fractions of elution peak 3 (Elution 2, 1-10; pH 4.8-3.8) a ~12 kDa protein was the major component which changed to RBCS (green arrow, 15 kDa) in the later fractions. Interestingly, none of the analyzed elution fractions contained RBCL (green arrow, 55 kDa). M – marker; L – load; FT – flow through; W – wash. Elution 1 and 2 ranges correspond to horizontal bars in Figure VI-44 A.

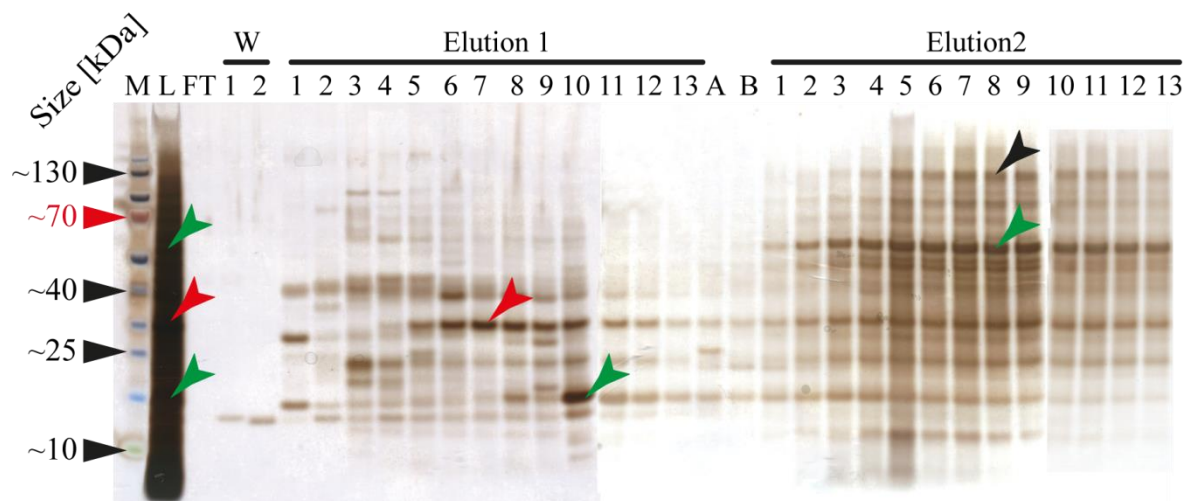


Figure VI-46: SDS-PAGE and silver-staining analysis of 1 mL fractions from MonoS pH gradient elution (IV.4.4, IV.11 and IX.7).

Only a ~12 kDa protein was found in the wash and none in the FT fraction. Protein composition in early fractions of elution 1 (1-5) changed considerably corresponding to the numerous peaks observed for this pH range in Figure VI-44 B. RBCS (green arrow, 15 kDa; ~pH 7.0) accounted for the first major peak during pH elution with DsRed (red arrow; pH 6.5) eluting at a slightly earlier. RBCL (green arrow, 55 kDa; ~pH 10.6) and ~130 kDa protein (black arrow) were found in elution 2 (1-13), however, also DsRed and RBCS were found in these fractions due to extensive tailing. A – fraction at 44 mL; B – fraction at 47 mL; M – marker; L – load; FT – flow through; W – wash. Elution 1 and 2 ranges correspond to horizontal bars in Figure VI-44 B.

VI.9.6 Implications for a purification strategy

The major tobacco HCP RuBisCO was shown to bind to AEX resins at various pH values, therefore FT mode promises the best separation of a target protein from this impurity (VI.9.1).

Due to the good separation achieved with Q operation in binding mode seemed also feasible and an increased pH, lower flow rates as well as shallower gradients can further improve peak resolution. The salt tolerant STAR AX resin can simplify the purification process because buffer exchange can be replaced by a simple 1:1 dilution with water and conditioning is unnecessary if the extract conductivity can be reduced to 14 mS cm^{-1} , e.g. by changing the salt concentration in the extraction buffer (VI.9.4.1). However, regardless of the operation-mode, the capacity of all AEX resins can be a limiting or cost-driving factor because of the abundance of RuBisCO and its tendency to bind these ligands.

In contrast, the CEX resin SP can be operated in binding mode at pH 6.5 (if the target protein adsorbs at this pH) preventing most of the HCPs from binding, including RuBisCO (VI.9.2). This concentrates the target protein after elution and avoids the capacity problem described for AEX. In addition, host cell DNA is unlikely to bind to the negatively-charged SP resin, helping to remove this impurity together with the HCPs in the FT fraction.

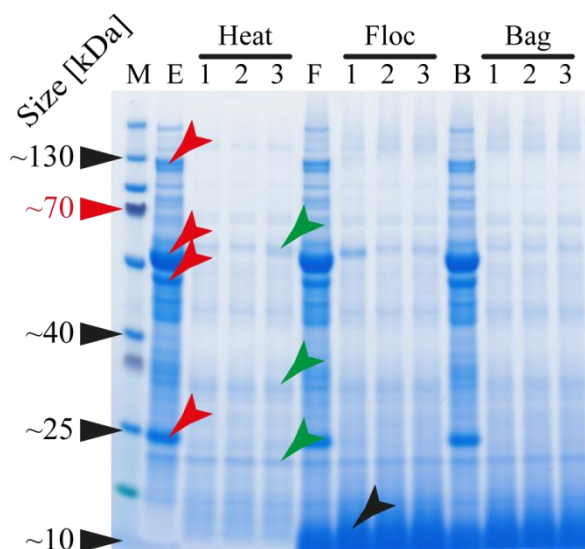
CaptoAdhere can be operated in FT and binding modes, but the latter allows low-pH and low-conductivity elution, making the elution fractions ideal for subsequent loading onto SP resin. RuBisCO binding can limit the column capacity and protein denaturation (especially at low pH) can result in the precipitation of tobacco HCPs. Instead, protein desorption can also be achieved by high salt concentrations or the addition of mobile phase modifiers [155, 341] making the elution compatible with HIC. In the results presented above, HIC was only evaluated as a first capture step. Therefore, additional experiments should be carried out in the future to determine how HIC performs if most RuBisCO has been removed. Only minimal conditioning for high-salt HIC loading would be necessary after an initial IEX capture step, resulting in a concentrated process intermediate and high conductivity.

Information about the elution pH of HCPs and target proteins will help to specify the operational conditions that need to be screened for specific separation problems, and also for intermediate purification steps [328, 344, 349]. Even so, the problems and ambiguities described above for tobacco HCPs (VI.9.5) must be addressed in order to exploit the full potential of this method.

The abundance of RuBisCO was a major challenge in most of the chromatographic methods investigated for the capture and purification of recombinant proteins from tobacco extracts. The abundance can be reduced along with many other HCPs by using an acidic extraction pH as described above (VI.5.3 and VI.9.2) and in the literature [75]. Heat precipitation may be used to remove HCPs as well (Figure VI-47 A) [350]. However, these methods are limited to target proteins that are extractable at low pH or stable at high

temperatures and this cannot be a generic solution to every process. Even if a target protein is soluble at low pH or does not precipitate when heated, these conditions can irreversibly affect protein folding, activity, efficacy and safety [351, 352] and should thus be carefully evaluated.

A - Heat



B - Membrane

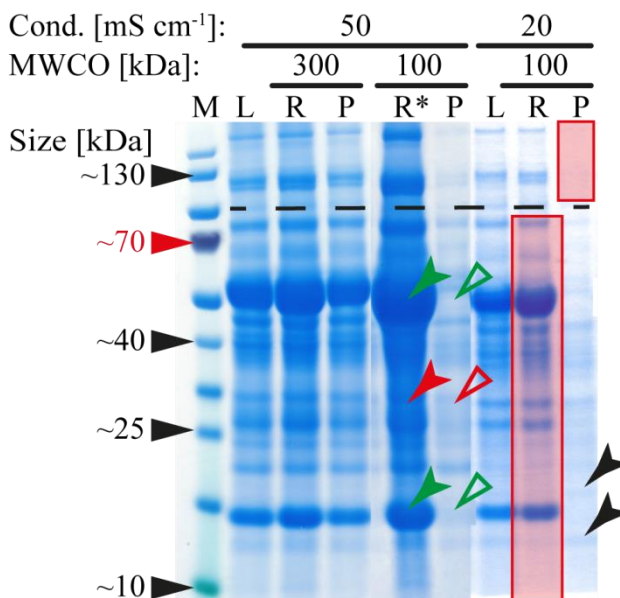


Figure VI-47: Methods of RuBisCO removal prior to chromatographic purification of target proteins (IV.4.4).

A. Heating plant extract to 65°C for 2 min (Heat) will denature and precipitate the most abundant HCPs (red arrows), while only a limited number of proteins remains in solution (green arrows). Use of flocculant (Floc, Polymin P, black arrow) or bag filters (Bag) does not alter the protein pattern after heat treatment. B. A 300 kDa MWCO membrane does not affect tobacco HCPs (even distribution in retentate R and permeate P) whereas a 100 kDa membrane retains RuBisCO (solid green arrows), DsRed (solid red arrow) and several other proteins despite molecular masses below 60 kDa and only few HCPs pass the membrane (black arrows). The HCP retention is observed at different conductivities. The gel areas highlighted in red would be free of proteins if HCPs are present only as monomers in extract. * undiluted retentate sample to demonstrate concentrating effect of the 100 kDa membrane. M – marker; E – extract; F – flocculated extract; B – bag filtrate; L – load; R – retentate; P – permeate.

In contrast, membrane filtration using suitable MWCO values is a gentle alternative for the non-chromatographic size fractionation of proteins. As shown above (VI.9.1.3), the RuBisCO subunits form oligomers even under extraction conditions in the pH range 4.5–8.5, which prevents the small subunit (~15 kDa) from passing through a 30 kDa MWCO membrane. Both RuBisCO subunits were also quantitatively retained by a 100 kDa membrane at 20 and 50 mS cm⁻¹ indicating that extract conductivity and thus electrostatic interactions do not influence its association within the investigated range (Figure VI-47). No retention of RuBisCO was observed when a 300 kDa membrane was used for separation, thus RuBisCO was not present in its native hetero hexadecameric form (~560 kDa). Instead, an intermediate oligomerization status seemed likely, e.g. 4:4 (~280 kDa) or 2:2 (~140 kDa), which is still large compared to common non-antibody biopharmaceutical proteins such as LicKM E7 GGG (~37 kDa) [353]. A 100 kDa membrane could therefore be useful to separate RuBisCO from

smaller target proteins. Interestingly, many other HCPs <100 kDa were also retained by this membrane, indicating their oligomerization under the investigated conditions (Figure III.6).

Future experiments should investigate in more detail the oligomeric structure of RuBisCO (and other abundant proteins such as RuBisCO activase) in tobacco extracts at different pH and conductivity values, e.g. by SEC analysis preferably in combination with multi-angle laser light scattering (MALLS). Membranes with different MWCOs than those tested here may further improve separation of HCPs and target proteins.

The severity of the restricted column capacity discussed above (VI.9.1 and VI.9.4) would diminish if most of the RuBisCO could be removed from the tobacco extract, allowing AEX resins to be used in binding mode. This would expand the separation options available because the different selectivities of the resins could be exploited, e.g. a target protein could be separated from impurities by shifting one of them into the FT by selecting appropriate conditions.

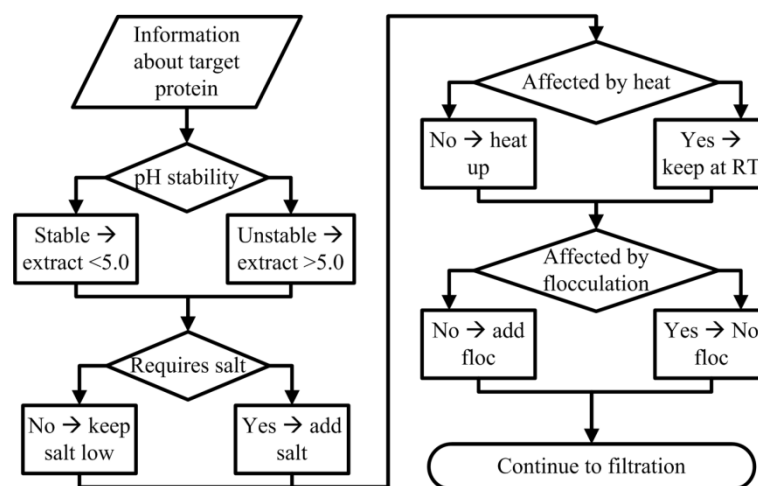


Figure VI-48: Generic selection scheme for extraction conditions of biopharmaceutical proteins produced in tobacco plants.

Extraction of a POI at low pH and low conductivity reduces the amount of co-extracted HCPs and increases compatibility with subsequent IEX (e.g. SP) respectively. A heat treatment can further reduce the concentration of tobacco HCPs and addition of flocculants greatly enhances capacity of depth filters.

Different generic purification processes were developed based on the results presented in this chapter (VI.9) and previous chapters (VI.2 to VI.6) and are shown in Figure VI-48, Figure VI-49 and Figure VI-50. At first, several conditioning steps can be used to reduce the number of HCPs in the extract and improve process performance in dependence of the POI properties (Figure VI-48). In case all HCPs are present at the time of target capture (Figure VI-49), POIs with a basic pI should be purified after a buffer exchange by capture on CEX resins such as SP at a pH of ~6.5 because this will direct most HCPs to the FT fraction (VI.9.2). Depending on the elution salt concentration, HIC or MMC can be used as a second purification step. POIs with a neutral pI can be captured on salt-tolerant STAR AX after HCPs have been

bound to an anionic mixed-mode resin such as CaptoAdhere. Sample dilution or re-buffering may be necessary to promote target protein binding. Buffer exchange and SP or a combination of HIC and SP can be used as subsequent purification steps, depending on the elution buffer salt concentration.

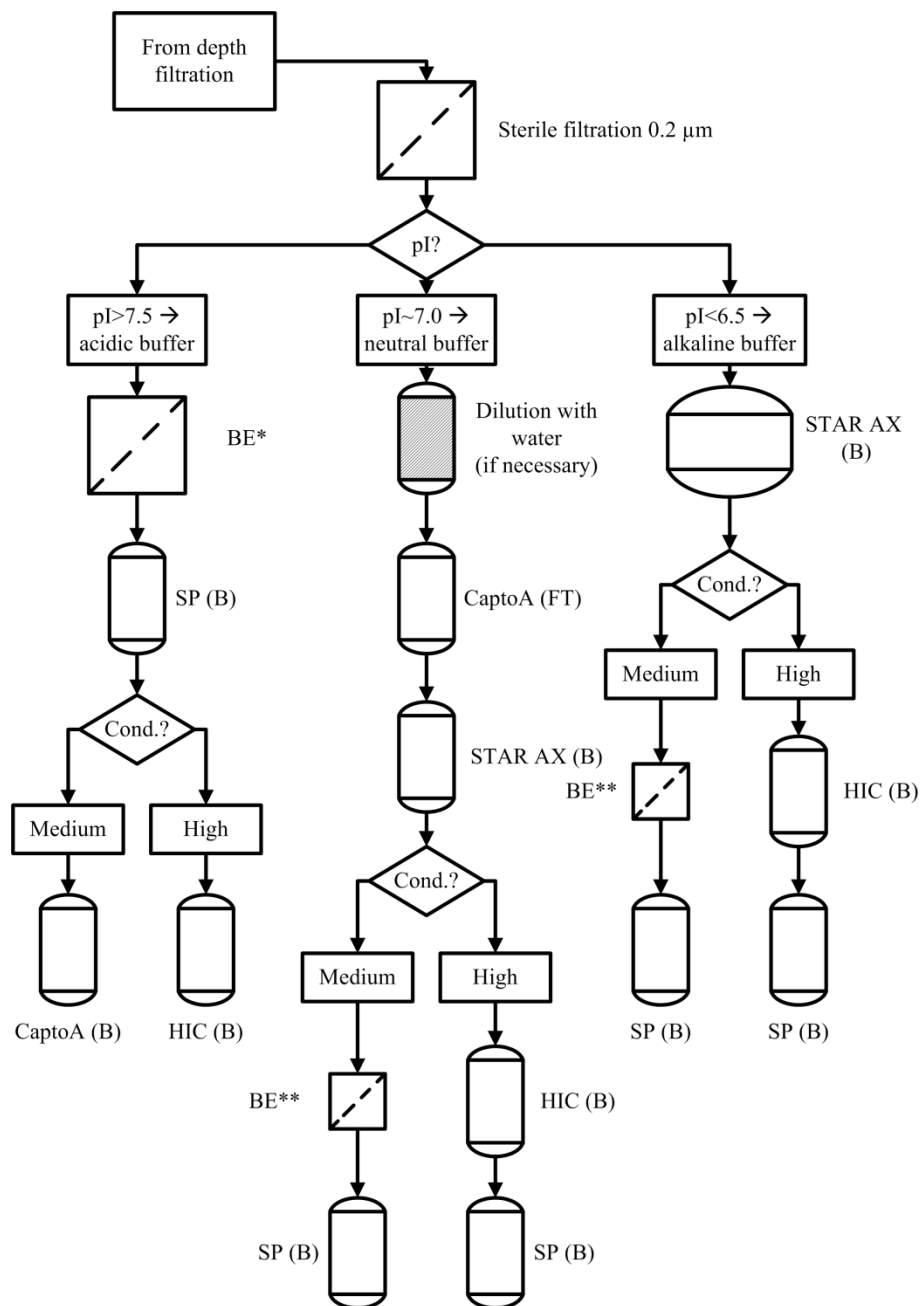


Figure VI-49: Generic purification scheme for filtered tobacco extract (IV.11 and VI.9).

Acidic proteins can be purified after buffer exchange to pH 6.5 and conductivity <5.0 mS cm^{-1} (BE*) through binding to SP and depending on the elution salt concentration subsequent binding to CaptoA or HIC. Some HCPs can be removed by CaptoA before a neutral protein is bound to STAR AX at pH 8.5. Depending on the elution conductivity, buffer exchange to low salt and low pH (BE**) followed by SP or a combination of HIC and SP can be applied for further purification. Basic proteins require a large column volume for capture on STAR AX because RuBisCO will also bind. Depending on the elution conductivity, buffer exchange to low salt and low pH (BE**) followed by SP or a combination of HIC and SP can be applied for further purification.

A high-capacity resin is required to capture acidic target proteins because most tobacco HCPs will bind under the same conditions (VI.9.1). The salt-tolerant AEX resin STAR AX can be used for this purpose because it is also compatible with high salt concentrations in plant extracts, circumventing the need for initial sample conditioning. The subsequent purification steps would be the same as those used for POIs with a neutral pI.

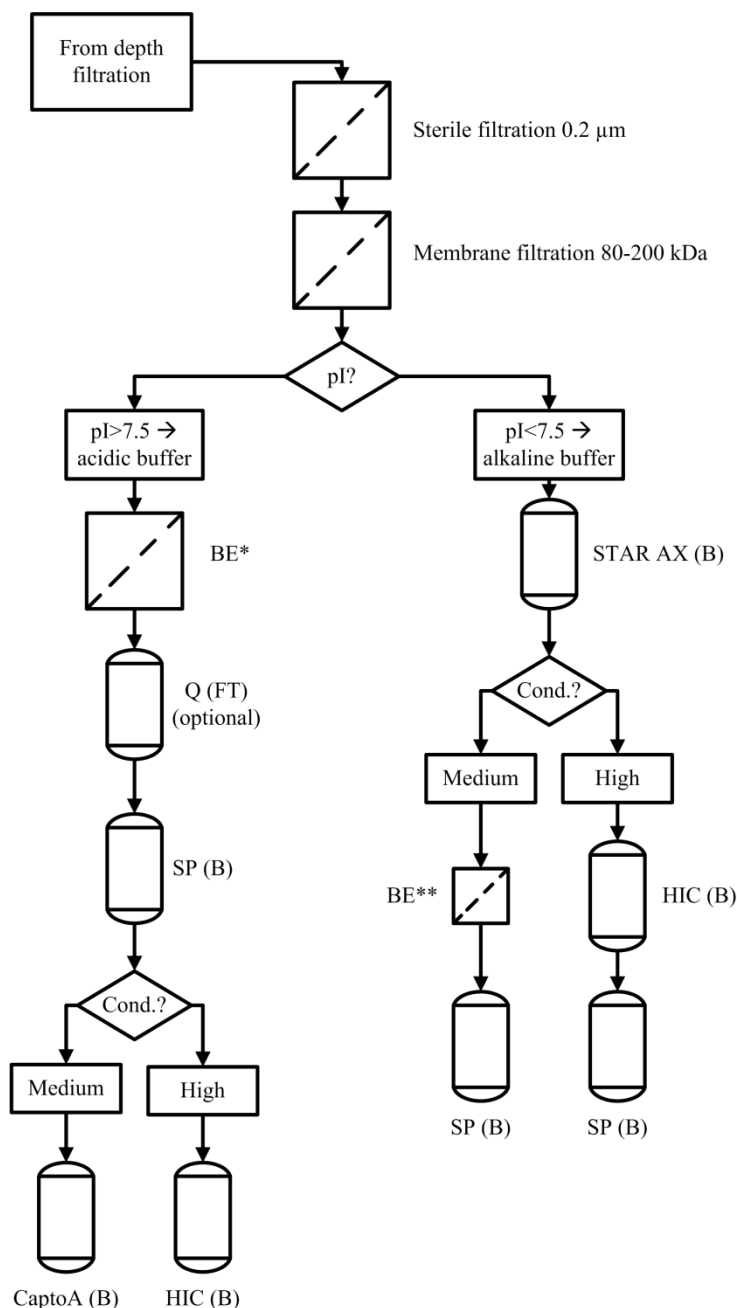


Figure VI-50: Generic purification scheme for RuBisCO depleted filtered tobacco extract (IV.11 and VI.9). After adequate membrane filtration the majority of RuBisCO is removed from tobacco extract (Figure VI-47) facilitating further processing. Acidic proteins can be captured with SP after buffer exchange to pH 6.5 and conductivity $< 5.0 \text{ mS cm}^{-1}$ (BE*). Operating a Q column in FT-mode before SP can further reduce the amount of HCP. Neutral and basic target proteins can be captured using STAR AX. Depending on the elution conductivity, buffer exchange to low salt and low pH (BE**) followed by SP or a combination of HIC and SP can be applied for further purification.

The purification scheme can be simplified if HCPs such as RuBisCO are removed by membrane filtration prior to chromatography (Figure VI-50). Targets with a pI of <7.5 can be captured on a small STAR AX column followed by purification as described above. For POIs with a pI >7.5 , the same purification strategy can be applied, but a Q column may be used to capture trace amounts of negatively-charged HCPs at pH 6.5. Several other empirical and rational selection criteria for assembling chromatography separation steps have been proposed recently [167, 328, 354-357].

In future experiments additional inexpensive buffer substances and mobile phase modifiers [155] should be screened because they can alter the selectivity of chromatography, as has been observed for Q resins (VI.9.1.3).

VI.10 Predicting chromatographic separation

The results presented in this section have been published as the following manuscript:

1. Buyel JF, Woo JA, Cramer SM, Fischer R. 2013. The use of quantitative structure-activity relationship models to develop optimized processes for the removal of tobacco host cell proteins during biopharmaceutical production. *Journal of Chromatography A* 1322:18-28.

Identifying the most suitable conditions for chromatographic separations can be time consuming, laborious and expensive (III.6.5) as shown by the results presented above (VI.9) and in the literature [328]. Therefore, an attempted was undertaken to model the behavior of tobacco HCPs under different chromatography conditions.

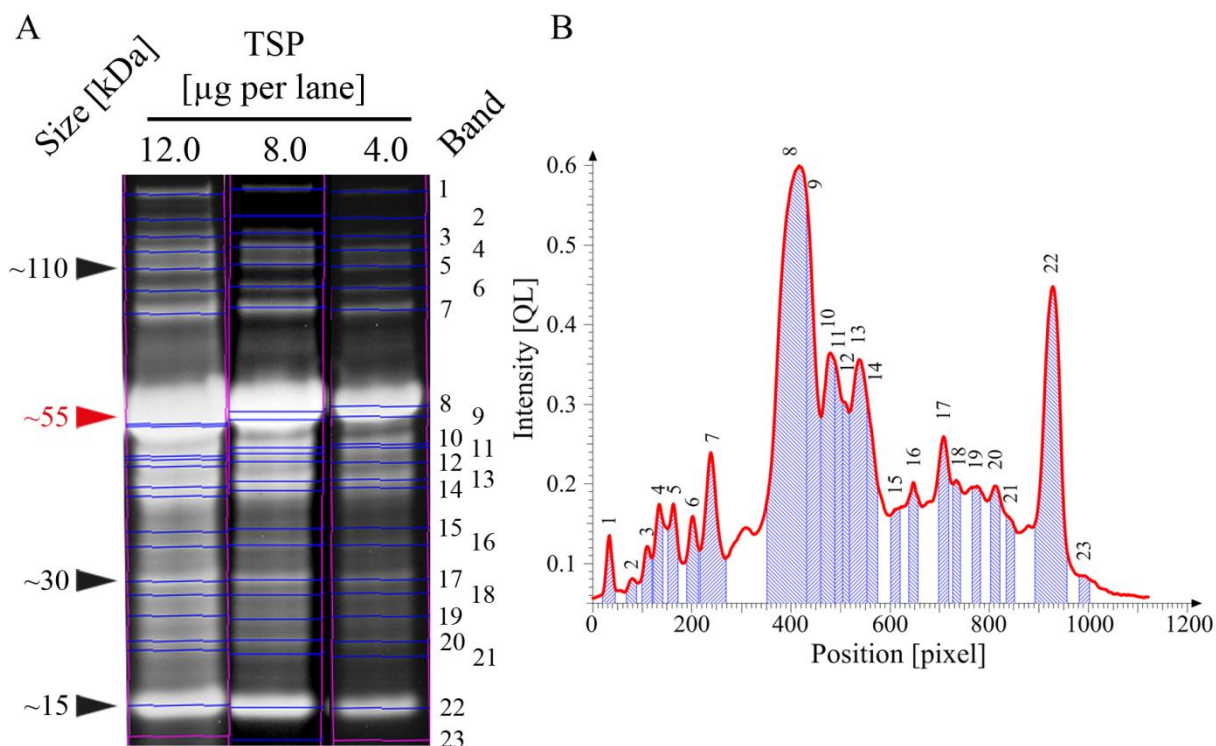


Figure VI-51: Protein band isolation and quantity assignment (IV.4.4 and IV.10.1).

A. Varying amounts of TSP were separated by SDS-PAGE and 23 protein bands were identified each time. Dominant bands at 55 kDa and 15 kDa correspond to the RuBisCO large and small chains, respectively. B. Densitometric analysis of 8.0 µg TSP lane from A. The TSP fraction within each band was calculated as the relative area ratio of the corresponding densitogram peak and averaged over all three dilutions (lanes in A).

VI.10.1 Identification and homology modeling of tobacco HCPs

HCPs were extracted from tobacco leaves at a pH of 7.5 and a conductivity of 45 mS cm⁻¹ as described above (IV.5 and VI.1). These conditions were chosen because (i) they guaranteed the efficient extraction of a broad range of proteins and (ii) they represented standard conditions established for monoclonal antibodies [255]. A total of 107 individual proteins identified by PMFs using the Mascot database (IX.5) was present in 23 distinct bands (Figure VI-51 A) isolated from SDS-PAA gels (IV.4.4). As expected, the large and small subunits of RuBisCO were dominant proteins in the tobacco extract accounting for 19% and 10% of TSP respectively. Other abundant proteins included RuBisCO activase (2% TSP), glyceraldehyde-3-phosphate dehydrogenase isomers (3% TSP) and sedoheptulose-1,7-bisphosphatase, cysteine synthase isomers and mRNA binding proteins (1% TSP each).

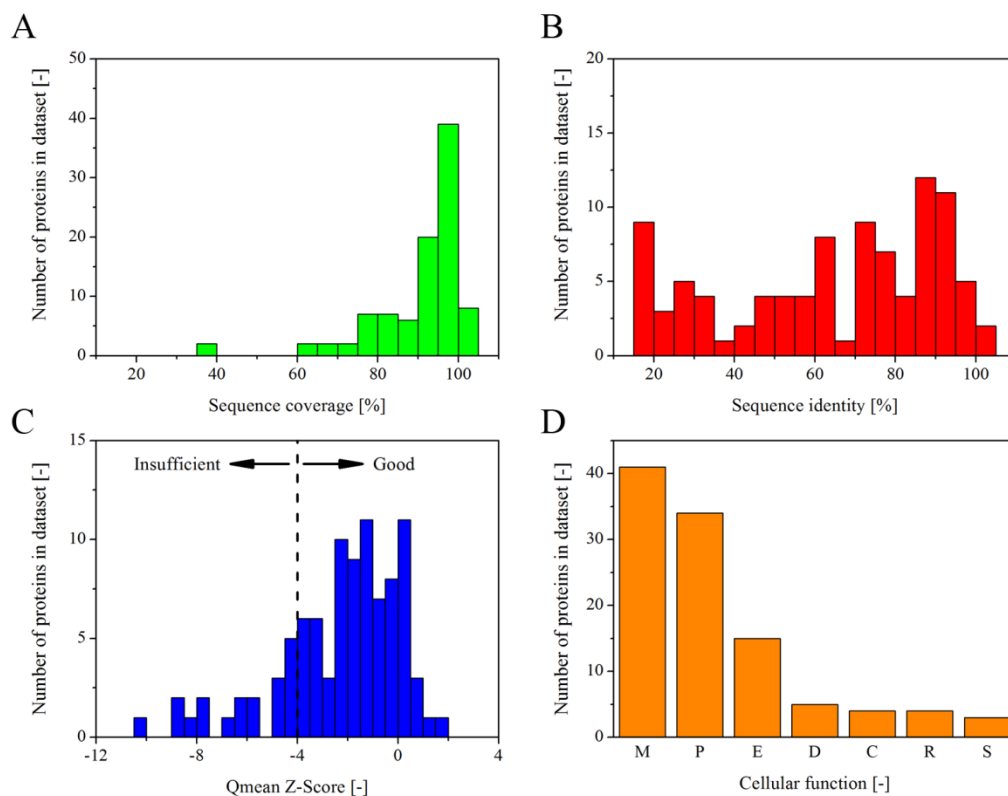


Figure VI-52: Homology model quality evaluation.

A. Half of the homology models for identified tobacco HCPs had a very high sequence coverage of >95%. B. No significant bias was observed for the sequence identity between template and model sequences. C. An acceptable Z-score of >-4 was calculated for 80% of the homology models. D. Most of the identified tobacco HCPs were related to cell metabolism (M), protein synthesis and turnover (P) and energy (E), whereas only few proteins belonging to DNA and genome organization (D), cell structure (C), resistance and stress response (R) and signaling (S) were found in aqueous leaf extracts.

The proteins were subdivided into seven functional classes as shown in Figure VI-52. Most were related to cellular metabolism (e.g. carbohydrate turnover), protein synthesis and energy metabolism. Corresponding amino acid sequences were retrieved from UniProt [256], nine of which contained non-annotated potential leader/signal peptides that were predicted using the TargetP 1.1 webserver [260, 261] and excluded from the modeled structures.

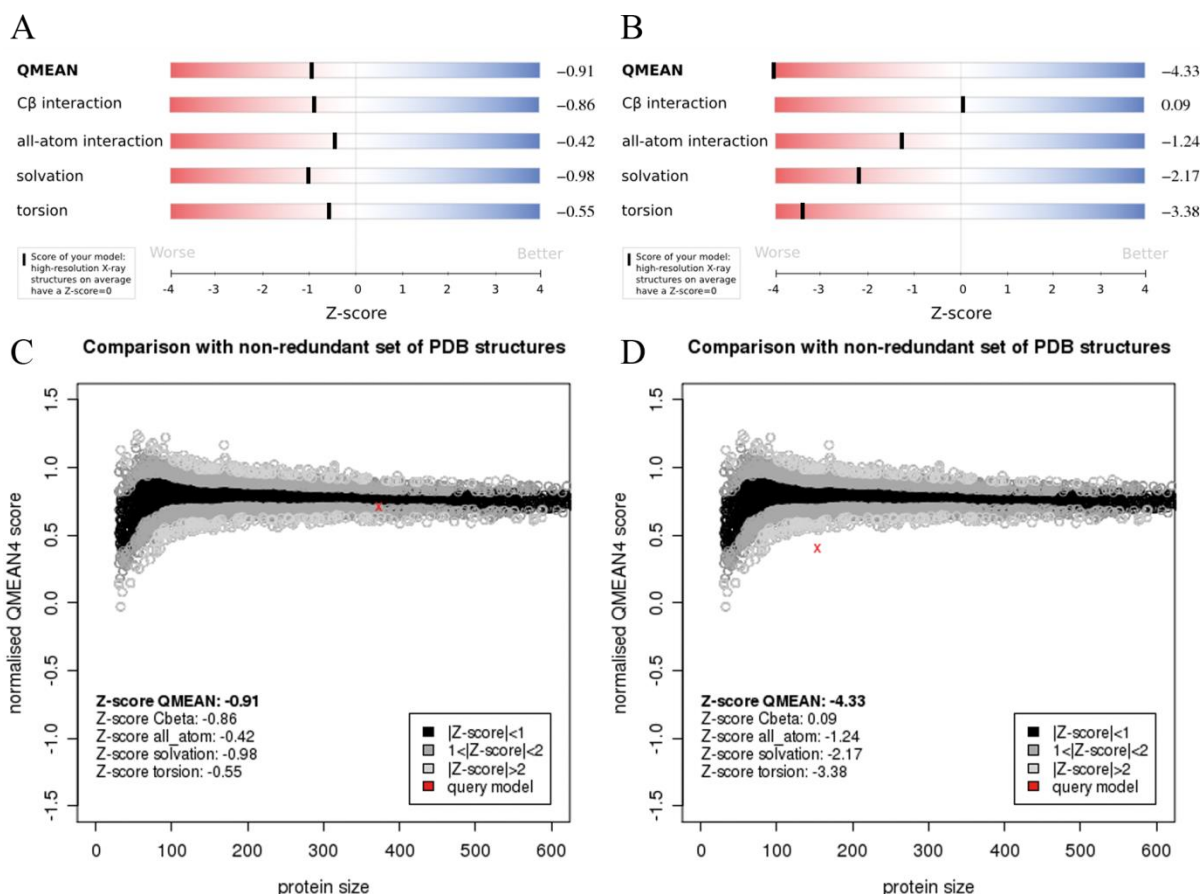


Figure VI-53: Comparison of “good” and “bad” homology models (IV.10.2).

Good homology models (A and C, actin 7, UniProtKB accession number Q6F4H4) exhibit values close to zero for all scoring functions (A) indicating an insignificant deviation from values obtained for high-quality structures (C, resolution below 2 Å). Bad homology models (B and D, 40S ribosomal protein S5, UniProtKB accession number O24111) exhibit values of $>|2|$ for one or more of the scoring functions (B) indicating significant deviations from values expected according to high resolution structures (D). Note that the value of all scoring functions are transformed to a standard normal distribution, hence a value of $>|2|$ indicates a score that is two-times the standard deviation away from the expected (zero) value and thus has a significant difference assuming a 5% significance level. The panels in this figure were generated as part of the homology modeling using SWISS-Model [198].

Two proteins lacked a suitable template and for 10 only partial models covering less than 30% of the sequence were retrieved when the SWISS-MODEL alignment algorithm was applied. Among the remaining proteins, 67 models achieved an acceptable coverage of more than 90% with 13 having a very high ($>99\%$), 28 having a high ($99\% > x > 95\%$) and 26 having a moderate ($95\% > x > 90\%$) coverage. The z-score of 38 of these models was above -2.00 indicating an insignificant difference in quality compared to high resolution X-ray structures (5% significance level). Only 12 models had a z-score below -4.00 indicating a poor model quality. Thus, without solving any structures *de novo*, it was possible to generate high-coverage homology models for more than 60% of the identified HCPs, more than 55% of which were high-quality models.

VI.10.2 QSAR model generation

QSAR models for different modes of chromatography can accurately predict the retention times of various model proteins based on their 3D structures [266-273]. The molecular descriptors included in these models resembled the mode of action of the corresponding chromatographic separations, e.g. descriptors representing the net charge, shape/charge-distribution and hydrogen bond capacity of a protein were selected in models for IEC [272]. Recently published chromatographic retention data [155, 266-274] were compiled into a single, non-redundant training set comprising 45 different model proteins (IX.6) for which descriptors were calculated in the pH range 4.0–8.0 (IV.10.3). This training set was used to generate QSAR models for protein retention on Capto MMC (MMC), Q Sepharose FF (AEX) and SP Sepharose FF (CEX) (IV.10.4). The descriptors selected for the model together with their statistics are shown in Table VI.11. A detailed explanation of the different selected descriptors is provided in Table VI.12 and Table VI.13. A graphical representation of the model quality is shown in Figure VI-54 depicting actual-versus-predicted as well as y-randomization plots.

All models correlated well with the actual data (r^2 and adj. $r^2 > 0.85$, Table VI.11) and thus appeared suitable to predict the elution behavior of tobacco HCPs described in the next paragraph (VI.10.3). However, the correlation was lower compared to previously reported QSAR models [266, 268, 273]. First, this was because the retention data from different sources were compiled to a single dataset, introducing 5–8% ‘batch-to-batch’ variability into the data. Furthermore, the different original datasets required transformation to a uniform response (elution salt concentration rather than %buffer XY or retention volume) resulting in a further degree of uncertainty. Second, a more thorough cross-evaluation was used because more sophisticated QSAR modeling software is now available (YAMS vs. ROMS, <http://reccr.chem.rpi.edu/Software/modeling/index.html>). This prevented the selection of over-fitted models with artificially high correlation to the actual data. Such over-fitted models can describe a given dataset well but their predictive value for compounds not included in the set is limited [213].

Descriptors related to protein charge and electrostatic potential (e.g. EP.H7 and PEOE_VSA_FPPOS) and hydrophobicity (e.g. MLP2.W5 and MLP1.W17) were chosen for the QSAR model of MMC during SVM selection (Table VI.11, explanations in Table VI.12 and Table VI.13). This agreed well with expectations for the mixed-mode character of this ligand as well as previous publications using similar descriptors [266]. The majority of descriptors were related to charge in the QSAR models for SP (e.g. RECON_FPIP7 and Fcharge) and Q (e.g. Fcharge and PEOE_VSA_FPPOS) as was also recently reported [273, 358].

A broad pH range was covered for SP, whereas the pH range of the MMC and Q models was limited due to the lack of data (Table VI.11) [266-273]. A limited pH applicability domain can be a drawback for practical applications because pH often has a critical impact on protein separation, especially for IEC [344, 349]. It is thus unlikely that optimal separation conditions can be predicted for models that only cover a limited set of pH values, e.g. pH 8.0 and 8.5 are not included in the current QSAR model for Q (VI.9.1.3).

A larger number of proteins will also improve the predictive power of the QSAR models because this will increase the likelihood that descriptors correlating with the retention mechanism can be identified.

Table VI.11: Descriptors selected during QSAR modeling of protein chromatography including model statistics (IV.9).

Model		MMC	Q	SP
Selected descriptors	1	EP.H7*	RECON_SIGMin	PEOE_VSA.4.1
	2	PEOE_VSA_FPPOS	FCharge	RECON_FPIP7
	3	RECON_FDel.K.NA2	RECON_SIKIA	Fcharge
	4	ANGLE.B04	PEOE_VSA_FPPOS	PEOE_VSA.3.1
	5	MLP2.W5*	RECON_FLapI6	RECON_PIPMin
	6	MLP2.AVGP*	RECON_SIEPMin	vsurf_DW13
	7	PEOE_VSA.3.1	FASA.	RECON_FPIP20
	8	MLP2.W24*	---	vsurf_IW3
	9	MLP2.W16*	---	RECON_FSIEPA2
	10	RECON_FDel.K.NA9	---	---
	11	HydroSurf_H2	---	---
	12	MLP1.W17*	---	---
Data points in training set		34	24	107
Applicable at pH [-]		6.0-7.0	7.0-7.5	5.0-8.0
Best model	r^2	0.9329	0.9555	0.8847
	Adj. r^2	0.9147	0.9434	0.8802
Best y-random.	Based on r^2	r^2	0.447	0.5123
		Adj. r^2	0.2172	0.4225
	Based on RMSE	r^2	0.4224	0.5123
		Adj. r^2	0.2955	0.4225

* Descriptors calculated using the PEST algorithm (Table VI.12).

Table VI.12: Descriptor types used in QSAR modeling of protein chromatography (Table VI.11).

Descriptor type	Explanation	Reference
EP	Electrostatic potential	[359]
Fcharge	Total charge of the molecule (sum of formal charges)	[360]
MLP	Molecular lipophilicity potential	[361, 362]
PEOE	Partial equalization of orbital electronegativities calculates atomic partial charge distribution between bonded atoms at equilibrium	[363]
PEST	Property-encoded surface translator uses a technique akin to ray-tracing to explore the volume enclosed by the solvent-accessible surface of a protein.	[265]
RECON	Rapid reconstruction of molecular charge densities and charge density-based electronic properties of molecules, using atomic charge density fragments	[265, 364]
vsurf	VolSurf is a computational procedure to produce 2D molecular descriptors from 3D molecular interaction energy grid maps	[365]

Table VI.13: Explanations to descriptors selected during QSAR modeling of protein chromatography (Table VI.11).

Code	Description
Angle	Molecular shape descriptor
DW	Ratio of hydrophobic/hydrophilic regions
ED.min	Lowest hydrophobic energy
F.DeI.K	Surface integral of the rate of change of the K electronic kinetic energy density normal to and away from the molecular surface
FASA*	Fractional CASA- calculated as CASA- / ASA.
FLap	Fingerprints for Ligands and Proteins
FPIP	Fractional Politzer Ionization Potential
FPPOS	Fraction of positive polar van der Waals surface area
HydroSurf H2	Surface Hydrophobicity using Hearn Scales 2
IW	Integy Moment
PIP_min	Local average ionization potential minimum
SIEP	Surface integral of electrostatic potential
SIG	Surface integral of G electronic kinetic energy density
SIKIA	Surface integral of the K kinetic energy density average
VSA	Van-der-Waals surface area

* CASA – charged accessible surface area.

The selection of additional proteins should aim to close gaps in the coverage of molecular masses and elution times in the training set, e.g. it will be helpful to include proteins eluting in the 500–1000 mM NaCl range in the MMC and SP models whereas the training set for Q should be expanded to encompass proteins eluting at >250 mM NaCl (Figure VI-54).

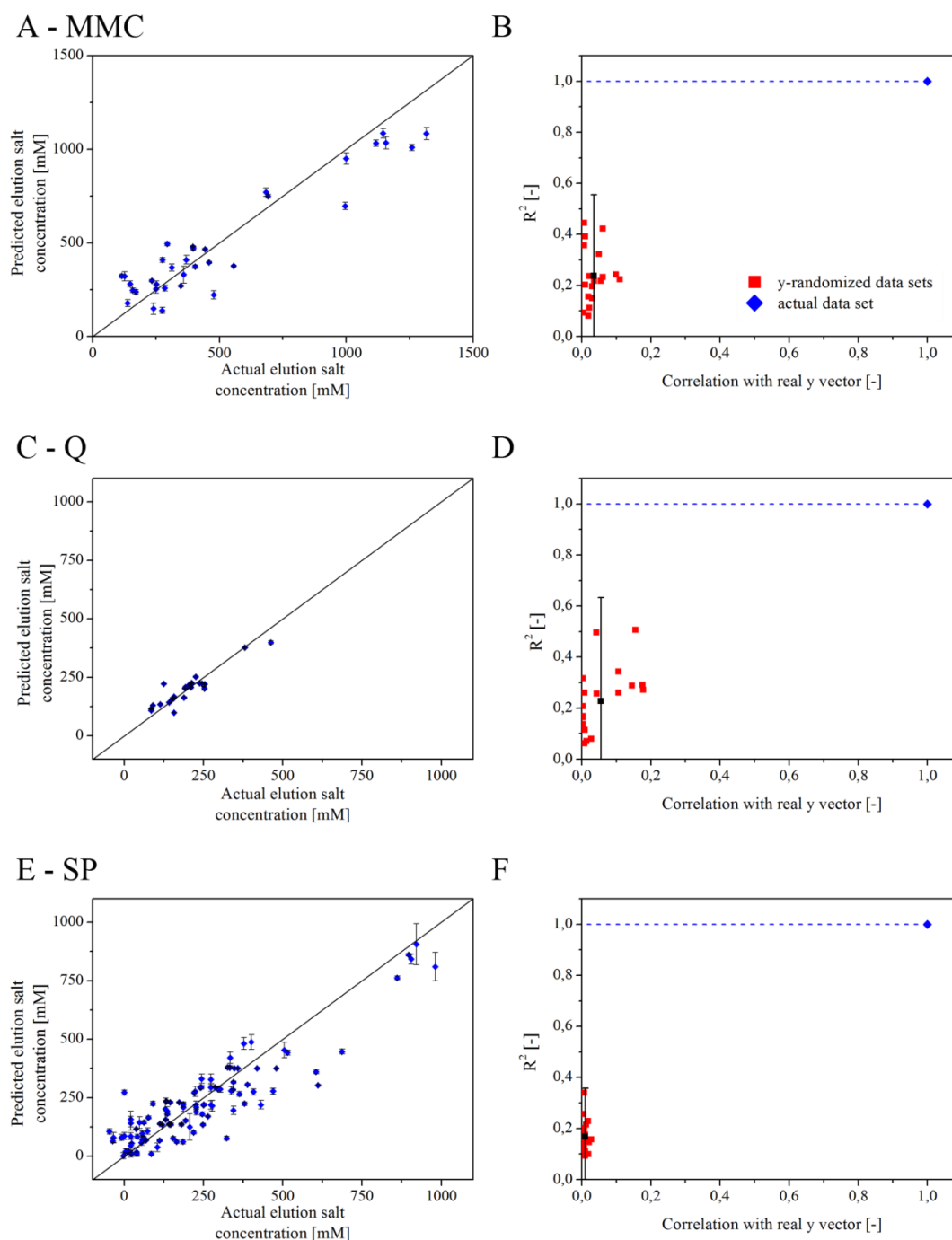


Figure VI-54: Graphical evaluation of QASR models for CptoMMC (A and B), Q Sepharose FF (C and D) and SP Sepharose FF (III.6.5.4 and IV.9).

A, C, D. Actual-versus-predicted plots of elution salt concentrations for proteins included in the training set. The predicted elution salt concentrations are in good agreement with the ideal model, which is indicated by the black diagonal. Error bars indicate prediction variability based on 10-fold cross-validation. B, D, E. Models based on y-randomized data sets (red squares) deviate significantly from the real models which are based on the actual data sets (blue diamonds) as the interval covering 99% (six times σ) of all random models (black bar) in the R^2 (coefficient of determination) space does not overlap with the actual model, indicating meaningful descriptor selection. This is a desired outcome because randomized data sets (red squares) should have (i) a low correlation to the real y vector and (ii) a low R^2 value. The black squares indicate the average correlation of random models to the respective real model.

An important factor that has not been addressed in previous QSAR chromatography models is protein oligomerization [191], which determines which parts of the protein surface are exposed to the surrounding aqueous medium (mobile phase) allowing interactions with the resin, thus indicating the binding strength and elution salt concentration [182]. RuBisCO can form oligomers under typical extraction and chromatography conditions, even if these oligomers are not biologically active (VI.9.1.3 and VI.9.6). Therefore, oligomerization is also likely to occur for model proteins used to generate QSAR training data sets (Figure VI-56) and if this phenomenon is not taken into account, the descriptors will be calculated based on improper 3D models (e.g. monomer instead of dimer) and the descriptor values will be compromised, especially surface charge or surface hydrophobicity descriptors. The selection of relevant descriptors will therefore be flawed and any physicochemical interpretations will be misleading. Figure VI-55 shows the impact of using different multimeric models for some tobacco HCPs that can form oligomers. With the exception of AOC, oligomerization increases the retention of proteins bound to SP at pH 5.0 or 6.0, and dimers increase the elution salt concentration by 2.5 ± 1.2 -fold ($n = 8$).

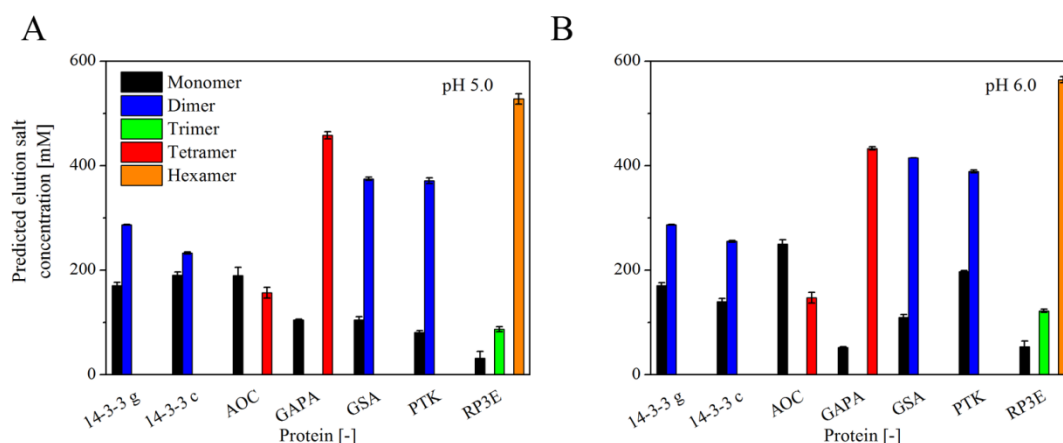


Figure VI-55: Impact of oligomeric state of tobacco HCPs on predicted elution salt concentration on SP Sepharose FF (IV.9).

An average 2.5-fold increase of the predicted elution salt concentration was observed when dimeric 3D-models were used to calculate protein descriptors at pH 5.0 (A) and 6.0 (B), only AOC was an exception. Other oligomerizations had a similar effect. 14-3-3 g - 14-3-3 g-1 protein; 14-3-3 c - 14-3-3-like protein C; AOC - Allene oxide cyclase; GAPA - Glyceraldehyde-3-phosphate dehydrogenase B, chloroplastic; GSA - Glutamate-1-semialdehyde 2,1-aminomutase, chloroplastic; PTK - Plastid transketolase; RP3E - Ribulose-phosphate 3-epimerase, chloroplastic.

In order to prevent oligomerization-based errors in future experiments, the elution salt concentrations for model proteins should be determined by including analytical techniques that allow oligomeric states to be characterized, e.g. SEC and/or MALLS [366-369]. Because these techniques were not used during the original experiments, it was not possible to correct for oligomerization of the model proteins in the QASR models described here.

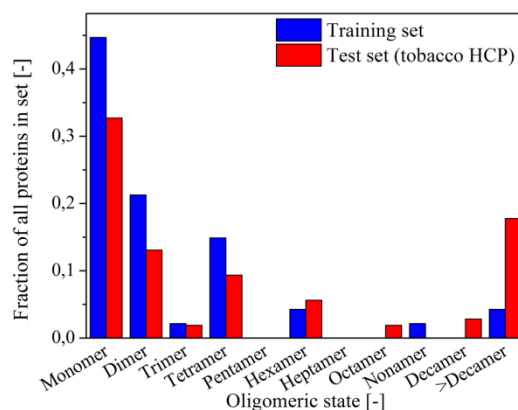


Figure VI-56: Comparison of oligomeric states of proteins in the training (IX.6) and tobacco HCP (IX.5) data sets.

Clearly, the training data set did not consist solely of monomeric proteins and thus, using only monomeric 3D-models of these proteins flawed the descriptor calculation and selection. Additionally, monomeric and dimeric proteins were overrepresented in the training set compared to the test set, reducing the predictive power of the resulting QSAR models.

In summary, the models described above are likely to predict elution salt concentrations more accurately than previous models for proteins not included in the training dataset because (i) they contain more data points than the previous models, increasing the likelihood that relevant descriptors will be identified, and (ii) over-fitting was avoided by applying rigorous cross-evaluation to prevent the selection of descriptors only relevant for the particular training dataset [217]. However, a broader pH range and the incorporation of oligomeric proteins is necessary to improve the model's performance even further.

VI.10.3 Generation of “synthetic” chromatograms

The models established above (VI.10.2) were used to predict the elution salt concentrations not only of model proteins as has been done before [273] but for all identified tobacco HCPs (IX.5) based on descriptors calculated according to their 3D structures. The predicted elution salt concentrations were used as the mean μ of a Gaussian-shaped elution function for each protein (Equation 44) although other functions can also be used [336]. Multiplication of this function with the individual protein amounts determined above (VI.10.1, IX.5) yielded elution profiles scaled to the actual protein abundance in tobacco extracts. Summing these elution profiles up for all proteins yielded “synthetic” chromatograms (Figure VI-57, a comparison with experimental runs is given in section VI.10.4). A global σ parameter was used to describe the non-ideal elution behavior of the proteins causing the broadening of elution peaks. Values in the range 10–25 mM were assessed with smaller σ -values corresponding to sharper peaks and thus higher resolution. A value of 20–25 mM resulted in chromatograms closely resembling

experimental runs (VI.10.4). Proteins with elution salt concentrations of <50 mM NaCl were pooled as a single FT peak, with a retention volume of 3.5 mL and a σ value of 1 mL.

$$c_i = \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma}\right)^2}$$

Equation 44: Gaussian function used to calculate protein elution peaks with c_i as the relative elution peak height for protein i , x the actual salt concentration, μ_i the predicted elution salt concentration for protein i and σ the global peak broadening parameter.

VI.10.4 Comparison with experimental separation

A good agreement was achieved between the predicted and observed HCP separation for SP Sepharose FF (Figure VI-57 A). The elution salt concentrations predicted for peaks 1 and 2 were 24 and 37 mM higher than the observed values. The second predicted peak featured a shoulder due to the assumed separate elution of the RuBisCO large and small subunits. However, as discussed above, this model assumption was incorrect because the RuBisCO subunits formed oligomers under the extraction and chromatography conditions (VI.9.6). Attempts to incorporate a realistic oligomeric RuBisCO state into the model were hampered for two reasons: (i) it was not possible to calculate descriptors for the biological assembly (8:8 hetero hexadecamer) or parts thereof (4:4 or 2:2) because the 3D-model contained too many atoms for the algorithm to handle, and (ii) RuBisCO was not present in its assumed biological assembly (it passed through a 300 kDa membrane, whereas the size of the biological assembly is ~560 kDa and should have been retained). The precise stoichiometry of the subunits was not known. Therefore, determining the oligomeric state of RuBisCO and incorporating an adequate 3D structure (including descriptors) into the model is likely to resolve the difference between predicted and observed retention volumes.

A larger discrepancy was observed for the predicted and actual elution of peak 3 (405 and 570 mM NaCl, respectively). Oligomerization could also explain this discrepancy because many ribosomal proteins are predicted to elute at salt concentrations of ~400 mM, but this may increase if ribosomes or their subunits are intact under the chosen chromatography conditions, especially taking into account that ribosomes have evolved to bind RNA, a negatively charged ‘ligand’ such as SP. Other abundant proteins such as RP3E also underwent oligomerization and shifted towards higher salt concentrations for elution (Figure VI-55).

The reduced area of peak 3 can be explained by the missing proteins that accounted for 34% of TSP in the SP model. These proteins were not part of the model either because they were not present in the protein bands used for HCP identification (VI.10.1) or it was not possible to calculate all their descriptors.

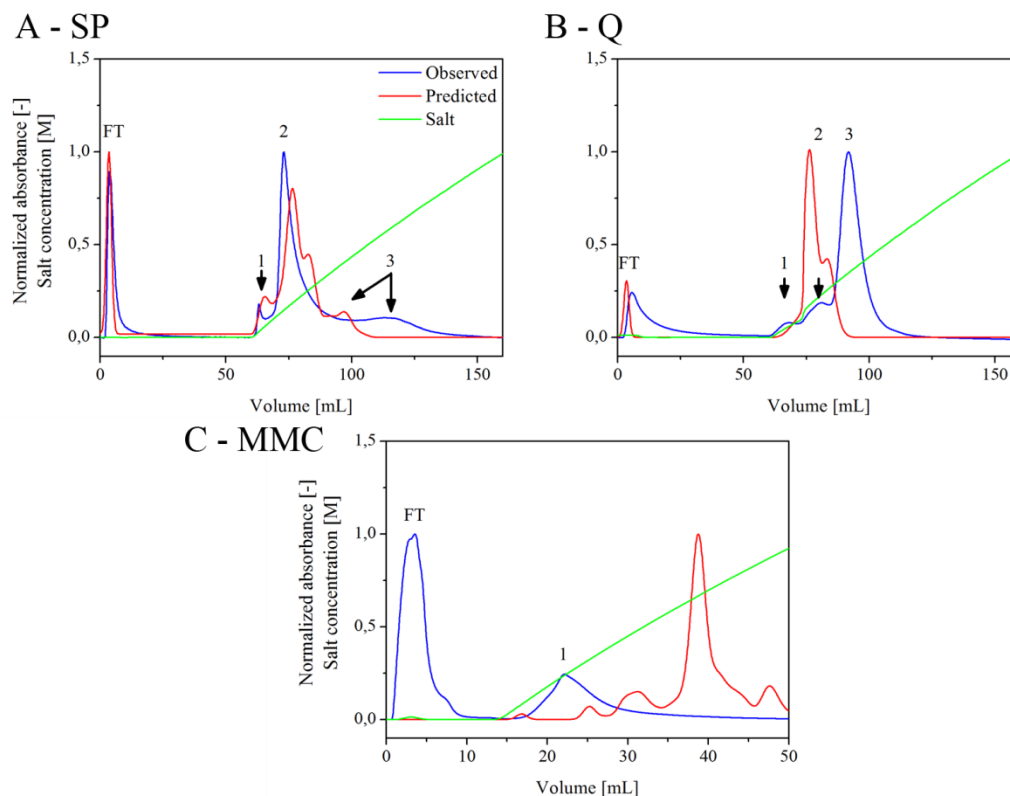


Figure VI-57: Comparison of observed and predicted tobacco HCP separation on different chromatographic resins (IV.9, IV.11 and VI.10).

A. Observed and predicted HCP separation were in good agreement for SP resin and showed both three elution peaks. B. For Q, one major elution peak was predicted and observed, however, the observed peak eluted at a higher salt concentration than predicted and two additional smaller peaks appeared. C. Predictions for MMC were poor as only a major FT peak and a smaller elution peak instead of four predicted elution peaks were observed. See text for detailed discussion. Buffer pH was 6.0 for SP and 7.0 for Q and MMC.

The above limitations also applied to the QSAR model for Q Sepharose FF and can explain the deviation of 172 mM NaCl between the predicted and observed elution of RuBisCO (highest peaks in Figure VI-57 B). Interestingly, two additional peaks (peak 1 and 2) were eluted before the major RuBisCO elution peak (peak 3) and these contained the large RuBisCO subunit as their major protein component. Hence peaks 1 and 2 may represent different oligomeric states, e.g. a dimer in peak 1 and a monomer in peak 2. Although ~65% of HCPs were included in the model, the limited number of model proteins (Table VI.11) reduced the precision of predictions achieved with the Q model compared to the SP model.

More model proteins were included for descriptor selection over a broader pH range for MMC, but it was not possible to calculate all of the additional descriptors required to model the hydrophobic interactions for the RuBisCO small subunit so only 56% of TSP was covered by the model. The high FT peak in the actual run probably reflected the limited resin capacity rather than an incorrect model prediction (Figure VI-57 C). However, these data did not entirely explain the poor correlation between the MMC model and the observed HCP separation.

Instead, the selected pH of 7.0 may explain the early elution of RuBisCO (peak 1) at ~240 mM NaCl compared to the predicted elution at ~670 mM. This is because electrostatic interactions dominate protein binding under low-salt loading conditions (VI.9.4.2) [341] and are affected by pH, but the MMC model was missing e.g. the 'FCharge' descriptor depicting the total protein charge (Table VI.11). In this context, the descriptors of protein hydrophobicity used here may have been inadequate even though they have been used successfully to model protein retention during HIC [265, 269]. Mixed-mode chromatography is not a linear combination of hydrophobic and electrostatic interactions but is dependent on the relative position of charged and hydrophobic patches on a protein surface, facilitating binding to the sterically-defined ligand. Therefore, other descriptors of protein hydrophobicity [370, 371] or descriptors featuring a combination of hydrophobic and electrostatic interactions including the steric constraints of the ligand (Siddarth Parimal, personal communication) may be advantageous. The oligomerization of proteins (VI.10.2) can also have a critical effect on the predictive power of protein hydrophobicity descriptors because protein oligomerization is often driven by hydrophobic surface patches which are not available for protein ligand interactions after oligomerization [372]. The extent of protein retention on media exhibiting hydrophobic interaction characteristics can therefore be overestimated.

In summary, for the first time QSAR models of chromatographic separation have been successfully applied to the crude protein mixture of plant extracts. In this context the accuracy of the predictions for SP are remarkable given the fact that protein oligomerization and competitive binding were not accounted for in the model. In general, the predictive power of models increased with the number of reference proteins in the training set and with decreasing complexity of the interaction mechanism (plain IEX vs. MMC) (Table VI.11). The r^2 and adjusted r^2 values did not show a good correlation with the quality of the model predictions. Predicted and observed elution salt concentrations matched well for SP Sepharose FF, indicating that QSAR modeling can be applied successfully to the separation even of crude protein mixtures which has not been tested before. Therefore, QSAR modeling can be used to guide and accelerate process development for expression platforms with known HCPs. This can be achieved with minimal experimental effort, simply by exploiting current protein databases and modeling tools [198, 219, 256, 360]. The limited size of the training data set diminished the accuracy of predictions for Q Sepharose FF, but additional reference data are likely to compensate for this shortcoming. Additional changes will be required to improve the MMC model as discussed in more detail in section VI.10.5.

VI.10.5 Future improvements of prediction quality

Several changes can be incorporated in the method presented above to improve the predictive power of the “synthetic” chromatography model. First, additional proteins, if possible including authentic HCPs, can be included in the training data set used to generate the QSAR models. This will help to overcome the bias in the current training set towards small, monomeric proteins with IEC elution salt concentrations typically below 400 mM (Figure VI-54). As discussed above (III.6.5.4) such a bias in the training data narrows down the applicability domain of a QSAR model [213]. The same holds true for the pH range of the current models.

Second, the quality of homology models for HCPs can be increased as (i) more X-ray and NMR structures are added to the structure databases and thus become available for modeling, (ii) the resolution of the available structures increases, and (iii) the sequence coverage of the models becomes complete.

Third, more HCPs can be included in the predictions as the 108 proteins currently present in the model only account for 56–66% of the TSPs (depending on the QSAR model). More HCPs in the model will help to complete the elution profile. Fourth, the accuracy of the peak profile can be improved by using individually-assigned protein amounts (e.g. by 2D-gel analysis) rather than averaged amounts based on band intensity and the number of proteins per band. Fifth, protein-specific instead of global σ parameters will improve the description of the elution behavior and resin efficacy (Figure III-5). This concept can be extended to asymmetric elution peaks to model “tailing” during elution. Sixth, the impact of competitive binding/elution of more than one model protein species should be investigated as this may affect the binding orientation of bound proteins and thus the strength of their interaction with the resin and the resulting elution salt concentration/conditions [346, 373-376].

Finally, protein models can be used that elaborate to the biologically-active state of the different proteins (dimers, tetramers etc.), rather than simply using a monomeric form. This is important because protein association can affect accessible surface areas and thus ligand binding sites [377, 378].

Additionally, it will be most useful to include the individual target proteins in the QSAR models because this will allow a direct analysis of beneficial separation conditions that yield highly pure product. Moreover, potential impurities co-eluting with the target in a first chromatographic step can be identified at an early stage of process development allowing a rational selection of subsequent separation steps. In the course of this thesis incorporation of target proteins in the models was hampered by three reasons that also pointed out the limits of the QSAR method. First, no adequate structure template for E25T, a potential Malaria vaccine

candidate (provided by Alexander Boes), was found in the SWISS-Model database preventing any homology modeling attempt. Second, it was not possible to perform energy minimization calculations for the 3D homology structure of the monoclonal antibody 2G12 due the size of the protein and resulting software errors. Third, descriptor calculations were incomplete for the biological assembly of DsRed (tetramer) and thus prevented reliable predictions for this protein even though a refined 3D structure was available.

VII. Conclusion and scope

VII.1 Process design and control for plant-derived biopharmaceuticals

The use of suitable bag and depth filters was investigated for the clarification of plant extracts containing biopharmaceutical proteins, with and without flocculants and other additives (VI.2, VI.3, VI.5 and VI.6). The direct consumable costs for depth filters was cut by 50% solely by identifying a more suitable filter combination, reducing the number of depth filtration steps from three to one. Flocculants and cellulose-based additives incorporated during depth filtration again reduced the costs for filters by 50% each, leaving only ~10% of the initial costs for depth filters. The production of biopharmaceuticals in plants will benefit from these changes in process design due to the lower production costs, reduced contamination risks (because of the simplified process stream) and shorter processing time, thus improving the product quality and reducing the risk of target protein degradation/denaturation. Further improvements will be possible by optimizing the extraction buffer pH and other conditioning steps dependent on the target protein such as heat treatment and/or membrane filtration (VI.10.5) [75, 350]. Certificates allowing contact with food products were available for all substances, facilitating their use in GMP-compliant processes.

The flocculants and additives were also compatible with subsequent chromatography steps based on resins and buffer conditions tested for the effective removal of major tobacco proteins such as RuBisCO (VI.9) under a variety of process conditions. This helped to identify generic process steps suitable for any target protein, which will reduce the development time for new processes. QSAR modeling of chromatography can also help to reduce development times (VI.10). These models can be applied successfully to crude plant extracts given an adequate amount of training data, known HCPs and high-quality descriptors. With the growing number of protein structures in the databases and the increasing sophistication of modeling tools that are freely available via the internet, the characterization and modeling of HCPs for individual expression platforms will become a routine procedure for the acceleration of process development [198, 200, 257, 258, 379].

VII.2 The basis for a Quality-by-Design approach

The methodology described in this thesis is as important as the data. Previous studies have often focused on single factors affecting the outcome of experiments, or neglected factor interactions if more than one parameter was studied [75, 380, 381]. Therefore, valuable information was not recorded and resources were used inefficiently. In contrast, the DoE approach used throughout

this thesis allowed the systematic and cost-effective analysis of design spaces, resulting in a more detailed understanding of the process steps under investigation. The DoE approach also allowed the development of predictive models forming the basis of a QbD approach by offering (i) a quantitative correlation between process parameters and the quality of certain process steps, and (ii) a benchmark to define the working space in future processes [382-384]. The QbD approach seeks to determine process conditions that will result in reproducible product yields, purity and efficacy, thus producing safe, high-quality drugs, in contrast to the quality by control approach in which quality is ensured by extensive testing of the final product [383, 385].

The models presented above also revealed which process parameters (factors) influenced the outcome/quality/performance of a specific process step and indicated their individual leverage. This will facilitate the setup of a failure mode and effects analysis (FMEA) including the assignment of severity, occurrence and detection probability and ultimately a risk priority number (RPN) to each relevant process parameter [382, 383, 386].

The use of advanced PAT will help to close gaps between current models for the plant-based production of biopharmaceutical proteins as described here, resulting in a holistic description of the process. For example, it has not been possible thus far to link depth filter capacity to process parameters such as extract turbidity, probably because turbidity is a sum parameter of particle size and number. This limitation could be overcome by incorporating new inline devices that can determine particle size distribution [387-389], allowing a more accurate calculation of the required filter area and hence the filtration time, and maybe also selection of the most suitable depth filter type ‘on-the-fly’ during production.

The ultimate goal should be a model describing the complete production and purification process by a (small) number of critical process parameters, linking process conditions to the critical quality attributes of the product and taking into account the specific properties of the product, e.g. pH stability. Such a global model can be achieved by combining the models described here with those established for different upstream operations [247, 278, 331, 390]. The model can then be used in a feedback setup where the impact of process parameters on product quality is determined in a first step, and then product quality data can be used to define the operation range for process parameters that result in the reliable production of a high-quality target protein. Such a fundamental understanding of a new process can help to convince regulatory authorities of the safety of plant-derived biopharmaceuticals and will elevate this technology to the level of established production platforms such as CHO cells.

VIII. References

1. Gagneux, S., *Host-pathogen coevolution in human tuberculosis*. Philosophical Transactions of the Royal Society B-Biological Sciences, 2012. **367**(1590): p. 850-859.
2. Dawkins, R. and J.R. Krebs, *Arms Races between and within Species*. Proceedings of the Royal Society of London Series B-Biological Sciences, 1979. **205**(1161): p. 489-511.
3. Borchardt, J.K., *The beginnings of, drug therapy: Ancient Mesopotamian medicine*. Drug News & Perspectives, 2002. **15**(3): p. 187-192.
4. Licciardi, P.V. and J.R. Underwood, *Plant-derived medicines: A novel class of immunological adjuvants*. International Immunopharmacology, 2011. **11**(3): p. 390-398.
5. Raskin, I., et al., *Plants and human health in the twenty-first century*. Trends in Biotechnology, 2002. **20**(12): p. 522-531.
6. Bentley, R., *Different roads to discovery; Prontosil (hence sulfa drugs) and penicillin (hence beta-lactams)*. Journal of Industrial Microbiology & Biotechnology, 2009. **36**(6): p. 775-786.
7. Levesque, H. and O. Lafont, *[Aspirin throughout the ages: a historical review]*. La Revue de medecine interne, 2000. **21 Suppl 1**: p. 8s-17s.
8. Müller-Jahncke, W.D., C. Friedrich, and U. Meyer, *Arzneimittelgeschichte*. 2005: WVG, Wissenschaftliche Verlagsgesellschaft.
9. Wachtel-Galor, S. and I.F.F. Benzie, *Herbal Medicine: An Introduction to Its History, Usage, Regulation, Current Trends, and Research Needs*, in *Herbal Medicine: Biomolecular and Clinical Aspects*, I.F.F. Benzie and S. Wachtel-Galor, Editors. 2011, Llc.: Boca Raton FL.
10. Macnalty, A.S., *Emil von Behring, born March 15, 1854*. British medical journal, 1954. **1**(4863): p. 668-70.
11. Raju, T.N.K., *Emil Adolf von Behring and serum therapy for diphtheria*. Acta Paediatrica, 2006. **95**(3): p. 258-259.
12. Fleming, A., *On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae*. Bulletin of the World Health Organization, 2001. **79**(8): p. 780-790.
13. Exposito, O., et al., *Biotechnological Production of Taxol and Related Taxoids: Current State and Prospects*. Anti-Cancer Agents in Medicinal Chemistry, 2009. **9**(1): p. 109-121.
14. Roche, *Roche Position on Biotechnology – Safety, Health and Environmental Aspects*. Corporate Sustainability Committee, 2011.
15. Chassin, C. and P. Pollak, *Outlook for chemical and biochemical manufacturing*. Pharmaceuticals / Biopharmaceuticals, 2004(January/February): p. 23-26.
16. Šmelcerović, A. and S. Šmelcerović, *The Importance and Advantages of the Biotechnological Methods for the Production of Antibiotics and for the Discovery of new Structures*. Facta Universitatis, 2004. **1**,(No 5): p. 109-112.
17. Stottmeister, U., et al., *White biotechnology for Green chemistry: fermentative 2-oxocarboxylic acids as novel building blocks for subsequent chemical syntheses*. Journal of Industrial Microbiology & Biotechnology, 2005. **32**(11-12): p. 651-664.
18. Willke, T. and K.D. Vorlop, *Industrial bioconversion of renewable resources as an alternative to conventional chemistry*. Applied microbiology and biotechnology, 2004. **66**(2): p. 131-142.

19. Otto, C., V. Yovkova, and G. Barth, *Overproduction and secretion of alpha-ketoglutaric acid by microorganisms*. Applied microbiology and biotechnology, 2011. **92**(4): p. 689-95.
20. Wilke, D., *Chemicals from biotechnology: molecular plant genetics will challenge the chemical and the fermentation industry*. Applied microbiology and biotechnology, 1999. **52**(2): p. 135-145.
21. Hartmann, T., *From waste products to ecochemicals: Fifty years research of plant secondary metabolism*. Phytochemistry, 2007. **68**(22-24): p. 2831-2846.
22. Wilson, S.A. and S.C. Roberts, *Recent advances towards development and commercialization of plant cell culture processes for the synthesis of biomolecules*. Plant Biotechnology Journal, 2012. **10**(3): p. 249-268.
23. Lipinski, C.A., et al., *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings*. Advanced Drug Delivery Reviews, 1997. **23**(1-3): p. 3-25.
24. Smith, G.R., M.J.E. Sternberg, and P.A. Bates, *The relationship between the flexibility of proteins and their conformational states on forming protein-protein complexes with an application to protein-protein docking*. Journal of Molecular Biology, 2005. **347**(5): p. 1077-1101.
25. Anderson, D.E., W.J. Becktel, and F.W. Dahlquist, *Ph-Induced Denaturation of Proteins - a Single Salt Bridge Contributes 3-5 Kcal Mol to the Free-Energy of Folding of T4-Lysozyme*. Biochemistry, 1990. **29**(9): p. 2403-2408.
26. Jana, S. and J.K. Deb, *Strategies for efficient production of heterologous proteins in Escherichia coli*. Applied microbiology and biotechnology, 2005. **67**(3): p. 289-298.
27. Schmidt, F.R., *Recombinant expression systems in the pharmaceutical industry*. Applied microbiology and biotechnology, 2004. **65**(4): p. 363-372.
28. Ma, J.K.C., P.M.W. Drake, and P. Christou, *The production of recombinant pharmaceutical proteins in plants*. Nature Reviews Genetics, 2003. **4**(10): p. 794-805.
29. Walsh, G., *Post-translational modifications of protein biopharmaceuticals*. Drug Discovery Today, 2010. **15**(17-18): p. 773-780.
30. Chusainow, J., et al., *A Study of Monoclonal Antibody-Producing CHO Cell Lines: What Makes a Stable High Producer?* Biotechnology and Bioengineering, 2009. **102**(4): p. 1182-1196.
31. Costa, A.R., et al., *Guidelines to cell engineering for monoclonal antibody production*. European Journal of Pharmaceutics and Biopharmaceutics, 2010. **74**(2): p. 127-138.
32. Scolnik, P.A., *mAbs A business perspective*. MAbs, 2009. **1**(2): p. 179-184.
33. Reichert, J. and A. Pavlou, *Monoclonal antibodies market*. Nature Reviews: Drug Discovery, 2004. **3**(5): p. 383-384.
34. Fischer, R. and N. Emans, *Molecular farming of pharmaceutical proteins*. Transgenic Research, 2000. **9**(4-5): p. 279-99.
35. Rodrigues, M.E., et al., *Technological Progresses in Monoclonal Antibody Production Systems*. Biotechnology Progress, 2010. **26**(2): p. 332-351.
36. Herrera-Estrella, L., et al., *Expression of chimaeric genes transferred into plant cells using a Ti-plasmid-derived vector*. Nature, 1983. **303**(5914): p. 209-213.
37. Twyman, R.M., et al., *Molecular farming in plants: host systems and expression technology*. Trends in Biotechnology, 2003. **21**(12): p. 570-578.
38. Fischer, R., et al., *Towards molecular farming in the future: transient protein expression in plants*. Biotechnology and Applied Biochemistry, 1999. **30**: p. 113-116.
39. Tzfira, T. and V. Citovsky, *Agrobacterium-mediated genetic transformation of plants: biology and biotechnology*. Current opinion in biotechnology, 2006. **17**(2): p. 147-54.

40. Sheludko, Y.V., et al., *Comparison of several Nicotiana species as hosts for high-scale Agrobacterium-mediated transient expression*. Biotechnology and Bioengineering, 2007. **96**(3): p. 608-14.
41. Pogue, G.P., et al., *Production of pharmaceutical-grade recombinant aprotinin and a monoclonal antibody product using plant-based transient expression systems*. Plant Biotechnology Journal, 2010. **8**(5): p. 638-54.
42. Sheludko, Y.V., *Agrobacterium-mediated transient expression as an approach to production of recombinant proteins in plants*. Recent Patents on Biotechnology, 2008. **2**(3): p. 198-208.
43. Gelvin, S.B., *Agrobacterium-mediated plant transformation: the biology behind the "gene-jockeying" tool*. Microbiology and Molecular Biology Reviews, 2003. **67**(1): p. 16-37, table of contents.
44. Vezina, L.P., et al., *Transient co-expression for fast and high-yield production of antibodies with human-like N-glycans in plants*. Plant Biotechnology Journal, 2009. **7**(5): p. 442-55.
45. Wydro, M., E. Kozubek, and P. Lehmann, *Optimization of transient Agrobacterium-mediated gene expression system in leaves of Nicotiana benthamiana*. Acta Biochimica Polonica, 2006. **53**(2): p. 289-98.
46. Maclean, J., et al., *Optimization of human papillomavirus type 16 (HPV-16) L1 expression in plants: comparison of the suitability of different HPV-16 L1 gene variants and different cell-compartment localization*. Journal of General Virology, 2007. **88**(Pt 5): p. 1460-9.
47. Conley, A.J., et al., *Recombinant protein production in a variety of Nicotiana hosts: a comparative analysis*. Plant Biotechnology Journal, 2011. **9**(4): p. 434-44.
48. Shoji, Y., et al., *A plant-based system for rapid production of influenza vaccine antigens*. Influenza and Other Respiratory Viruses, 2012. **6**(3): p. 204-210.
49. Regnard, G.L., et al., *High level protein expression in plants through the use of a novel autonomously replicating geminivirus shuttle vector*. Plant Biotechnology Journal, 2010. **8**(1): p. 38-46.
50. Gelvin, S.B., *Improving plant genetic engineering by manipulating the host*. Trends in Biotechnology, 2003. **21**(3): p. 95-8.
51. Penna, S. and T.R. Ganapathi, *Engineering the plant genome: prospects of selection systems using non-antibiotic marker genes*. GM Crops, 2010. **1**(3): p. 128-36.
52. Miki, B. and S. McHugh, *Selectable marker genes in transgenic plants: applications, alternatives and biosafety*. Journal of Biotechnology, 2004. **107**(3): p. 193-232.
53. Sperb, F., et al., *Molecular Cloning and Transgenic Expression of a Synthetic Human Erythropoietin Gene in Tobacco*. Applied Biochemistry and Biotechnology, 2011. **165**(2): p. 652-665.
54. Pawlowski, W.P. and D.A. Somers, *Transgene inheritance in plants genetically engineered by microprojectile bombardment*. Molecular Biotechnology, 1996. **6**(1): p. 17-30.
55. Fischer, R., et al., *GMP issues for recombinant plant-derived pharmaceutical proteins*. Biotechnology Advances, 2012. **30**(2): p. 434-9.
56. Caplan, A., et al., *Introduction of genetic material into plant cells*. Science, 1983. **222**(4625): p. 815-21.
57. Weising, K., J. Schell, and G. Kahl, *Foreign Genes in Plants - Transfer, Structure, Expression, and Applications*. Annual Review of Genetics, 1988. **22**: p. 421-477.
58. Hiatt, A., R. Cafferkey, and K. Bowdish, *Production of antibodies in transgenic plants*. Nature, 1989. **342**(6245): p. 76-8.

59. Schillberg, S., R.M. Twyman, and R. Fischer, *Opportunities for recombinant antigen and antibody expression in transgenic plants--technology assessment*. *Vaccine*, 2005. **23**(15): p. 1764-9.
60. Ma, J.K., et al., *Plant-derived pharmaceuticals--the road forward*. *Trends in plant science*, 2005. **10**(12): p. 580-5.
61. Khan, I., et al., *Using storage organelles for the accumulation and encapsulation of recombinant proteins*. *Biotechnology Journal*, 2012. **7**(9): p. 1099-108.
62. Tremblay, R., et al., *Tobacco, a highly efficient green bioreactor for production of therapeutic proteins*. *Biotechnology Advances*, 2010. **28**(2): p. 214-21.
63. Wilken, L.R. and Z.L. Nikolov, *Recovery and purification of plant-made recombinant proteins*. *Biotechnology Advances*, 2012. **30**(2): p. 419-33.
64. Commandeur, U., R.M. Twyman, and R. Fischer, *The biosafety of molecular farming in plants*. *AgBiotechNet*, 2003. **5**: p. 1-9.
65. Warzecha, H., *Biopharmaceuticals from plants: a multitude of options for posttranslational modifications*. *Biotechnology & Genetic Engineering Reviews*, 2008. **25**: p. 315-30.
66. Giritch, A., et al., *Rapid high-yield expression of full-size IgG antibodies in plants coinfecting with noncompeting viral vectors*. *Proceedings of the National Academy of Sciences, USA*, 2006. **103**(40): p. 14701-6.
67. Obembe, O.O., et al., *Advances in plant molecular farming*. *Biotechnology Advances*, 2011. **29**(2): p. 210-222.
68. Ahmad, A., et al., *Green Biofactories: Recombinant Protein Production in Plants*. *Recent patents on biotechnology*, 2010.
69. Bakker, H., et al., *An antibody produced in tobacco expressing a hybrid beta-1,4-galactosyltransferase is essentially devoid of plant carbohydrate epitopes*. *Proceedings of the National Academy of Sciences, USA*, 2006. **103**(20): p. 7577-7582.
70. Nikolov, Z.L. and S.L. Woodard, *Downstream processing of recombinant proteins from transgenic feedstock*. *Current Opinion in Biotechnology*, 2004. **15**(5): p. 479-86.
71. Goldstein, D.A. and J.A. Thomas, *Biopharmaceuticals derived from genetically modified plants*. *QJM*, 2004. **97**(11): p. 705-16.
72. Nichols, M.E., et al., *Disruption of leaves and initial extraction of wildtype CPMV virus as a basis for producing vaccines from plants*. *Journal of Biotechnology*, 2002. **92**(3): p. 229-35.
73. Menkhaus, T.J., et al., *Considerations for the recovery of recombinant proteins from plants*. *Biotechnology Progress*, 2004. **20**(4): p. 1001-14.
74. Wang, W., F. Tai, and S. Chen, *Optimizing protein extraction from plant tissues for enhanced proteomics analysis*. *Journal of Separation Science*, 2008. **31**(11): p. 2032-9.
75. Hassan, S., et al., *Considerations for extraction of monoclonal antibodies targeted to different subcellular compartments in transgenic tobacco plants*. *Plant Biotechnology Journal*, 2008. **6**(7): p. 733-48.
76. Holler, C. and C. Zhang, *Purification of an acidic recombinant protein from transgenic tobacco*. *Biotechnology and Bioengineering*, 2008. **99**(4): p. 902-9.
77. Fischer, R., et al., *Plant-based production of biopharmaceuticals*. *Current Opinion in Plant Biology*, 2004. **7**(2): p. 152-8.
78. Holler, C., D. Vaughan, and C.M. Zhang, *Polyethyleneimine precipitation versus anion exchange chromatography in fractionating recombinant beta-glucuronidase from transgenic tobacco extract*. *Journal of Chromatography A*, 2007. **1142**(1): p. 98-105.
79. Gregory, J. and S. Barany, *Adsorption and flocculation by polymers and polymer mixtures*. *Advances in Colloid and Interface Science*, 2011. **169**(1): p. 1-12.
80. Zhou, Y. and G.V. Franks, *Flocculation mechanism induced by cationic polymers investigated by light scattering*. *Langmuir*, 2006. **22**(16): p. 6775-6786.

81. Barany, S. and A. Szepesszentgyorgyi, *Flocculation of cellular suspensions by polyelectrolytes*. *Advances in Colloid and Interface Science*, 2004. **111**(1-2): p. 117-129.
82. Roush, D.J. and Y.F. Lu, *Advances in primary recovery: Centrifugation and membrane technology*. *Biotechnology Progress*, 2008. **24**(3): p. 488-495.
83. Russel, W.B., D.A. Saville, and W.R. Schowalter, *Colloidal Dispersions*. 1989, Cambridge, UK: Cambridge University Press.
84. Runkana, V., P. Somasundaran, and P.C. Kapur, *Mathematical modeling of polymer-induced flocculation by charge neutralization*. *Journal of Colloid and Interface Science*, 2004. **270**(2): p. 347-358.
85. Gregory, J., *Polymer adsorption and flocculation in sheared suspensions*. *Colloids and Surfaces*, 1988. **31**(0): p. 231-253.
86. Pelssers, E.G.M., M.A.C. Stuart, and G.J. Fleer, *Kinetics of bridging flocculation. Role of relaxations in the polymer layer*. *Journal of the Chemical Society, Faraday Transactions*, 1990. **86**(9): p. 1355-1361.
87. Pelssers, E.G.M., M.A.C. Stuart, and G.J. Fleer, *Kinetic Aspects of Polymer Bridging - Equilibrium Flocculation and Nonequilibrium Flocculation*. *Colloids and Surfaces*, 1989. **38**(1-3): p. 15-25.
88. Pearson, C.R., et al., *Zeta potential as a measure of polyelectrolyte flocculation and the effect of polymer dosing conditions on cell removal from fermentation broth*. *Biotechnology and Bioengineering*, 2004. **87**(1): p. 54-60.
89. Fisher, R.R. and C.E. Glatz, *Polyelectrolyte precipitation of proteins: II. Models of the particle size distributions*. *Biotechnol Bioeng*, 1988. **32**(6): p. 786-96.
90. Tang, P., J. Greenwood, and J.A. Raper, *A Model to Describe the Settling Behavior of Fractal Aggregates*. *Journal of Colloid and Interface Science*, 2002. **247**(1): p. 210-219.
91. Platis, D. and N.E. Labrou, *Development of an aqueous two-phase partitioning system for fractionating therapeutic proteins from tobacco extract*. *Journal of Chromatography A*, 2006. **1128**(1-2): p. 114-124.
92. Schugerl, K. and J. Hubbuch, *Integrated bioprocesses*. *Current Opinion in Microbiology*, 2005. **8**(3): p. 294-300.
93. Asenjo, J.A. and B.A. Andrews, *Aqueous two-phase systems for protein separation: Phase separation and applications*. *Journal of Chromatography A*, 2012. **1238**: p. 1-10.
94. Oelmeier, S.A., F. Dismer, and J. Hubbuch, *Application of an Aqueous Two-Phase Systems High-Throughput Screening Method to Evaluate mAb HCP Separation*. *Biotechnology and Bioengineering*, 2011. **108**(1): p. 69-81.
95. Gottschalk, U., *Process Scale Purification of Antibodies*. . 2009, Hoboken, New Jersey: John Wiley & Sons.
96. Laukel, M., P. Rogge, and G. Dudziak, *Disposable downstream processing for clinical manufacturing*. *BioProcess International*, 2011. **9**(S2): p. 14-21.
97. Whitford, W.G., *Single-use systems as principal components in bioproduction*. *BioProcess International*, 2010. **8**(11): p. 34-44.
98. Pegel, A., et al., *Evaluating disposable depth filtration platforms for mAb harvest clarification*. *BioProcess International*, 2011. **9**(9): p. 52-56.
99. O'Brien, T.P., et al., *Large-scale, single-use depth filtration systems for mammalian cell culture clarification*. *BioProcess International*, 2012. **10**(5): p. 50-57.
100. Bradford, S.A., et al., *Physical factors affecting the transport and fate of colloids in saturated porous media*. *Water Resources Research*, 2002. **38**(12): p. 1327.
101. Johnson, W.P., X.Q. Li, and G. Yal, *Colloid retention in porous media: Mechanistic confirmation of wedging and retention in zones of flow stagnation*. *Environmental Science & Technology*, 2007. **41**(4): p. 1279-1287.

102. Yigzaw, Y., et al., *Exploitation of the adsorptive properties of depth filters for host cell protein removal during monoclonal antibody purification*. Biotechnology Progress, 2006. **22**(1): p. 288-296.
103. Iwasaki, T., J. Slade, and W. Stanley, *Some notes on sand filtration*. American Water Works Association, 1937. **29**(10): p. 1591-1602.
104. Ison, C.R. and K.J. Ives, *Removal Mechanisms in Deep Bed Filtration*. Chemical Engineering Science, 1969. **24**(4): p. 717-&.
105. Jegatheesan, V. and S. Vigneswaran, *Mathematical modelling of deep bed filtration*. Modsim 2003: International Congress on Modelling and Simulation, Vols 1-4, 2003: p. 1805-1810.
106. Yao, K.M., M.M. Habibian, and C.R. Omelia, *Water and Waste Water Filtration - Concepts and Applications*. Environmental Science & Technology, 1971. **5**(11): p. 1105-&.
107. Tien, C. and B.V. Ramarao, *Granular Filtration of Aerosols And Hydrosols*. 2007: Elsevier.
108. Tufenkji, N. and M. Elimelech, *Correlation equation for predicting single-collector efficiency in physicochemical filtration in saturated porous media*. Environmental Science & Technology, 2004. **38**(2): p. 529-536.
109. Rajagopalan, R. and C. Tien, *Trajectory analysis of deep-bed filtration with the sphere-in-cell porous media model*. Aiche Journal, 1976. **22**(3): p. 523-533.
110. Nelson, K.E. and T.R. Ginn, *Colloid filtration theory and the Happel sphere-in-cell model revisited with direct numerical simulation of colloids*. Langmuir, 2005. **21**(6): p. 2173-2184.
111. Tufenkji, N. and M. Elimelech, *Breakdown of colloid filtration theory: Role of the secondary energy minimum and surface charge heterogeneities*. Langmuir, 2005. **21**(3): p. 841-852.
112. Happel, J., *Viscous Flow in Multiparticle Systems - Slow Motion of Fluids Relative to Beds of Spherical Particles*. Aiche Journal, 1958. **4**(2): p. 197-201.
113. Ma, H.L. and W.P. Johnson, *Colloid Retention in Porous Media of Various Porosities: Predictions by the Hemispheres-in-Cell Model*. Langmuir, 2010. **26**(3): p. 1680-1687.
114. Cushing, R.S. and D.F. Lawler, *Depth Filtration: Fundamental Investigation through Three-Dimensional Trajectory Analysis*. Environmental Science & Technology, 1998. **32**(23): p. 3793-3801.
115. Prieve, D.C. and E. Ruckenstein, *Effect of London forces upon the rate of deposition of Brownian particles*. Aiche Journal, 1974. **20**(6): p. 1178-1187.
116. Torkzaban, S., S.A. Bradford, and S.L. Walker, *Resolving the coupled effects of hydrodynamics and DLVO forces on colloid attachment in porous media*. Langmuir, 2007. **23**(19): p. 9652-9660.
117. Derjaguin, B.V. and L. Landau, *Theory of the Stability of Strongly Charged Lyophobic Sols and of the Adhesion of Strongly Charged Particles in Solutions of Electrolytes*. Acta Physico-Chimica, 1941. **14**: p. 633-662.
118. Verwey, E.J.W. and J.T.G. Overbeek, *Theory of the Stability of Lyophobic Colloids*. 1948: Elsevier Applied Science.
119. Tufenkji, N. and M. Elimelech, *Deviation from the classical colloid filtration theory in the presence of repulsive DLVO interactions*. Langmuir, 2004. **20**(25): p. 10818-10828.
120. Elimelech, M. and C.R. Omelia, *Effect of Particle-Size on Collision Efficiency in the Deposition of Brownian Particles with Electrostatic Energy Barriers*. Langmuir, 1990. **6**(6): p. 1153-1163.
121. Liang, Y., et al., *Interaction forces between colloidal particles in liquid: Theory and experiment*. Advances in Colloid and Interface Science, 2007. **134-35**: p. 151-166.

122. Elimelech, M. and C.R. Omelia, *Kinetics of Deposition of Colloidal Particles in Porous-Media*. Environmental Science & Technology, 1990. **24**(10): p. 1528-1536.
123. Redman, J.A., M.K. Estes, and S.B. Grant, *Resolving macroscale and microscale heterogeneity in virus filtration*. Colloids and Surfaces a-Physicochemical and Engineering Aspects, 2001. **191**(1-2): p. 57-70.
124. Logan, B.E., et al., *Clarification of Clean-Bed Filtration Models*. Journal of Environmental Engineering-Asce, 1995. **121**(12): p. 869-873.
125. Vigneswaran, S., V. Jegatheesan, and G. Keir, *Deep Bed Filtration: modeling theory and practice*. Waste Water Treatment Technologies. 2009, Oxford, UK: EOLSS Publishers.
126. Hunter, R.J., *Zeta Potential in Colloid Science: Principles and Applications*. 1981: Academic Press.
127. Franchi, A. and C.R. O'Melia, *Effects of Natural Organic Matter and Solution Chemistry on the Deposition and Reentrainment of Colloids in Porous Media*. Environmental Science & Technology, 2003. **37**(6): p. 1122-1129.
128. Hahn, M.W., D. Abadzic, and C.R. O'Melia, *Aquasols: On the role of secondary minima*. Environmental Science & Technology, 2004. **38**(22): p. 5915-5924.
129. Gregory, J. and A.J. Wishart, *Deposition of latex particles on alumina fibers*. Colloids and Surfaces, 1980. **1**(3-4): p. 313-334.
130. Tobiason, J.E., *Chemical effects on the deposition of non-Brownian particles*. Colloids and Surfaces, 1989. **39**(1): p. 53-75.
131. Song, L., P.R. Johnson, and M. Elimelech, *Kinetics of Colloid Deposition onto Heterogeneously Charged Surfaces in Porous Media*. Environmental Science & Technology, 1994. **28**(6): p. 1164-71.
132. Saiers, J.E. and J.N. Ryan, *Colloid deposition on non-ideal porous media: The influences of collector shape and roughness on the single-collector efficiency*. Geophysical Research Letters, 2005. **32**(21).
133. Bhattacharjee, S., C.H. Ko, and M. Elimelech, *DLVO interaction between rough surfaces*. Langmuir, 1998. **14**(12): p. 3365-3375.
134. Hahn, M.W. and C.R. O'Melia, *Deposition and reentrainment of Brownian particles in porous media under unfavorable chemical conditions: Some concepts and applications*. Environmental Science & Technology, 2004. **38**(1): p. 210-220.
135. Polyakov, Y.S., *Depth filtration approach to the theory of standard blocking: Prediction of membrane permeation rate and selectivity*. Journal of Membrane Science, 2008. **322**(1): p. 81-90.
136. Eaton, L.C., *Host cell contaminant protein assay development for recombinant biopharmaceuticals*. Journal of Chromatography A, 1995. **705**(1): p. 105-14.
137. Hagel, L. and L. Haneskog, *Size-exclusion Chromatography*, in *eLS*. 2001, John Wiley & Sons, Ltd.
138. Thommes, J. and M. Etzel, *Alternatives to chromatographic separations*. Biotechnology Progress, 2007. **23**(1): p. 42-45.
139. Burgess, R.R., *Protein precipitation techniques*. Methods in Enzymology, 2009. **463**: p. 331-42.
140. Berg, J.M., J.L. Tymoczko, and L. Stryer, *Protein Structure and Function*, in *Biochemistry*. 2002, W H Freeman: New York.
141. Hatti-Kaul, R., *Isolation and Purification of Proteins*. 2003.
142. Lee, T.S., et al., *A systematic approach to the large-scale production of protein crystals*. Enzyme and Microbial Technology, 2000. **26**(8): p. 582-592.
143. Hubbuch, J.J., et al., *The influence of homogenisation conditions on biomass-adsorbent interactions during ion-exchange expanded bed adsorption*. Biotechnology and Bioengineering, 2006. **94**(3): p. 543-553.

144. Anspach, F.B., et al., *Expanded-bed chromatography in primary protein purification*. Journal of Chromatography A, 1999. **865**(1-2): p. 129-144.
145. Hubbuch, J., J. Thommes, and M.R. Kula, *Biochemical engineering aspects of expanded bed adsorption*. Technology Transfer in Biotechnology: From Lab to Industry to Production, 2005. **92**: p. 101-123.
146. Ghosh, R., *Protein separation using membrane chromatography: opportunities and challenges*. Journal of Chromatography A, 2002. **952**(1-2): p. 13-27.
147. Guiochon, G. and L.A. Beaver, *Separation science is the key to successful biopharmaceuticals*. Journal of Chromatography A, 2011. **1218**(49): p. 8836-8858.
148. Healthcare, G., *Protein Purification: Handbook*.
149. Janson, J.C., *Protein Purification: Principles, high resolution methods, and applications*. 2011, Hoboken, NJ: John Wiley & Sons, Inc.
150. Wankat, P.C., *Rate-Controlled Separations*. 1994: Springer Netherlands.
151. McNaught, A.D. and A. Wilkinson, *IUPAC. Compendium of Chemical Terminology (the "Gold Book")*. 2nd ed. 1997, Oxford: Blackwell Scientific Publications.
152. Aboul-Enein, H.Y., *Selectivity versus specificity in chromatographic analytical methods*. Accreditation and Quality Assurance, 2000. **5**(5): p. 180-181.
153. Trathnigg, B., *Size-exclusion Chromatography of Polymers*. Encyclopedia of Analytical Chemistry, ed. R.A. Meyers. 2000, Chichester: John Wiley & Sons Ltd. 8008-8034.
154. Fritz, J.S., *Factors affecting selectivity in ion chromatography*. Journal of Chromatography A, 2005. **1085**(1): p. 8-17.
155. Holstein, M.A., et al., *Mobile phase modifier effects in multimodal cation exchange chromatography*. Biotechnology and Bioengineering, 2012. **109**(1): p. 176-186.
156. Lenhoff, A.M., *Protein adsorption and transport in polymer-functionalized ion-exchangers*. Journal of Chromatography A, 2011. **1218**(49): p. 8748-8759.
157. Braithwaite, A. and F.J. Smith, *Chromatographic Methods*. 1996: Blackie Academic & Professional.
158. Yamamoto, S., et al., *Ion exchange chromatography of proteins-prediction of elution curves and operating conditions. I. Theoretical considerations*. Biotechnology and bioengineering, 1983. **25**(6): p. 1465-83.
159. Martin, A.J. and R.L. Synge, *A new form of chromatogram employing two liquid phases: A theory of chromatography. 2. Application to the micro-determination of the higher monoamino-acids in proteins*. Biochemical journal, 1941. **35**(12): p. 1358-68.
160. Nelson, W.C., D.F. Silarski, and P.C. Wankat, *Continuous-Flow Equilibrium Staged Model for Cycling Zone Adsorption*. Industrial & Engineering Chemistry Fundamentals, 1978. **17**(1): p. 32-38.
161. Towler, G.P. and R.K. Sinnott, *Chemical Engineering Design: Principles, Practice and Economics of Plant and Process Design*. 2008, Oxford: Elsevier/Butterworth-Heinemann.
162. Vandemter, J.J., *Citation Classic - Longitudinal Diffusion and Resistance to Mass-Transfer as Causes of Non-Ideality in Chromatography*. Current Contents/Engineering Technology & Applied Sciences, 1981(3): p. 16-16.
163. Janson, J.-C. and J.Å. Jönsson, *Introduction to Chromatography*, in *Protein Purification*. 2011, John Wiley & Sons, Inc.: Hoboken, NJ. p. 23-50.
164. Orellana, C.A., C. Shene, and J.A. Asenjo, *Mathematical Modeling of Elution Curves for a Protein Mixture in Ion Exchange Chromatography Applied to High Protein Concentration*. Biotechnology and Bioengineering, 2009. **104**(3): p. 572-581.
165. Giddings, J.C. and H. Eyring, *A Molecular Dynamic Theory of Chromatography*. Journal of Physical Chemistry, 1955. **59**(5): p. 416-421.
166. Gu, T., *Mathematical Modeling and Scale-Up of Liquid Chromatography*. 1995, Berlin: Springer Berlin Heidelberg.

167. Sandoval, G., et al., *Extension of the selection of protein chromatography and the rate model to affinity chromatography*. Journal of Molecular Recognition, 2010. **23**(6): p. 609-617.
168. Hunter, A.K. and G. Carta, *Protein adsorption on novel acrylamido-based polymeric ion-exchangers - IV. Effects of protein size on adsorption capacity and rate*. Journal of Chromatography A, 2002. **971**(1-2): p. 105-116.
169. Yamamoto, S., *Electrostatic interaction chromatography process for protein separations: Impact of engineering analysis of biorecognition mechanism on process optimization*. Chemical Engineering & Technology, 2005. **28**(11): p. 1387-1393.
170. Gu, T. and Y. Zheng, *A study of the scale-up of reversed-phase liquid chromatography*. Separation and Purification Technology, 1999. **15**(1): p. 41-58.
171. Gu, T., et al., *Modeling of Gradient Elution in Multicomponent Nonlinear Chromatography*. Chemical Engineering Science, 1992. **47**(1): p. 253-262.
172. Adkins, C.J., *Equilibrium Thermodynamics*. 1983: Cambridge University Press.
173. Freundlich, H., *Über die Adsorption in Lösungen*. 1906, Leipzig: Wilhelm Engelmann.
174. Langmuir, I., *The constitution and fundamental properties of solids and liquids. Part I solids*. Journal of the American Chemical Society, 1916. **38**: p. 2221-2295.
175. Monod, J., *The Growth of Bacterial Cultures*. Annual Review of Microbiology, 1949. **3**: p. 371-394.
176. Michaelis, L. and M.L. Menten, *Die kinetik der invertinwirkung*. Biochemische Zeitschrift, 1913. **49**(333-369): p. 352.
177. Butler, J.A.V. and C. Ockrent, *Studies in Electrocapillarity. Part III. The Surface Tensions of Solutions Containing Two Surface-Active Solute*. The Journal of Physical Chemistry B, 1930(34): p. 2841-2845.
178. Choy, K.K.H., J.F. Porter, and G. McKay, *Langmuir isotherm models applied to the multicomponent sorption of acid dyes from effluent onto activated carbon*. Journal of Chemical and Engineering Data, 2000. **45**(4): p. 575-584.
179. Golshanshirazi, S., J.X. Huang, and G.A. Guiochon, *Comparison of an Experimental Competitive Isotherm and the Levan-Vermeulen Model and Prediction of Band Profiles in a Case of Selectivity Reversal*. Analytical Chemistry, 1991. **63**(11): p. 1147-1154.
180. LeVan, M.D. and T. Vermeulen, *Binary Langmuir and Freundlich isotherms for ideal adsorbed solutions*. The Journal of Physical Chemistry, 1981. **85**(22): p. 3247-3250.
181. Brunauer, S., P.H. Emmett, and E. Teller, *Adsorption of Gases in Multimolecular Layers*. Journal of the American Chemical Society, 1938. **60**(2): p. 309-319.
182. Brooks, C.A. and S.M. Cramer, *Steric Mass-Action Ion-Exchange - Displacement Profiles and Induced Salt Gradients*. Aiche Journal, 1992. **38**(12): p. 1969-1978.
183. Osberghaus, A., et al., *Determination of parameters for the steric mass action model-A comparison between two approaches*. Journal of Chromatography A, 2012. **1233**: p. 54-65.
184. Bentrop, D. and H. Engelhardt, *Chromatographic Characterization of Ion-Exchangers for High-Performance Liquid-Chromatography of Proteins .1. Chromatographic Determination of Loading Capacity for Low-Molecular and High-Molecular Mass Anions*. Journal of Chromatography, 1991. **556**(1-2): p. 363-372.
185. Velayudhan, A. and C. Horvath, *Preparative chromatography of proteins analysis of the multivalent ion-exchange formalism*. Journal of Chromatography, 1988. **443**: p. 13-29.
186. Regnier, F.E. and I. Mazsaroff, *A Theoretical-Examination of Adsorption Processes in Preparative Liquid-Chromatography of Proteins*. Biotechnology Progress, 1987. **3**(1): p. 22-26.

187. Xia, F., D. Nagrath, and S.M. Cramer, *Modeling of adsorption in hydrophobic interaction chromatography systems using a preferential interaction quadratic isotherm*. Journal of Chromatography A, 2003. **989**(1): p. 47-54.
188. Lienqueo, M.E., et al., *Current insights on protein behaviour in hydrophobic interaction chromatography*. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences, 2007. **849**(1-2): p. 53-68.
189. Mollerup, J.M., *Applied thermodynamics: A new frontier for biotechnology*. Fluid Phase Equilibria, 2006. **241**(1-2): p. 205-215.
190. Nfor, B.K., et al., *High-throughput isotherm determination and thermodynamic modeling of protein adsorption on mixed mode adsorbents*. Journal of Chromatography A, 2010. **1217**(44): p. 6829-6850.
191. Mollerup, J.M., *Modelling oligomer formation in chromatographic separations*. Journal of Chromatography A, 2011. **1218**(49): p. 8869-8873.
192. Machold, C., et al., *Hydrophobic interaction chromatography of proteins. I. Comparison of selectivity*. Journal of Chromatography A, 2002. **972**(1): p. 3-19.
193. Iyer, H., et al., *Use of the steric mass action model in ion-exchange chromatographic process development*. Journal of Chromatography A, 1999. **832**(1-2): p. 1-9.
194. Levene, P.A. and H.S. Simms, *Calculation of Isoelectric Points*. Journal of Biological Chemistry, 1923. **55**: p. 801-813.
195. Kendrew, J.C., et al., *A three-dimensional model of the myoglobin molecule obtained by x-ray analysis*. Nature, 1958. **181**(4610): p. 662-6.
196. Wuthrich, K., *Protein structure determination in solution by NMR spectroscopy*. Journal of Biological Chemistry, 1990. **265**(36): p. 22059-62.
197. Marti-Renom, M.A., et al., *Comparative protein structure modeling of genes and genomes*. Annual Review of Biophysics and Biomolecular Structure, 2000. **29**: p. 291-325.
198. Kiefer, F., et al., *The SWISS-MODEL Repository and associated resources*. Nucleic acids research, 2009. **37**(Database issue): p. D387-92.
199. Benkert, P., M. Kunzli, and T. Schwede, *QMEAN server for protein model quality estimation*. Nucleic Acids Research, 2009. **37**: p. W510-W514.
200. Arnold, K., et al., *The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling*. Bioinformatics, 2006. **22**(2): p. 195-201.
201. Kabsch, W. and C. Sander, *Dictionary of Protein Secondary Structure - Pattern-Recognition of Hydrogen-Bonded and Geometrical Features*. Biopolymers, 1983. **22**(12): p. 2577-2637.
202. Guvench, O. and A.D. MacKerell, Jr., *Comparison of protein force fields for molecular dynamics simulations*. Methods in molecular biology, 2008. **443**: p. 63-88.
203. Cornell, W.D., et al., *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995)*. Journal of the American Chemical Society, 1996. **118**(9): p. 2309-2309.
204. Brooks, B.R., et al., *Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations*. Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.
205. Brooks, B.R., et al., *CHARMM: The Biomolecular Simulation Program*. Journal of Computational Chemistry, 2009. **30**(10): p. 1545-1614.
206. Lide, D.R., *CRC Handbook of Chemistry and Physics: A Ready-reference Book of Chemical and Physical Data*. 2004: CRC Press.
207. Salsbury, F.R., *Molecular dynamics simulations of protein dynamics and their relevance to drug discovery*. Current Opinion in Pharmacology, 2010. **10**(6): p. 738-744.

208. Klepeis, J.L., et al., *Long-timescale molecular dynamics simulations of protein structure and function*. Current Opinion in Structural Biology, 2009. **19**(2): p. 120-127.
209. Durrant, J.D. and J.A. McCammon, *Molecular dynamics simulations and drug discovery*. BMC Biology, 2011. **9**.
210. Hildebrandt, A., et al., *Electrostatic potentials of proteins in water: a structured continuum approach*. Bioinformatics, 2007. **23**(2): p. E99-E103.
211. Beauchamp, K.A., et al., *Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements*. Journal of Chemical Theory and Computation, 2012. **8**(4): p. 1409-1414.
212. Dismer, F. and J. Hubbuch, *3D structure-based protein retention prediction for ion-exchange chromatography*. Journal of Chromatography A, 2010. **1217**(8): p. 1343-1353.
213. Tropsha, A. and A. Golbraikh, *Predictive QSAR Modeling workflow, model applicability domains, and virtual screening*. Current Pharmaceutical Design, 2007. **13**(34): p. 3494-3504.
214. Eriksson, L., et al., *Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs*. Environmental Health Perspectives, 2003. **111**(10): p. 1361-1375.
215. Gonzalez, M.P., et al., *Variable Selection Methods in QSAR: An Overview*. Current Topics in Medicinal Chemistry, 2008. **8**(18): p. 1606-1627.
216. Konovalov, D.A., et al., *Statistical confidence for variable selection in QSAR models via Monte Carlo cross-validation*. Journal of Chemical Information and Modeling, 2008. **48**(2): p. 370-83.
217. Tropsha, A., P. Gramatica, and V.K. Gombar, *The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models*. Qsar & Combinatorial Science, 2003. **22**(1): p. 69-77.
218. Gramatica, P., *Principles of QSAR models validation: internal and external*. Qsar & Combinatorial Science, 2007. **26**(5): p. 694-701.
219. Dehmer, M., et al., *Statistical Modelling of Molecular Descriptors in QSAR/QSPR*. 2012, Hoboken, NJ: Wiley-Blackwell.
220. Teodoro, M.L., G.N. Phillips, and L.E. Kaviraki, *Molecular docking: A problem with thousands of degrees of freedom*. 2001 IEEE International Conference on Robotics and Automation, Vols I-IV, Proceedings, 2001: p. 960-965.
221. Tait, A.S., et al., *Host cell protein dynamics in the supernatant of a mAb producing CHO cell line*. Biotechnology and Bioengineering, 2012. **109**(4): p. 971-982.
222. Butler, S., J. Luczak, and L.J. Schiff, *Outsourced Aseptic Fill/Finish and Stability Programs for Biopharmaceuticals*. BioProcess International, 2003: p. 44-49.
223. Levy, G. and E. Nelson, *Pharmaceutical Formulation and Therapeutic Efficacy*. Jama-Journal of the American Medical Association, 1961. **177**(10): p. 689-&.
224. Baldrick, P., *Pharmaceutical excipient development: The need for preclinical guidance*. Regulatory Toxicology and Pharmacology, 2000. **32**(2): p. 210-218.
225. Garçon, N., G. Leroux-Roels, and W.-F. Cheng, *Vaccine adjuvants*. Perspectives in Vaccinology, 2011. **1**(1): p. 89-113.
226. (ICH), I.C.o.H.o.T.R.f.R.o.P.f.H.U. *Quality Guidelines*. 2012; Available from: <http://www.ich.org/products/guidelines/quality/article/quality-guidelines.html>.
227. Food and Drug Administration, *International Conference on Harmonisation: Guidance on preclinical safety evaluation of biotechnology-derived pharmaceuticals*. Federal Register, 1997. **62**: p. 61515-61519.
228. Gad, S.C., *Handbook of Pharmaceutical Biotechnology*. 2007: Wiley-Interscience.
229. Booth, P.M., *FDA implementation of standards developed by the international conference on harmonisation*. Food and Drug Law Journal, 1997. **52**(2): p. 203-223.

230. Bestwick, D. and R. Colton, *Extractables and Leachables from Single-Use Disposables*. BioProcess International, 2009. **7**(S1): p. 88-94.
231. Aranha, H., *Viral clearance strategies for biopharmaceutical safety Part 1: General considerations*. Biopharm-the Applied Technologies of Biopharmaceutical Development, 2001. **14**(1): p. 28-35.
232. Aranha, H. and S. Forbes, *Viral clearance strategies for biopharmaceutical safety - Part 3. A multifaceted approach to process validation*. Biopharm-the Applied Technologies of Biopharmaceutical Development, 2001. **14**(5): p. 42-+.
233. Aranha, H., *Viral clearance strategies for biopharmaceutical safety - Part 2: Filtration for viral clearance*. Biopharm-the Applied Technologies of Biopharmaceutical Development, 2001. **14**(2): p. 32-+.
234. Atkinson, A.C., B. Bogacka, and A.A. ZhigljavskiÄ-, *Optimum design 2000*. 2001: Kluwer Academic Publishers.
235. Montgomery, D.C., *Design and Analysis of Experiments, Minitab Manual*. 2008: Wiley.
236. Mandenius, C.F. and A. Brundin, *Bioprocess optimization using design-of-experiments methodology*. Biotechnology Progress, 2008. **24**(6): p. 1191-203.
237. Abu-Absi, S.F., et al., *Defining process design space for monoclonal antibody cell culture*. Biotechnology and Bioengineering, 2010. **106**(6): p. 894-905.
238. Kleppmann, W., *Taschenbuch Versuchsplanung*. 6th edition ed. Produkte und Prozesse optimieren, ed. F.J. Brunner. 2009, München: Carl Hanser Verlag.
239. Myers, R.H., et al., *Response surface methodology: A retrospective and literature survey*. Journal of Quality Technology, 2004. **36**(1): p. 53-77.
240. Eriksson, L., *Design of Experiments: Principles and Applications*. 2008: Umetrics Academy.
241. Goos, P. and B. Jones, *Optimal Design of Experiments: A Case Study Approach*. 2011: John Wiley & Sons.
242. Gilmour, S.G. and L.A. Trinca, *Optimum design of experiments for statistical inference*. Journal of the Royal Statistical Society Series C-Applied Statistics, 2012. **61**: p. 345-401.
243. Brook, R.J. and G.C. Arnold, *Applied Regression Analysis and Experimental Design*. 1985: M. Dekker.
244. Weisberg, S., *Applied Linear Regression*. 2005: Wiley-Interscience.
245. Myers, R.H., *Classical and modern regression with applications*. 1990: PWS-KENT.
246. Winkelkemper, et al., *Purification performance index and separation cost indicator for experimentally based systematic downstream process development*. Vol. 72. 2010, Kidlington, ROYAUME-UNI: Elsevier. 6.
247. Buyel, J.F. and R. Fischer, *Predictive models for transient protein expression in tobacco (Nicotiana tabacum L.) can optimize process time, yield, and downstream costs*. Biotechnology and Bioengineering, 2012. **109**(10): p. 2575-88.
248. Osberghaus, A., et al., *Optimizing a chromatographic three component separation: A comparison of mechanistic and empiric modeling approaches*. Journal of Chromatography A, 2012. **1237**: p. 86-95.
249. Main, G.D., S. Reynolds, and J.S. Gartland, *Electroporation protocols for Agrobacterium*. Methods in Molecular Biology, 1995. **44**: p. 405-12.
250. Holland, T., et al., *Optimal nitrogen supply as a key to increased and sustained production of a monoclonal full-size antibody in BY-2 suspension culture*. Biotechnology and Bioengineering, 2010. **107**(2): p. 278-289.
251. Bradford, M.M., M.H. Simonian, and J.A. Smith, *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding*. Anal Biochem, 1976. **72**(10): p. 248-54.

252. Simonian, M.H. and J.A. Smith, *Spectrophotometric and colorimetric determination of protein concentration*. Current Protocols in Molecular Biology, 2006. **76**(Chapter 10): p. 10.1.1–10.1A.9.
253. Howell, S., et al., *High-density immobilization of an antibody fragment to a carboxymethylated dextran-linked biosensor surface*. Journal of Molecular Recognition, 1998. **11**(1-6): p. 200-203.
254. Piliarik, M., H. Vaisocherova, and J. Homola, *Surface plasmon resonance biosensing*. Methods in molecular biology, 2009. **503**: p. 65-88.
255. Lobedann, M., *Pilot-scale process development for the purification of the recombinant antibody 2G12 from transgenic tobacco*, in *Molecular Biotechnology, PhD thesis*. 2009, RWTH Aachen University: Aachen.
256. UniProt-Consortium, *Reorganizing the protein space at the Universal Protein Resource (UniProt)*. Nucleic Acids Research, 2012. **40**(D1): p. D71-D75.
257. Velankar, S., et al., *PDBe: Protein Data Bank in Europe*. Nucleic Acids Research, 2012. **40**(D1): p. D445-D452.
258. Bernstein, F.C., et al., *The Protein Data Bank: a computer-based archival file for macromolecular structures*. Journal of Molecular Biology, 1977. **112**(3): p. 535-42.
259. Peitsch, M.C., *Protein Modeling by E-Mail*. Bio-Technology, 1995. **13**(7): p. 658-660.
260. Emanuelsson, O., et al., *Predicting subcellular localization of proteins based on their N-terminal amino acid sequence*. Journal of Molecular Biology, 2000. **300**(4): p. 1005-1016.
261. Nielsen, H., et al., *Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites*. Protein Engineering, 1997. **10**(1): p. 1-6.
262. Benkert, P., T. Schwede, and S.C.E. Tosatto, *QMEANclust: estimation of protein model quality by combining a composite scoring function with structural density information*. BMC Structural Biology, 2009. **9**: p. 35.
263. Benkert, P., S.C.E. Tosatto, and D. Schomburg, *QMEAN: A comprehensive scoring function for model quality assessment*. Proteins-Structure Function and Bioinformatics, 2008. **71**(1): p. 261-277.
264. Benkert, P., M. Biasini, and T. Schwede, *Toward the estimation of the absolute quality of individual protein structure models*. Bioinformatics, 2011. **27**(3): p. 343-350.
265. Breneman, C.M., et al., *New developments in pest shape/property hybrid descriptors*. Abstracts of Papers of the American Chemical Society, 2002. **224**: p. U487-U487.
266. Hou, Y. and S.M. Cramer, *Evaluation of selectivity in multimodal anion exchange systems: A priori prediction of protein retention and examination of mobile phase modifier effects*. Journal of Chromatography A, 2011. **1218**(43): p. 7813-7820.
267. Nagrath, D., F. Xia, and S.M. Cramer, *Characterization and modeling of nonlinear hydrophobic interaction chromatographic systems*. Journal of Chromatography A, 2011. **1218**(9): p. 1219-1226.
268. Chen, J., T. Yang, and S.M. Cramer, *Prediction of protein retention times in gradient hydrophobic interaction chromatographic systems*. Journal of Chromatography A, 2008. **1177**(2): p. 207-214.
269. Ladiwala, A., et al., *Investigation of protein retention and selectivity in HIC systems using quantitative structure retention relationship models*. Biotechnology and Bioengineering, 2006. **93**(5): p. 836-850.
270. Breneman, C.M., et al., *Prediction of protein retention times in anion-exchange chromatography systems using support vector regression*. Chemometrics and Chemoinformatics, 2005. **894**: p. 111-125.
271. Song, M.H., et al., *Prediction of protein retention times in anion-exchange chromatography systems using support vector regression*. Journal of Chemical Information and Computer Sciences, 2002. **42**(6): p. 1347-1357.

272. Tugcu, N., et al., *Prediction of the effect of mobile-phase salt type on protein retention and selectivity in anion exchange systems*. Analytical Chemistry, 2003. **75**(14): p. 3563-3572.
273. Ladiwala, A., et al., *A priori prediction of adsorption isotherm parameters and chromatographic behavior in ion-exchange systems*. Proceedings of the National Academy of Sciences, USA, 2005. **102**(33): p. 11710-11715.
274. Ladiwala, A., et al., *Prediction of protein affinity in hydrophobic interaction chromatography using quantitative structure-retention relationship (QSRR) models*. Abstracts of Papers of the American Chemical Society, 2004. **227**: p. U1020-U1020.
275. McLachlan, G.J., K.-A. Do, and C. Ambrose, *Some Cluster Analysis Methods*, in *Analyzing Microarray Gene Expression Data*. 2004, John Wiley & Sons, Inc.: Hoboken, NJ. p. 61-97.
276. Peixoto, J.L., *Hierarchical Variable Selection in Polynomial Regression-Models*. American Statistician, 1987. **41**(4): p. 311-313.
277. Peixoto, J.L., *A Property of Well-Formulated Polynomial Regression-Models*. American Statistician, 1990. **44**(1): p. 26-30.
278. Buyel, J.F., *Manufacturing biopharmaceutical proteins by transient expression in Nicotiana tabacum (L.)*, in *Molecular Biotechnology*. 2013, RWTH Aachen University: Aachen. p. 98.
279. Anonymous, *Annex 15: Qualification and validation*. 2001, European Commission: Brussels. p. 11.
280. Filterbag.com. *Filter Bag Sizes*. [Web page] 2011 [cited 2013 06.02.2013]; Available from: <http://www.filterbag.com/Filter-Bag-Sizes-18.html>.
281. Hanen, F., et al., *Effect of salinity on growth, leaf-phenolic content and antioxidant scavenging activity in Cynara cardunculus L*, in *Biosaline Agriculture and High Salinity Tolerance*. 2008, Birkhuser: Basel. p. 335-343.
282. Neves, G.Y.S., et al., *Root growth inhibition and lignification induced by salt stress in soybean*. Journal of Agronomy and Crop Science, 2010. **196**(6): p. 467-473.
283. Huh, N.W., et al., *The interaction of polyphenols with bilayers: Conditions for increasing bilayer adhesion*. Biophysical Journal, 1996. **71**(6): p. 3261-3277.
284. Nakayama, T., et al., *Affinity of polyphenols for lipid bilayers*. Biofactors, 2000. **13**(1-4): p. 147-151.
285. McIntosh, T.J., et al., *Polyphenols increase adhesion between lipid bilayers by forming interbilayer bridges*. Plant Polyphenols 2: Chemistry, Biology, Pharmacology, Ecology, 2000. **66**: p. 451-470.
286. Graupner, N., *Application of lignin as natural adhesion promoter in cotton fibre-reinforced poly(lactic acid) (PLA) composites*. Journal of Materials Science, 2008. **43**(15): p. 5222-5229.
287. Pocius, A.V. and M. Chaudhury, *Surfaces, Chemistry and Applications: Adhesion Science and Engineering*. 2002: Elsevier.
288. Caffall, K.H. and D. Mohnen, *The structure, function, and biosynthesis of plant cell wall pectic polysaccharides*. Carbohydrate Research, 2009. **344**(14): p. 1879-1900.
289. Iraki, N.M., et al., *Alteration of the physical and chemical-structure of the primary-cell wall of growth-limited plant-cells adapted to osmotic-stress*. Plant Physiology, 1989. **91**(1): p. 39-47.
290. Iraki, N.M., et al., *Cell-walls of tobacco cells and changes in composition associated with reduced growth upon adaptation to water and saline stress*. Plant Physiology, 1989. **91**(1): p. 48-53.
291. Keutgen, A.J. and E. Pawelzik, *Cultivar-dependent cell wall modification of strawberry fruit under NaCl salinity stress*. Journal of Agricultural and Food Chemistry, 2007. **55**(18): p. 7580-7585.

292. Zsivanovits, G., et al., *Material properties of concentrated pectin networks*. Carbohydrate Research, 2004. **339**(7): p. 1317-1322.
293. Matia-Merino, L., K. Lau, and E. Dickinson, *Effects of low-methoxyl amidated pectin and ionic calcium on rheology and microstructure of acid-induced sodium caseinate gels*. Food Hydrocolloids, 2004. **18**(2): p. 271-281.
294. Janeway, C., *Immunobiology five*. 2001: Garland Pub.
295. Baird, G.S., D.A. Zacharias, and R.Y. Tsien, *Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral*. Proceedings of the National Academy of Sciences, USA, 2000. **97**(22): p. 11984-11989.
296. Sajadi, M.M., et al., *Signature biochemical properties of broadly cross-reactive HIV-1 neutralizing antibodies in human plasma*. Journal of Virology, 2012. **86**(9): p. 5014-5025.
297. Boulton, R.B., V.L. Singleton, and L.F. Bisson, *Principles and Practices of Winemaking* 1998: Springer.
298. Buyel, J.F., H.M. Gruchow, and R. Fischer, *Depth Filters Containing Diatomite Achieve More Efficient Particle Retention than Filters Solely Containing Cellulose Fibers*. Frontiers in Plant Science, 2015. **6**(1134): p. 1-11.
299. Dullien, F.A.L., *Maximizing the capacity and life of depth-type filters*. Canadian Journal of Chemical Engineering, 1989. **67**(4): p. 689-692.
300. Nečesaný, V., *The isoelectric point of lignified cell walls*. Holz als Roh- und Werkstoff, 1971. **29**(9): p. 354-357.
301. Buyel, J.F. and R. Fischer, *Synthetic polymers are more effective than natural flocculants for the clarification of tobacco leaf extracts*. Journal of Biotechnology, 2014. **195**: p. 37-42.
302. Buyel, J.F., P. Opdensteinen, and R. Fischer, *Cellulose-based filter aids increase the capacity of depth filters during the downstream processing of plant-derived biopharmaceutical proteins*. Biotechnology Journal, 2014. **10**(4): p. 584-591.
303. Buyel, J.F. and R. Fischer, *Downstream processing of biopharmaceutical proteins produced in plants: the pros and cons of flocculants*. Bioengineered, 2014. **5**(2): p. 138-42.
304. Menzel, S., et al., *Optimized Blanching Reduces the Host Cell Protein Content and Substantially Enhances the Recovery and Stability of Two Plant-Derived Malaria Vaccine Candidates*. Frontiers in Plant Science, 2016. **7**(159): p. 1-15.
305. Buyel, J.F., et al., *Rational design of a host cell protein heat precipitation step simplifies the subsequent purification of recombinant proteins from tobacco*. Biochemical Engineering Journal, 2014. **88**(15 July 2014): p. 162-170.
306. Voepel, N., et al., *Malaria vaccine candidate antigen targeting the pre-erythrocytic stage of Plasmodium falciparum produced at high level in plants*. Biotechnology Journal, 2014. **9**(11): p. 1435-1445.
307. Buyel, J.F. and R. Fischer, *Scale-down models to optimize a filter train for the downstream purification of recombinant pharmaceutical proteins produced in tobacco leaves*. Biotechnology Journal, 2014. **9**(3): p. 415-25.
308. Buyel, J.F. and R. Fischer, *Flocculation increases the efficacy of depth filtration during the downstream processing of recombinant pharmaceutical proteins produced in tobacco*. Plant Biotechnology Journal, 2014. **12**(2): p. 240-52.
309. Kim, T.D., et al., *Thermal behavior of proteins: Heat-resistant proteins and their heat-induced secondary structural changes*. Biochemistry, 2000. **39**(48): p. 14839-14846.
310. Kwon, S., Y. Jung, and D. Lim, *Proteomic analysis of heat-stable proteins in Escherichia coli*. Bmb Reports, 2008. **41**(2): p. 108-111.

311. Buyel, J.F. and R. Fischer, *Characterization of complex systems using the design of experiments approach: transient protein expression in tobacco as a case study*. Journal of Visualized Experiments, 2014. **1**(83): p. e51216.
312. Buyel, J.F. and R. Fischer, *A juice extractor can simplify the downstream processing of plant-derived biopharmaceutical proteins compared to blade-based homogenizers*. Process Biochemistry, 2014. **50**(5): p. 859-866.
313. Komarnytsky, S., et al., *Production of recombinant proteins in tobacco guttation fluid*. Plant Physiology, 2000. **124**(3): p. 927-933.
314. Drake, P.M.W., et al., *Development of rhizosecretion as a production system for recombinant proteins from hydroponic cultivated tobacco*. FASEB Journal, 2009. **23**(10): p. 3581-3589.
315. Turpen, T.H., *Tobacco mosaic virus and the virescence of biotechnology*. Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences, 1999. **354**(1383): p. 665-673.
316. Kingsbury, N.J. and K.A. McDonald, *Quantitative Evaluation of EI Endoglucanase Recovery from Tobacco Leaves Using the Vacuum Infiltration-Centrifugation Method*. Biomed Research International, 2014. **2014**: p. 483596.
317. Beiss, V., et al., *Heat-precipitation allows the efficient purification of a functional plant-derived malaria transmission-blocking vaccine candidate fusion protein*. Biotechnology and Bioengineering, 2015. **112**(7): p. 1297-1305.
318. Buyel, J.F., *Numeric simulation can be used to predict heat transfer during the blanching of leaves and intact plants*. Biochemical Engineering Journal, 2016. **109**: p. 118-126.
319. Mandal, M.K., et al., *Inhibition of protease activity by antisense RNA improves recombinant protein production in Nicotiana tabacum cv. Bright Yellow 2 (BY-2) suspension cells*. Biotechnology Journal, 2014. **9**(8): p. 1065-1073.
320. Welty, J.R., C.E. Wicks, and R.E. Wilson, *Fundamentals of momentum, heat, and mass transfer*. 1976: Wiley.
321. Lowe, D., et al., *Aggregation, stability, and formulation of human antibody therapeutics*. Adv Protein Chem Struct Biol, 2011. **84**: p. 41-61.
322. Gong, R., et al., *Engineered human antibody constant domains with increased stability*. Journal of Biological Chemistry, 2009. **284**(21): p. 14203-10.
323. Rouet, R., D. Lowe, and D. Christ, *Stability engineering of the human antibody repertoire*. FEBS Letters, 2014. **588**(2): p. 269-77.
324. Stabel, J.R. and A. Lambertz, *Efficacy of pasteurization conditions for the inactivation of Mycobacterium avium subsp paratuberculosis in milk*. Journal of Food Protection, 2004. **67**(12): p. 2719-2726.
325. Wichers, H., *Minimal processing and the allergenic properties of food*, in *Managing Allergens in Food*, C. Mills, H. Wichers, and K. Hoffmann-Sommergruber, Editors. 2006, Elsevier Science: New York, NY. p. 336.
326. Davis, P.J. and S.C. Williams, *Protein modification by thermal processing*. Allergy, 1998. **53**: p. 102-105.
327. Dubois, M.F., A.G. Hovanessian, and O. Bensaude, *Heat-shock-induced denaturation of proteins. Characterization of the insolubilization of the interferon-induced p68 kinase*. J Biol Chem, 1991. **266**(15): p. 9707-11.
328. Nfor, B.K., et al., *Multi-dimensional fractionation and characterization of crude protein mixtures: Toward establishment of a database of protein purification process development parameters*. Biotechnology and Bioengineering, 2012. **109**(12): p. 3070-83.
329. Egelkrout, E., V. Rajan, and J.A. Howard, *Overproduction of recombinant proteins in plants*. Plant Sci, 2012. **184**: p. 83-101.

330. Zilberstein, A., et al., *Characterization of Dnase(S) Activity in Tobacco Leaf Extracts*. Plant Science, 1987. **52**(1-2): p. 57-65.
331. Buyel, J.F., et al., *Predictive models for the accumulation of a fluorescent marker protein in tobacco leaves according to the promoter/5'UTR combination*. Biotechnology and Bioengineering, 2013. **110**(2): p. 471-82.
332. Spreitzer, R.J. and M.E. Salvucci, *Rubisco: Structure, regulatory interactions, and possibilities for a better enzyme*. Annual Review of Plant Biology, 2002. **53**: p. 449-475.
333. Tao, Y.Y., et al., *Adsorption of deamidated antibody variants on macroporous and dextran-grafted cation exchangers: II. Adsorption kinetics*. Journal of Chromatography A, 2011. **1218**(11): p. 1530-1537.
334. Stone, M.C. and G. Carta, *Protein adsorption and transport in agarose and dextran-grafted agarose media for ion exchange chromatography*. Journal of Chromatography A, 2007. **1146**(2): p. 202-215.
335. Almodovar, E.X.P., Y.Y. Tao, and G. Carta, *Protein Adsorption and Transport in Cation Exchangers with a Rigid Backbone Matrix with and without Polymeric Surface Extenders*. Biotechnology Progress, 2011. **27**(5): p. 1264-1272.
336. Staby, A., et al., *Comparison of chromatographic ion-exchange resins III. Strong cation-exchange resins*. Journal of Chromatography A, 2004. **1034**(1-2): p. 85-97.
337. Jackewitz, A., *Development of a High Capacity, Mixed-Mode Resin for High Conductivity mAb Feedstocks*. Biopharm International, 2012. **10**(7): p. 82-84.
338. Pall Corporation, *Instructions for use: HyperCel™ STAR AX 1 mL and 5 mL PRC*, in *Instructions for use: HyperCel™ STAR AX*. 2012, Pall Corporation: Port Washington. p. 1-2.
339. Michaelis, L., M.P. Schubert, and C.V. Smythe, *Potentiometric Study of the Flavins*. Journal of Biological Chemistry, 1936. **116**(2): p. 587-607.
340. GE Healthcare, *Capto™ adhere Data file*, in *Multimodal media: Capto™ adhere Data file*. 2011, GE Healthcare: Uppsala. p. 1-8.
341. Yang, T., et al., *Evaluation of multi-modal high salt binding ion exchange materials*. Journal of Chromatography A, 2007. **1157**(1-2): p. 171-7.
342. Pall Corporation, *AcroSep™ HEA and PPA HyperCel™ Columns*, in *Mixed-Mode Chromatography Columns for Protein Separation*. 2008, Pall Corporation: Ann Arbor. p. 1-12.
343. Pall Corporation, *AcroSep™ MEP HyperCel™ Columns*, in *Mixed-Mode Chromatography Columns for Protein Separation*. 2008, Pall Corporation: Ann Arbor. p. 1-16.
344. Ahamed, T., et al., *Selection of pH-related parameters in ion-exchange chromatography using pH-gradient operations*. Journal of Chromatography A, 2008. **1194**(1): p. 22-29.
345. Bates, R.C. and D.D. Frey, *Quasi-linear pH gradients for chromatofocusing using simple buffer mixtures: local equilibrium theory and experimental verification*. Journal of Chromatography A, 1998. **814**(1-2): p. 43-54.
346. Dismer, F. and J. Hubbuch, *A novel approach to characterize the binding orientation of lysozyme on ion-exchange resins*. Journal of Chromatography A, 2007. **1149**(2): p. 312-20.
347. Burgos, I., S.A. Dassie, and G.D. Fidelio, *Thermodynamic Model for the Analysis of Calorimetric Data of Oligomeric Proteins*. Journal of Physical Chemistry B, 2008. **112**(45): p. 14325-14333.
348. Burgos, I., et al., *Thermodynamic and structural analysis of homodimeric proteins: model of beta-lactoglobulin*. Biochimica et Biophysica Acta, 2012. **1824**(2): p. 383-91.

349. Ahamed, T., et al., *pH-gradient ion-exchange chromatography: An analytical tool for design and optimization of protein separations*. Journal of Chromatography A, 2007. **1164**(1-2): p. 181-188.
350. Scheller, J., et al., *Production of spider silk proteins in tobacco and potato*. Nature Biotechnology, 2001. **19**(6): p. 573-577.
351. Theodossiou, T., et al., *Thermally induced irreversible conformational changes in collagen probed by optical second harmonic generation and laser-induced fluorescence*. Lasers in Medical Science, 2002. **17**(1): p. 34-41.
352. Zichel, R., et al., *Aptamers as a Sensitive Tool to Detect Subtle Modifications in Therapeutic Proteins*. PLoS One, 2012. **7**(2): p. 1-11.
353. Buyel, J.F., et al., *Extraction, purification and characterization of the plant-produced HPV16 subunit vaccine candidate E7 GGG*. Journal of Chromatography B: Biomedical Sciences and Applications, 2012. **880**(1): p. 19-26.
354. Winkelkemper, T. and G. Schembecker, *Purification performance index and separation cost indicator for experimentally based systematic downstream process development*. Separation and Purification Technology, 2010. **72**(1): p. 34-39.
355. Leser, E.W., M.E. Lienqueo, and J.A. Asenjo, *Implementation in an expert system of a selection rationale for purification processes for recombinant proteins*. Annals of the New York Academy of Sciences, 1996. **782**: p. 441-55.
356. Asenjo, J.A., J. Parrado, and B.A. Andrews, *Rational design of purification processes for recombinant proteins*. Annals of the New York Academy of Sciences, 1991. **646**: p. 334-56.
357. Nfor, B.K., et al., *Model-based rational strategy for chromatographic resin selection*. Biotechnology Progress, 2011. **27**(6): p. 1629-1643.
358. Song, M.H., et al., *Prediction of protein retention times in anion exchange chromatography systems using support vector machine regression*. Abstracts of Papers of the American Chemical Society, 2002. **224**: p. U507-U508.
359. Weiner, P.K., et al., *Electrostatic Potential Molecular-Surfaces*. Proceedings of the National Academy of Sciences, USA, 1982. **79**(12): p. 3754-3758.
360. ChemicalComputingGroup, *Molecular Operating Environment (MOE)*, in *Molecular Operating Environment (MOE)*. 2011, Chemical Computing Group Inc.: Montreal, QC, Canada.
361. Audry, E., et al., *A New Approach of Structure-Activity-Relationships - the Potential of Molecular Lipophily*. European Journal of Medicinal Chemistry, 1986. **21**(1): p. 71-72.
362. Heiden, W., G. Moeckel, and J. Brickmann, *A New Approach to Analysis and Display of Local Lipophilicity Hydrophilicity Mapped on Molecular-Surfaces*. Journal of Computer-Aided Molecular Design, 1993. **7**(5): p. 503-514.
363. Gasteiger, J. and M. Marsili, *Iterative Partial Equalization of Orbital Electronegativity - a Rapid Access to Atomic Charges*. Tetrahedron, 1980. **36**(22): p. 3219-3228.
364. Baroni, M., et al., *A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for ligands and proteins (FLAP): Theory and application*. Journal of Chemical Information and Modeling, 2007. **47**(2): p. 279-294.
365. Cruciani, G., et al., *Molecular fields in quantitative structure-permeation relationships: the VolSurf approach*. Journal of Molecular Structure: THEOCHEM, 2000. **503**(1-2): p. 17-30.
366. Oliva, A., M. Llabres, and J.B. Farina, *Applications of multi-angle laser light-scattering detection in the analysis of peptides and proteins*. Current Drug Discovery Technologies, 2004. **1**(3): p. 229-42.
367. Mogridge, J., *Using light scattering to determine the stoichiometry of protein complexes*. Methods in Molecular Biology, 2004. **261**: p. 113-8.

368. Wang, T. and J.A. Lucey, *Use of multi-angle laser light scattering and size-exclusion chromatography to characterize the molecular weight and types of aggregates present in commercial whey protein products*. Journal of Dairy Science, 2003. **86**(10): p. 3090-101.
369. Wen, J., T. Arakawa, and J.S. Philo, *Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions*. Analytical Biochemistry, 1996. **240**(2): p. 155-66.
370. Chennamsetty, N., et al., *Design of therapeutic proteins with enhanced stability*. Proceedings of the National Academy of Sciences, USA, 2009. **106**(29): p. 11937-42.
371. Chennamsetty, N., et al., *Prediction of protein binding regions*. Proteins, 2011. **79**(3): p. 888-97.
372. Ali, M.H. and B. Imperiali, *Protein oligomerization: How and why*. Bioorganic & Medicinal Chemistry, 2005. **13**(17): p. 5013-5020.
373. Evans, S.T., et al., *The effect of feed composition on the behavior of chemically selective displacement systems*. Journal of Chromatography A, 2010. **1217**(8): p. 1249-1254.
374. Freed, A.S., S. Garde, and S.M. Cramer, *Molecular Simulations of Multimodal Ligand-Protein Binding: Elucidation of Binding Sites and Correlation with Experiments*. Journal of Physical Chemistry B, 2011. **115**(45): p. 13320-13327.
375. Dismer, F., M. Petzold, and J. Hubbuch, *Effects of ionic strength and mobile phase pH on the binding orientation of lysozyme on different ion-exchange adsorbents*. Journal of Chromatography A, 2008. **1194**(1): p. 11-21.
376. Teske, C.A., et al., *Competitive adsorption of labeled and native protein in confocal laser scanning microscopy*. Biotechnology and Bioengineering, 2006. **95**(1): p. 58-66.
377. Miller, S., et al., *The Accessible Surface-Area and Stability of Oligomeric Proteins*. Nature, 1987. **328**(6133): p. 834-836.
378. Goodsell, D.S. and A.J. Olson, *Structural symmetry and protein function*. Annual Review of Biophysics and Biomolecular Structure, 2000. **29**: p. 105-153.
379. Berman, H.M., et al., *Past, present, and future of the protein data bank as an enabling resource*. Abstracts of Papers of the American Chemical Society, 2000. **219**: p. U112-U112.
380. Furtado, A., R.J. Henry, and F. Takaiwa, *Comparison of promoters in transgenic rice*. Plant biotechnology journal, 2008. **6**(7): p. 679-93.
381. Ranjan, R., et al., *Efficient chimeric promoters derived from full-length and sub-genomic transcript promoters of Figwort mosaic virus (FMV)*. Journal of Biotechnology, 2011. **152**(1-2): p. 58-62.
382. Harms, J., et al., *Defining process design space for biotech products: Case study of Pichia pastoris fermentation*. Biotechnology Progress, 2008. **24**(3): p. 655-662.
383. Rathore, A.S., *Roadmap for implementation of quality by design (QbD) for biotechnology products*. Trends in Biotechnology, 2009. **27**(9): p. 546-553.
384. Lebrun, P., et al., *Development of a new predictive modelling technique to find with confidence equivalence zone and design space of chromatographic analytical methods*. Chemometrics and Intelligent Laboratory Systems, 2008. **91**(1): p. 4-16.
385. Anonymous, *Pharmaceutical Quality for the 21st Century A Risk-Based Approach Progress Report*, D.o.H.a.H. Services, Editor. 2007, U.S. Food and Drug Administration: Silver Spring.
386. Doblhoff-Dier, O. and R. Bliem, *Quality control and assurance from the development to the production of biopharmaceuticals*. Trends in Biotechnology, 1999. **17**(7): p. 266-270.
387. Beutel, S. and S. Henkel, *In situ sensor techniques in modern bioprocess monitoring*. Applied microbiology and biotechnology, 2011. **91**(6): p. 1493-1505.

388. Bluma, A., et al., *In-situ imaging sensors for bioprocess monitoring: state of the art*. Analytical and Bioanalytical Chemistry, 2010. **398**(6): p. 2429-2438.
389. Justice, C., et al., *Process control in cell culture technology using dielectric spectroscopy*. Biotechnology Advances, 2011. **29**(4): p. 391-401.
390. Buyel, J.F. and R. Fischer, *Processing heterogeneous biomass: overcoming the hurdles in model building*. Bioengineered, 2013. **4**(1): p. 21-24.

IX. Appendix

IX.1 List of publications

1. Buyel JF, Fischer R. 2014. Flocculation increases the efficacy of depth filtration during the downstream processing of recombinant pharmaceutical proteins produced in tobacco. *Plant Biotechnology Journal* 12(2):240-52.
2. Buyel JF, Fischer R. 2014. Generic chromatography-based purification strategies accelerate the development of downstream processes for biopharmaceutical proteins produced in plants. *Biotechnology Journal* In Press.
3. Buyel JF, Fischer R. 2014. Scale-down models to optimize a filter train for the downstream purification of recombinant pharmaceutical proteins produced in tobacco leaves. *Biotechnology Journal* 9(3):415-25.
4. Buyel JF, Woo JA, Cramer SM, Fischer R. 2013. The use of quantitative structure-activity relationship models to develop optimized processes for the removal of tobacco host cell proteins during biopharmaceutical production. *Journal of Chromatography A* 1322:18-28.
5. Buyel JF., Hubbuch J, Fischer R. 2016. Comparison of Tobacco Host Cell Protein Removal Methods by Blanching Intact Plants or by Heat Treatment of Extracts. *Journal of Visualized Experiments*. 2016, e54343, 9.

IX.2 Register of equipment

Name	Type/Use	Manufacturer
-20°C premium	-20°C freezer	Liebherr, Germany
---	phytotron	Ilka Zell, Germany
0.2 and 0.45 µm filter	filter	Carl Roth GmbH, Germany
1.5 and 2.0 mL tubes	reaction tubes	Sarstedt, Germany
15 mL and 50 mL tubes	reaction tubes	Greiner Bio-One, Austria
2.0 mL cryotube	cryo tube	Carl Roth GmbH, Germany
2720 Thermal cycler	PCR cycler	Applied Biosystems, CA, USA
96 half area flat bottom black	96 well plate	Greiner Bio-One, Austria
ÄKTA explorer	chromatography device	GE Healthcare, UK
ÄKTA purifier	chromatography device	GE Healthcare, UK
Allegra 25R	centrifuge	BeckmanCoulter, CA, USA
Amicon 15	concentrator tube	Millipore, MA, USA
Aquarius	deionized water supply device	membraPure, Germany
Biophotometer	photometer	Eppendorf, Germany
BioWizard	sterile bench	Kojair, Finland
BP 121 S	scale	Sartorius, Germany
BP 610	scale	Sartorius, Germany
Cellstar	96 well plate	Greiner Bio-One, Austria
Centrifuge 5415D	centrifuge	Eppendorf, Germany
Commercial Blender	blender	Warring, CT, USA
Cond 315i	conductometer	WTW, Germany
Forma -86C ULT freezer	-80°C freezer	ThermoFisher, MA, USA
HiTrap column	pre-packed IEC columns	GE Healthcare, UK
Innova 4230	incubator/shaker	New Brunswick Scientific, CT, USA
KMO 2 basic	stirrer	IKA, Germany
M-Power	scale	Sartorius, Germany
Masterflex	peristaltic pump	Cole-Parmer, IL, USA
Masterflex L/S	peristaltic pump	Cole-Parmer, IL, USA
Mikro 220R	centrifuge	Hettich, Germany
MiniGyroRocker SSM3	rocker	Barloworld Scientific, UK
MiniSubCell GT	gel electrophoresis chamber	BioRad, CA, USA
MiraCloth 1R	filter tissue	Merck, Germany
Multiporator	electroporation device	Eppendorf, Germany
N816	vacuum pump	KNF, Germany
NanoDrop ND-1000	spectrometer	peqlab, Germany
pH 340i	pH meter	WTW, Germany
Polytron PT3000	mixer	Kinematica, Switzerland
Polytron PT6100	mixer	Kinematica, Switzerland
PowerPac300	DC source	BioRad, CA, USA
PowerPacBasic	DC source	BioRad, CA, USA
Premium	refrigerator	Liebherr, Germany
PVDF membrane	blotting membrane	Millipore, MA, USA
Reactron RT50	mixer	Kinematica, Switzerland
RTC basic	stirrer	IKA, Germany
RZR1	stirrer	Heidolph, Germany
SenTix 41	pH electrode	WTW, Germany
Slice200	cross-flow device	Sartorius, Germany
Synergy HT	96-well spectrometer	BioTek, VT, USA
Thermomixer compact	temperature-controlled mixer	Eppendorf, Germany
Type 6732-61	mixer	Jungheinrich, Germany
Universal Hood II	gel scanning device	BioRad, CA, USA
Varioklav	autoclave	H+P, Germany
Vortex-Genie 2	vortex	Scientific Industries, IL, USA
Whatman paper	blotting paper	Whatman Inc., UK
XCell sure lock	electrophoresis chamber	Invitrogen, CA, USA
XCell II	blot module	Invitrogen, CA, USA

IX.3 List of chemicals

Name	Type/Use	Manufacturer
Acetosyringone.....	phytohormone	Duchefa, The Netherlands
Carbenicilin	antibiotic.....	Duchefa, The Netherlands
Catofast GM	flocculant.....	BASF, Germany
Catofast VSH.....	flocculant.....	BASF, Germany
CelluFluxx F15.....	additive.....	Erbslöh, Germany
CelluFluxx F25.....	additive.....	Erbslöh, Germany
CelluFluxx F45.....	additive.....	Erbslöh, Germany
CelluFluxx P30.....	additive.....	Erbslöh, Germany
Citric acid	buffer	Carl Roth GmbH, Germany
Dipotassium hydrogen phosphate...	buffer component	Carl Roth GmbH, Germany
Disodium hydrogen phosphate	buffer component	Carl Roth GmbH, Germany
EDTA	buffer component	Carl Roth GmbH, Germany
Ethanol	solution component	Carl Roth GmbH, Germany
Ferty2Mega	fertilizer	Kammlott, Germany
Goat α -human H+L AP.....	antibody.....	Dianova, Germany
Goat α -mouse Fc AP.....	antibody.....	Jackson, UK
Goat α -rabbit H+L AP	antibody.....	Jackson, UK
Kanamycin	antibiotic.....	Duchefa, The Netherlands
Lupamin 9095.....	flocculant.....	BASF, Germany
Lupasol PS.....	flocculant.....	BASF, Germany
LuvoZell C200	additive.....	Lehmann & Voss, Germany
LuvoZell C90	additive.....	Lehmann & Voss, Germany
Magnafloc LT 37.....	flocculant.....	BASF, Germany
Magnafloc LT 38.....	flocculant.....	BASF, Germany
Methanol.....	solution component	Carl Roth GmbH, Germany A
Murashige & Skoog salts.....	solution component	Duchefa, The Netherlands
Paragas	flocculant.....	BASF, Germany
Polymin P	flocculant.....	BASF, Germany
Polymin SK	flocculant.....	BASF, Germany
Polymin VT	flocculant.....	BASF, Germany
Potassium chloride	buffer component	Carl Roth GmbH, Germany
Potassium dihydrogenphosphate ...	buffer component	Carl Roth GmbH, Germany
Praestol 2350.....	flocculant.....	Ashland, KY, USA
Praestol 2610.....	flocculant.....	Ashland, KY, USA
Praestol 2640.....	flocculant.....	Ashland, KY, USA
Praestol 822 BS	flocculant.....	Ashland, KY, USA
Praestol 851 BC.....	flocculant.....	Ashland, KY, USA
Praestol 855 BS	flocculant.....	Ashland, KY, USA
Praestol 859 BS	flocculant.....	Ashland, KY, USA
Rabbit α -DsRed.....	antibody.....	MBL, MA, USA
Rabbit α -His.....	antibody.....	Genscript, NJ, USA
Rifampicin	antibiotic.....	Duchefa, The Netherlands
Sedipur CL 950	flocculant.....	BASF, Germany
Sedipur CL 951	flocculant.....	BASF, Germany
Sodium acetate	buffer component	Carl Roth GmbH, Germany
Sodium chloride	buffer component	Carl Roth GmbH, Germany
Sodium dihydrogenphosphate	buffer component	Carl Roth GmbH, Germany
Sodium disulfite	antioxidant.....	Carl Roth GmbH, Germany
Sodium hydroxide	base	Carl Roth GmbH, Germany
Sucrose	buffer component	Duchefa, The Netherlands
Tris base	buffer	Carl Roth GmbH, Germany
Tris-HCl	buffer	Carl Roth GmbH, Germany
Trisodium phosphate	buffer	Carl Roth GmbH, Germany
Tween-20.....	non-ionic detergent.....	Carl Roth GmbH, Germany
ZETAG 7109.....	flocculant.....	BASF, Germany
ZETAG 7587.....	flocculant.....	BASF, Germany

IX.4 List of buffers

Name	Component	Final concentration [mM] ([g L ⁻¹])	Comment	
2-fold infiltration medium	Sucrose	300.0 (100.0)	pH 5.6	
	Glucose	20.0 (3.6)		
	Murashige & Skoog salts	--- (8.6)		
	Acetosyringone	0.2 (0.04)		
AP Buffer	Tris	100.0 (12.1)	pH 9.6	
	NaCl	100.0 (5.8)		
	MgCl ₂	5.0 (0.48)		
Blotting Buffer	Tris	25.0 (3.0)	pH 8.2	
	Glycine	192.0 (14.4)		
	Methanol	--- (160.0)		
Extraction Buffer A	Na ₂ HPO ₄	50 (7.1)	pH 6.0-8.0*	
	NaCl	10-500* (0.6-28.9)		
	NaS ₂ O ₅	10 (1.9)		optional
Extraction Buffer B	Trisodium citrate	50 (12.9)	pH 4.0-6.0*	
	NaCl	10-500* (0.6-28.9)		
	NaS ₂ O ₅	10 (1.9)		optional
HBS-EP+	HEPES	10.0 (2.4)	pH 7.4	
	EDTA	3.0 (0.9)		
	NaCl	150.0 (8.8)		
	Tween-20	--- (0.5)		
IEX Buffer A and HIC Buffer B	Na ₂ HPO ₄	15-50 (2.1-7.1)	pH 6.0-8.5*	
IEX Buffer B and HIC Buffer A	Na ₂ HPO ₄ NaCl	15-50 (2.1-7.1) 1000.0 (58.4)	pH 6.0-8.5*	
Lysogeny broth (LB)	Tryptone	--- (10.0)	pH 7.0	
	Yeast extract	--- (5.0)		
	NaCl	170.0 (10.0)		
	Agar	--- (15.0)		optional
	Ampicillin	0.13 (0.05)		optional
MES Buffer	MES	50.0 (9.76)	pH 7.3	
	Tris	50 (6.05)		
	SDS	3.5 (1.0)		
	EDTA	1.0 (0.3)		

List of buffers continued

Name	Component	Final concentration [mM] ([g L ⁻¹])	Comment
PBS(-T)	NaCl	137.0 (8.0)	pH 7.4
	KCl	2.7 (0.2)	
	Na ₂ HPO ₄	10.1 (1.44)	
	KH ₂ PO ₄	1.7 (0.24)	
	Tween-20	--- (1.0)	optional
Yeast extract broth (YEB)	Beef extract	--- (5.0)	pH 7.0
	Yeast extract	--- (1.0)	
	Peptone	--- (5.0)	
	Sucrose	14.5 (5.0)	
	MgSO ₄	2.0 (0.5)	
	Agar	--- (15.0)	optional
	Carbenicillin	0.13 (0.05)	optional
	Kanamycin	0.05 (0.025)	optional
	Rifampicin	0.03 (0.025)	optional

* values varied according to experimental/DoE setup

IX.5 List of identified tobacco HCPs

Protein name	UniProt accession number	Found in band ...	Seq. coverage [%]	Seq. identity [%]	Qmean Z [-]	Qmean score 4 [-]	Frac. of TSP [-]	Cellular function [-]
(E)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase	Q6RI20	7	93.64	19.47	-10.16	0.18	0.64	m
14-3-3 g-1 protein	Q947K7	15	93.65	78.81	-1.31	0.69	0.24	d
14-3-3-like protein C	P93343	1	90.38	97.87	-1.05	0.72	0.56	d
24K germin like protein	Q7XZV3	19	100.00	34.10	-3.98	0.49	0.42	m
30S ribosomal protein S5. chloroplastic	P93014	18	74.41	24.88	-4.96	0.41	0.15	p
30S ribosomal protein S7. chloroplastic	P62732	17	99.35	95.46	-2.20	0.60	0.18	p
30s ribosomal protein S8	P06363	25	100.00	85.82	-1.79	0.63	0.25	p
31 kDa ribonucleoprotein. chloroplastic	D6PZY7	23	38.12	47.06	-1.03	0.66	0.22	p
40S ribosomal protein S5	O24111	17	100.00	95.46	-4.33	0.41	0.18	p
40s ribosomal protein S8	Q3HRZ6	15	95.54	52.94	-4.10	0.49	0.24	p
50S ribosomal protein L10. chloroplastic	O80362	18	96.11	31.07	-2.35	0.60	0.15	p
50S ribosomal protein L12. chloroplastic	P24929	18	93.98	46.09	-1.22	0.68	0.15	p
50S ribosomal protein L3. chloroplastic	O80360	16	94.14	42.86	-2.16	0.63	0.34	p
60S ribosomal protein L12	Q6RJY1	18	77.11	92.19	-3.20	0.50	0.15	p
60S ribosomal protein L13a-like protein	Q3HRW1	17	100.00	85.44	-2.32	0.62	0.18	p
60S ribosomal protein L5-A	Q38HU5	14	96.36	73.20	-6.95	0.33	0.24	p
Actin	C7F8N2	23	100.00	88.89	0.38	0.81	0.22	c
Actin 7	Q6F4H4	23	98.94	90.35	-0.92	0.71	0.22	c
ADP ribosylation factor	Q08IJ1	19	97.79	94.35	0.32	0.82	0.42	s
Alanine:glyoxylate aminotransferase	C1IGP4	10	91.52	28.42	-3.28	0.57	0.69	m
Allene oxide cyclase	Q711R1	17	94.02	67.82	-0.50	0.75	0.18	m
Aminomethyltransferase. mitochondrial	P54260	10	98.41	50.81	-0.55	0.74	0.69	m
Anthocyanidin synthase	A9ZMI6	21	74.73	77.65	0.17	0.79	---	m
AT4g29060/F19B15_90	Q9SZD6	2	---	---	---	---	0.15	p
ATP synthase CF1 epsilon subunit	P00834	25	100.00	61.65	-1.51	0.66	0.25	e
ATP synthase delta chain. chloroplastic	P32980	19	96.28	17.13	-6.42	0.28	0.42	e
ATPase subunit I	Q36600	25	37.04	15.00	-0.93	0.59	0.25	e
ATP-dependent Clp protease proteolytic subunit	P12210	18	90.31	44.07	-3.06	0.54	0.15	p
β -carbonic anhydrase	Q8W183	16	64.49	77.78	-2.14	0.62	0.34	m
Carbamoyl phosphate synthase large subunit	Q8L6J9	2	---	---	---	---	0.15	m
Carbonic anhydrase. chloroplastic	P27141	15; 16; 23	93.67	79.71	-1.98	0.63	0.79	m

CDSP32 protein	O04002	15	72.64	15.32	-3.57	0.53	0.24	m
List of identified tobacco HCPs continued								
Protein name	UniProt accession number	Found in band ...	Seq. coverage [%]	Seq. identity [%]	Qmean Z [-]	Qmean score 4 [-]	Frac. of TSP [-]	Cellular function [-]
Chaperonin 21	Q9M5A8	17	---	---	---	---	0.18	p
Chlorophyll a-b binding protein. chloroplastic	Q40481	19	82.50	92.12	-8.52	0.05	0.42	e
Chloroplast elongation factor TuB	P68158	16	97.80	72.14	0.20	0.78	0.34	p
Chloroplast photosynthetic oxygen-evolving protein 33 kDa subunit	Q40459	12	95.55	48.78	-4.30	0.47	0.28	e
Chloroplast pigment-binding protein CP29	Q0PWS7	14	80.00	33.19	-10.27	0.03	0.24	e
Chloroplast ribosomal protein	C9EFD1	15	---	---	---	---	0.24	p
Chloroplast sedoheptulose-1,7-bisphosphatase	C5IU71	13; 14	76.65	38.70	-3.35	0.56	0.99	m
ClpC	O48931	5	91.58	51.96	-2.73	0.60	0.33	p
Cytochrome b6	P06247	25	98.60	90.09	-7.80	0.21	0.25	e
Cytosolic aconitase	Q9FVE9	5	99.22	61.28	-1.03	0.70	0.33	m
Dicarboxylate/tricarboxylate carrier	Q8SF02	23	93.67	23.65	-7.90	0.23	0.22	m
Elicitor-inducible LRR receptor-like protein	Q9SLS3	7	79.55	27.83	-3.98	0.53	0.64	s
Elongation factor EF-2	Q9SGT4	5	99.53	60.59	-1.82	0.65	0.33	p
Elongation factor Tu	B5JK07	9	98.23	71.98	0.45	0.79	0.79	p
Elongation factor Tu. chloroplastic	P68158	9	97.80	72.14	0.20	0.78	0.79	p
Eukaryotic initiation factor 4A-10	P41382	9	90.56	70.13	-1.05	0.70	0.79	p
Eukaryotic initiation factor 4A-9	Q40471	9	90.56	70.40	-1.11	0.70	0.79	p
Eukaryotic translation initiation factor 3 subunit	B9RVA6	2	---	---	---	---	0.15	p
Eukaryotic translation initiation factor 5A-1	P69040	18	88.68	83.69	1.55	0.94	0.15	p
Ferredoxin-dependent glutamate synthase 1	Q7M242	1; 2	---	---	---	---	0.71	m
Ferredoxin-NADP reductase. leaf isozyme 1. chloroplastic	Q9FKW6	14	94.86	91.19	1.36	0.87	0.24	e
Fructose-1,6-bisphosphatase	Q9XF82	11; 12	82.11	88.06	-1.20	0.70	0.60	m
Fructose-bisphosphate aldolase-like protein	Q2PYX3	12	91.34	63.22	-0.65	0.74	0.28	m
GcpE	Q8GZR6	7	87.57	19.62	-8.59	0.27	0.64	m
Glutamate-1-semialdehyde 2,1-aminomutase. chloroplastic	P31593	9	97.47	72.34	0.01	0.77	0.79	m
Glutamine synthetase	Q42951	11	98.31	89.43	-0.62	0.70	0.32	m
Glutathione reductase. chloroplastic	P80461	9	95.11	45.53	-2.09	0.64	0.79	m

Glyceraldehyde-3-phosphate dehydrogenase A. chloroplastic	P09043	11; 12; 13; 14	100.00	89.29	-0.52	0.75	1.59	m
List of identified tobacco HCPs continued								
Protein name	UniProt accession number	Found in band ...	Seq. coverage [%]	Seq. identity [%]	Qmean Z [-]	Qmean score 4 [-]	Frac. of TSP [-]	Cellular function [-]
Glyceraldehyde-3-phosphate dehydrogenase B. chloroplastic	P09044	11; 12; 13; 14	88.31	92.35	-1.89	0.66	1.59	m
Glycine dehydrogenase [decarboxylating]. mitochondrial	O49954	4	---	---	---	---	0.77	m
Glycolate oxidase	O82077	12	96.50	90.50	0.66	0.79	0.28	m
Guanine nucleotide-binding protein subunit beta-like protein	P49026	14	97.55	79.31	-1.58	0.68	0.24	s
Haloacid dehalogenase-like hydrolase family protein	B9T3Q7	15	94.16	20.90	-3.62	0.52	0.24	m
Heat shock protein 70-3	Q67BD0	7	85.65	81.08	-0.02	0.75	0.64	r
Heat shock protein 90	Q14TB1	6	95.14	61.26	-4.45	0.50	0.28	r
Histone H2B	P93354	18	86.21	64.00	-2.59	0.55	0.15	d
Hsr203J	Q8H954	7	94.93	28.09	-3.46	0.56	0.64	r
Hydroxypyruvate reductase	A1EGU2	11	96.63	25.20	-4.54	0.49	0.32	m
Light-harvesting complex I protein Lhca2	Q40512	19	78.70	94.12	-8.34	0.06	0.42	e
Manganese superoxide dismutase	Q84QX4	17	98.53	57.21	-1.31	0.70	0.18	m
Methionine synthase	Q069K2	6	99.22	87.35	-2.04	0.64	0.28	m
MRNA-binding protein	Q7X998	13; 14	91.24	16.87	-3.92	0.52	0.99	p
Oxygen-evolving enhancer protein 2-1. chloroplastic	Q7DM39	17	91.94	70.76	-0.48	0.76	0.18	e
Peptidyl-prolyl cis-trans isomerase TLP20. chloroplastic	B2BF99	18	80.68	70.06	0.61	0.85	0.15	m
Phosphoenolpyruvate carboxylase	P27154	4	96.89	75.40	-2.07	0.64	0.77	m
Phosphoglycerate kinase. chloroplastic	Q42961	10	98.52	59.10	-0.42	0.74	0.69	m
Phosphoglycerate kinase. cytosolic	Q42962	10	97.76	61.83	0.54	0.80	0.69	m
Phosphoribulokinase	C3URC2	12	80.17	15.48	-5.84	0.38	0.28	m
Phosphoribulokinase. chloroplastic	P25697	11	81.66	15.39	0.00	0.00	0.32	m
Phylloplanin	Q56S59	16	---	---	---	---	0.34	r
Plastid transketolase	C3RXI5	7	97.34	87.41	0.00	0.00	0.64	m
Plastidic aldolase NPALDP1	Q9SXX4	12	99.43	53.82	-1.34	0.69	0.28	m
Presequence protease 1. chloroplastic/mitochondrial	Q9LJL3	3	98.29	99.90	-0.26	0.74	0.36	p
Presequence protease 2. chloroplastic/mitochondrial	Q8VY06	3	98.29	89.48	-0.23	0.74	0.36	p
Probable alanyl-tRNA synthetase. chloroplastic	Q9FFC7	5	76.51	24.84	-4.75	0.48	0.33	m

Protein thylakoid formation 1. chloroplastic	Q7XAB8	16	---	---	---	---	0.34	c
Putative beta 3 proteasome subunit	Q93X33	18	100.00	55.77	-1.93	0.62	0.15	p
List of identified tobacco HCPs continued								
Protein name	UniProt accession number	Found in band ...	Seq. coverag e [%]	Seq. identit y [%]	Qmea n Z [-]	Qmea n score 4 [-]	Frac. of TSP [-]	Cellular functio n [-]
Putative cytosolic cysteine synthase 7; oas1	Q3LAG6	13; 14	87.64	85.89	-0.37	0.76	0.99	m
Putative cytosolic cysteine synthase 7; oas7	Q3LAG5	13; 14	97.54	83.91	-0.88	0.73	0.99	m
Putative photosystem I subunit III precursor	Q84QE7	25	80.21	87.01	-5.58	0.29	0.25	e
Putative spindle disassembly related protein CDC48	Q1G0Z1	5	91.58	78.68	-1.10	0.69	0.33	c
Ribonucleoprotein	Q08940	16; 23	98.28	17.03	-3.67	0.52	0.55	d
Ribosomal protein L11-like	Q9FSF6	18	93.92	90.59	-4.53	0.41	0.15	p
Ribulose biphosphate carboxylase small chain S41. chloroplastic	P69249	20; 26	100.00	100.00	-0.49	0.76	9.62	e
Ribulose biphosphate carboxylase/oxygenase activase 1. chloroplastic	Q40460	10; 11; 13; 14; 15; 23	76.30	89.76	-1.79	0.66	2.46	e
Ribulose biphosphate carboxylase/oxygenase activase 2	Q40565	23	76.64	85.27	-1.91	0.65	0.22	e
Ribulose-1.5-biphosphate carboxylase/ oxygenase large subunit	P00876	2; 3; 8; 9; 16	96.21	99.78	-1.15	0.69	19.1 0	e
Ribulose-phosphate 3- epimerase. chloroplastic	Q43157	16; 17	97.87	91.30	0.34	0.78	0.52	m
Serine-glyoxylate aminotransferase	Q56YA5	10	95.51	30.08	-2.97	0.59	0.69	m
Thaliana 60S ribosomal protein L7	Q38HU4	15	97.93	77.64	-4.17	0.48	0.24	p
Thioredoxin peroxidase	Q8RVF8	17; 18	95.48	62.11	-0.94	0.72	0.33	m
TNF receptor-associated protein 1-like	B6V765	6	---	---	---	---	0.28	d
Translation elongation factors	B9TE78	6	68.37	27.82	-2.20	0.59	0.28	p
Translation initiation factor IF-2. chloroplastic	Q9SI75	6	98.15	58.35	-3.50	0.55	0.28	p
Translation initiation factor IF-2. chloroplastic	Q9SHI1	6	---	---	---	---	0.28	p

IX.6 List of proteins included in QSAR training sets

Protein name	PDB code	Resolution [Å]	Oligomeric state [-]
Adenosine deaminase	1vfl	1.8	2
Alkaline phosphatase	1ajc	2.5	2
Alpha-chymotrypsin a	5cha	1.6	3
Alpha-lactalbumin	1f6r	2.2	6
Asparaginase type II	3eca	2.4	4
Aspartate aminotransferase	1ajr	1.7	2
Avidin	1vyo	1.4	4
Beta-galactosidase	1f4h	2.8	4
Beta-lactoglobulin	1bsq	2.2	1
Calmodulin	1qiw	2.3	1
Carbonic anhydrase II	1v9e	1.9	1
Carboxylesterase	1auo	1.8	2
Catalase	4blc	2.3	4
Chymotrypsinogen a	2cga	1.8	2
Citramalate synthase	3ble	2.0	2
Cytochrome c	2b4z	1.5	1
Elastase	1lvy	1.8	1
Endoglucanase I	1egl	3.6	4
Ferritin	1ies	2.6	24
Glucoamylase	1lf6	2.1	1
Glucose oxidase	1cf3	1.9	1
Glutamate dehydrogenase 1	1nr7	3.3	6
Glycogen phosphorylase b	1gpb	1.9	2
Insulin (chain a)	4ins	1.5	12
Invertase	2ac1	2.1	1
Lactoferrin	1bka	2.4	1
Lipoxygenase-1	1f8n	1.4	1
Lysozyme	1aki	1.5	1
Ovalbumin	1ova	1.9	2
Ovotransferrin	1aiv	3.0	1
Papain	9pap	1.6	1
Peanut lectin	2pel	2.2	4
Pepsin	3pep	2.3	1
Phospholipase a2	1poc	2.0	2
Pyruvate kinase	1a49	2.1	4
Ribonuclease a	1rbx	1.6	1
Ribonuclease b	1rbb	2.5	2
Serum albumin	1ao6	2.5	1

List of proteins included in QSAR training sets continued

Protein name	PDB code	Resolution [\AA]	Oligomeric state [-]
Serum transferrin	1a8e	1.6	1
Triacyl-glycerol acylhydrolase	3tgl	1.9	1
Trypsin	1s81	1.7	1
Trypsin inhibitor	1pit	1.3	n.a.
Trypsinogen	1s0q	1.0	1
Ubiquitin	1ubq	1.8	1
Urease	1fwe	2.0	9

IX.7pH gradient buffers for IEX

Name	Component	Concentration [mM] ([g L ⁻¹])	pK _a
AEX buffer pH 10.5 to 3.5	Hydroxylamine 50%	7.7 (0.51)	5.67
	Methylamine 50%	9.8 (0.76)	10.75
	1-Methylpiperazine	6.4 (0.64)	9.19/4.82
	Ethylenediamine	9.1 (0.55)	9.93/6.99
	1,4-Dimethylpiperazine	13.7 (1.56)	8.15/4.04
	Bis-Tris	5.8 (1.21)	6.22
	NaCl	10 (0.58)	---
CEX buffer pH 4.0 to 11.0	MESx1H ₂ O	11 (2.35)	6.15
	Formic acid	9.9 (0.46)	3.75
	Acetic acid	13 (0.81)	4.76
	MOPSO	8.7 (1.96)	6.90
	HEPPSO	9.9 (2.66)	8.04
	TAPS	4.6 (1.13)	8.44
	CHES	9.4 (1.95)	9.39
	CAPS	15.6 (3.45)	10.50
	NaCl	10 (0.58)	---

Note: all pH buffer compositions according to Frieder Kröner (personal communication).