

Manufacturing biopharmaceutical proteins by transient expression in *Nicotiana tabacum* (L.)

**(Herstellung von biopharmazeutisch aktiven Proteinen durch transiente
Expression in *Nicotiana tabacum* (L.))**

**Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der
RWTH Aachen University zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften genehmigte Dissertation**

vorgelegt von

Johannes Felix Buyel, M.Sc.

aus Mönchengladbach

Berichter: Universitätsprofessor Dr. Rainer Fischer

Professor Dr. Jürgen Hubbuch

Tag der mündlichen Prüfung: 24.07.2013

Diese Dissertation ist auf den Internetseiten der Hochschulbibliothek online verfügbar

Acknowledgements

I would first like to thank my supervisor Prof. Dr. Rainer Fischer for giving me the opportunity to conduct my PhD studies in such an excellent research environment as the Institute for Molecular Biotechnology, RWTH Aachen University, in cooperation with the Fraunhofer Institute for Molecular Biology and Applied Ecology. I am also very grateful for his support and scientific input, which I could count on whenever needed. In this context I would like to acknowledge funding from the European Research Council Advanced Grant Future-Pharma, proposal number 269110.

I thank Prof. Hubbuch for accepting the position of second supervisor for my thesis and I am grateful to him and Prof. Steven M. Cramer for allowing me to visit their laboratories and gain insight into their fascinating research.

I would also like to thank Dr. Jürgen Drossard for his supervision, support and advice. I am grateful to him for setting out the rough framework of my thesis and allowing me to fill it with such a high degree of freedom of action.

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, providing guidance (especially in the early phase of my PhD) and for simply being great people to work and BBQ with. Stay as you are folks...I know you will =).

It was also a formidable experience and a great pleasure to supervise the Bachelors/Masters students Claudia Haase, Joschka Buyel and Thomas Kaefer. I thank you people for handling all my idiosyncrasies so well and for your great contribution to the whole effector/protein expression part as I also acknowledge in the body of the thesis (now you have to read it :p).

Also many thanks to the whole IPP crew for being the most helpful colleagues I could think of and for keeping the lab up and running.

Thank you Dr. Richard Twyman for checking my English ☺.

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 both!

"...essentially, all models are wrong, but some are useful..."

George E. P. Box

A. Table of contents

I.	Introduction	1
I.1	Diseases and healthcare	1
I.2	Biopharmaceuticals	1
I.3	Production of biopharmaceutical proteins	2
I.4	Plant-made pharmaceuticals	2
I.5	Protein expression strategies in plants	4
I.6	Plant–bacterial interactions.....	5
I.6.1	Economic considerations for downstream process design	7
I.7	Design of experiments	8
I.8	Aim of this thesis and workflow	12
II.	Materials and methods.....	14
II.1	Equipment and chemicals.....	14
II.2	Expression vectors and cloning	14
II.2.1	General cloning procedures and cultivation of bacteria	14
II.2.2	<i>Agrobacterium tumefaciens</i> infiltration of plants.....	14
II.2.3	pGFD vector	16
II.2.4	Promoter/5'UTR combinations for DsRed expression	16
II.2.5	Type III effectors	16
II.3	Plant growth.....	17
II.3.1	Plant species	17
II.3.2	Greenhouse	17
II.3.3	Post-infiltration incubation	17
II.4	Sample preparation and analysis	17
II.4.1	Sampling from infiltrated plants.....	17
II.4.2	Protein quantitation	17
II.4.3	SDS-PAGE analysis	18
II.4.4	Western blot and immunodetection.....	18
II.4.5	Trypan blue staining of dead cells.....	19
II.4.6	Callose staining	19
II.4.7	Microscopy	20
II.4.8	Metabolic analysis	20
II.5	Design of experiments	21
II.5.1	Transient protein expression model.....	21

II.5.2	Promoter strength model	22
III.	Results and discussion	23
III.1	Transient protein expression model	23
III.1.1	Model theory	23
III.1.1.1	Model concept.....	23
III.1.1.2	Plant parameters	23
III.1.1.3	Volume model.....	24
III.1.1.4	Concentration model	26
III.1.1.5	Yield model.....	29
III.1.1.6	Cost function	29
III.1.2	Model building and key factors	30
III.1.2.1	System consistency and repeatability	30
III.1.2.2	Volume model.....	32
III.1.2.3	Concentration model	33
III.1.2.4	Yield model.....	36
III.1.2.5	Model testing.....	37
III.1.2.6	Cost function	37
III.1.3	Model implications for a production process.....	39
III.2	Promoter and 5'UTR strength in transient expression	41
III.2.1	Vectors	41
III.2.2	Effects of promoter and 5'UTR	42
III.2.2.1	The predictive model captures promoter/5'UTR-dependent differences in expression	42
III.2.2.2	The optimal combination is the CaMV 35SS promoter and CHS 5'UTR, which achieved the highest level of DsRed accumulation within 8 days	44
III.2.2.3	DsRed accumulation rates change over time and are dependent on the promoter, 5'UTR and leaf age	47
III.2.2.4	ER localization increases the accumulation of DsRed over long incuba- tion times and in combination with the <i>nos</i> promoter.....	49
III.2.2.5	Different 5'UTRs form compatible pairs for stoichiometric protein ex- pression when combined with the CaMV 35SS and <i>nos</i> promoters	51
III.2.3	Implications for the design of expression cassettes	53
III.3	Transient expression of type III effectors	53
III.3.1	Vectors	53

III.3.2	Effects on plant tissue and defense responses	54
III.3.2.1	Reactions to bacterial injection.....	54
III.3.2.2	Altering reactions by the expression of effectors	58
III.3.3	Effects on co-expressed target proteins	59
IV.	Conclusion and scope	62
IV.1	Process design and control for plant-derived bio-pharmaceuticals	62
IV.2	The basis for a Quality-by-Design approach.....	62
V.	Summary.....	64
VI.	References	65
VII.	Appendix	79
VII.1	List of publications	79
VII.2	Register of equipment.....	80
VII.3	List of chemicals.....	81
VII.4	List of buffers	82

B. Abbreviations

(v/v).....	volume per volume
(w/v).....	weight per volume
ADP.....	adenosine diphosphate
Amp.....	ampicillin
API.....	active pharmaceutical ingredient
Avr.....	avirulence
BCIP.....	5-bromo-4-chloro-3-indolyl phosphate
BiP.....	binding protein
cAMP.....	cyclic adenosine monophosphate
CaMV.....	<i>Cauliflower mosaic virus</i>
CCD.....	central composite design
CDPK.....	calcium-dependent protein kinase
cGMP.....	current good manufacturing practice
CHO.....	Chinese hamster ovary
CHS.....	chalcone synthase
CI.....	confidence interval
CNV.....	<i>Cucumber necrosis virus</i>
C.V.....	coefficient of variation
d.....	days
Da.....	Dalton
DCF.....	downstream cost factor
df.....	degree of freedom
DNA.....	deoxyribonucleic acid
DoE.....	design of experiments
dpi.....	days post injection
dps.....	days post seeding
DSP.....	downstream processing
EDC.....	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA.....	ethylenediaminetetraacetic acid
ELISA.....	enzyme-linked immunosorbent assay
EP.....	polysorbate-20 (Tween-20)
ER.....	endoplasmic reticulum
ETI.....	effector-triggered immunity
FDA.....	Food and Drug Administration
FMEA.....	failure mode and effects analysis
FPLC.....	fast protein liquid chromatography
GCF.....	growth cost factor
GOI.....	gene of interest
h.....	hours
HBS.....	HEPES-buffered saline
HEPES.....	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Hop.....	Hrp outer protein
Hrp.....	hypersensitive response and pathogenicity
IC.....	infiltration costs
IEF.....	isoelectric focusing
IgA.....	immunoglobulin A
IgG.....	immunoglobulin G
IRES.....	internal ribosome entry site
Kan ^r	kanamycin (resistant)

LacI.....	<i>lac</i> -operon repressor protein
LB.....	left border (vector description)
LB.....	lysogeny broth
LDS.....	lithium dodecylsulfate
LMO.....	leave-many-out
LOF.....	lack-of-fit
LPH.....	leader peptide sequence of heavy chain from murine mAb24
MAPK.....	mitogen-activated protein kinase
MES.....	2-(N-morpholino)ethanesulfonic acid
NB-LRR.....	nucleotide binding–leucine-rich repeat
NBT.....	nitroblue tetrazolium
NEB.....	New England Biolabs
NHS.....	N-hydroxysuccinimide
nos.....	nopaline synthase
OD _{600nm}	optical density at 600 nm
OFAT.....	one factor at a time
Pa.....	Pascal
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
PCD.....	programmed cell death
PCR.....	polymerase chain reaction
pH.....	negative decadic logarithm of H ⁺ concentration
POI.....	protein of interest
polyA.....	polyadenylate
PRESS.....	predicted residual sum of squares
PRR.....	pattern recognition receptor
PTI.....	pathogen-triggered immunity
PTM.....	post-translational modification
PVPP.....	polyvinylpyrrolidone
QbD.....	quality by design
RB.....	right border (vector description)
RE.....	restriction endonuclease
RF.....	recovery factor
ROS.....	reactive oxygen species
RP-HPLC.....	reversed-phase high-performance liquid chromatography
rpm.....	revolutions per minute
RPN.....	risk priority number
R-protein.....	resistance protein
RSM.....	response surface methodology
SA.....	salicylic acid
SAR.....	scaffold attachment region
SARS.....	severe acute respiratory syndrome
SDS.....	sodium dodecylsulfate
sIgA.....	secretory IgA
SPR.....	surface plasmon resonance
SRP.....	signal recognition particle
TBSV.....	<i>Tomato bushy stunt virus</i>

T-DNA	transfer DNA
TEV	<i>Tobacco etch virus</i>
TF	transcription factor
TL	TEV leader sequence
TLF	two-level factorial
TMV	<i>Tobacco mosaic virus</i>
Tris	tris (hydroxymethyl)aminomethane
TSP	total soluble protein
TTE	type III effector
TTSS	type III secretion system
USP	upstream processing
UTR	untranslated region
YEB	yeast extract broth

I. Introduction

I.1 Diseases and healthcare

Pathogens have co-evolved with their hosts in a so called ‘arms race’ [1]. Mutations in either the host or pathogen are favored if they confer improved resistance or pathogenicity [2]. After each mutation, a new equilibrium between host and pathogen is established. From the beginning of society, humans have tried to shift this equilibrium in their favor using potions and herbs [3]. The motivation to do so was the improvement in health through the absence of disease, which has ultimately led to modern medicine, pharmacology and healthcare systems.

I.2 Biopharmaceuticals

In pre-industrialized times, drugs were often prepared as a crude mixture of substances isolated from natural biological sources such as plants [4, 5]. These were the earliest biopharmaceuticals. With the emergence of natural sciences in the 18th and 19th centuries, the concept that drugs must contain a certain active pharmaceutical ingredient (API) was developed, whereupon traditional drugs fell into disrepute because it was impossible to identify specific APIs in their formulations [5]. Many subsequent drugs therefore contained chemically-synthesized molecules such as acetylsalicylic acid or sulfonamides as APIs until the mid-20th century [6-8]. Biopharmaceuticals were re-established as scientifically-sound drugs from the late 19th to mid-20th century e.g. by the works of Emil von Behring (diphtheria antitoxin) and Alexander Fleming (antibiotics) [9-12]. The advantage of biopharmaceuticals over chemically-synthesized APIs is that their structure can be much more complex, allowing complex disease mechanisms to be targeted [13-18]. This is because biopharmaceutical production uses 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].

Biopharmaceuticals can be divided into two major groups. The first consists of small molecules (often secondary metabolites) that are synthesized naturally by cells [9, 21, 22]. Metabolic engineering allows alternative hosts to produce the same molecules. For example, paclitaxel (Taxol) is isolated from *Taxus brevifolia* (Pacific yew tree) and is used to treat breast cancer, but it can also be produced in optimized plant cell cultures [13]. The second group consists of proteins, and since most medically-relevant proteins are from humans they are usually produced in genetically-engineered cells or organisms.

I.3 Production of biopharmaceutical proteins

The first recombinant biopharmaceutical proteins (such as insulin) were produced in the bacterium *Escherichia coli* [23, 24]. However, bacteria have only a limited potential for post-translational modifications (PTMs) such as phosphorylation and glycosylation, which are necessary for the functional efficacy of many human proteins [25]. Bacteria can also produce endotoxins that pose a serious risk to human health (e.g. inducing septic shock) when present in the final drug product [26]. Therefore, many recombinant proteins are now produced in mammalian cell cultures [24]. The most prominent examples are monoclonal antibodies, which are often produced in Chinese hamster ovary (CHO) cells with a median sales price of \$US 8000 per gram and a total market value of more than \$US 10 billion [27-30]. Even so, mammalian cells suffer disadvantages including the potential for contamination with human pathogens such as retroviruses (requiring sterile conditions to be maintained at all times), the expensive sterile cultivation equipment including stainless-steel fermenters and expensive media [31]. The up-front costs are therefore high despite the trend towards disposable technologies aiming to replace most of the fixed-process installations [32].

I.4 Plant-made pharmaceuticals

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 [33-35]. Several characteristics have to be considered when selecting a plant species and tissue for production [36, 37]. Cereal crops, such as wheat (*Triticum* L. spp.), barley (*Hordeum vulgare* L.) or maize (*Zea mays* L.), allow the long-term storage of recombinant proteins in seeds but the potential exists to contaminate the human food chain [38, 39]. *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 [40]. 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 [35] although since then many other pharmaceutical proteins have been produced in a diverse range of plants [41].

Plants offer numerous advantages for the production of biopharmaceutical proteins including the low cost of upstream production, the potential for large-scale cultivation, and inherent safety reflecting the inability of human pathogens to replicate in plants [31, 36-38, 42]. Additionally, plant cells can assemble oligomeric proteins correctly and introduce PTMs

[43-45]. The glycosylation of proteins in plants can be controlled to improve API efficacy and reduce immunogenicity in humans [46-48].

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). There are many technological and economic reasons for this delay [48, 49]. One technological issue is the low and variable yields of recombinant protein in many plant-based platforms (compared to established systems such as CHO cells) resulting in batch-to-batch variation [41, 50]. This may be caused by gene silencing, particularly when two identical promoters are used to drive the expression of different genes, such as those encoding the heavy and light chains of a monoclonal antibody [51]. Many strategies have been proposed to prevent homology-based gene silencing including the use of synthetic promoters and the generation of combinatorial diversity by shuffling different elements from several promoters [52, 53]. The 5'UTR sequences should be included in these strategies because sequence repeats shorter than 50 bp may still be sufficient to induce gene silencing [51]. The 5'UTRs are known to affect protein synthesis, so they are also interesting candidates to improve and/or balance the expression of recombinant proteins in their own right [54-59]. Therefore, promoter/5'UTR combinations not only affect the level of recombinant protein expression in plants but may also play a crucial role in the quality and consistency of a biopharmaceutical product. Promoter/5'UTR combinations should therefore be evaluated before designing such a process.

Mechanistic models have been established for certain promoters in bacteria [60], yeast [61, 62] and *Drosophila melanogaster* embryos [63, 64]. For single-celled organisms these models often describe promoter activity in terms of their dependence on one or more regulatory molecules (e.g. lactose, LacI and cAMP in case of the *lac* operon) [60]. Models for eukaryotic promoters require more precise knowledge about the transcription factors (TFs) involved and their spatiotemporal expression profiles [63, 64]. However, the size and complexity of tobacco plants makes it difficult to determine the distribution of TFs and the costs of mechanistic modeling are likely to outweigh the benefits. Therefore, the use of descriptive rather than mechanistic models seems more appropriate. These models should not only consider the promoter and 5'UTR as factors, but should include parameters such as leaf age and incubation time because these factors also affect recombinant protein expression and will thus improve the model quality if included in the experimental setup [65-68]. However, previous studies have either focused on promoter analysis alone [69, 70] or have used cell suspension cultures to compare different sequences [54, 56, 57, 71-73]. Some studies

considered the temporal accumulation of a reporter protein in seeds or buds rather than leaves [74, 75]. Others did not resolve differences among leaves of different ages when analyzing promoters [76, 77] or investigated expression in chloroplasts [78]. A model is therefore required that links the impact of promoters and 5'UTRs on protein accumulation on one hand with the spatiotemporal constraints of transient protein expression in whole plants on the other. Such a model would allow the rational evaluation of promoter/5'UTR combinations in terms of their suitability for the reproducible, high-level or balanced expression of one or several recombinant proteins, thus improving batch-to-batch reproducibility and regulatory compliance.

1.5 Protein expression strategies in plants

There are two main strategies for the production of recombinant proteins in plants: stable expression (transgenic plants) and transient expression [37, 79]. Transgenic plants can be generated by the introduction of exogenous DNA into plant cells using particle bombardment and other direct transfer methods, or biological delivery with a natural pathogen such as *Agrobacterium tumefaciens*, in each case followed by selection for stable integration events [80]. To facilitate selection, the introduced DNA includes a marker gene such as neomycin phosphotransferase (*nptII*) which provides resistance to aminoglycoside antibiotics [81]. Transformed cells are usually selected in callus culture, followed by regeneration into fertile plants [82]. These are self-pollinated for several generations to produce homozygous transgenic plants which will not segregate the transgene nor lose it due to genetic instability [83, 84]. The duration of this process depends on the generation time of the plant species, but about 24 months are required when using the Petit Havana SR1 variety of tobacco, assuming five rounds of self-pollination (T_5) [37]. The advantages of stable transgenic plants are the reproducible transgene expression levels, the defined master seed bank and the ease of scale-up [37, 49, 85].

Transient expression allows the rapid production of large amounts of recombinant protein because wild-type plants are used, avoiding the time-consuming transformation, regeneration and self-pollination steps [66, 85-87]. Transient expression is achieved by injecting or vacuum infiltrating leaves with *A. tumefaciens* carrying either a standard transfer DNA (T-DNA) expression construct or a hybrid construct containing elements from plant viruses [88]. The T-DNA containing the gene of interest (GOI) is then exported to nearby plant cells via a type IV secretion system [89], and imported into the nucleus of recipient cells by the combined activity of plant and bacterial proteins, so that the GOI is expressed resulting

in the synthesis of the recombinant protein (protein of interest, POI) [88]. Then plants, or parts thereof, are harvested after an incubation period of typically 3–8 days [67, 90-92]. Using this method, it is possible to obtain the recombinant protein less than two months after the DNA construct is provided, and within 3 weeks if wild-type tobacco plants of a suitable age are always kept ready [93]. The ability to scale up production rapidly is a significant advantage over both transgenic plants and conventional production platforms allowing the provision of vaccines in response to sudden epidemic outbreaks, e.g. of influenza or SARS. Transient expression also allows the production of proteins that are toxic to plants if they accumulate during vegetative growth [94].

One drawback of transient expression systems is that they are considered to suffer greater batch-to-batch yield variation than transgenic plants, thus conflicting with the requirements of current good manufacturing practice (cGMP) [95-97]. The cGMP guidelines apply to the production of all biopharmaceuticals and emphasize quality criteria such as batch-to-batch consistency and reproducibility [38, 49]. To fulfill cGMP requirements for transient expression systems, the sources of variation need to be identified and measures must be implemented to reduce or eliminate their impact on the production process. Less batch-to-batch variation increases the likelihood of regulatory acceptance and makes it possible to predict the outcome of a production process more accurately. Factors such as incubation temperature and leaf age can affect the accumulation of recombinant proteins in plant tissues during transient expression (I.4), but the precise impact and correlations have not been quantified [65, 67]. Therefore, a model describing the dependence of transient protein expression in tobacco leaves on process parameters such as post-infiltration incubation temperature, plant and leaf age as well as the optical density of the *A. tumefaciens* solution would improve our understanding of the transient expression system substantially. Such a model could predict protein expression in future batches and thereby form a process analytical technology (PAT) database for parameter and out-of-specification limits under a cGMP-compliant process. Integrating this model with those for other upstream and/or downstream operations (I.4, [98]) could form the basis of a quality by design (QbD) concept for the whole production process and would significantly enhance the likelihood of regulatory approval.

I.6 Plant–bacterial interactions

As stated above (I.4), the accumulation of recombinant proteins in plants often falls short of the levels achieved in more traditional protein expression systems such as mammalian cell cultures, e.g. monoclonal antibody manufacturing in CHO cells [41, 50]. Transient expression

achieves higher protein yields than expression in transgenic plants [85-87], and a broad set of approaches has been developed to improve the transient system further, including the use of stronger and/or synthetic promoters [56, 57], the inclusion of 5'UTRs and introns [58], combinations of viral and bacterial transformation vectors [99, 100] and the use of silencing suppressors [101].

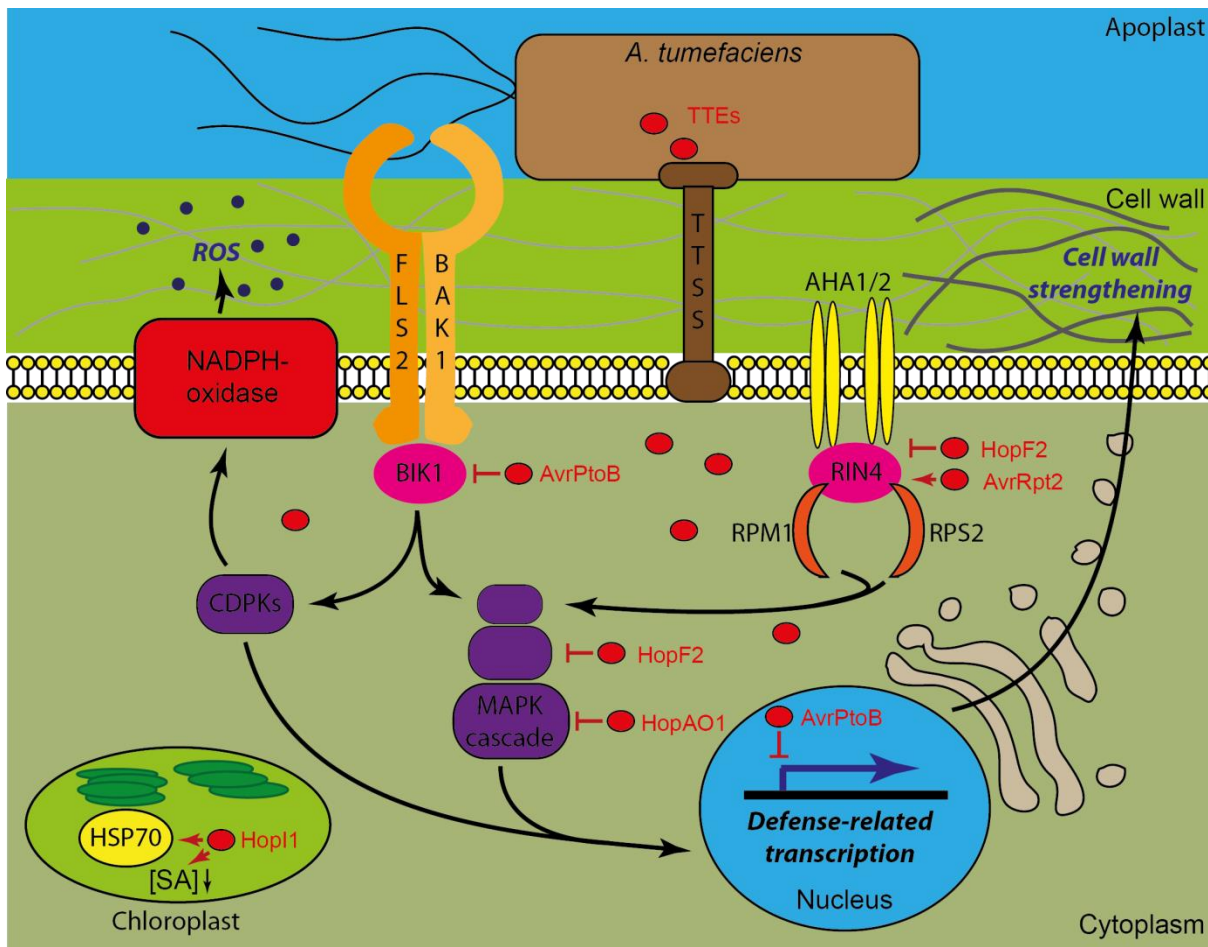


Figure I.1: Known plant cell targets of the TTEs AvrPtoB, AvrRpt2, HopF2, HopAO1 and HopI1.

TTE – type III effector; TTSS – type III secretion system; ROS – reactive oxygen species; CDPK – calcium-dependent protein kinase; MAPK – mitogen-activated protein kinase; HSP70 – heat shock protein 70; SA – salicylic acid. Adapted from [102].

Type III effectors (TTEs) from the plant pathogen *Pseudomonas syringae* can also boost transient protein expression when induced in genetically-modified *Arabidopsis thaliana* plants prior to treatment with *A. tumefaciens* [103]. TTEs are synthesized by the pathogen and exported into the plant cell using the bacterial type III secretion system (TTSS) [102, 104-106]. The purpose of TTEs is to interfere with pathogen-triggered immunity (PTI) and effector-triggered immunity (ETI), and ultimately to block plant countermeasures against the pathogen [102, 107-110]. PTI is induced in plants following contact with pathogen-associated molecular patterns (PAMP) such as bacterial flagellins [111]. The perception of PAMPs is dependent on membrane receptors known as pattern recognition receptors (PRRs) [111]. PTI

is often characterized by reactions such as cell wall strengthening through the formation of callose deposits or the synthesis of phenolic compounds such as salicylic acid [112, 113]. ETI reflects the binding of TTEs to so-called resistance proteins (R-proteins) that are often intracellular receptors of the nucleotide binding leucine-rich repeat (NB-LRR) family [114]. ETI may have evolved in plants as a secondary response to TTEs [114]. However, some TTEs can block ETI, such as AvrPtoB [115]. ETI is accompanied by programmed cell death (PCD), in this context also referred to as the hypersensitive response (HR) [114]. Recent studies indicate that PTI and ETI may overlap in terms of triggers and responses [114, 116]. Pathogens use a broad set of TTEs to target different host cell proteins and suppress plant defenses (Figure I.1) [102, 110].

AvrPtoB (Avr = avirulence) prevents plant defense signaling upstream of the MAPK (mitogen-activated protein kinase) cascade by binding to receptors or associated proteins [102]. This TTE can also ubiquitinylate R-proteins using its C-terminal E3-ligase domain, thereby targeting the R-proteins for proteasomal degradation, preventing PCD [115]. AvrRpt2 probably interferes with salicylic acid-dependent signaling and auxin metabolism in the host through its cysteine protease activity [117]. The presence of AvrRpt2 is detected by the plant protein RIN-4 and this elicits PCD, but HopF2 (hypersensitive response and pathogenicity outer protein) can prevent detection [102, 118]. HopF2 also interferes with plant cell signaling by the ADP-ribosylation of MAPKKs [119]. HopAO1 has a similar effect on signaling by phosphorylating and thus inactivating MAPKs [120]. HopI1 is localized in the chloroplasts and modulates salicylic acid defense signaling [108].

PTI/ETI also occurs if *A. tumefaciens* is injected/infiltrated into tobacco plants, resulting in the formation of callose deposits, the induction of secondary metabolism and cell death [121-124]. Co-expressing TTEs together with a POI might reduce these effects and thus improve protein expression levels by (i) reducing callose deposition to enhance the transfer of T-DNA into the plant cell, (ii) avoiding the induction of secondary metabolism leaving more resources for recombinant protein synthesis, and (iii) preventing cell death.

I.6.1 Economic 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 [41]. 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 [31, 37]. On the other hand, plant secondary metabolites and polymers may interfere with DSP by clogging equipment due to

their tendency to undergo oxidization and precipitation, so the actual costs may be higher [125]. The DSP costs can be reduced if less biomass needs to be processed to manufacture a certain amount of target protein. This can be achieved if expression levels are increased by using better promoters, and by optimizing codon usage or other genetic elements [54, 56, 58], but expression levels remain a problem associated with plant-based systems as stated above (I.4). However, another approach is to reduce the process volume by rational and knowledge-based selection of the processed biomass, e.g. by selecting only those plant parts with the highest expression levels. For transgenic plants, this approach tends to have only a minimal impact because the target proteins accumulate in different tissues throughout the growth period, resulting in homogenous levels of the target protein in all parts of the plant [56, 126, 127]. However, selective biomass processing can prove advantageous in transient expression systems if the variability of expression can be linked to certain properties of the plant allowing the guided selection of the biomass for processing.

I.7 Design of experiments

Experiments or observations aim to gather information about a subject or system and use the information to make predictions about its future behavior. For example, a process is predictable when parameters causing significant variation are identified and controlled. Varying these parameters within meaningful ranges in a set of experiments spans the so-called design space [128, 129]. When the information from the different data points (experiments) within a design space is combined, a model of the process can be built. The classical approach to span a design space is to vary one factor at a time (OFAT), but this concentrates the data points along lines within the design space reflecting the variation of a single process parameter (Figure I.2 A) [130]. Accordingly, coverage of the design space is poor, the information output is low and interactions between factors are unlikely to be found, resulting in poor predictions and suboptimal processes.

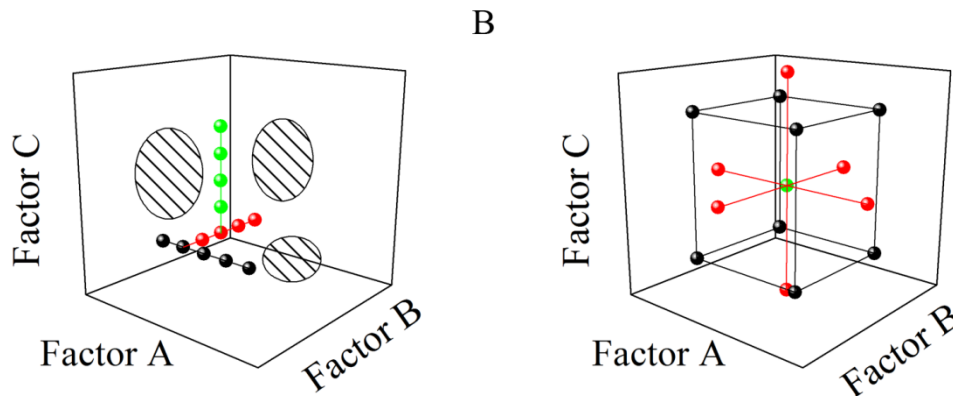


Figure I.2: Comparison of OFAT and DoE designs.

A. Experiments using 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 information obtained about the system and the likelihood of finding a desired operation point.

In contrast, much more information can be obtained from a statistical design of experiments (DoE), which 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 (Figure I.2 B) [130, 131]. Parameters included in the design space are called *factors*, 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 during a DoE, e.g. if only a limited number of temperatures can be tested. The measured variable is called a *response*, e.g. the recombinant protein concentration, fluorescence intensity, filter lifetime or extract turbidity.

The ranges of a factor tested in a DoE are often determined in initial experiments or known from experience or the literature. The ranges 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 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 1) [132]. Selecting a meaningful range for each factor is crucial because otherwise an important effect may remain concealed, as seen in Figure I.3. TLFs can be augmented with central and optional star points to form a so-called central-composite design (CCD) to avoid this problem

albeit at the cost of additional experiments [132]. However, 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 optimal criteria are listed in Table I.1, 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 [133].

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

Equation 1: 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, $k = l$.

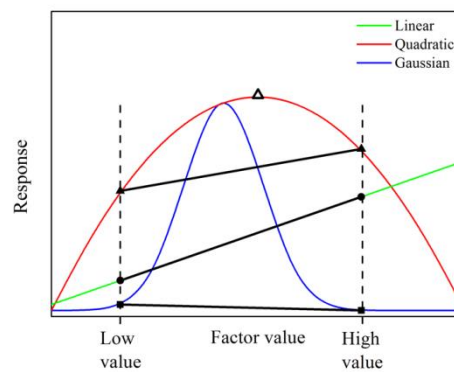


Figure I.3: The importance of meaningful factor range selection.

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 I.1: 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 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 2), adjusted r^2 and predicted r^2 (Equation 3 and Equation 4) [134]. 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 uses the predicted residual sum of squares (PRESS, Equation 5) and is similar to a leave-one-out analysis and thus measures the predictive power of a model [134]. Other measures to identify low-quality models may not be suitable to evaluate a DoE because data redundancy is low by design and thus leave-many-out (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 deviance of model and experimental data (Equation 6) [134-137]. 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 allow the non-quantitative evaluation of the model [134].

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

Equation 2: 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 3: 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 4: Predicted r^2 calculation, with predicted residual sum of squares (PRESS, Equation 5), 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 5: 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 ith element of a data set but excluding that specific element from the calculation [138, 139].

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

Equation 6: 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 [140]. 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 [141].

I.8 Aim of this thesis and workflow

This thesis considers different aspects of upstream production (USP) and how they can affect DSP and process costs for a plant-derived biopharmaceutical protein (Figure I.4). Initially, parameters affecting transient protein expression in *Nicotiana* species were identified and their impact on protein accumulation was quantified using DoE and modeling tools. The tested parameters included cultivation conditions, genetic elements and plant properties, as outlined in sections I.4 and I.5. Additionally, the impact of type-III effectors on co-expressed target proteins by means of defense response modulation was assessed (I.6). The rationale was that such effectors might inhibit the defense responses elicited by *A. tumefaciens*, thereby increasing the capacity of plant cells to produce target recombinant proteins. The monoclonal antibody 2G12 (specific for the HIV gp120 glycoprotein) and the fluorescent marker protein DsRed were used as model proteins for the work described in this thesis. An antibody was chosen because this is the most prevalent type of biopharmaceutical protein (section I.3), whereas DsRed was selected because it is easy to detect and quantify by fluorescence analysis. Furthermore, a GMP-compliant production process for these two proteins was recently established providing background data for comparison with the results described in this thesis.

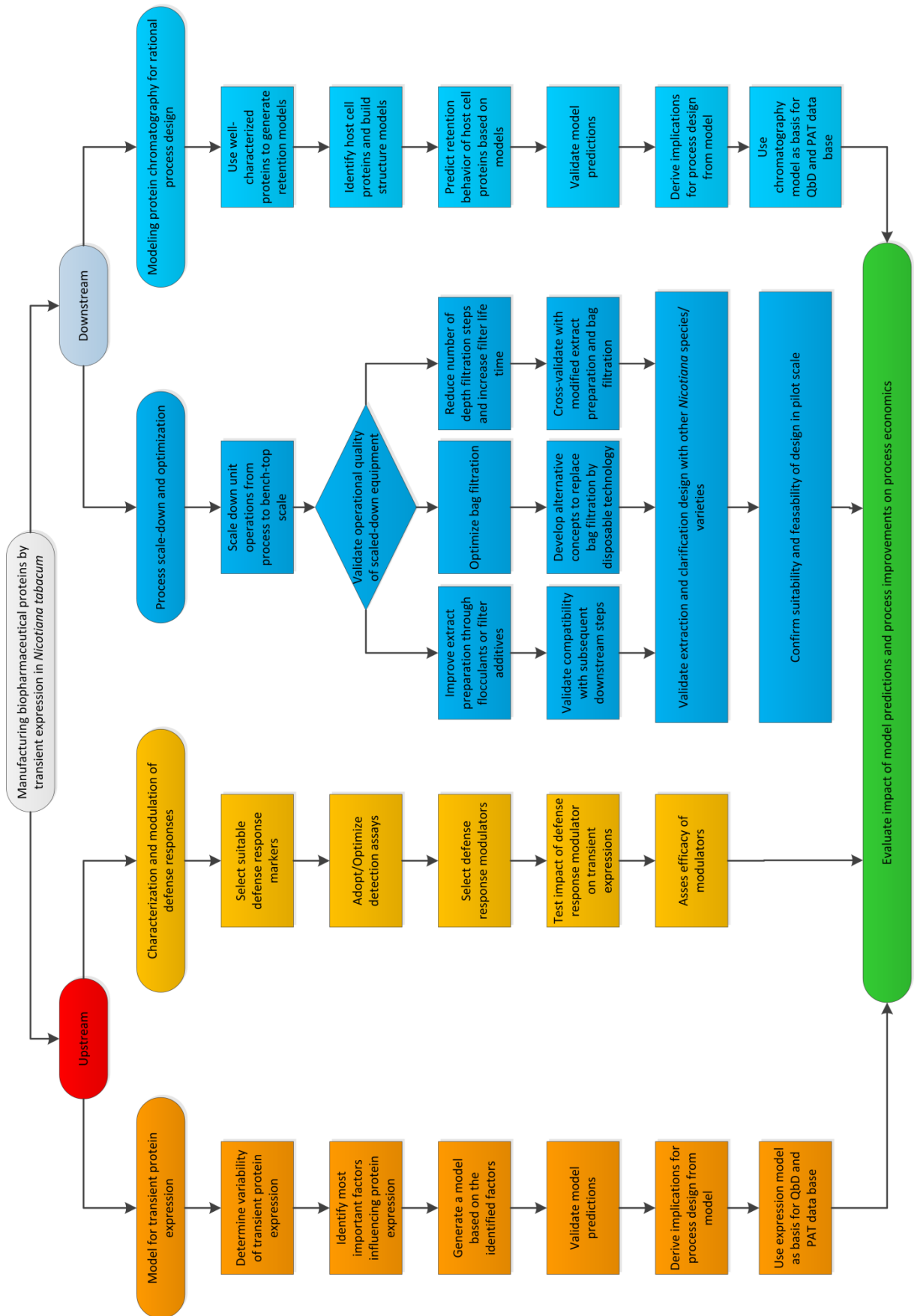


Figure I.4: Workflow during this PhD thesis, as described in section I.8.

The work packages in blue are part of an accompanying PhD thesis entitled “Manufacturing biopharmaceutical proteins in tobacco” [98].

II. Materials and methods

II.1 Equipment and chemicals

All equipment is listed together with the manufacturers' information in the appendix, section VII.1. All chemicals and buffers are listed in the appendix, sections VII.4 and VII.5, respectively.

II.2 Expression vectors and cloning

II.2.1 General cloning procedures and cultivation of bacteria

Recombinant plasmids were propagated in *Escherichia coli* K12 strain DH5 α . Bacteria were cultured in lysogeny broth (LB) containing 50 $\mu\text{g}/\text{mL}$ ampicillin for the selection of recombinant clones at 37°C. Plasmid DNA was isolated using NucleoSpin® Plasmid columns from MN (Düren, Germany) according to the manufacturer's protocol. All restriction endonucleases (RE) were ordered from New England Biolabs (NEB) (Ipswich, MA, USA) and used according to the manufacturer's recommendations. DNA fragments were separated by agarose gel electrophoresis after RE treatment in 0.8–2.0% (w/v) gels at 100 V for ~45 min. DNA fragments were purified using NucleoSpin® Gel and PCR Clean-up columns from MN according to the manufacturer's protocol. KAPAHiFi™ polymerase from PEQLAB (Erlangen, Germany) was used to amplify DNA sequences by PCR using an annealing temperature set 5°C below the melting temperature calculated for the primer pairs using Clone Manager 9.0 by Sci-Ed Software (Cary, NC, USA). All primers were ordered from Eurofins MWG Operon (Ebersberg, Germany). Synthetic genes were ordered from GeneArt (now part of Life Technologies GmbH, Darmstadt, Germany) and codon optimized for expression in *N. tabacum*. DNA fragments were ligated using the Quick Ligation™ Kit from NEB. Manipulated plasmids were introduced into *E. coli* RbCl-competent cells by heat shock transformation at 42°C for 1.5 min [142].

II.2.2 *Agrobacterium tumefaciens* infiltration of plants

A. tumefaciens strain GV3101:pMP90RK was transformed with the plasmids (section II.2.1 as well as II.2.3, II.2.4 and II.2.5) by electroporation [143]. 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. tabacum* or *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_{600nm} . An OD_{600nm} of 1.0 corresponded to $1.43 \pm 0.12 \times 10^9$ colony forming units per mL ($n = 6$). For the co-expression of TTEs and pGFD, the OD_{600nm} of the bacteria carrying each construct was adjusted to twice the final desired value and the bacterial suspensions were mixed 1:1 prior to use (Table III.9). Bacteria were inoculated either by manual injection into intercostal fields using 1-mL syringes or by vacuum-infiltration of whole plants at ~ 50 Pa absolute pressure for 15 min followed by sudden vacuum release. Leaves 1–8 of tobacco plants were used for this treatment with leaf 1 defined as the oldest non-cotyledon leaf and leaf 8 the youngest fully-expanded leaf (Figure II.1 A). Each leaf was subdivided into four positions to resolve differences in protein expression along the mid-vein axis (Figure II.1 B).

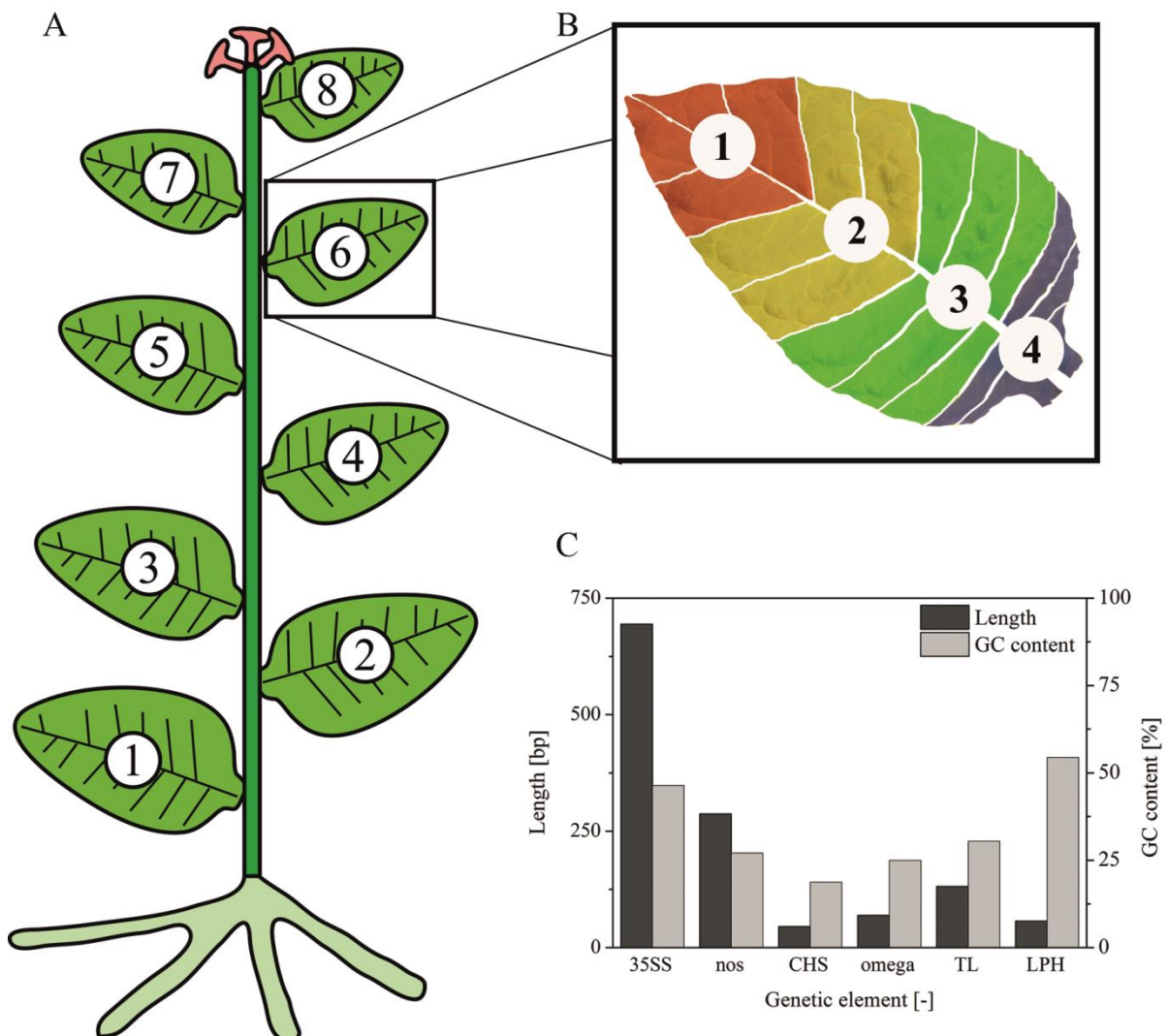


Figure II.1: Plant segmentation and promoter/5'UTR properties (section II.2.4).

A. schematic representation of a tobacco plant at 47 days after seeding (time of harvest) with numbers indicating the leaf age as used throughout this thesis, 1 being the oldest and 8 the youngest leaf. B. Subdivision of each leaf into four positions along the mid-vein axis, as used during infiltration and model building. C. Length and GC-content of the promoter and 5'UTR sequences describe in section II.2.4 and III.2.

II.2.3 pGFD vector

The pGFD vector, a derivative of pPAM (GenBank AY027531), was kindly supplied by Dr. Thomas Rademacher, Fraunhofer IME [144]. 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.

II.2.4 Promoter/5'UTR combinations for DsRed expression

Eight vectors containing different promoter/5'UTR combinations were constructed to express a modified *DsRed* gene, as shown in Figure III.11. The vector backbone was as described above. Coding sequences for a His₆ tag and a KDEL retention/retrieval signal for localization in the endoplasmic reticulum (ER) were added in-frame to the 3'-end of the *DsRed* coding sequence. All constructs contained the CaMV 35S polyadenylation (polyA) signal. The length and GC-content of the genetic elements are shown in Figure II.1 C. The folding energies of the different mRNAs at 25°C were calculated using the RNAfold webserver [145] based on the mRNA sequence from the reported transcriptional start sites of the CaMV 35SS [146] and *nos* [147] promoters to the first 50 bp of the *DsRed* coding sequence. These vectors were constructed and cloned by Thomas Kaever as part of his bachelors thesis [148].

II.2.5 Type III effectors

Five vectors containing different TTE genes were constructed using vector pAIX-2, which is also based on pPAM and shares the same backbone features as described above. The T-DNA contained one of the TTE genes flanked by RB7 SARs and expression was driven by a combination of the *nos* promoter, the omega-prime 5'UTR and *nos* polyA signal. A schematic representation of the expression cassette is shown in the results section (Figure III.16). Some of the cloning was carried out by Claudia Haase as part of her bachelors thesis [149].

II.3 Plant growth

II.3.1 Plant species

The *N. tabacum* variety Petit Havana SR1 (for brevity referred to as SR1 hereafter) was used for transient protein expression in this thesis (II.2.2).

II.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 Ferty 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.

II.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 15, 17, 20, 22, 25, 28 or 30°C. Treated leaves were sampled 5 days post injection (dpi) when plants reached 40 or 47 days post seeding (dps) corresponding to the growing and mature bud stages, respectively.

II.4 Sample preparation and analysis

II.4.1 Sampling from infiltrated plants

Samples were taken from infiltrated leaf parts (II.2.2) 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 .

II.4.2 Protein quantitation

The quantity of total soluble protein (TSP) in the supernatants was determined using the Bradford method [150, 151]. 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 [152, 153]. 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.

II.4.3 SDS-PAGE analysis

Pre-cast 4–12% (w/v) continuous Bis-Tris gels and additional equipment from Life Technologies were used for sample (II.4.1) analysis by reducing sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-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.

II.4.4 Western blot and immunodetection

Samples separated by SDS-PAGE (II.4.3) 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 II.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 II.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)	
c-myc	Mouse α-c-myc	Goat- α-mouse-Fc-AP (Jackson 115-005-008)

II.4.5 Trypan blue staining of dead cells

Leaf samples transiently expressing TTEs were stained with Trypan blue using a modified form of the original protocol (Table II.2) [154, 155]. Briefly, leaf samples were heated in staining solution (70% (v/v) ethanol, 10% (v/v) glycerol, 10% (v/v) lactic acid, 10% (v/v) Milli-Q water, 0.0125% (w/v) Trypan blue), washed in chloral hydrate (2.5 g mL⁻¹) and mounted in 50% (v/v) glycerol on glass slides for microscopy.

Table II.2: Trypan blue staining protocol parameter settings.

Protocol parameter	Original value ¹	Optimized value ²
Heating temperature [°C]	80	90
Heating time [s]	60	90
Incubation temperature [°C]	30	15
Incubation time [min]	10	22
Trypan blue concentration [mg mL ⁻¹]	0.250	0.125
De-staining temperature [°C]	30	37

¹ Values from [154], ² Values from [155].

II.4.6 Callose staining

Callose deposits in leaf samples treated with *A. tumefaciens* were stained at 2 dpi as described [156]. Samples were covered with de-staining solution (6:1 v/v ethanol/acetic acid) and heated to 45°C for 60 min, and the process was repeated with fresh de-staining solution. The samples were rinsed with water, equilibrated in phosphate buffer (K₂HPO₄, 150 mM, pH 9.0) for 30 min and stained in the same buffer containing 0.05% (w/v) aniline blue for 24 h. The samples were then rinsed with water and mounted in 50% (v/v) glycerol on glass slides for microscopy.

II.4.7 Microscopy

Stained leaf samples and particles in extracts were analyzed using a LEICA DRM light microscope (LEICA, Wetzlar, Germany) or an Opera confocal microscope (PerkinElmer, Waltham, MA, USA). A list of objectives and filters is provided in Table II.3 and Table II.4. Images were captured using a LEICA DFC 320 camera or the built-in optics of the Opera confocal microscope.

Table II.3: Objectives used with the LEICA DRM.

Objective	Magnification ¹	Medium	Aperture
HCX PL Fluostar	1.6	Air	0.05
HC PL Fluostar	10	Air	0.30
HCX PL Fluostar	20	Air	0.40
PL Fluostar	40	Oil	1.00 – 0.50
HCX PL APO	63	Oil	1.32 – 0.6
HCX PL APO	100	Oil	1.4

¹ The mounted ocular HC Plan 10x/25 provided an additional 10-fold magnification to be multiplied with these values to yield the total magnification.

Table II.4: Filters used with the LEICA DRM.

Filter cube	Excitation filter [nm]	Beam splitter [nm]	Emission filter [nm]	Substance to detect
Chroma 400 DCLP	350/50	400	445/40	Callose
Y3 (DsRed)	545/30	565	610/75	DsRed
GFP	470/40	500	525/50	GFP/autofluorescence

II.4.8 Metabolic analysis

Plant secondary metabolites were separated using an ÄKTA explorer equipped with a Chromolith FastGradient RP-18 endcapped 50-2 column (Merck, Darmstadt, Germany) and a Security Guard Cartridge Holder (Phenomenex, Aschaffenburg, Germany) with a C18 cartridge AJ0-4286. Leaf samples were prepared as described above (II.4.1) by extracting in a 40/60 (v/v) mixture of methanol/water acidified to pH 3.6 with 0.1% (v/v) acetic acid, followed by centrifugation as above but at -15°C . A Phenex RC 4-mm syringe 0.2 μm filter AF0-3203-52 (Phenomenex, Aschaffenburg, Germany) was used to clear remaining particles from the samples before chromatography. A 10- μl sample was injected at a flow rate of 0.8 mL min^{-1} and eluted in a 0–80% B gradient in 6.5 min after a 0.5-min wash. Sample elution was monitored at 260, 320 and 360 nm as described [157]. Solvent A was water (WFI quality) acidified with 0.01% (v/v) TFA to pH 3.0, and solvent B was HPLC-grade methanol. Authentic standards used to identify plant secondary metabolites are listed in Table II.5.

Table II.5: Authentic standards used for identification of plant secondary metabolites.

Standard	Final concentration [$\mu\text{g/mL}$]
4-Aminopyridine	5
Nicotine	80
Quinoline	75
2,4'-Dipyridyl	20
Caffeic acid	5
Chlorogenic acid	25
Rutin	10
Cinnamic acid	5

II.5 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.

II.5.1 Transient protein expression model

For plants harvested after 40 or 47 days, stepwise augmented IV-optimal response surface designs were used to build models for transient protein expression depending on the factors listed in Table II.6. Categorical balance was forced for the factors. DsRed and 2G12 concentrations in extracts from plants harvested at 40 dps were modeled based on 142 samples, and 465 samples were used to model the concentrations of the same proteins in plants harvested at 47 dps. Factors showing a significant influence on protein expression were pre-selected from a fourth-order 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 if a factor was needed to maintain the model hierarchy [158, 159]. The resulting models for plants harvested after 40 or 47 days were compared and a consensus model was established including all factors present in both models.

Table II.6: Factors considered during DoE and for model building including the range of their set points.

Factor	Unit	Lower boundary	Upper boundary
Post-injection temperature	[$^{\circ}\text{C}$]	15	30
Plant age at time of injection	[dps]	35	42
Leaf age	[-]	1	8 ¹
Position on leaf	[-]	1	4
Optical density of injected <i>A. tumefaciens</i>	[-]	0.13	2.00

¹ 5 for plants injected 35 dps

II.5.2 Promoter strength model

A D-optimal response surface design was used to build a model for transient protein expression depending on the numeric factors listed in Table II.7. Promoters and 5'UTRs were treated as categorical factors with the different sequences corresponding to the factor levels (Figure III.11). Categorical balance was forced for the factors. A total of 630 samples was used to model DsRed concentrations in extracts from plants harvested 2, 5 or 8 dpi. Factors showing a significant influence on protein expression were pre-selected from a 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.

Table II.7: Numeric factors considered during DoE and for promoter model building including the range of their set points.

Factor	Unit	Lower boundary	Upper boundary
Position on leaf	[-]	1	4
Incubation time after <i>A. tumefaciens</i> injection	[d]	2	8
Leaf age	[-]	1	8

III. Results and discussion

III.1 Transient protein expression model

The results presented in this section have been published as:

1. Buyel JF, Fischer R. 2012. Predictive models for transient protein expression in tobacco (*Nicotiana tabacum* L.) can optimize process time, yield, and downstream costs. *Biotechnology and Bioengineering* 109(10):2575-88.
2. Buyel JF, Fischer R. 2013. Processing heterogeneous biomass: Overcoming the hurdles in model building. *Bioengineered* 4(1).

III.1.1 Model theory

III.1.1.1 Model concept

Protein quantitation often involves two steps, an initial target-specific step (e.g. ELISA or fluorescence measurement) to determine the concentration, and a second non-specific step to measure the volume of the bulk extract [160]. A routine extraction procedure using a fixed volume of extraction buffer per unit biomass allows the protein concentration and corresponding extract volume to be determined for homogenous source materials, e.g. a clarified cell pellet extract. When the concentration (C) and volume (V) are known, the quantity of recombinant protein (m) can be calculated as shown in Equation 7. However, where the source material is heterogeneous (e.g. leaves of a plant at different stages of maturity) different parts of the material will contain significantly different amounts of recombinant protein. The quantity and distribution of the protein in the source material must therefore be determined to improve the process layout, because this allows the optimal process material to be selected. To determine the quantity and distribution of the recombinant protein (P), the concentration and volume terms of Equation 7 are required for each portion of the source material.

$$m[P] = V \cdot C[P]$$

Equation 7: Calculation of the mass of a recombinant protein $m[P]$ from a given volume V and concentration of the recombinant protein $C[P]$.

III.1.1.2 Plant parameters

Tobacco leaves transiently expressing a fluorescent protein (DsRed) and a monoclonal antibody (2G12) were used as an example of heterogeneous source material. Two parameters were introduced to describe the heterogeneity of the leaves: (i) the variable k denoting the individual leaves of each plant (Figure III.1 A); and (ii) the variable p indicating different

positions within a leaf. Although the classification into leaves (k) is intuitive, the second classification seems artificial but is based on experience with protein expression patterns in leaves. The number of levels for k is specified by the habit and age of the tobacco plants whereas four different positions were defined for p (Figure III.1 D). This definition was advantageous, first because it facilitated second-order (quadratic) polynomial fitting of the resulting volume and concentration data over each leaf, and second because it allowed the size of each leaf to be normalized into four relative length units with each position spanning an interval of one unit. Assigning the integers 1–4 as the intervals means that 0.5 and 4.5 represent the leaf apex and node, respectively. Length normalization from zero to unity is also possible but the volume and concentration data would then need to be transformed prior to further calculations and integration.

III.1.1.3 Volume model

Having subdivided the plant, a model was built for the volume term of Equation 7 as shown in Figure III.1 A–F. Three experimentally-determined ratios (η_k , ε and $\psi_{k,p}$) were used to convert the average total biomass of a leaf ($m_{t,k}$) into position-specific extract volumes for that leaf ($V_{k,p}$) as shown in Equation 8. Second-order polynomial fits through the four $V_{k,p}$ values ($V_{k,1}$ to $V_{k,4}$) were used to obtain continuous volume functions for each leaf as shown in Equation 9. This allowed subsequent multiplication with protein concentration functions and integration to determine the protein yield. Because the volume model is not dependent on the properties of a recombinant protein it can be combined with the concentration function of any protein of interest.

$$V_{k,p} = m_{t,k} \cdot \eta_k \cdot \psi_{k,p} \cdot \varepsilon$$

Equation 8: Calculation of extract volumes $V_{k,p}$ for the positions p of a given leaf k , where $m_{t,k}$ is the average total leaf biomass of leaf k [g], η_k is the effective biomass ratio for each leaf k (intercostal field biomass [g]/total leaf biomass [g]), $\psi_{k,p}$ is the position ratio for each leaf k (biomass of position p [g]/effective leaf biomass [g]) and ε is the extract ratio (volume of solids-free extract [mL]/biomass [g]).

$$f(V_{k,p}, \beta_k) = \beta_{k,1} + \beta_{k,2}V_{k,p} + \beta_{k,3}(V_{k,p})^2; \beta_k = (\beta_{k,1}, \beta_{k,2}, \beta_{k,3})$$

Equation 9: Linear second-order polynomial fit of extract volumes ($V_{k,p}$ [mL]) over the four positions (p) of an individual leaf (k), where $\beta_{k,1}$ through $\beta_{k,3}$ represent the fitted coefficients for each leaf k . $V_{k,p}$ values were obtained from Equation 8.

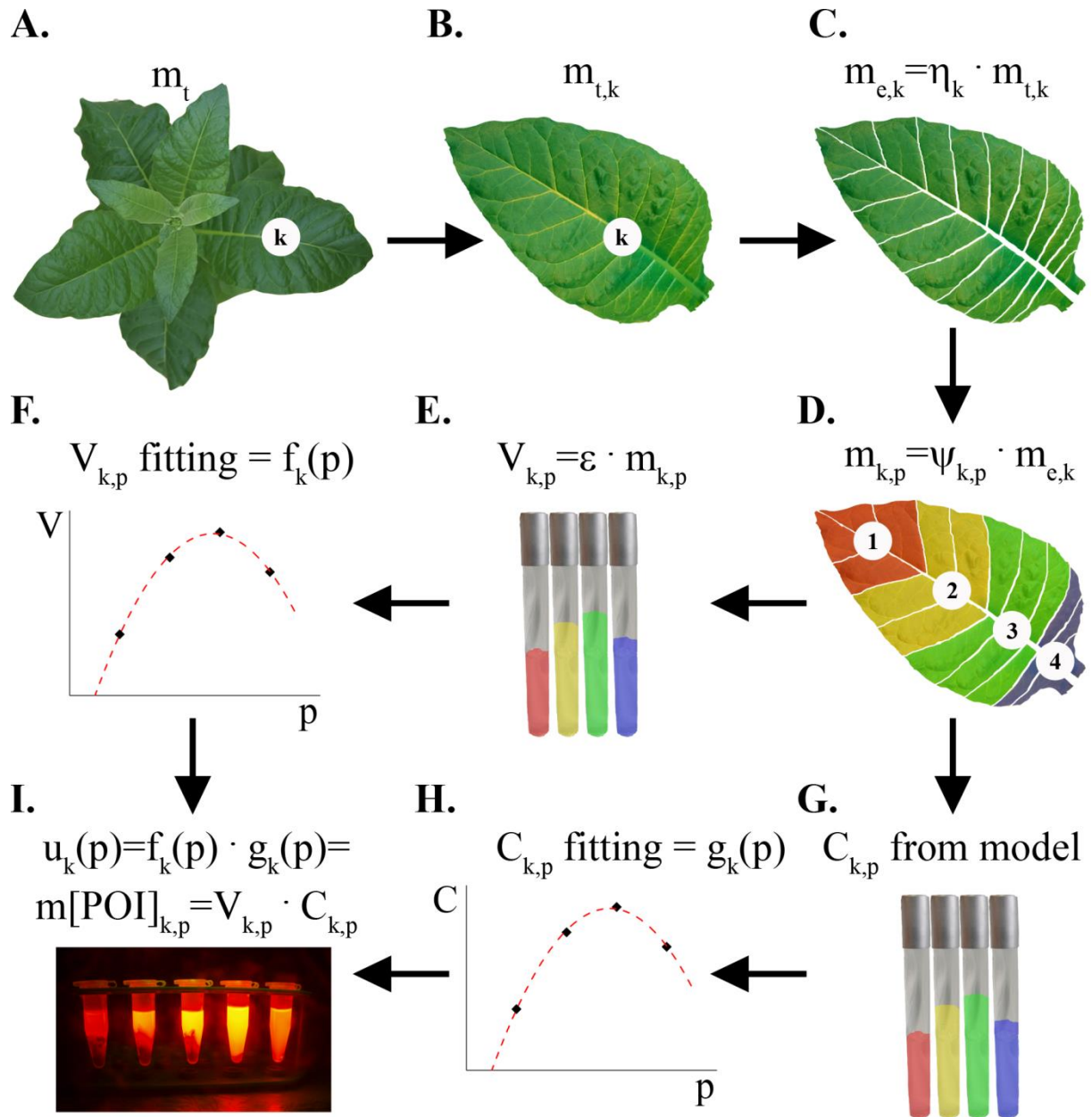


Figure III.1: Procedure for calculating the amount of recombinant protein in each leaf of an infiltrated plant (II.2.2, II.4.1, and II.4.2).

A. Each leaf is assigned a number k that declines in older leaves. B. Total biomass $m_{t,k}$ is determined for every leaf k . C. Veins are removed and the effective biomass $m_{e,k}$ is determined. The ratio of these two biomasses is the effective biomass ratio η_k . D. Effective leaf biomasses are divided into four segment classes and used to calculate the ratios of these classes to the effective biomass of each leaf $\psi_{k,p}$. Given these two sets of ratios, η_k and $\psi_{k,p}$, the effective biomass of each leaf position can be calculated for every leaf in a given plant, $m_{k,p}$, or plant population (Equation 8). E. When proteins are extracted from biomass, each gram of biomass will yield approximately 2.75 mL of extract under routine conditions, giving an extract ratio ε of 2.75 mL g^{-1} . G. Concentrations of recombinant protein are predicted using the model Equation 10. F. and H. Linear second-order polynomial fitting is used to transform discrete position-dependent volumes and concentrations into continuous functions for each leaf (Equation 9 and Equation 11). I. Integrating the product of volume and concentration functions over all positions of a leaf (p) yields the amount of recombinant protein for an individual leaf. The sum of integrals over all leaves (k) predicts the amount of recombinant protein that can be isolated from one tobacco plant under the conditions described (compare Figure III.3 B and Equation 12).

III.1.1.4 Concentration model

Concentration functions for DsRed and 2G12 were built in a second step (Figure III.1 A-D, G, H). Concentrations were initially considered to be dependent on parameters k and p as well as the following additional process parameters: incubation temperature (θ), optical density of the injected *A. tumefaciens* at 600 nm (OD_{600nm}) and plant age at time of harvest (dps). These three process parameters were selected instead of (for example) humidity or light source because they can be controlled in most laboratories and would be more valuable for the standardization of process conditions. Parameter combinations were selected to span the design space using DesignExpert 8.0 according to an IV-optimal design plan (II.5.1). The same software was used to fit a model on the experimental data to generate Equation 10 (an example of a graphical representation of this equation is shown in Figure III.2). The significance of the model and selected factors are shown in Table III.1 and Table III.2, respectively. To reflect the conditions during a given production process, fixed values for incubation temperature (22°C) and OD_{600nm} (1.0) were then used to predict the leaf (k)- and position (p)-dependent protein concentrations ($C_{k,p}$) for plants at two growth stages (harvested at 40 or 47 dps, respectively). Second-order polynomial fits were once again used to obtain continuous concentration functions for both recombinant proteins as shown in Equation 11.

$$\begin{aligned} \text{Log}_{10}(C_{k,p}[P]) &= c_{k,1} + c_{k,2}\theta + c_{k,3}p + c_{k,4}OD_{600nm} + c_{k,5}\theta \cdot p + c_{k,6}\theta^2 + c_{k,7}p^2 \\ &+ c_{k,8}OD_{600nm}^2 + c_{k,9}\theta^2 \cdot p + c_{k,10}\theta^3 + c_{k,11}\theta^4 \end{aligned}$$

Equation 10: Empirical formula for the calculation of protein concentration ($C_{k,p}[P]$) in the extract of a tobacco leaf (k) at position (p) based on the evaluation of DoE data, where k is the leaf age, p the position on leaf (see also Figure III.1 D), θ is the temperature in the range 15–30°C, and OD_{600nm} is the optical density of the injected *A. tumefaciens* culture at 600nm. The model statistics are shown in Table III.1.

$$g(C_{k,p}, \gamma_k) = \gamma_{k,1} + \gamma_{k,2}C_{k,p} + \gamma_{k,3}(C_{k,p})^2; \gamma_k = (\gamma_{k,1}, \gamma_{k,2}, \gamma_{k,3})$$

Equation 11: Linear second-order polynomial fit of protein concentrations ($C_{k,p}$, [µg/mL]) over the four positions (p) of an individual leaf (k), where $\gamma_{k,1}$ through $\gamma_{k,3}$ represent the fitted, protein-dependent coefficients. $C_{k,p}$ values were obtained from Equation 10.

Table III.1: Validation statistics of a consensus model for the two recombinant proteins at both harvesting times.

Plant age [dps]	40		47	
Protein	DsRed	2G12	DsRed	2G12
r^2	0.96	0.91	0.94	0.88
Adjusted r^2	0.95	0.89	0.93	0.87
Predicted r^2	0.93	0.86	0.93	0.85

Table III.2: Model factors with a significant impact on transient protein expression in *N. tabacum*.

Source	Sum of squares	Degrees of freedom	F-value	p-value
Model	29.47	26	44.25	< 0.0001
Temperature [°C] (A)	0.21	1	8.26	0.0048
Leaf number [-] (B)	7.68	4	75.00	< 0.0001
Position on leaf [-] (C)	0.24	1	9.49	0.0026
OD 600nm [-] (D)	0.04	1	1.45	0.2304
AB	2.17	4	21.19	< 0.0001
AC	0.08	1	3.22	0.0754
BC	0.52	4	5.05	0.0009
A ²	14.63	1	571.23	< 0.0001
C ²	0.0003	1	0.015	0.9038
D ²	0.14	1	5.45	0.0213
A ² B	1.10	4	10.72	< 0.0001
A ² C	0.13	1	4.88	0.0292
A ³	0.16	1	6.25	0.0138
A ⁴	0.015	1	0.57	0.4501
Residual	2.92	114	n.a.	n.a.
Lack of Fit	2.53	92	1.56	0.1172
Pure Error	0.39	22	n.a.	n.a.

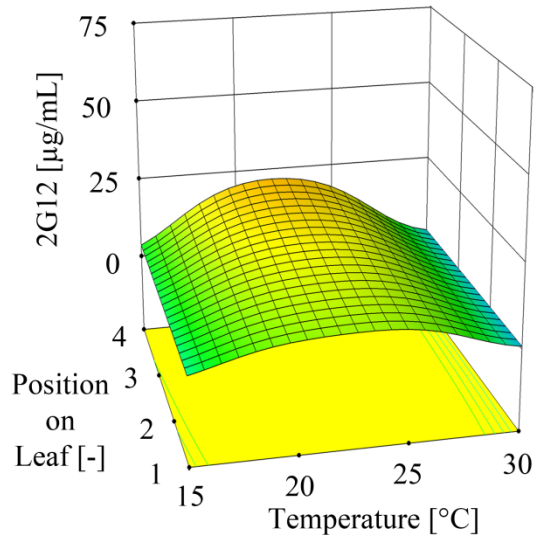
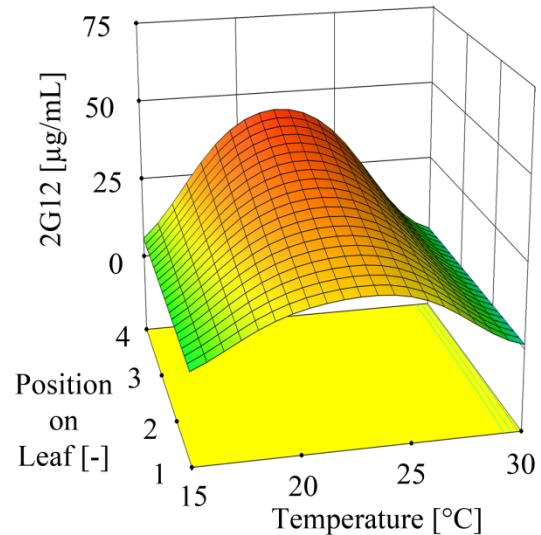
A. Leaf 4**B. Leaf 5**

Figure III.2: The concentration of 2G12 [$\mu\text{g mL}^{-1}$] in leaf extracts from tobacco plants injected with *A. tumefaciens* ($\text{OD}_{600\text{nm}} = 1.0$) at 42 dps is dependent on the developmental stage of a leaf (II.4.1 and II.5.1).

A distinct increase in 2G12 concentration was observed between leaf 4 (A) and leaf 5 (B). The concentration of DsRed is presented as a function of post-infiltration incubation temperature and position on a leaf.

Table III.3: Costs for energy, consumables and labor required for a transient protein expression process in tobacco. Note that GCF_A , IC_A and DCF in Equation 13 correspond to the summed costs for upstream production, infiltration and downstream processing respectively. For GCF_A and IC_A there are two values depending on the age of the plants that are processed: GCF_{40dps} or GCF_{47dps} and IC_{40dps} or IC_{47dps} .

Step	Factor	Position	Required units per plant		Unit cost	Costs per plant [€]	
			40 dps	47 dps	[€/L] or [€/h] or [€/kWh]	40 dps	47 dps
Upstream	Energy	Light [kWh]	8.53	12.94	0.14	1.19	1.81
		Climate control [kWh]	3.58	5.42	0.14	0.50	0.76
	Consumables	Growth support [-] (rockwool)	1.00	1.00	0.15	0.15	0.15
		Fertilizer [L]	5.73	10.44	0.002	0.01	0.02
		Fresh water [L]	5.86	10.64	0.01	0.03	0.05
	Labor	Labor [h]	0.13	0.14	30.00	3.79	4.27
				Sum		(GCF_A)	5.67

Step	Factor	Position	Required units per batch		Unit cost	Costs per batch [€]	
			40 dps	47 dps	[€/L] or [€/h] or [€/kWh]	40 dps ^a	47 dps
Infiltration	Energy	Fermentation [kWh]	281.42	432.96	0.14	39.40	60.61
		Infiltration/vacuum [kWh]	5.20	8.00	0.14	0.73	1.12
		Medium temperature adjustment [kWh]	7.55	11.61	0.14	1.06	1.63
	Consumables	YEB medium [L]	99.67	153.33	1.15	114.20	175.70
		Infiltration medium [L]	325.00	500.00	2.22	721.17	1109.49
		Fresh water [L]	1074.67	1653.33	0.01	5.37	8.27
	Labor	Inactivation chemicals [L]	6.50	10.00	21.63	140.60	216.30
Labor	Labor [h]	36.00	36.00	30.00	1080.00	1080.00	
			Sum		(IC_A)	2102.53	2653.12

Step	Factor	Position	Required units per kg processed biomass	Unit cost	Costs per kg processed biomass [€]
				[€/L] or [€/h] or [€/kWh]	
Downstream ^b	Energy	Homogenization [kWh]	0.15	0.14	0.02
		Filtration [kWh]	0.02	0.14	0.002
		Chromatography [kWh]	0.11	0.14	0.02
	Consumables	Bags [-]	n.a.	n.a.	37.77
		Filters [-]	n.a.	n.a.	41.60
		Chromatography resin [-]	n.a.	n.a.	44.57
		Tubing [-]	n.a.	n.a.	21.58
		Chemicals [-]	n.a.	n.a.	6.50
		Water for buffers and cleaning [L]	20.00	0.01	0.10
	Labor	Labor [h]	1.69	30.00	50.55
			Sum	(DCF)	202.71

^a Infiltration costs for plants harvested 40 dps were 65% compared to plants harvested 47 dps because the former were shorter and required less infiltration solution, only the labor remained the same; ^b The required units and costs per kg biomass calculations were based on data from a 200-kg GMP process producing the same proteins in transgenic plants.

III.1.1.5 Yield model

In order to calculate the average amount of protein ($m[P]$) that can be isolated from all the leaves of an infiltrated tobacco plant under the described conditions, the volume and concentration functions (f and g respectively) were multiplied to give the mass function u (Equation 12, Figure III.1 I). This function was then integrated over the whole interval of the control variable p (0.5–4.5) and summed for all leaves (k) to predict the amount of protein produced per plant.

$$m[P] = \sum_{k=1}^n \int_{0.5}^{4.5} (f(V_{k,p}, \beta_k) \cdot g(C_{k,p}, \gamma_k)) dp = \sum_{k=1}^n \int_{0.5}^{4.5} (u_{k,p}) dp$$

Equation 12: Final equation for the calculation of total recombinant protein mass ($m[P]$, [μg]) from a single tobacco plant infiltrated with *A. tumefaciens* ($OD_{600 \text{ nm}} = 1.0$) and incubated at 22°C for 5 d. The parameter n varies for plants at different ages and corresponds to the total number leaves of a plant ($n = 5$ or 8 for plants harvested 40 and 47 dps, respectively). The value of k can vary in the range $1 \leq k < n$ for the different harvesting schemes, corresponding to the number of the oldest leaf to be harvested in that specific scheme.

III.1.1.6 Cost function

The costs of a plant-based production process were separated into three sections: (i) upstream costs for plant growth; (ii) infiltration costs for treatment with *A. tumefaciens*; and (iii) downstream costs for the preparation of extracts and purification of the recombinant protein. This strategy was chosen because upstream costs can be calculated easily on a per-plant basis, whereas infiltration is performed batch-wise and downstream costs are usually calculated on a per-liter or per-kg basis because the process begins with bulk plant biomass (Table III.3). Furthermore, the biomass and leaf count of plants differs according to age and harvest scheme, so the number of plants alone is an inappropriate estimator for the costs that will arise during extraction and purification. The infiltration costs were assumed to be constant for the anticipated batch sizes of 2500–5000 plants. However, for plants harvested at 40 dps instead of 47 dps the infiltration costs were reduced by ~20% because the plants were ~35% shorter and required less infiltration medium but the same amount of labor (Table III.3). Process costs were calculated according to Equation 13 for 5 g of purified recombinant protein, including a recovery factor (RF) accounting for ~30% losses during purification. This factor is based on a 200-kg GMP process producing the same recombinant proteins in transgenic plants (Dr. Jürgen Drossard, unpublished data). The first term in Equation 13 corresponds to the number of plants needed to produce the desired amount of protein ($m_f[P]$) depending on plant age (A) and harvest scheme (k). The second term combines upstream costs (which are dependent on plant number and age) and downstream costs (which are dependent on plant number, age and harvest scheme). The last term in Equation 13 represents the age-dependent infiltration costs.

$$\text{Production costs} = \frac{m_f[P]}{RF \cdot m[P]_{kA}} \cdot [GCF_A + (DCF \cdot m_{kA})] + IC_A$$

Equation 13: Cost function for the production of DsRed or 2G12 by transient expression in tobacco, where $m_f[P]$ is a specified final amount of recombinant protein required after purification [g], RF is the average recovery factor accounting for losses during purification, $m[P]_{kA}$ is the amount of recombinant protein per plant depending on harvest scheme k and harvest time A [g/plant], GCF_A is the growth cost factor depending on harvest time A [€/plant], DCF is the downstream cost factor [€/kg biomass], m_{kA} is the biomass factor depending on harvest scheme k and harvest time A [kg biomass/plant], and IC_A is the infiltration costs depending on harvest time A [€]. Note that the value of DCF was based on a two-step antibody purification process using Protein A.

With the strategy described in sections III.1.1.1 to III.1.1.6, predictive models based on individual sampling and process parameters can be built for both the concentration and volume terms in Equation 7, leading to a yield function. With these models in hand, different production platform settings can be compared, and can be combined with a cost function to optimize processes duration, yield and/or downstream costs given a reasonable degree of reproducibility and repeatability.

III.1.2 Model building and key factors

III.1.2.1 System consistency and repeatability

A model that predicts the outcome of a process is useful if the outcome is dependent on certain parameters, but the predictive accuracy would be reduced by further experimental variation. It is therefore necessary to investigate the consistency of the process and the repeatability of its performance.

Transient protein expression is noted for inter-batch variations in yield [95, 96]. Therefore, parameters that affect consistency were investigated to determine their influence on tobacco plants cultivated in a greenhouse/phytotron setting (II.3), focusing on intra-plant, inter-plant and inter-batch variation in the yield of two recombinant proteins. Plants co-expressing a secreted monoclonal antibody (2G12) and a plastid-targeted fluorescent marker protein (DsRed) were studied and compared to transgenic plants expressing the same proteins as part of the Pharma-Planta project [161]. The leaf biomass distribution was used as a second indicator for the degree of variation among plants and batches (Figure III.3 A), thus allowing comparison to the protein yields (Figure III.3 B).

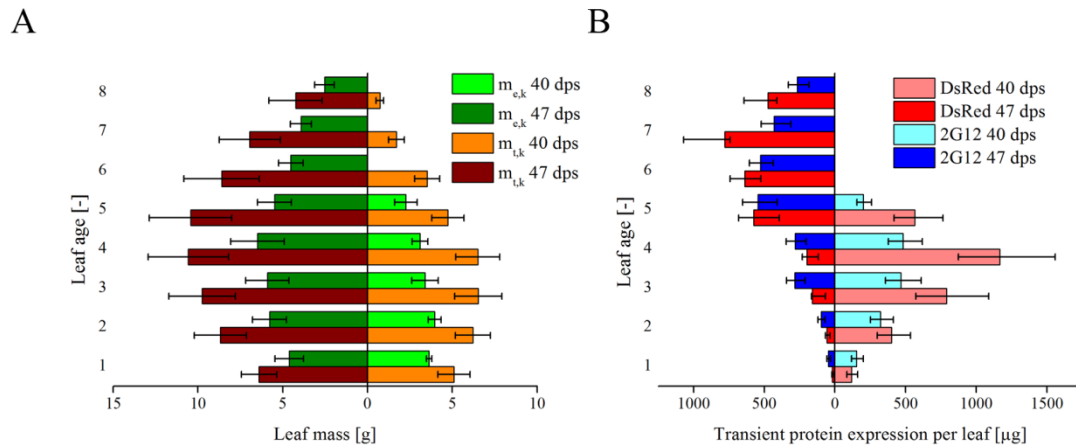


Figure III.3: Dependence of leaf biomass and transient protein expression on leaf and plant age (II.3, II.4.1 and II.4.2).

A. Averages of total leaf and intercostal field biomasses ($m_{t,k}$ and $m_{e,k}$, respectively) are shown for plants harvested 40 or 47 dps (right and left side of the y-axis, respectively). B. Amount of DsRed or 2G12 produced in leaves of plants harvested 40 or 47 dps (right and left side of the y-axis, respectively). Numbers on y-axis correspond to the leaf age (1 = the oldest, non-cotyledon leaf).

Significant intra-plant variation for protein expression was observed among leaves of different ages and among different positions within a single leaf (Figure III.3 B). Therefore, these two significant sources of variation were included as model parameters inherent to the production platform (concentration model). For individual leaves, the average coefficient of variation (C.V.) between neighboring intercostal fields separated by the midrib was protein-dependent but was in both cases <10% and was therefore considered insignificant (Figure III.4 A).

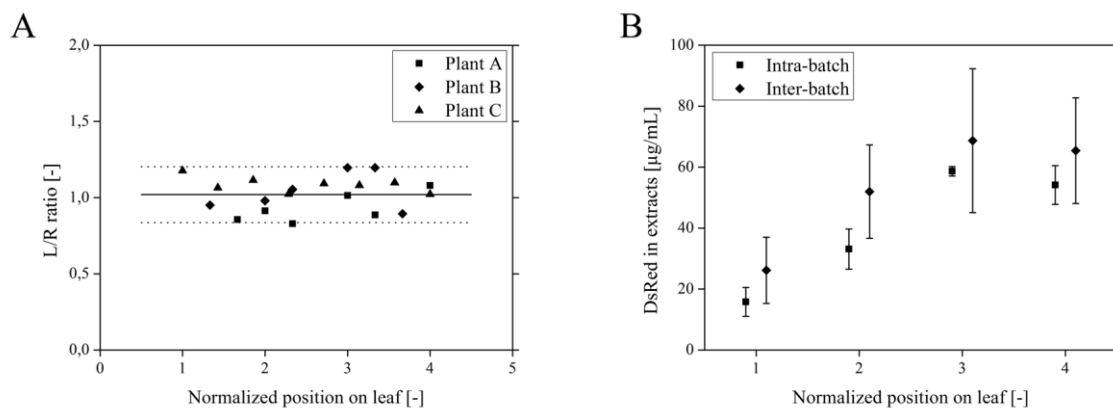


Figure III.4: Variation in DsRed levels within a single leaf (II.2.2).

A. Axial symmetry of transient protein expression in tobacco leaves. Data points represent the ratios of protein concentrations in extracts from sections of left side (relative to the mid vein) divided by those of corresponding sections from right side at different positions on the leaf (see Figure II.1 and Figure III.1 D for position numbering). Dashed lines indicate average inter-plant coefficient of variation ($n = 9$) over all leaf positions. Protein expression not biased by the leaf side will result in random scattering around the value 1, as observed here. B. Variation in DsRed levels at different positions within the leaf among different plants in a single batch (intra-batch, squares; $n = 3$) and among plants of different batches (inter-batch, diamonds; $n = 3$). Note that x-axis positions for both data series are shifted by 0.1 from the corresponding integer for clarity.

Although treated as insignificant, the C.V. for intra-batch variation was marginally greater than 15% (Figure III.4 B), which was attributed in part to non-avoidable sampling artifacts. Individual variations in leaf morphology mean that the corresponding intercostal fields from any two leaves differ slightly in position, increasing the C.V., and it becomes impossible to separate intra-plant and inter-plant variation. A C.V. of 18–30% for transient protein expression in *N. benthamiana* has been reported in recent publications [66, 67] and even transgenic plants exhibit a similar degree of inter-plant variation with a C.V. $\geq 18\%$ (Dr. Jürgen Drossard, unpublished data) [162].

Inter-batch variation was found to be more significant, with a C.V. of $\sim 33\%$ (Figure III.4 B). However, similar values (30–35%) have been found in transgenic plants (Dr. Jürgen Drossard, unpublished data) and have also been reported in the literature [162]. Indeed, because the inter-batch variation was $< 50\%$, introducing the batch number as a block factor could partially compensate for this variability during model building. Larger batch sizes are also likely to reduce C.V. values because individual plants would have a smaller impact. The intra-plant, inter-plant and inter-batch variation in transient expression therefore appeared similar to that in transgenic plants. The conclusion was that a predictive model for transient protein expression will yield valuable information about the system in terms of process design.

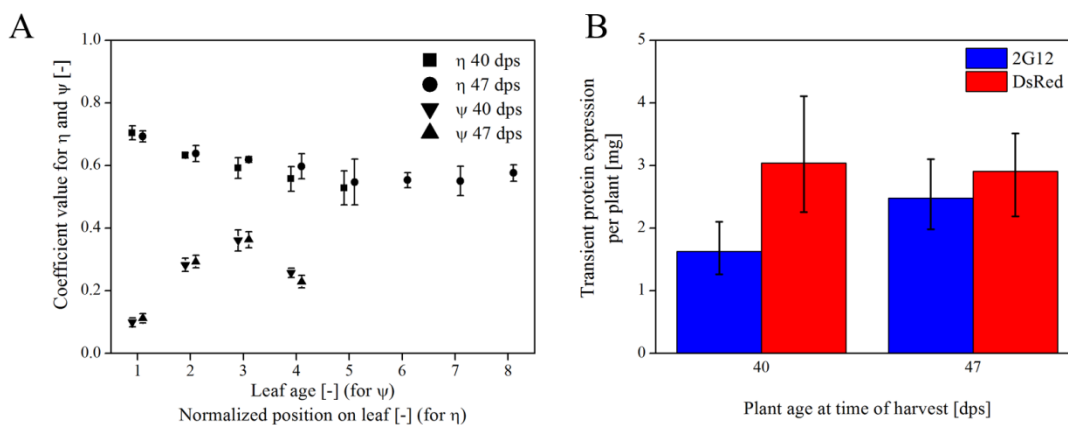


Figure III.5: Experimentally determined ratios η and ψ and the prediction of total recombinant protein produced per plant (II.5.1).

A. Ratios were determined for each leaf of plants harvested 40 and 47 dps, respectively. Error bars indicate standard deviation, $n = 3$ for plants harvested at 40 dps and $n = 20$ for those harvested at 47 dps. B. Amounts of DsRed and 2G12 antibody produced per plant based on model prediction. Error bars indicate 95% C.I.

III.1.2.2 Volume model

As discussed above, the values of three ratios were determined because they were necessary for the transformation of leaf biomass into extract volume (η_k , ε and $\psi_{k,p}$) at two different growth stages, 40 and 47 dps (Figure III.5 A). Transformation ratios η_k and $\psi_{k,p}$ were similar

for plants harvested at both stages and the extract ratio ε was the same in both groups, with a value of 2.75 ± 0.16 ($n = 50$). As expected, older plants had a greater per leaf biomass than younger plants, and therefore produced larger extract volumes (Figure III.3 A). Second-order polynomial fits through the four position-specific volumes of each leaf resulted in r^2 values of ~ 0.98 for most leaves at both growth stages. The lowest r^2 value was 0.93.

III.1.2.3 Concentration model

A consensus model was then built for all four combinations of recombinant protein and plant age by including all factors meeting the 5% significance level in at least one of the individual models. This final model also contained the factor p^2 (quadratic effect of the position on leaf) although it did not meet the stated significance level. This factor was included because the analysis of single leaves clearly indicated that leaf position resulted in a quadratic effect (Figure III.4 B and Figure III.5 A). The inter-batch C.V. of $\sim 33\%$ discussed above probably obscured the impact of this factor. However, including p^2 in the model will improve the overall quality because it reflects the actual concentration distribution on a leaf. The final model was then used to calculate the concentrations of both 2G12 and DsRed at both growth stages (Equation 12, Figure III.6).

Temperature was found to have the most significant impact on transient protein expression. The optimal temperature range for 2G12 production was broad, with a maximum at $\sim 21^\circ\text{C}$ (Figure III.2), whereas DsRed expression increased steadily up to a maximum temperature of $\sim 25^\circ\text{C}$ followed by a steep reduction at higher temperatures (Figure III.7). The parabolic dependence of protein expression on temperature may reflect the limited activation energy of participating enzymes and the onset of temperature stress, e.g. the increased synthesis of chaperones, limiting protein synthesis within the low and high temperature zones, respectively. The efficiency of transformation by *A. tumefaciens* may also be reduced at elevated temperatures [163]. These data are in good agreement with previous findings [65].

Leaf age was another important factor influencing protein expression, probably reflecting the greater potential for protein synthesis in young, growing leaves compared to mature ones (Figure III.8) [66, 67, 164]. This reflects the finding presented here that protein yields were higher in leaves from plants harvested at 40 dps compared to those harvested seven days later. However, the pattern may shift for other plant species [68]. The spike in transient protein expression levels between the fourth and fifth leaves of plants harvested at 47 dps presumably reflects a sink-source transition (Figure III.3 B, Figure III.8 B and C).

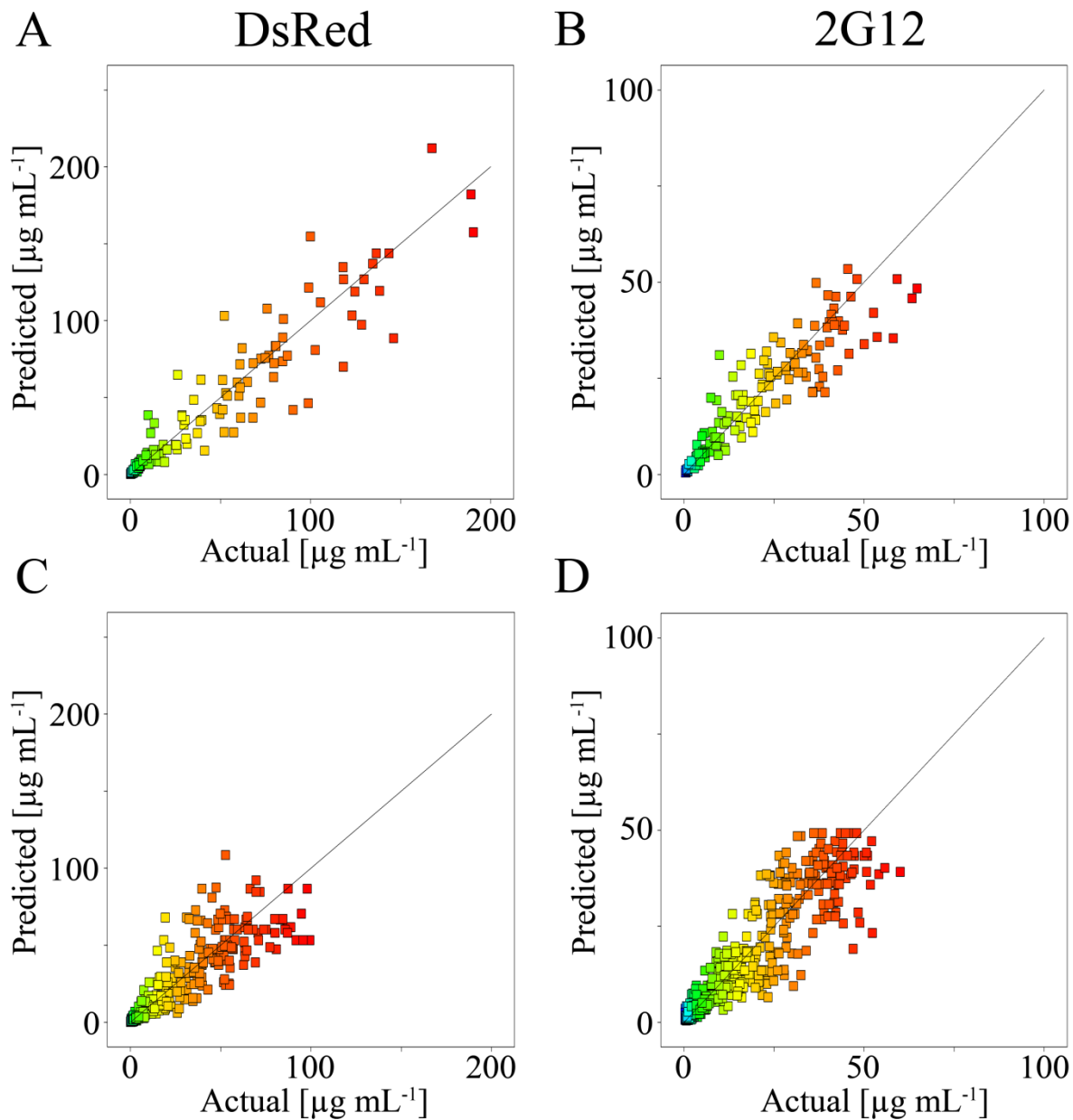


Figure III.6: Predicted versus actual plots for models of transient DsRed (A+C) and 2G12 (B+D) expression in plants harvested at 40 dps (top row) or 47 dps (bottom row) (II.4.1 and II.5.1). Gray diagonal corresponds to the ideal model, i.e. perfect match between actual and predicted values.

The third major factor influencing transient expression was the age of the plant, with lower yields in older plants. The basis of this observation is that young plants expressed large amounts of the recombinant proteins in all leaves (Figure III.3 B), whereas in older plants there was a clear division between younger and older leaves, as already discussed. Plant age-dependent recombinant protein expression has been noted by others [68]. Expression levels increase with age in the related tobacco species *N. benthamiana*, which was attributed to species-dependent differences [66].

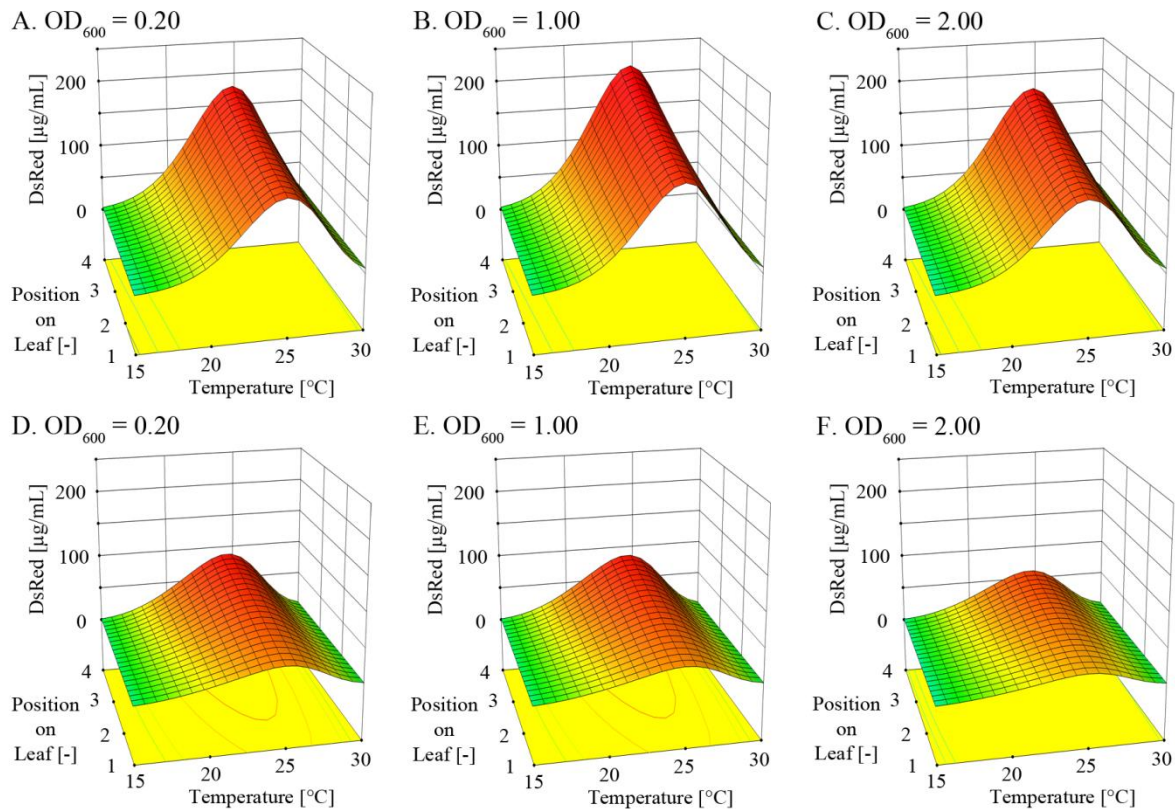


Figure III.7: The concentration of DsRed [$\mu\text{g mL}^{-1}$] in leaf extracts as a function of temperature and position on the leaf (II.4.1 and II.5.1).

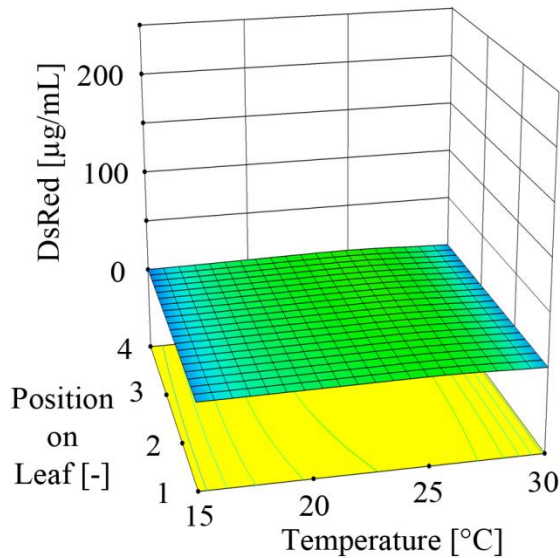
A. tumefaciens solutions of different $OD_{600\text{nm}}$ values during injection showed only minor effects protein expression levels, whereas a distinct temperature optimum was found. Top row: injection into leaf 4 of tobacco plants at 35 dps (harvest at 40 dps); bottom row: injection into leaf 6 of tobacco plants at 42 dps (harvest at 47 dps).

The position within the leaf also had a significant impact on transient expression levels at least in the middle range. The youngest leaves showed uniformly high expression and the oldest leaves showed uniformly low expression, but between these extremes there was an age-dependent gradient of expression from node to tip with expression beginning to decline first at the node and then progressively towards the leaf tip over time (Figure III.8). This might reflect the underlying age-dependent pattern of mitotic activity which correlates with reduced protein synthesis [165]. Others have also reported position-specific transient protein expression in leaves, although as above the observed patterns appear to be species-dependent [65].

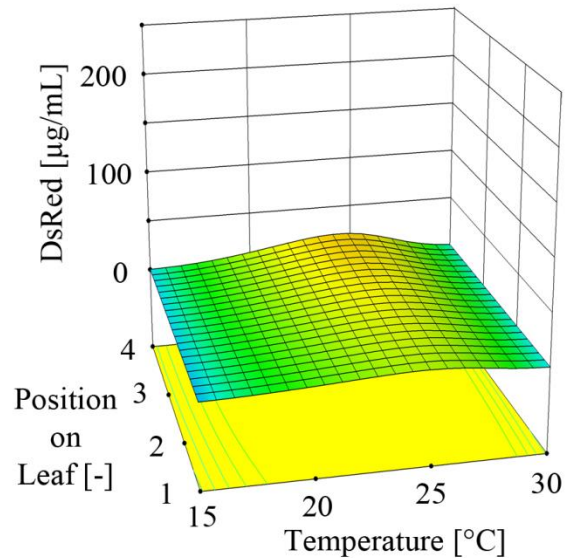
Finally, the yield of recombinant protein depended on the $OD_{600\text{nm}}$ of the injected *A. tumefaciens* culture, with the highest levels generated following infiltration with cultures at an $OD_{600\text{nm}}$ of ~ 1.0 (Figure III.7). The influence of this factor was low compared to the others discussed above and large changes only shifted the recombinant protein yields by $\pm 10\text{--}15\%$. This suggests that *A. tumefaciens* solutions with an $OD_{600\text{nm}}$ of 0.25 could be used for infiltration without a significant loss of yield, thus lowering production costs by reducing the

scale of bacterial cultures required for infiltration. Similar OD_{600nm} effects have been reported for *N. benthamiana* and lettuce [65, 67].

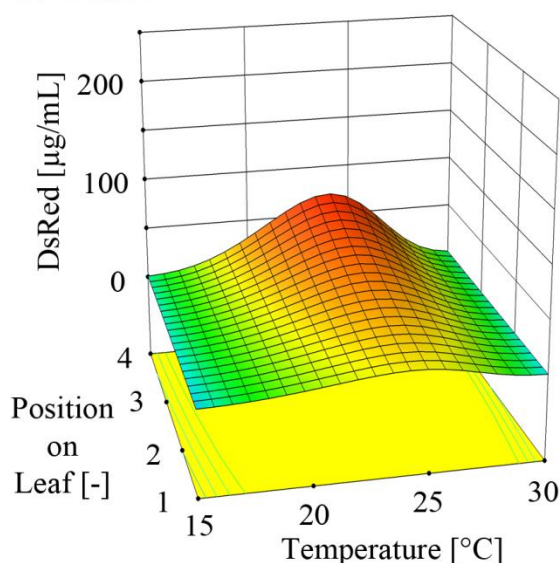
A. Leaf 1



B. Leaf 4



C. Leaf 5



D. Leaf 8

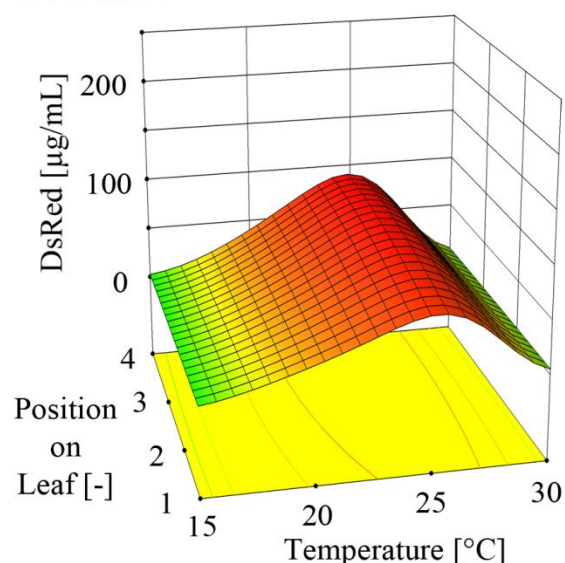


Figure III.8: The concentration of DsRed [$\mu\text{g mL}^{-1}$] in leaf extracts from tobacco plants injected with *A. tumefaciens* ($OD_{600nm} = 1.0$) at 42 dps is dependent on the developmental stage of a leaf (II.4.1 and II.5.1).

Sections A–D illustrate calculations for leaves at different stages: leaf 1 is the first non-cotyledon leaf, leaf 8 is the youngest fully expanded leaf. The concentration of DsRed is presented as a function of post-infiltration incubation temperature and position on a leaf.

III.1.2.4 Yield model

The yield model was built by combining the volume and concentration models as described above (III.1.1.5), allowing the yield of recombinant protein to be predicted for individual leaves (Figure III.3 B) and whole plants (Figure III.5 B). Protein expression per unit biomass

was higher in younger than older plants, especially in the case of DsRed. Younger plants produced the same amount of DsRed as older plants despite having only half the biomass, and they produced 66% as much 2G12. The difference in expression levels relative to biomass may reflect the unique properties of each protein but could also be explained by the secretion of the antibody to the apoplast versus the import of DsRed into the plastids, i.e. each protein will be exposed to environments that influence stability in different ways [40, 43, 91].

III.1.2.5 Model testing

The yield models (including the volume and concentration models) were tested for their predictive accuracy in tobacco leaves vacuum infiltrated with *A. tumefaciens* carrying the DsRed/2G12 expression vector and harvested at 40 dps (5 days after infiltration). The yield of DsRed in each position of leaf 3 was found to fall well within the 95% confidence interval (CI) predicted by the model, whereas the yield of 2G12 was slightly lower than the 95% CI at one leaf position (Figure III.9). All averages of 2G12 expression fell within the lower range of the 95% CI. This trend can mainly be attributed to a batch effect because all plants used for model validation came from the same batch, whereas the models were built using samples from different batches. The use of agroinfiltration rather than injection may also have influenced transgene expression [166]. However, observed and predicted average yields matched well for both recombinant proteins, confirming the validity of the yield models.

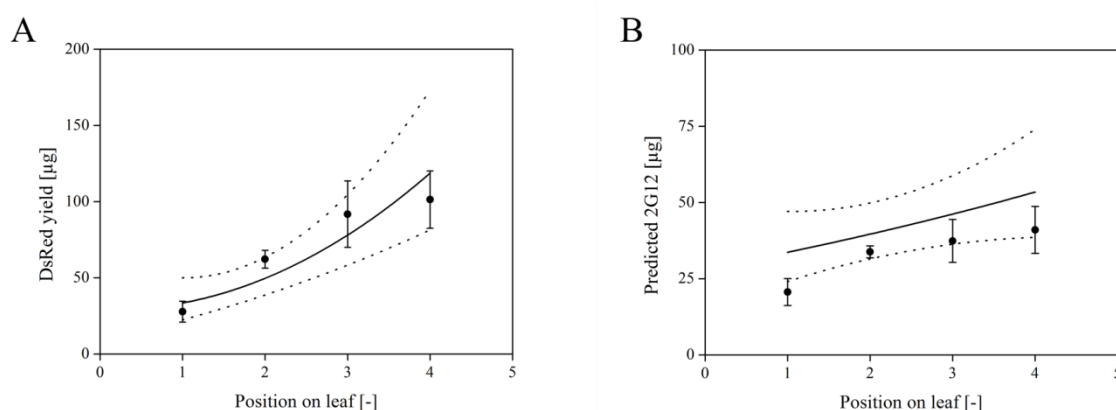


Figure III.9: Validation of the DsRed (A) and 2G12 (B) yield models (II.2.2 and II.4.2).

Extracts from whole infiltrated leaves (stage 3) from tobacco plants harvested 40 dps were used for model validation. Observed (circles, bars indicate standard deviation, $n = 3$) and predicted (solid line marks prediction average, dotted lines mark 95% CI) expression levels for both model proteins matched well.

III.1.2.6 Cost function

The cost function as described above (III.1.1.6) was applied to the yield models to provide insight into production optimization strategies. According to the yield models, the greatest proportion of each recombinant protein was located in young leaves (Figure III.3 B) allowing

the evaluation of the economic impact of several harvest schemes differing in the number and type of leaves processed per plant. In consecutive harvest schemes, the oldest leaves were iteratively excluded from processing. A process output of 5 g purified recombinant protein was used as the basis for these calculations, corresponding to approximately 7.14 g of bulk recombinant protein assuming a *RF* of 0.7. The most significant cost savings were achieved by using plants harvested at 47 dps and excluding the oldest four leaves (Table III.4). This reduced production costs for DsRed by 23% and for 2G12 by 12%. The major benefit from this scheme was to reduce downstream processing costs by 45% for DsRed and 35% for 2G12, which absorbed the moderate increase in upstream cultivation costs (Figure III.10 A and Table III.4). Reducing the number of harvested leaves per plant further did not achieve additional cost savings overall because downstream cost savings were offset by higher upstream costs. The exclusion of older leaves did not result in such significant savings when plants were harvested at 40 dps. The greatest impact was achieved by excluding only one leaf, the oldest, but this reduced the costs of producing DsRed only marginally (5%) whereas the costs of 2G12 production were not affected. These differences reflect the fact that younger plants produce higher levels of each protein per unit biomass so leaves cannot be discarded without also discarding significant amounts of the target product. Direct comparison of plants harvested at 40 and 47 dps revealed that production costs were lower for the young plants (45% for DsRed and 8% for 2G12 assuming the processing of all leaves) and this was without taking into account the shorter production cycle with younger plants, meaning that more batches can be processed each year. The cost function presented here can easily be adapted to other downstream unit operations by replacing the downstream cost factor (*DCF*) in Equation 13 with alternatives [167].

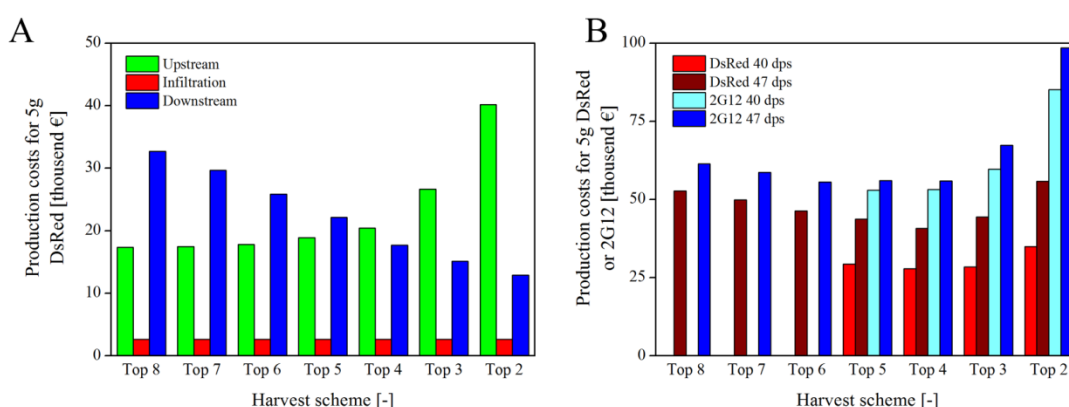


Figure III.10: Process costs for the production of 5 g DsRed or 2G12 based on different harvest schemes and age at the infiltration stage (II.2.2, II.4.2 and III.1.1.6).

A. The different cost positions accounting for the total production costs of DsRed in plants harvested 47 dps. B. Harvest schemes for the two recombinant proteins are compared in terms of production costs and age of the harvested plants.

Table III.4: Batch size and process costs according to harvest scheme. Calculations represent a process yielding 5 g of purified recombinant protein (7.14 g bulk antibody assuming an *RF* of 0.7). Amounts of recombinant protein per plant were calculated using Equation 12 taking into account the different harvest schemes. Upstream production, infiltration and downstream processing costs were calculated with Equation 13 using GCF_A , IC_A and DCF from Table III.3.

Plant age	40 dps				47 dps							
	Top 5	Top 4	Top 3	Top 2	Top 8	Top 7	Top 6	Top 5	Top 4	Top 3	Top 2	
Harvest scheme												
2G12 per plant [μ g]	1627	1474	1151	684	2478	2431	2334	2051	1769	1224	696	
Required plants [-]	4391	4847	6205	10438	2883	2938	3060	3483	4038	5838	10268	
Biomass processed per plant [g]	29	24	18	11	66	59	51	41	30	20	11	
Biomass processed per batch [kg]	128	116	110	117	189	174	155	142	122	116	115	
Upstream costs [€]	2491	2750	3520	5922	1635	1667	1736	1976	2291	3312	5826	
Infiltration costs [€]	5	3	7	9	9	0	5	4	5	6	0	
Downstream costs [€]	2103	2103	2103	2103	2653	2653	2653	2653	2653	2653	2653	
Downstream costs [€]	2590	2358	2236	2380	3834	3526	3135	2879	2473	2341	2328	
Total costs [€]	6	0	1	4	4	6	3	7	9	8	2	
Total costs [€]	5292	5318	5967	8513	5735	5458	5137	5121	5030	5919	8419	
Total costs [€]	3	5	0	5	6	9	1	3	7	7	6	

Plant age	40 dps				47 dps							
	Top 5	Top 4	Top 3	Top 2	Top 8	Top 7	Top 6	Top 5	Top 4	Top 3	Top 2	
Harvest scheme												
DsRed per plant [μ g]	3041	2923	2523	1733	2907	2887	2831	2669	2470	1894	1255	
Required plants [-]	2349	2443	2831	4122	2457	2474	2524	2676	2892	3771	5691	
Biomass processed per plant [g]	29	24	18	11	66	59	51	41	30	20	11	
Biomass processed per batch [kg]	68	59	50	46	161	147	128	109	87	75	64	
Upstream costs [€]	1332	1386	1606	2338	1736	1747	1782	1890	2043	2664	4020	
Infiltration costs [€]	9	4	6	7	0	9	8	5	0	5	8	
Infiltration costs [€]	2103	2103	2103	2103	2653	2653	2653	2653	2653	2653	2653	
Downstream costs [€]	1385	1188	1020	9399	3268	2969	2585	2212	1771	1512	1290	
Downstream costs [€]	9	7	4		2	9	4	3	4	8	5	
Total costs [€]	2929	2785	2837	3488	5269	4983	4633	4368	4079	4442	5576	
Total costs [€]	0	3	2	9	5	2	5	1	7	6	7	

III.1.3 Model implications for a production process

The data presented here indicate that the degree of variation for transient protein expression in tobacco can be reduced to ~15%, which is in the same range as transgenic tobacco plants (unpublished data and [162]) and transient expression in *N. benthamiana* [66, 67]. There was significant inter-batch variation in protein yields, but this can be reduced by the careful

control of growth conditions (e.g. by cultivation in a phytotron) and by the use of larger batch sizes. Even with this amount of variation, it was possible to predict transient protein expression levels with reasonable accuracy using the concentration models presented here. According to these models, expression levels strongly depend on the cultivation temperature although the optimal temperature may need to be tailored for specific proteins and/or for proteins targeted to specific subcellular compartments [91]. The temperature dependence will be interesting to follow up in future experiments because the data suggest that yields could be increased by at least 10% using a temperature-optimization strategy. The volume and concentration models showed that transient protein expression levels declined as plants aged and generated more biomass, reflecting the greater potential for protein synthesis in young leaves. The yield model showed that the higher expression levels in young plants offset or even exceeded the benefits of more biomass in older plants, allowing young plants to be harvested for processing and allowing the production cycle to be shortened resulting in a higher overall annual output. This finding indicates that harvest times cannot be transferred between processes based on stable transformation and transient expression. The long-term accumulation of recombinant proteins in transgenic plants leads to higher overall yields, whereas in transient expression systems the best results are generated by taking advantage of young, rapidly-growing tissues with high levels of protein synthesis. The cost function highlighted the fact that young plants reduce downstream costs by 10–30% as specific protein expression is higher and less biomass needs to be processed. These cost savings can be increased by harvesting selected leaves.

In addition to the process optimization described above, the data collected here can form the basis of a QbD concept to develop a cGMP-compliant process for the transient expression of biopharmaceutical proteins in plants. For example, critical process parameters such as temperature were identified and their impact on product quantity was determined. In the future, it will be necessary to determine the effects of different growth conditions and leaf parts on the quality of the protein product rather than focusing on the yield alone, allowing the definition of an operating space for a production process. The generic character of the approach will allow its use with other expression platforms.

III.2 Promoter and 5'UTR strength in transient expression

The results presented in this section have been published as:

1. Buyel JF, Kaefer T, Buyel JJ, Fischer R. 2013. Predictive models for the accumulation of a fluorescent marker protein in tobacco leaves according to the promoter/5'UTR combination. *Biotechnology and Bioengineering* 110(2):471-82.

III.2.1 Vectors

Vectors were constructed as described in sections II.2.1 and II.2.4 as part of the bachelor thesis of Thomas Kaefer [148]. For promoter and UTR testing, two constitutive promoters were used to drive the *DsRed* transgene: the double-enhanced *Cauliflower mosaic virus* 35S promoter (CaMV 35SS) as an example of a strong promoter [146], and the *A. tumefaciens* Ti-plasmid nopaline synthase (*nos*) promoter as an example of a weak promoter [147, 168]. Each promoter was combined with three different 5'UTRs, namely the *Petroselinum hortense* chalcone synthase (CHS) gene 5'UTR [169], the omega prime sequence from *Tobacco mosaic virus* (hereafter called 'omega' for simplicity) [170] and the *Tobacco etch virus* leader sequence (TL) [171] (Figure III.11). A fourth construct was tested that consisted of the CHS 5'UTR followed by the leader peptide sequence of murine mAb24 [172] codon optimized for tobacco (LPH). This sequence ensures that nascent polypeptides are targeted to the secretory pathway through the endoplasmic reticulum (ER). The *DsRed* coding sequence was identical in all vectors and was modified to contain C-terminal His₆ and KDEL tags, the latter working in concert with the LPH sequence to ensure the protein accumulates in the ER rather than passing through the Golgi body for secretion to the apoplast [173]. Both tags were present in all constructs even when anticipated to provide no function, in order to minimize the differences in mRNA and protein sequences and therefore remove potential sources of variation. A reference gene was not used for normalization of *DsRed* expression because position effects require T-DNA integration and are therefore unlikely to influence transient expression [174]. There can also be interference between test and reference promoters in transient expression experiments [175].

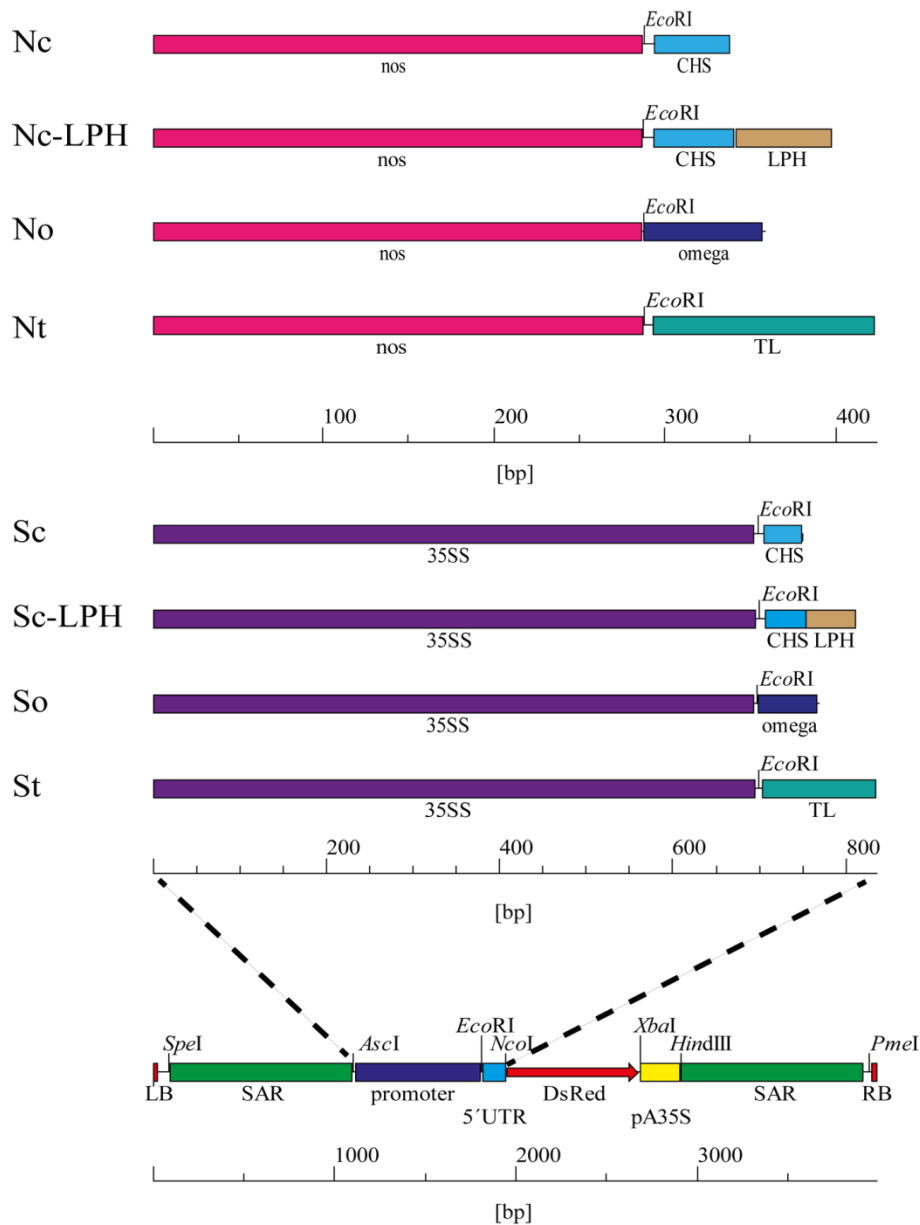


Figure III.11: Schematic representation of expression cassettes.

The promoter/5'UTR elements of the eight different vectors are shown at the top (II.2.1 and II.2.4). Starting with vector So, the omega 5'UTR was replaced by one of the three other UTRs (CHS, CHS-LPH or TL). In a second step the CaMV 35SS promoter was replaced with the *nos* promoter in each of the four initial constructs, yielding a total of eight different promoter/5'UTR combinations. The lower panel shows the T-DNA map.

III.2.2 Effects of promoter and 5'UTR

III.2.2.1 The predictive model captures promoter/5'UTR-dependent differences in expression

A model was established that accurately predicted the accumulation of DsRed in tobacco leaf extracts. The significance of the multiple linear regression model was confirmed using standard quality attributes such as r^2 (Table III.5). The incubation time was the most important factor affecting the accumulation of DsRed, but leaf age and position on a leaf were also significant factors as discussed in section III.1.2.3. Seventeen factors and factor

interactions had a significant impact on DsRed expression (Table III.6). Letter duplets in this table indicate an interaction between two of the five model factors: A (position on leaf), B (incubation time), C (leaf age), D (promoter) and E (5'UTR). For example “BD” denotes an interaction between factor B (incubation time) and factor D (promoter). Accordingly, letter triplets indicate three-factor interactions. In this context, an interaction means that the effect of two factors, e.g. B (incubation time) and D (promoter), on the observed result (DsRed concentration), are interdependent.

The model revealed that 11 of the 17 significant model terms (factors and interactions) involved the promoter, the 5'UTR or a combinatorial interaction between them (Table III.6). One consequence of these significant promoter/5'UTR interactions was that a given 5'UTR, although not considered strong by itself, might become a more important determinant of expression when combined with a certain promoter. For example the CHS 5'UTR achieved the highest DsRed accumulation when combined with the CaMV 35SS promoter but the lowest when combined with *nos*. The most significant interaction of promoter or 5'UTR involved the incubation time, e.g. BD and BE (Table III.6). The model therefore provides valuable information about transient protein expression patterns in tobacco leaves by making predictions based on interactions between factors that are often not considered in such experiments. The model is not mechanistic like some promoter models in prokaryotes [60], yeasts [61, 62] and *Drosophila melanogaster* embryos [63, 64] because the tobacco plant is too large and complex. However, it links promoter/5'UTR effects on protein accumulation with the spatiotemporal constraints of transient protein expression in whole plants. This will allow product yields in transient expression platforms to be estimated when concentration data from the model presented here are combined with the reported model for leaf mass or extract volumes described above (III.1). The model also reveals which leaves are most suitable for comparative experiments (in terms of expression levels and yields), which will help to standardize comparative tests. In contrast, previous studies have either focused on promoter analysis alone [69, 70] or have used cell suspension cultures to simplify the comparison [54, 56, 57, 71-73]. The temporal accumulation profiles of reporter proteins have only been studied in transgenic plants [74, 75, 77] and in some cases have not resolved differences among leaves of different ages [76].

Table III.5: Parameters used to confirm the significance of the predictive model.

Evaluation parameter	Value
r^2	0.993
Adjusted r^2	0.987
Predicted r^2	0.976
PRESS ¹	4.054
Lack of Fit	0.267

¹ Predicted residual sum of squares**Table III.6: Factors included in the model to predict reporter protein expression. Factors involving the promoter and 5'UTR are highlighted in bold and italics, respectively. A p-value of 0.05 indicates a significance (alpha) level of 5%.**

Factor	F-value	p-value
<i>Model</i>	<i>182.124</i>	<i>< 0.0001</i>
A (position on leaf)	16.056	0.0001
B (incubation time)	11128.220	< 0.0001
C (leaf age)	215.390	< 0.0001
D (promoter)	2324.292	< 0.0001
<i>E (5'UTR)</i>	<i>30.610</i>	<i>< 0.0001</i>
AC	3.382	0.0026
BC	20.640	< 0.0001
BD	224.509	< 0.0001
<i>BE</i>	<i>28.744</i>	<i>< 0.0001</i>
CD	4.108	0.0005
<i>CE</i>	<i>2.036</i>	<i>0.0093</i>
<i>DE</i>	<i>15.730</i>	<i>< 0.0001</i>
B ²	627.288	< 0.0001
BCD	13.424	< 0.0001
<i>BCE</i>	<i>1.831</i>	<i>0.0230</i>
<i>BDE</i>	<i>5.370</i>	<i>0.0017</i>
B²D	147.927	< 0.0001

III.2.2.2 The optimal combination is the CaMV 35SS promoter and CHS 5'UTR, which achieved the highest level of DsRed accumulation within 8 days

The CaMV 35SS promoter achieved an average 10-fold higher yield of DsRed than the *nos* promoter at both 5 and 8 dpi. The difference in DsRed accumulation also depended on which 5'UTR was combined with the promoter (Figure III.12). A 110-fold difference in protein levels was reported previously, but each promoter was also linked to a significant amount of intrinsic 5'UTR sequence >50 bp in length [176]. In contrast, differences in accumulation can be attributed directly to the promoter because the residual intrinsic 5'UTRs were only 6 bp for the CaMV 35SS promoter and 21 bp for the *nos* promoter, whereas the promoter-independent 5'UTRs tested here were up to 131 bp in length. Therefore, it was assumed that the 10-fold

difference in recombinant protein accumulation more closely reflects the differences in promoter activity. Some bias may remain in these results due to the short intrinsic 5'UTRs as well as differences based on interactions between the 5'UTRs and the 3'UTR. The difference between two promoters was greatest (12–14-fold) when the promoters were combined with the CHS 5'UTR and was less remarkable (4–7-fold) when the promoters were combined with the CHS-LPH 5'UTR. Averaged over all leaves in the plant, the CaMV 35SS promoter/CHS 5'UTR combination achieved the highest DsRed yields at 5 and 8 dpi (Figure III.13 A and C and Figure III.14 A and C). Thus, the shortest 5'UTR with the lowest GC content (Figure II.1 C) resulted in the highest accumulation of recombinant protein (Figure III.12) contrasting with a previous publication attributing higher expression levels to longer 5'UTRs [177].

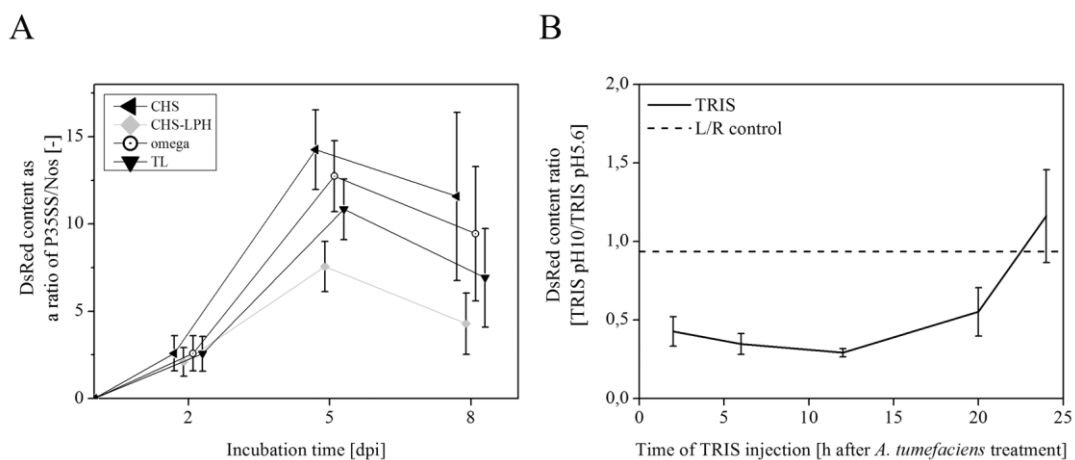


Figure III.12: A. Comparison of CaMV 35SS and *nos* promoter activity in terms of DsRed accumulation over time in tobacco leaves (II.2.2 and II.4.2).

Symbols indicate DsRed accumulation as a 35SS:*nos* ratio over all leaves in an individual plant at leaf position 2.5. Error bars represent standard deviations calculated across all leaves. The large standard deviations at 5 and 8 dpi were caused by DsRed accumulation varying significantly according to the age of the leaf. Abbreviations for 5'UTRs: CHS – chalcone synthase; CHS-LPH – chalcone synthase with leader peptide; omega – TMV omega sequence; TL – TEV leader sequence. Data points are shifted by 0.1 or 0.3 dpi away from the actual x-axis value for clarity. B. DsRed accumulation in tobacco leaves can be inhibited by injecting Tris buffer (pH 10.0) after infiltration, although the impact is time-dependent. Symbols represent the ratio of DsRed content in leaves treated with Tris pH10 to controls treated with Tris pH 5.6. Error bars indicate standard deviation ($n = 3$). Effects caused by differences on the left and right hand side of a leaf (L/R control, dashed line) do not deviate significantly from unity.

Others have shown that protein expression can vary significantly despite similar 5'UTR lengths and numbers of transcripts [55]. However, the results presented here agree with previous reports stating that high protein expression levels are often associated with short 5'UTRs with a low GC content [178, 179]. A high GC content can result in stable mRNA secondary structures, hindering effective translation if the resulting Gibbs energy (ΔG) is below -200 kJ mol^{-1} ($-50 \text{ kcal mol}^{-1}$) [180-182]. Data from the RNAfold webserver [145], indicated that only the TL 5'UTR combined with the 35SS promoter exhibited a folding

energy in this range (approximately -180 kJ mol^{-1}), whereas all the other 5'UTRs combined with the CaMV 35SS or *nos* promoter showed weaker ΔG values, all approximately -80 kJ mol^{-1} (Table III.7).

Table III.7: Gibbs energies of folding calculated for the 5' end of mRNAs including the 5'UTR and the first 50 bp of the DsRed coding region at 25°C.

$\Delta G \text{ [kJ mol}^{-1}\text{]}$	Promoter	
5'UTR	35SS	<i>nos</i>
CHS	63.55	92.66
CHS-LPH	88.97	88.93
Omega	65.60	94.71
TL	179.62	126.69

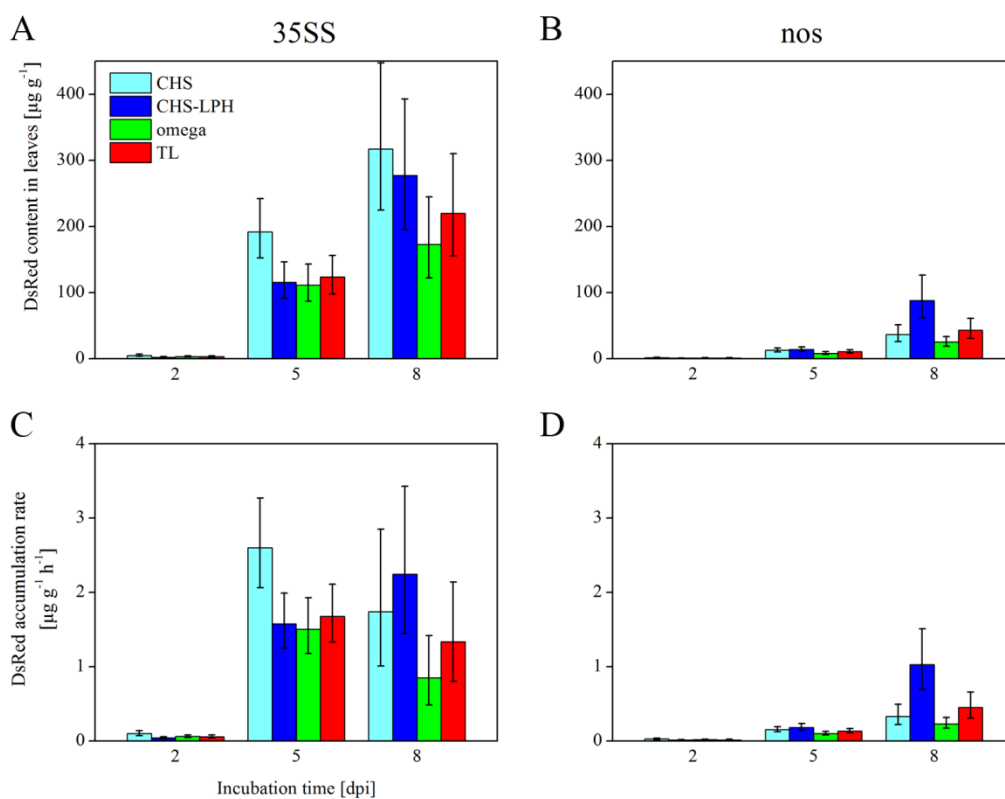


Figure III.13: Comparison of promoter/5'UTR combinations in terms of predicted DsRed content and accumulation rates over time in the fifth leaf of a tobacco plant at leaf position 2.5 (II.2.2, II.4.2 and II.5.2).

A. and B. show DsRed content achieved by 5'UTRs combined with CaMV 35SS or *nos* respectively. Corresponding accumulation rates are shown in C. and D. Error bars indicate 95% confidence interval of the model prediction. Abbreviations for 5'UTRs: CHS – chalcone synthase; CHS-LPH – chalcone synthase with leader peptide; omega – TMV omega sequence; TL – TEV leader sequence.

Accordingly, one would expect the CaMV 35SS-TL combination to achieve limited DsRed synthesis, but the observed expression levels were similar to those of the CaMV 35SS-omega combination (Figure III.13 A and C and Figure III.14 A and C). This may reflect the presence of an internal ribosome entry site (IRES) in the TL 5'UTR [183], which mediates

cap-independent translational initiation thereby circumventing the need for the ribosome to pass through mRNA secondary structures before reaching the start codon [184, 185]. Interestingly, the differences in the folding energies among the 5'UTRs were smaller when expression was driven by the *nos* promoter.

None of the constructs contained upstream AUG nucleotide triplets, so these were excluded as a possible reason for differences in the performance of the 5'UTRs [178, 180]. However, it is possible that the eight promoter/5'UTR constructs could generate mRNAs with different *cis*-acting elements that could either recruit different sets of translation factors, or the same factors with differing efficiency and stability, thus contributing to the different DsRed expression levels [186-188].

These results suggest that the sequence and structure of the 5'UTR are more important than its length in terms of the regulation of recombinant protein expression. Additional promoter/5'UTR combinations should be investigated in the future to quantify the effect of factors, such as folding energy, which influence the 'strength' of a 5'UTR.

III.2.2.3 DsRed accumulation rates change over time and are dependent on the promoter, 5'UTR and leaf age

During the first two days after agroinfiltration, less than 10 µg of recombinant protein per gram fresh biomass was detected in any of the tobacco leaves, regardless of the promoter/5'UTR combination (data not shown). This delayed onset of DsRed accumulation (Figure III.13 C and D) resembled the lag phase that occurs in cell cultures and probably reflects the initial transformation events that take place prior to protein synthesis, such as the transfer of T-DNA from *A. tumefaciens* into the tobacco cells and its subsequent import into the nucleus [189]. This assumption is supported by the finding that transient protein expression can be impaired at least within the first 20 h after agroinfiltration by injecting 50 mM pH 10.0 Tris buffer into the apoplast to increase the pH (Figure III.12 B). Others have achieved similar results with different inhibition techniques [190, 191]. Measures that accelerate the accumulation of a recombinant protein during the lag phase may therefore help to improve the overall yield, e.g. optimized infiltration protocols, *vir* gene inducers that are more potent than acetosyringone or engineered VirD2 sequences that improve the efficiency of T-DNA import.

After 5 dpi, most DsRed was detected in leaves 5–7 (Figure III.14 A and B) in agreement with the results presented in section III.1.2.3 and earlier studies using transgenic plants [56]. For combinations including the CaMV 35SS promoter and any of the three 5'UTRs, the rate of DsRed accumulation rate peaked at 5 dpi in leaves 4–7 and declined

thereafter (Figure III.13 C). The accumulation rate continued to increase up to 8 dpi in the older leaves (1–3) and in the youngest leaf (8), but this was not sufficient to overcome the trend in the other leaves so the overall effect was to slow down protein accumulation. The flattening of DsRed accumulation after 5 dpi may reflect the onset of plant defense mechanisms such as gene silencing [101]. Viral silencing suppressors such as p19 from *Tomato bushy stunt virus* (TBSV) have been used successfully to counteract these defense mechanisms and ensure the high-level transient expression of recombinant proteins [192-194]. However, the use of silencing suppressors is limited to *N. benthamiana* because p19 from TBSV and other viruses elicits the hypersensitive response (HR) in *N. tabacum*, which offsets any benefits [195, 196]. Other suppressors have been discovered with similar activity to p19 [197], and some such as p20 from *Cucumber necrosis virus* (CNV), do not elicit the HR in *N. tabacum* [196]. Hence, these suppressors may be used in future experiments to improve recombinant protein expression in tobacco.

Alternatively, the decline in DsRed accumulation may reflect convergence towards steady-state equilibrium between synthesis and degradation [198, 199]. It is also possible that the model slightly emphasizes the degree of flattening due to the second-order polynomial fit of the impact of incubation time. As stated above, DsRed continued to accumulate at least up to 8 dpi in leaves 2 and 3 of the same plant, particularly those transformed with the CaMV 35SS/CHS combination, leading to a more homogenous DsRed content across the different leaves (Figure III.13 C). Therefore, prolonging the incubation time may help to improve recombinant protein yields in older leaves that would otherwise increase downstream processing costs without increasing product recovery (section III.1.2.6). The higher DsRed accumulation rates at 5 dpi may also explain the variability of up to 40% reported for transient expression [66]. This is because small errors in sampling times (independent variable) of ± 4 h will result in changes in protein concentration (dependent variable) of up to 20%. Sampling times should therefore be scheduled carefully when transient protein expression is used to investigate promoter/5'UTR properties. In this context, the prolonged incubation times may also help to improve batch-to-batch consistency in the production of biopharmaceutical proteins using transient expression systems.

Unlike the plants discussed above, no flattening of DsRed accumulation was observed in plants transformed with *nos* promoter constructs over the same incubation periods. Furthermore, when comparing the three 5'UTRs (without a targeting sequence) their ranking changed in terms of activity from CHS>TL>omega during early expression to TL>CHS>omega at 8 dpi (Figure III.13 B and D and Figure III.14 B and D). The highest

DsRed levels were found in leaves 4–7, in contrast to earlier reports showing that the youngest leaves have the lowest expression levels [126, 127]. This difference between transgenic plants and transient expression can reflect the duration of the production cycle, i.e. in transgenic plants proteins accumulate during the entire growth period so older leaves have time to accumulate more protein (40–50 days in the case of tobacco variety Petit Havana SR1). In contrast, the transient expression production cycle lasts 5–10 days and the capacity of the protein biosynthesis machinery is more important than long-term accumulation, i.e. the rate of protein synthesis is higher in young, rapidly-growing leaves [66, 67, 164, 200].

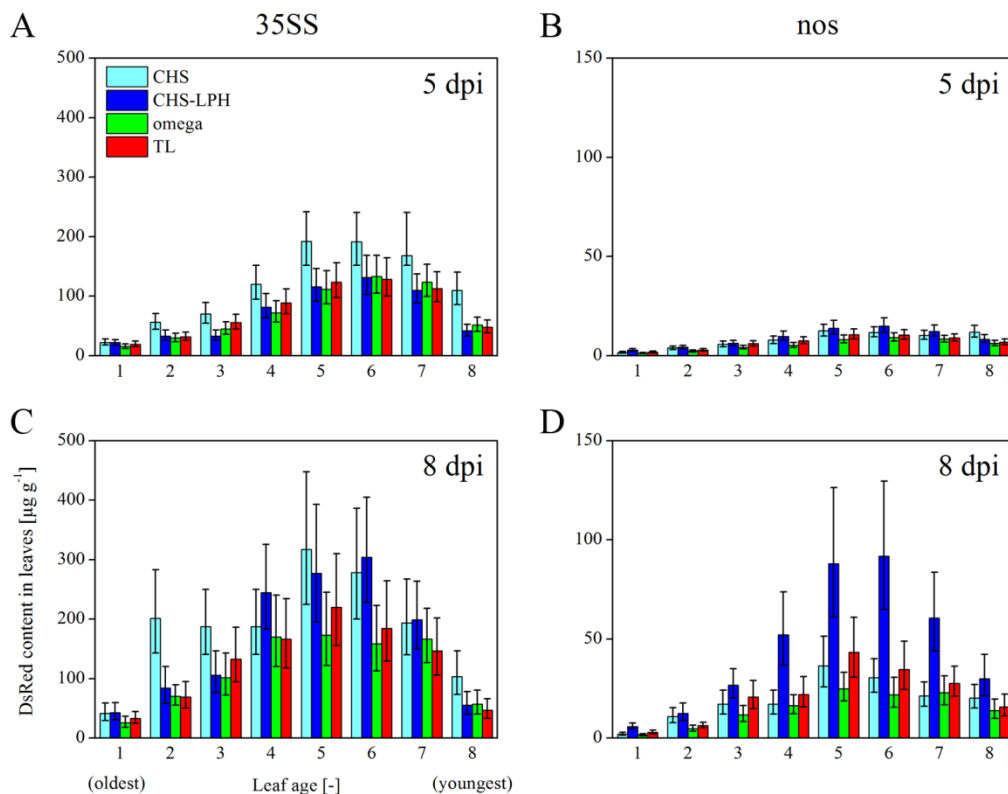


Figure III.14: Predicted DsRed content in tobacco leaves (1 = oldest, 8 = youngest) over the time course of transient expression is dependent on the promoter/5'UTR combination (II.2.2, II.4.2 and II.5.2). Expression driven by CaMV 35SS is higher at 5 and 8 dpi (A and C respectively) compared to *nos* (B and C respectively). Error bars indicate 95% confidence interval of the predicted values. Abbreviations for 5'UTRs: CHS – chalcone synthase; CHS-LPH – chalcone synthase with leader peptide; omega – TMV omega sequence; TL – TEV leader sequence.

III.2.2.4 ER localization increases the accumulation of DsRed over long incubation times and in combination with the *nos* promoter

The combination of the CaMV 35SS promoter and the CHS-LPH 5'UTR (which includes an ER-targeting signal) reduced DsRed accumulation to ~60% of the level achieved without the targeting signal (CHS 5'UTR) by the 5 dpi time point (Figure III.14 A). There was also no flattening off in the accumulation of DsRed, in contrast to the plants transformed with CaMV

35SS constructs combined with any of the three 5'UTRs but lacking a targeting signal. Indeed, plants transformed with the CaMV 35SS/CHS-LPH combination continued to accumulate DsRed and reached parity in overall expression levels by 8 dpi (Figure III.13 A and Figure III.14 C). When combined with the *nos* promoter, CHS-LPH even exceeded the performance of CHS.

Several factors may have delayed the accumulation of DsRed in the ER. First, mRNA stability might be affected by the LPH coding sequence, which may contain additional *cis*-acting elements [201, 202] although no known elements were identified using the RegRNA webserver [203]. However, other mechanisms such as miRNA-mediated repression/decay could also be involved [204]. Second, the number of translation events that can be completed in a given time may be lower for the CHS-LPH 5'UTR because the leader peptide increases the length of the polypeptide by 8% and might therefore also increase the total synthesis time for each polypeptide. Third, the affinity of signal recognition particles (SRP) towards leader peptides depends on the sequence of the leader [205]. The 35SS/CHS-LPH signal peptide competes directly with endogenous signal peptides for SRPs, and only a fraction of the 35SS/CHS-LPH leader peptide pool will bind to SRPs, dock to translocons in the ER membrane and finalize translation. This fraction will diminish if the 35SS/CHS-LPH leader has to compete with higher-affinity leader peptides such as those found on the immunoglobulin binding protein (BiP) [205]. The tobacco genome encodes several BiP proteins, all of which belong to the HSP70 family and are inducible by stress [206, 207]. Accordingly, they are likely to be activated by agroinfiltration and their synthesis and translocation into the ER may block leader peptides competing for the same SRPs. This is supported by the observation that DsRed accumulation driven by CHS-LPH is delayed compared to CHS without the leader peptide (Figure III.13 and Figure III.14). Finally, the strict quality control in the ER may reduce protein accumulation during the early expression phase by removing partially folded proteins [208]. The same mechanism may promote DsRed accumulation during longer incubation periods because proteins are more likely to be folded correctly.

When combined with the *nos* promoter, the CHS-LPH 5'UTR resulted in the highest DsRed accumulation in leaves 2–8 at both 5 and 8 dpi (Figure III.14 B and D). A 2–4-fold increase in DsRed fluorescence compared to the other 5'UTRs was observed at the later time point, and the different accumulation rates among the 5'UTRs were more prominent at this time point when combined with the *nos* promoter (Figure III.13 D). DsRed fluorescence in different leaves varied most at 8 dpi, with up to 20-fold differences in concentration between

young and old leaves. ER targeting can thus be used to achieve high rates of DsRed accumulation during transient expression if the incubation times are prolonged to account for the extended lag phase and to take advantage of the lack of flattening, perhaps reflecting a different steady-state dynamic in the ER favoring protein synthesis and accumulation over degradation. It has been reported that ER targeting can increase recombinant protein accumulation by more than 10-fold in transgenic plants [209, 210] which may reflect the long-term effect of the maintained high accumulation rate that was observed in the short-term transient expression experiments described here. However, prolonged incubation times can also be disadvantageous if silencing is triggered by the expression of the recombinant protein, as discussed above.

III.2.2.5 Different 5'UTRs form compatible pairs for stoichiometric protein expression when combined with the CaMV 35SS and *nos* promoters

It is often necessary to produce more than one protein at the same time while maintaining a certain stoichiometry, e.g. for protein-protein interaction studies or the production of heteromeric recombinant proteins such as antibodies [211-214]. This can be difficult to achieve by transient expression because protein contents vary according to leaf age, incubation time, promoter and 5'UTRs as discussed above (Table III.6). For example, the ideal stoichiometric ratio of proteins for the expression of a secretory IgA (sIgA) is 4:4:1:1, representing the IgA light chain, IgA heavy chain, joining chain and secretory component. However, only 45–50% of the recombinant proteins assemble into the full-size sIgA when the same promoter/5'UTR is used for all proteins [215, 216]. This low yield may be caused by the relatively high concentrations of joining chain and secretory component, which saturate the corresponding binding sites on IgA molecules (2:2:1:1 ratio) thereby preventing the assembly of two IgA molecules with one molecule of joining chain and secretory component.

Therefore, a pair of promoter/5'UTR combinations was identified that would support the most stable DsRed accumulation ratio in the different leaves of a tobacco plant and in the course of the incubation period. All 5'UTR pairs sharing the same promoter are compared in Figure III.15 A (e.g. 35SS/CHS:35SS/omega = CHS/omega) and 5'UTR pairs with different promoters are compared in Figure III.15 B (e.g. 35SS/CHS:nos/CHS = CHS/CHS).

The lowest fluctuations in DsRed accumulation were observed when the omega and TL 5'UTRs were paired with the CaMV 35SS promoter (Figure III.15 A). DsRed accumulation was approximately 10% higher when the TL 5'UTR was used, in agreement with data from transgenic plants [209]. In combination with the *nos* promoter the most consistent DsRed accumulation profiles were achieved when comparing the CHS and omega

5'UTRs. DsRed accumulation was approximately 50% higher when the CHS 5'UTR was used. The ratio of TL and omega exhibited the least variability across all leaves for each individual time point during incubation but the value of the ratio steadily increased. When 5'UTRs combined with different promoters were compared, DsRed accumulation ratios varied more than before according to the incubation time and leaf age, and the ratios at 2 dpi in particular were smaller compared to later time points (Figure III.15 B). This probably reflects the delayed onset of DsRed synthesis discussed above (sections III.2.2.3 and III.2.2.4). However, excluding the values for 2 dpi, the CaMV 35SS/CHS to *nos*/CHS, CaMV 35SS/omega to *nos*/CHS and CaMV 35SS/omega to *nos*/omega comparisons resulted in acceptable variability of approximately 30% over time and in different leaves. These data suggest that recombinant proteins can be produced at specific ratios by transient expression in plants using different promoters, which provides a valuable toolkit for the investigation of protein interactions, the stoichiometry between signaling proteins in complexes and metabolic enzymes, and for the production of heteromeric proteins such as sIgA.

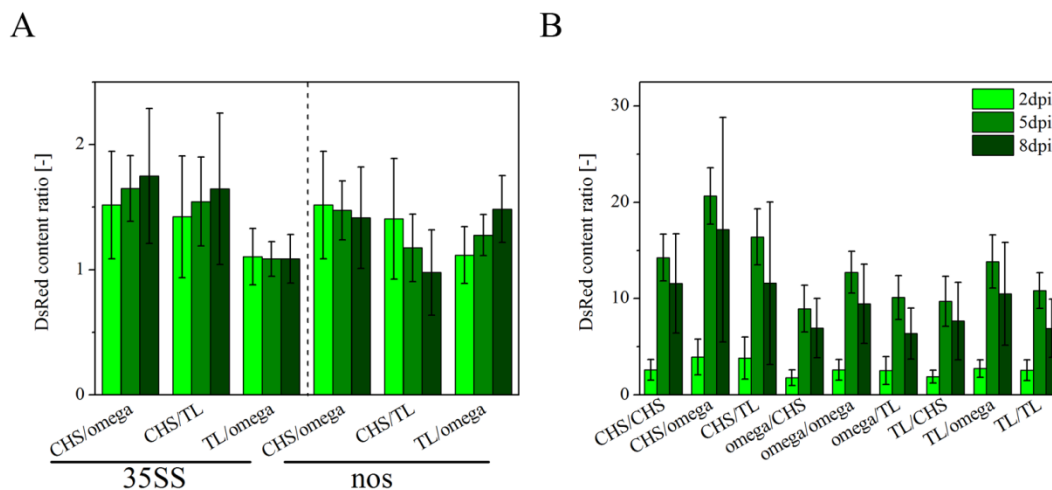


Figure III.15: Comparison of DsRed content resulting from different promoter/5'UTR combinations (II.2.2, II.4.2 and II.5.2).

A. Ratios of /5'UTR pairs with the same promoter (35SS – left of the dashed line; *nos* – right of the dashed line). B Ratios of 5'UTR pairs with different promoters. The 5'UTR in front of the slash was combined with the 35SS promoter whereas the 5'UTR after the slash was combined with the *nos* promoter. All columns represent averages over all leaves at position 2.5 and were calculated for the three incubation times. Error bars indicate the standard deviation of values averaged over all eight leaves of a plant. The ER-targeted construct was excluded from the comparison. Abbreviations for 5'UTRs: CHS – chalcone synthase; omega – TMV omega sequence; TL – TEV leader sequence.

The extrapolation of data from reporter proteins to others can be error-prone, but when DsRed was co-expressed with a monoclonal antibody by agroinfiltration, the two proteins showed highly similar expression profiles, both in terms of the accumulation and the spatiotemporal distribution (III.1.2.3). The antibody concentration was about half that of the DsRed protein (a homotetramer), as expected for a heterotetramer assembling from the

products of two different transgenes. Future studies should compare DsRed genes coding for the same amino acid sequence but using different nucleotide sequences to determine how the coding sequence of the reporter might affect protein accumulation. Additionally, the promoters/5'UTRs presented here should be combined with other genes to confirm the transferability of the model to other proteins and to test whether the high-level co-expression of several recombinant proteins interferes with protein accumulation.

III.2.3 Implications for the design of expression cassettes

Factors that influence recombinant protein levels in plants include promoter choice, 5'UTR choice, post-infiltration incubation time, subcellular localization and leaf age. These have been investigated individually but this is the first study to look at combinations of these factors in a unified model. The model presented here is able to capture the effect of promoter/5'UTR combinations, and characteristics of the expression system such as incubation time or leaf age, on the accumulation of the reporter protein DsRed. The model successfully predicted that the CaMV35SS/CHS combination would achieve the highest DsRed yield in all leaves, and that incubation for more than 5 days after agroinfiltration would improve the reproducibility of the transient expression platform. The collected data provide strong evidence that two recombinant proteins could be transiently expressed at a specific ratio in all the leaves of a tobacco plant and over a period of 8 days by the rational choice of promoter/5'UTR combinations. In future models, further regulatory elements should be included, such as 5'UTR introns and 3'UTRs [56, 58, 209, 217, 218]. This will improve the understanding of the factors that affect transient protein expression in whole plants providing a powerful tool for the development of manufacturing processes for plant-derived pharmaceutical proteins.

III.3 Transient expression of type III effectors

The results presented in this section are being prepared for the following publication:

1. Buyel JF, Buyel JJ, Haase C, Fischer R. Exploiting the potential of type III effectors from *Pseudomonas syringae* to enhance transient protein expression in plants.

III.3.1 Vectors

Vectors were constructed as described in sections II.2.1 and II.2.5. The *nos* promoter and omega 5'UTR were selected to prevent elevated TTE expression (III.2), and five TTEs were selected based on the broad coverage of different plant defense pathways (Figure I.1).

AvrPtoB interferes with PAMP recognition and the defense-related initiation of transcription [102, 115], whereas HopF2 and HopAO1 downregulate the MAPK cascade [119, 120]. HopI1 was used to prevent salicylic acid synthesis [108] and AvrRpt2 was used to modulate auxin levels [117]. The genes encoding AvrPtoB, AvrRpt2, HopAO1 and HopI1 were cloned in the pAIX-2 expression vector by Claudia Haase as part of her Bachelors thesis.

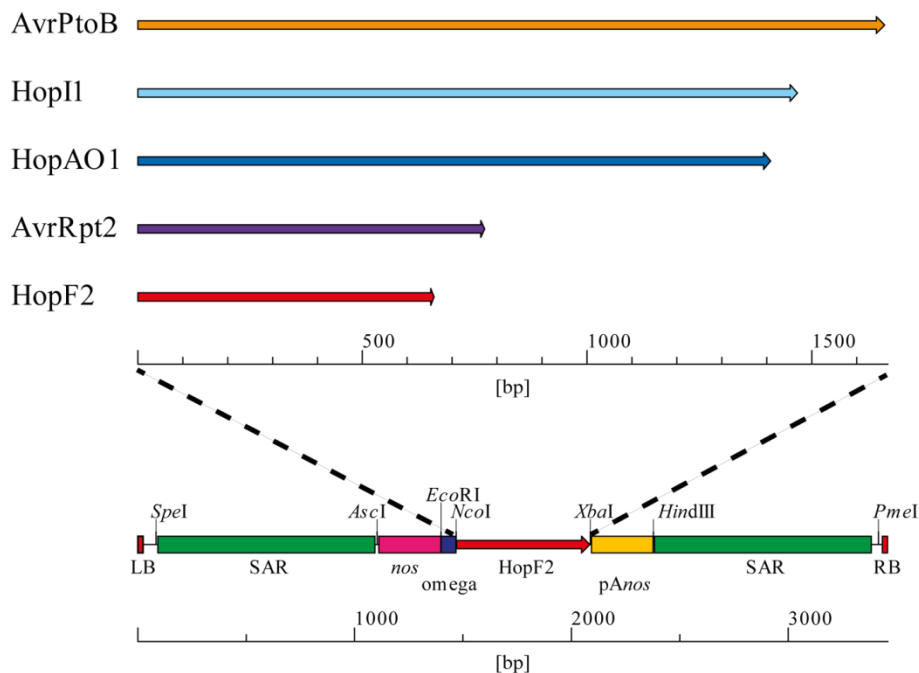


Figure III.16: Schematic representation of TTE genes in vector pAIX-2, which were used for single and multiple transient expression experiments in *N. tabacum* (II.2.2 and II.2.5).

The HopF2 vector was constructed first, and the coding region replaced with the other four genes encoding AvrRpt2, HopAO1, HopI1 or AvrPtoB. The expression cassette comprised the *nos* promoter, omega 5'UTR and pAnos poly(A) signal. The lower panel shows the T-DNA map.

III.3.2 Effects on plant tissue and defense responses

III.3.2.1 Reactions to bacterial injection

Before testing the impact of the five selected TTEs, the default sequence of defense responses elicited by the injection of bacteria was investigated. One response is the synthesis of proteins encoded by the introduced T-DNA (sections III.1 and III.2), but the plant also responds to the tissue damage caused by injection or infiltration and to the presence of the bacteria, which are likely to be recognized by pathogen receptors thus triggering the corresponding defense responses (I.6). Mechanical wounding occurs naturally when pests feed on leaves. *Nicotiana* species have evolved to induce the synthesis of the alkaloid nicotine under these circumstances so the leaves become unpalatable (hence the genus name). An increase in nicotine was detected 5 days after infiltration with *A. tumefaciens* (hereafter A_+ for brevity) or mock infiltration medium without bacteria (Figure III.17 A arrow 1). The mock injection did not induce any other changes, whereas A_+ also induced the synthesis of chlorogenic acid and

suppressed the synthesis of rutin (quercetin-3-O-rutinoside) (Figure III.17 A arrows 2 and 3 respectively). The identity of these metabolites was confirmed by comparison to authentic standards (Figure III.17 B). Chlorogenic acid and rutin are known defense-related metabolites in other plant species [219]. A_+ also induced the synthesis of several other metabolites whose identity is unknown (Figure III.17 A arrows labeled 4). The upregulation of toxic secondary metabolites during transient expression would be a matter of concern for the regulators and specific process steps to remove these substances might be required to satisfy safety demands when transient expression in tobacco is used for the production of biopharmaceutical proteins [220-222].

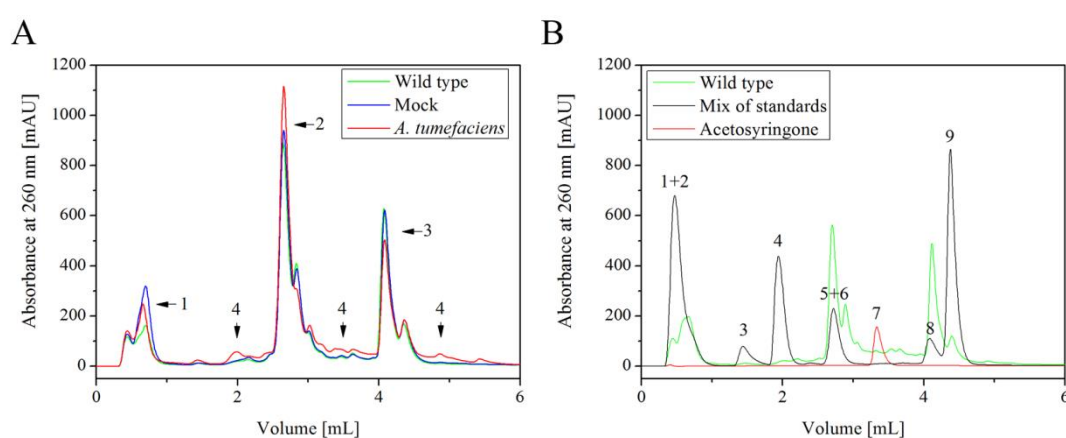


Figure III.17: Changes in the metabolic profile of tobacco leaves following injection with *A. tumefaciens*.

A. FPLC chromatograms from untreated, mock treated and *A. tumefaciens* (A_+) treated samples (II.2.5 and II.4.8). Elevated nicotine concentrations were found in mock and A_+ samples (arrow 1), whereas the levels of chlorogenic acid (arrow 2), rutin (arrow 3) and diverse uncharacterized metabolites (arrows 4) were modulated specifically in the A_+ samples. B. Authentic standards were used to identify the metabolites (Table II.5).

Table III.8: Elution order of authentic standards used for identification of tobacco metabolites.

Elution number ¹	Name of the substance
1	4-Aminopyridine
2	Nicotine
3	Quinoline
4	2,4-Dipyridyl
5	Caffeic acid
6	Chlorogenic acid
7	Acetosyringone
8	Rutin
9	Cinnamic acid

¹ Elution number according to Figure III.17 B.

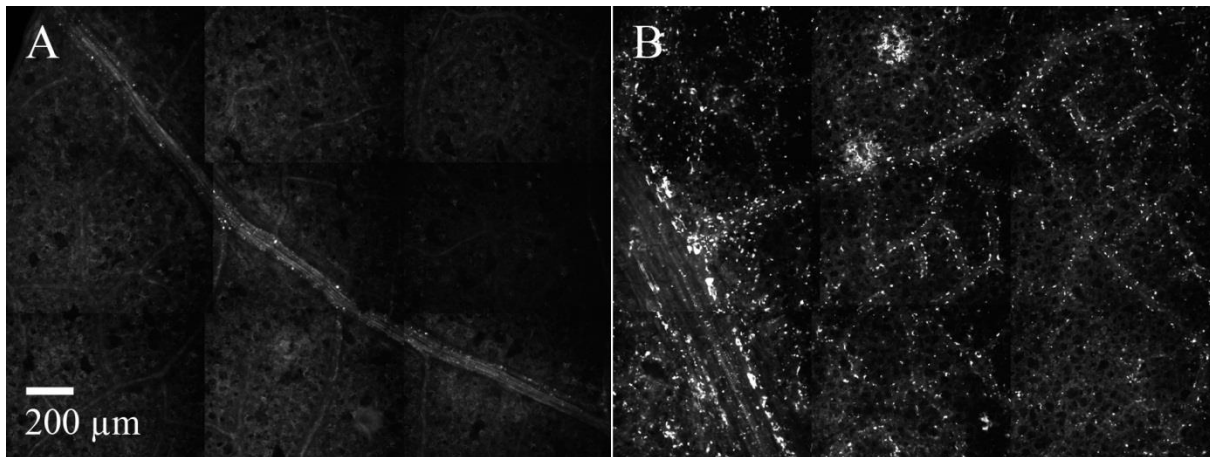


Figure III.18: Callose deposits (white spots) in tobacco leaves observed by automated confocal microscopy (II.2.5, II.4.6 and II.4.7).

A. Few callose deposits were observed in the untreated leaves and those detected were located close to veins. B. A_+ samples contained many callose deposits, with intense staining close to the veins but also across the intercostal fields.

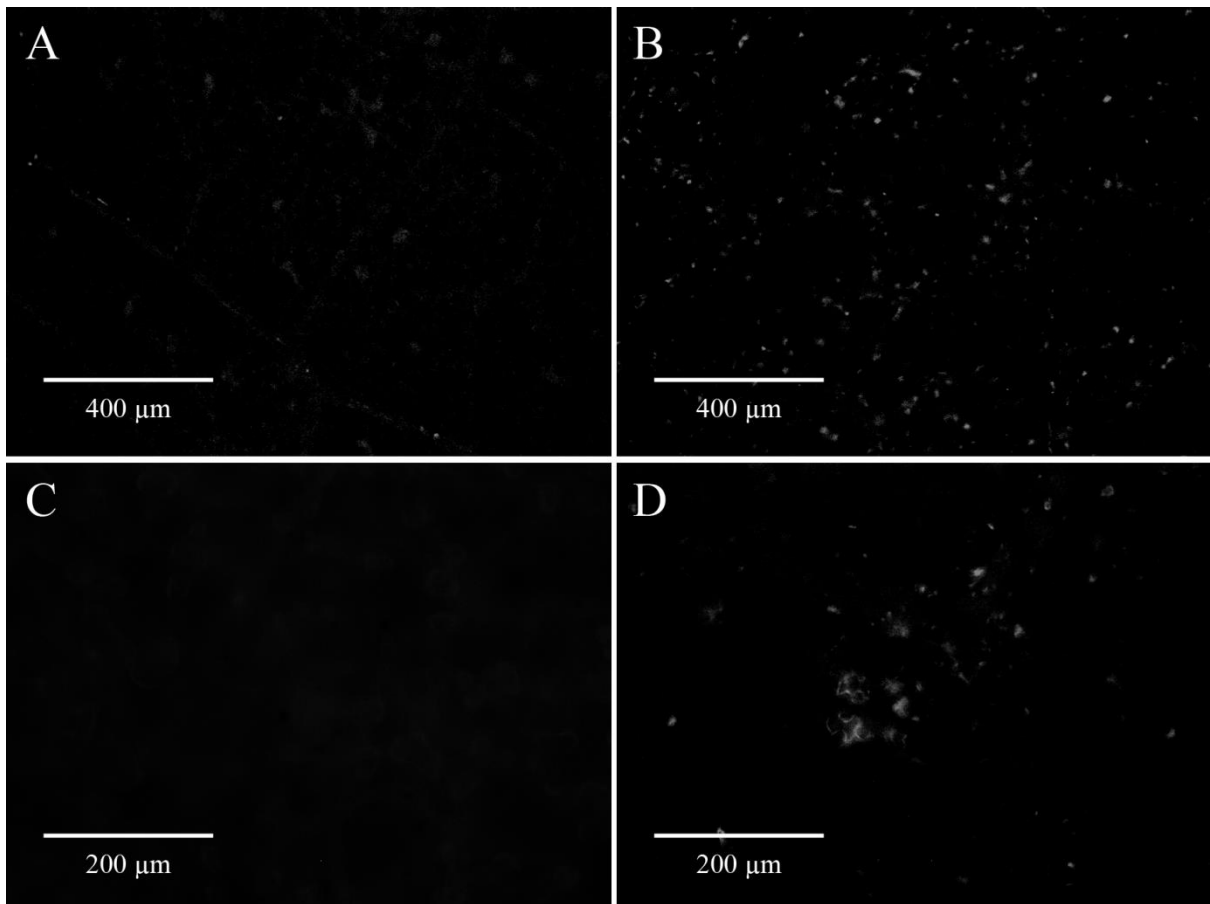


Figure III.19: Callose deposits (white spots) in tobacco leaves investigated by fluorescence microscopy at different levels of magnification (II.2.5, II.4.6 and II.4.7, all micrographs in black-white pseudo-colors).

A+C. As in Figure III.18, untreated samples contained few deposits. B+D. In A_+ samples, the number and intensity of callose deposits increased significantly.

As well as the metabolic changes in A_+ samples, there was also a massive increase in callose deposits at 2 dpi (Figure III.18 B) which could be seen at different levels of

magnification (Figure III.19). Staining was concentrated along the leaf venation, perhaps indicating a sieving effect of these tissues on *A. tumefaciens* during injection. Only a few callose deposits were found in the untreated samples (Figure III.18 A and Figure III.19 A and C) or in the mock-treated samples (data not shown). They appeared close to veins, a natural site for callose deposition [223, 224]. These results suggested that the callose deposits in A_+ samples represent a defense response, involving the strengthening of the plant cell wall (I.6). The attachment of bacteria to the plant cells and subsequent T-DNA transfer may therefore have been hampered, reducing the number of transfected cells and thus the final product yield, with an overall negative impact on process economy (I.5).

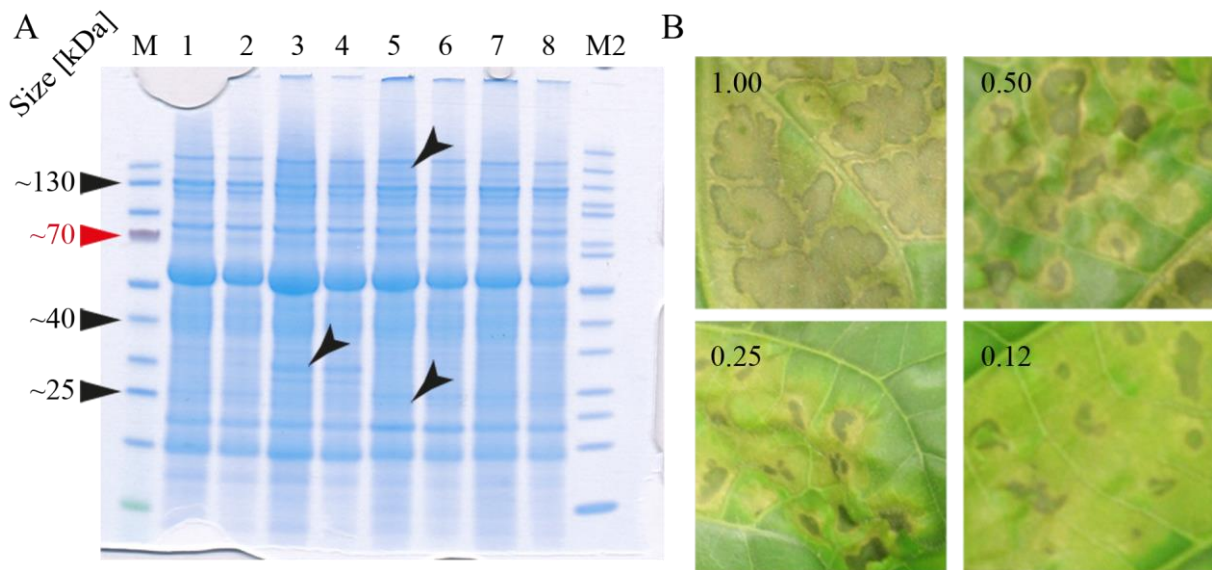


Figure III.20: Influence of *A. tumefaciens* injection and effector expression (II.2.2, II.2.5 and II.4.3).

A. Comparison of TSP in wild type (lanes 1+2), GFD transgenic plants (lane 3+4), A_0 samples (lanes 5+6) and A_+ samples (lanes 7+8). Only marginal differences were observed. The intensity of the 35-kDa band corresponding to DsRed was more intense in samples from transgenic plants (black arrow in lane 3) compared to A_+ . Two uncharacterized bands appeared in the A_0 and A_+ samples (arrows in lane 5). M – pre-stained protein ladder, M2 – unstained protein ladder. B. Deleterious effect of HopF2 on leaf tissue at 5 dpi depending on OD_{600nm} of *A. tumefaciens* during injection. The severity of necrosis declines with decreasing OD_{600nm} .

The difference between A_+ and controls was less striking when the TSP composition of leaf extracts was compared (Figure III.20 A). Two faint bands (~150 kDa and 23 kDa) were found in A_0 extracts (leaves injected with *A. tumefaciens* carrying a control vector) and A_+ extracts (Figure III.20 A lane 5+6 and 7+8 respectively, black arrows) in addition to those observed in untreated wild-type and transgenic controls (note that a ~35 kDa band also appeared in samples from transgenic plants, corresponding to recombinant DsRed). This was anticipated because the resolution of 1D SDS-PAGE is limited and a second dimension, such as isoelectric focusing (IEF), in combination with a more sensitive detection method such as fluorescence labeling, may be required to resolve differences among the samples. Furthermore, it is likely that many of the proteins involved in the defense reaction were

already present in the control samples and were activated by post-translational modification, e.g. phosphorylation in the case of MAPKs [122, 225, 226].

III.3.2.2 Altering reactions by the expression of effectors

The plant defense responses discussed above may attract regulatory scrutiny when biopharmaceutical proteins are produced by transient expression in tobacco, therefore measures to counteract the responses should be considered. The co-expression of candidate TTE genes could be used to suppress the defense responses and potentially increase product yields (I.6). An initial characterization of transient TTE expression in tobacco was carried out by Claudia Haase as part of her Bachelors thesis, which showed that a bacterial OD_{600nm} of 0.5–1.0 caused necrotic lesions in the leaves indicating that a lower OD_{600nm} is required even though the concentration of TTE proteins was below the detection limit of a western blot [149]. The OD_{600nm} range was therefore expanded in this investigation to identify useful parameter values for each effector while avoiding necrosis (Table III.9). The degree of necrosis declined visibly as the OD_{600nm} was reduced (Figure III.20 B). Conditions were identified for each effector that resulted in the absence of macroscopic necrosis (Table III.9) and the results were confirmed by microscopic analysis after Trypan blue staining (data not shown, II.4.5). Those OD_{600nm} values were selected for subsequent co-expression experiments with pGFD, containing the coding sequences for DsRed and 2G12 (III.3.3).

Table III.9: OD_{600nm} ranges for *A. tumefaciens* carrying TTE expression cassettes injected into tobacco leaves.

Effector	Necrotic potential	OD_{600nm} range		Selected for co-expression
		Lower boundary	Upper boundary	
AvrPtoB	++	0.0078	0.25	0.031
AvrRpt2	+++	0.0078	0.25	0.031
HopAO1	+	0.031	0.25	0.125
HopF2	+++	0.0078	0.25	0.031
HopI1	0	0.031	0.25	0.125

At the selected OD_{600nm} values, TTE expression had a significant impact on the metabolic profiles of the treated tobacco leaves. HopAO1 and HopI1 (Figure III.21 A) generated metabolic profiles similar to A_+ (Figure III.20 A) but the concentration of rutin was lower in the TTE samples (Figure III.21 A, arrow 3). The metabolite pattern from AvrPtoB samples resembled A_+ (Figure III.21 B), but AvrRpt2 and HopF2 induced the accumulation of several uncharacterized metabolites (Figure III.21 B, arrows labeled 4) which correlated well with the potential to induce necrosis (Table III.9).

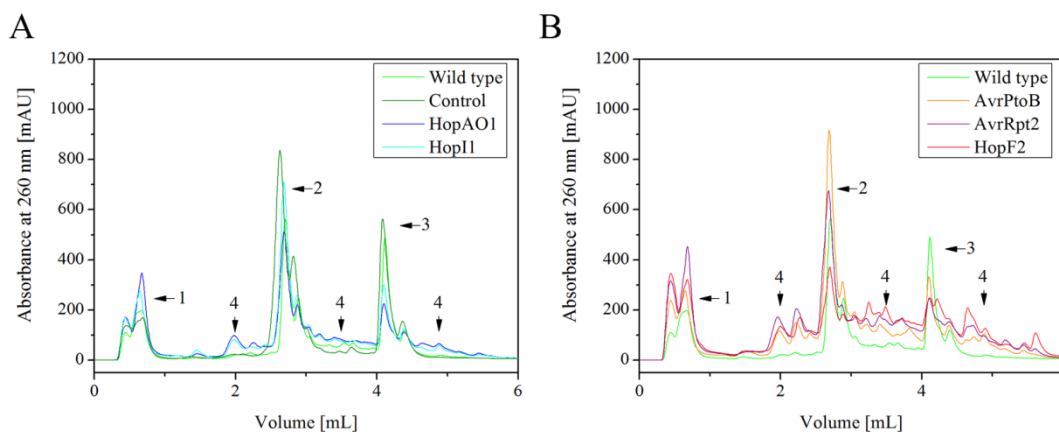


Figure III.21: Impact of TTEs on the metabolic profiles of tobacco leaves after transient expression (II.2.2, II.2.5 and II.4.3).

A. HopAO1 and HopI1 induced the accumulation of nicotine (arrow 1) and other metabolites (arrows 4) similar to pGFD and mock treatments (Figure III.20 A), but suppressed the production of rutin (arrow 3) while chlorogenic acid (arrow 2) was not significantly affected. The untreated control from Figure III.20 is also shown to estimate batch-dependent sample buffer differences. B. AvrPtoB produced a similar profile to HopAO1 and HopI1, whereas HopF2 and AvrRpt2 induced the accumulation of diverse uncharacterized substances (arrows 4).

These data suggested that HopAO1, HopI1 and AvrPtoB are better candidates for co-expression experiments reflecting their inherent lower tendency to cause necrosis, thus avoiding yield losses due to reduced cell viability.

III.3.3 Effects on co-expressed target proteins

The co-expression of TTEs and pGFD using the OD_{600nm} values shown in Table III.9 resulted in lower 2G12 and DsRed levels in most cases (Figure III.22). Only HopAO1, HopI1 and AvrPtoB had a positive impact on the expression of these proteins. The positive impact of HopAO1 was seen at both temperatures whereas that of AvrPtoB was only seen at 22°C, and the positive impact of HopI1 was only seen at 25°C and was restricted to 2G12. Negative effects of the TTEs were generally less severe at 22°C, and 2G12 expression generally suffered to a lesser degree than DsRed. This was attributed primarily to the lower yields of the TTEs at 22°C, which reduced necrosis and yield losses, and also to the fact that 2G12 is secreted and therefore protected from the proteases induced by cell death whereas DsRed in the chloroplasts is more likely to be degraded. This was supported by the finding that TTEs with the greatest necrotic potential (HopF2, AvrPtoB and AvrRpt2) had a more negative impact on target protein concentrations at elevated temperatures (Figure III.22 A and B). In turn this suggested that the OD_{600nm} selected for HopI1 may be too low to induce a beneficial effect at 22°C because such an effect was only observed at 25°C, corresponding to a higher level of HopI1 expression.

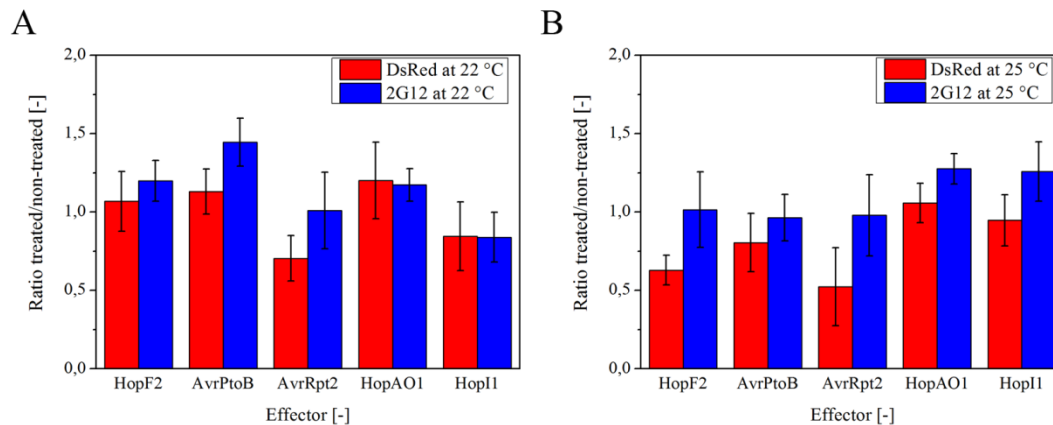


Figure III.22: The impact of TTEs on co-expressed DsRed and 2G12 (II.2.2, II.2.5 and II.4.2).

A. At 22°C, HopF2, HopAO1 and AvrPtoB boosted the accumulation of the target proteins, whereas HopI1 and AvrRpt2 had the opposite effect. B. At 25°C, only HopAO1 had a positive impact on both target proteins, and HopI1 selectively boosted the levels of 2G12.

Additional experiments with different TTE OD_{600nm} values were carried out at 25°C to confirm these findings. Reducing the OD_{600nm} of HopF2, AvrPtoB and HopAO1 had a positive impact by increasing the accumulation or at least limiting the reduction of DsRed (Figure III.23 A). HopAO1 also boosted 2G12 levels (Figure III.23 B). Reducing the OD_{600nm} for HopI1 caused the amount of 2G12 to decline, which is in agreement with the statement above that HopI1 OD_{600nm} may be a limiting factor.

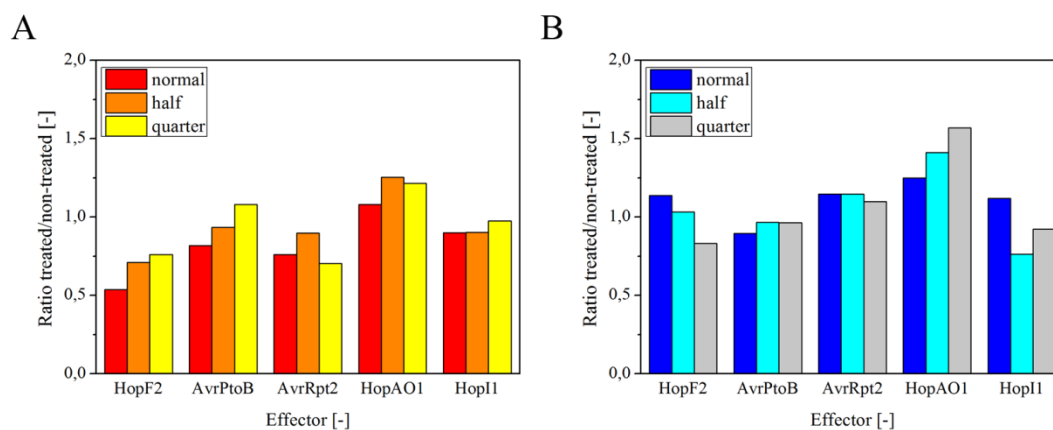


Figure III.23: Impact of varying the OD_{600nm} of TTEs on levels of co-expressed DsRed and 2G12 at 25°C (II.2.2, II.2.5 and II.4.2).

A. Reducing the OD_{600nm} of HopF2, AvrPtoB and HopAO1 had a beneficial impact on DsRed accumulation. B. Reducing the OD_{600nm} of HopAO1 promoted 2G12 accumulation but reducing the OD_{600nm} of HopI1 reduced the 2G12 levels indicating that the OD_{600nm} is a limiting factor for this TTE.

These experiments were continued in the Bachelors thesis of Joschka Buyel [155], showing that despite the positive results presented here, (i) TTEs are not present at the time of bacterial injection because they are synthesized only after T-DNA transfer and are therefore

unable to achieve their full potential in the suppression of plant defense responses, and (ii) transient expression causes the TTEs to accumulate to harmful levels which ultimately offset any beneficial effect on recombinant protein expression. These drawbacks caused the limited beneficial impact of TTEs on the levels of co-expressed proteins observed in this investigation. Therefore, transgenic plants expressing TTE genes under the control of an inducible promoter will be developed to circumvent both of the shortcomings discussed above. The co-transfer of intact effector proteins together with the *A. tumefaciens* T-DNA may also resolve these challenges. In future experiments, co-transfer can be achieved by (i) expressing a fully functional TTSS in *A. tumefaciens* [110], or (ii) by fusing a type-IV secretion signal to the effector genes and thus using the same system for the effectors and T-DNA [227, 228].

IV. Conclusion and scope

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

The impact of genetic elements and cultivation parameters such as temperature on transient protein expression was quantified using the model host *N. tabacum* (III.1 and III.2). The resulting data will facilitate the development of GMP-compliant processes allowing the reproducible production of plant-derived biopharmaceuticals as well as the balanced expression of different target proteins. For example, strict temperature control and defined post-infiltration harvest times were found to be crucial for reproducible expression levels. In future experiments, the dependence of protein expression on (i) the coding sequence of the gene of interest and (ii) the amino acid sequence of the target protein should be investigated in more detail to gain a deeper understanding of factors that determine transient expression levels in plants [229, 230].

The use of *P. syringae* type III effectors was investigated as a strategy to boost transient protein expression by suppressing plant defense responses (III.3). These experiments suggested that if such effectors were delivered along with the T-DNA, using type IV specific recognition sequences, a significant increase in expression could be achieved which would make plant-based production platforms much more competitive [227, 228]. However, more recent data has indicated that DsRed expression levels are not affected by type-III effector proteins co-transferred with T-DNA. Thus, the plant defense responses triggered by the injection of *A. tumefaciens* during transient expression may not be a limiting factor.

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

In addition to the data described in this thesis, the methodology is also particularly important. Previous studies have often focused only on single factors affecting the outcome of experiments, or neglected factor interactions if more than one parameter was studied [70, 71, 231]. 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 [232-234]. The QbD approach seeks to determine process conditions that will result in reproducible product yields, purity and efficacy and thus a safe, high-quality drug in contrast to the quality by control approach in which quality is ensured by extensive testing of the final product [233, 235].

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 [232, 233, 236].

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. To achieve this goal, the models presented here can be combined with models covering different DSP steps that have been developed recently and described in an accompanying PhD project entitled “Manufacturing biopharmaceutical proteins in tobacco” [98]. Such a global 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.

V. Summary

Interest in plant-derived biopharmaceuticals has increased over the last few years, and will continue to do so following the full regulatory approval of the first plant-derived pharmaceutical protein in May 2012. There have been several reports of biopharmaceutical proteins produced in intact plants by *Agrobacterium*-mediated transient expression. However, the variable expression levels in this system make it difficult to develop robust production processes that will achieve regulatory approval. The potential high downstream processing costs have raised questions about the economic viability of plant-based biopharmaceuticals. Both of these critical aspects have been addressed in this thesis. In the first part, parameters were identified that affected transient protein expression in tobacco plants. The quantitative impact of these parameters was determined and modeled using a design of experiments approach. The post-infiltration incubation temperature, plant and leaf age and incubation time were found to be major factors influencing protein yields and variation. Therefore, carefully controlling these parameters in future production processes will significantly reduce batch-to-batch variability and will improve compliance with regulatory guidelines. In the second part the influence of genetic elements such as promoters on the spatiotemporal expression levels was determined and summarized in predictive models. These models indicated that recombinant gene expression is not only dependent on promoters and 5'UTRs but also on the interaction between these elements. Furthermore, the models implied that a rational combination of promoters and 5'UTRs can result in balanced levels of two or more recombinant proteins throughout the duration of transient expression in plants which can improve yields of complex targets such as sIgA. The last section of this thesis focuses on the effect that plant-bacteria interactions have on the yield of transiently expressed target proteins and investigated instruments that can alter these interactions. Additionally, the models developed in this thesis can serve as the basis for a QbD concept in future processes, helping to define, evaluate and control critical process parameters.

VI. References

1. 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.
2. Gagneux, S., *Host-pathogen coevolution in human tuberculosis*. Philosophical Transactions of the Royal Society B-Biological Sciences, 2012. **367**(1590): p. 850-859.
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. 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.
24. Schmidt, F.R., *Recombinant expression systems in the pharmaceutical industry*. Applied microbiology and biotechnology, 2004. **65**(4): p. 363-372.
25. Walsh, G., *Post-translational modifications of protein biopharmaceuticals*. Drug Discovery Today, 2010. **15**(17-18): p. 773-780.
26. 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.
27. 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.
28. 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.
29. Scolnik, P.A., *mAbs A business perspective*. MAbs, 2009. **1**(2): p. 179-184.
30. Reichert, J. and A. Pavlou, *Monoclonal antibodies market*. Nature Reviews Drug Discovery, 2004. **3**(5): p. 383-384.
31. Fischer, R. and N. Emans, *Molecular farming of pharmaceutical proteins*. Transgenic research, 2000. **9**(4-5): p. 279-99; discussion 277.
32. Rodrigues, M.E., et al., *Technological Progresses in Monoclonal Antibody Production Systems*. Biotechnology Progress, 2010. **26**(2): p. 332-351.
33. Caplan, A., et al., *Introduction of genetic material into plant cells*. Science, 1983. **222**(4625): p. 815-21.
34. 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.
35. Hiatt, A., R. Cafferkey, and K. Bowdish, *Production of antibodies in transgenic plants*. Nature, 1989. **342**(6245): p. 76-8.

36. 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.
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. Ma, J.K., et al., *Plant-derived pharmaceuticals--the road forward*. *Trends in Plant Science*, 2005. **10**(12): p. 580-5.
39. Khan, I., et al., *Using storage organelles for the accumulation and encapsulation of recombinant proteins*. *Biotechnology Journal*, 2012.
40. Tremblay, R., et al., *Tobacco, a highly efficient green bioreactor for production of therapeutic proteins*. *Biotechnology Advances*, 2010. **28**(2): p. 214-21.
41. Wilken, L.R. and Z.L. Nikolov, *Recovery and purification of plant-made recombinant proteins*. *Biotechnology Advances*, 2012. **30**(2): p. 419-33.
42. Commandeur, U., R.M. Twyman, and R. Fischer, *The biosafety of molecular farming in plants*. *AgBiotechNet*, 2003. **5**: p. 9.
43. Warzecha, H., *Biopharmaceuticals from plants: a multitude of options for posttranslational modifications*. *Biotechnology & Genetic Engineering Reviews*, 2008. **25**: p. 315-30.
44. 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*, 2006. **103**(40): p. 14701-6.
45. Obembe, O.O., et al., *Advances in plant molecular farming*. *Biotechnology Advances*, 2011. **29**(2): p. 210-222.
46. Ahmad, A., et al., *Green Biofactories: Recombinant Protein Production in Plants*. *Recent patents on biotechnology*, 2010.
47. 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 of the United States of America*, 2006. **103**(20): p. 7577-7582.
48. Fischer, R., et al., *Commercial aspects of pharmaceutical protein production in plants*. *Current Pharmaceutical Design*, 2012. **in Press**.
49. Fischer, R., et al., *GMP issues for recombinant plant-derived pharmaceutical proteins*. *Biotechnology Advances*, 2012. **30**(2): p. 434-9.
50. Kelley, B., *Industrialization of mAb production technology: the bioprocessing industry at a crossroads*. *MAbs*, 2009. **1**(5): p. 443-52.
51. Peremarti, A., et al., *Promoter diversity in multigene transformation*. *Plant Molecular Biology*, 2010. **73**(4-5): p. 363-378.
52. Mehrotra, R., et al., *Designer promoter: an artwork of cis engineering*. *Plant Molecular Biology*, 2011. **75**(6): p. 527-536.
53. Bhullar, S., et al., *Strategies for development of functionally equivalent promoters with minimum sequence homology for transgene expression in plants: cis-elements in a novel DNA context versus domain swapping*. *Plant Physiology*, 2003. **132**(2): p. 988-998.

54. Bhattacharyya, S., N. Dey, and I.B. Maiti, *Analysis of cis-sequence of subgenomic transcript promoter from the Figwort mosaic virus and comparison of promoter activity with the cauliflower mosaic virus promoters in monocot and dicot cells*. Virus Research, 2002. **90**(1-2): p. 47-62.
55. Cazzonelli, C.I., et al., *Characterization of a strong, constitutive mung bean (*Vigna radiata* L.) promoter with a complex mode of regulation in planta*. Transgenic Research, 2005. **14**(6): p. 941-967.
56. Mitsuhashi, I., et al., *Efficient promoter cassettes for enhanced expression of foreign genes in dicotyledonous and monocotyledonous plants*. Plant and Cell Physiology, 1996. **37**(1): p. 49-59.
57. Horstmann, V., et al., *Quantitative promoter analysis in *Physcomitrella patens*: a set of plant vectors activating gene expression within three orders of magnitude*. BMC biotechnology, 2004. **4**: p. 13.
58. Chung, B.Y., et al., *Effect of 5'UTR introns on gene expression in *Arabidopsis thaliana**. BMC Genomics, 2006. **7**: p. 120.
59. Lu, J.L., et al., *Activity of the 5' regulatory regions of the rice polyubiquitin *rub1* gene in transgenic rice plants as analyzed by both GUS and GFP reporter genes*. Plant Cell Reports, 2008. **27**(10): p. 1587-1600.
60. Landrain, T.E., et al., *Modular model-based design for heterologous bioproduction in bacteria*. Current Opinion in Biotechnology, 2009. **20**(3): p. 272-9.
61. Gertz, J., E.D. Siggia, and B.A. Cohen, *Analysis of combinatorial cis-regulation in synthetic and genomic promoters*. Nature, 2009. **457**(7226): p. 215-8.
62. Gertz, J. and B.A. Cohen, *Environment-specific combinatorial cis-regulation in synthetic promoters*. Molecular Systems Biology, 2009. **5**: p. 244.
63. Segal, E., et al., *Predicting expression patterns from regulatory sequence in *Drosophila* segmentation*. Nature, 2008. **451**(7178): p. 535-40.
64. Zinzen, R.P., et al., *Computational models for neurogenic gene expression in the *Drosophila* embryo*. Current Biology, 2006. **16**(13): p. 1358-65.
65. Joh, L.D., et al., *High-level transient expression of recombinant protein in lettuce*. Biotechnology and bioengineering, 2005. **91**(7): p. 861-71.
66. 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.
67. 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.
68. Wroblewski, T., A. Tomczak, and R. Michelmore, *Optimization of *Agrobacterium*-mediated transient assays of gene expression in lettuce, tomato and *Arabidopsis**. Plant biotechnology journal, 2005. **3**(2): p. 259-73.
69. Joersbo, M., J.D. Mikkelsen, and J. Brunstedt, *Relationship between promoter strength and transformation frequencies using mannose selection for the production of transgenic sugar beet*. Molecular Breeding, 2000. **6**(2): p. 207-213.
70. Furtado, A., R.J. Henry, and F. Takaiwa, *Comparison of promoters in transgenic rice*. Plant biotechnology journal, 2008. **6**(7): p. 679-93.

71. 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.
72. Rathus, C., R. Bower, and R.G. Birch, *Effects of promoter, intron and enhancer elements on transient gene expression in sugar-cane and carrot protoplasts*. Plant Molecular Biology, 1993. **23**(3): p. 613-8.
73. Fuentes, A., et al., *Development of a highly efficient system for assessing recombinant gene expression in plant cell suspensions via Agrobacterium tumefaciens transformation*. Biotechnology and applied biochemistry, 2004. **39**(Pt 3): p. 355-61.
74. Kavita, P. and P.K. Burma, *A comparative analysis of green fluorescent protein and beta-glucuronidase protein-encoding genes as a reporter system for studying the temporal expression profiles of promoters*. Journal of Biosciences, 2008. **33**(3): p. 337-343.
75. Qu le, Q. and F. Takaiwa, *Evaluation of tissue specificity and expression strength of rice seed component gene promoters in transgenic rice*. Plant Biotechnology Journal, 2004. **2**(2): p. 113-25.
76. Cal, M., et al., *A rice promoter containing both novel positive and negative cis-elements for regulation of green tissue-specific gene expression in transgenic plants*. Plant Biotechnology Journal, 2007. **5**(5): p. 664-674.
77. Furtado, A., R.J. Henry, and A. Pellegrineschi, *Analysis of promoters in transgenic barley and wheat*. Plant Biotechnology Journal, 2009. **7**(3): p. 240-253.
78. Rasala, B.A., et al., *Improved heterologous protein expression in the chloroplast of Chlamydomonas reinhardtii through promoter and 5' untranslated region optimization*. Plant Biotechnology Journal, 2011. **9**(6): p. 674-683.
79. 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.
80. Gelvin, S.B., *Improving plant genetic engineering by manipulating the host*. Trends in Biotechnology, 2003. **21**(3): p. 95-8.
81. 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.
82. Miki, B. and S. McHugh, *Selectable marker genes in transgenic plants: applications, alternatives and biosafety*. Journal of Biotechnology, 2004. **107**(3): p. 193-232.
83. 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.
84. Pawlowski, W.P. and D.A. Somers, *Transgene inheritance in plants genetically engineered by microprojectile bombardment*. Molecular Biotechnology, 1996. **6**(1): p. 17-30.
85. Fischer, R., et al., *Towards molecular farming in the future: transient protein expression in plants*. Biotechnology and Applied Biochemistry, 1999. **30**: p. 113-116.
86. 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.

87. 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.
88. Tzfira, T. and V. Citovsky, *Agrobacterium-mediated genetic transformation of plants: biology and biotechnology*. Current opinion in biotechnology, 2006. **17**(2): p. 147-54.
89. 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.
90. 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.
91. 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.
92. 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.
93. 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.
94. 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.
95. Larsen, J.S., Curtis, W. R., *Reducing Batch-to-Batch Variability of Agrobacterium-Mediated Transient Protein Expression in Plant Tissue Culture*, in *The 2008 Annual Meeting 2008*: Philadelphia, PA.
96. O'Neill, K.M., J.S. Larsen, and W.R. Curtis, *Scale-up of Agrobacterium-mediated transient protein expression in bioreactor-grown Nicotiana glutinosa plant cell suspension culture*. Biotechnology Progress, 2008. **24**(2): p. 372-6.
97. Lin, J.J., *Electrotransformation of Agrobacterium*. Methods in Molecular Biology, 1995. **47**: p. 171-8.
98. Buyel, J.F., *Manufacturing biopharmaceutical proteins in tobacco*, RWTH Aachen University: Aachen.
99. Lindbo, J.A., *High-efficiency protein expression in plants from agroinfection-compatible Tobacco mosaic virus expression vectors*. BMC Biotechnology, 2007. **7**: p. 52.
100. Gleba, Y., V. Klimyuk, and S. Marillonnet, *Magniffection--a new platform for expressing recombinant vaccines in plants*. Vaccine, 2005. **23**(17-18): p. 2042-8.
101. Johansen, L.K. and J.C. Carrington, *Silencing on the spot. Induction and suppression of RNA silencing in the Agrobacterium-mediated transient expression system*. Plant physiology, 2001. **126**(3): p. 930-8.
102. Block, A. and J.R. Alfano, *Plant targets for Pseudomonas syringae type III effectors: virulence targets or guarded decoys?* Current Opinion in Microbiology, 2011. **14**(1): p. 39-46.

103. Tsuda, K., et al., *An efficient Agrobacterium-mediated transient transformation of Arabidopsis*. Plant Journal, 2012. **69**(4): p. 713-719.
104. Buttner, D. and U. Bonas, *Port of entry - the type III secretion translocon*. Trends in Microbiology, 2002. **10**(4): p. 186-192.
105. Roine, E., et al., *Hrp pilus: An hrp-dependent bacterial surface appendage produced by Pseudomonas syringae pv tomato DC3000*. Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(7): p. 3459-3464.
106. Galan, J.E. and H. Wolf-Watz, *Protein delivery into eukaryotic cells by type III secretion machines*. Nature, 2006. **444**(7119): p. 567-573.
107. Kim, M.G., et al., *Two Pseudomonas syringae type III effectors inhibit RIM-regulated basal defense in Arabidopsis*. Cell, 2005. **121**(5): p. 749-759.
108. Jelenska, J., et al., *A J domain virulence effector of Pseudomonas syringae remodels host chloroplasts and suppresses defenses*. Current Biology, 2007. **17**(6): p. 499-508.
109. Shan, L., P. He, and J. Sheen, *Intercepting host MAPK signaling cascades by bacterial type III effectors*. Cell Host & Microbe, 2007. **1**(3): p. 167-174.
110. Guo, M., et al., *The majority of the type III effector inventory of Pseudomonas syringae pv. tomato DC3000 can suppress plant immunity*. Molecular Plant-Microbe Interactions, 2009. **22**(9): p. 1069-80.
111. Akira, S., S. Uematsu, and O. Takeuchi, *Pathogen recognition and innate immunity*. Cell, 2006. **124**(4): p. 783-801.
112. Luna, E., et al., *Callose Deposition: A Multifaceted Plant Defense Response*. Molecular Plant-Microbe Interactions, 2011. **24**(2): p. 183-193.
113. An, C.F. and Z.L. Mou, *Salicylic Acid and its Function in Plant Immunity*. Journal of Integrative Plant Biology, 2011. **53**(6): p. 412-428.
114. Jones, J.D.G. and J.L. Dangl, *The plant immune system*. Nature, 2006. **444**(7117): p. 323-329.
115. Abramovitch, R.B. and G.B. Martin, *AvrPtoB: A bacterial type III effector that both elicits and suppresses programmed cell death associated with plant immunity*. Fems Microbiology Letters, 2005. **245**(1): p. 1-8.
116. Thomma, B.P.H.J., T. Nurnberger, and M.H.A.J. Joosten, *Of PAMPs and Effectors: The Blurred PTI-ETI Dichotomy*. Plant Cell, 2011. **23**(1): p. 4-15.
117. Chen, Z.Y., et al., *Pseudomonas syringae type III effector AvrRpt2 alters Arabidopsis thaliana auxin physiology*. Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(50): p. 20131-20136.
118. Wilton, M., et al., *The type III effector HopF2(Pto) targets Arabidopsis RIN4 protein to promote Pseudomonas syringae virulence*. Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(5): p. 2349-2354.
119. Wang, Y.J., et al., *A Pseudomonas syringae ADP-Ribosyltransferase Inhibits Arabidopsis Mitogen-Activated Protein Kinase Kinases*. Plant Cell, 2010. **22**(6): p. 2033-2044.
120. Underwood, W., S.Q. Zhang, and S.Y. He, *The Pseudomonas syringae type III effector tyrosine phosphatase HopAO1 suppresses innate immunity in Arabidopsis thaliana*. Plant Journal, 2007. **52**(4): p. 658-672.

121. Pruss, G.J., E.W. Nester, and V. Vance, *Infiltration with Agrobacterium tumefaciens Induces Host Defense and Development-Dependent Responses in the Infiltrated Zone*. *Molecular Plant-Microbe Interactions*, 2008. **21**(12): p. 1528-1538.
122. Djamei, A., et al., *Trojan horse strategy in Agrobacterium transformation: Abusing MAPK defense signaling*. *Science*, 2007. **318**(5849): p. 453-456.
123. Anand, A., et al., *Salicylic acid and systemic acquired resistance play a role in attenuating crown gall disease caused by Agrobacterium tumefaciens*. *Plant Physiology*, 2008. **146**(2): p. 703-715.
124. Ludwig, A.A., et al., *Ethylene-mediated cross-talk between calcium-dependent protein kinase and MAPK signaling controls stress responses in plants*. *Proceedings of the National Academy of Sciences of the United States of America*, 2005. **102**(30): p. 10736-10741.
125. Menkhaus, T.J., et al., *Considerations for the recovery of recombinant proteins from plants*. *Biotechnology Progress*, 2004. **20**(4): p. 1001-14.
126. An, G., et al., *Organ-Specific and Developmental Regulation of the Nopaline Synthase Promoter in Transgenic Tobacco Plants*. *Plant Physiology*, 1988. **88**(3): p. 547-552.
127. An, G., M.A. Costa, and S.B. Ha, *Nopaline synthase promoter is wound inducible and auxin inducible*. *Plant Cell*, 1990. **2**(3): p. 225-33.
128. Atkinson, A.C., B. Bogacka, and A.A. Zhigljavski, *Optimum design 2000*. 2001: Kluwer Academic Publishers.
129. Montgomery, D.C., *Design and Analysis of Experiments, Minitab Manual*. 2008: Wiley.
130. Mandenius, C.F. and A. Brundin, *Bioprocess optimization using design-of-experiments methodology*. *Biotechnology Progress*, 2008. **24**(6): p. 1191-203.
131. Abu-Absi, S.F., et al., *Defining process design space for monoclonal antibody cell culture*. *Biotechnology and Bioengineering*, 2010. **106**(6): p. 894-905.
132. Kleppmann, W., *Taschenbuch Versuchsplanung*. 6th edition ed. *Produkte und Prozesse optimieren*, ed. F.J. Brunner. 2009, München: Carl Hanser Verlag.
133. Myers, R.H., et al., *Response surface methodology: A retrospective and literature survey*. *Journal of Quality Technology*, 2004. **36**(1): p. 53-77.
134. Eriksson, L., *Design of Experiments: Principles and Applications*. 2008: Umetrics Academy.
135. Goos, P. and B. Jones, *Optimal Design of Experiments: A Case Study Approach*. 2011: John Wiley & Sons.
136. 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.
137. Brook, R.J. and G.C. Arnold, *Applied Regression Analysis and Experimental Design*. 1985: M. Dekker.
138. Weisberg, S., *Applied Linear Regression*. 2005: Wiley-Interscience.
139. Myers, R.H., *Classical and modern regression with applications*. 1990: PWS-KENT.

140. 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.
141. 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.
142. Inoue, H., H. Nojima, and H. Okayama, *High-Efficiency Transformation of Escherichia-Coli with Plasmids*. Gene, 1990. **96**(1): p. 23-28.
143. Main, G.D., S. Reynolds, and J.S. Gartland, *Electroporation protocols for Agrobacterium*. Methods in Molecular Biology, 1995. **44**: p. 405-12.
144. 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.
145. Gruber, A.R., et al., *The Vienna RNA websuite*. Nucleic acids research, 2008. **36**(Web Server issue): p. W70-4.
146. Fang, R.X., et al., *Multiple Cis Regulatory Elements for Maximal Expression of the Cauliflower Mosaic-Virus S-35 Promoter in Transgenic Plants*. Plant Cell, 1989. **1**(1): p. 141-150.
147. Bevan, M., W.M. Barnes, and M.D. Chilton, *Structure and transcription of the nopaline synthase gene region of T-DNA*. Nucleic acids research, 1983. **11**(2): p. 369-85.
148. Kaefer, T., *Characterization of transient protein expression patterns in Nicotiana tabacum treated with recombinant Agrobacterium tumefaciens*, in *Molecular Biotechnology*2011, RWTH Aachen University: Aachen.
149. Haase, C., *Cloning and Expression of Pseudomonas syringae Type III Effectors in Nicotiana tabacum*, in *Molecular Biotechnology*2010, RWTH Aachen University: Aachen.
150. 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*. Analytical Biochemistry, 1976. **72**(10): p. 248-54.
151. Simonian, M.H. and J.A. Smith, *Spectrophotometric and colorimetric determination of protein concentration*. Current Protocols in Molecular Biology, 2006. **Chapter 10**: p. Unit 10 1A.
152. 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.
153. Piliarik, M., H. Vaisocherova, and J. Homola, *Surface plasmon resonance biosensing*. Methods in molecular biology, 2009. **503**: p. 65-88.
154. Vierheilig, H., et al., *Imaging arbuscular mycorrhizal structures in living roots of Nicotiana tabacum by light, epifluorescence, and confocal laser scanning microscopy*. Canadian Journal of Botany-Revue Canadienne De Botanique, 2001. **79**(2): p. 231-237.

155. Buyel, J.J., *Optimization of Pseudomonas syringae type III effector expression in Nicotiana tabacum for use in transient protein expression systems*, in *Molecular Biotechnology* 2011, RWTH Aachen University: Aachen.
156. Adam, L. and S.C. Somerville, *Genetic characterization of five powdery mildew disease resistance loci in Arabidopsis thaliana*. *Plant Journal*, 1996. **9**(3): p. 341-356.
157. Keinänen, M., N.J. Oldham, and I.T. Baldwin, *Rapid HPLC screening of jasmonate-induced increases in tobacco alkaloids, phenolics, and diterpene glycosides in Nicotiana attenuata*. *Journal of Agricultural and Food Chemistry*, 2001. **49**(8): p. 3553-3558.
158. Peixoto, J.L., *A Property of Well-Formulated Polynomial Regression-Models*. *American Statistician*, 1990. **44**(1): p. 26-30.
159. Peixoto, J.L., *Hierarchical Variable Selection in Polynomial Regression-Models*. *American Statistician*, 1987. **41**(4): p. 311-313.
160. Bindschedler, L.V. and R. Cramer, *Quantitative plant proteomics*. *Proteomics*, 2011. **11**(4): p. 756-775.
161. Sparrow, P.A.C., et al., *Pharma-Planta: Road testing the developing regulatory guidelines for plant-made pharmaceuticals*. *Transgenic Research*, 2007. **16**(2): p. 147-161.
162. Colgan, R., et al., *Optimisation of contained Nicotiana tabacum cultivation for the production of recombinant protein pharmaceuticals*. *Transgenic research*, 2010. **19**(2): p. 241-56.
163. Dillen, W., et al., *The effect of temperature on Agrobacterium tumefaciens-mediated gene transfer to plants*. *The Plant Journal*, 1997. **12**(6): p. 1459-1463.
164. Galbraith, D.W., et al., *Rapid flow cytometric analysis of the cell cycle in intact plant tissues*. *Science*, 1983. **220**(4601): p. 1049-51.
165. Poethig, R.S. and I.M. Sussex, *The developmental morphology and growth dynamics of the tobacco leaf*. *Planta*, 1985. **165**(2): p. 158-169.
166. Simmons, C.W., J.S. VanderGheynst, and S.K. Upadhyaya, *A model of Agrobacterium tumefaciens vacuum infiltration into harvested leaf tissue and subsequent in planta transgene transient expression*. *Biotechnology and Bioengineering*, 2009. **102**(3): p. 965-70.
167. 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
168. An, G.H., *Development of Plant Promoter Expression Vectors and Their Use for Analysis of Differential Activity of Nopaline Synthase Promoter in Transformed Tobacco Cells*. *Plant Physiology*, 1986. **81**(1): p. 86-91.
169. Kreuzaler, F., et al., *UV-induction of chalcone synthase mRNA in cell suspension cultures of Petroselinum hortense*. *Proceedings of the National Academy of Sciences*, 1983. **80**(9): p. 2591-3.
170. Gallie, D.R., et al., *The 5'-Leader Sequence of Tobacco Mosaic-Virus Rna Enhances the Expression of Foreign Gene Transcripts Invitro and Invivo*. *Nucleic Acids Research*, 1987. **15**(8): p. 3257-3273.

171. Carrington, J.C. and D.D. Freed, *Cap-Independent Enhancement of Translation by a Plant Potyvirus 5' Nontranslated Region*. Journal of Virology, 1990. **64**(4): p. 1590-1597.
172. Fischer, R., et al., *Expression and characterization of bispecific single-chain Fv fragments produced in transgenic plants*. European Journal of Biochemistry, 1999. **262**(3): p. 810-816.
173. Munro, S. and H.R. Pelham, *A C-terminal signal prevents secretion of luminal ER proteins*. Cell, 1987. **48**(5): p. 899-907.
174. Filipecki, M. and S. Malepszy, *Unintended consequences of plant transformation: a molecular insight*. Journal of Applied Genetics, 2006. **47**(4): p. 277-86.
175. Bhullar, S., et al., *Analysis of promoter activity in transgenic plants by normalizing expression with a reference gene: anomalies due to the influence of the test promoter on the reference promoter*. Journal of Biosciences, 2009. **34**(6): p. 953-962.
176. Sanders, P.R., et al., *Comparison of cauliflower mosaic virus 35S and nopaline synthase promoters in transgenic plants*. Nucleic acids research, 1987. **15**(4): p. 1543-58.
177. Dowson Day, M.J., et al., *Plant viral leaders influence expression of a reporter gene in tobacco*. Plant molecular biology, 1993. **23**(1): p. 97-109.
178. Kochetov, A.V., et al., *Eukaryotic mRNAs encoding abundant and scarce proteins are statistically dissimilar in many structural features*. Febs Letters, 1998. **440**(3): p. 351-355.
179. Dey, N. and I.B. Maiti, *Structure and promoter/leader deletion analysis of mirabilis mosaic virus (MMV) full-length transcript promoter in transgenic plants*. Plant Molecular Biology, 1999. **40**(5): p. 771-782.
180. Mignone, F., et al., *Untranslated regions of mRNAs*. Genome biology, 2002. **3**(3): p. REVIEWS0004.
181. Ringner, M. and M. Krogh, *Folding free energies of 5'-UTRs impact post-transcriptional regulation on a genomic scale in yeast*. Plos Computational Biology, 2005. **1**(7): p. e72.
182. Livingstone, M., et al., *Mechanisms governing the control of mRNA translation*. Physical Biology, 2010. **7**(2): p. 021001.
183. Zeenko, V. and D.R. Gallie, *Cap-independent translation of tobacco etch virus is conferred by an RNA pseudoknot in the 5'-leader*. Journal of Biological Chemistry, 2005. **280**(29): p. 26813-24.
184. Hellen, C.U. and P. Sarnow, *Internal ribosome entry sites in eukaryotic mRNA molecules*. Genes & Development, 2001. **15**(13): p. 1593-612.
185. Jackson, R.J., C.U.T. Hellen, and T.V. Pestova, *The mechanism of eukaryotic translation initiation and principles of its regulation*. Nature Reviews Molecular Cell Biology, 2010. **11**(2): p. 113-127.
186. Wilkie, G.S., K.S. Dickson, and N.K. Gray, *Regulation of mRNA translation by 5'- and 3'-UTR-binding factors*. Trends in Biochemical Sciences, 2003. **28**(4): p. 182-8.
187. Keene, J.D. and S.A. Tenenbaum, *Eukaryotic mRNPs may represent posttranscriptional operons*. Molecular Cell, 2002. **9**(6): p. 1161-1167.

188. Muench, D.G., C. Zhang, and M. Dahodwala, *Control of cytoplasmic translation in plants*. Wiley Interdisciplinary Reviews-Rna, 2012. **3**(2): p. 178-194.
189. Gelvin, S.B., *Agrobacterium and plant genes involved in T-DNA transfer and integration*. Annual Review of Plant Physiology and Plant Molecular Biology, 2000. **51**: p. 223-256.
190. Virts, E.L. and S.B. Gelvin, *Analysis of transfer of tumor-inducing plasmids from Agrobacterium tumefaciens to Petunia protoplasts*. Journal of bacteriology, 1985. **162**(3): p. 1030-8.
191. Sykes, L.C. and A.G. Matthysse, *Time required for tumor induction by Agrobacterium tumefaciens*. Applied and environmental microbiology, 1986. **52**(3): p. 597-8.
192. Voinnet, O., Y.M. Pinto, and D.C. Baulcombe, *Suppression of gene silencing: A general strategy used by diverse DNA and RNA viruses of plants*. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(24): p. 14147-14152.
193. Ma, P.D., et al., *A Viral Suppressor P1/HC-Pro Increases the GFP Gene Expression in Agrobacterium-mediated Transient Assay*. Applied Biochemistry and Biotechnology, 2009. **158**(2): p. 243-252.
194. Saxena, P., et al., *Improved foreign gene expression in plants using a virus-encoded suppressor of RNA silencing modified to be developmentally harmless*. Plant Biotechnology Journal, 2011. **9**(6): p. 703-712.
195. Scholthof, H.B., K.B.G. Scholthof, and A.O. Jackson, *Identification of Tomato Bushy Stunt Virus Host-Specific Symptom Determinants by Expression of Individual Genes from a Potato-Virus-X Vector*. Plant Cell, 1995. **7**(8): p. 1157-1172.
196. Angel, C.A., Y.C. Hsieh, and J.E. Schoelz, *Comparative Analysis of the Capacity of Tombusvirus P22 and P19 Proteins to Function as Avirulence Determinants in Nicotiana species*. Molecular Plant-Microbe Interactions, 2011. **24**(1): p. 91-99.
197. Arzola, L., et al., *Transient Co-Expression of Post-Transcriptional Gene Silencing Suppressors for Increased in Planta Expression of a Recombinant Anthrax Receptor Fusion Protein*. International Journal of Molecular Sciences, 2011. **12**(8): p. 4975-4990.
198. Lackner, D.H., et al., *A network of multiple regulatory layers shapes gene expression in fission yeast*. Molecular Cell, 2007. **26**(1): p. 145-55.
199. Pradet-Balade, B., et al., *Translation control: bridging the gap between genomics and proteomics?* Trends in Biochemical Sciences, 2001. **26**(4): p. 225-9.
200. 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.
201. Abreu, R.D., et al., *Global signatures of protein and mRNA expression levels*. Molecular Biosystems, 2009. **5**(12): p. 1512-1526.
202. Goodarzi, H., et al., *Systematic discovery of structural elements governing stability of mammalian messenger RNAs*. Nature, 2012. **485**(7397): p. 264-8.
203. Huang, H.Y., et al., *RegRNA: an integrated web server for identifying regulatory RNA motifs and elements*. Nucleic Acids Research, 2006. **34**: p. W429-W434.

204. Shyu, A.B., M.F. Wilkinson, and A. van Hoof, *Messenger RNA regulation: to translate or to degrade*. *Embo Journal*, 2008. **27**(3): p. 471-481.
205. Flanagan, J.J., et al., *Signal recognition particle binds to ribosome-bound signal sequences with fluorescence-detected subnanomolar affinity that does not diminish as the nascent chain lengthens*. *Journal of biological chemistry*, 2003. **278**(20): p. 18628-37.
206. Denecke, J., et al., *The Tobacco Luminal Binding-Protein Is Encoded by a Multigene Family*. *Plant Cell*, 1991. **3**(9): p. 1025-1035.
207. Kalinski, A., et al., *Binding-Protein Expression Is Subject to Temporal, Developmental and Stress-Induced Regulation in Terminally Differentiated Soybean Organs*. *Planta*, 1995. **195**(4): p. 611-621.
208. Idiris, A., et al., *Engineering of protein secretion in yeast: strategies and impact on protein production*. *Applied microbiology and biotechnology*, 2010. **86**(2): p. 403-417.
209. Richter, L.J., et al., *Production of hepatitis B surface antigen in transgenic plants for oral immunization*. *Nature biotechnology*, 2000. **18**(11): p. 1167-71.
210. Conrad, U. and U. Fiedler, *Compartment-specific accumulation of recombinant immunoglobulins in plant cells: an essential tool for antibody production and immunomodulation of physiological functions and pathogen activity*. *Plant molecular biology*, 1998. **38**(1-2): p. 101-9.
211. Ruth, C. and A. Glieder, *Perspectives on Synthetic Promoters for Biocatalysis and Biotransformation*. *Chembiochem*, 2010. **11**(6): p. 761-765.
212. Carothers, J.M., J.A. Goler, and J.D. Keasling, *Chemical synthesis using synthetic biology*. *Current Opinion in Biotechnology*, 2009. **20**(4): p. 498-503.
213. Mukherji, S. and A. van Oudenaarden, *Synthetic biology: understanding biological design from synthetic circuits*. *Nature Reviews Genetics*, 2009. **10**(12): p. 859-871.
214. Mijakovic, I., D. Petranovic, and P.R. Jensen, *Tunable promoters in systems biology*. *Current Opinion in Biotechnology*, 2005. **16**(3): p. 329-335.
215. Ma, J.K.C., et al., *Generation and Assembly of Secretory Antibodies in Plants*. *Science*, 1995. **268**(5211): p. 716-719.
216. Wycoff, K.L., *Secretory IgA antibodies from plants*. *Current Pharmaceutical Design*, 2005. **11**(19): p. 2429-2437.
217. Fischer, M., et al., *Evolution of vitamin B2 biosynthesis: structural and functional similarity between pyrimidine deaminases of eubacterial and plant origin*. *Journal of Biological Chemistry*, 2004. **279**(35): p. 36299-308.
218. Sivamani, E. and R. Qu, *Expression enhancement of a rice polyubiquitin gene promoter*. *Plant Molecular Biology*, 2006. **60**(2): p. 225-239.
219. Kroner, A., et al., *Nicotiflorin, rutin and chlorogenic acid: phenylpropanoids involved differently in quantitative resistance of potato tubers to biotrophic and necrotrophic pathogens*. *Plant Physiology and Biochemistry*, 2012. **57**: p. 23-31.
220. Meyer, M.R. and H.H. Maurer, *Absorption, distribution, metabolism and excretion pharmacogenomics of drugs of abuse*. *Pharmacogenomics*, 2011. **12**(2): p. 215-233.

221. Toledano, A., M.I. Alvarez, and A. Toledano-Diaz, *Diversity and variability of the effects of nicotine on different cortical regions of the brain - therapeutic and toxicological implications*. Central Nervous System Agents in Medicinal Chemistry, 2010. **10**(3): p. 180-206.
222. Schep, L.J., R.J. Slaughter, and D.M. Beasley, *Nicotinic plant poisoning*. Clinical Toxicology, 2009. **47**(8): p. 771-81.
223. McNairn, R.B. and H.B. Currier, *Sieve Plate Callose . A Factor in Blockage of Axial Phloem Transport*. Naturwissenschaften, 1967. **54**(22): p. 591-&.
224. Chen, X.Y. and J.Y. Kim, *Callose synthesis in higher plants*. Plant Signaling and Behavior, 2009. **4**(6): p. 489-92.
225. Mishra, N.S., R. Tuteja, and N. Tuteja, *Signaling through MAP kinase networks in plants*. Archives of Biochemistry and Biophysics, 2006. **452**(1): p. 55-68.
226. Zhang, S. and D.F. Klessig, *Pathogen-induced MAP kinases in tobacco*. Results and problems in cell differentiation, 2000. **27**: p. 65-84.
227. Vergunst, A.C., et al., *Recognition of the Agrobacterium tumefaciens VirE2 translocation signal by the VirB/D4 transport system does not require VirE1*. Plant Physiology, 2003. **133**(3): p. 978-988.
228. Vergunst, A.C., et al., *Positive charge is an important feature of the C-terminal transport signal of the VirB/D4-translocated proteins of Agrobacterium*. Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(3): p. 832-837.
229. Gustafsson, C., et al., *Engineering genes for predictable protein expression*. Protein Expression and Purification, 2012. **83**(1): p. 37-46.
230. Welch, M., et al., *Designing Genes for Successful Protein Expression*. Synthetic Biology, Pt B, 2011. **498**: p. 43-66.
231. 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.
232. 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.
233. Rathore, A.S., *Roadmap for implementation of quality by design (QbD) for biotechnology products*. Trends in Biotechnology, 2009. **27**(9): p. 546-553.
234. 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.
235. 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.
236. 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.

VII. Appendix

VII.1 List of publications

1. Buyel JF, Fischer R. 2012. Predictive models for transient protein expression in tobacco (*Nicotiana tabacum* L.) can optimize process time, yield, and downstream costs. *Biotechnology and Bioengineering* 109(10):2575-88.
2. Buyel JF, Fischer R. 2013. Processing heterogeneous biomass: Overcoming the hurdles in model building. *Bioengineered* 4(1). (Epub ahead of print).
3. Buyel JF, Kaever T, Buyel JJ, Fischer R. 2013. Predictive models for the accumulation of a fluorescent marker protein in tobacco leaves according to the promoter/5'UTR combination. *Biotechnology and Bioengineering* 110(2):471-82.
4. Fischer R, Schillberg S, Buyel JF, Twyman RM. 2013. Commercial aspects of pharmaceutical protein production in plants. *Current Pharmaceutical Design* (in press).

VII.2 List of publications in preparation

1. Buyel JF, Buyel JJ, Haase C, Fischer R. Exploiting the potential of type III effectors from *Pseudomonas syringae* to enhance transient protein expression in plants, in preparation.

VII.3 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
ÅKTA explorer	chromatography device	GE Healthcare, UK
Allegra 25R	centrifuge	BeckmanCoulter, CA, 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 Blener	blender	Warring, CT, USA
Cond 315i	conductometer	WTW, Germany
Forma -86C ULT freezer	-80°C freezer	ThermoFisher, MA, USA
Innova 4230	incubator/shaker	New Brunswick Scientific, CT, USA
KMO 2 basic	stirrer	IKA, Germany
LEICA DRM+DCF	microscope	Leica, Germany
M-Power	scale	Sartorius, Germany
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
OPERA	automated microscope	PerkinElmer, MA, USA
pH 340i	pH meter	WTW, Germany
PowerPac300	DC source	BioRad, CA, USA
PowerPacBasic	DC source	BioRad, CA, USA
Premium	refrigerator	Liebherr, Germany
PVDF membrane	blotting membrane	Millipore, MA, USA
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

VII.4 List of chemicals

Name	Type/Use	Manufacturer
2,4-Dipyridyl	FPLC standard.....	Sigma-Aldrich, MO, USA
4-Aminopyridine	FPLC standard.....	Sigma-Aldrich, MO, USA
Acetosyringone.....	phytohormone	Duchefa, The Netherlands
Aniline-blue	dye.....	Merck, Germany
Ampicillin.....	antibiotic.....	Carl Roth GmbH, Germany
Caffeic acid	FPLC standard.....	Sigma-Aldrich, MO, USA
Carbenicilin	antibiotic.....	Duchefa, The Netherlands
Chlorogenic acid.....	FPLC standard.....	Sigma-Aldrich, MO, USA
Cinnamic acid.....	FPLC standard.....	Sigma-Aldrich, MO, USA
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
Methanol.....	solution component	Carl Roth GmbH, Germany A
Murashige & Skoog salts.....	solution component	Duchefa, The Netherlands
Nicotine	FPLC standard.....	Sigma-Aldrich, MO, USA
Potassium chloride	buffer component	Carl Roth GmbH, Germany
Potassium dihydrogenphosphate ...buffer component		Carl Roth GmbH, Germany
Quinoline	FPLC standard.....	Sigma-Aldrich, MO, USA
Rabbit α -DsRed.....	antibody.....	MBL, MA, USA
Rabbit α -His	antibody.....	Genscript, NJ, USA
Rifampicin.....	antibiotic.....	Duchefa, The Netherlands
Rutin	FPLC standard.....	Sigma-Aldrich, MO, USA
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
Trifluoroacetic acid	acid.....	Carl Roth GmbH, Germany
Tris base	buffer.....	Carl Roth GmbH, Germany
Tris-HCl	buffer.....	Carl Roth GmbH, Germany
Trisodium phosphate	buffer.....	Carl Roth GmbH, Germany
Trypan blue	dye.....	Carl Roth GmbH, Germany
Tween-20.....	non-ionic detergent.....	Carl Roth GmbH, Germany

VII.5 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)		
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)		
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

List of buffers continued

Name	Component	Final concentration [mM] ([g L ⁻¹])	Comment
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	

Curriculum Vitae

Personal details

Name:	Johannes Felix Buyel
Date of birth:	29.01.1983
Place of birth:	Mönchengladbach
Nationality:	German
Family status:	single

Contact information

Address:	Schulring 3, 41812 Erkelenz, Deutschland
Phone:	+49-01632536915
E-Mail:	Johannes.buyel@rwth-aachen.de

Professional experience and projects

Since 10/2009	Process development for the purification of biopharmaceutical proteins (e.g. vaccine candidates) under QbD aspects (DoE, PAT), Institute for Molecular Biotechnology (BioVII), Prof. Dr. Rainer Fischer, Rheinisch-Westfälisch Technischen Hochschule Aachen (RWTH)/ Fraunhofer Institute for Molecular Biology und Applied Ecology (IME), Aachen, Germany
09/2009-10/2009	Production of the anti-HIV antibody 2G12 for clinical phase I trials as part of the PharmaPlanta Project, BioVII, RWTH Aachen
09/2008-08/2009	Expression, purification and characterization of an E7-based therapeutic HPV-vaccine candidate, Center for Molecular Biotechnology (CMB), Prof. Dr. Vidadi Yusibov, FhG USA in Newark, DE, USA
04/2008-07/2008:	Expression and localization of the sox-11 cancer marker in different tumor cell lines and primary tumor tissue, Department for Immunetechnology, Prof. Dr. Carl Borrebaeck, University of Lund, Lund, Sweden
05/2007-07/2007:	Cloning of an antibody-/antigen-vector and recombinant expression in <i>Nicotiana tabacum</i> , BioVII, RWTH Aachen
10/2006-02/2008:	Laboratory manager, development and evaluation of the BioLector technology, m2p-labs, Aachen
07/2006-09/2006:	Quantification of the impact of medium ionic strength on the growth of <i>Escherichia coli</i> , Chair for Bioengineering (BioVT), Prof. Dr.-Ing. Jochen Büchs, RWTH Aachen
05/2006-08/2006	Cloning, transient expression, purification and thermodynamic characterization of fluorescent protein HIV-epitope fusions, RWTH Aachen
05/2005-06/2006:	Development and maintenance of the RAMOS shaking flask system (Respiration activity monitoring system), BioVT, RWTH Aachen

Scientific education

- Since 10/2009: PhD studies at the Institute for Molecular Biotechnology, Prof. Dr. Rainer Fischer, RWTH Aachen
- 09/2008-08/2009: Master thesis at CMB, Prof. Dr. Vidadi Yusibov, FhG USA
- 10/2006-09/2009: M. Sc. in Biotechnology/Molecular Biotechnology (with honors), RWTH Aachen
- 10/2003-09/2006: B. Sc. in Biotechnology/Molecular Biotechnology, BioVII, RWTH Aachen

Stipends and awards

- 12/2012: Fraunhofer IME Cutting Edge Award
- 05/2010: Springorum Denkmünze of the RWTH
- 09/2008-03/2009: Stipend of the Fraunhofer USA Inc.
- 04/2008-07/2008: ERASMUS Stipend

Publications/Posters/Talks

- 2013: Advanced Analytics for Therapeutic Proteins, Kloster Irsee, Germany
PREP Symposium, Boston, USA
Plant-Based Vaccines, Antibodies & Biologics, Verona, Italy
Protein and Antibody Engineering Summit, Boston, USA
Predictive models for the accumulation of a fluorescent marker protein in tobacco leaves according to the promoter/5'UTR combination. Buyel *et al.*, Biotechnol. Bioeng.
Processing heterogeneous biomass: overcoming the hurdles in model building. Buyel und Fischer, Bioengineered.
- 2012: Invited speaker at the 4. European DOE User Meeting, Vienna, Austria
Predictive models for transient protein expression in tobacco (*Nicotiana tabacum* L.) can optimize process time, yield, and downstream costs. Buyel und Fischer, Biotechnol. Bioeng.
- 2011: International Conference on Biopartitioning and Purification, Puerto Vallarta, Mexico
Extraction, purification and characterization of the plant produced HPV16 subunit vaccine candidate E7 GGG. Buyel *et al.*, J Chromatogr B.

Additional activities

10/2006-09/2007: DECHEMA Chemcar-Contest
06/2004: Local elections candidate in NRW
05/2003-09/2003: Ancillary employee at notary Christoph Binger in Erkelenz
Since 05/2003: Voluntary engagement for youth work in the St. Adelgundis congregation in Arsbeck

Education

07/2002-04/2003: Alternative service
Youth work at the Johann Baptist congregation Wildenrath
07/1993-06/2002: Cusanus-Gymnasium Erkelenz
07/1989-06/1993: Franziskus-Schule Erkelenz

Language skills

English (fluent)
French (basic knowledge)
Latinum

Interests

Badminton
Basketball
Kung Fu
Politics
Sailing
Soccer

Erkelenz, 29.07.2013

Johannes Felix Buyel