



Analysis of a Tyrosine-and Lysine-Rich Fusion Protein Engineered for the Production of Cell Wall-Bound Recombinant Proteins

Nils Saillard

► To cite this version:

Nils Saillard. Analysis of a Tyrosine-and Lysine-Rich Fusion Protein Engineered for the Production of Cell Wall-Bound Recombinant Proteins. Agricultural sciences. 2013. dumas-01088584

HAL Id: dumas-01088584

<https://dumas.ccsd.cnrs.fr/dumas-01088584>

Submitted on 28 Nov 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**AGROCAMPUS
OUEST**

☒ **CFR Angers**

☐ **CFR Rennes**



Année universitaire :

2012-2013

Spécialité :

Horticulture

Spécialisation :

BIOTECH (Suivie à AgroParisTech)

Mémoire de Fin d'Études

- ☒ d'Ingénieur de l'Institut Supérieur des Sciences agronomiques, agroalimentaires, horticoles et du paysage
- ☐ de Master de l'Institut Supérieur des Sciences agronomiques, agroalimentaires, horticoles et du paysage
- ☐ d'un autre établissement (étudiant arrivé en M2)

Analysis of a Tyrosine- and Lysine- Rich Fusion Protein Engineered for the Production of Cell Wall-Bound Recombinant Proteins

Par : Nils SAILLARD

Soutenu à Paris le 19/09/2013

Devant le jury composé de :

Enseignant référent : Sylvain Chaillou

Sophie Landaud

Aurélie Baliarda

Les analyses et les conclusions de ce travail d'étudiant n'engagent que la responsabilité de son auteur et non celle d'AGROCAMPUS OUEST

Analysis of a Tyrosine- and Lysine- Rich Fusion Protein Engineered for the Production of Cell Wall-Bound Recombinant Proteins

List of abbreviations (and glossary)

A.A. = Amino Acids

BSA = Bovine Serum Albumin

C5a = Complement factor 5a

CaMV = Cauliflower Mosaic Virus

to CIP = to perform a Calf Intestinal Phosphatase treatment

DNA = Deoxyribo-Nucleic Acid

ELISA = Enzyme-Linked Immunosorbent Assay, a test that uses antibodies to detect and sometimes quantify, in aqueous samples, a substance (a specific protein for instance)

EGFP = Enhanced Green Fluorescence Protein

Hemizygous = A diploid individual or cell is hemizygous for a gene when there is only one copy of this gene present on a pair of chromosomes at one locus. This is different from a heterozygote which displays two different alleles of the gene in matter on the two chromosomes at the same locus when a hemizygous has simply no corresponding gene or sequence on the pairing chromosome.

IL6 = Interleukine 6

LeTLRP = Tyrosine- and Lysine- Rich Protein isolated from *Lycopersicon esculentum* (Tomato) = the cell wall anchor in this study

min = minutes

ML = Mature Leaf tissue

MQ = water filtered by the Milli-Q® system from Millipore© corporation = Ultra-pure water

M.S. = Murashige and Skoog

MS = Mature Stem tissue

NIC = Nearly Isogenic Control= Non transgenic and nearly isogenic line used as a control (and for transformation)

NtTLRP = Tyrosine- and Lysine- Rich Protein isolated from *Nicotiana tabacum* (Tobacco)

PBS = Phosphate-Buffered Saline

pbs = pairs of base

PCR = Polymerase Chain Reaction

rpm = Rotations Per Minute

RT = Room Temperature $\approx 20 \pm 5^\circ\text{C}$

T0 = Original transgenic generation, here comprised of plants regenerated on artificial medium from leaf tissue transformed with *A. tumefaciens*.

T1 = The first filial generation of transgenic plants, which is comprised of offspring resulting from the self-cross of the T0 generation.

T2 = The second filial generation of transgenic plants, which is comprised of offspring resulting from the self-cross of the T1 generation.

TCA = TriChloroacetic Acid

WT = Wild Type

Tris = tris(hydroxymethyl)aminomethane

YL = Young Leaf tissue

TSP = Total Soluble Protein

YS = Young Stem tissue

Résumé

C5a et IL6 sont de petites protéines qui sont déjà produites à l'échelle industrielle par fermentation microbienne, ceci à des fins pharmaceutiques. Il a déjà été démontré qu'il était également possible de faire produire ces protéines par des lignées de tabac génétiquement transformées. Une nouvelle construction génétique adressant ces protéines vers l'apoplasme pour les stocker dans la paroi cellulaire a été mise au point, et ce dans le but de faciliter le processus de purification et ainsi réduire les coûts de production. Une telle construction fut obtenue en réalisant une protéine de fusion entre l'IL6 ou le C5a et une protéine pariétale : la TLRP (Protéine Riche en Tyrosine et en Lysine). Une fois cette construction intégrée à une lignée de tabac, les rendements potentiels de ces plantes, ainsi que la localisation subcellulaire des protéines recombinantes produites furent étudiées afin de vérifier si le système ainsi créé fonctionnait de manière efficace. Différentes lignées, tissus et fractions cellulaires furent ainsi étudiés afin de déterminer comment tirer profit au maximum de cette technique. Il fut possible d'isoler, à partir de la paroi végétale de ces plantes, une partie des protéines recombinantes produites mais pas plus de 50% de celles-ci, la majorité des protéines extraites se trouvant sous forme libre dans l'apoplasme. De plus les quantités de protéines recombinantes globalement produites par ces plantes sont relativement faibles ; les jeunes feuilles constituent, d'après les analyses réalisées, la matière première la plus prometteuse, avec un rendement maximum de protéines isolées à partir de la paroi végétale se situant aux environs de 1 pg d'IL6 par mg de poids frais. Une tentative de créer une construction utilisant la EGFP (Protéine Fluorescente Verte Améliorée) afin de réaliser une analyse plus complète de ce système de production fut également explorée mais un phénomène de délétion lors du clonage empêcha sa réalisation.

Abstract

C5a and IL6 are small proteins that are already industrially produced for pharmaceutical purposes by microbial fermentation. It was shown that producing them in stably transformed tobacco was possible. Going further, a new genetic construct addressing those proteins to the apoplast to bind and store them in the cell wall was made. This system was achieved using a fusion protein which included the TLRP (Tyrosine- and Lysine- Rich Protein). The purpose of this construct was to ease the purification process and reduce its cost. The yields that this technique could produce and the actual localizations of the recombinant proteins were then investigated in order to know if this system was working efficiently. Different lines, tissues and cellular localizations were investigated to find out the best that could be achieved out of this system. Some of the recombinant protein could, as expected, be recovered from the cell wall but no more than 50% of the whole production as most of the collected recombinant protein was found free in the apoplast. The overall production of those plants was anyway low. The maximum yield of IL6 gathered from the cell wall being around 1 pg/mg of fresh weight for young leaves, which is the most promising tissue analyzed. A new construct using EGFP was as well prepared in order to further analyze this system but the cloning failed due to some deletion phenomenon which occurred during the cloning step.

Key Words

Molecular farming, cell wall anchoring, IL6, C5a, Tyrosine- and Lysine- Rich Protein

Table of Contents

List of abbreviations (and glossary)	i
Résumé	ii
Abstract	ii
Key Words.....	iii
Table of Contents	iii
Table of Figures.....	v
Table of Tables.....	v
I. Introduction	1
A] Producing C5a and IL6.....	1
B] The Tyrosine- and Lysine- Rich Protein (TLRP), a protein anchored to the cell wall.....	2
1. Presentation of the Tyrosine- and Lysine- Rich Protein.....	2
2. The TLRP in a fusion protein.....	2
C] Genetic constructs used in this study.....	3
D] Reporter genes.....	4
E] Stably transformed tobacco (<i>N.tabacum</i>) as the expression system.....	5
II. Material and Methods.....	6
A] Evaluation of the C5a and IL6 productions in <i>N. tabacum</i> using the TLRP system.....	6
1. Tobacco growing conditions	6
2. Identification and characterization of the transgenic plants and lines.....	6
3. Extraction and sampling protocols	9
4. Total soluble protein content assessment by Bradford Protein Assay.....	11
5. ELISA method for quantification of C5a and IL6.....	12
B] Visualizing the expression of TLRP in plant.....	12
1. Creation of the transformation vector: pLH-TLRP-t-EGFP.....	12
2. Selection and verifications of the clones	12
III. Results	13
A] Evaluation of the C5a and IL6 production in <i>N. tabacum</i> using the TLRP system	13
1. Segregation pattern of the transgenes.....	13
2. Preliminary experiments.....	16
3. Recovering the IL6 content from the plant material.....	18
B] Visualizing the expression of TLRP in plant.....	21

1.	Failure to identify a valid pLH-TLRP-t-EGFP	21
2.	Localization of the deleted region	22
IV.	Discussion	23
A]	Evaluation of the C5a and IL6 productions in <i>N. tabacum</i> using the TLRP system.....	23
1.	Identification and characterization of the transgenic plants and lines	23
2.	Stability and handling of the protein samples	24
3.	The production of IL6	27
4.	TLRP properties	31
B]	Visualizing the expression of TLRP in plant.....	32
	Conclusion.....	33
	Bibliography:.....	34
	Appendices:	37
	Appendix A: Sequences of the primers used in this study	37
	Appendix B: Examples of resistance to antibiotic test's results	38
	Appendix C: Table of the probabilities of the T2 resistance test's outcomes.....	39
	Appendix D: Table of probabilities of the possible genetic segregation pattern in T1.....	41
	Appendix E: IL6-ELISA data for the tissue yields of the different lines (full set of data).....	43
	Appendix F: Yields of IL6 recovered from the cell wall and the other compartments	44
	Appendix G: Proportions of IL6 recovered from the cell wall	44
	Appendix H: Digestion of the origin vectors.....	45
	Appendix I: Theoretical restriction pattern of pLH-LeTLRP-t-EGFP's digestion by FokI	46
	Appendix J: Digestion of a pLH-LeTLRP-t-EGFP candidate by FokI.....	47
	Appendix K: Analysis of the result of the PCRs performed on pLH-LeTLRP-t-EGFP	48
	Appendix L: Analysis of the restriction patterns of the pLH-LeTLRP-t(...) vectors	49
	Appendix M: Mapping of the missing sequence of pLH-LeTLRP-t-EGFP	50

Table of Figures

Figure 1: IL6 3D structure	1
Figure 2: Immunochemical histolocalization of NtTLRP	2
Figure 3: pLH-LeTLRP-t-EGFP genetic map	3
Figure 4: PCR testing the presence of the transgenes in the F1 generation	14
Figure 5: Total IL6 Yields of the line 16: full set of data.....	19
Figure 6: Comparison of the IL6 yields of the different tissues sampled.....	20
Figure 7: Comparison of homozygote and hemizygote plants' yields	20
Figure 8: Yields of IL6 recovered from the cell wall and the other compartments.....	21
Figure 9: Mapping of the missing sequence of pLH-LeTLRP-t-EGFP	22

Table of Tables

Table 1: PCR Mix used for the control PCR	7
Table 2: List of the primers used for the control PCR.....	7
Table 3: Thermocyclor protocol used to for the nptII ^{***+} , C5a ^{*++} and IL6 ^{*+} PCRs	7
Table 4: Results of the resistance tests performed on the T2 generation	14
Table 5: Zigosity distribution pattern in T1 and comparison to the expected model	15
Table 6: Average TSP concentrations and yields.....	16
Table 7: IL6 yields primarily detected in the protein fractions	18

I. Introduction

In the field of biotechnology the use of plants for the production of therapeutic proteins is emerging thanks to its advantages that are currently being discovered and explored (Joshi & Lopez, 2005). Although microbial fermentation still remains the industrial process that is most widely used, other organisms such as plants can be used in an attempt to decrease the costs of production and industrial processes through the creation of particular systems. In the present work, we studied how to use the plant's cell wall as a compartment to accumulate and store the recombinant proteins IL6 and C5a which are small pharmaceutical proteins industrially produced by microbial fermentation. Engineering this system would allow us to use the cell wall as a substitute for the costly chromatographic columns which are usually used for purification (Sharma, 1989). In order to have the expected outcome a Tyrosine- and Lysine- Rich Protein (TLRP) sequence was fused to the sequence of either C5a or IL6 and further transformed and expressed into tobacco (*N. tabacum*). In between the sequence of the cell wall anchor (the TLRP) and the pharmaceutical protein (IL6 or C5a), the cleavage site of a protease was inserted in order to be able to release and recover the recombinant proteins once the impurities were washed away from the cell wall. This technique, referred to as the TLRP system in this study, could have some industrial advantages: using it, it was hoped to decrease the production costs of the recombinant proteins by obtaining good yields of recombinant proteins and simplifying the processing of the plant material.

A] Producing C5a and IL6

In this study, the therapeutic proteins produced as a test are the complement factor 5a (C5a) and the interleukin 6 (IL6). Both of them are small proteins (75 and 212 A.A.) which have a role in many biological processes and pathologies (Nausch et al., 2012a, 2012b, 2012c and 2013). They are mainly known to be primary messengers involved in the transmission of the information in the immune system. Succeeding in producing them efficiently and for a low cost could have many applications. For example, IL6 and C5a can be used to produce antibodies to treat septic shock or some cancers (Fulciniti et al., 2009; Riedemann et al., 2003). Producing such antibodies is realized by injecting human C5a or IL6 in some farm animals such as rabbits but to do so the C5a and IL6 antigens must be produced first (Czermak et al., 1999).

Producing C5a and IL6 especially represents a challenge as they are small proteins involved in the signaling in the blood stream, which means they are particularly unstable and have a tendency to aggregate. Some post-harvesting storage solutions have already been found (Kenis et al., 2002). However, preventing their degradation while producing and purifying them from plants seems especially challenging (Nausch et al., 2012b and 2012c).

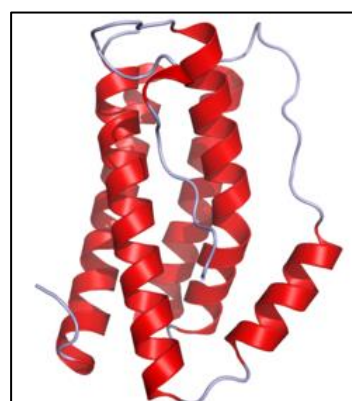


Figure 1: IL6 3D structure

Somers, Stahl & Seehra. 1997

B] The Tyrosine- and Lysine- Rich Protein (TLRP), a protein anchored to the cell wall

1. Presentation of the Tyrosine- and Lysine- Rich Protein

The TLRP is a small protein (less than 150 A.A.) expressed in several vascular plants. It was discovered in tomato first. Some analog genes of this TLRP were then found in tobacco and arabidopsis (Domingo et al., 1994 and 1999). The native TLRP has been found to bind to the secondary cell wall and to accumulate in the secondary xylem (Domingo et al., 1994).

Some of the proteins bound to the cell wall can be extracted out of it. For instance, HPRGs (Hydroxy-Proline Rich Glycoproteins) are bound to the cell wall through their massive glycosylation pattern. Interactions are thus based on ionic and hydrogen bonds (Showalter, 1993). These interactions can be broken using the salting out effect of a highly salted extraction buffer (Christoph Unger, personal communication, Department of Agrobiotechnology, Agricultural and Environmental Faculty, University of Rostock). Yet the TLRP seems embedded in the cell wall and it has not been possible to extract it from the plant material so far (Domingo et al., 1999). However, the TLRP is indeed present in the cell wall since it has been possible to detect its presence through in situ immunohistochemical localization (Figure 2). These observations lead to the conclusion that the binding of the TLRP to the cell wall structure must be covalent. Given the TLRP structure, its binding would be the result of the crosslinking of tyrosine residues with the lignin components of the cell wall. Tyrosine dimerizations (or isodityrosine linkages) and the formation of disulfide bounds in between cysteine residues are also very likely to happen. Therefore, the TLRP would not only bind to the lignin but also to other proteins and itself, participating in the scaffolding of the cell wall. Both reactions of crosslinking involving tyrosine residues should be dependent on the activity of cell wall peroxydases which is fed by hydrogen peroxide (H_2O_2) (Domingo et al., 1994 and 1999).

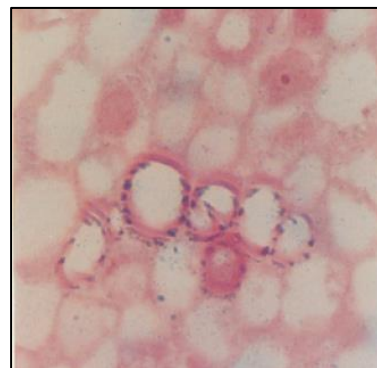


Figure 2:
Immunohistochemical
localization of
NtTLRP

Domingo et al. 1999

The marker (blue) shows the accumulation of the TLRP protein along the xylem tissues among all cell walls (pink)

2. The TLRP in a fusion protein

Fusing the TLRP with another protein adds in to the chimeric protein obtained several domains; especially a signal peptide that will address the protein to the apoplast and the domains responsible for the crosslinks (Domingo et al., 1999). Domingo et al. (1999) have already made the proof of concept that such a fusion protein ends up tightly bound to the cell wall. However, in this study it is the ability to recover the recombinant protein (freed from its TLRP part by a protease) which will be tested. To do so it was planned to fuse the IL6 or C5a sequence with the one of the TLRP and to add the restriction site of the thrombin in between. The thrombin enzyme was chosen because it is an animal protease acting in the blood stream of mammals, there was then few chances that unwanted interactions happen within this system (Coughlin, 2000).

The interest of this study lies in the idea to use this system to produce and purify IL6 and C5a: the TLRP part of the fusion protein would serve to hold it attached to the insoluble cell wall while

ANALYSIS OF A TYROSINE—AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS

washing the soluble impurities from the homogenized plant material. Then the IL6 or C5a would be recovered by resuspending the washed and centrifuged cell wall directly into a buffer where the recombinant proteins could be extracted and stored. In this solution the IL6 or C5a would be released by performing a thrombin digestion. After the digestion the solution which would contain the IL6 or the C5a from this moment would be centrifuged to remove the cell wall. Extended purification steps to remove the other proteins than the ones produced with this system and the metabolic contaminants would then not be necessary. Theoretically the only contaminant that could interfere would be the thrombin but with a well thought protocol the riddance of the thrombin could be easily performed.

When the TLRP system was conceived it was hoped that this system would also increase the recombinant protein yields compared to other systems because the TLRP would stabilize the IL6 or the C5a proteins. Those wishes were based on the fact that fusing an additional sequence to the N-ter extremity of a short and unstable protein can sometimes increase its half-life (Hondred et al., 1999). Another way the TLRP could also have protected the recombinant proteins from degradation would have been by simply addressing and binding them to the cell wall, isolating them from proteases and other threats to their integrity that could have been encountered in the apoplast or elsewhere in the cell.

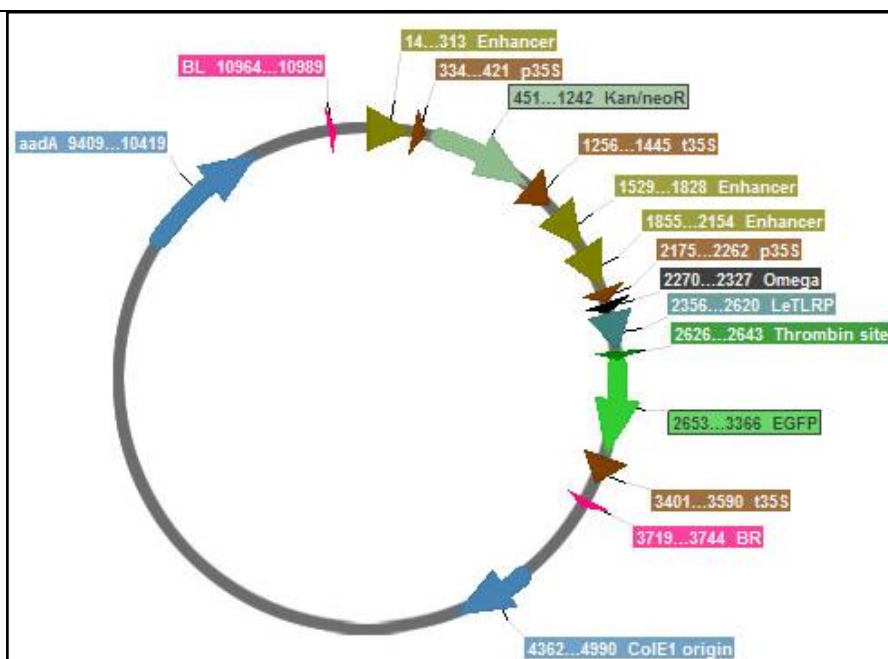
C] Genetic constructs used in this study

Genetic constructs to express the TLRP system and produce C5a and IL6 in plants have previously been designed by Henrik Nausch (Department of Agrobiotechnology, Agricultural and Environmental Faculty, University of Rostock). Several constructs used in this study are similar in many ways to each other as they are all derivatives. Figure 3, representing one of these derivatives: pLH-TLRP-EGFP, can be used as a reference to visualize the spatial organization of the structures explained below.

Figure 3:

pLH-LeTLRP-t-EGFP genetic map

This figure was realized with Serial Cloner©. It is an annotated map of the functional regions of pLH-LeTLRP-EGFP. Each label contains the name of a gene/genetic region and its position on the map (number of the first and the last nucleotide of the labelled region). BL and BR are the Right and Left Borders of the T-DNA region. There are 2 ORF (Open Reading Frame) in this T-DNA. They are both surrounded by a promoter and a terminator: p35S, enhancers and omega are all parts of the promoters and t35S marks the terminators. One of the ORF is the chimeric protein: the cell wall anchor (LeTLRP), separated from the fluorescent marker (EGFP) by the restriction site of the thrombin (thrombin site). The other ORF is the resistance gene Kan/NeoR (also called nptII) which is a maker for plant selection. Outside the T-DNA region, delimited by the left and the right borders (BL and BR), there are; the bacterial gene which confer resistance to spectinomycin and spectromycin (aadA); ColE1, the origin of replication for amplification in *E.coli*. and VS1, the origin of replication for amplification in *Agrobacterium*. VS1 is not labelled on the map though it is present in between aadA and ColE1 on the map.



The sequences expressing the transgenes were optimized to have a high transcriptional and translational rate: the transcription of the chimeric protein was set under the control of a 35S promoter from the cauliflower mosaic virus (CaMV) of which enhancer has been duplicated (Omirulleh et al., 1993). After the 35S region an Ω region was added in order to improve the transcription as well (Gallie, Lucas & Walbot, 1989). The Ω was followed by the sequence coding for the fusion protein, which has been codon optimized for expression in tobacco (Rouwendal et al., 1997). The fusion protein sequence is composed by the *TLRP* from tomato (*LeTLRP*) linked to the sequence coding for the protease cleavage site (the thrombin site) and the gene of interest (*IL6* or *C5a*). To finish the sequence ends with a 35S terminator from the same virus (the CaMV).

The cassette described above was inserted into the backbone of pLH9000. This vector contains genetic material for replication and transformation in *E. coli* (ColE1) and in *Agrobacterium* (VS1), as well as selection markers for bacterial (*aadA*) and plant (*nptII*) selection. The vector containing the *IL6* version of the protein construct was called pLH-LeTLRP-t-IL6 and the one containing *C5a* was called pLH-LeTLRP-t-C5a.

D] Reporter genes

Several questions have been raised concerning the plants expressing the TLRP protein system: does the fusion protein accumulate in the cell wall as expected? How does it settle exactly in the cell wall? In Domingo's study (1999), the accumulation of TLRP was very likely the result of both its ability to bind to the lignin and the promoter used but what is the pattern of accumulation of the fusion protein in the different tissues and organs with the construct used in this study? Do the pH and the oxidative burst affect the binding of TLRP to the cell wall and its accumulation (Domingo et al., 1994)? Using thrombin, which proportions of recombinant protein can be recovered from the cell wall and which proportions stay bound to it?

Thanks to techniques such as ELISA or immunochemical histolocalization, the C5a and IL6 constructs could be used to try to answer these questions but it would be complicated, expensive and time consuming to do so. It was then decided to use reporter genes to study the TLRP fusion proteins' behavior. Though using a marker gene may not exactly reflect the production of C5a or IL6 because protein markers are mostly longer, more stable and have their own properties, using a marker protein will have the advantage of being easier to manipulate and will allow a preliminary characterization of the TLRP system. Several candidates were considered for the marker including fluorescent proteins and the glucuronidase GUS.

Replacing IL6 or C5a by a fluorescent marker would allow us to investigate different aspects of the hereby created system. A TLRP fused with a fluorescent protein could be observed *in vivo* using confocal microscopy, making it possible to visualize how it is deposited in the cell wall (Haseloff et al., 1999). It would also permit measuring the concentrations of the recombinant protein quickly and for a low cost using a fluorometer (Remans et al., 1999). A leakage or a loss before, during and after the extraction process could then be easily identified and measured, revealing flaws in the system or the extraction protocol.

For instance, the fluorescence would be a very useful technique to test if a part of the produced proteins stays bound to the cell wall even after digesting the cell wall with some thrombin. Indeed it is a possibility that, for some technical or theoretical reasons, the thrombin treatment would fail or work partially. For instance the recombinant proteins could be embedded in the dense and impermeable matrix of the lignified cell wall which would keep them from being digested, and so the protein of

interest would not be released. In that situation in which the recombinant proteins were to be sterically inaccessible for the thrombin then the access for other solutes useful for detection (dyes, antibodies, probes ...) could be compromised as well. Meanwhile a laser would have better chances to go through the cell wall to reveal a fluorescent signal.

It could also be that some proteins get lost during the washing steps. A fluorescent marker would be handy in that situation too as it would be easy to detect and quantify losses in the eluates.

The most obvious choice of fluorescent protein would be the EGFP (Enhanced Green Fluorescent Protein) that has been widely used so far for its stability and the strong intensity of its signal (Hraška, Rakouský & Čurn, 2006). In case EGFP would not be suitable, there are other alternatives. mCherry for instance, a DsRed derivative, offers a good alternative to EGFP (Patterson, Day & Piston, 2001; Shaner et al., 2004).

E] Stably transformed tobacco (*N.tabacum*) as the expression system

Though the production of C5a and IL6 is already available from microbial fermentation (which gave good yields and has already been studied to be improved - Nausch et al., 2012a and 2013), using a plant as a producer instead could reduce production costs and allow an easy scaling up (Twyman et al., 2003). The tobacco plant was chosen to test the TLRP system because it is an organism with promising potential for molecular farming: it has a good biomass yield, being a crop it is easy to cultivate and gather, however it is not a food culture so there is little risk of misuses, it is easy to transform, grow and regenerate *in vitro*, the risks of unwanted cross-pollination are scarce when good precautions are taken, it produces numerous offspring and it already has been used in many studies, including some which aimed to produce C5a and IL6 (Tremblay et al., 2010; Nausch et al., 2012b and 2012c). Nausch et al. (2012b; 2012c) thus revealed that addressing the C5a to the vacuole and the IL6 to the endoplasmic reticulum (ER) gave higher yields compared to addressing them to the apoplast. Nevertheless using the TLRP system could still have given good yields as the TLRP does not only provide a simple addressing to the apoplast but also a cell wall binding tag. Moreover the main issue with proteins produced in plants is the downstream processing which can represent 80% of the costs, and the TLRP system is an attempt to improve this particular aspect (Roque, Lowe and Taipa, 2008).

Some tobacco leaf discs from the *N.tabacum* 'Geudertheimer' cultivar were then transformed with pLH-LeTLRP-t-IL6 and pLH-LeTLRP-t-C5a. Only the transformed T0 individuals with one copy of the transgene in their genome and the best yields were kept for use in this study. These individuals were identified using southern blot and ELISA (Unpublished data). The use of plants having a single transgenic locus greatly simplified the analysis of the plant and happened to give the best yields in other studies (Nausch et al., 2012b and 2012c). Going further, tobacco seeds harvested from this T0 were the starting point of this study. The lines used for the C5a experiments derived from the T0 plant 39 and the plant characterized by the event 16 gave the material for the IL6 tests.

II. Material and Methods

A] Evaluation of the C5a and IL6 productions in *N. tabacum* using the TLRP system

1. Tobacco growing conditions

In December, *N. tabacum* seeds from T0 lines 39 for the C5a construct and 16 for the IL6 construct were sowed in 5L pots filled with potting soil. For each line 10 pots were sowed. After few weeks the pots were thinned out to let only one plant develop per pot. The plants were kept in the greenhouse during the whole period. A whitening of the pane and an intensification of the watering were performed in summer but the settings in the greenhouse were not intensively controlled or monitored. The plant tissues were harvested for the first time in March. At the beginning of April, the plants started to flower due to the conditions induced by long days. The tobacco being an annual crop, after finishing its flowering it starts to wither and dies. Hence a repropagation by cutting was necessary before the seed harvest in the end of April. In June a second tissue harvest was performed on the cuttings from April while flowering bud started to appear again, the reproduction cycle being shortened by the summer weather conditions (temperature, light intensity, hours of sunshine).

Mature infructescences were gathered and stored in paper bags to dry for a few weeks after being cut. Once the fruits were dry, they were cut open to get the many seedlings that they can contain. The gathered seeds were then stored to dry once more in open ependorf tubes for few days before those were closed.

2. Identification and characterization of the transgenic plants and lines

a. Strategy to assess the zygosity of the T1 plants

The plants of the T0 generation are normally hemizygous for the transgenes. When there is only one insertion event, as in this study, the product of the self-crossing of this hemizygous T0 generation, called T1, should be a mixture of individuals which completely lack the transgene (Wild Type-homozygotes), individuals carrying a single copy of the gene (Transgenic hemizygotes) and finally individuals containing two copies of the transgene (Transgenic homozygous). According to the Mendelian segregation laws, it can be predicted that those categories should be found respectively in 25-50-25 % proportions.

A PCR targeting the transgenic sequences can be performed to reveal which plants are transgenic and which are wild type. The T1 generation was then tested by PCR. Meanwhile, it is not as simple to distinguish the two transgenic categories since both would give a positive result if a PCR was performed. The distinction between homo- and hemi- zygotes can however be done by analyzing the segregation pattern in the next self-crossed generation (T2): further to a Mendelian segregation, a transgenic hemizygous plant would give 75% transgenic offspring whereas transgenic homozygous plants would engender only transgenic plants. The T2 generation was then tested to reveal the zygosity of the T1 plants they originated from.

The T-DNA inserted in the genome of the plants in this study included an antibiotic resistance gene to kanamycin (*nptII*). It is then possible to reveal which plants are transgenic by selecting them on a media containing antibiotics. A resistance test is not as demonstrative as a PCR but it is more

ANALYSIS OF A TYROSINE—AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS

costly and it requires more handling time to perform PCR. A resistance test is therefore fitter for large scale screenings. This is why it was used to assess the proportions of transgenic plants in the T2 sprouts.

b. Identification of the transgenic plants by PCR

i. DNA extraction

The DNA extractions were performed using a fast sampling protocol: For each plant a 100 mg leaf sample was frozen in liquid nitrogen and grinded with a little plastic pestle directly into an ependorf tube. After the homogenization, 400 μ L of a lysis buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl) was added to each sample. Once melted, the samples were vortexed, incubated 10min at 95°C and put back to thaw on ice. The samples were then centrifuged and the supernatant, being the DNA template, was kept.

ii. Control PCR

The following PCR mix was used:

Table 1: PCR Mix used for the control PCR

Solution	Amount (μ L)
PCR master mix (Promega®)	5
Forward Primer (10 mM)	1
Reverse Primer (10 mM)	1
DNA sample (1:5)	1
H ₂ O	2
Total	10

For each tested gene a different set of primers was used (all the sequences of the primers used in this study can be found in the *Appendix A*):

Table 2: List of the primers used for the control PCR

Gene tested	Forward Primer	Reverse Primer
nptII	2npt-fw	2npt-rv
C5a	c5a-fw	c5a-rv
IL6	IL6fw	IL6rv

The following cycle was used:

Table 3: Thermocyclor protocol used to for the nptII^{*}, C5a^{***} and IL6^{***} PCRs**

Step	Temperature (°C)	Time	Repeats
Preheating	94	2 min	
Denaturation	94	30 sec	x 35
Annealing	55*or 60**	30 sec	
Polymerization	72	1 min ⁺ or 45 sec ⁺⁺	
Finalization	72	7 min	
Storage	4	∞	

The PCR products were loaded into a 2% agarose gel containing 0.001% ethidium bromide and then spread by electrophoresis before being revealed under U.V. light (Ultra-Violet).

ANALYSIS OF A TYROSINE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS*c. Antibiotic resistance test**i. Sterilization and growth on selective media of the tobacco seeds*

The seeds were sterilized by dipping them in a 0.5 kg/L sodium hypochlorite solution for 5 min. They were then carefully washed in sterile water. The sterilized seeds were sowed one by one on petri dishes using a pattern that allowed plating 50 seeds per dish. The petri dishes used were filled of agar-M.S. medium supplemented with 100 µg/mL kanamycin. In addition, plates without antibiotic were used to plate some seeds of all origins as negative controls.

ii. Identification of the resistant sprouts

Resistant sprouts were identified visually. The transgenic offspring was dark green and well developed (which was also the condition observed on seeds grown in the control plates) whereas the susceptible wild type plants grew yellow and puny (See *Appendix B*).

d. Statistical analysis of the results

The hypothesis that the segregation pattern was following a simple Mendelian model was made. Therefore probabilities of the different draws were theoretically calculated. The probabilities of each draw were then sorted by increasing value of the likelihood in order to help to determine whether an obtained draw was more or less likely to be obtained compared to the other draws. In some case the repartitions of the draws were compared as well without computation. The data generated for the statistical analysis are displayed in the *Appendices C and D*.

3. Extraction and sampling protocols*a. Harvesting and post harvesting considerations*

The tissues were sampled in the greenhouse directly from the plants using a pruner and/or a scalpel. The plant tissues had to be sampled fresh at the beginning of the study: ice expansion and crystal formation during the freezing would destroy the cells' wholeness, which is necessary in order to separate the proteins from the apoplast from the cytoplasmic proteins. Later on, when it was not a concern to separate those fractions, the tissues were quickly dipped in liquid nitrogen after the harvest to be stored at -80°C until use.

As the plants used in this study were too old at the time of the sampling and had developed different architectures over time, it was not possible to sample the tissues by developmental stage based on their position on the plant. The most recently developed tissues which were soft, light green and close to the apical meristem(s) were then considered as young leaves or stems whereas the mature tissues were picked dark green and harder. The mature samples were fully developed tissues which would have been ready to be harvested in an agronomical context. They were collected close to the base of the plant while sampling any senescent yellowish leaves was avoided. The different tissues were collected on a same plant in the following order: Young Leaf (YL), Young Stem (YS), Mature Leaf (ML) and Mature Stem (MS). When the stem(s) were sampled they were taken from above or below the node(s) from which the sampled leave(s) belonged.

Except for the preliminary ELISA (III.A.3.a), the tissue samples used in this study were always weighted prior to being submitted to any extraction protocol in order to be able to determine the yields per gram of fresh weight. The weight of the samples ranged between 200 and 1000 mg.

b. Extraction protocol of the protein fractions

One of the aims of this study was to localize recombinant proteins *in planta*. Several localizations were investigated. Sometimes different fractions were pooled together for technical or practical reasons during the analysis.

Repeated washing steps in between each fraction's collection were carried out to ensure there was no contamination from the other fractions. A washing step usually consisted of adding 2mL of extraction buffer, vortexing, centrifuging and discarding the supernatant. These series of steps were repeated 3 times to complete a wash before passing to the next extraction step.

Samples were wished to be used immediately but it was not always possible. In that case, the samples were stored at 4°C for short term use or -20°C for long term storage when necessary.

- *The apoplastic proteins*

The apoplastic fraction was obtained after vacuum infiltrating some PBS (Phosphate-Buffered Saline) into raw tissue pieces. The infiltrated tissues were then vortexed using two sets of centrifuge tubes; the first was holding back the tissue pieces and it was pierced to let the apoplastic fraction run into the second tubes, the collector ones. This fraction was about 200-500 µL depending on the sample extracted. The collection of this fraction was not followed by a washing step as describe above; but by two other similar vacuum infiltrations along with centrifugations in order to wash the remaining apoplastic proteins.

- *The cytoplasmic proteins*

The cytoplasmic fraction was obtained by homogenizing the plant material, either using a blender (with a lot of extraction buffer, up to 10 mL of a K₂PO₄ solution at 50 mM) or using procellys® (2x20sec at 6800 rpm) after freezing the samples in liquid nitrogen with 1-1,5 mL of PBS.

When the hand blender was used few ice cubes were added to the sample when necessary to keep it from getting warm. On the other hand, when procellys® was used, , the samples were kept cold by placing them on ice or in a cold steel block in between the homogenization steps.

- *The proteins bound to the cell wall by intermolecular forces*

The “ionic” fraction was obtained by extracting cell wall samples with a PBS buffer enriched with 1 M NaCl (called high salt buffer) to obtain a salt out effect. Either 1 or 3 mL of this solution was use to resuspend the homogenized cell wall. The cell wall was then centrifuged back to collect this fraction.

Occasionally, the “ionic” fraction was also sampled before the homogenization by vacuum infiltration some “high salt buffer” in the tissues and collecting it back the same way the apoplastic fraction was sampled.

ANALYSIS OF A TYROSINE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS

- *The thrombin digestion*

The thrombin digestion was done by suspending the washed pellet of blended cell wall into 1 or 2 mL of extraction buffer (3X TBS, 1 mM NaCl) along with 20 μ L of thrombin solution (1 U/ μ L, 10% BSA). The mix was incubated overnight (\approx 16h) at RT (Room Temperature) on a rotating wheel. After the incubation the sample was centrifuged and the supernatant was analyzed.

For the very first thrombin digestions a leftover of thrombin solution (10 U/ μ L, no BSA) from the stock (-20°C) was used. Given that it was 10 times more concentrated than what was going to be used after, only 2 μ L of the thrombin solution were used in the first experiments (this concerns the protein samples extracted to be analyzed during the first ELISA - III.A.4.a).

When this first thrombin batch from GE Healthcare© was finished, a new batch was ordered from Sigma-Aldrich© (product reference: T4648-1KU). When the product was received, it was dissolved in a 10% albumin solution up to a 1 U/ μ L concentration. The solution made in this way was aliquoted and stored at -20°C .

- c. *Total protein extraction*

In order to obtain the maximum concentration of recombinant protein and compare the different lines on a global level, a total extraction protocol was set. It combines all of the other extraction methods to ensure to get the most recombinant protein as possible:

The plant tissues previously frozen in liquid nitrogen were roughly grinded with a pestle in a mortar then poured into 2mL ependorfs and weighted. 1mL extraction buffer (3XPBS, 1M NaCl) was then added to the tube. Lastly ceramic beads were added on top in order to use Precellys® for a complete homogenization (2x15s at 6500 rpm). Once the sample homogenized, 20 μ L of thrombin solution (1U/ μ L and 10% albumin) was added to the resulting slime. The mix was vortexed, and incubated at room temperature on a rotating wheel for 16h. After incubation the slimy mix was centrifuged and the supernatant was immediately analyzed.

4. Total soluble protein content assessment by Bradford Protein Assay

The Total Soluble Protein (TSP) content of the samples was assessed by Bradford assay. Two types of Bradford assays were performed in this study, one specific to high protein concentrations and one specific to low concentrations. The low range assay was used to determine the TSP concentration of samples within a 0.02-0.2 mg/mL range whereas the high range assay covered the 0.2-2 mg/mL range. If the samples had a TSP concentration higher than 2 mg/mL they could be diluted to fit within the range. The high range Bradford assay was done by adding 10 μ L of sample to 300 μ L of Bradford Reagent (Sigma-Aldrich®) while the low range one used 100 μ L of sample with 200 μ L reagent. The samples were mixed with the reagent directly in a transparent 96 well plate and the absorbance at 495nm was measured with a KC4© fluorometer from Bio-Tek®.

It was possible to assess the TSP concentrations of the samples using a polynomial to model the relation between the optical density at 495nm (OD_{495}) and the total protein concentration of the samples. In order to build and compute the standard curve's equation, a dilution range of an albumin solution of a known concentration was used (2mg/mL, Bovine Serum Albumin [BSA] Standard from Thermo Scientific®).

5. ELISA method for quantification of C5a and IL6

The quantification and detection of C5a was performed using the C5a-ELISA kit manufactured by R&D systems©. The protocol was done according to the instruction provided by the manufacturer. A similar system was used for IL6, using an IL6-ELISA kit and its protocol made available by eBioscience©

Each sample was technically measured twice to ensure the reliability of the obtained result. For a sample, the average of the two measured values was considered along with the associated errors for the analysis and the comparison with the other samples. In the analysis of the results these errors are not mentioned as they were faint compared to the variations experimentally observed.

B] Visualizing the expression of TLRP in plant

1. Creation of the transformation vector: pLH-TLRP-t-EGFP

The *EGFP* gene was amplified via PCR using the expand high fidelity PCR kit by Roche Applied System© according to their protocol. The vector pBI121-MCS-Egfp (from the stock of the Department of Agrobiotechnology, Agricultural and Environmental Faculty, University of Rostock) was used as a template with the couple of primers NruI-EGFP-fw and EGFP-NruI-rv (See *Appendix A* for the sequences) at the annealing temperature of 56°C. The PCR product, the EGFP with NruI restriction sites added on the extremities, was directly digested with the NruI enzyme, along with the pLH-LeTLRP-t-IL6 vector. The pLH-LeTLRP-t- vector obtained was CIPed after digestion to avoid self-ligation. The ligation of the NruI-EGFP-NruI fragment and pLH-LeTLRP-t- was then realized. The product of this ligation was then used to transform *Escherichia coli* using a heat shock based method.

2. Selection and verifications of the clones

Once the transformation done, the transformed bacteria were plated on a selective media containing streptomycin (25µg/mL) and spectinomycin (50 µg/mL) to select the resistant bacteria, that is to say the ones transformed with pLH which contain the resistance gene to those antibiotics. But contamination, empty vectors and especially vectors having integrated the *EGFP* sequence in the wrong orientation would still grow on the dishes, hence the need to check several colonies before to transfer the construct to *Agrobacterium*. Eight colonies were then investigated.

a. Check PCR

First the integration of the EGFP was tested by colony PCR using primers specific to the *EGFP* sequence or close to the cloning site. Different combinations of primers targeting both the vector and the insert together were used to see which clones had their EGFP insert rightfully oriented.

The DNA templates needed for the colony PCRs were obtained either by boiling a bacterial suspension of each colony or by realizing vector purifications using GenElute© from Sigma-Aldrich©. The PCR protocol used was the same as the one used for *C5a* and *IL6* check PCR (see II.A.2.b.ii)

b. Restriction enzyme digestions

Failing to obtain conclusive results from the PCRs, some digestions were performed as well. The vectors purified from each eight colonies were digested using enzymes and protocols provided by Fermentas©. Both the FastDigest® system and the classic system were used depending on the enzyme available.

One digestion can cut the vector at several loci which gives several fragments. For our analysis of the restriction pattern it was necessary to distinguish the different fragment and restriction sites. To do so, an arbitrary nucleotide was chosen on the vector's sequence to be the origin on the vector's map or the "nucleotide n°1". The restriction sites for each cut were then named after their position: the name of each site was composed of the name of the restriction enzyme which cut it followed by a number representative from their position on the vector's sequence, the n°1 being the first restriction site of the enzyme in question met starting from the 1st nucleotide and moving along the 5'-3' orientation on the DNA sequence of the vector. (Example: The enzyme *Ban II* has 3 restriction sites on the pLH-TLRP-t-EGFP vector, they start from the nucleotides 937, 2336 and 2658, they are respectively referred to as *BanII(1)*, *BanII(2)* and *BanII(3)*). Based on the same concept, the bands on the restriction patterns were named as well according to the order of the different fragments on the vector before the digestion. (Example: When the band n°2 of the *BanII* digestion is mentioned, it is the fragment cut between *BanII(2)* and *BanII(3)* which is in question.)

The expected restriction patterns were predicted using Serial Cloner® or/and Clone Manager®.

III. Results

A] Evaluation of the C5a and IL6 production in *N. tabacum* using the TLRP system

1. Segregation pattern of the transgenes

a. PCR of the T1-generation plants

No antibiotic selection was done during the sowing of the T0 generation. For this reason the T1 generation, which sprout from it, contained some non-transgenic plants since T0 is hemizygous by definition. To identify which individuals were transgenic a PCR was performed on DNA samples extracted from the T1 plants. Both the transgenes of interest and the resistance gene to Kanamycin (*nptII*) were tested in order to confirm the obtained results.

It was determined by PCR that the individuals 16-1, 16-2, 16-3, 16-4, 16-5, 16-6, 16-7, 16-8 and 39-1, 39-3, 39-4, 39-5 were transgenic, and that 16-9, 16-10 and 39-2, 39-6, 39-7, 39-8, 39-9, 39-10 were not (Figure 4).

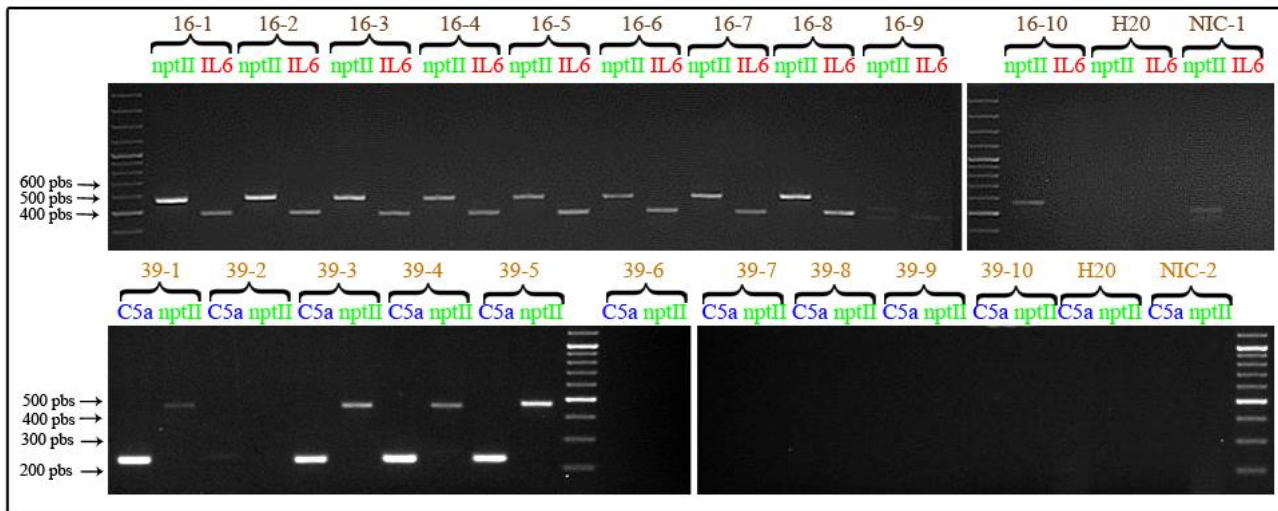


Figure 4: PCR testing the presence of the transgenes in the F1 generation

Gel electrophoresis revealed under U.V. light of the PCR performed on the genomic DNA of the different individuals in F1. The presence of a band allows the identification of the individuals containing the transgenes. Primers specific from nptII (resistance marker), IL6 and C5a were used. According to the sequences of the constructs the expected sizes of the fragments should be 469 pbs for the nptII fragment, 560 pbs for the IL6 band and 233 pbs for C5a. Above each lane (in brown) is the identification number of the tested individual: (T0 mother line n°)-(F1 individual n°), then on the second line (in green, red or blue) is written the gene targeted by the set of primer used. Water samples (H20) and non-transgenic plant DNA samples (NIC) were used as well as negative controls.

b. Segregation pattern in T2

Table 4: Results of the resistance tests performed on the T2 generation

This table displays the result of the resistance and germination tests performed using seeds of the T1 generation meaning T2 sprouts. Plants of both lines containing IL6 and C5a constructs were tested, but only the lines predetermined transgenic in III.A.1.a were tested. Non-transgenic plants (NIC) and plates without antibiotics (Kanamycin -) were used as controls. For each plant line tested the proportion of sprouts which seemed resistant is given as a percentage of the seeds which germinated (% resistant phenotype). On the plates with kanamycin (+), a proportion of more or less 75% resistant seed indicates that the T1 line used was hemizygous for the transgenes whereas all sprouts should be resistant if the T1 plant which gave the seeds was homozygous

Transgene (T0 line)	None (NIC)				C5a (39)								IL6 (16)															
Kanamycin	-		+		-				+				-								+							
Plant line (T1 seeds)	1	2	1	2	39-1	39-3	39-4	39-5	39-1	39-3	39-4	39-5	16-1	16-2	16-3	16-4	16-5	16-6	16-7	16-8	16-1	16-2	16-3	16-4	16-5	16-6	16-7	16-8
% Resistant phenotype	100	100	0	0	100	100	100	100	72	80	80	82	100	100	100	100	100	100	100	100	100	100	92	100	74	100	76	100

Once the T1 generation gave mature seeds, they were tested by an antibiotic resistance test in order to determine if the seed were coming from a homozygous or hemizygous plant (Table 4).

The proportions of resistant plant in T2-generation sprouts allowed us to identify the T1 plants 16-1, 16-2, 16-4, 16-6, 16-8 as homozygotes and the plants 39-1, 39-3, 39-4, 39-5 and 16-3, 16-5, 16-7 as hemizygotes (Table 4).

If you consider that the outcome of this experiment follows a binomial distribution $B(50;0.75)$, all data hereby obtained seem coherent except the result obtained for the line 16-3. Indeed, there should be about 75 % of resistant plantlets in the offspring of a hemizygous plant and the line 16-3 displays a proportion of resistant plantlets of about 92%. Considering the parameters of the experiment getting a proportion closer to 75% would be 99.7% more likely if the line 16-3 were hemizygous and the pattern of segregation Mendelian (See Appendix C) and impossible if 16-3 was homozygous.

a. Segregation pattern in T1

Table 5: Zigosity distribution pattern in T1 and comparison to the expected model

Information given by the outcome of the PCR in T1 (Figure 4) and the resistance test in T2 (Table 4) are resumed and summed up in this table to reveal the zygosity segregation pattern of the T0 offspring. T1 plants can then be classified in 3 categories: non transgenic or wild type (WT-homozygote), transgenic homozygote and transgenic hemizygote. Both the line 39 for C5a and 16 for IL6 are compared to the theoretical result (Expected) of a simple Mendelian pattern that could be expected. In each cell is displayed both the proportion of each category of individual and the actual count (in brackets)

	Zygosity	C5a: 39	Expected	IL6: 16
Percent (Number of individuals)	WT-Homozygote	60 (6)	25 (2.5)	20 (2)
	Transgene-Homozygous	0 (0)	25 (2.5)	50 (5)
	Transgene-Hemizygote	40 (4)	50 (5)	30 (3)

If you consider the Mendelian model of a purely autogamous reproduction the probability of getting a draw such as 6/0/4 (WT-homozygote/Transgenic Homozygote /Transgenic Hemizygote) or a less probable draw is below 3% (see *Appendix D*). A 3% value is below what most statistical test would accept, the significance level often being 5%. Then obtaining the draw of the line 39 is quite unlikely with the hypotheses made. On the other hand, the chance of obtaining a draw of 2/5/3 or something less likely is close to 20% which is a less surprising and then an acceptable draw considering the parameters and hypotheses made here (see *Appendix D*).

2. Preliminary experiments

a. *Total soluble protein content*

i. Total Soluble Content of the protein fractions

In previous experiments performed by Henrik Nausch, some recombinant proteins were recovered using the TLRP system but only in very small amounts (unpublished data). These experiments were performed on the cell wall of leaves of the T0 plants that gave the lines 39 and 16. It was then questioned if the TRLP anchoring worked efficiently. The other fractions in which the protein could have ended up were then investigated. Therefore the total soluble protein contents of the apoplastic, cytoplasmic and “ionic” fractions were investigated (Table 6). The decision was made to investigate the stem tissue as well since the TLRP binds to the secondary cell wall of xylem vessels (Domingo et al., 1999) which should be more abundant in stems than in leaves.

Table 6: Average TSP concentrations and yields

The data displayed in this table were acquired from three replicates: an untransformed plant (a NIC plant), a plant from the C5s line (39) and one from the Il6 line (16). The samples taken from those plants were submitted to a Bradford assay to measure TSP in both the stem and leaf samples (Tissue: Stem of Leaf) of the different protein fractions (Fraction: Apoplastic, cytoplasmic and ionic). Leaves and stems sections of each plant were sampled in equivalent amounts: about 12 cm² of leaf surface and 1.3 cm³ of stem tissue were respectively sampled. This table gives the average protein concentrations measured in each kind of sample (Approximate TSP concentration of the sample) along with errors based on both standard deviations and technical accuracy. The average total protein yields of each sample (TSP yield) is deduced from the measured concentration. For each fraction a different volume of buffer (Volume of Buffer used) was used depending on the extraction method, this is why a higher TSP concentration does not mean a higher yield. Finally a sum of the amounts of TSP found in each fraction (Sum) was calculated and used to assess which proportions of the total TSP content (% TSP) each fraction contains.

Tissue Fraction	Approximate TSP concentration of the sample (mg/mL)		Volume of Buffer used (mL)	TSP Yield (mg TSP/sample)		% TSP	
	Stem	Leaf		Stem	Leaf	Stem	Leaf
Apoplastic	0.48 ± 0.14	0.34 ± 0.26	0.3	0.14	0.10	17.4	16.1
Cytoplasmic	0.06 ± 0.03	0.04 ± 0.01	10	0.60	0.40	72.7	63.8
Ionic	0.03 ± 0.01	0.04 ± 0.01	3	0.08	0.13	10.0	20.1
Sum			13.3	0.83	0.63	100.0	100.0

It was not always possible to determine the total soluble protein (TSP) concentrations during the first measurement due to some technical issues. These issues led to measurements being taken a second and a third time. However after the third measurement the problem was solved, the data could not be validated as it seemed that the protein concentrations progressively decreased over the measurements (data not shown). Indeed, proteins can be degraded by proteases or by microbial contaminants, be broken down by ice crystals during the freezing cycles or simply physically degraded over time. Consequently these preliminary results were acquired with a low accuracy. They are nevertheless displayed since they helped to explain some experimental results and set new protocols (Table 6).

Given the protocol and calculations applied, the data displayed in the Table 6 are naturally not representative of a biological context but of the concentrations in our samples and the yields of our extraction technique. However, this table reveals that most proteins are in the cytoplasm as expected. Nevertheless, a non-negligible amount of soluble proteins that need to be investigated is also present in the apoplast and the cell wall (since the “ionic” fraction contains the proteins bound to the cell wall by intermolecular forces).

ii. Complementary data about the protein sample's stability

Protein samples are especially unstable when their concentrations are below 1 mg/mL as they will degrade rapidly in such conditions (Wang et al., 1999). From the data acquired it is obvious that the samples used in Table 6 have too low TSP concentrations to be properly stored. Considering that only the extraction protocol takes more than a day (only the thrombin digestion takes 16h at RT) and that the samples were on hold to be subjected several times to the Bradford assay, a better extraction or preservation method should be considered as the samples are likely to have already been subjected to degradation at the time of the measurement.

It is not advisable to keep fresh protein samples more than a day at 4°C. Nevertheless, the pH of some protein samples was tested at some point and a decreased of the pH values was measured after few days of storage at 4°C: the pH of the samples went from a 6-8 to a 5-7 range (the pH was dependent on the type of sample though the extraction solution was buffered at a pH of 7,4 - data not shown). This decrease in pH suggests a microorganism proliferation. Freezing was not the best option either since freezing cycles affect the quality of the samples as well. This is mainly due to the formation of ice crystals and the denaturation that happens during the freezing.

The compartment which gave the most concentrated samples and then likely the more stable ones was the apoplast. It is not because it was the compartment with the highest TSP content but because very small amount of buffer were used to extract it (only what could be vacuum infiltrated) whereas a lot of buffer was necessary to homogenize properly the tissues, diluting a lot the cytoplasmic fraction. This fact inspired to use less buffer for the protein extractions during the following experiments. Having more concentrated samples should not only help to preserve the samples but also to detect smaller amounts of protein using the same techniques.

b. ELISA C5a

Some tissues sampled on the T1 generation of the line 39 were extracted using the fractioning method and each fraction was analyzed with the C5a-ELISA. The dilutions of the standard permitted constructing a calibration curve but no signal was observed for any of the sample (data not shown). The reasons that no C5a was detected could be (i) the C5a was not expressed highly enough to be

detected, (ii) the plant tested did not express C5a at all, (iii) the thrombin treatment did not work or (iv) the C5a was already degraded by the time the analysis was made.

When the IL6 ELISA test was performed on the plants of the line 16 (II.A.3.a), some of the fractions gave some weak signals. Besides IL6 gave higher yields *in planta* in previous studies (Nausch et al., 2012b, 2012c). The higher yields of the IL6 plants were thought to be the result of the fact that IL6 was a more stable protein in this configuration (Nausch et al., 2012b, 2012c and Inge Broer, personal communication). The IL6-ELISA is also more precise than the C5a-ELISA which allowed detecting lower amounts of recombinant proteins. It was then decided to keep working with the IL6 plants.

3. Recovering the IL6 content from the plant material

a. *A first attempt to localize the IL6 in planta*

The protein fraction separation protocol was used to obtain different samples from the IL6 plants from the line 16 and compare them. A total extraction protocol was performed as well. ELISA was used to measure the quantity of IL6 in each sample.

Table 7: IL6 yields primarily detected in the protein fractions

The samples used to acquire the data presented in this table came half from 16-1 and half from 16-5: both tissues were extracted at the same time using 500mg of 16-1 tissue with 500mg of 16-5 tissue pooled in a single sample. Stem and leaf tissues were taken and extracted following two protocols, one separating the different protein fractions ('apoplastic', 'cytoplasmic and thrombin' and 'ionic') and another extracting all the proteins from the tissue at once (total). The IL6 concentrations were then measured in the different samples using ELISA. Finally the concentrations found were converted into yields of IL6 per mg of fresh weight (Tissue Yield).

Fraction \ Tissue	Tissue Yield (pg/mg)		Volume of Buffer used (mL)
	Stem	Leaf	
Apoplastic	0.031	0.020	0.2
Ionic	0	0	2
Cytoplasmic and Thrombin	0.002	0.006	2
Total	0.042	0.020	2

Overall the concentrations of IL6 found in the samples are low (Table 7). We cannot compare precisely the fractions studied in the Tables 6 and 7, however considering the amounts of buffer used for each fraction in both experiments and the TSP contents measured in the Table 7, it is possible to evaluate more or less the TSP concentrations' magnitude of the samples in the Table 7 and it should be between 0.04 and 0.75 mg/mL. Considering this rough method of comparison, the total and apoplastic samples should be the ones with the highest TSP concentrations and it can be noticed that they also are the samples giving the highest IL6 yields. This could be explained by the fact that the apoplastic and total fractions simply contain more IL6 than the other fractions but it could also be a consequence to the fact that the lower the TSP concentration is the faster the protein samples degrade. Even though

ANALYSIS OF A TYROSINE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS

less buffer was used in the experiment of the Table 7, it may still not be enough to keep the samples from degradation. It was then decided to use procellys® for homogenization which allowed reducing even more the amounts of buffer used and some albumin (BSA) was sometimes added to the samples as a carrier to increase their TSP concentrations to an acceptable level.

Knowing that the values displayed in Table 7 may have been influenced by degradation and that it could have occurred at different rates for the different samples, those values cannot be entirely trusted. However, they still offer useful information about the different samples, even if exact values are not part of it. Indeed it can be noticed that some IL6 is indeed found in these plants, that most of it is located in the apoplast and that the IL6 does not seem to bind to the cell wall via intermolecular forces since no IL6 was detected in the “ionic” fraction.

b. Analysis of the different IL6 individuals and tissues

Even though it was expected to obtain higher yields using an improved protocol, the expected readings would stay quite low. Hence, it became important to determine which tissues and lines were giving the best yields in order to facilitate the detection of IL6 and the analysis of the samples. The eight transgenic T1 plants of the line 16 were then sampled to be submitted to total protein extractions and an ELISA in which young and mature tissues of both stems and leaves were tested (Figure 5).

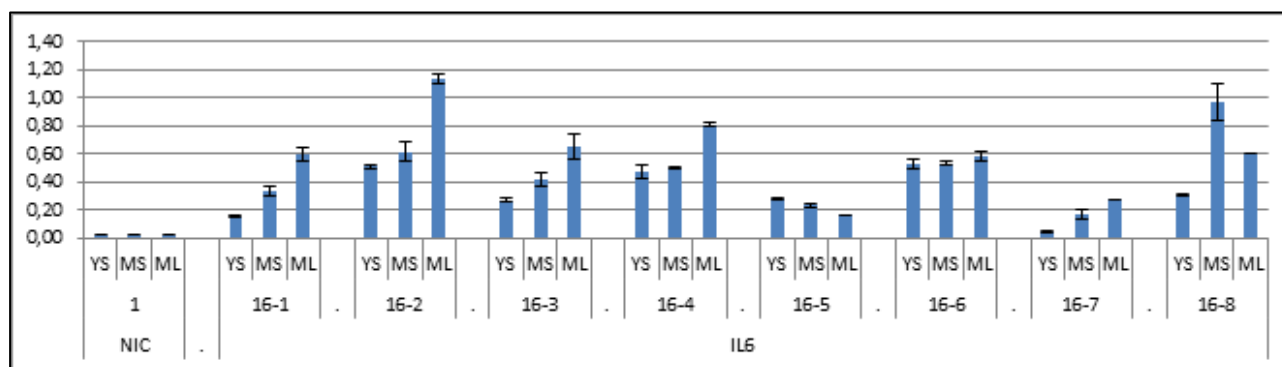


Figure 5: Total IL6 Yields of the line 16: full set of data

Yields in this figure are expressed in pg of IL6 per mg of tissue sampled (same as ng/g). Each sample is characterized by the plant it is taken from (16-1, 16-2 ...) and the tissue it was sampled from (YS=Young Stem, MS=Mature Stem, ML=Mature Leaf). A non-transgenic plant (NIC-1) was used as a control. (See *Appendix E* for the exact values)

Using BSA with the new batch of thrombin, much greater amounts of IL6 were obtained. The young leaf tissues gave the highest IL6 yields (data not shown), their lowest yield being 3.9 pg/mg for the 16-5 tobacco plant. The concentrations of IL6 in the samples were sometimes so high that it was not possible to determine the yields of some of the young leaf samples such as 16-2 or 16-4 because they were above the range of detection of the test, and it was not possible to remeasure it after dilution due to some technical issues that happened with the storing system (this is why they are not displayed in the Figure 5). However, based on the range of validity of the ELISA and the information about the biological material used, some yields up to 8 pg/mg could still have been measured; some of the young leaf samples had then an IL6 yield higher than 8 pg per mg of fresh weight.

The average yield of each tissue was calculated (Figure 6) to determine the best tissue to sample, YL as discussed hereinabove gave significantly higher yields than the other tissues which technically have a too high variability to be significantly ranked. However it seems that the yields of

ANALYSIS OF A TYROSINE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS

ML are slightly higher than the ones of MS which are themselves better than the ones of YS. Otherwise it was noticed that the homozygote individuals had higher yields than the hemizygous ones, to check this theory the averages of the yields of homozygous and hemizygous were calculated to be compared (Figure 7). However the different tissues were not pooled together as it did not make sense to pool together samples that were that different in our analysis. Moreover separating both the three kinds of tissue and the two zigosities allowed reducing the errors (based on standard deviations) which supports the theory that those kinds of sample could contain significantly different amounts of IL6. Even making those adjustments the errors are still very high and the differences of concentrations are consequently not “significant”, even though the different kinds of sample constantly give IL6 yields at least two times higher for the samples from homozygous plants than the ones from hemizygous plants (Figure 7).

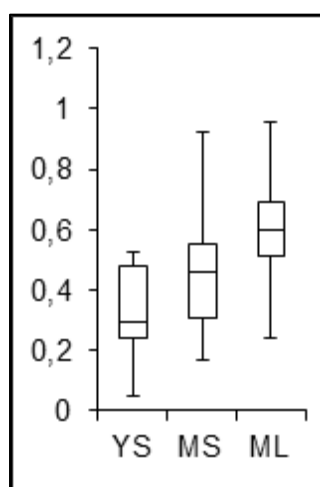


Figure 7: Comparison of the IL6 yields of the different tissues sampled

The data displayed in figure 5 were used to make this box plot. Yields are still expressed in mg/pg or ng/g.

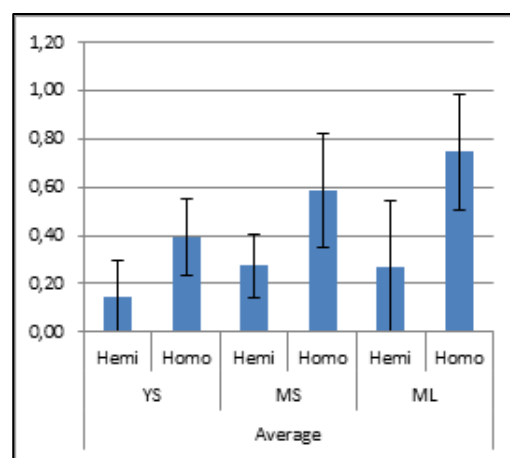


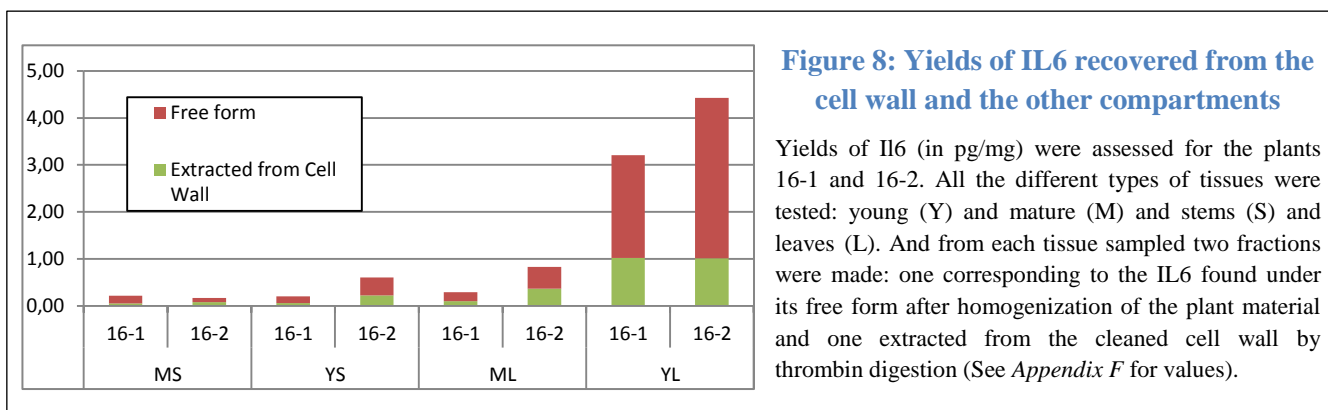
Figure 6: Comparison of homozygote and hemizygote plants' yields

For each tissue (YS, MS and ML), the average yield (in mg/pg) was calculated splitting homozygote (homo) and hemizygote (hemi) plants (identified in II.A.1.b)

c. The IL6 bound to the cell wall

Given that much higher IL6 yields were detected using the new fresh batch of thrombin that contains BSA as a carrier protein, another attempt to recover some IL6 from the cell wall was performed. This was done by repeating the protein extraction and measuring the IL6 concentrations but in only two fractions this time (Figure 8). The first fraction corresponds to the proteins not-bound to the cell wall (obtained by simple homogenization in a “high salt” buffer), meaning the IL6 in its free form (virtually the sum of apoplastic, ionic and cytoplasmic proteins) and the other fraction analyzed corresponds to the protein extracted from the cell wall by the thrombin digestion (which correspond to the “thrombin” fraction in the previous experiments). The latter is particularly interesting, it is the one isolated from the washed cell wall; it theoretically only contains the IL6, the thrombin and the BSA (though the purity was not tested by comparing IL6 amounts to TSP contents because when the BSA was added, it was added in amounts that hid the original TSP contents from the biological samples, the Bradford assay could then not be used in a relevant way).

ANALYSIS OF A TYROSINE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS



The proportion of IL6 bound to the cell wall vary for each sample analyzed but the proportions of IL6 recovered from the cell wall are all ranging between 20 and 50% (See *appendix G* to read the calculated proportions). It means that most of what is extracted is still found in its free form. It is not the best configuration as the more IL6 binds to the cell wall and is recovered from it, the more the system can produce purified IL6. Nevertheless a proportion of IL6 can be purified from the cell wall making this experiment a “proof of concept”.

B] Visualizing the expression of TLRP in plant

1. Failure to identify a valid pLH-TLRP-t-EGFP

A new vector using the TLRP system to express the EGFP was built. Once the cloning of the EGFP sequence in the pLH-TLRP-t vector was done, several PCRs and digestion were performed on the pLH-LeTLRP-t-EGFP candidates to check the success of the experiment. The outcomes of these control tests were not fitting the expected results: although it was detected that part of the *EGFP* sequence was integrated in the vector using primers internal to its sequence, it was not always possible to obtain a DNA fragment of the expected size or any PCR product at all using other primers. The digestions by restriction enzymes did not work as expected either. For instance, the digestions performed with the *NruI* enzyme never gave anything but a single linear fragment whereas *NruI* was the enzyme used to digest the vector and the insert at both 3' and 5' ends. Similar unexpected results were obtained testing a variety of primers and restriction enzymes.

During those PCRs and digestions the vectors purified from eight different clones were tested and all eight clones selected for testing gave identical patterns when submitted to the same tests, suggesting that all of them were impaired with the same problem.

The original vector used in this experiment (pLH-LeTLRP-t-IL6), which was used as the backbone that would receive the EGFP gene was then checked to see if its integrity could be the root of the issue. Some substitute vectors, such as pLH-LeTLRP-Thrombin and pLH-LeTLRP-t-C5a, were then checked as well to replace it in the cloning protocol in case the problem would originate from pLH-TLRP-IL6. But in the end, none of the original vectors was altered as every test performed on them gave the expected results, even those which did not work properly with pLH-LeTLRP-t-EGFP (See *Appendix H* and *L*).

ANALYSIS OF A TYROSINE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS

All the results obtained during the PCRs and vector digestions (discussed above) suggests that the EGFP sequence was partly integrated but that during the cloning deletion of a part of the vector happened after the EGFP was integrated.

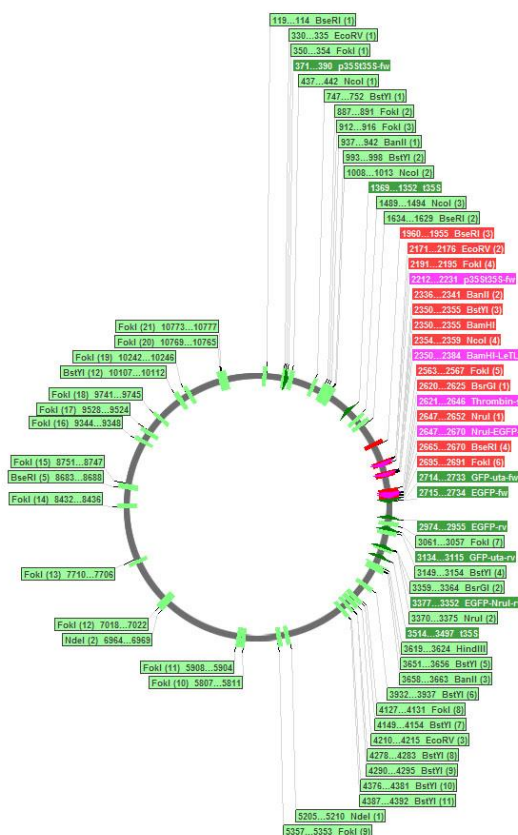
All the experiments performed did not allow finding a proper clone. Nevertheless, it is looking back at those results which allowed us to make the hypothesis that a part of the pLH-LeTLRP-t-EGFP's expected sequence was missing and to assess which part was missing. This hypothesis of a missing region was confronted back to the outcomes of the digestions and PCR to identify the missing part (see *Appendix J* for an example of analysis of a restriction pattern). All the information gathered from the digestions and the PCRs were pooled together (*Appendix K* and *L*), and used in order to map the missing region (Figure 9 and *Appendix M*).

2. Localization of the deleted region

The Figure 9 makes it possible to restrict the beginning of the missing area to a region that starts between the pairs of base 1629 and 1960 and ends between the pairs of bases 2691 and 2714. Consequently the missing region includes most of the promoter of the chimeric construct and goes up to the beginning of the EGFP sequence, including the LeTLRP anchor's sequence (Figure 3). This suggests that the recurrent issue of the deletion was triggered by putting site to side the sequence of the *EGFP* gene with the ones of the *TLRP* and the promoter specially designed by Henrik Nausch to enhance the protein transcription (see I.C.).

Figure 9: Mapping of the missing sequence of pLH-LeTLRP-t-EGFP

This figure represent the map of the vector pLH-LeTLRP-t-EGFP on which has been located and annotated all the restriction sites and primers studied. The color of the labelled areas indicates if this part of the vector is believed to be missing or not. When the labelled area is in green the experiments confirmed that the part of the vector was not deleted and if it is red or pink it means that the area is suspected to be missing. An enlarged and detailed version of this map is available in *Appendix M*.



IV. Discussion

A] Evaluation of the C5a and IL6 productions in *N. tabacum* using the TLRP system

1. Identification and characterization of the transgenic plants and lines

a. Segregation of the characters of the line IL6-16 in T2

Although an experiment testing 100 seeds per line would have secured the obtained result and could avoid unconvincing data, using only 50 seeds seemed reliable enough here for the purpose of assessing which plants were hemizygous. Moreover, all data analysed in T2 are coherent with the emitted theories, except the 92% value which is a bit surprising (Table 4).

Investigating what could be behind this unexpected value, three hypotheses have been raised:

- First of all, the “5% rule” used to decide if a result is coherent or not is a statistical principle, no possibility is excluded: though it is not very likely to obtain that draw, it does not make it impossible.
- Second, it has been questioned whether cross-pollination could have happened or not. If that were to happen it would be complex to determine in which proportions the cross pollination would affect the outcome of the resistance test but a hemizygous plant partially crossed with a homozygous plant could explain a percentage of resistant offspring higher than 75%. In the end, cross-contamination is unlikely because it is very scarce in tobacco and good precautions were taken in the greenhouse, such as bagging the inflorescence before the blooming.
- Third, the plant lines used in this study have been recently genetically modified. Plants which have been transformed are known to have potentially unstable genetic structures during the first generations. The genomes are especially mangled during the karyogamy and consequently this instability affects the transmission of the transgenes during the first generations. When that is the case, the segregation pattern does not follow the Mendelian rules of repartition anymore (Tizaoui & Kchouk, 2012). It is possible that the 16-3 did not reach stability yet, even though most of the other lines seem to have a Mendelian segregation and then reached stability at the T2 generation. However, just because the other lines gave a result compatible with a Mendelian repartition, it does not mean they necessarily follow it.

Applying Occam’s razor principle, the 92% is more likely a random draw which happened though it was unexpected and it might have been avoided using a larger number of replicates.

b. Transgene segregation in T1

The same reasoning used to analyze the results from the T2 generation can be applied to the T1 generation. However in this configuration the instability of the transgene seems a better

explanation to the observed segregation patterns. Indeed the T1 generation should be the first affected by a genetic disruption given that it is the first filial generation after the transgenesis. The T1 is then the result of the first karyogamy which is usually a key moment for the genomic rearrangements (Nacry et al., 1998). Moreover there are no other draws to support the Mendelian segregation pattern (though there are no other draws that undermine it either).

It could be that some structures are particularly unstable and therefore are not carried to the next generation as frequently as some other. For instance the hemizygous plants, presenting a genetic asymmetry could be disadvantaged. Thus it can be noticed that the amount of hemizygous is inferior to the number of homozygous for both lines 16 and 39 whereas theoretically there should be about equal amounts of each. The integration locus could be important as well (Tizaoui, & Kchouk, 2012). For instance we can notice that no homozygote transgenic C5a plants were found. The locus of the insertion of the transgene in the line 39 might then be right on a gene indispensable for the plant survival, reproduction or development. If so, one functional copy of the gene could be enough to ensure its role but one copy would still be necessary. That would explain why no transgenic homozygotes were found.

Given that 10 seeds per line were used here, the statistical worth of this experiment is easily put in doubt. Consequently the interpretations made here are to be considered carefully because the data obtained do not allow proving what is assessed.

c. Conclusion

The aim of this experiment was not to test those segregation hypotheses but to identify good candidates to manipulate and to explain some experimental results and this objective was fulfilled. Therefore the numbers of replicates are way below what would be necessary to draw real statistical conclusions. More replicates would be necessary to do so. But the outcome of this experiment might not be indispensable for our study. Moreover it would take over a year and a considerable amount of space to grow enough plants from the T1 generation to test them.

However it was interesting to analyze the spotted anomalies to get to know the genetic characteristics of the manipulated plant material and to make sure there was no problem with the studied lines.

Finding which individuals were homozygotes proved to be important to analyze the IL6 production and the absence of C5a homozygotes indicates that we should investigate more T1 individuals of the line 39 or use a different line derived from a different event. Indeed if homozygous C5a plants were producing more recombinant proteins the same way the IL6 plants do (Figure 7), it might be easier to detect some C5a production in those plants.

2. Stability and handling of the protein samples

The total amount of protein yielded from the sample was sometimes less than a milligram (Table 6), which is quite low and could make the samples unstable (Wang et al., 1999). Moreover it was observed that the sample could be affected by different parameters such as microbial proliferation, freezing cycles and physico-chemical degradation. It is important to stabilize the samples in order to obtain data which are more accurately representative of the system and obtain better yields from this technique. Moreover obtaining better yields could ease the analysis of the sample by providing higher amounts of recombinant proteins that would be easier to detect and quantify.

a. The issue of preserving the samples from degradation

The information taken from observations on the behavior of the protein samples in different conditions in all the experiments in this study (data not shown) highlighted that to be accurate the protein samples' tests such as Bradford assay and ELISA should be performed as fast as possible. Indeed, neither storing at 4°C nor -20°C seems to have ensured the complete stability of the samples. The problem is that it is not always possible to analyze the samples very fast as several manipulations need to be completed depending on the extraction and analysis protocols that were used. Moreover most manipulations are done at RT. Even though it should be possible to realize all the manipulations in sterile conditions and at low temperature, it is technically complicated and it might not even be sufficient to preserve the samples from degradation.

Stabilizing the samples is also important for allowing comparison with other samples. All the samples were different since they were extracted from different sources, at a different time point and with different methods. Consequently the different samples could degrade at different rates since they were made of proteins of different natures which were kept in different environmental settings. Then, the different samples could have different stabilities and different half-lives whether that may apply for different samples of the same experiment or the same samples in different experiments. Nevertheless, taking better measures to preserve the samples should help minimize those differences.

b. Formulation of the extraction buffers

Improving the extraction buffer formulation is one measure that could efficiently improve the stability of the samples and it is easier to put into practice than changing the protocols.

For instance the amount of salt used in the extraction protocols has been brought back to a normal level (1 mM) given that the IL6 did not seem to have intermolecular interactions with the cell wall, at least using a 'high salt buffer' did not help recovering more IL6 (Table 7 and data not shown). It was even harmful to the analytical process (data not shown).

i. The use of Albumin

- The advantages of using albumin*

Another efficient improvement of the buffer composition that was done during this study was the use of BSA as an additive to stabilize the protein samples with a low TSP content:

Because the samples had very low TSP concentrations (Table 6) and IL6 is very unstable, the samples were suspected of having degraded over time (III.A.2.a.ii). To avoid these instability issues, it is advisable to add at least 2mg/mL of BSA as a carrier when the TSP concentrations of the sample were below 1-2 mg/mL (Wang et al., 1999). Applying these recommendations seems efficient as the magnitude of the reported IL6 yields has been more or less multiplied by 10 in between the Table 7 and the Figures 5 and 8.

From the moment the BSA was added, it was then possible to analyze the samples which used to have too low TSP concentrations to be properly detected (Table 6, Figure 5 and 8). And this is very important because as seen the Table 6, the concentrations of IL6 in the samples do not depend as much on the original quantities or concentrations of IL6 in vivo as they depend on the extraction technique and the volumes of buffer used.

BSA is especially necessary to realize the extraction using the thrombin which gives the fraction called “thrombin” or “bound to the cell wall” in this study. Indeed the aim of the study is to set up a system that would provide some recombinant protein (IL6) that would not need to be purified much further. For this reason, during the “thrombin” extraction, the plant material is homogenized, washed from all solutes and then resuspended in the extraction buffer where there are no more proteins in solution. At this point, any protein added in low amounts to the solution would become very unstable. Without the BSA, the thrombin would then degrade, probably too fast for the digestion to be fulfilled. And even if some IL6 was still released it would, anyway, quickly degrade too. This is likely what has been observed during the preliminary experimentation (Table 6).

Other studies proved that it is possible to make the IL6 very stable under some conditions. For instance IL6 is particularly stable in blood plasma even when it is stored at temperatures as high as 30°C (Kenis et al., 2002). Then it should be possible to obtain stable IL6 in our samples too. It could be achieved by somehow transferring some properties of the plasma to our extraction buffer even though a plant extract and the blood plasma are obviously very different media and it might be complicated.

In conclusion, BSA, as a protein carrier, has proven its usefulness in our protocol: the use of BSA probably stabilized the samples which had low protein concentrations. In addition it should have preserved the activity of the thrombin as well.

- *The inconvenient of using BSA*

Theoretically, once the thrombin treatment performed, the only proteins present in the extract would be the recombinant proteins (IL6 or C5a), the thrombin and the albumin. The addition of the albumin could then be considered to affect the purity of the sample. But the albumin is already used as a neutral excipient in some pharmaceutical formulations since it is the natural carrier protein of the human blood. So in most cases it could even not be useful to remove it. However, even if the albumin had to be removed, it is likely that it would be easy to separate it from the recombinant proteins even in native conditions because they have distinctively different properties (size for instance: C5a is about 8.6 kDa and IL6 is 23.3kDa whereas albumin is about 66 kDa). Using albumin in this system should therefore not be a problem considering that a simple polishing purification would be sufficient to concentrate the IL6 and eliminate the albumin.

The use of BSA in excess had a downside however. In the original configuration, it was planned to use the Bradford assay on the different samples to measure the amounts of TSP extracted and determine if the higher amounts of IL6 found in some of the samples would be the result of a globally more intense protein synthesis or if the recombinant proteins were produced or retained in different proportions in the different tissues. But the concentrations of BSA used have flattened the TSP concentrations' differences in between the samples (data not shown). The TSP amounts originating from the plant material were so low compared to the amount of BSA added that it was sometimes not possible to tell whether the BSA was added to a sample with a low TSP concentration or to a sample without protein at all from the Bradford assay point of view. Using BSA more parsimoniously in the future and carefully monitoring how much of it is added to each sample might help to track back the original TSP concentrations to the extent that the samples do not have too low TSP concentrations. But other protein measurement protocols should be explored as well in order to get better results as measuring the TSP could be important for comparing the different samples and assessing their purity. HPLC could be a good solution for measuring the amounts of IL6, BSA and albumin but this technique requires much more material and means than simple Bradford and ELISA

tests. Using western blots to obtain the same answers should be possible as well, though it would require a lot of manipulations and it might give less precise quantification.

ii. Protein buffer additives

Even though some progress has been made in this study the best parameters for handling and storing the samples are yet to be determined

Some additives could prove useful for solving the different issues mentioned about preserving the protein samples. Several types of additives could be used: protease inhibitors and antioxidants to prevent chemical and physical degradation, preservatives to prevent the development of microorganisms, emulsifiers to prevent aggregation, detergents to increase yields, cryoprotectants (such as trehalose or glycerol) to preserve the proteins during the storage below 0°C, etc. However those additives should be tested first, as they might have negative effects on the samples. For instance, if protease inhibitors are added, these inhibitors should be selected so as not to affect the activity of the thrombin. There are different classes of protease and the inhibitors are often specific to them but protease inhibitors are often commercially provided as mixes of different active principles that cover a wide range of proteases. Nevertheless it should be possible to find a mix or make one that is not effective against the thrombin, even if it would mean diminishing the efficiency of the protease inhibitor mix.

3. The production of IL6

a. *An important intra- and inter- individual variability*

There is a lot of variability between the different lines tested. This high variability in itself is not an asset here since it makes it more complicated to analyze the samples and draw conclusion about which plants and tissues are the best producers and which are the worst.

i. The IL6 yield variability due to the experimental conditions

A high variability of the IL6 yields is noticeable between the different plants but it is very likely that the variability would stay high even within the samples of a same origin as well. Indeed, if we compare the yields of the lines 16-1 and 16-2 in different experiments (Figure 5 and 8, even though they are not exactly the same experiment), we can see that the total yield reads vary from 20% to 70% for the same kinds of tissue sample.

There are many parameters that could influence the amounts of IL6 detected. Not only the genetic background and its properties could be factors acting on the IL6 production, the physiological stage of the tissue, the concentration of the sample, the analytic errors, the growth conditions, etc. could matter as well. The observed variability is therefore likely due to the technical errors accumulated all along the analysis protocol but it is also linked to the sampling of the plant material. Indeed the production of all and every protein can vary from one physiological stage to another and from one growing condition to another.

Though the tobacco plants were grown in very similar conditions and the tissues were sampled in conditions that were aimed to be close to each other, it was not technically possible to sample exactly the same organs at the exact same age every time and for every plant. This is especially true in this study because the studied organisms were fully grown tobacco plants with different architectures, sometimes sampled at different times of the year for the different experiments and which were

sometimes grown out of cuttings. Making several measurements should however help to assess the magnitude of this variability and allow us to compare the average yields. But we did not have enough time in this study to compile enough data to do it. For this reason, using only the measurements made so far, it does not seem reliable to select some individuals over others since most of the different kinds of sample have been measured once only and most plant lines do not display differences that are high enough to be fully significant. If a selection must be done with a great level of confidence, it can only be done making more measurements.

ii. The IL6 yield variability between the different lines

Even though there is a high variability due to the experimental conditions, some differences are still obvious and allow us to draw preliminary assumptions.

For instance, when you compare the yields of the homozygous and the hemizygous plants (Figure 7), there is an obvious difference; homozygous plants display IL6 yields about 2 to 2.5 times higher than the hemizygous ones. It might be confusing looking at the set of data (Figure 5) because the yields of the best hemizygous producers meet the yields of the worst homozygous producers but this is just an artefact of the sample's high variability. Nevertheless this result seems logical: the doubled production could be a direct consequence of a 2 times higher transcription rate for the homozygotes since they have two copies of the transgene in their genome when hemizygotes have only one. In the end using homozygotes for the coming experiments would be an advantage as they give better yields and will always give homozygous offspring which simplify the analysis and characterization of the samples.

Another observation to be made is that the plant 16-1 seems to display a yield consistently lower than the other homozygote plants (Figure 5 and 8). That situation is not surprising; in fact it is usual that different individuals of a same transgenic line produce different amounts of recombinant proteins (Jana Huckauf, personal communication, Department of Agrobiotechnology, Agricultural and Environmental Faculty, University of Rostock.). This variability can even be used to "conventionally" breed the transgenic plants. That way the recombinant protein yields are sometimes increased up to 70 times within the six first generations of a transgenic line (Hood, Woodard & Horn, 2002), even when the transgenic lines are characterized by a single locus event (Nausch et al., 2012b and 2012c). The acquired data suggest that such variability exist between the different IL6 plants of the line 16. Selecting the best IL6 producers over the generation could then increase the yields over the successive generations. From our set of data the plant 16-1 could be discarded from the study but the other homozygous plants do not display differences that are high enough to be discriminated without further tests.

iii. The link between TSP and IL6 yields

It was possible to evaluate with a low accuracy the TSP content of some samples and it was noticeable that it was quite variable similarly to the IL6 yields (data not shown). It is then possible that the IL6 yield variations depend on the protein metabolism of the sampled tissues. Indeed, considering all the other parameters identical, the higher the TSP content is, the more IL6 should be yielded. Being able to quantify the TSP yields with good accuracy and for all the samples would have been useful to see if that was the case. If it was the case a part of the variability issue could have been solved by comparing the IL6 yields as a proportion of the TSP (Nausch et al. 2012b and 2012c).

The reasons why two samples would have different TSP yields are multiple: the extraction protocol, the handling of the samples, the nature of the tissue sampled, the amount of tissue sampled,

the inter- and intra- individual variability ... Measuring the TSP content is then important to be able to compare different samples between each other. For instance knowing the TSP contents from two tissue samples presenting a different concentration of a specific protein (such as IL6 or C5a) would indicate whether the observed difference is only due to a difference of TSP concentrations or to the fact that one of the tissues sampled expresses or accumulates the protein in question in greater proportions. Often both facts are to be considered at the same time and are useful information to improve and understand a production system. This is why, though some answers about the samples were obtained, it is difficult to draw conclusions about the TLRP system in this study missing this information. However it should be possible to perform Bradford assays (Table 6) to obtain the TSP yields of new samples if the experiments were repeated, but in this repetition the TSP concentrations of the samples should be tested before adding the thrombin and the BSA to the sample. Moreover from the measurements made in this study (data not shown) we can evaluate the magnitude of the TSP yields to expect which should ease the analysis: leaf samples should have TSP yields that approximate 10 mg/g whereas TSP yields of stem samples should be lower than 1 mg/g. (data not shown).

The degradation of the samples may be the reason why the TSP yields were lower than what was expected in the first experiments (Table 6) but the protein metabolism could also partly explain the obtained results considering that Szedljak., Szántai-Köhegyi & Kosáry (2010) showed that the TSP content in mature leaves can vary from 1 mg/g to 7 mg/g within few weeks. This fact adds one more factor to consider when analyzing the data in this study; it means that there is one more parameter which brings variability when comparing two samples in two experiments in which the tissues have been sampled at different times.

Previous studies have given variable outcome to what the protein yields of tobacco leaves should be. In some studies the yield values range between 1 and 8 mg/g of fresh weight (Szedljak., Szántai-Köhegyi & Kosáry, 2010), for others it varies between 10 and 20 mg/g (Kung et al., 1980) and some values even suggest yields of to 80 mg/g (Nausch et al, 2012a). All those values depend on the extraction and analysis methods of course but the cultivar and the age of the plant are also factors of variation (Szedljak., Szántai-Köhegyi & Kosáry, 2010). There are then many other parameters that would likely allow us to improve our TSP yielding. It seems possible to obtain higher total protein yields from leaves either by improving the extraction protocol or by mastering some other parameters such as the optimum harvesting time for instance (Szedljak., Szántai-Köhegyi & Kosáry, 2010). Finding the extraction, preservation and production methods that would give the best TSP yields could theoretically also help produce more recombinant proteins via the TLRP system.

b. The IL6 production in the different tissues

Determining which tissues were the best producers might be especially important to be able to facilitate the use of techniques with low resolution such as immuno-histochemical localization or in situ fluorescence that were considered to be used for further study of the system. Despite all the parameters coming into play it is possible to notice some difference between the different tissues and hypothesize some trends. For instance the most obvious observation is that young leaves are the best producers of all but other differences are noticeable as well.

i. Leaf and stem tissues

Leaves give higher IL6 yields than stems (Figures 5,6,7,8 and 10); ML produces in average about 30% more IL6 than MS and 90% more IL6 than YS (Figure 5); YL would provide at least 15 times more IL6 than both YS and MS.

Although it was not possible to evaluate exactly the TSP yields of stems and leaves, it was possible to evaluate that stem TSP yields are below 1mg/g whereas leaf yields were close to 10 mg/g for the sample displayed in Figure 5. Given the lack of reliable values for the yields of YL no further comments can be made on this matter but for ML it would mean that, even though, relative to the fresh weight ML is a better producer, relative to the TSP content, YS and MS produce greater proportions of IL6 than ML.

Be that as it may the high protein production of the leaf tissues counters the advantages that a stem tissue could present. As tobacco is cultivated for its leaf production and the leaves are easier to homogenize than the stems, this situation is not detrimental.

ii. Young and mature tissues

It was possible to see the effects of the developmental stage on the IL6 yields as well: young leaves have significantly higher IL6 yields than mature ones; they are at least 10 times higher. For the stems it is the opposite but the difference is not as great (Figure 5) since YS produces about 30% less IL6 than MS.

Because the Bradford assay was compromised by the use of BSA it is not possible to say if YS and MS have different TSP yields. However since leaf tissues had much higher TSP yields it was possible to approximate the TSP yields of those tissues. That way the TSP yield of YL was estimated close to 12mg/g whereas the TSP yield of ML was estimated near 10mg/g. Although young leaves produce more TSP than mature ones, this cannot explain why YL contains at least 10 times more IL6 than the ML since young leaves accumulate only 20% more total protein than the mature leaves. These data prove that the differences of IL6 yields observed are beyond a simple total protein metabolism matter.

iii. The IL6 yield is dependent on the 35S CaMV promoter expression

Though the 35S promoter of CaMV is a massively used promoter for its high expression levels and its allegedly constitutive expression, Hraška, Rakouský, & Čurn (2008) reported tissue specific levels of expression. Their study reveals that as in this study the expression of the transgene under the control of the 35S promoter of CaMV is stronger in the young leaves than in the mature ones but rather stable along the stem. This explains why more IL6 is found in the YL than in the ML.

The 35S CaMV promoter is also especially active in the vascular tissues, which is logic considering that CaMV is a virus; it targets vascular bundles in order to spread in the plant. The high amount of vascular tissues in the stems could explain why, proportionally to the TSP content, they produce more IL6.

iv. Conclusion

Analyzing the different samples it was possible to understand that the different yields of IL6 in the different tissues were consequences of both the protein metabolism and the activity of the 35S promoter of CaMV which are both tissue specific. Both the organ specific total protein yields and the promoter's activity are factors that are independent of the TLRP system, some other parameters directly linked to the properties of this system may take a part in the explanation of the different yields observed but it is difficult to evaluate if that is the case because of the incomplete data. Nevertheless most of it seems interpretable through explanations that are not directly linked to the use of the TLRP.

4. TLRP properties

a. *An overall low productivity*

However using the TLRP anchor could be an advantage in the purification and processing steps, the IL6 yields of the plants used in this study (LeTLRP-IL6) are overall low. Plants with a simple addressing of the non-chimeric IL6 to the endoplasmic reticulum (ER-IL6) displayed yields about 50 000 times higher than the LeTLRP-IL6 plants. Apparently storing IL6 in the ER (which contains few proteases and some chaperon proteins) seems to stabilize the recombinant protein *in planta*, increasing the yields (Nausch et al., 2012b). For this reason ER-IL6 plants would likely be preferred as there is little chance that the TLRP-system compensate the disadvantage of a 50 000 time lower yield. At some point it was thought that the TLRP peptide could protect the IL6 by binding it to the cell wall and that for this reason the difference between LeTLRP-IL6 and ER-IL6 would not be so great. But that is not the case since the IL6 yields of TLRP-IL6 are very much lower than what could have been expected; its yields are 5 000 times lower than the ones of a simple apoplastic addressing (Apo-IL6 - Nausch et al., 2012b).

Moreover, research is still more focused on improving the production of IL6 in *E.coli* which remains the industrial mean of production of IL6 (Nausch et al., 2013). Microbial fermentation is costly but usually leads to higher yields of proteins (considering amounts of recombinant protein per amount of biomass, extraction buffer and TSP). It would not be worthy to use this system, even if the LeTLRP-IL6 plants were as productive as the ER-IL6 plant and the production of tobacco was inexpensive; it would require treating 250 kg of tobacco leaves with 1m³ of BSA solution to obtain 10 mg of IL6. And this is without considering the solutions needed for the washing steps and the difficulty of homogenizing the plant material. Meanwhile treating less than 4 liters of a bacterial culture is enough to obtain the same result (Nausch et al., 2013). Finally homogenization, growing and purification techniques are already available for bacterial cultures and it is common to use tools adapted to it in the field of biotechnology whereas transgenic plants do not benefit from such an industry.

b. *Efficiency of the TLRP system*

Even though the TLRP system proved it could work by first producing and retaining some IL6 and then releasing it, its efficiency is not optimal since the yields are low (Figure 8 and *Appendix F*) and only 20 to 50% of the extractible IL6 was released from the cell wall (*Appendix G*). Moreover, the amounts of IL6 under its free form, presented in the Figure 8 and the *Appendix F*, are probably underestimated as we showed that the presence of the TLRP peptide disrupts the ELISA test (data not shown). To be more accurately quantified the “free fraction” protein samples should be digested by some thrombin as well.

The IL6 which is not bound to the cell wall would be lost in a process that uses the TLRP to purify the recombinant proteins. This is especially unfortunate given that it was already established that the TLRP system was not giving the best yields. If you consider that the proportion of IL6 lost during the purification represent about 60% of the yield of ML, the IL6 productivity using some ER-IL6 plants would now be 100 000 times higher than what is achieved in this study. An exhaustive analysis of the costs of production in each situation (LeTLRP-IL6, ER-IL6 and *E.coli* productions, etc.) could reveal if the savings made over the production and purification costs could offset those low yields but a 100 000 fold difference is very likely to not be offset.

It was surprising to find some IL6 free (Figure 8) in the apoplast (Table 7) since according to Domingo et al (1999) a chimeric protein using TLRP could not be detected in the homogenized samples. It was then assumed that all the TLRP expressed was binding to the xylem's cell wall. This was supported by the fact that the chimeric protein could be detected via in situ immunohistochemical localization (Domingo et al., 1999). It is not known why the binding of the LeTLRP is not total in this study or why the amounts of IL6 extracted in this study are so small compared to what was achieved without using the TLRP (Nausch et al., 2012b). But it could be that the system is not working properly and that only a small proportion of the IL6 produced is recovered from the cell wall. That would explain the low IL6 yields; the IL6 would be expressed in normal amounts but it would stay bound to the cell wall. In that case the IL6 found under its "free" form could be a trace from the protein anabolism pathway. The ELISA test used to measure IL6 in this study is very accurate and such concentrations may not have been detected in Domingo's studies (1994 and 1999).

c. Conclusion

In the end it would be possible to concentrate the solution and remove the albumin with a polishing step. The TLRP system could then technically be a substitute for other purification systems; it could provide almost pure IL6 without using much purification material and the tobacco is less expensive to grow compared to bacteria. However the amounts of IL6 recovered are extremely low, the TLRP system is therefore not sustainable. It would be useful to be able to check the residual cell wall for some IL6 as all of it could not have been released by the thrombin digestion. This would be done using some of the evocated marker (see I. D.). If a lot more of IL6 could be bound and release from the cell wall, then the TLRP system would have a chance to be worthy.

Even though the idea to use a cell wall anchor to produce some recombinant protein is promising, the TLRP might not be the best candidate to do so. For instance, the fact that the TLRP binds to lignin is not an advantage since lignified tissues are harder to homogenize and are present in limited amounts in the different organs, therefore a cell wall anchor which binds to another the material of the cell wall could be a more efficient system. The fusion protein used in such a system may also be more stable than the TLRP-T-IL6. A first step to improve this system could then be to use a different anchor. There are many different proteins which bind to cell wall and there are different ways for them to do it (Showalter, 1993). In the end, investigating new cell wall anchors may help to achieve both higher yields and a better binding to the cell wall or a more efficient recovery of the produced protein.

It is also possible that the whole capacity of the system has already been explored and that the low yields observed are linked to the nature of the system. This matter should be further investigated to figure out if it would be possible to obtain better yields by improving either the extraction technique or the genetic construct.

B] Visualizing the expression of TLRP in plant

Though we succeeded in obtaining some results about the TLRP system, many questions about how it works remain unanswered. Using IL6 or C5a might not be the best way to answer these questions. This is why some strategies for visualizing how the TLRP behaves in the transgenic tobacco have been set up. Unfortunately these strategies have not given results yet.

Unfortunately it was not possible to obtain a full and functional pLH-TLRP-t-EGFP vector because part of the vector was deleted during the cloning. It is known by experienced molecular biologists that some constructs fail to be made and that from the attempt to bind two sequences an undesirable reconfiguration of the DNA can result (Inge Broer, personal communication, Department of Agrobiotechnology, Agricultural and Environmental Faculty, University of Rostock). The deletion observed on the pLH-TLRP-t-EGFP vector might be due to the use of the doubled enhancer (Figure 3) since repetitive sequences sometimes increase the risks of these kinds of events to happen (Bzymek & Lovett, 2001). The studied deletion is not a random event anyway since all the analyzed clones gave the same profile. It is then likely that the sequence we aimed to obtain cannot be created using conventional methods. Though more attempts to achieve the cloning can be made adjusting different parameters (bacterial growth temperatures, ligation temperatures, number of colonies screened, cutting and assembly of the sequence, etc.), using another marker with a totally different sequence would likely solve the problem faster and has a better chance to work. In order to obtain a functional vector mCherry (kindly provided by Karl Josef Dietz, Biochemistry and Physiology of Plants, Bielefeld) was prepared to be integrated in the pLH-TLRP-t vector. Other alternatives to the EGFP, such as DsRed, or GUS have been explored as well (*Appendix A*) but those constructs were not complete as there was not enough time to do so.

Conclusion

This study allowed us to characterize more clearly the context in which the TLRP system can be used. Although the result is not promising because the yields are relatively low and more productive systems have already been explored, the TLRP system has proven to work. However some adjustments on both the system and the techniques used could still be done to improve it. It would therefore be interesting to keep exploring this system and use the constructs which have been prepared to understand its mechanisms especially as it should be simpler now that directions about how to handle the different samples have been explored. It should also be easier to compare and analyze further results knowing that protein yields are dependent on the tissue sampled, the age of the plant and the promoter expression and that those three factors can be the origin of variability in this experiment.

Bibliography:

- Bzymek, M., & Lovett, S. T. (2001). Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. *Proceedings of the National Academy of Sciences*, 98(15), 8319-8325.
- Ckurshumova, W., Caragea, A. E., Goldstein, R. S., & Berleth, T. (2011). Glow in the dark: fluorescent proteins as cell and tissue-specific markers in plants. *Molecular plant*, 4(5), 794-804.
- Coughlin, S. R. (2000). Thrombin signalling and protease-activated receptors. *Nature*, 407(6801), 258-264.
- Czermak, B. J., Sarma, V., Pierson, C. L., Warner, R. L., Huber-Lang, M., Bless, N. M., ... & Ward, P. A. (1999). Protective effects of C5a blockade in sepsis. *Nature medicine*, 5(7), 788-792.
- Domingo, C., Gomez, M. D., Cañas, L., Hernández-Yago, J., Conejero, V., & Vera, P. (1994). A novel extracellular matrix protein from tomato associated with lignified secondary cell walls. *The Plant Cell Online*, 6(8), 1035-1047.
- Domingo, C., Saurí, A., Mansilla, E., Conejero, V., & Vera, P. (1999). Identification of a novel peptide motif that mediates cross-linking of proteins to cell walls. *The Plant Journal*, 20(5), 563-570.
- De Rodríguez, D. J., Romero-García, J., Rodríguez-García, R., & Angulo-Sanchez, J. L. (2002). Characterization of proteins from sunflower leaves and seeds: relationship of biomass and seed yield. *Trends in new crops and new uses*. ASHS Press, Alexandria, VA, 143-149.
- Fulciniti, M., Hideshima, T., Vermot-Desroches, C., Pozzi, S., Nanjappa, P., Shen, Z., ... & Munshi, N. C. (2009). A high-affinity fully human anti-IL-6 mAb, 1339, for the treatment of multiple myeloma. *Clinical Cancer Research*, 15(23), 7144-7152.
- Gallie, D. R., Lucas, W. J., & Walbot, V. (1989). Visualizing mRNA expression in plant protoplasts: factors influencing efficient mRNA uptake and translation. *The Plant Cell Online*, 1(3), 301-311.
- Haseloff, J., Siemering, K. R., Prasher, D. C., & Hodge, S. (1997). Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly. *Proceedings of the National Academy of Sciences*, 94(6), 2122-2127.
- Hondred, D., Walker, J. M., Mathews, D. E., & Vierstra, R. D. (1999). Use of ubiquitin fusions to augment protein expression in transgenic plants. *Plant physiology*, 119(2), 713-724.
- Hood, E. E., Woodard, S. L., & Horn, M. E. (2002). Monoclonal antibody manufacturing in transgenic plants—myths and realities. *Current opinion in biotechnology*, 13(6), 630-635.
- Hraška, M., Rakouský, S., & Čurn, V. (2006). Green fluorescent protein as a vital marker for non-destructive detection of transformation events in transgenic plants. *Plant cell, tissue and organ culture*, 86(3), 303-318.
- Hraška, M., Rakouský, S., & Čurn, V. (2008). Tracking of the CaMV-35S promoter performance in GFP transgenic tobacco, with a special emphasis on flowers and reproductive organs, confirmed its predominant activity in vascular tissues. *Plant Cell, Tissue and Organ Culture*, 94(3), 239-251.
- Joshi, L., & Lopez, L. C. (2005). Bioprospecting in plants for engineered proteins. *Current opinion in plant biology*, 8(2), 223-226.
- Kenis, G., Teunissen, C., De Jongh, R., Bosmans, E., Steinbusch, H., & Maes, M. (2002). Stability of interleukin 6, soluble interleukin 6 receptor, interleukin 10 and CC16 in human serum. *Cytokine*, 19(5), 228-235.

ANALYSIS OF A TYROSINE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS

- Kung, S.D., Saunders, J.A., Tso, T.C., Vaughan, D.A., Womack, M., Staples, R.C. & Beecher, G.R. (1980): Tobacco as a potential food source and smoke material: nutritional evaluation of tobacco leaf protein. *Journal of Food Science* 45: 320–327.
- Nacry, P., Camilleri, C., Courtial, B., Caboche, M., & Bouchez, D. (1998). Major chromosomal rearrangements induced by T-DNA transformation in *Arabidopsis*. *Genetics*, 149(2), 641-650.
- Nausch, H., Huckauf, J., Koslowski, R., Meyer, U., Broer, I., & Mischofsky, H. (2012a). Recombinant production of the human complement factor 5 a in *Escherichia coli*. *African Journal of Biotechnology*, 11(81), 14685-14695.
- Nausch H, Mischofsky, H., Koslowski, R., Meyer, U., Broer, I., & Huckauf, J. (2012b). High-level transient expression of ER-targeted human interleukin 6 in *Nicotiana benthamiana*. *PloS one*, 7(11), e48938.
- Nausch, H., Mischofsky, H., Koslowski, R., Meyer, U., Broer, I., & Huckauf, J. (2012c). Expression and Subcellular Targeting of Human Complement Factor C5a in *Nicotiana* species. *PloS one*, 7(12), e53023.
- Nausch, H., Huckauf, J., Koslowski, R., Meyer, U., Broer, I., & Mischofsky, H. (2013). Recombinant Production of Human Interleukin 6 in *Escherichia coli*. *PloS one*, 8(1), e54933.
- Omirulleh, S., Ábrahám, M., Golovkin, M., Stefanov, I., Karabev, M. K., Mustárdy, L., & Dudits, D. (1993). Activity of a chimeric promoter with the doubled CaMV 35S enhancer element in protoplast-derived cells and transgenic plants in maize. *Plant molecular biology*, 21(3), 415-428.
- Patterson, G., Day, R. N., & Piston, D. (2001). Fluorescent protein spectra. *Journal of Cell Science*, 114(5), 837-838.
- Remans, T., Schenk, P. M., Manners, J. M., Grof, C. P., & Elliott, A. R. (1999). A protocol for the fluorometric quantification of mGFP5-ER and sGFP (S65T) in transgenic plants. *Plant Molecular Biology Reporter*, 17(4), 385-395.
- Riedemann, N. C., Neff, T. A., Guo, R. F., Bernacki, K. D., Laudes, I. J., Sarma, J. V., ... & Ward, P. A. (2003). Protective effects of IL-6 blockade in sepsis are linked to reduced C5a receptor expression. *The Journal of Immunology*, 170(1), 503-507.
- Roque, A. C. A., Lowe, C. R., & Taipa, M. Â. (2004). Antibodies and genetically engineered related molecules: production and purification. *Biotechnology progress*, 20(3), 639-654
- Rouwendal, G. J., Mendes, O., Wolbert, E. J., & de Boer, A. D. (1997). Enhanced expression in tobacco of the gene encoding green fluorescent protein by modification of its codon usage. *Plant molecular biology*, 33(6), 989-999.
- Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N., Palmer, A. E., & Tsien, R. Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature biotechnology*, 22(12), 1567-1572.
- Sharma, S. K. (1989). Key issues in the purification and characterization of recombinant proteins for therapeutic use. *Advanced Drug Delivery Reviews*, 4(2), 87-111.
- Showalter, A. M. (1993). Structure and function of plant cell wall proteins. *The Plant Cell*, 5(1), 9.
- Somers, W., Stahl, M., Seehra, J.S. 1.9 A crystal structure of interleukin 6: implications for a novel mode of receptor dimerization and signaling. *EMBO J.* v16 pp.989-997 , 1997.
- Szedljak, I., Szántai-Kőhegyi, K., & Kosáry, J. (2010). Study of tobacco plant as a possible nutritive protein source. *Acta Alimentaria*, 39(2), 149-156.
- Tizaoui, K., & Kchouk, M. E. (2012). Genetic approaches for studying transgene inheritance and genetic recombination in three successive generations of transformed tobacco. *Genetics and molecular biology*, 35(3), 640-649.

ANALYSIS OF A TYROSINE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS

- Tremblay, R., Wang, D., Jevnikar, A. M., & Ma, S. (2010). Tobacco, a highly efficient green bioreactor for production of therapeutic proteins. *Biotechnology Advances*, 28(2), 214-221.
- Twyman R. M., Stoger, E., Schillberg, S., Christou, P., & Fischer, R. (2003). Molecular farming in plants: host systems and expression technology. *TRENDS in Biotechnology*, 21(12), 570-578.
- Wang, W. (1999). Instability, stabilization, and formulation of liquid protein pharmaceuticals. *International Journal of Pharmaceutics*, 185(2), 129-188.

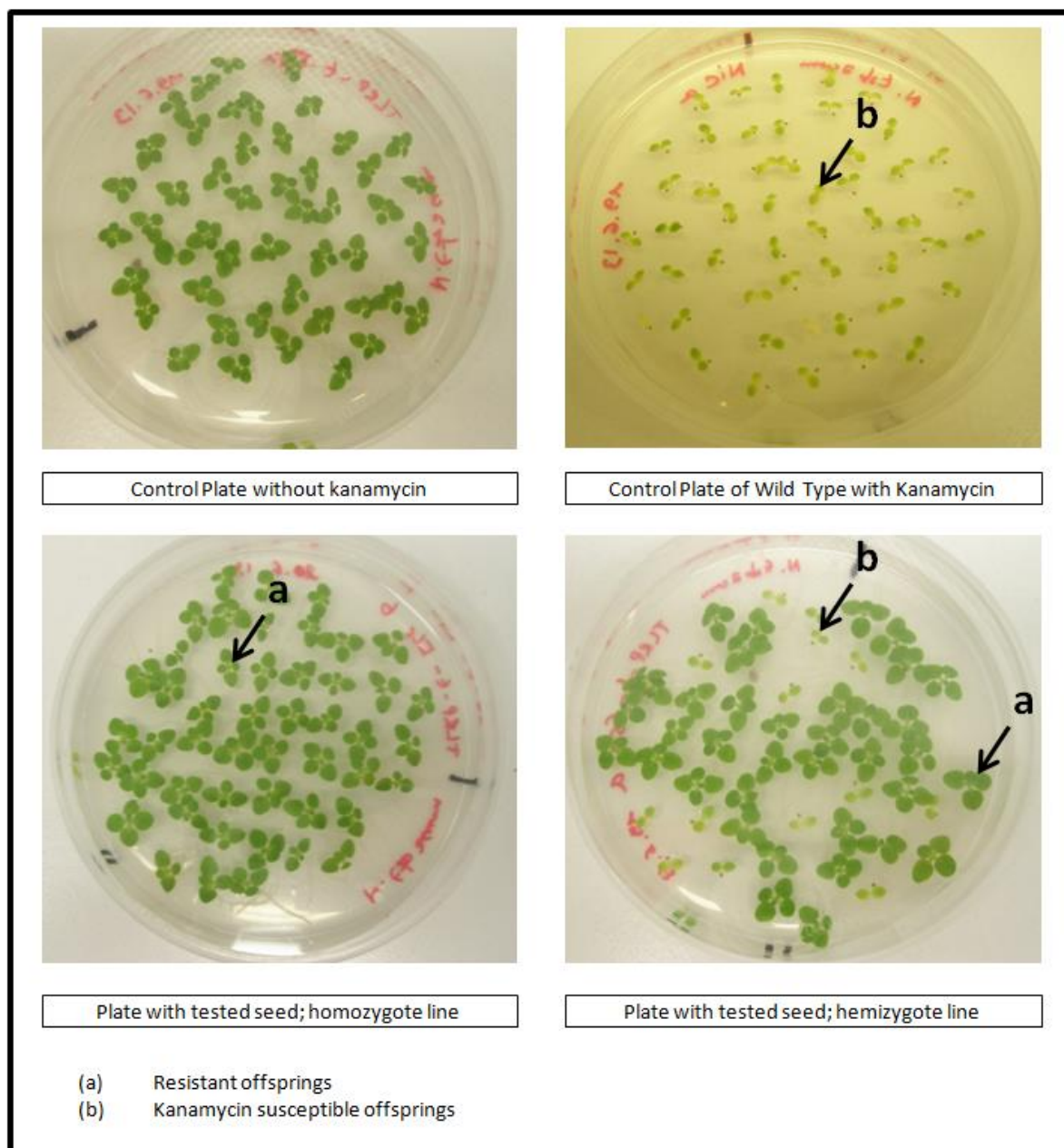
ANALYSIS OF TYROSYNE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS

Appendices:

Appendix A: Sequences of the primers used in this study

The restriction sites added to the constructs thanks to those primers were written in different colors to highlight them. Four different sites are here considered: NruI (red), BamHI (purple), EcoRI (green) and NsiI (blue).

Number	Name	Sequence (5'-3')
1	2npt-fw	TCC GGC CGC TTG GGT GGA GAG
2	2npt-rv	CTG GCG CGA GCC CCT GAT GCT
3	c5afw	AAG AAA TTG CGG CGA AAT ATA A
4	c5arv	GGC CCA GCT GCA TAT CTT TA
5	IL6fw	GGG CGA AGA TAG CAA AGA TG
6	IL6rv	TAA AGC TGC GCA GAA TCA GA
7	NruI-EGFP-fw	ATA TTC GCG AGT GAG CAA GGG CGA GGA G
8	EGFP-NruI-rv	GCC GTC GCG ATT ACT TGT ACA GCT CGTC
9	Thrombin-sense	GTA CAC TTG TAC CAC GTG GTA GTT CG
10	BamHI-LeTLRP	GCA CGG GAT CCA TGG GTT CCA AGG CAA TTT TGT TTC TTG G
11	EGFP-fw	CGT AAA CGG CCA CAA GTT CA
12	EGFP-rv	TCT TGT AGT TGC CGT CGT CC
13	p35St35S-fw	TCC TTC GCA AGA CCC TTC CT
14	t35S (rv)	CAA CAC ATG AGC GAA ACC
15	GFP-uta-fw	ACG TAA ACG GCC ACA AGT TC
16	GFP-uta-rv	ATG CCG TTC TTC TGC TTG TC
17	NruI-DsRed2-fw	AGT TCG CGA GCC TCC TCC GAG AAC GTC A
18	DsRed2+20-NruI-rv	ATC TCG CGA GTA ATG GTA GCG ACC GGC
19	NruI-mCherry-fw	TAG TTC GCG AGT GAG CAA GGG CGA GGA G
20	mCherry-NruI-rv	GTG ATC TCG CGA TTA CTT GTA CAG CTC
21	NruI/EcoRI-GUS-fw	CGA GAA TTC TTA CGT CCT GTA GAA ACC CC
22	GUS-NsiI/NruI-rv	CGA ATG CAT CGA TGG TGC GCC AGG

ANALYSIS OF A TYROSINE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS*Appendix B:* Examples of resistance to antibiotic test's results

ANALYSIS OF A TYROSINE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS

Appendix C: Table of the probabilities of the T2 resistance test's outcomes

The probability of each draw (Probability of the Draw) was calculated based on the law of a binomial distribution considering 50 draws with a 0.75 success rate. Each draw (Transgenic Sprouts out of Fifty) was as well displayed as a percentage of transgenic sprouts over the total (Outcome in %). The probabilities of the draws were summed in the order of their increasing values in spite of evaluating the likelihood that a less probable draw could have been obtained (Likelihood in Percent to obtain a rarer Draw). However the draws are ordered by decreasing values of their probabilities in the table below to evidence which draws are more likely to be obtained. Given that there was seven transgenic plants in F1 (III.A.1.a), seven draws were made. The obtained draws are displayed (Number of Obtained draws) and compared to what would have been obtained if the output strictly followed a binomial repartition (Theoretical Number of draws). However the number of theoretical draws being integers it is not that easy to make the comparison. This is why they have been summed in the order of increasing values as well (Sum of theoretical Draws), to be compared to the sum of the draws really obtained (Sum of obtained Draws).

Transgenic Sprouts out of Fifty	Probability of the Draw	Outcome in %	Likelihood in Percent to obtain a rarer Draw	Number of Obtained Draws	Theoretical Number of Draws	Sum of obtained Draws	Sum of theoretical Draws
38	1.29E-01	76	87.06	1	0.906	1.000	0.906
37	1.26E-01	74	74.46	1	0.882	2.000	1.788
39	1.19E-01	78	62.52		0.836	2.000	2.624
36	1.11E-01	72	51.41	1	0.777	3.000	3.401
40	9.85E-02	80	41.56	2	0.690	5.000	4.091
35	8.88E-02	70	32.68		0.622	5.000	4.713
41	7.21E-02	82	25.47	1	0.505	6.000	5.217
34	6.48E-02	68	18.99		0.453	6.000	5.671
42	4.63E-02	84	14.36		0.324	6.000	5.995
33	4.32E-02	66	10.04		0.302	6.000	6.297
32	2.64E-02	64	7.40		0.185	6.000	6.482
43	2.59E-02	86	4.81		0.181	6.000	6.663
31	1.48E-02	62	3.33		0.104	6.000	6.767
44	1.23E-02	88	2.10		0.086	6.000	6.853
30	7.65E-03	60	1.33		0.054	6.000	6.907
45	4.94E-03	90	0.84		0.035	6.000	6.941
29	3.65E-03	58	0.47		0.026	6.000	6.967
46	1.61E-03	92	0.31	1	0.011	7.000	6.978
28	1.60E-03	56	0.15		0.011	7.000	6.989
27	6.50E-04	54	0.09		0.005	7.000	6.994
47	4.11E-04	94	0.05		0.003	7.000	6.997
26	2.44E-04	52	0.02		0.002	7.000	6.999
25	8.45E-05	50	0.01		0.001	7.000	6.999

ANALYSIS OF A TYROSINE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS

Transgenic Sprouts out of Fifty	Probability of the Draw	Outcome in %	Likelihood in Percent to obtain a rarer Draw	Number of Obtained Draws	Theoretical Number of Draws	Sum of obtained Draws	Sum of theoretical Draws
48	7.71E-05	96	0.00		0.001	7.000	7.000
24	2.71E-05	48	0.00		0.000	7.000	7.000
49	9.44E-06	98	0.00		0.000	7.000	7.000
23	8.02E-06	46	0.00		0.000	7.000	7.000
22	2.20E-06	44	0.00		0.000	7.000	7.000
50	5.66E-07	100	0.00		0.000	7.000	7.000
21	5.56E-07	42	0.00		0.000	7.000	7.000
20	1.30E-07	40	0.00		0.000	7.000	7.000
19	2.79E-08	38	0.00		0.000	7.000	7.000
18	5.52E-09	36	0.00		0.000	7.000	7.000
17	1.00E-09	34	0.00		0.000	7.000	7.000
16	1.67E-10	32	0.00		0.000	7.000	7.000
15	2.55E-11	30	0.00		0.000	7.000	7.000
14	3.54E-12	28	0.00		0.000	7.000	7.000
13	4.46E-13	26	0.00		0.000	7.000	7.000
12	5.09E-14	24	0.00		0.000	7.000	7.000
11	5.22E-15	22	0.00		0.000	7.000	7.000
10	4.78E-16	20	0.00		0.000	7.000	7.000
9	3.89E-17	18	0.00		0.000	7.000	7.000
8	2.78E-18	16	0.00		0.000	7.000	7.000
7	1.72E-19	14	0.00		0.000	7.000	7.000
6	9.14E-21	12	0.00		0.000	7.000	7.000
5	4.06E-22	10	0.00		0.000	7.000	7.000
4	1.47E-23	8	0.00		0.000	7.000	7.000
3	4.17E-25	6	0.00		0.000	7.000	7.000
2	8.70E-27	4	0.00		0.000	7.000	7.000
1	1.18E-28	2	0.00		0.000	7.000	7.000

ANALYSIS OF A TYROSINE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS

Appendix D: Table of probabilities of the possible genetic segregation pattern in T1

A draw is displayed as q/r/s, “q” being the number of “WT-homozygote” obtained in the draw, “r” the number of “Transgenic Homozygote” and “s” the number of “Transgenic Hemizygote”. The probability (p) of each draw was calculated according to the formula:

$$p = \frac{10!}{q! \cdot r! \cdot s!} \cdot \left(\frac{1}{4}\right)^{q+r} \cdot \left(\frac{1}{2}\right)^s$$

The probabilities were ordered by decreasing values and summed either in the decreasing or the increasing order to determine the probability to get a more or a less likely draw than the last one considered in the sum.

The draws experimentally obtained and the draws as likely to happen are highlighted in yellow. Furthermore the draws obtained are displayed in bold.

Draw	Probability of the draw	Probability to get a more likely draw in %	Probability to get this draw or a less likely one in %
2 / 2 / 6	7.69E-02	0.00	100.00
2 / 3 / 5	7.69E-02	7.69	92.31
3 / 2 / 5	7.69E-02	15.38	84.62
3 / 3 / 4	6.41E-02	23.07	76.93
1 / 3 / 6	5.13E-02	29.48	70.52
3 / 1 / 6	5.13E-02	34.61	65.39
2 / 4 / 4	4.81E-02	39.73	60.27
4 / 2 / 4	4.81E-02	44.54	55.46
1 / 2 / 7	4.39E-02	49.35	50.65
2 / 1 / 7	4.39E-02	53.74	46.26
1 / 4 / 5	3.85E-02	58.14	41.86
4 / 1 / 5	3.85E-02	61.98	38.02
3 / 4 / 3	3.20E-02	65.83	34.17
4 / 3 / 3	3.20E-02	69.03	30.97
1 / 1 / 8	2.20E-02	72.24	27.76
1 / 5 / 4	1.92E-02	74.43	25.57
2 / 5 / 3	1.92E-02	76.35	23.65
5 / 1 / 4	1.92E-02	78.28	21.72
5 / 2 / 3	1.92E-02	80.20	19.80
0 / 3 / 7	1.46E-02	82.12	17.88
3 / 0 / 7	1.46E-02	83.59	16.41
0 / 4 / 6	1.28E-02	85.05	14.95
4 / 0 / 6	1.28E-02	86.33	13.67
4 / 4 / 2	1.20E-02	87.62	12.38
0 / 2 / 8	1.10E-02	88.82	11.18
2 / 0 / 8	1.10E-02	89.92	10.08

ANALYSIS OF A TYROSINE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS

Draw	Probability of the draw	Probability to get a more likely draw in %	Probability to get this draw or a less likely one in %
3 / 5 / 2	9.61E-03	91.01	8.99
5 / 3 / 2	9.61E-03	91.98	8.02
0 / 5 / 5	7.69E-03	92.94	7.06
5 / 0 / 5	7.69E-03	93.71	6.29
1 / 6 / 3	6.41E-03	94.48	5.52
6 / 1 / 3	6.41E-03	95.12	4.88
0 / 1 / 9	4.88E-03	95.76	4.24
1 / 0 / 9	4.88E-03	96.25	3.75
2 / 6 / 2	4.81E-03	96.73	3.27
6 / 2 / 2	4.81E-03	97.21	2.79
0 / 6 / 4	3.20E-03	97.70	2.30
6 / 0 / 4	3.20E-03	98.02	1.98
4 / 5 / 1	2.40E-03	98.34	1.66
5 / 4 / 1	2.40E-03	98.58	1.42
3 / 6 / 1	1.60E-03	98.82	1.18
6 / 3 / 1	1.60E-03	98.98	1.02
1 / 7 / 2	1.37E-03	99.14	0.86
7 / 1 / 2	1.37E-03	99.27	0.73
0 / 0 / 10	9.77E-04	99.41	0.59
0 / 7 / 3	9.16E-04	99.51	0.49
7 / 0 / 3	9.16E-04	99.60	0.40
2 / 7 / 1	6.87E-04	99.69	0.31
7 / 2 / 1	6.87E-04	99.76	0.24
5 / 5 / 0	2.40E-04	99.83	0.17
4 / 6 / 0	2.00E-04	99.85	0.15
6 / 4 / 0	2.00E-04	99.87	0.13
0 / 8 / 2	1.72E-04	99.89	0.11
1 / 8 / 1	1.72E-04	99.91	0.09
8 / 0 / 2	1.72E-04	99.93	0.07
8 / 1 / 1	1.72E-04	99.95	0.05
3 / 7 / 0	1.14E-04	99.96	0.04
7 / 3 / 0	1.14E-04	99.97	0.03
2 / 8 / 0	4.29E-05	99.99	0.01
8 / 2 / 0	4.29E-05	99.99	0.01
0 / 9 / 1	1.91E-05	99.99	0.01
9 / 0 / 1	1.91E-05	100.00	0.00
1 / 9 / 0	9.54E-06	100.00	0.00
9 / 1 / 0	9.54E-06	100.00	0.00
0 / 10 / 0	9.54E-07	100.00	0.00
10 / 0 / 0	9.54E-07	1.00E+02	0.00

ANALYSIS OF TYROSYNE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS

Appendix E: IL6-ELISA data for the tissue yields of the different lines (full set of data)

Sample			OD 495nm		Sample Concentration (pg/mL)		Tissue Yield (pg protein/mg plant fresh weight)			
Plant line	Individual	Tissue	Replicate 1	Replicate 2	Replicate 1	Replicate 2	Replicate 1	Replicate 2	Average	Standard deviation
NIC	1	YS	0.12	0.13	7.5	8.2	0.02	0.02	0.02	0.00062
		MS	0.157	0.175	10.3	11.7	0.02	0.02	0.02	0.00087
		ML	0.088	0.09	5.2	5.3	0.02	0.02	0.02	0.00039
IL6: 16-	1	YS	0.702	0.728	61.2	64.0	0.15	0.16	0.15	0.00234
		MS	1.295	1.473	126.9	147.9	0.32	0.34	0.33	0.01736
		ML	1.265	1.414	123.4	140.9	0.58	0.62	0.60	0.02712
	2	YS	2.05	2.135	219.2	230.0	0.50	0.51	0.51	0.00859
		MS	1.872	2.155	196.7	232.6	0.59	0.64	0.62	0.03517
		ML	2.005	2.071	213.5	221.9	1.12	1.15	1.14	0.01535
	3	YS	1.1	1.16	104.5	111.3	0.26	0.27	0.27	0.00591
		MS	1.481	1.69	148.9	174.2	0.40	0.44	0.42	0.02257
		ML	1.784	1.527	185.8	154.4	0.68	0.62	0.65	0.04405
	4	YS	2.018	2.252	215.1	245.1	0.46	0.49	0.47	0.02118
		MS	2.491	2.53	276.4	281.5	0.49	0.50	0.50	0.00324
		ML	1.839	1.796	192.6	187.3	0.81	0.80	0.81	0.00810
	5	YS	1.069	1.111	101.0	105.8	0.28	0.28	0.28	0.00449
		MS	1.142	1.078	109.3	102.0	0.24	0.23	0.23	0.00578
		ML	0.587	0.581	49.5	48.9	0.17	0.16	0.17	0.00072
	6	YS	1.832	1.969	191.7	208.9	0.52	0.54	0.53	0.01568
		MS	1.988	1.913	211.3	201.9	0.54	0.53	0.53	0.00871
		ML	1.412	1.52	140.7	153.6	0.57	0.60	0.59	0.01788
	7	YS	0.288	0.334	21.2	25.3	0.05	0.05	0.05	0.00295
		MS	0.55	0.701	45.8	61.1	0.15	0.18	0.17	0.01586
		ML	0.818	0.834	73.5	75.2	0.27	0.28	0.27	0.00222
	8	YS	1.356	1.392	134.0	138.3	0.30	0.31	0.31	0.00335
		MS	2.763	2.361	312.6	259.3	1.01	0.92	0.96	0.06612
		ML	1.395	1.393	138.6	138.4	0.60	0.60	0.60	0.00036

ANALYSIS OF TYROSYNE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS*Appendix F:* Yields of IL6 recovered from the cell wall and the other compartments

Sample		Tissue Yield (= pg of IL6/mg FW*)	
Individual	Tissue	Free form	Extracted from the Cell Wall
16-1	MS	0.17	0.05
16-2		0.09	0.08
16-1	YS	0.15	0.06
16-2		0.38	0.23
16-1	ML	0.19	0.10
16-2		0.46	0.37
16-1	YL	2.19	1.02
16-2		3.41	1.01

*FW=Fresh Weight

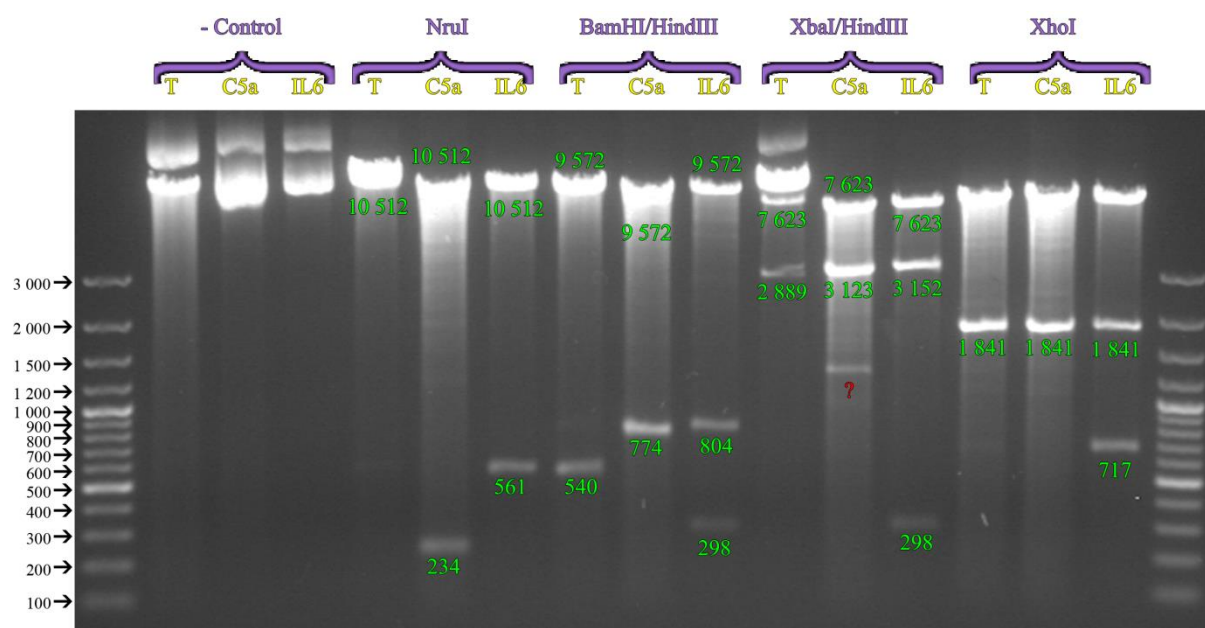
Appendix G: Proportions of IL6 recovered from the cell wall

The following percentages are directly calculated from the data showed in *Appendix F* and Figure 8; it is the amount of IL6 in the fraction digested from the cell wall relatively to the total amount of IL6 extracted (free form + recovered from the cell wall).

Tissue	Individual	% Bound to cell wall
MS	16-1	23
	16-2	46
YS	16-1	28
	16-2	37
ML	16-1	35
	16-2	44
YL	16-1	32
	16-2	23

ANALYSIS OF A TYROSINE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS

Appendix H: Digestion of the origin vectors



This picture represents the outcome of the digestion of the original vectors after migration by electrophoresis. The three vectors: pLH-LeTLRP-Thrombin (T), pLH-LeTLRP-t-C5a (C5a) and pLH-LeTLRP-t-IL6 (IL6) were digested by different sets of enzymes NruI alone, BamHI with HindIII (BamHI/ HindIII), XbaI with HindIII (XbaI/HindIII) and XhoI alone. The negative control (- Control) are samples of undigested vectors. On both sides of the gel molecular ladders have been loaded to evaluate the sizes of the fragment. The sizes of the fragments from the ladder are displayed in black and pointed out with an arrow. The expected sizes of the bands from the samples were calculated using SerialCloner®, these sizes are displayed below or above their localization on the gel. Experimental results fitting to the experiment are written in green, any unexpected result or missing band would be written in red.

All the restriction patterns observed on this gel correspond to the expected ones except for the digestion of pLH-LeTLRP-t-C5a by XhoI and HindII which present an unexplained extra band on its profile. This proves that the background vector is not behind the encountered cloning issues.

ANALYSIS OF A TYROSINE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS

Appendix I: Theoretical restriction pattern of pLH-LeTLRP-t-EGFP's digestion by FokI

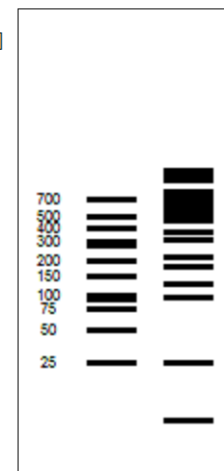
This figure was obtained using Serial Cloner®. It displays information about the pattern that that should be obtained if the expected pLH-LeTLRP-t-EGFP vector was digested by FokI: the number of fragments; the size of each fragment, the first and last nucleotide of each fragments (From “FokI[n° of the first nucleotide of the fragment] To “FokI [n° of the last nucleotide of the fragment]). It also shows (on the right side) what the pattern of the electrophoresis of the digestion should look like. The software ranks the fragments by length and not by order in the sequence as used in this study.

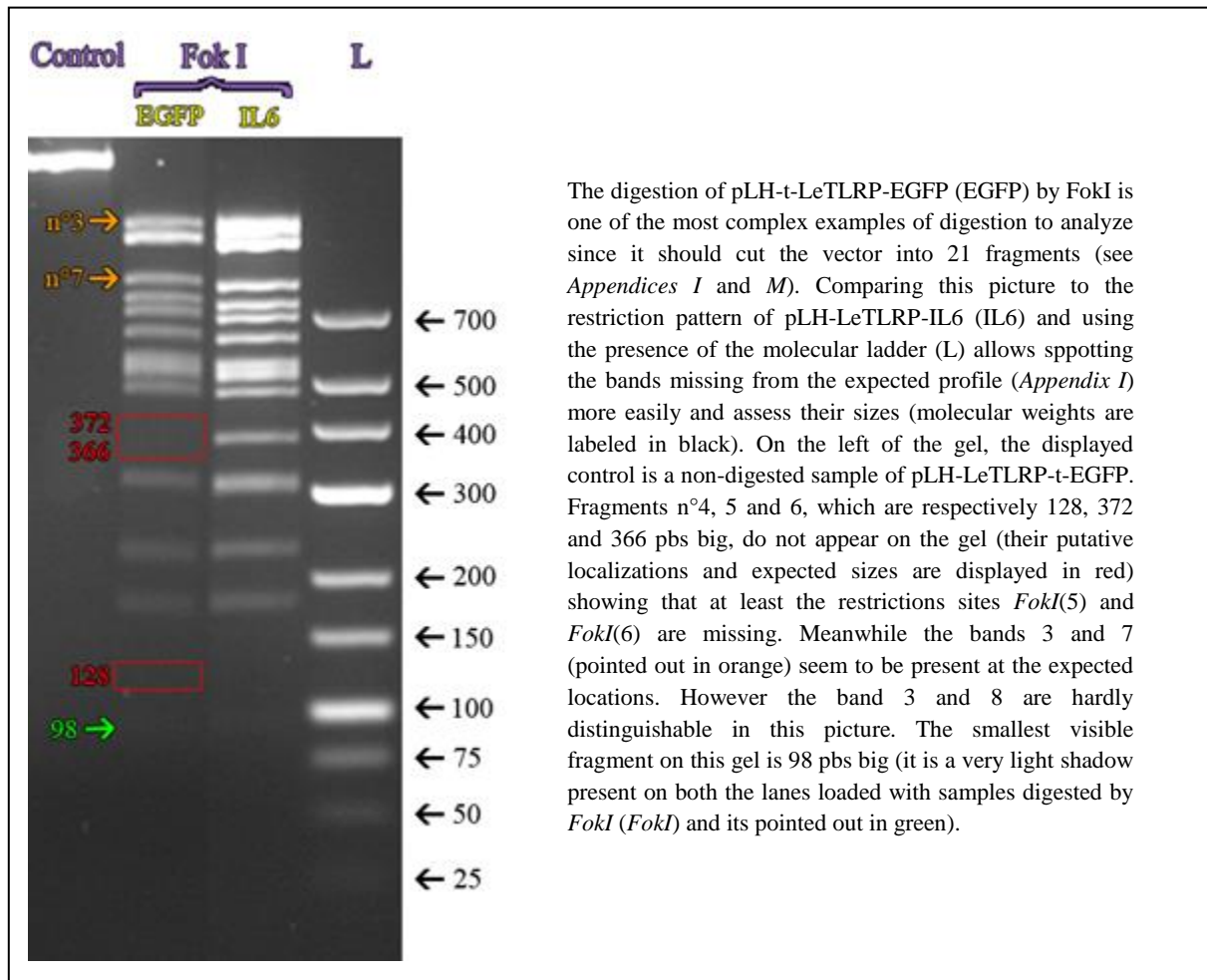
<Serial Cloner V2.5> -- <Fr, 2. Aug 2013 21:09>

Restriction analysis of pLH-LeTLRP-T-EGFPFASTA.xdna [Circular]
Incubated with FokI

21 fragments generated.

1:	1.279 bp	-	From #FokI[912]	To #FokI[2191]
2:	1.226 bp	-	From #FokI[4127]	To #FokI[5353]
3:	1.114 bp	-	From #FokI[5904]	To #FokI[7018]
4:	1.070 bp	-	From #FokI[3057]	To #FokI[4127]
5:	818 bp	-	From #FokI[10773]	To #FokI[350]
6:	726 bp	-	From #FokI[7706]	To #FokI[8432]
7:	688 bp	-	From #FokI[7018]	To #FokI[7706]
8:	597 bp	-	From #FokI[8747]	To #FokI[9344]
9:	537 bp	-	From #FokI[350]	To #FokI[887]
10:	523 bp	-	From #FokI[10242]	To #FokI[10765]
11:	501 bp	-	From #FokI[9741]	To #FokI[10242]
12:	454 bp	-	From #FokI[5353]	To #FokI[5807]
13:	372 bp	-	From #FokI[2191]	To #FokI[2563]
14:	366 bp	-	From #FokI[2691]	To #FokI[3057]
15:	315 bp	-	From #FokI[8432]	To #FokI[8747]
16:	217 bp	-	From #FokI[9524]	To #FokI[9741]
17:	180 bp	-	From #FokI[9344]	To #FokI[9524]
18:	128 bp	-	From #FokI[2563]	To #FokI[2691]
19:	97 bp	-	From #FokI[5807]	To #FokI[5904]
20:	25 bp	-	From #FokI[887]	To #FokI[912]
21:	8 bp	-	From #FokI[10765]	To #FokI[10773]



Appendix J: Digestion of a pLH-LeTLRP-t-EGFP candidate by FokI

There are several facts that make the analysis of the restriction patterns complicated and inconclusive:

- The missing region probably did not start and end exactly right after of before the studied restriction sites. There is likely at least a few dozen nucleotides leeway, and such difference cannot be seen on too small or too big DNA fragments. So the restriction site *FokI*(7), which is 5 nucleotide long could very well be missing as well on the 7th fragment and it could not be seen. The same thing goes for the 3rd fragment, which presence would indicate the presence of both *FokI*(3) and *FokI*(4) in theory but does not actually prove the presence of *FokI*(4).

- Moreover, there should be DNA fragments that are 25, 8 and 4 pbs on the gel according to the restriction pattern prediction (*Appendix I*) but those are not visible and it is not likely that they are missing due to the deletion. Those fragments are just too small to emit a sufficiently strong signal to be seen under those conditions. Therefore there is a possibility that some fragments are there but remain unseen or that some fragments are not there and we cannot see it either.

- Furthermore many bands are overlapping; it is hard to distinguish exactly every band. Going further with the analysis of the restriction profile it was discovered that the 3rd band that should be overlapping with the 8th is in fact not present either. But it cannot be guessed or even supposed from this picture.

ANALYSIS OF A TYROSINE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS

This is why it was necessary to cross the information from the different PCRs (*Appendix K*) and digestions (*Appendix L*) to conclude.

In the end it was theorized that the sequence removed by the deletion included *FokI*(4), *FokI*(5), and *FokI*(6) (*Appendix M*). Defining the lost area, it was possible to say that an extra band, about 600 pbs long, should have appeared as a fusion of the theoretical beginning of the 3rd band and the end of the 6th one. Observing back the gel, it is very much possible that an extra band of 600 pbs would overlap with the other bands. This result is then fitting the emitted theory deduced from crossing all the information obtained from the digestions (*Appendix L*) and the PCRs (*Appendix K*).

The double feedback control of the information as demonstrated in this appendix was used to build the *Appendix M*.

Appendix K: Analysis of the result of the PCRs performed on pLH-LeTLRP-t-EGFP

This table sum up the results of the different PCRs performed on the pLH-LeTLRP-t-EGFP candidates. The sequence of the different primers used, referred to by names inspired from the sequences they bind to (Primer), can be found in *Appendix A*. A number was assign to every set of primer which has been used. It means that for every couple of primers numbered at least one PCR was performed. Some primers (such as t35S (rv)) can bind to two different loci in the vector, when that was the case the loci (locus) were named after a distinctive sequence close to it. The two different products given by one set of primer could be differentiated by their sizes and therefore it was possible to determine which has been amplified successfully when the other was not. The green color indicates that a PCR product was obtained, proving the presence of both primer's sequences on the vector, while a failure (no band), in red, means that at least one of the primer's sequences was compromised by the deletion. In the end, in order to determine which sequences were indeed present or missing (Presence: Yes/No), the PCR information was cross-checked. And finally the information was confronted to the restriction pattern information to conclude on a configuration that could explain the results obtained (Theory).

Primer	locus	Sets (sucess/failure)				Presence	Theory
NruI-EGFP-fw		1				?	No
EGFP-NruI-rv		1				?	Yes
Thrombin-sense			3			No	No
BamHI-LeTLRP				4		No	No
EGFP-fw		2	3	4	7	Yes	Yes
EGFP-rv		2	6			Yes	Yes
p35St35S-fw	nptII	5	6			Yes	Yes
	EGFP	5	6			No	No
t35S (rv)	nptII	5	-	-	-	Yes	Yes
	EGFP	5	9	10	7	Yes	Yes
GFP-uta-fw		8	9			Yes	Yes
GFP-uta-rv		8		10		Yes	Yes

ANALYSIS OF A TYROSINE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS

Appendix L: Analysis of the restriction patterns of the pLH-LeTLRP-t(...) vectors

On the same principle as the *appendix K*, this table sum up the information obtained through the vector digestions. The table is divided in two parts based on the analyzed vectors (Vector): the digestions of pLH-TLRP-EGFP which did not always give the expected patterns and provided data to analyze and the other vectors for which it was not necessary to distinguish the number of cuts observed in the vector (Number of cuts – determined visually on the picture of the gel) and the number of cuts that was expected in theory (expected number of cuts - determined using Serial Cloner®). Indeed based on the results displayed in *Appendix H*, pLH-TLRP-Thrombin, pLH-TLRP-t-C5a, pLH-TLRP-t-IL6 were not affected by the deletion issue.

Meanwhile if some bands were suspected to be missing from the digestion of pLH-TLRP-EGFP, their numbers (Missing band n°) were written down (Read). Example: Digesting pLH-LeTLRP-t-EGFP by BseRI should have given 5 bands, there but it was not possible to see the bands 3 and 4 on the gel. This suggest that *BseRI(4)* is missing and that *BseRI(3)* and *BseRI(5)* could be missing too. More information than what is here reported could be available from those digestions but it was not acquainted in this table. As explained in *Appendix J*, the read of each digestion separately was not always conclusive. Only the analysis made in the end, crossing the different information to find which restriction sites were really missing, allowed giving a satisfactory conclusion (Theory) from an analytical point of view.

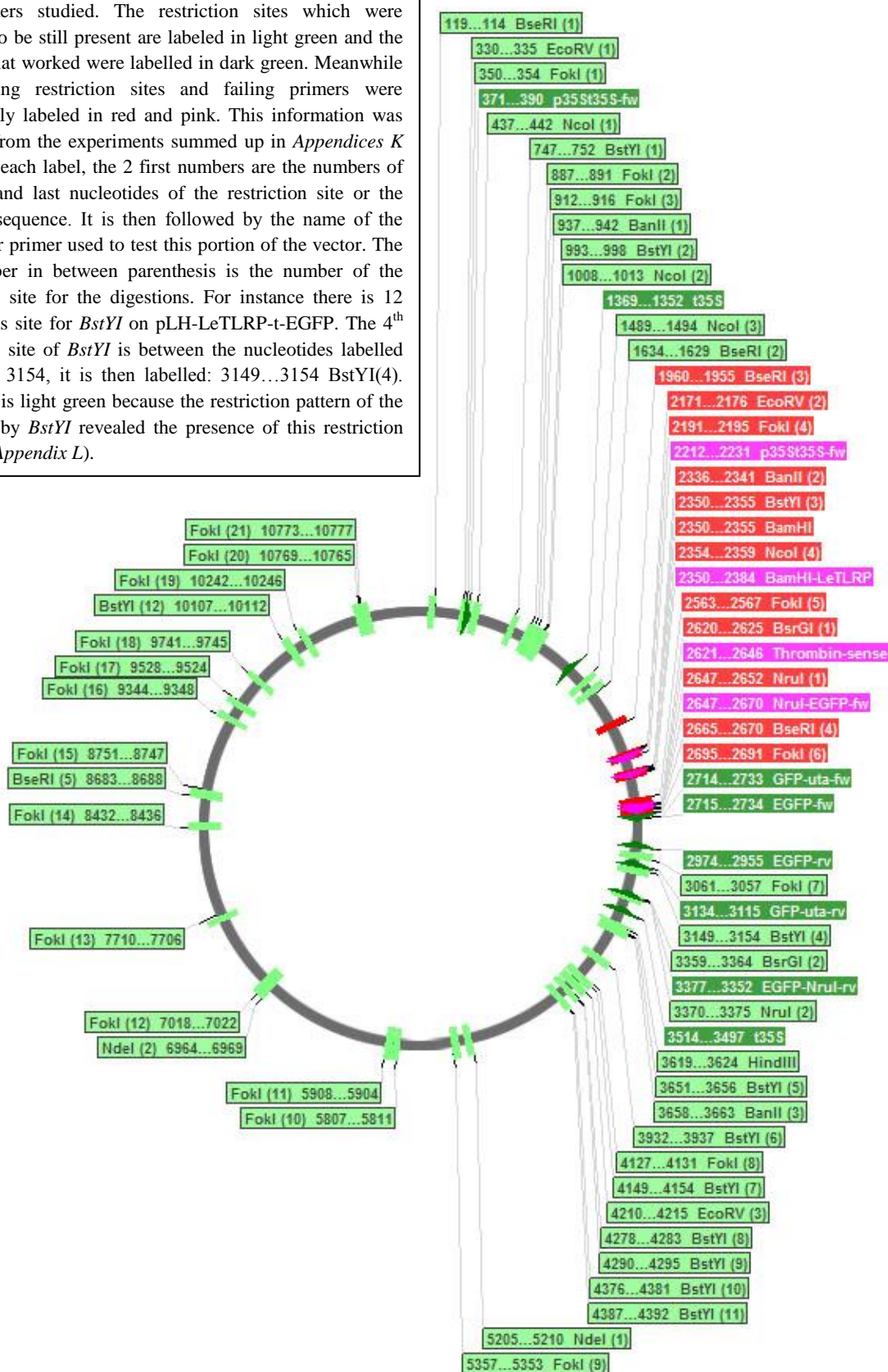
For every digestion a negative control (Control) was made without using any enzyme in case the vector remained undigested (no cut) which showed a specific migration profile different from a single band profile (1 cut). The “-” symbol indicates that the information is lacking, because the experiment was not performed. When two different restriction enzymes were used in a same digestion the enzymes used (Enzyme) were written down separated by a slash. Example: The digestion by both the BamHI and HindIII enzymes is written BamHI/HindIII. When one enzyme was named differently depending on the manufacturer or the literature, the different names were listed with an equal sign in between. Example: BstYI and XhoII are the same enzyme, it was listed as BstYI=XhoII.

Vector :	pLH-TLRP-EGFP			pLH-TLRP-Thrombin	pLH-TLRP-T-C5a	pLH-TLRP-T-IL6
Enzyme	Expected number of cuts	Missing bands (n°)		Number of cuts		
		Read	Theory			
Control	0	none	none	0	0	0
BamHI	0	none	none	1	1	1
BamHI/HindIII	2	1 or 2	1	2	2	3
BanII	3	2	2	-	-	-
BseRI	5	3 and 4	3, 4	-	-	-
BsrGI=Bsp1407I	2	1 or 2	1	-	1	1
BstYI=XhoII	12	3	3	-	n	11
EcoRV	3	2 or 3	2	-	-	-
FokI=BseGI	21	5 and 6	4,5,6	-	19	21
HindIII	1	none	none	1	1	2
NcoI	4	3 and 4	3,4	-	-	-
NdeI	2	none	none	-	-	-
NruI	2	1 or 2	1	1	2	2
SpuI	0	none	none	-	-	-
XbaI/HindIII	-	-	none	2	2	3
XhoI	-	-	2	2	2	3

ANALYSIS OF A TYROSINE- AND LYSINE- RICH FUSION PROTEIN ENGINEERED
FOR PRODUCTION OF CELL WALL-BOUND RECOMBINANT PROTEINS

Appendix M: Mapping of the missing sequence of pLH-LeTLRP-t-EGFP

This figure has been generated using Serial Cloner©. It represent the map of the vector pLH-LeTLRP-t-EGFP on which has been located and annotated all the restriction sites and primers studied. The restriction sites which were deduced to be still present are labeled in light green and the primers that worked were labelled in dark green. Meanwhile the missing restriction sites and failing primers were respectively labeled in red and pink. This information was deduced from the experiments summed up in *Appendices K* and *L*. In each label, the 2 first numbers are the numbers of the first and last nucleotides of the restriction site or the primer's sequence. It is then followed by the name of the enzyme or primer used to test this portion of the vector. The last number in between parenthesis is the number of the restriction site for the digestions. For instance there is 12 restrictions site for *BstYI* on pLH-LeTLRP-t-EGFP. The 4th restriction site of *BstYI* is between the nucleotides labelled 3149 and 3154, it is then labelled: 3149...3154 *BstYI*(4). The label is light green because the restriction pattern of the digestion by *BstYI* revealed the presence of this restriction site (See *Appendix L*).



	Diplôme : Ingénieur Agronome Spécialité : Horticulture Spécialisation / option : BIOTECH (Suivie à AgroParisTech) Enseignant référent : Sylvain CHAILLOU	
Auteur(s) : Nils SAILLARD		Organisme d'accueil : Universität Rostock Agrar und Umweltwissenschaftliche Fakultät Agrobiotechnologie und Begleitforschung zur Bio und Gentechnologie Adresse : Justus von Liebig Weg 8 18059 Rostock ALLEMAGNE Maître de stage : Christoph Unger
Date de naissance : 11/08/1989		
Nb pages : 33	Annexe(s) : 17	
Année de soutenance : 2013		
Titre français : Analyse d'une protéine de fusion utilisant la TLRP (Protéine Riche en Tyrosine et en Lysine) pour produire des protéines recombinantes en les adressant à la Paroi Cellulaire. Titre anglais : Analysis of a Tyrosine- and Lysine- Rich Fusion Protein Engineered for the Production of Cell Wall-Bound Recombinant Proteins		
<p>Résumé : C5a et IL6 sont de petites protéines déjà produites à l'échelle industrielle par fermentation microbienne, ceci à des fins pharmaceutiques. Il a déjà été démontré qu'il était également possible de faire produire ces protéines par des lignées de tabac génétiquement transformées. Une nouvelle construction génétique permettant de les stocker dans la paroi cellulaire a été mise au point, et ce dans le but de faciliter le processus de purification et ainsi réduire les coûts de production de ces protéines. Une telle construction fut obtenue en réalisant une protéine de fusion entre l'IL6 ou le C5a et une protéine pariétale : la TLRP (Protéine Riche en Tyrosine et en Lysine). Une fois cette construction intégrée à une lignée de tabac, les rendements potentiels de ces plantes, ainsi que la localisation subcellulaire des protéines recombinantes produites par cette construction furent étudiées en analysant différentes lignées, tissus et fractions cellulaires de ces plantes. Moins de 50% des protéines recombinantes extraites de ces plantes purent être isolées à partir de leur paroi. De plus les quantités de protéines recombinantes globalement produites par ces plantes sont relativement faibles avec un rendement en protéines isolées à partir de la paroi de leurs jeunes feuilles qui atteint environ 1 pg d'IL6 par mg de poids frais seulement. Une tentative de créer une construction utilisant la EGFP afin de réaliser une analyse plus complète de ce système de production fut également explorée mais un phénomène de délétion lors du clonage empêcha sa réalisation.</p>		
<p>Abstract: C5a and IL6 are small proteins that are already industrially produced for pharmaceutical purposes by microbial fermentation. It was shown that producing them in stably transformed tobacco plants was possible as well. Going further, a new construct genetic construct addressing those proteins to the apoplast to bind and store them in the cell wall was made. This system was achieved using a fusion protein which included the TLRP (Tyrosine- and Lysine- Rich Protein). The purpose of this construct was to ease the purification process and reduce its cost. The yields and the actual localizations of the recombinant proteins that this technique could produce were then investigated in order to know if this system was working efficiently. Different lines, tissues and cellular localizations were investigated to find out the best that could be achieved out of this system. Some of the recombinant protein could, as expected, be recovered from the cell wall but no more than 50% of the whole production as most of the collected recombinant proteins remained free in the apoplast. The overall production of those plants was anyway low. The maximum yield of IL6 extracted from the cell wall being around 1 pg/mg of fresh weight for young leaves, which is the most promising tissue analyzed. A new construct using EGFP was as well prepared in order to further analyze this system but the cloning failed due to some deletion phenomenon which occurred during the cloning step.</p>		
Mots-clés : Molecular farming, paroi cellulaire, IL6, C5a, Protéine Riche en Tyrosine et en Lysine Key Words: Molecular farming, cell wall anchoring, IL6, C5a, Tyrosine- and Lysine- Rich Protein		