Heidelberg Institute for Plant Science

Targeting to compartments of the endomembrane system for the accumulation of HIV-1 p24 in tobacco plants

Maria Goretti Virgili López

Ruprecht-Karls-Universität Heidelberg

Dissertation

Submitted to the

Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany

for the degree of
Doctor of Natural Sciences

presented by

Master of Science

Maria Goretti Virgili López

born in
Tarragona, Spain

Oral examination: 24.10.2008

Targeting to compartments of the endomembrane system for the accumulation of HIV-1 p24 in tobacco plants

Gutachter: Prof. Dr. David G. Robinson

Prof. Dr. Karin Schumacher

Quando você quer alguma coisa, todo o Universo conspira para que você realize o seu desejo

When you really want something to happen, the whole universe conspires so that your wish comes true

É justamente a possibilidade de realizar um sonho que torna a vida interessante

It's the possibility of having a dream come true that makes life interesting

Paulo Coelho, The Alchemist

To my beloved family

Acknowledgments

I am very grateful to many people helping me during my studies. Without them, it is really hard to imagine how I would have been able to finish my thesis.

First, I wish to thank my three supervisors, Prof. Dr. David G. Robinson, Prof. Dr. Jean-Marc Neuhaus and Dr. Alessandro Vitale, for giving me such a precious opportunity to join their labs and for valuable advice, guidance and encouragement.

I would like to thank Prof. Dr. Karin Schumacher, Prof. Dr. Benhard Dobberstein and Prof. Dr. Ralf Bartenschlager to accept to be the committee of my examination.

I am grateful to the Pharma-planta consortium for granting me with the PhD scholarship.

I also really appreciate all the useful advice and help from Dr. Guillaume Gouzerh, Dr. Markus Langhans, Dr. Julia Bubeck, Dr. Maddalena de Virgilio and Dr. Alessandra Barbante, their encouragement and trust make me feel so lucky to have such wonderful colleagues.

I also give my thanks to the different lab people I have met in the three laboratories I have worked. In Neuchâtel: Dr. Ana Slaughter, Dr. Marc Creus, Dr. Anita Ivanova, Dr. Zhanna Kravchuk, Dr. Ilian Simidjiev, Sophie Marc-Martin, Dr. Laurent Zimmerli, Dr. Okmeni N'Guemelieu, Dr. Olivier Zava, Dr. Anne Gouget, Romain Dubresson, Michael, Raphaël Terrier, Farid Mosbaoui, Myriam Scaffidi, Roberta Lopes-Ventura, Letitia Ischer, Noa Ferrari. In Heidelberg: Dr. Julia Bubeck, Dr. Markus Langhans, Dr. Giselbert Hinz, Dr. Peter Pimpl, David Scheuring, Silke Niemes, Christina Larenz, Barbara Jesenofsky, Steffie Gold, Dr. Corrado Viotti, Dr. Stefan Hillmer, Sarah Colanesi, Nicole Frey, Dr. Peter Oliviuson, Dr. Susanne Holstein, Eva Besemfelder-Butz. In Milan: Dr. Maddalena de Virgilio, Dr. Alessandra Barbante, Dr. Anna Giulini, Dr. Emanuela Pedrazzini, Marie Maîtrejean, Andrea Pompa, Giusseppe, Dr. Angelo Viotti, Dr. Aldo Cerotti, Dr. Floriana Gavazzi, Dr. Diego Breviario, Dr. Silvia Giani, Ida Brambilla, Flavia Branfi, Antonia Allevi for their kind advice and friendliness.

My thanks to the gardener in Heidelberg, Mr. Michael Schilbach, who took care of my tobacco plants. I may also thank the tobacco plants and the different bacteria I used during this work as without them I could not achieve any results.

I wish to thank my family and friends. The people I have met while doing my PhD rotation and have become my closest and dearest friends and counsellors, and to all of you I give my love and thanks. In Neuchâtel: Elena Morales, François Zurcher, Doramys and Gregor Hodel, Anita Ivanova, Marc Creus, Ana Slaughter, Marco d'Alessandro, Cécile Fatteberg. In Heidelberg: Julia Bubeck, Nuno Santos, Vitor Vieira, Arunkumar Kota, Alessandro de Mauro, Magdalena Dyniszuk, Roberto Goya, Tiago de Oliveira, Juraj Paule, Valeria Pettorino, Iain Brown, Clarissa Falcao, Svitlana Zhukovska, Chris Byrnes, Juraj Paule, Noemí León, Nicole Schädlich, Katja Batschulat, Pia Oepen. In Milan: Cristina Cristilli, Marie Maîtrejean, Sara, Alessandra Barbante, Maddalena de Virgilio, Filippo, Antonella, Floriana, Monica, Alessio.
In Catalonia and around the world: Alicia Montes, Laura López, Anna Rogero, Xavi Ramírez, Jordi Picas, Maria Laroque, Irene Puga, Corinne Pubill, Marina Castellarnau, María Martín, Elías Pezzat, Juanjo Moreno, Lilian Li, Armin Djamei, Omar and Norma Sued, Harold Oliva and Carla Ferrada, Núria Climent, Chema Martínez, Massimiliano Molinari and Jessica Beckett, Mariarosa Mettifogo and Diego Buja.

I would like to give a special thank to my beloved parents, Miquel-Àngel and Neus, who have always believed in me and helped me to reach my goals. Their unconditional support forged my desire to achieve all that I wish in life. I owe them everything and wish I could show them just how much I love and appreciate them. My sister Laia, who has always encouraged me in my journeys "outside home" and for her inconditional love and trust. My boyfriend Jordi, whose love and encouragement allowed me to finish this journey, and already has my heart and love. Finally, I would also like to dedicate this work to my beloved grandfather Antonio and to the departed soul of my beloved grandmother Magdalena. I hope that this work makes you proud.

Summary

Molecular farming aims at producing high-value proteins in plants for pharmaceutical or other industrial use. Tobacco is one of the most used systems in this field of research, offering developed technology for gene transfer and protein expression. Foreign protein stability is a major issue in molecular farming and is strongly dependent on the subcellular compartment of accumulation. We are screening various compartments of the plant endomembrane system for the high accumulation of the HIV-1 p24 nucleocapsid protein, a potential vaccine against the human immunodeficiency virus.

Previous work has shown that the localisation of fusions of GFP (Green Fluorescent Protein) to the full length or truncated versions of the transmembrane domain (TMD) of human LAMP1 (Luminal Associated Membrane Protein 1) in tobacco, was detected in the lumen of different compartments of the plant secretory pathway (endoplasmic reticulum -ER-, Golgi Apparatus or plasma membrane), depending on the length of the TMD.

In this work, different p24 fusion proteins were designed to accumulate in different compartments of the secretory pathway in tobacco cells. Therefore, the HIV-1 p24 was fused at the N- or C-terminus of the Red Fluorescent Protein (RFP) followed by the different TMDs. Moreover, p24 was also N- or C-terminally fused to the N-terminal domain of maize prolamin γ-zein (zein-p24 and p24-zein). Zein proteins are originally accumulated in ER-derived protein bodies in seeds and previous studies showed the potential of accumulating heterologous proteins in this compartment protecting them from proteases and enhancing their stability. Finally, p24 was fused to the C-terminal tail-anchor of cytochrome b5 (p24-TA), which is expected to be anchored in the ER membrane facing the cytosol.

Localisation studies in tobacco protoplasts showed that the constructs containing RFP at the C-terminus of the p24 (p24RFP-TMD) are targeted to the expected compartments (ER, Golgi Apparatus or plasma membrane). However, when the RFP is placed at the N-terminus of p24 (RFPp24-TMD) the fluorescence appears in the tonoplast and the vacuolar lumen, indicating vacuolar delivery and cleavage from the membrane anchor.

Transgenic tobacco plants expressing the p24RFP-TMD fusion proteins with the correct targeting to the ER, Golgi and plasma membrane, and also expressing zein-p24,

p24-zein and p24-TA were produced. The highest accumulation levels (1% of total soluble protein, TSP) were achieved for p24 containing zein in either N-terminal or C-terminal position. Fusion proteins targeted to the ER showed different accumulation levels if the protein was exposed on the luminal side (p24RFP-TMD, 0.3% TSP) or in the cytosolic side (p24-TA, 0.15% TSP). p24RFP-TMD fusion proteins accumulating in the Golgi apparatus and the plasma membrane showed accumulation levels around 0.15% TSP. The zein fusions formed polymers that were in part difficult to denature even in the presence of SDS, a feature that suggests protein body formation. Pulse-chase experiments indicated that the difference in accumulation of the constructs was mainly due to difference in protein stability. However, RNA blot analysis showed that the zein fusions also lead to increased RNA accumulation. In all cases, the p24 could be released from the fusion tags by digestion with thrombin as the p24 fusion proteins were designed to have a thrombin cleavage site for purification purposes.

On the whole, these results highlight the promising approach of targeting HIV-1 p24 to the ER by fusing to the zein domain and provide new information on the relationship between subcellular localisation and stability of integral membrane proteins.

Zusammenfassung

Das Ziel des “Molecular Farmings” ist es hoch qualitative Proteine für den pharmazeutischen oder weiteren industriellen Gebrauch in Pflanzen zu produzieren. Die meist verwendete Modellpflanze für das “Molecular Farming” stellt Tabak dar, da für diesen Organismus die molekularen Techniken zum Gentransfer bzw. für Proteinexpression optimiert wurden. Ein Problem des “Molecular Farmings” stellt die Stabilität exprimierter Fremdproteine je nach subzellulärer Lokalisation dar. Diese Arbeit untersucht die Auswirkung der Expression des Nukleokapsid-Proteins p24 aus HIV-1 (“Human Immunodeficiency Virus”) in den verschiedenen Kompartimenten des Endomembransystems der Pflanze. Dabei handelt es sich um einen möglichen Impfstoff gegen das HIV.

Mittels der Fusion des Fluorophores GFP an die Transmembrandomäne (TMD) des humanen Proteins LAMP1 (“Luminal Associated Membrane Protein 1”) bzw. an Deletionskonstrukte dieser LAMP1-TMDs konnte bereits gezeigt werden, dass deren Lokalisation innerhalb der verschiedenen Kompartimente des sekretorischen Systems der Tabak-Pflanze (Endoplasmatisches Retikulum, Golgi-Apparat oder Plasmamembran) abhängig ist von der Länge der jeweiligen Transmembrandomäne.

Im Rahmen dieser Arbeit wurden unterschiedliche Fusionsproteine hergestellt um die Expression des p24 in verschiedenen Kompartimenten der Pflanzenzelle in Tabak zu untersuchen. Zunächst wurde an das Protein p24 aus HIV-1 C- oder N-terminal ein RFP (“Red Fluorescent Protein”) fusioniert. Anschließend wurden an den C-Terminus dieses Fusionsproteins die verschiedenen TMDs fusioniert. Des Weiteren wurde p24 N- oder C-terminal mit der N-terminalen Domäne von Prolamin γ-Zein aus Mais fusioniert (Zein-p24 oder p24-Zein). Das Prolamin γ-Zein reichert sich Ursprünglicherweise in ER-gereiften Proteinkörpern (“protein bodies”) im Samen an. Vorhergehende Studien zeigten, dass in diesen Proteinkörpern akkumulierte heterologe Proteine vor Proteasen geschützt sind und dadurch deren Stabilität gefördert wird. Für ein weiteres Fusionsprotein wurde p24 an die endständige Transmembrandomäne von Cytochrom b5 fusioniert (p24-TA). Dadurch wird gewährleistet, dass das rekombinante Protein in der Membran des ERs verankert wird, wobei das p24 sich im Cytosol befindet.

Anhand von Lokalisierungsstudien in Tabak-Protoplasten konnte festgestellt werden, dass die p24-Fusionsproteine, welche das RFP am C-Terminus tragen (p24RFP-TMD),

im jeweils erwarteten Zielkompartiment lokalisiert waren (ER, Golgi-Apparat oder Plasmamembran). Jedoch wurden die p24-Fusionsproteine, die das RFP am N-Terminus tragen (RFPp24-TMD) am Tonoplasten und in der Vakuole detektiert. Diese Mislokalisierung in die Vakuole ist wahrscheinlich auf eine Abspaltung der Transmembrandomäne zurückzuführen.

Es wurden transgene Tabakpflanzen erstellt, die die Fusionsproteine p24RFP-TMD mit Lokalisationssignalen für entweder ER, Golgi-Apparat oder Plasmamembran enthielten. Darüber hinaus wurden auch Pflanzen hergestellt, welche die Proteine Zein-p24, p24-Zein und p24-TA enthielten. Die höchste Ansammlung (*1% total soluble protein, TSP)* wurde bei den N- bzw. C-terminal Zein-fusionierten p24 Pflanzen erreicht. Fusionsproteine mit einem ER-Signal zeigten ein unterschiedliches Akkumulationsniveau, je nachdem, ob p24 luminal (p24RFP-TMD, 0.3% vom TSP), oder cytosolisch (p24-TA, 0.15% vom TSP) exponiert wurde. Das Akkumulationsniveau von p24RFP-TMD im Golgi-Apparat und in der Plasmamembran betrug 0.15% der TSP. Die Zein-Fusionsproteine formten Polymere, die teilweise schwer zu denaturieren waren, selbst unter Verwendung von SDS. Dieses Verhalten lässt den Schluss zu, dass die Zein-Fusionsproteine zu Proteinkörpern fusionieren und damit sehr stabil sind. Das unterschiedliche Akkumulationsniveau dieser Konstrukte konnte durch Pulse-chase Experimente auf eine unterschiedliche Proteinstabilität zurückgeführt werden. RNA-blot-Analysen zeigten jedoch, dass diese Zein-Fusionen ebenfalls zu einer erhöhten RNA-Anreicherung führten. Bei allen Experimenten konnte aufgrund eingefügter Thrombin-Schnittstellen zwischen p24 und der daran fusionierten Sequenzen das virale Protein p24 aus dem jeweiligen Fusionsprotein mittels eines Thrombinverdaus gelöst werden.

Zusammenfassend ergaben die Ergebnisse dieser Arbeit einen viel versprechenden Ansatz, wie das Protein p24 zielgerichtet durch Fusionen an die Zein-Domäne zum ER transportiert werden kann. Darüber hinaus wurden neue Informationen über das Verhältnis zwischen subzellulärer Lokalisation und Stabilität von Membranproteinen präsentiert.

Abbreviations

5'-UTR	5'-Unstranslated region
Amp	Ampicillin
Amp^r	Ampicillin resistant
APS	Ammonium Persulfate
BAP	6-bencylaminopurine
BASTA®	Glufosinate-ammonium
CaMV	Cauliflower Mosaic virus
CLSM	Confocal Laser Scanning Microscope
Cyt b5	Cytochrome b5
EB	Electroporation Buffer
EBA	Expanded bed adsorption
ER	Endoplasmic Reticulum
FXa	Factor Xa
Gen	Gentamycin
Gen^r	Gentamycin resistance
GFP	Green Fluorescent Protein
GUS	Glucuronidase
h	hour
HIV	Human immunodeficiency virus
IPTG	Isopropyl-thio-2-D-galactopyranoside
Kan	Kanamycin
Kan^r	Kanamycin resistance
LAMP1	Luminal associated membrane protein 1
LB	Luria-Bertani medium
LUC	Luciferase
µl	Microliter
min	Minute
mRFP	Monomeric Red Fluorescent Protein
MS	Murashige and Skoog medium
NAA	N-acetyl-t-aspartic acid
Nef	Negative factor
O.D.	Optical density
O/N	Over Night
PB	Protein bodies
PEG	Polyethylenglycol
PXV	Potato-X-virus
RT	Room Temperature
sec	second
ST	Syalil Transferase
TA	Tail anhor
Tet	Tetracyclin
Tet^r	Tetracyclin resistant
TEV	Tobacco Etch virus
TEX	Transient expression buffer
TMV	Tobacco Mosaic virus
Th	Thrombin
TMD	Transmembrane domain
TSP	Total soluble protein
Xgal	indolyl-0-D-galactopyranoside

INTRODUCTION

I Molecular farming - plants as expression system

For thousands of years, plants have provided humans with food, ornamentals, wood, and more recently, with industrial products. They have been used as natural medicines for human health since long time and plants are still used to treat diseases all over the world (Chea et al. 2007; Muthaura et al. 2007). Plants are a source of secondary metabolites which confer many therapeutic effects. They are produced by plants to defend themselves from insects, herbivores, fungi, microbes and plant competitors and to attract pollinators (Wink 1988) but also show anti-inflammatory, antimicrobial and psychoactive properties.

Biotechnology methods using plants as expression system offer the possibility to produce not only high amounts of natural therapeutic secondary metabolites but also foreign recombinant proteins for medical or veterinary applications. There are many advantages in producing these proteins in plants rather than in mammalian cell cultures, bacteria, yeast or transgenic animals (table I). In comparison with other expression systems, plants require low initial investment, medium time scale (months) and very low production costs (< 5€/gram biomass); they have very high scale-up potential, they are faster to grow than transgenic animals and safer, as there is no risk of pathogen contamination, they can produce active complex proteins, and edible crops can even be directly used. The main disadvantages are the early stage development of the expression system and the incomplete biological containment (Fischer 2005).

In 1986, the first pharmaceutical relevant protein was produced in plants: the human growth hormone (Barta et al. 1986). The first recombinant antibody was produced in tobacco in 1989 (Hiatt et al. 1989). From then on, many other recombinant proteins of medical and industrial interest have been produced in different plant species.

The production system of a recombinant protein in plants is a long process when compared with other expression systems. The preparation of the expression vectors, the transformation and regeneration of transgenic plants and the analysis of the different lines and generations can take from months to years. However, there are some transient expression systems that allow obtaining small amounts of product in a short time (Sparkes et al. 2006). With a viral vector, transient expression can also lead to rapid and abundant production of a protein (McCormick et al. 2008).

	Transgenic plants	Plant cell cultures	Bacteria	Yeast	Mammalian cell culture	Transgenic animals
Costs						
Production costs	Low	Medium	Low	Medium	High	High
Time effort	High	Medium	Low	Medium	High	High
Scale-up costs	Low	High[1]	High[1]	High[1]	High[1]	High[1]
Propagation	Easy	Easy	Easy	Easy	Limited	Possible
Productivity	High	Medium	Medium	Medium	Medium	High
Quality						
Product quality	High	High	Low	Medium	High	High
Glycosylation	Similar[2]	Similar[2]	None	Incorrect	Correct	Correct
Contamination risk	No	No	Yes[3]	No	Yes[4]	Yes[4]
Practical application						
Data monitoring	Difficult	Easy	Easy	Easy	Easy	Difficult
Ethical concerns	Medium	Low	Low	Low	Medium	High
GMP[5] conformity	Difficult	Possible	Possible	Possible	Possible	Possible
Storage cost	Inexpensive RT[6]	Inexpensive -20ºC	Inexpensive -20ºC	Inexpensive -20ºC	Expensive N_2	Expensive N_2

Table I. Comparison of different expression systems for the production of recombinant proteins (Ma et al. 2003; Schillberg et al. 2003).

[1] Expensive media, expensive facilities for cultivation and livestock husbandry

[2] Glycan chains produced in animal systems are authentic but may differ from those produced in humans. Plant glycans are similar to those of animals but have plant-specific groups such as α-fucose and β-xylose

[3] Endotoxins

[4] Virus, prions and oncogenic DNA

[5] Good Manufacturing Practice

[6] Room temperature

II Factors affecting recombinant protein expression

One of the main aims in molecular farming is the production of recombinant proteins at high yields. To achieve high yields, it is important to optimise the expression vector design for the different stages of gene expression, from transcription to protein stability (Ma et al. 2003).

II.1 Selection of the promoter

The choice of the promoter will determine the tissue, the levels and the time point at which recombinant proteins are expressed. The 35S Cauliflower Mosaic Virus (CaMV) promoter is the most popular choice for dicotyledonous plants. It is a strong and constitutive promoter (Odell et al. 1985) and its activity can be enhanced by duplication of the enhancer region (Kay et al. 1987; Dowson Day et al. 1993). However, in

monocotyledonous plants this promoter has low activity and other promoters like ubiquitin-1 for maize and *gos-2* for rice are preferred (de Pater et al. 1992; Christensen and Quail 1996). An alternative to enhance dicotyledonous constitutive promoters in monocotyledonous plants is the inclusion of an intron, usually within the 5'-UTR (untranslated region). This allows intron-mediated enhancement, enhancing up to 100 times the activity of the CaMV 35S promoter in *Zea mays* and *Poa pratensis* (Vain et al. 1996).

Although constitutive expression allows high levels of recombinant protein accumulation in many plant tissues, it can adversely affect the normal growth and development of vegetative parts of the plant as well as other organisms in contact with the tissue (herbivores, microorganisms and insects). Moreover, gene silencing can occur leading to the decrease or absence of recombinant protein accumulation (Taylor 1997). Restriction of protein accumulation to seeds using tissue-specific promoters can solve these disadvantages and enhance the stability of the recombinant proteins allowing long-term storage without degradation (Stoger et al. 2000; Ma et al. 2003). Inducible promoters can be also used to limit recombinant protein expression to just before or after harvest (Zuo and Chua 2000).

An important issue is an efficient initiation of the translation for the transgene expression. The 5'-UTR and other sequences around the AUG translation initiation codon play an important role on the translation efficiency in plants (Kozak 1986). There is an enhancement of recombinant protein expression when an appropriate 5'-UTR sequence is cloned before the transgene. The most commonly used sequences are the 5'-UTR sequences from the capsid protein from plant viruses like Tobacco Mosaic virus (TMV) and Tobacco Etch virus (TEV) (Gallie et al. 1991; Niepel and Gallie 1999; Hongmin et al. 2000) but other sequences have shown the same advantages (Zou et al. 2003).

II.2 Codon usage

Transgenes expressed in plants are often originally from heterologous species like viruses, bacteria or mammals. These organisms often have a different codon bias than the host plant, which might result in pausing at disfavoured codons and truncation, misincorporation or frameshifting. These problems can be solved by introducing point mutations in the transgene coding sequence to have the codon usage in line with that of

the host. This will result in maximizing the rate of protein synthesis and eliminating cryptic introns and instable sequences (Ma et al. 2003; Twyman et al. 2003).

II.3 Glycosylation

An advantage plants offer with respect to other expression systems for recombinant protein production is that the protein-synthesis pathway is highly conserved with animals, so human transgenes expressed in plants yield proteins structurally identical to their native counterparts. However, there are some differences in post-translational modification, particularly in the Asn-linked, Golgi-modified glycan structures of secretory glycoproteins (Fig.1). Human *N*-glycosylated proteins expressed in plants lack the terminal galactose and sialic acid residues that are normally found in mammals, but have α(1,3)-fucose and β(1,2)-xylose which are absent in mammals (Ma et al. 2003; Twyman et al. 2003; Gomord et al. 2005). In figure 1 the glycosylation differences between plants and mammals are presented (Gomord et al. 2005). The glycan composition can vary between plant species and between different tissues of the same plant.

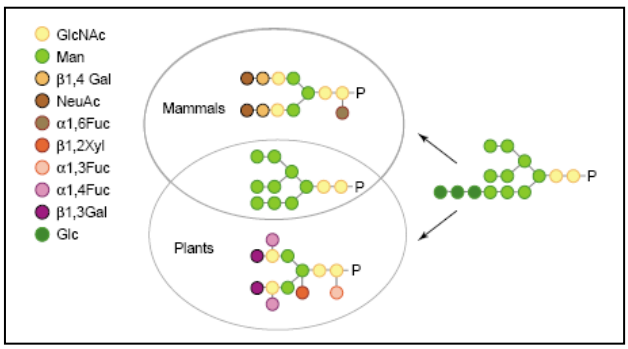


Figure 1. Structure of complex long-chain glycan in plants and mammals. In plants and mammals, two main classes of *N*-glycans are synthesized from the same oligosaccharide precursor. The first class is high-mannose glycans and is present in plants and mammals. Their structure consists in two molecules of *N*-acetylglucosamine (GlcNAc) and between 5 and 9 mannose residues (Man). The second class comprises complex glycans: they derive from the other class and are produced by removal of several mannoses and addition of other sugar residues by Golgi glycosidases and glycosyltransferases, some of the latter being different in plants and mammals. As a result, complex type *N*-glycans are structurally different in plants and mammals. In plants, the proximal *N*-acetylglucosamine of the core is substituted by an α(1,3)-fucose whereas in mammals it is an α(1,6)-fucose. Moreover, the plant core β-mannose is substituted by a β(1,2)-xylose. In addition, there are some differences in the terminal sugars. In plants, β(1,3)-galactose (β1,3 Gal) and α(1,4)-fucose (α1,4Fuc) are linked to the terminal *N*-acetylglucosamine, whereas in mammals the residues linked are β(1,4)-galactose (β1,4Gal) combined with sialic acid (NeuAc). Abbreviation: P, protein.

There is a general concern that the differences in complex glycans could affect the activity, biodistribution and stability of the recombinant proteins compared with the

native forms (Ma et al. 2003). Moreover, some studies have shown the immunogenicity of plant glycans administered to different mammals although their effect has not been proven in humans (Bardor et al. 2003; Jin et al. 2008). To solve these problems, several strategies have been developed to *humanize* the glycosylation patterns of recombinant proteins produced in plants (Bakker et al. 2001; Schahs et al. 2007; Strasser et al. 2008).

II.4 Subcellular protein targeting

Eukaryotic cells are characterized by having complex endomembrane structures through which a large subset of newly synthesized proteins and nutrients travel until they reach their final destination. There are different membrane-bound compartments and organelles with specific functions in the cell. These are the endoplasmic reticulum (ER), Golgi apparatus, peroxisomes, endosomes, lysosomes/vacuoles, mitochondria, and, in plants, plastids (Alberts et al. 2002).

The accumulation of recombinant proteins in a specific cell compartment may greatly facilitate its isolation and purification. The subcellular localisation of recombinant proteins plays a critical role for their correct folding, assembly and accumulation, and the compartment in which they accumulate also determines the possibility of post-translational modifications. Recombinant proteins can be directed to a specific cellular compartment by adding targeting signals. The compartment might protect the recombinant proteins from proteolytic degradation, preserve their integrity and increase the accumulation levels (Kamenarova et al. 2005).

The ER is the port of entry of the secretory pathway. Most secretory proteins are directed to the ER by a specific transient sequence at the amino terminus, termed signal peptide. The signal peptide performs its function as soon as it emerges from the ribosome, and subsequent translocation of the nascent protein into the ER occurs co-translationally (Alberts et al. 2002). Removal of the signal peptide is also co-translational. In all eukaryotes, soluble proteins that enter the ER and lack additional signals traffic to the Golgi complex and are then secreted from the cell by default (Vitale and Pedrazzini 2005).

The ER has been used as a compartment for foreign protein accumulation due to its low hydrolytic activity in comparison to other compartments. In genetic engineering, the most used ER localisation signal is the H/KDEL tetrapeptide, which is present in most soluble ER resident proteins (Vitale and Pedrazzini 2005). Although with this system

levels of one to two orders of magnitude increase of recombinant protein accumulation have been reported, the mechanism can be saturated as the H/KDEL must bind to a receptor located in the Golgi apparatus, being therefore a Golgi-to-ER retrieval mechanism rather than a true retention system (Crofts et al. 1999). In plants, several studies have shown the ER as an optimal compartment for the expression of recombinant proteins. Anti-HBsAg [anti-(hepatitis B virus surface antigen)] mouse IgG1 monoclonal antibody has been expressed in the ER of transgenic tobacco plants using the KDEL retention signal with accumulation levels of 0.5% of total soluble protein (TSP) (Ramirez et al. 2003). In another study bispecific single-chain variable fragment (scFv)-KDEL antibody was expressed in tobacco plants achieving accumulation levels of 1.65% TSP (Fischer et al. 1999). The expression and proper accumulation of two protein components of KDEL-tagged spider dragline silk in tobacco plants has also been described (Menassa et al. 2004).

Despite the high capacity of the ER for recombinant proteins accumulation, inefficient retention has been described. For example, this is the case of the expression of a monoclonal antibody (IgG1) harbouring a KDEL. The heterologous protein was correctly retained in the ER in tobacco leaves, however in seeds the protein was secreted or accumulated in protein storage vacuoles (Petruccelli et al. 2006). In another study, scFv-Fc antibodies fused to KDEL were produced in Arabidopsis. In seeds, the recombinant protein was detected in the periplasmic space instead of the ER, suggesting that overproduction of recombinant scFv-Fc disturbs normal ER retention and protein-sorting mechanisms in the secretory pathway at least in seeds (Van Droogenbroeck et al. 2007). Mistargeting of KDEL-tagged recombinant human serum albumin was observed in wheat endosperm cells in which the recombinant protein was deposited in prolamin aggregates within the vacuole (Arcalis et al. 2004). All these results demonstrate that recombinant proteins containing the KDEL ER retention signal are sometimes deposited in an unexpected compartment, especially in seeds, probably because of the unique storage properties of this organ.

A further organelle offering many advantages for accumulation of recombinant proteins is the chloroplast. As described in the transformation methods (IV.1.2), they can produce high amounts of protein and offer biosafety advantages of transgene containment due to maternal inheritance (Twyman et al. 2003). The highest chloroplast expression has been achieved in maize expressing bacterial genes reaching levels up to 46.1% TSP (De Cosa et al. 2001). However, chloroplasts do not glycosylate proteins,

limiting their use to the production of non-glycosylated proteins. For glycoproteins, there is no alternative to the secretory pathway.

Protein storage vacuoles (Robinson et al. 2005; Vitale and Hinz 2005; Frigerio et al. 2008) and protein bodies (Oparka and Harris 1982; Coleman et al. 1996; Müntz 1998) are protein storage compartments of the secretory pathway. Most storage proteins of seeds and virtually all vegetative storage proteins accumulate in storage vacuoles after synthesis in the ER and traffic through the Golgi complex. Seed storage proteins of the prolamin class, typical of cereals, instead form large polymers within the ER, termed protein bodies. Storage vacuoles and protein bodies have been also used to accumulate recombinant proteins. To produce recombinant proteins in vacuoles, it is important to consider its high hydrolytic activity in vegetative tissues and to understand the mechanisms of deposition of vacuolar storage proteins (Vitale and Pedrazzini 2005). Using this approach, cereals like maize, rice and wheat have been used to accumulate pharmaceutical proteins. For example, a partially humanized secretory immunoglobulin was expressed in rice endosperm cells (Stoger et al. 2000; Nicholson et al. 2005). The non-assembled light chain, heavy chain and secretory component accumulated predominantly within ER-derived protein bodies, while the assembled antibody, with antigen-binding function, accumulated specifically in protein storage vacuoles (Nicholson et al. 2005).

Recombinant proteins can be also targeted to the apoplast. For example, the plant secretory pathway has been described as a proper place for protein folding and glycosylation of secreted recombinant antibodies in tobacco (Schillberg et al. 1999; Nuttall et al. 2005). Another strategy is the secretion of a target protein in the hydroponic medium in roots, termed rhizosecretion, which provides an alternative manufacturing platform that simplifies the downstream purification procedure and increases protein yield. Using this system, *Agrobacterium rhizogenes* induces the formation of large amounts of root tissue called hairy roots and the cells are able to produce recombinant proteins and deliver them to the extracellular medium (root exudates) (Gaume et al. 2003). An example using this technique is the full-length monoclonal antibody complex expressed in transgenic *Nicotiana tabacum* roots which showed a yield of 11.7 μg per gram root dry weight per day per plant (Drake et al. 2003).

III Choice of the host plant

Several plants species have been used for molecular farming and each of them present specific advantages and disadvantages for the produced proteins (table II). Many factors must be considered before choosing the expression host system (Ma et al. 2003; Schillberg et al. 2003). The most used plant production systems have been arranged in four groups: leafy crops, cereals and legumes, fruits and vegetables; and fibre and oil crops.

III.1 Leafy crops

Tobacco has a long history as a successful crop system for molecular farming and is the main tool for testing production in laboratories. The main advantages of tobacco include the well-established technology for gene transfer and expression, the high biomass yield (more than 100,000 kg per hectare for close-cropped tobacco), the potential for rapid scale-up owing to prolific seed production, and the availability for large-scale infrastructure processing. Moreover, tobacco is neither a food nor a feed plant and there is low risk for contamination of the food chain (Ma et al. 2003; Twyman et al. 2003). Due to the advantages offered by tobacco, many recombinant proteins and vaccine candidates have been produced with this expression system (table III).

Other leafy species including alfalfa, soybean, spinach and lettuce are being investigated. Alfalfa is particularly useful because it has a large dry biomass yield per hectare and can be harvested up to nine times a year (Twyman et al. 2003). It has been used to express recombinant antibodies giving the advantage that they are produced as a single glycoform rather than different glycoforms found in other expression systems. *In vitro* galactosidation was successfully achieved to proof that alfalfa plants have the ability to produce recombinant IgG1 having a *N*-glycosylation that is suitable for *in vitro* or *in vivo* glycan remodelling into a human-compatible plantibody (Khoudi et al. 1999; Bardor et al. 2003). Recombinant phytase from *Aspergillus ficuum* has also been produced (Ullah et al. 2002). Many vaccine candidates produced in transgenic alfalfa have shown potential to protect humans and animals from some diseases (table III). The Canadian Medicago Inc. uses this crop for molecular farming.

A rabies subunit vaccine produced in transgenic spinach plants was demonstrated to be efficient in human volunteers, especially in those who had previously been

immunized with a conventional rabies vaccine (Yusibov et al. 2002). However, a HIV vaccine candidate against Tat has been produced in spinach but did not induce antibody production in mice (Karasev et al. 2005).

Lettuce is another broadly used crop which can be grown year-round in greenhouses, and technology is already in place for harvesting, washing and chopping (Joh et al. 2005). Several recombinant edible vaccines have been produced in lettuce against different diseases (table III). Although nuclear or chloroplast transformation are normally used in molecular farming, transient expression in lettuce has been suggested as a good alternative for high yield production of recombinant proteins. Joh et al. (2005) described that agroinfiltrated lettuce produced 1.1% TSP recombinant β-glucuronidase (GUS) which is higher than the commercially-manufactured GUS in transgenic corn seeds (0.7% water-soluble protein). Furthermore, proteins of industrial interest like Miraculin, a taste-modifying protein considered a natural alternative to artificial sweeteners and flavour enhancers, have been produced in this plant (Sun et al. 2006).

The main disadvantage of leafy crops is that recombinant proteins are expressed in an aqueous water environment in which they are often unstable, resulting in low yields. After harvesting, the leaves must be dried or frozen for transport, or processed soon to extract useful amounts of the product. Moreover, in the case of tobacco many cultivars produce high levels of toxic alkaloids: these phenolic compounds are released during the extraction process and can interfere with the downstream processing. Fortunately, varieties exist with low-alkaloid content and can be used for molecular farming (Twyman et al. 2003).

III.2 Cereals and legumes

Cereals and legumes can solve some problems found during the production of recombinant proteins in leafy crops. Seeds offer an appropriate environment for the stable accumulation of recombinant proteins due to the presence of specialised storage compartments derived from the secretory pathway such as protein bodies or protein storage vacuoles. They enable long-term storage, even at room temperature, and as they are desiccated the exposure of stored proteins to non-enzymatic hydrolysis and protease degradation is reduced (Ma et al. 2003; Twyman et al. 2003). Moreover, seeds do not present phenolic compounds like tobacco leaves and the downstream processing is improved (Ma et al. 2003). In terms of biosafety, special precautions may be taken to

avoid pollen spread and seed contamination with non transgenic crops. A main disadvantage is that the overall yield of recombinant proteins in seed crops are much lower than in tobacco, and the most appropriate expression system must be determined on a case-by-case basis. For example, a single-chain Fv antibody expressed in rice, wheat, pea and tobacco showed that tobacco presented the greater amount of biomass produced per hectare (Stoger et al. 2002). Therefore, many factors like biomass yield per hectare, yield of recombinant protein per unit biomass, the ease transformation and the speed of scale-up must be considered before choosing the expression system (Twyman et al. 2003).

Several crops are being used for seed-based production, including cereals (maize, rice and wheat) and legumes (soybean and pea). Maize has been used for the commercial production of avidin and β-glucuronidase (Hood et al. 1997; Witcher et al. 1998), also some recombinant antibodies (Woodard et al. 2003; Karnoup et al. 2005; Rademacher et al. 2008) and antigens against different diseases (table III). The main concern of using maize to produce recombinant proteins is the potential contamination of other corn crops (Sparrow et al. 2007).

Rice has a well-developed gene transfer system, it is self-pollinating (Sparrow et al. 2007) and it has been used to express some subunit vaccines (table III). Moreover, a genetic variation of rice termed Golden rice has been developed but it is not considered a biofarming project in terms that no recombinant protein is purified at the end of the process. However, it is the culmination of many molecular biology techniques used for recombinant protein production but in this case for human nutrition improvement. This rice expresses β-carotene in the grain to treat vitamin A deficiency that often has lethal effects, especially in infants of developing countries (see Golden rice reference). It has been donated for humanitarian use in developing countries and is expected to save many lives.

Wheat has been used for the expression of malaria parasite proteins (Tsuboi et al. 2008) and barley for the expression of β-glucanase (Jensen et al. 1996). At present, ORF Genetics uses a natural built-in containment system for transgenic barley, which is grown in Iceland where the pollen could not survive outside cultivated fields or cross-pollinate (Sparrow et al. 2007).

Soybean presents many advantages including the potential yield, the familiarity with cultivation and the general crop infrastructure. Moreover, it fixes atmospheric nitrogen

which reduces the need for chemical fertilizers (Sparrow et al. 2007). Soybean has been mainly used to produce subunit vaccines (table III). Moreover, transgenic soybean plants can be directly used as a diet complement. Denbow et al. (1998) described a study in which broilers were fed with transgenic soybean expressing recombinant phytase leading to an improvement of the animals in growth and bone strength.

III.3 Fruits and vegetables

The major advantage of protein expression in fruit and vegetable crops is that the edible organs can, in theory, be consumed as uncooked, unprocessed or partially-processed material, making them ideal for the production of recombinant subunit vaccines, nutriceuticals and antibodies designed for topical application (Twyman et al. 2003). However, a disadvantage of edible vaccines is the uncontrollable amount of recombinant protein in each organ from the same plant. Therefore, this approach is unlikely to be allowed because control over the pharmaceutical dosage can be very difficult in this case.

Potatoes have been widely used for the production of many plant-derived vaccines and some of them have been administered to humans in clinical trials (Ma et al. 2003; Twyman et al. 2003). Potato-derived vaccine proteins have been produced as single subunit vaccines and as chimeras containing antigens from two different disease agents (table III). Although some clinical trials have been successful, the widespread development of potato for oral vaccines is hindered by the necessity to cook the tubers to destroy toxins leading to the degradation of thermolabile products (Sparrow et al. 2007). Potatoes have been used for the production of proteins from human milk (Chong et al. 1997), human lactoferrin (Salmon et al. 1998; Chong and Langridge 2000), glucanases (Dai et al. 2000) and diagnostic antibody-fusion protein (Schunmann et al. 2002). Tomatoes are more palatable than potatoes and offer many advantages like greenhouse growth for containment and high biomass yield (~68,000 kg per hectare) (Ma et al. 2003; Twyman et al. 2003). The ripe fruit does not contain toxins and can be dried as a powder and administered in capsule form (Sparrow et al. 2007). It is the first plant in which a plant-derived rabies vaccine was produced (McGarvey et al. 1995). In other studies, recombinant vaccine subunits have been expressed as well as biologically active and glycosylated human alpha-1-antitrypsin showing accumulation levels up to 1.55% TSP (Agarwal et al. 2008). Carrots may be a useful crop because the taproot is a

storage organ that can be consumed raw (Sparrow et al. 2007). They have been used for the production of some plant-derived vaccines (table III). Bananas, and more recently papayas, are attractive fruits to produce recombinant proteins as they are widely grown and are consumed by both children and adults (Twyman et al. 2003). Recombinant HBsAg and an anti-cysticercosis vaccine have been produced in transgenic bananas and papayas, respectively (Kumar et al. 2005; Hernandez et al. 2007).

Species	Advantages	Disadvantages
Model plants		
Arabidopsis thaliana	Range of available mutants, accessible genetics, ease of transformation	Not useful for commercial production (low biomass)
Simple plants		
Physcomitrella patens, Chlamydomonas reinhasdtii, Lemna	Containment, clonal propagation, secretion into medium, regulation compliance, homologous recombination in *Physcomitrella*	Scalability
Leafy crops		
Tobacco	High yield, established transformation and expression technology, rapid scale-up, non-food/feed	Low protein stability in harvested material, presence of alkaloids
Alfalfa, clover	High yield, useful for animal vaccines, clonal propagation, homogenous N-glycans (alfalfa)	Low protein stability in harvested material, presence of oxalic acid
Lettuce	Edible, useful for human vaccines	Low protein stability in harvested material
Cereals		
Maize, rice	Protein stability during storage, high yield, ease to transform and manipulate	
Wheat, barley	Protein stability during storage	Low yields, difficult to transform and manipulate
Legumes		
Soybean	Economical, high biomass, expression in seed coat	Low yields, difficult to transform and manipulate
Pea, pigeon pea	High protein content	Low expression levels
Fruits and vegetables		
Potato, carrot	Edible, protein stable in storage tissues	Potato needs to be cooked
Tomato	Edible, containment in greenhouse	More expensive to grow, must be chilled after harvest

Table II. Characteristics of different plant expression systems.

Single subunit vaccine candidate	Plant species	Reference
Hepatitis B	Tobacco	(Huang et al. 2005)
	Potato	(Richter et al. 2000; Joung et al. 2004; Shulga et al. 2004)
	Lettuce	(Kapusta et al. 1999)
	Soybean	(Huang et al. 2005)
	Tomato	(Gao et al. 2003; Lou et al. 2007)
	Banana	(Kumar et al. 2005)
	Rice	(Qian et al. 2008)
Cholera	Potato	(Arakawa et al. 1997)
	Tobacco	(Jani et al. 2004; Mishra et al. 2006)
	Rice	(Oszvald et al. 2008)
Rotavirus	Tobacco	(Birch-Machin et al. 2004; Perez Filgueira et al. 2004)
	Potato	(Wu et al. 2003; Li et al. 2006)
	Alfalfa	(Wigdorovitz et al. 2004; Dong et al. 2005)
E.coli	Potato	(Haq et al. 1995; Mason et al. 1998; Lauterslager et al. 2001)
	Soybean	(Garg et al. 2007)
	Carrot	(Rosales-Mendoza et al. 2007; Rosales-Mendoza et al. 2008)
	Maize	(Chikwamba et al. 2002; Chikwamba et al. 2003; Streatfield et al. 2003; Tacket et al. 2004)
	Tobacco	(Wen et al. 2006)
Human papillomavirus	Potato	(Biemelt et al. 2003; Warzecha et al. 2003)
	Tobacco	(Massa et al. 2007; Fernandez-San Millan et al. 2008; Lenzi et al. 2008)
Norwalk virus	Potato,	(Mason et al. 1996; Tacket et al. 2000)
	tobacco	(Zhang et al. 2006)
	Tomato	(Santi et al. 2008)
	Tobacco	
Infectious bronchitis virus (IBV)	Potato	(Zhou et al. 2004)
Foot and mouth disease (FMDV)	Potato	(Carrillo et al. 2001)
	Alfalfa	(Wigdorovitz et al. 1999; Dus Santos et al. 2005)
	Tobacco	(Wigdorovitz et al. 1999; Wu et al. 2003; Li et al. 2006)
Transmissible gastroenteritis	Potato	(Gomez et al. 2000)
coronavirus (TGEV)	Maize	(Lamphear et al. 2004)
Respiratory syncytial virus (RSV)	Tobacco	(Belanger et al. 2000)
Human immunodeficiency virus (HIV)	Tobacco	(Zhang et al. 2000; Zhang et al. 2002; Perez-Filgueira et al. 2004; Obregon et al. 2006; Marusic et al. 2007; Sack et al. 2007; Barbante et al. 2008; de Virgilio et al. 2008; Strasser et al. 2008)
	Tomato	(Ramirez et al. 2007)
	Spinach	(Karasev et al. 2005)
	Maize	(Rademacher et al. 2008; Ramessar et al. 2008)
Diabetes	Lettuce	(Ruhlman et al. 2007)
	Carrot	(Porceddu et al. 1999)
	Tobacco	(Avesani et al. 2003)
	Rice	(Yasuda et al. 2005)
Rabies	Spinach	(Yusibov et al. 2002)
	Tomato	(McGarvey et al. 1995)
	Maize	(Loza-Rubio et al. 2008)
	Tobacco	(Modelska et al. 1998; Yusibov et al. 2002; Girard et al. 2006)
Avian reovirus (ARV)	Alfalfa	(Huang et al. 2006)
SARS	Lettuce	(Li et al. 2006)
	Tobacco	(Li et al. 2006)
Measles	Lettuce	(Webster et al. 2006)
	Carrot	(Marquet-Blouin et al. 2003)
	Tobacco	(Huang et al. 2001)
Genital herpes	Soybean	(Zeitlin et al. 1999)
Cysticercosis	Papaya	(Hernandez et al. 2007)
Tuberculosis	Carrot	(Wang et al. 2001)
	Tobacco	(Zelada et al. 2006; Dorokhov et al. 2007)
Newcastle disease virus	Maize	(Guerrero-Andrade et al. 2006)
	Tobacco	(Zhao and Hammond 2005)
	Rice	(Yang et al. 2005; Yang et al. 2007)
Smallpox	Tobacco	(Golovkin et al. 2007)
Poliovirus	Tobacco	(Fujiyama et al. 2006)
Lyme disease	Tobacco	(Navarre et al. 2006)
Simian Immunodeficiency virus (SIV)	Potato	(Kim et al. 2004)

Chimera vaccine candidate	Plant species	Reference
Cholera and periodontal disease	Potato	(Shin et al. 2006)
Cholera and rotavirus	Potato	(Choi et al. 2005)
Cholera and SIV	Potato	(Kim et al. 2004)
HIV and rotavirus	Potato	(Kim and Langridge 2004)
Measles and tetanus	Carrot	(Bouche et al. 2005)
HIV and hepatitis B	Tobacco, Arabidopsis	(Greco et al. 2007; Guetard et al. 2008)
FMDV and hepatitis B	Tobacco	(Huang et al. 2005)
Cholera and diabetes	Tobacco, lettuce	(Ruhlman et al. 2007)
HIV and rabies	Tobacco	(Yusibov et al. 1997)
E.coli and porcine diarrhoea	Rice	(Oszvald et al. 2007)

Table III. Vaccine subunits and chimera vaccines produced in different crops.

III.4 Fibre and oil crops

Oil bodies are small droplets containing a triacylglycerol matrix surrounded by a monolayer of phospholipids embedded with alkaline proteins termed oleosins. SemBioSys have developed a technology based on the fusion of recombinant proteins to oleosin which allows the accumulation of the recombinant proteins in oil bodies simplifying the extraction and purification process. The main crop used is safflower which is generally considered to be self-pollinating (Sparrow et al. 2007). Oilseed rape/canola, flax and cotton have also been considered to produce recombinant proteins, however they present some disadvantages like open pollination, existence of compatible wild relatives and the fibre and oil can interfere with the downstream processing (Twyman et al. 2003; Sparrow et al. 2007).

IV Plant transformation

IV.1 Stable transformation

IV.1.1 Nuclear transformation

There are two methods by which the foreign gene is transferred and incorporated into the host plant nuclear genome in a stable manner: *Agrobacterium tumefaciens* – mediated transformation or particle bombardment. The first method is mainly used for dicotyledonous plants and particle bombardment is mainly used for monocotyledonous plants in which *Agrobacterium*-mediated transformation is less efficient (Ma et al. 2003). Other symbiotic bacteria like *Rhizobium*, *Sinorhizobium* and *Mesorhizobium* are

being tested for the transfer of recombinant DNA in plants cells (Broothaerts et al. 2005).

The main advantage of nuclear transformation is that once the DNA is inserted into the host genome, an almost unlimited and sustainable production capacity can be reached using the established agricultural infrastructure (Boehm 2007). However, the production of transgenic plants is time consuming, from a few to many months depending on the plant species. Moreover, the introduction of foreign DNA into the plant nuclear genome sometimes results in the introduction of superfluous DNA. In *Agrobacterium*-mediated transformation, it is usually due to an inefficient processing of the T-DNA border sequences resulting in the co-transfer of flanking vector sequences or the whole plasmid (Martineau et al. 1994). In the case of particle bombardment, the superfluous DNA is transferred because the whole plasmid is usually used to coat the projectiles (Christou 1996). Some strategies have been developed to solve this problem, like the incorporation of the *barnase* gene outside the T-DNA border in *Agrobacterium*-mediated transformation (Hanson et al. 1999). The expression of this gene is lethal for plant cells, therefore it blocks the proliferation of cells which have integrated part of the vector. To avoid the superfluous DNA using bombardment, instead of the whole plasmid only the essential expression cassettes are attached to the microprojectiles (Breitler et al. 2002).

IV.1.2 Chloroplast transformation

Chloroplasts can be also transformed by incorporation of the gene of interest in their genome. In this case, transplastomic plants are obtained rather than transgenic plants (Boehm 2007). The most commonly method used for chloroplast transformation is the bombardment of the plant tissue with the expression vector, but tobacco appears to be the only species in which this methodology has been established as routine (Svab and Maliga 1993; Daniell 2006). However, the range of transformable plants species is increasing, including cotton, soybean, oilseed rape, tomato and potato (Boehm 2007). Plant cells can contain 100 plastids with 100 identical copies of their genome which means 10,000 copies of the transgenes can be expressed in transplantomic plants leading to large amounts of proteins (up to 47% TSP) never achieved by nuclear transformation (Daniell 2006). Due to the prokaryotic nature of the chloroplasts, gene silencing has not been observed and many genes can be expressed at the same time in

operons. Moreover, chloroplasts as other organelles or compartments limit the possible toxicity of the recombinant protein to the host plant by having it compartmentalized (Daniell 2006; Boehm 2007). Because genes in chloroplast genomes are not transmitted through pollen, recombinant genes are easier to contain avoiding escape to the environment (Horn et al. 2004).

A disadvantage of chloroplasts resulting from their prokaryotic origin is the lack of post-translational modifications like glycosylation, therefore they are similarly limited as the bacterial expression systems (Boehm 2007).

This production strategy has been commercialized by Chlorogen Inc. (St. Louis, MO, USA) to produce a protein of the TGF-ß superfamily which is expected to have the market approval by 2013; and by Icon Genetics (Halle, Germany) which expect to have products on the market by 2012.

IV.2 Transient transformation

IV.2.1 *Agrobacterium*-mediated transient expression (agroinfiltration)

Agroinfiltration is also used to transiently transform leaves in order to test the efficiency of the expression vectors before performing stable transformation. In this case, the recombinant bacteria carrying the expression vector is transferred into the leaf by vacuum infiltration or by direct injection. The bacterial proteins catalyze the transfer of the T-DNA into the host cell nucleus where the transgene will be transcribed without integration into the host genome. After 2-4 days there is the maximum accumulation of the recombinant protein and later there is a decrease in transcription due to post-transduction gene silencing (Kapila et al. 1997; Johansen and Carrington 2001).

In contrast to viral vectors, agroinfiltration does not lead to a systemic expression of the transgene which is only expressed in the agroinfiltrated leaves. Moreover, this system allows the transfer of large genes (>2 kb) which are genetically unstable in viral vectors (Porta and Lomonossoff 1996). Another advantage of agroinfiltration is that different transgenes can be simultaneously expressed by delivering a mixture of recombinant Agrobacteria (Johansen and Carrington 2001).

IV.2.2 Protoplasts

In this system protoplasts (“naked cells”) are obtained from a tissue by degradation of the cell wall, and the plasma membrane is the only barrier between the cytoplasm and

its immediate external environment (Davey et al. 2005). They are exclusively used for preliminary tests on protein synthesis in basic scientific research. Reliable procedures are available to isolate and culture protoplasts from a range of plants, including both monocotyledonous and dicotyledonous crops. Protoplasts can be transformed by electroporation or by polyethylenglycol (PEG)-mediated transformation. The transformation efficiency can reach 90% for *Arabidopsis* and tobacco mesophyll protoplasts and 75% for maize, and co-transfection of multiple plasmids expressing different constructs is very efficient (Sheen 2001; Fisk and Dandekar 2005).

Some experimental conditions need to be optimized for each plant species when transforming protoplasts such as plasmid DNA purity, DNA to protoplast ratio, and protoplast culture density. Moreover, the activities of single cells can also be easily monitored and visualized by vital markers, such as green fluorescent protein (GFP) and luciferase (LUC). Although responses in transient expression assays can be monitored as early as 1 to 2 h after DNA transfection, optimal assay conditions need to be established experimentally (Sheen 2001).

IV.2.3 Virus-mediated

Another method is the use of modified plant viruses like TMV or Potato X virus (PXV) to infect plants and transiently express the recombinant protein. High levels of protein expression are achieved in a short time as the virus is systemically spread however some limitations are that green plant matter must be processed immediately, this system cannot be scaled-up for high amounts (kg) of protein needed, and viruses cannot hold large size transgenes (Porta and Lomonossoff 1996; Fischer et al. 2004; Horn et al. 2004). Despite this, Icon Genetics (www.icongenetics.com) has developed a technology called *magnifection* which combines the advantages of recombinant virus expression and the transgene transfer capacity of *Agrobacterium* achieving in some cases accumulation levels of 80% TSP (Gleba et al. 2005).

V Strategies for recombinant protein accumulation

In this thesis, different strategies to accumulate recombinant proteins in different membranes of the plant secretory pathway have been investigated. Two of these strategies consist in the fusion of the protein of interest to different transmembrane

domains (TMD) and another strategy consists in the fusion to a γ-zein domain to induce the formation of protein bodies. The different TMDs and the γ-zein domain are explained in more detail in the next sections.

V.1 Membrane proteins

V.1.1 Types of membrane proteins

Cell membranes are crucial to the life of a cell. Although the basic structure of biological membranes is provided by the lipid bilayer, most of the specific functions are performed by membrane proteins (Chou 2001). The way that a membrane bound protein is associated with the lipid bilayer usually reflects the function of the protein (Chou 2001).

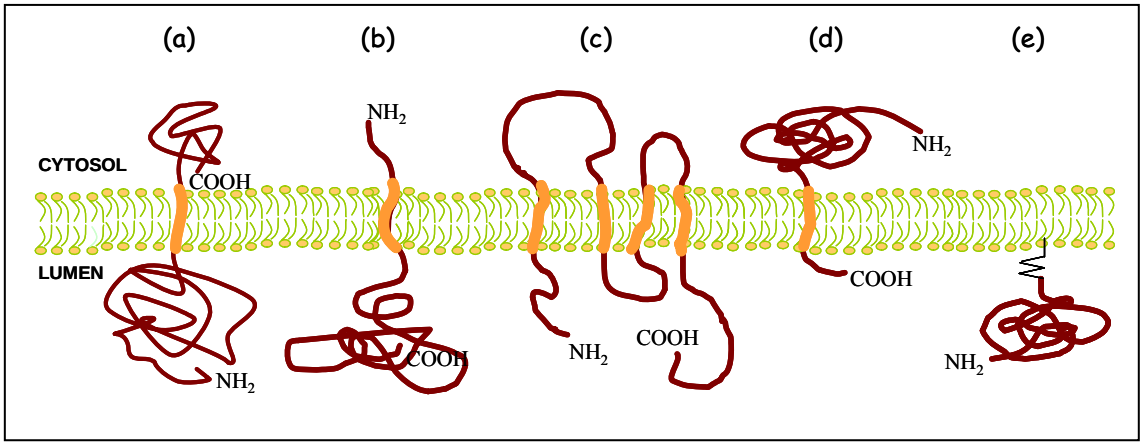


Figure 2. Schematic drawing showing the five types of membrane proteins. (a) type I transmembrane, (b) type II transmembrane, (c) type III multipass transmembrane, (d) type IV transmembrane and (e) GPI-anchored membrane. Type I transmembrane proteins and GPI-anchored proteins contain their N-terminus in the lumen whereas in the type II and type IV transmembrane proteins it remains in the cytosol.

There are five types of membrane proteins: (a) type I transmembrane proteins (Fig. 2a); (b) type II transmembrane proteins (Fig. 2b); (c) multipass transmembrane proteins (Fig. 2c); (d) type IV transmembrane proteins (Fig. 2d); and (e) GPI-anchored membrane proteins (Fig. 2e). The first three types of membrane proteins contain one or more transmembrane domains which pass once or several times through the membrane and interact with the hydrophobic tails of the lipids in the interior of the bilayer (illustrated in Fig. 2 a-c). Type I membrane proteins have a luminal or extracellular N-terminus and a cytosolic C-terminus; type II membrane proteins have a cytosolic N-

terminus and a luminal or extracellular C-terminus; type III proteins have more than one transmembrane domain and therefore cross the membrane at least twice. The membranes of the different compartments of the endomembrane system have different lipid-protein compositions which determine distinct thickness, becoming thicker and more rigid from the ER towards the cell surface (Bretscher and Munro 1993; Sprong et al. 2001). The retention of proteins in the ER, Golgi or plasma membrane depends in many cases on the TMD length, as first inferred by statistical analyses in which features of the TMD are correlated to specific subcellular localisation, and then demonstrated experimentally by expressing recombinant proteins with shortened or extended TMD (Munro 1995; Pedrazzini et al. 1996; Fu and Kreibich 2000; Pedrazzini et al. 2000; Brandizzi et al. 2002). However, the retention mechanisms are still not fully characterised and there are no general rules to accumulate proteins in a specific compartment based on TMD length (Twyman et al. 2003; Fischer et al. 2004; Kamenarova et al. 2005).

The last two types concerning proteins anchored to membrane include type IV transmembrane proteins and GPI-anchored membrane proteins. The type IV membrane proteins are a well-known class of ER resident proteins constituted by enzymes with cytosolically exposed N-terminal catalytic domains and C-terminal membrane anchors (Fig. 2d). Proteins with this kind of topology, referred to as "tail-anchored" are inserted posttranslationally into the ER membrane by a signal recognition particle-independent mechanism (Pedrazzini et al. 1996). For GPI-anchored proteins, the glycophosphatidylinositol (GPI) anchor is attached to the C-terminus of some membrane proteins destined for the plasma membrane. This linkage forms in the ER lumen, where, at the same time, the transmembrane segment of the protein is cleaved off (Fig. 2e). As the protein travels through the secretory pathway, it is transferred via vesicles to the Golgi Apparatus and finally reaches the plasma membrane being attached only by its GPI-anchor to the external face of the cell membrane. The cleavage of the group by phospholipases will result in controlled release of the protein from the membrane to the extracellular space (Vidugiriene and Menon 1994).

V.1.1.1 Human LAMP1

The fusion of a target protein to the TMD of human lysosomal associated protein 1 (LAMP1) and shortened versions has been suggested as a potential tool to accumulate

recombinant proteins in different compartments of the plant secretory pathway by sorting the proteins by the TMD. The resulting recombinant proteins are expected to be type I transmembrane proteins with their N-terminal domain facing the lumen of the compartment in which they accumulate.

LAMP1 is a type I membrane glycoprotein localised in lysosomes and late endosomes. It contains a highly glycosylated luminal domain, a transmembrane domain and a short cytoplasmic tail (Cook et al. 2004). At steady state, LAMP1 is highly enriched in these compartments but a previous work of Brandizzi et al. (2002) demonstrated that different versions of LAMP1 TMD, with different length sequences, fused to GFP and expressed in tobacco plants, were localised in different membranes of the plant secretory pathway, from the ER and Golgi stacks to the plasma membrane. In animal cells, this sorting depends on the length of the TMD and the different composition of intracellular membranes as proposed by Munro et al. (1995). Therefore, we wanted to investigate if the different TMD facilitate this differential accumulation for a target protein.

V.1.1.2 Rabbit cytochrome b5 tail anchor

Another TMD that offers potential for recombinant protein accumulation is the TMD of rabbit cytochrome b5 (cyt b5) tail anchor (TA). TA proteins are a class of integral membrane proteins that are inserted into membranes by a hydrophobic sequence close to the C-terminus and they have the entire functional N-terminal domain facing the cytosol. These proteins do not contain an N-terminal signal sequence and they are sorted to their target membrane by posttranslational mechanisms only upon termination of translation. They are type IV membrane proteins (illustrated Fig. 2d) and in eukaryotes they carry out many fundamental cell functions (Borgese et al. 2003). One example is the mammalian membrane-bound flavoprotein NADH-cytochrome b(5) reductase which exists in two homologous isoforms specifically localised to the mitochondria outer membrane (MOM) and the ER membrane (for review see Borgese et al. 1993). The lack of cyt b5 is responsible for a rare but incapacitating genetic disease, type II hereditary methemoglobinemia (Jaffé and Hultquist 1995; Shirabe et al. 1995). The cyt b5 family includes 15–23 kDa polypeptides which consist of an N-terminal, globular, cytosolic haem-binding domain, a short connecting region, and a hydrophobic transmembrane domain followed by a few polar luminal residues at the extreme C-terminus. In mammals, the net charge of the C-terminal polar region determines the

sorting to the correct target membrane, a positive charge leading to mitochondria and a negative or null charge to the ER (Mitoma and Ito 1992; Kuroda et al. 1998; Korsmeyer et al. 2000; Borgese et al. 2001; Hwang et al. 2004). However, it was observed that the expression of two rabbit cyt b5 isoforms in plants resulted in the expected localisation of the ER form into the ER membrane but the MOM form was sorted to the chloroplast outer envelope (COE). It also seems that the plant ER is less selective than mammalian ER with regard to TA protein targeting, because it is able to accommodate cyt b5 forms with opposite net charges. These findings highlighted that plant cells have a specific and yet uncharacterized mechanism to sort TA proteins with the same positive C-terminal charge to different membranes (Maggio et al. 2007).

It has been demonstrated that the translocation of the C-terminus of a tagged form of mammalian cyt b5 carrying an *N*-glycosylation site in its C-terminal domain to the ER might occur by a mechanism distinct from that of signal peptide-dependent translocation (Yabal et al. 2003). This site was added as a criterion for the correct translocation of the protein across the ER membrane. Moreover, it has been shown that this TA C-terminal domain can be used as a tool to increase the stability of recombinant proteins that are naturally expressed in the cytosol by accumulating them on the cytosolic face of the ER (Barbante et al. 2008).

V.1.2 Protein polymerization and formation of ER protein bodies

V.1.2.1 *Zea mays* γ-zein

As described in II.4, another strategy to produce recombinant proteins consists in their accumulation in protein bodies. To explore it, in this thesis we used the N-terminal domain of maize γ-zein to fuse the protein of interest and induce the formation of protein bodies. The 27 kDa γ-zein protein is a seed storage protein (maize prolamin) soluble in water only under reducing conditions (Vitale et al. 1982). It contains an N-terminal domain mainly composed of 8 repeats of the hexapeptide PPPVHL, followed by a short Proline-rich but not repeated domain, and finally a C-terminal Cysteine-rich domain that contains 8 out of the total 15 Cys residues of the polypeptide (Prat et al. 1985). A special feature of prolamins is the accumulation in ER as very large polymers called protein bodies (Vitale and Ceriotti 2004). In this case the proteins are not accumulated by direct anchoring to the ER membrane. Each maize protein body contains zeins of different classes. By expression in transgenic plants, it has been shown

that those of the γ and β class are able to efficiently form protein bodies also when expressed individually, whereas those of the α and δ class are much more unstable unless co-expressed with γ– or β-zeins (Coleman et al. 1996; Bagga et al. 1997). Expression of mutated forms of γ-zein indicated that the accumulation in the ER and the formation of protein bodies are independent events. While the retention in the ER is directed by the N-terminal domain with or without the Pro-rich domain, the information for protein body formation remains in the C-terminal domain (Geli et al. 1994). Mainieri et al. (2004) described the accumulation of bean phaseolin in ER-derived protein bodies by the fusion to the N-terminal γ-zein domain and the recombinant protein was named zeolin. Although the required information for protein body formation is in the C-terminal domain of γ-zein, the recombinant protein was able to form protein bodies, most likely due to an unexpected effect of the added phaseolin sequence. The fusion of phaseolin with the zein sequence makes the protein very stable and the accumulation (3.5 % TSP) is much higher than if KDEL is added (0.5 %TSP).

As described previously in II.4, the ER constitutes a convenient compartment for accumulation of recombinant proteins and the most common strategy is the fusion to the H/KDEL tetrapeptide (Vitale and Pedrazzini 2005). The fusion to the N-terminal domain of γ-zein constitutes an alternative strategy and a potential biotechnological tool to accumulate recombinant proteins in the ER (Vitale and Pedrazzini 2005). De Virgilio et al. (2008) described the fusion of HIV-1 Nef (Negative factor) to the N-terminal domain of γ-zein and to zeolin. The zein fusions (N-terminal or C-terminal-tagged) were not successful but the zeolin fusion achieved accumulation levels of 1.5% TSP and formation of ER-derived protein bodies in transgenic tobacco plants.

VI Recombinant protein purification

The purification process of recombinant proteins from a plant tissue is the most expensive step in an expression system based on transgenic plants. The main steps can be divided in: (1) fractionation of the plant tissue and protein extraction (2) solid-liquid separation and (3) recombinant protein purification (Kusnadi et al. 1997). The first step includes seed grinding and/or green tissue homogenization followed by cold buffer extraction and centrifugation. The objective of this step is to reduce the biomass volume and prepare the material for protein extraction. The composition of the extraction buffer

(pH, detergents, salts, protease inhibitors, reducing agents) determines the capacity of protein solubilization which also depends on the transgenic plants and the nature of the recombinant protein (Kusnadi et al. 1997). The main problems of this step are the proteolytic degradation and reactions with phenolic compounds and oils. An alternative to avoid these problems is to produce proteins that are secreted as disruption of cells is not necessary and these compounds will not be released (Kusnadi et al. 1997; Fischer et al. 2004). Moreover, for species with a high oil content, hexanes can be included in the extraction step without affecting the activity of the recombinant protein (Kusnadi et al. 1998).

In the second step, the separation of solid (insoluble) and liquid (soluble) phases is achieved by centrifugation or filtration. Another option is the use of EBA (expanded bed adsorption) resins however the size of the material is very important as it can collapse the expanded bed (Menkhaus et al. 2004).

The protein purification (step 3) is usually carried out by chromatographic techniques by using ion exchange, hydrophobic interaction, and size exclusion chromatography (Kusnadi et al. 1998). Another useful technique is the addition of affinity tags to facilitate the recovery of the recombinant protein in which the tags are removed after purification to restore the native structure of the protein (Fischer et al. 2004). Usually these tags are released by protease cleavage, which increases the final purification costs. To avoid cleavage with enzymes, it is useful to fuse the protein of interest to intein, a self-cleaving protein domain (Perler 2005). Recombinant proteins can be also accumulated in different cell compartments: for example, by fusion to TMD, in ER-derived oil bodies by fusion to oleosin, and in ER-derived protein bodies by fusion to γ-zein (Brandizzi et al. 2002; Mainieri et al. 2004; Capuano et al. 2007). The different compartments can be isolated and the recombinant proteins recovered.

It is important to have a purification system which is commercially competitive and allows not highly expensive purification of functional recombinant protein in large-scale. In summary, the final utilization of the recombinant protein, the plant species in which is produced and the tissue in which is expressed play key roles in determining the best purification method.

VII Human Immunodeficiency Virus (HIV)

VII.1 HIV infection

In 2007, 33.2 million people worldwide were living with HIV, about 2.5 million people became newly infected with HIV and an estimated 2.1 million lost their lives to AIDS. Despite the international community's best efforts, the HIV pandemic continues unabated. On average, people require life-saving antiretroviral treatment (ARVs) 7-10 years after becoming infected. While there has been recent progress in increasing access to treatment and prevention programs, HIV continues to outpace the global response with at least 70% of those in clinical need of ARVs worldwide not receiving them (UNAIDS).

Every day, nearly 7,000 people become newly infected with HIV and about 6,000 people die from AIDS which accelerates the pressure to develop a vaccine that must be a global health and development priority (IAVI). It is generally accepted that a first generation of successful HIV vaccines will offer much less than 100% of protection (as any new vaccine offers). Further development on these vaccines will lead to a preventive HIV vaccine which will be improved with time. It has been suggested that even partially effective vaccines would (i) protect some vaccinated individuals against HIV infection, (ii) reduce the probability that a vaccinated individual who later becomes infected will transmit the infection to others, and (iii) slow the rate of progression to AIDS for those who later become infected with HIV.

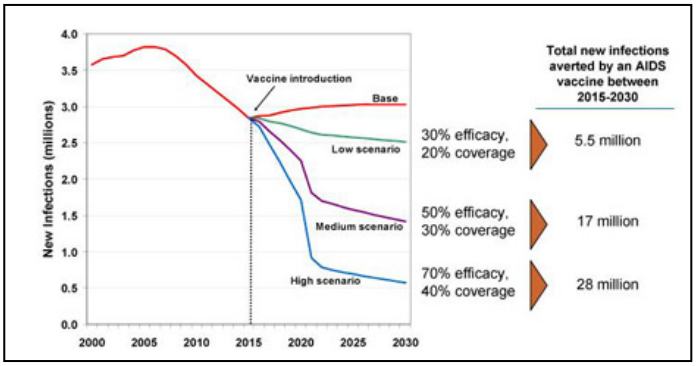


Fig.3. HIV infection development after a massive worldwide vaccination. New adult HIV infections in low-, middle- and high-income countries by year and vaccine scenario. The introduction of a vaccine at 2015 was chosen for illustrative purposes as a vaccine is not guaranteed by 2015.

IAVI organisation estimates that – even assuming that other programs for treatment and prevention have been scaled up – an HIV vaccine could substantially alter the course of the AIDS pandemic and reduce the number of newly infected people, even if vaccine efficacy and population coverage levels are relatively low (Fig. 3).

VII.2 HIV life cycle and structure

The HIV is a member of the *Retroviridae* family which comprises an important number of species that mostly infect vertebrates. This family is divided into three subfamilies: *Oncovirinae*, *Lentivirinae* and *Spumavirinae*. The HIV-1 is a Lentivirus and the name means they infect slowly and latently. In this subfamily there are other immunodeficiency virus that infect apes or monkeys (SIV) and cats (FIV), as well as another HIV-2 that remains localised in East Africa, (Turner and Summers 1999).

The HIV is a spherical particle of 80-100 nm, enveloped, with two copies of RNA. It contains three layers: the internal layer which contains the RNA, the nucleoproteins and the viral enzymes, the icosahedral capsid, and the envelope derived from the host cell membrane. The genome consists of two identical 9.6 kb long RNA chains, with positive polarity, which encode nine open reading frames (ORF). Three of these encode the Gag, Pol, and Env polyproteins common to all retroviruses, which are subsequently proteolysed into individual proteins. The four Gag proteins, p17 MA (matrix), p24 CA (capsid), p7 NC (nucleocapsid), and p6, and spacer peptides (SP) are structural components that make up the core of the virion. The two Env proteins, SU (surface or gp120) and TM (transmembrane or gp41) are structural components of the outer membrane envelope (Frankel and Young 1998; Ganser-Pornillos et al. 2008). The three Pol proteins, PR (protease), RT (reverse transcriptase), and IN (integrase), provide essential enzymatic functions and are also encapsulated within the particle. HIV-1 encodes six additional proteins, often called accessory proteins, three of which (Vif, Vpr, and Nef) are found in the viral particle. Two other accessory proteins, Tat and Rev, provide essential gene regulatory functions, and the last protein, Vpu, indirectly assists in assembly of the virion (Frankel and Young 1998). The HIV-1 virion structure is shown in figure 4 and the life cycle in Figure 5.

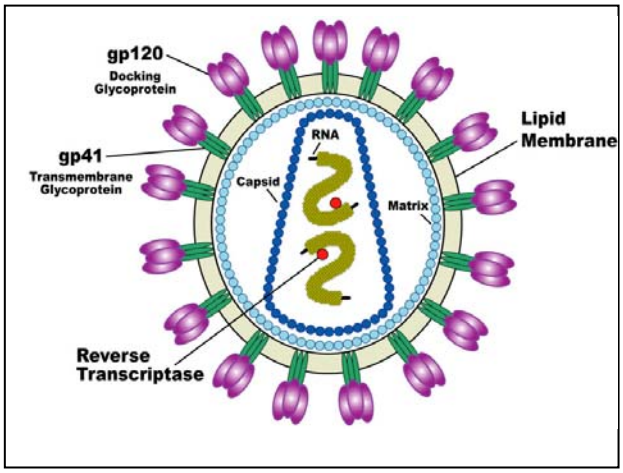


Figure 4. Structure of the HIV-1 virion. Envelope proteins: gp41 and gp120. Core proteins: capsid (p24), matrix (p17). Polymerase proteins: RT

VII.3 HIV vaccine candidates

The Gag protein is a single multi-domain viral protein which is sufficient for assembly of retrovirus-like particles in mammalian cells (Datta et al. 2007). Gag was shown to be a conserved gene in contrast to envelope protein genes and various T-cell epitopes have been mapped. Cytotoxic T lymphocyte (CTL) responses are responsible for bringing down the initial burst of viremia. In many studies with long term survivors CTL epitopes have been mapped to the *gag* region showing the maintenance of memory to *gag* epitopes (Chugh 2003).

After the failure of many vaccines including HIV envelope proteins, several Gag vaccines have been produced as single protein subunits or together with an attenuated viral or bacterial vector (Hanke et al. 2002; Xu et al. 2003; Song et al. 2006; Xin et al. 2007).

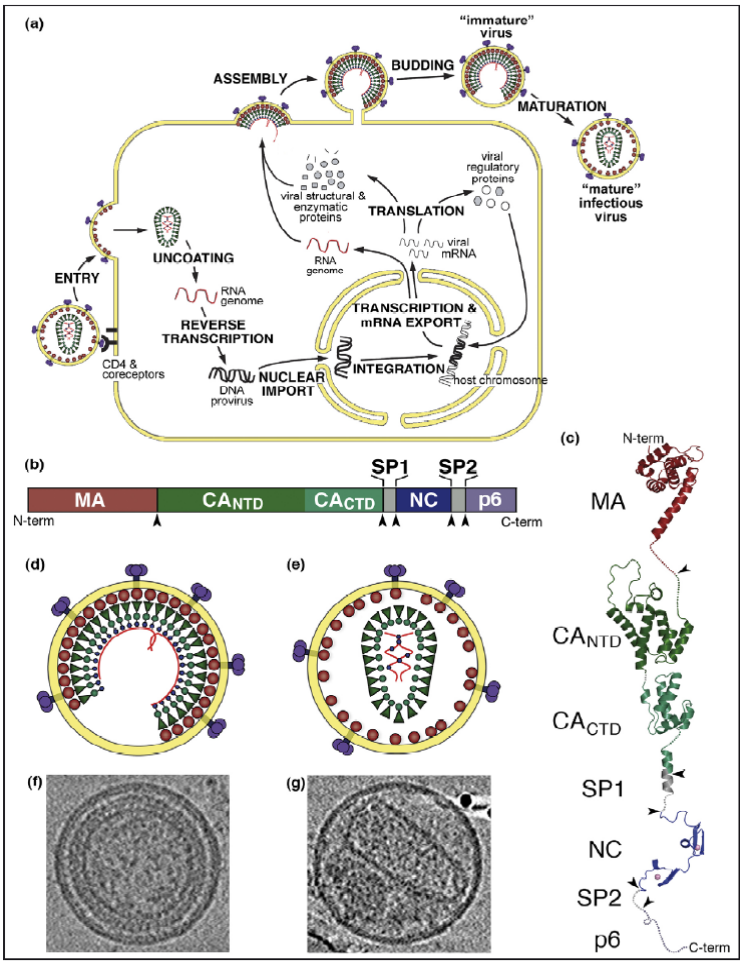


Figure 5. HIV life cycle and structure of Gag. (a) Summary of the HIV-1 replication cycle. (b) HIV-1 Gag polyprotein domain structure, showing the locations of MA, CA_{NTD}, CA_{CTD}, SP1, NC, SP2, and p6. (c) Structural model of the extended Gag polypeptide, derived from high-resolution structures and models of isolated domains. PR cleavage sites are indicated by the arrowheads in (b) and (c). Schematic models of the immature (d) and mature (e) HIV-1 virions. Tomographs of immature (f) and mature (g) HIV-1 particles taken by electron cryotomography. The spherical virions are approximately 130 nm in diameter. From (Ganser-Pornillos et al. 2008).

For example, a polyvalent vaccine in which Gag was expressed in an auxotrophic Bacillus Calmette-Guérin strain showed potential immunity against HIV (due to Gag) and against tuberculosis (due to the bacterial vector) (Im et al. 2007). An important domain of Gag is the p24 capsid protein. This protein forms the conical core of viral particles and cyclophilin A, a human cellular proline isomerase, has been demonstrated to interact with the p24 region Gag leading to its incorporation into HIV particles. The interaction between Gag and cyclophilin A is essential because the disruption of this interaction inhibits viral replication (Franke et al. 1994; Franke and Luban 1996). Moreover, formylation of the *N*-terminal proline of HIV-1 p24 was found by proteomics. The role of formylation of HIV-1 p24 is still unclear, but it is thought to play a critical role in the formation of the HIV-1 core and for infectivity (Misumi et al. 2007).

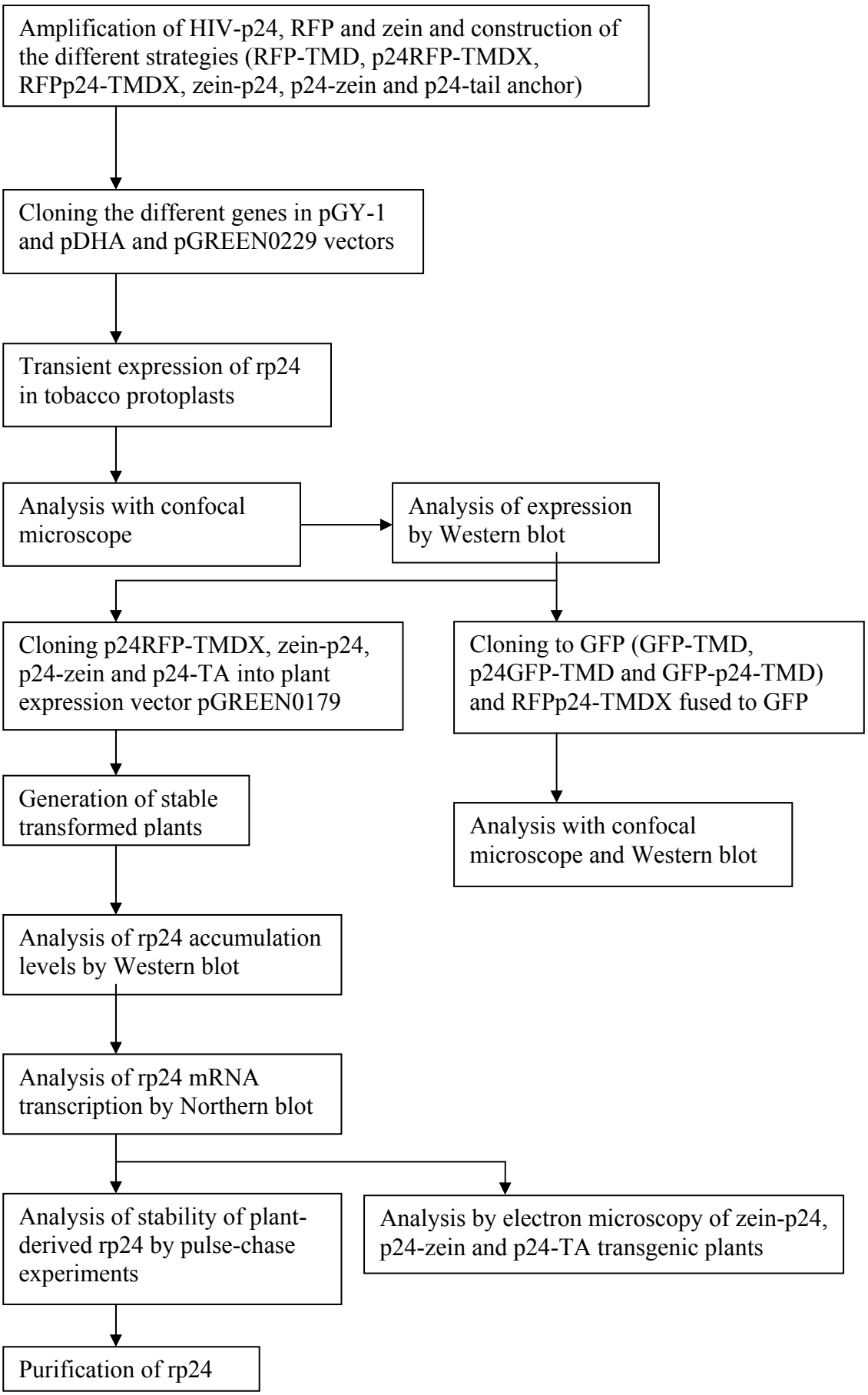
p24 was proved to trigger both cellular and humoral host responses highlighting its suitability as a subunit vaccine candidate. Moreover, p24, like Gag, is much more conserved among HIV clades and other immunodeficiency viruses than envelope proteins (Matsuo et al. 1992; Hilpert et al. 1999).

The HIV-1 p24 capsid protein has been produced in different expression systems including bacteria (Gupta et al. 1997; Qoronfleh 1999; Castilho et al. 2005; Bhardwaj et al. 2006; Mahboudi et al. 2006; Im et al. 2007), yeast (Vlasuk et al. 1989; Jiang et al. 2005), insect cells (Joshua et al. 2000), mammalian cells (Chen et al. 2007), and also plants (table III). In plants, it has been produced in tomato, however, in transgenic tomatoes using tomato bushy stunt virus, the p24 ORF was not maintained in the viral vector (Zhang et al. 2000). In contrast, transgenic tobacco plants showed accumulation levels of 0.35% TSP in the cytosol (Zhang et al. 2002). In another study, p24 from HIV-1 clade C was inoculated into *N.benthamiana* plants using TMV as a vector, obtaining yields of 100 mg per kg fresh leaf weight (Perez-Filgueira et al. 2004). On the basis of the observation that immunoglobulin proteins are usually expressed at high levels in plants, a fusion molecule containing the p24 from HIV-1 clade B fused to the Cα2 and Cα3 constant region domains of a human IgA α-chain was prepared. The fusion protein was targeted to the secretory pathway and presented levels of accumulation of 1.4% TSP in transgenic tobacco plants (Obregon et al. 2006).

Furthermore, in a recent study, two HIV-1 clade C proteins, p24 and p17/p24 chimera, were targeted to the ER and to the chloroplasts in tobacco plants (Meyers et al. 2008).

Transient expression of tobacco chloroplasts achieved maximum levels of accumulation, about 0.3% TSP, for both p24 and p17/p24.

In the present work different strategies have been tested exploring the possibility to accumulate the HIV-1 p24 protein in different membranes of the plant secretory pathway. This protein was fused to different domains and/or proteins to target the recombinant protein to different compartments, trying to increase its stability and therefore its accumulation.


Amplification of HIV-p24, RFP and zein and construction of the different strategies (RFP-TMD, p24RFP-TMDX, RFPp24-TMDX, zein-p24, p24-zein and p24-tail anchor)
Cloning the different genes in pGY-1 and pDHA and pGREEN0229 vectors
Transient expression of rp24 in tobacco protoplasts
Analysis with confocal microscope
Analysis of expression by Western blot
Cloning p24RFP-TMDX, zein-p24, p24-zein and p24-TA into plant expression vector pGREEN0179
Cloning to GFP (GFP-TMD, p24GFP-TMD and GFP-p24-TMD) and RFPp24-TMDX fused to GFP
Generation of stable transformed plants
Analysis with confocal microscope and Western blot
Analysis of rp24 accumulation levels by Western blot
Analysis of rp24 mRNA transcription by Northern blot
Analysis of stability of plant-derived rp24 by pulse-chase experiments
Analysis by electron microscopy of zein-p24, p24-zein and p24-TA transgenic plants
Purification of rp24

RESULTS

The objective of the present work was to produce the recombinant HIV-1 p24 protein in different subcellular compartments of the plant secretory pathway, in an effort to achieve high accumulation in tobacco plants. The p24 coding sequence was fused to red fluorescent protein (RFP) or GFP and to different transmembrane domains (TMD) derived from human lysosomal associated protein 1 (LAMP1), to the N-terminal domain of γ-zein from *Zea mays*, or to the C-terminal transmembrane domain (tail-anchor, TA) of rabbit cytochrome b5 (cyt b5). The first attempt was to transform tobacco protoplasts to analyze the expression of the different recombinant p24 proteins. The constructs containing RFP or GFP with or without a TMD were also analysed using confocal microscopy to determine the intracellular localisation of the fusion proteins. The expression cassettes of some constructs were cloned into binary vectors to perform *Agrobacterium*-mediated transient expression. The selected constructs, six in total, were cloned into another binary vector and *Agrobacterium tumefaciens* was transformed to allow plant transformation and generate stable transgenic tobacco plants. Transgenic tobacco lines that accumulate the recombinant protein in different amounts were then selected to determine mRNA levels. Stability of the different recombinant p24 fusion proteins was analysed by pulse-chase labelling of leaf protoplasts isolated from transgenic plants and purification was achieved *in vitro* by digesting the fusion proteins with thrombin protease. Furthermore, the intracellular localisation of the constructs containing γ-zein and TA domains was determined by electron microscopy.

In this Results chapter, the structure of the different fusion proteins expressed in tobacco plants is illustrated followed by the analysis of the subcellular localisation of some of the constructs in transiently transformed tobacco protoplasts and their level of expression. Next the production of transgenic lines expressing six of the fusion proteins is detailed. The stability of the fusion proteins and a first purification of recombinant p24 from the transgenic lines are assessed, as well as a preliminary electron microscopy study.

I Design of the fusion proteins

In this thesis, different proteins and TMD were fused to the HIV-1 p24 to accumulate it in different membranes of the plant secretory pathway. In the first strategy, fluorescent proteins were fused to p24 in the constructs containing TMD derived from LAMP1 TMD and cytosolic and secreted versions were also prepared. These fluorescent proteins were introduced to allow the detection of the recombinant proteins expected in different cell compartments by confocal microscope analysis. In another strategy, the p24 protein was fused to the N-terminal domain of maize γ-zein to induce the accumulation of the recombinant proteins in ER-derived protein bodies. Finally, the p24 was fused to the cyt b5 TA to accumulate the recombinant protein in the cytosolic face of the ER. The fusions to γ-zein and TA did not contain any fluorescent marker.

To conduct all the experiments, a total of fifty-two constructs were prepared as described in more detail in *Materials and methods* section V. The final cloning vectors to transform tobacco protoplasts were pGY1 and pDHA. These plasmids contain the 35S CaMV promoter and terminator, an origin of replication in *E.coli* and an ampicillin resistance gene. Next are detailed the preparation of the different constructs.

I.1 Constructs containing XFP

I.1.1 Mutagenesis of RFP for RFP-TMD constructions (positive controls)

We first wanted to test if the different TMD fused to RFP followed the same pattern as previously described when fused to GFP in tobacco plants. Human LAMP1 is a type I membrane glycoprotein composed of a highly glycosylated luminal domain, a transmembrane domain and a short cytoplasmic tail. At steady state, LAMP1 is highly enriched in late endosomes and lysosomes. Previous work of Brandizzi et al. (2002), in which the original TMD and shortened versions of it were fused to GFP and expressed in tobacco plants, showed that the localisation of the fluorescent protein changed from ER to Golgi stacks and to the plasma membrane depending on the length of the TMD. Based on these results, we investigated whether these different versions of LAMP1 TMD could be used as tools to accumulate large quantities of recombinant proteins in different compartments of the plant secretory pathway.

DsRed, a protein from *Discosoma sp.*, is used as natural red chromophore. DsRed, as other fluorescent proteins, mature quite slowly and incomplete maturation gives rise to

residual green fluorescence. Moreover, it tends to aggregate in tetramers leading to cellular toxicity. Therefore we wanted to use another form of the protein which is stable and tolerated by the host cell when overexpressed. Monomeric RFP1 (mRFP1) derives from DsRed, in which a total of 33 mutations were introduced to increase ten times the maturation rate of the protein and obtain a monomeric fluorescent protein by breaking the subunit interactions while preserving fluorescence (Miyawaki et al. 2003). The 0.7 kb long mRFP coding sequence (CDS) was mutated using four primers in three separated PCR reactions as described in *Materials and methods* V.1.1. The purpose was to eliminate a *Hin*dIII site and introduce a *Nhe*I and *Nco*I sites at the 5' end, to eliminate an internal *Pst*I site and to introduce a *Pst*I and a *Sal*I sites at the 3' end (Figure 1A). The new mutated RFP sequence was verified by sequencing to ensure the new sequence did not contain any mutation. Different constructs were then prepared by fusing the new RFP to different TMD.

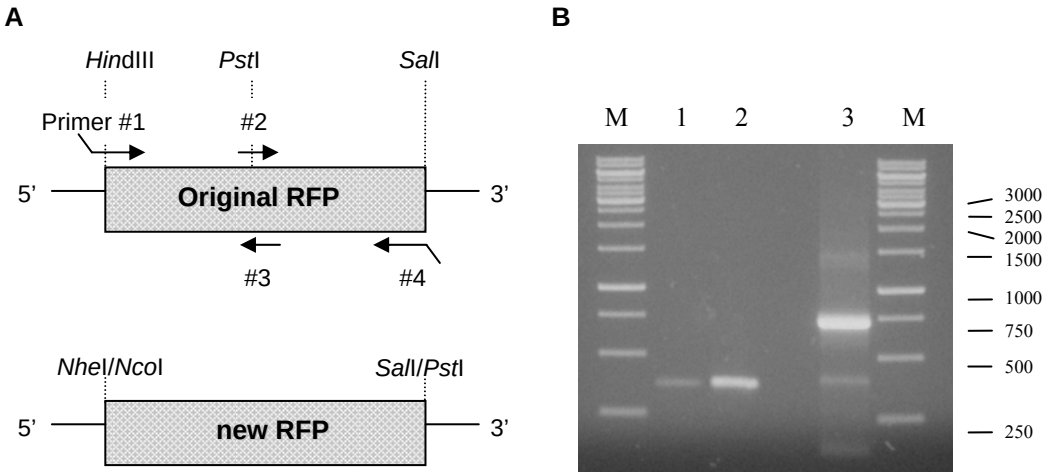
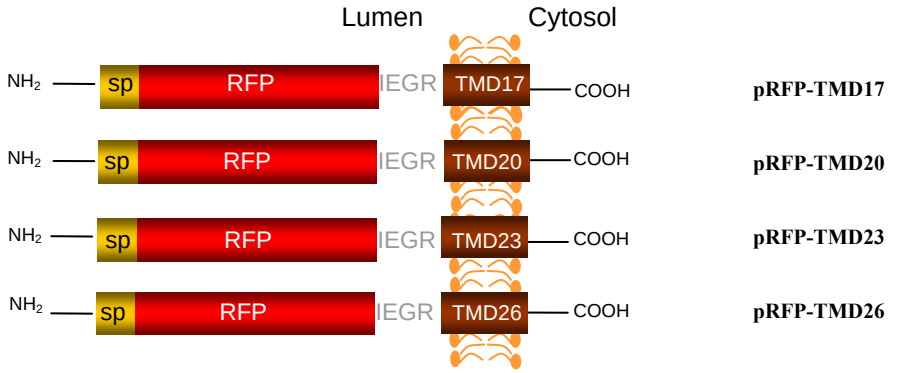


Figure 1. Site-directed mutagenesis of the mRFP CDS by overlap extension. A) Diagram of the original mRFP CDS with restriction sites mentioned in the text and the 4 primers used for the different. B) Diagram of the modified mRFP CDS with the new restriction sites mentioned in the text. C) Analysis of the different PCR fragments by agarose gel electrophoresis. Lane 1, 340 bp fragment of PCR1 product (primers #1 and 3); lane 2, 363 bp fragment of PCR 2 product (primers #2 and 4); lane 3, 700 bp full length mutated RFP (primers #1 and 4); M: Gene ruler 1 kb DNA ladder.

The original LAMP1 TMD, called TMD23, is 23 amino acids long and accumulated in the plasma membrane. In the TMD20 and TMD17 DNA sequences, 9 and 18 bases were deleted and the encoded amino acid sequences are accordingly shorter, 20 and 17 amino acids, and they accumulated in Golgi stacks and ER, respectively. Nine bases

were also added to the TMD23 DNA sequence resulting in a 26 amino acid long (TMD26) and is expected to accumulate in plasma membrane (Nadine Paris, unpublished).

In Figure 2 (upper panel) the different DNA constructs containing RFP and a TMD are represented. They were named pRFP-TMD17, pRFP-TMD20, pRFP-TMD23 and pRFP-TMD26 depending on their TMD. The DNA alignment shown in Figure 2 (lower panel) corresponds to the sequences encoding the TMD and emphasizes the position of the insertion or deletions. Silent point mutations that we identified in the TMD are highlighted in red.




```
TMD17           TATCGAAGGTAGAGAAGCTGAAGCTCTGATCCCCATCGCTGTGGGTGGCGCGCTAGCGGGG
TMD20           TATCGAAGGTAGAGAAGCTGAAGCTCTGATCCCCATCGCTGTGGGTGGCGCGCTAGCGGGG
TMD23           TATCGAAGGTAGAGAAGCTGAAGCTCTGATCCCCATCGCTGTGGGTGGCGCGCTTGCGGGG
TMD26           TATCGAAGGTAGAGAAGCTGAAGCTCTGATCCCCATCGCTGTGGGTGGCGCGCTAGCTGGC
                 I  E  G  R  E  A  E  A  L  I  P  I  A  V  G  G  A  L  A  G

TMD17           CTG--------------------------GCCTACCTCGTCGGCAGGAAGAGATCTTAA
TMD20           CTG------------------GTCCTCATCGCCTACCTCGTCGGCAGGAAGAGATCTTAA
TMD23           CTG---------GTCCTCATCGTCCTCATCGCCTACCTCGTCGGCAGGAAGAGATCTTAA
TMD26           CTTGTACTGATCGTCCTCATCGTCCTCATCGCCTACCTCGTCGGCAGGAAGAGATCTTAA
amino acids     L  V  L  I  V  L  I  V  L  I  A  Y  L  V  G  R  K  R  S  *
```

Figure 2. Schematic representation of the different RFP-TMD DNA constructs. The new mutated RFP was cloned into the pGY1 vector containing the signal peptide (sp) of tobacco chitinase. Between the RFP and the TMD there is a Factor Xa cleavage site (IEGR are the amino acids recognized by Factor Xa protease) for purification purposes. In the lower panel the DNA sequences of the different transmembrane domains are aligned. The TMD23 DNA sequence is originally from the LAMP1 TMD and the TMD17 and TMD20 contain deletions of the original sequence whereas TMD26 contains additional bases (grey). The red letters represent silent point mutations identified in the different sequences.

I.1.2 Mutagenesis of HIV p24 for insertion N-terminally to RFP and TMD

The p24 protein is derived from a precursor polyprotein of HIV, p55, that is cleaved into shorter polypeptides, namely p24 (core), p17 (matrix) and p7/p9 (nucleocapsid)

proteins. The p24 protein is synthesized in the cytosol as part of the GAG polyprotein and has an approximate size of 24 kDa. It forms the core of the virus and encapsulates the viral genomic RNA.

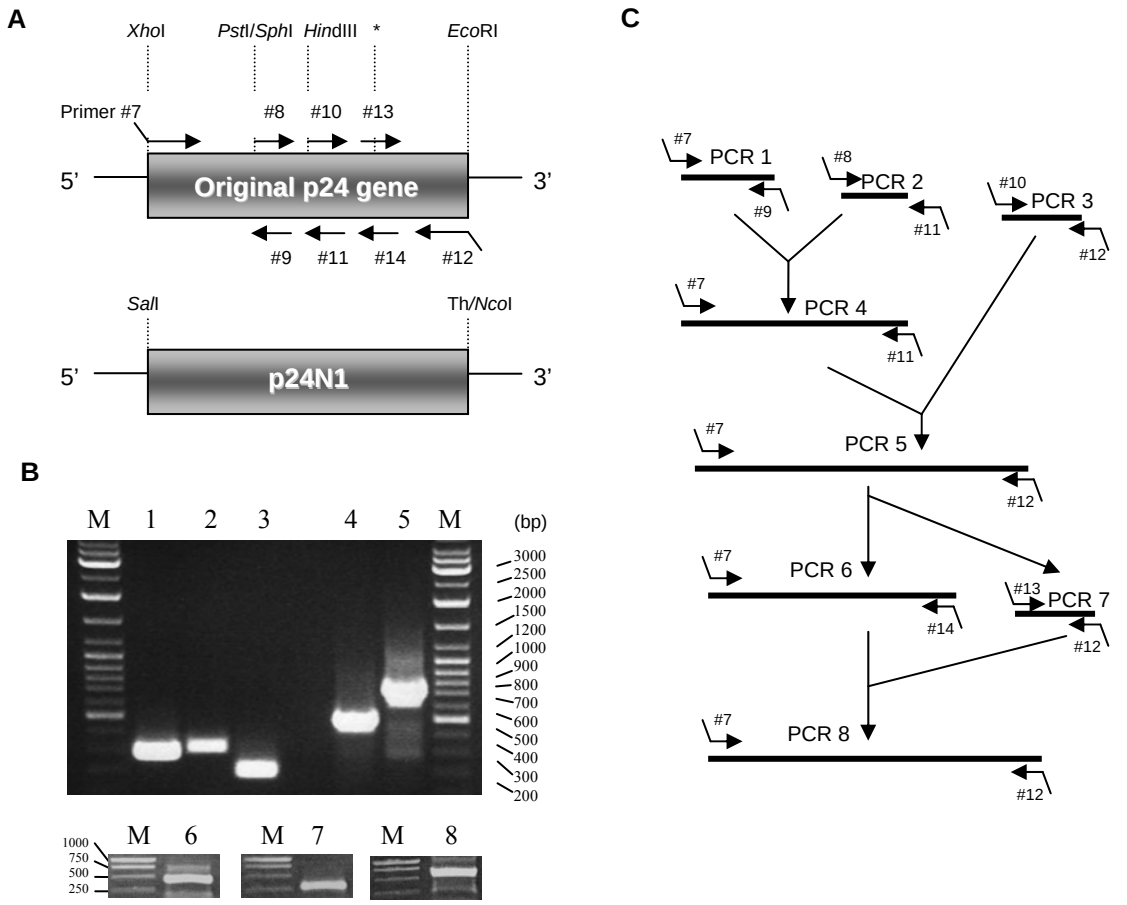


Figure 3. Mutagenesis of the original p24 gene to obtain p24N1. A) Diagram of site-directed mutagenesis by overlap extension of p24 CDS. The original p24 CDS is illustrated with the unwanted restriction sites and the 8 primers used for the different PCR. Below, the diagram shows the p24N1 CDS with the new restriction sites and the added sequence encoding the thrombin cleavage site (Th). B) Analysis of the different PCR fragments by agarose gel electrophoresis. Lane 1, PCR 1, 271 bp fragment; lane 2, PCR 2, 307 bp; lane 3, PCR 3, 192 bp; lane 4, PCR 4, 547 bp fragment; lane 5, first full length mutated p24, 720 bp; lane 6, correction PCR 6, 459 bp; lane 7, correction PCR 7, 283 bp; lane 8, final mutated p24N1, 724 bp. M: Gene ruler DNA ladder mix. C) Representation of the different mutagenic and assembly PCR reactions. Asterisk in A: point mutation.

Several changes were introduced into the original HIV-1 p24 gene in preparation for the different constructs. For a first mutated p24 CDS, eight primers were designed to (i) remove an *Xho*I site and introduce *Nhe*I and *Sal*I sites at the 5'end; (ii) remove internal *Sph*I, *Pst*I and *Hin*dIII sites; (iii) remove an *Eco*RI site, and (iv) introduce the sequence encoding the thrombin cleavage site followed by an *Nco*I site at the 3' end (Figure 3 A shows the restriction map before and after the mutagenesis). Mutagenesis by overlap

extension (Fig. 3C) was performed using a restriction fragment obtained from the original plasmid as template for the first reactions. The expected fragment sizes were obtained in each reaction (Fig 3B). Sequencing the end product revealed an unwanted mutation which had to be corrected by an additional mutagenesis step (Figure 3 A and C). The first full length mutated p24 CDS (produced by PCR 5) was cloned into pGEM®. Sequencing revealed 4 unwanted point mutations when comparing with the original p24 sequence. The original p24 clone used as template DNA may have already contained these mutations. Three of these four mutations were silent. Therefore, only the fourth point mutation was corrected by site-specific mutagenesis (reactions PCR 6, 7 and 8, Fig 3 C).

The new sequence was called p24N1 and was used for the constructs in which the RFP and the different TMDs were fused to the C-terminus of p24 (details in *Materials and methods* V.2; Figure 4). Between the p24 and the RFP proteins a thrombin cleavage site was introduced. Moreover, between the RFP and the TMD there is a Factor Xa cleavage site that was already in the TMD constructs. Both sites were introduced to cleave p24 from the fusion proteins in the final purification steps.

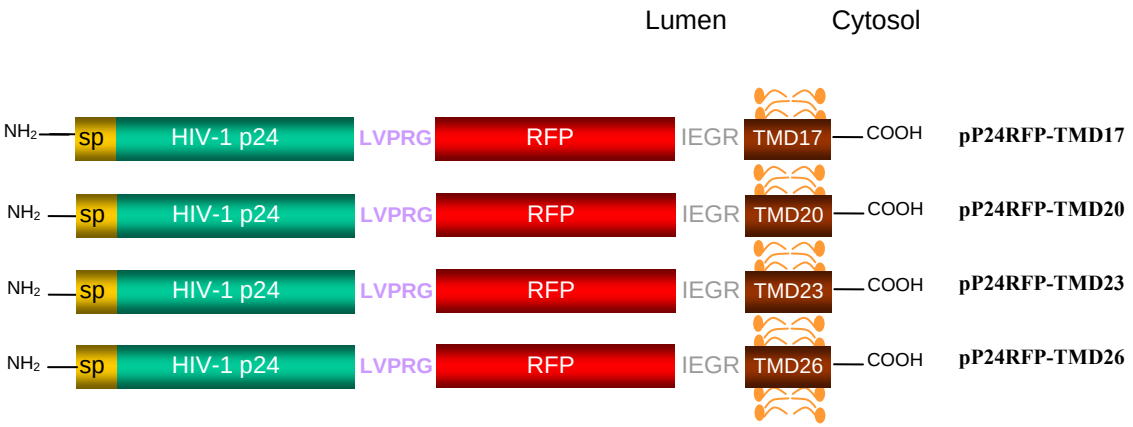


Figure 4. Schematic representation of the fusion proteins (p24RFP-TMD) expressed in tobacco protoplasts. SP: signal peptide; LVPRG represents the thrombin protease cleavage site; IEGR represents the Factor Xa protease cleavage site.

I.1.3 p24 fused to C-terminus of RFP and TMD

To clone the p24 gene at the C-terminus of RFP additional mutations were necessary, as restriction sites and an additional thrombin cleavage site had to be introduced into the p24N1 CDS: an *Xho*I and a thrombin cleavage sites were introduced at the 5' end and a

*Sal*I site was introduced at the 3' end. This new p24 was called p24N2. The *Xho*I and *Sal*I sites are both compatible with the *Sal*I site of the vector. As the insert could be inserted in both orientations, bacterial colonies were screened by colony PCR to identify the correct orientation of the gene. p24N2 was inserted between the RFP and the TMDs (Figure 5). The details of the cloning steps are described in *Materials and methods* V.3.

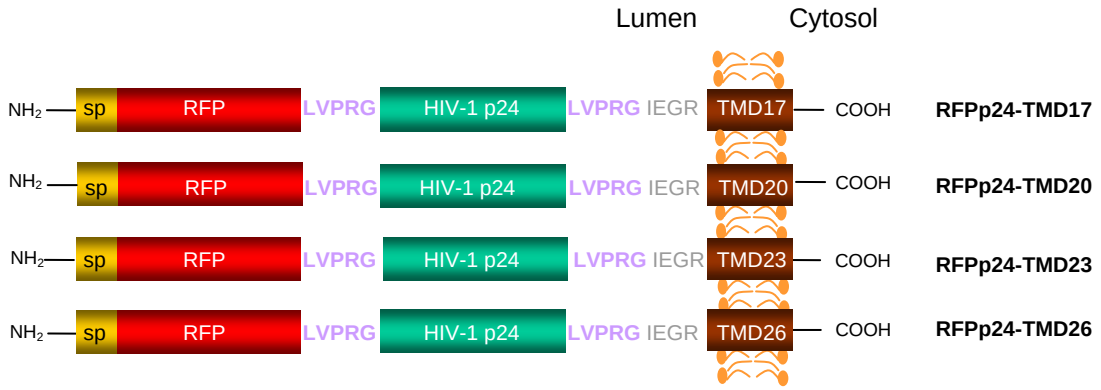


Figure 5. Schematic representation of the fusion proteins (RFPp24-TMD) expressed in tobacco protoplasts. Two thrombin cleavage sites were introduced flanking the p24 gene for purification purposes.

I.1.4 Other constructs

Additional TMD constructs were prepared after the first results with the RFP-containing constructs were obtained and evaluated (section II). RFP was replaced by GFP, and secreted and cytosolic p24RFP and RFPp24 were also produced. A schematic representation of the constructs is shown in Figure 6.

I.1.4.1 Zein and cytochrome b5 tail anchor constructions

In the constructs described above, p24 is part of a transmembrane protein and is exposed into the lumen of the compartments of destination. Two additional strategies were tested for the accumulation of p24 in the secretory pathway. The first one involves fusion with a domain of maize γ-zein. This seed storage protein is synthesized in maize endosperm during seed development and accumulates as part of large polymers termed protein bodies (PB) in the lumen of the ER. γ-zein is fully luminal and forms PB also when expressed in transgenic plant vegetative tissues in the absence of other maize storage proteins. The N-terminal domain of γ-zein that we used here has been

previously shown to promote the formation of PB when fused to the vacuolar storage protein phaseolin (Mainieri et al. 2004). Our hypothesis was that fusion of the same zein domain to p24 could also promote stable PB formation, leading to high accumulation of recombinant p24 (Fig. 7).

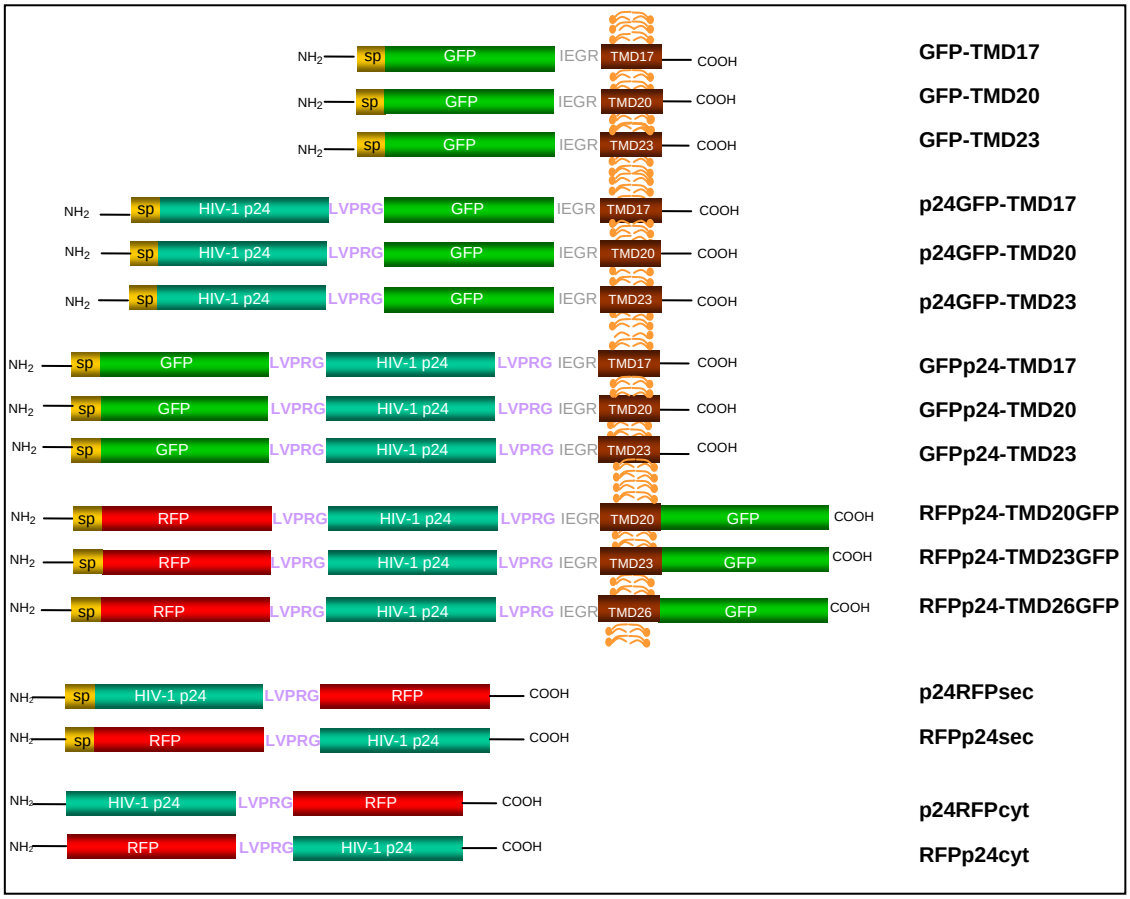


Figure 6. Representation of additional fluorescent protein p24 constructs. The p24 gene was also fused to different fluorescent proteins containing or not a TMD. The signal peptide (SP) is the tobacco chitinase SP.

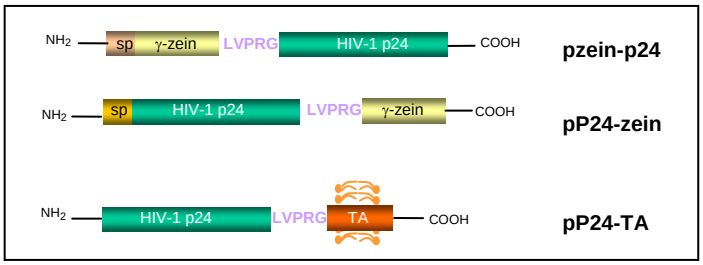


Figure 7. Representation of zein and TA constructs. The p24 was fused to the C-terminus or N-terminus of the N-terminal domain of γ-zein called pzein-p24 and pP24-zein, respectively. The signal peptide (SP) of pzein-p24 is from γ-zein. The p24 was also fused to the C-terminal domain of rabbit cyt b5 called pP24-TA.

The second strategy was devised to promote stable attachment of p24 to the cytosolic face of the ER membrane by the addition, at the C-terminus, of the TA of rabbit cyt b5. Like the other TA proteins (also called type IV integral membrane proteins), cyt b5 is post-translationally targeted to the membrane (in this case the ER) due to a transmembrane domain that is very close to its C-terminus. This domain is followed by a few C-terminal amino acids that in the final topology are luminal, whereas most of the protein remains exposed to the cytosol. The addition of a TA strategy could extend the half-life of p24 with respect to its synthesis as a soluble cytosolic protein but, contrarily to other strategies used here, would maintain its folding process in the cytosol, which is the natural folding environment of p24 (Figure 7). The different constructs are also listed in table I with the expected localisation according the fused sequences.

Construct name	Fluorescent protein	Expected localisation
Pzein-p24	None	ER protein bodies
pP24-zein	None	ER protein bodies
pP24-TA	None	ER (cytosolic)
pRFP-TMD17	RFP	ER (lumen)
pRFP-TMD20	RFP	Golgi (lumen)
pRFP-TMD23	RFP	PM (apoplast)
pRFP-TMD26	RFP	PM (apoplast)
pP24RFP-TMD17	RFP	ER (lumen)
pP24RFP-TMD20	RFP	Golgi (lumen)
pP24RFP-TMD23	RFP	PM (apoplast)
pP24RFP-TMD26	RFP	PM (apoplast)
pRFPp24-TMD17	RFP	ER (lumen)
pRFPp24-TMD20	RFP	Golgi (lumen)
pRFPp24-TMD23	RFP	PM (apoplast)
pRFPp24-TMD26	RFP	PM (apoplast)
GFP-TMD17	GFP	ER (lumen)
GFP-TMD20	GFP	Golgi (lumen)
GFP-TMD23	GFP	PM (apoplast)
P24GFP-TMD17	GFP	ER (lumen)
P24GFP-TMD20	GFP	Golgi (lumen)
P24GFP-TMD23	GFP	PM (apoplast)
GFPp24-TMD17	GFP	ER (lumen)
GFPp24-TMD20	GFP	Golgi (lumen)
GFPp24-TMD23	GFP	PM (apoplast)
RFPp24TMD20-GFP	RFP and GFP	Golgi (lumen)
RFPp24TMD23-GFP	RFP and GFP	PM (apoplast)
RFPp24TMD26-GFP	RFP and GFP	PM (apoplast)
P24RFPsec	RFP	Secreted
RFPp24sec	RFP	Secreted
P24RFPcyt	RFP	Cytosol
RFPp24cyt	RFP	Cytosol

Table I. Hybrid protein constructs, their fluorescent proteins and their expected localisation

I.2 Prediction of the signal peptide cleavage

Signal peptides are short transient sequences located at the N-terminus of a nascent secretory or membrane protein. They contain the necessary information to direct proteins to the ER and are usually cleaved off co-translationally by signal peptidase. When a nascent protein contains an ER signal peptide, it is translocated across the ER membrane during translation and folded and modified there. Unless there is a specific signal for ER retention or retrieval, the protein will then travel along the secretory pathway. A soluble protein will be eventually secreted, unless it has another sorting signal to direct it to another compartment of the endomembrane system, such as a vacuole, or has a retrieval system for ER localisation. A membrane anchored protein will travel along the pathway towards the plasmalemma or the tonoplast but its final residence membrane also depends on the length of its TMD (Brandizzi et al. 2002).

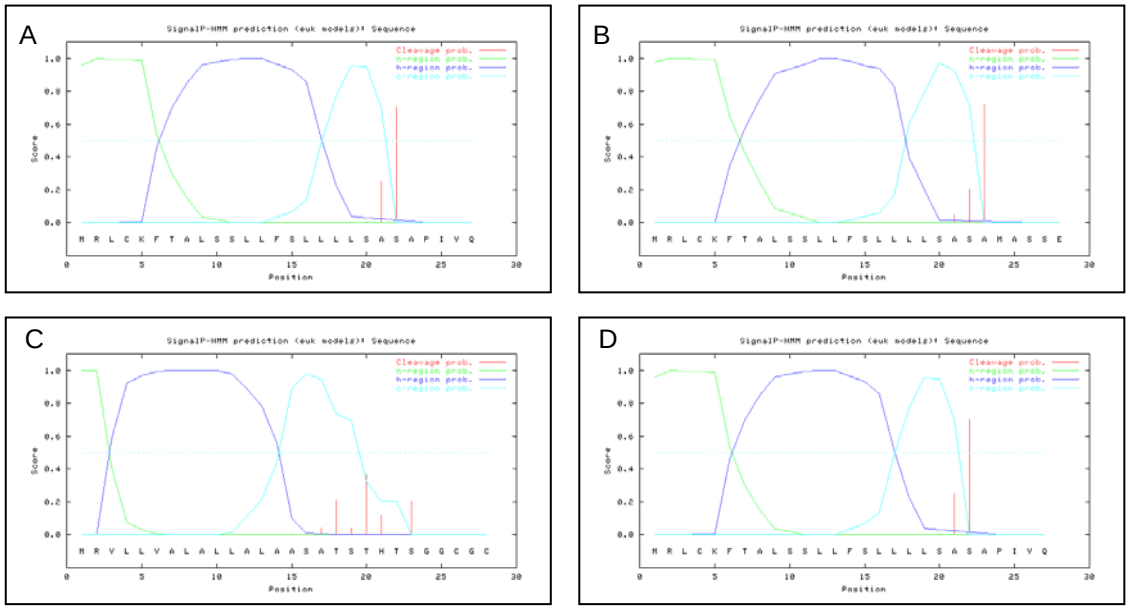


Figure 8. SignalP-HMM output analysis of tobacco chitinase and γ-zein signal peptides in the pP24RFP-TMD, pRFPp24-TMD, pzein-p24 and pP24-zein constructs. The output shows the n-region of the signal peptide in green, the h-region in purple and the c-region in blue. The most probable cleavage sites are shown in red: Alanine (A) and Proline (P) residues in pP24RFP-TMD constructs (panel A); Serine (S) and Glutamic acid (E) in pRFPp24TMD constructs (panel B); Threonine (T) or Serine (S) residues in the pzein-p24 construct (panel C) and Alanine (A) and Proline (P) residues in the pP24-zein construct (panel D).

It is useful to check the reliability of the signal peptide used in cloning strategies to avoid problems due to possible lack of cleavage. Several algorithms have been developed to predict the cleavage probability of signal peptides. Using SignalP

software, the protein sequence of pP24RFP-TMD17, pRFPp24-TMD17, pzein-p24 and pP24-zein constructs were used to predict the cleavage site of the tobacco chitinase and γ-zein signal peptides (Figure 8). The SignalP-HMM output shows a coloured graph for the n-region (positively charged region at N-terminus of the signal peptide), the h-region (hydrophobic region in the middle) and the c-region (more polar region at the C-terminus) and also marks in red the region were the cleavage is most probable occurring.

The program predicted that the cleavage of the tobacco chitinase signal peptide in the pP24RFP-TMD constructs and the pP24-zein construct would most probably occur after the amino acids Ala or Pro (Fig. 8 panels A and D). However, in the constructs in which the RFP follows immediately the signal peptide, the pRFPp24-TMD constructs, the cleavage of the tobacco chitinase signal peptide would occur after the Ser or Glu (Fig. 8 panel B). The pzein-p24 construct contains the natural γ-zein signal peptide and for this construct the cleavage is predicted to occur either after the Thr or the Ser (Fig.8 panel C). In all cases the signal peptides are predicted to be cleaved off from the fusion protein.

I.3 Construction of recombinant binary vectors

To transiently or stably transform tobacco plants with *Agrobacterium*, it is necessary to clone the sequences of interest into expression cassettes in a binary vector. For our studies, the RFP-TMD cassettes with or without the p24 CDS (twelve constructs), the two cassettes containing the γ-zein -p24 hybrids (zein-p24 and p24-zein) and the cassette containing the p24-TA hybrid (p24-TA), were cloned into the pGREEN0229 and/or pGREEN0179 binary vectors. These vectors contain two origins of replication, one for *E.coli* (ColEI), one for *Agrobacterium tumefaciens* (pSa), and a bacterial kanamycin resistance gene. The T-DNA region of pGREEN0229 contains the *bar* gene for plant resistance to ammonium glufosinate (BASTA®) under the control of the nopaline synthetase (nos) promoter. The pGREEN0179 contains a hygromycin resistance gene under the control of the 35S CaMV promoter.

When the blue/white (Xgal/IPTG) selection is used to identify clones containing an insert in the vector, false positives can be found. Therefore, a quick method to detect real positive recombinant *E.coli* colonies containing the new recombinant pGREEN0179 vectors was used (Figure 9), the cracking gel (*Materials and methods*

III.7). Putative positive recombinant *E.coli* colonies (white colonies) were picked together with a control colony (blue colony), lysed and loaded to an agarose gel. A difference in plasmid size could be appreciated in the photograph of the gel as the control blue colony contains the empty plasmid (the ligation did not work out). The positive colonies were re-confirmed by regular miniprep isolation of the binary vector and restriction digestion analysis.

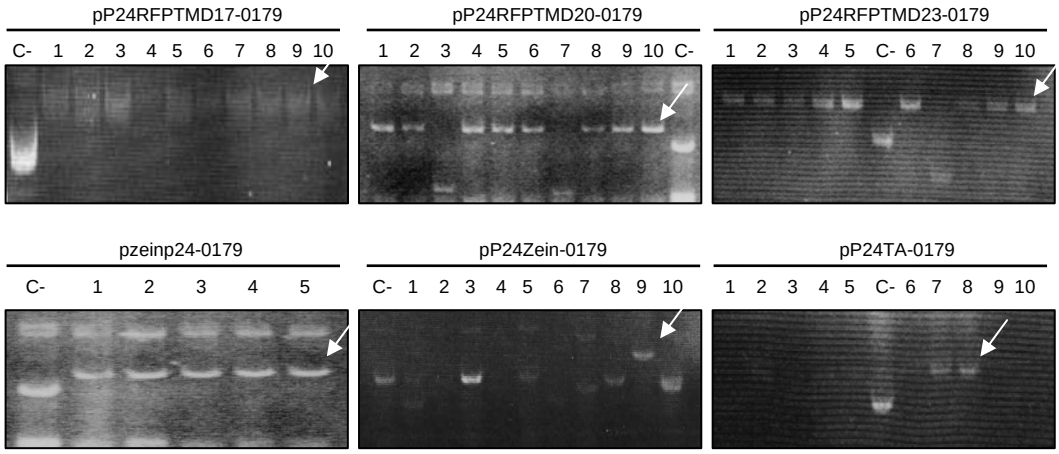


Figure 9. Cracking gel analysis of white *E.coli* colonies transformed with the ligation mixture of pGREEN0179 plasmid and the six different cassettes for plant transformation. Putative recombinant white colonies of pP24RFPTMD17-0179, pP24RFPTMD20-0179, pP24RFPTMD23-0179, pzeinp24-0179, pP24zein-0179 and pP24TA-0179 run on an agarose gel. C-: negative control corresponding to a blue colony. Upper bands compared to the negative control are putative positives (arrows).

After confirmation of the correct cloning in the different binary plasmids, all the new recombinant binary vectors were transformed into *Agrobacterium tumefaciens* GV3101 strain. This strain has chromosomal resistance to rifampicin (Rif) and already contained the T-helper plasmid, which confers resistance to gentamycin (Gen), and had been transformed with the disarmed Ti plasmid pSOUP which confers tetracyclin (Tet) resistance. The pGREEN vectors conferred Kan resistance to the bacteria, therefore when *A.tumefaciens* GV3101 was transformed, the recombinant colonies were selected on a quadruple selection medium (Rif, Gen, Tet and Kan). Recombinant *cultures* of each construct were prepared from single colonies and used for transient expression in *N.benthamiana* plants and stable transformation of *N.tabacum* plants. To trigger the expression of the *vir* genes, acetosyringone which is a monocyclic phenolic inducer, was added to the agroinfiltration medium for transient expression.

The different recombinant Agrobacteria used for transient agroinfiltration are listed in table II and the recombinant Agrobacteria used for stable transformation of tobacco plants are listed in table III.

Recombinant binary construct name	Binary vector backbone	Cassette from	Selection in bacteria	Selection in plants
pGRFP-TMD17	pGREEN0229	pRFP-TMD17	Kan	glufosinate
pGRFP-TMD20	pGREEN0229	pRFP-TMD20	Kan	glufosinate
pGRFP-TMD23	pGREEN0229	pRFP-TMD23	Kan	glufosinate
pGRFP-TMD26	pGREEN0229	pRFP-TMD26	Kan	glufosinate
pGP24RFP-TMD17	pGREEN0229	pP24RFP-TMD17	Kan	glufosinate
pGP24RFP-TMD20	pGREEN0229	pP24RFP-TMD20	Kan	glufosinate
pGP24RFP-TMD23	pGREEN0229	pP24RFP-TMD23	Kan	glufosinate
pGP24RFP-TMD26	pGREEN0229	pP24RFP-TMD26	Kan	glufosinate
pGRFPp24-TMD17	pGREEN0229	pRFPp24-TMD17	Kan	glufosinate
pGRFPp24-TMD20	pGREEN0229	pRFPp24-TMD20	Kan	glufosinate
pGRFPp24-TMD23	pGREEN0229	pRFPp24-TMD23	Kan	glufosinate
pGRFPp24-TMD26	pGREEN0229	pRFPp24-TMD26	Kan	glufosinate
pGzein-P24	pGREEN0229	pzein-P24	Kan	glufosinate
pGP24-zein	pGREEN0229	pP24-zein	Kan	glufosinate
pGP24-TA	pGREEN0229	pP24-TA	Kan	glufosinate

Table II. Description of the different recombinant binary constructs used for *Agrobacterium*-mediated transient expression.

Recombinant binary construct name	Binary vector backbone	Cassette from	Selection in bacteria	Selection in plants
pGP24RFP-TMD17	pGREEN0229	pP24RFP-TMD17	Kan	glufosinate
pGP24RFP-TMD20	pGREEN0229	pP24RFP-TMD20	Kan	glufosinate
pGP24RFP-TMD23	pGREEN0229	pP24RFP-TMD23	Kan	glufosinate
pGzein-P24	pGREEN0229	pzein-P24	Kan	glufosinate
pGP24-zein	pGREEN0229	pP24-zein	Kan	glufosinate
pGP24-TA	pGREEN0229	pP24-TA	Kan	glufosinate
pP24RFPTMD17-0179	pGREEN0179	pP24RFP-TMD17	Kan	Hyg
pP24RFPTMD20-0179	pGREEN0179	pP24RFP-TMD20	Kan	Hyg
pP24RFPTMD23-0179	pGREEN0179	pP24RFP-TMD23	Kan	Hyg
Pzeinp24-0179	pGREEN0179	pzein-P24	Kan	Hyg
pP24zein-0179	pGREEN0179	pP24-zein	Kan	Hyg
pP24TA-0179	pGREEN0179	pP24-TA	Kan	Hyg

Table III. Description of the different recombinant binary constructs containing the six cassettes of interest used for stable plant transformation.

II Characterization of the XFP fusion proteins by transient expression

II.1 Analysis of the subcellular distribution of XFP fusion proteins in transformed tobacco protoplasts and in agroinfiltrated leaves

Transient expression of recombinant proteins allows a first test of their subcellular localisation and accumulation *in vivo* before producing stable transformed plants. To study the subcellular localisation of the fusion proteins containing RFP and TMD, wild-type *N.tabacum* cv Petit Havana mesophyll protoplasts were transformed with the different constructs by DNA electroporation, and *N.benthamiana* plants were transformed using recombinant *Agrobacteria.* Tobacco protoplasts were transfected either with the single constructs alone or co-transfected with the soluble marker for the ER lumen GFP-HDEL, the membrane-bound *trans*-Golgi marker ST-YFP (Sialyl transferase-YFP) or with the membrane-bound pre-vacuolar compartment (PVC) marker GFP-BP80 (GFP-Binding Protein 80). The transformed cells were imaged using a confocal scanning microscope.

II.1.1 The RFP fluorescence is localised in the expected compartments in the RFP-TMD constructs

Brandizzi et al. (2002) reported that, in GFP fusions, the shorter TMD of 17 amino acids, TMD17, led to ER retention in tobacco epidermal cells whereas the TMD20 accumulated in Golgi stacks and the TMD23 accumulated in the plasma membrane. We investigated whether this differential retention also occurs when the same TMD were fused to RFP and the constructs were transiently expressed in tobacco protoplasts by electroporation and in tobacco plants by agroinfiltration.

Tobacco mesophyll protoplasts were electroporated with pRFP-TMD17, pRFP-TMD20, pRFP-TMD23 or pRFP-TMD26 DNA constructs and 24 hours after transformation the cells were observed with a confocal microscope. The RFP fluorescence was predominantly found in the expected compartment (Figure 10). Confocal images from protoplasts expressing RFP fused to TMD17 (pRFP-TMD17 construct) presented a reticular pattern corresponding to ER labelling, where intense labelling of the nuclear envelope, which is part of the ER, is clearly detected (Fig. 10 A arrow). Also some punctate structures were visible suggesting either Golgi labelling by a proportion of

protein that might have escaped the retention or concentration of the recombinant protein at particular locations in the ER cisternae. In the protoplasts transformed with pRFP-TMD20, in which Golgi accumulation is expected, the RFP fluorescence was present in punctate structures scattered in the cytoplasm. These structures were mobile as expected for Golgi (data not shown). The RFP fluorescence in both pRFP-TMD23 and pRFP-TMD26 was predominantly found in the plasma membrane (Fig. 10 C-D). Some punctate structures were also observed in the vicinity of the plasma membrane suggesting partial localisation in the Golgi or in endosomes.

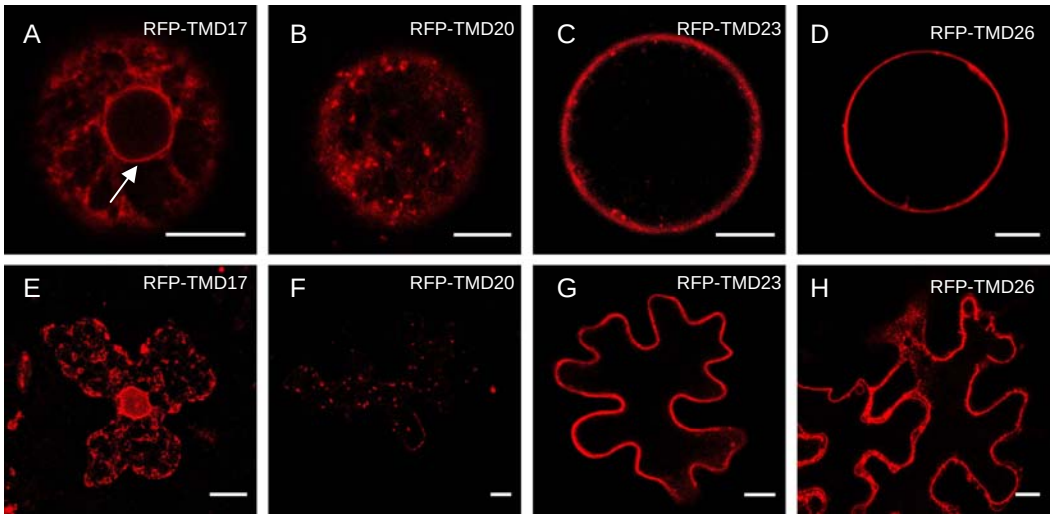


Figure 10. Intracellular localisation of RFP-TMD proteins in tobacco protoplasts (A-D) and in agroinfiltrated leaves (E-H). *N.tabacum* protoplasts transformed with the different pRFP-TMD DNA constructs were scanned through a Zeiss confocal microscope LSM510META equipped with laser for 543-nm excitation. RFP-TMD17 accumulated in ER and in the nuclear envelope (arrow) (A), RFP-TMD20 in Golgi stacks (B) and RFP-TMD23 and RFP-TMD26 in the plasma membrane (C-D). Lower panels show CLSM images of *N.benthamiana* leaves infiltrated with the corresponding cultures of agrobacteria (E-H). The epidermal infiltrated leaves harvested at 2-3 days post inoculation (d.p.i.) showed the same intracellular localisation of the fluorescent proteins. Scale bars correspond to 10 µm for A-D and to 20 µm for E-H.

Recombinant Agrobacteria containing the same cassettes in the pGREEN0229 vector were inoculated into *N.benthamiana* leaves and fluorescence was examined between 48 to 72 hours post inoculation to determine whether the pattern of fluorescence observed in tobacco leaves resembled the pattern reported in protoplasts.

Fluorescence was observed in the same expected compartments (Fig. 10 E-H) but sometimes it was difficult to observe as clear a pattern i.e. in the pGRFP-TMD17 in which the ER labelling is shown in a 3D image stack (Fig. 10 E). Moreover, sometimes

the time needed to be able to observe fluorescence differed in one or two days, as the fusion proteins required more time to be expressed and become visible.

With the aim to better study the association of the different recombinant fluorescent proteins with the previously visualized compartments, tobacco protoplasts were co-transfected with the different RFP-TMD constructs and with endomembrane compartment markers. As an ER marker, the chimeric soluble construct carrying the GFP fused to the ER retrieval signal HDEL (GFP-HDEL) was used. The pattern of RFP-TMD17 fluorescence and the pattern of GFP-HDEL fluorescence were closely comparable and superposition of the images demonstrated major overlap between the two fluorescent proteins (Fig. 11 A). The conclusion was that RFP-TMD17 is targeted to the ER as expected. The ST-YFP is a chimeric construct containing the spectral variant YFP and the transmembrane domain of the Golgi enzyme rat sialyl transferase. The fusion protein labels the *trans*-Golgi stacks. The protoplasts presented a punctated pattern (Fig. 11 B) and the labelled compartments were mobile (data not shown). In this case there was no colocalization with the RFP-TMD17. In a few cases colocalization could be observed when the protoplasts were co-transfected with the PVC marker GFP-BP80 (Fig. 11C). This construct is a chimera between GFP and the transmembrane domain of BP80 which is a plant vacuolar sorting receptor and is mainly located in the Golgi complex and endosomes that constitute the prevacuolar compartment (PVC).
As shown in Figure 11 D-F, when the RFP-TMD20 was co-transfected with the different markers, the red fluorescence was localised in punctate structures distributed all over the cytoplasm. There was an almost perfect colocalization with the yellow fluorescence of ST-YPF corresponding to Golgi labelling (Fig. 11E). In contrast, there was no colocalization with the ER marker, as expected (Fig. 11D) but there was some co-localisation with the PVC marker (Fig. 11F). These results suggest that some of the recombinant protein is correctly targeted to the Golgi apparatus whereas a part of the recombinant proteins en route to the vacuole was detected in the PVC. Moreover, there was some plasma membrane labelling suggesting some recombinant protein continue travelling the secretory pathway exiting Golgi (Fig. 11 D).

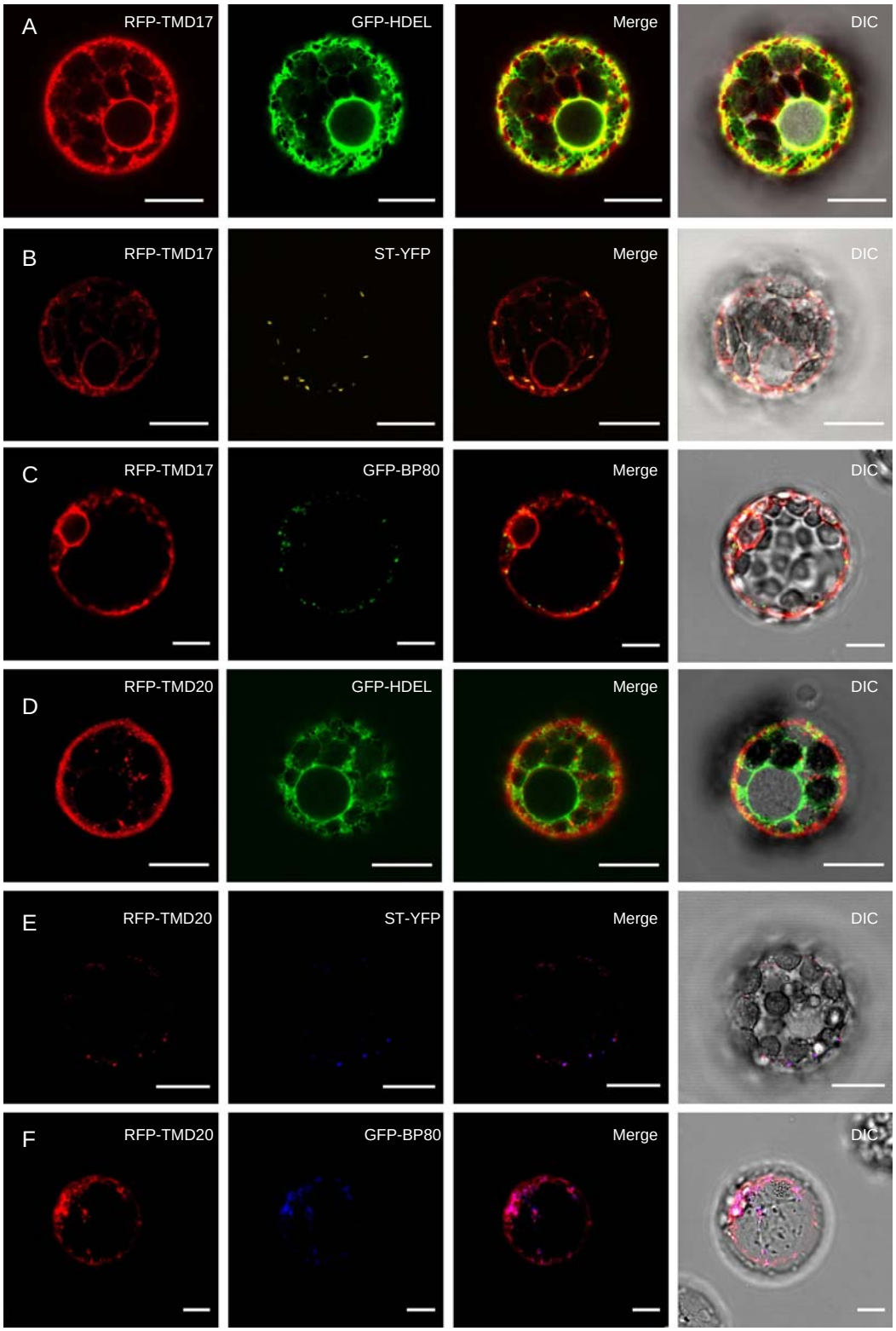


Figure 11. Intracellular localisation of RFP-TMD17 and RFP-TMD20 compared to different co-transformed markers in tobacco protoplasts. Single optical sections were observed by confocal microscopy. The first two images show the single-channel fluorescence, the third the merged images and the right image includes the image taken by differential interference contrast (DIC). Transient expression experiment showing the subcellular localisation of co-transfected pRFP-TMD17 or pRFP-TMD20 constructs and GFP-HDEL (A, D), ST-YFP (B, E) and GFP-BP80 (C, F), respectively. The localisation

of the reporter proteins was observed 24 h after transformation. Notice the almost complete co-localisation of the RFP-TMD17 and GFP-HDEL showing an ER network, and the p24RFP-TMD20 and ST-YFP in a punctate pattern. Scale bars correspond to 10 µm.

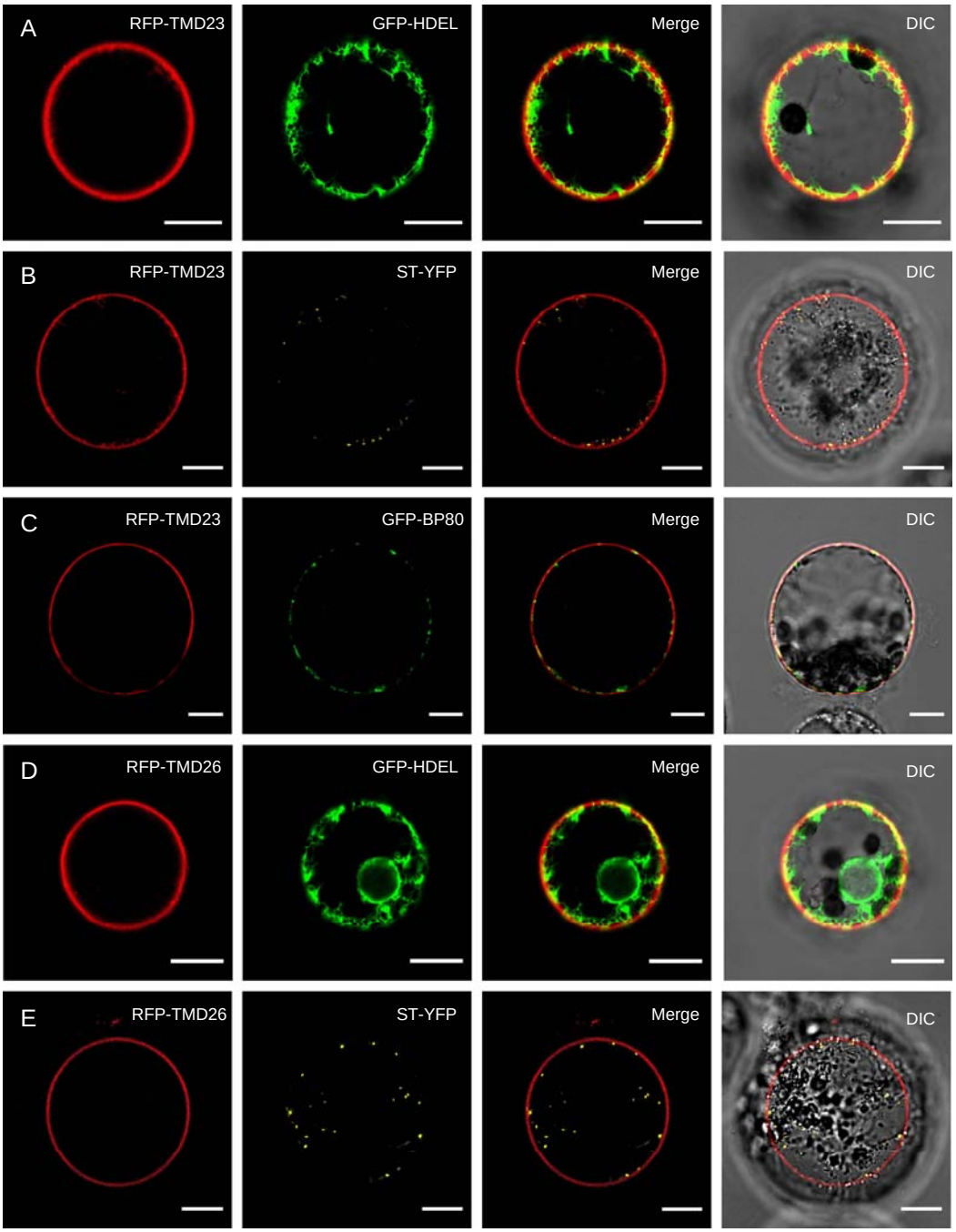


Figure 12. RFP fusion proteins containing TMD23 and TMD26 labelled plasma membrane in tobacco protoplasts. RFP fusion proteins were transiently expressed in tobacco protoplasts, and their localisation was visualized by confocal microscopy 24 h post-transformation. The different images represent the fluorescent signal associated with expression of RFP fusions containing TMD23 (A-C) or TMD26 (D-E). Images A and D represent the fluorescence signals associated with the co-expression of RFP-TMD23 and RFP-TMD26 with GFP-HDEL, an ER marker. Images B and E represent the fluorescence signals associated with the co-expression of RFP-TMD23 and RFP-TMD26 with ST-YFP, a

Golgi marker. Image C represents the fluorescence signals associated with the co-expression of RFP-TMD23 with GFP-BP80, a PVC marker. Bars represent 10 µm.

Tobacco protoplasts transiently expressing RFP-TMD23 and RFP-TMD26 showed clear fluorescence at the plasma membrane (Fig. 12). No co-localisation was observed with either constructs and the ER marker protein GFP-HDEL or the *trans*-Golgi marker ST-YFP (Fig. 12 A, B, D, E). Moreover, the RFP-TMD23 did not colocalise with the PVC marker GFP-BP80 (Fig. 12 C). Some red fluorescence was occasionally detected in punctate structures in the vicinity of the plasma membrane probably corresponding to proteins transiting the Golgi en route to the plasma membrane.

II.1.2 The C-terminally RFP-tagged HIV-1 p24 constructs are targeted to the expected compartments

Tobacco protoplasts and *N.benthamiana* leaves were transformed with the different pP24RFP-TMD constructs. In all cases, the recombinant fluorescent proteins showed the expected intracellular localisation, similarly to the RFP-TMD proteins (Fig. 13).

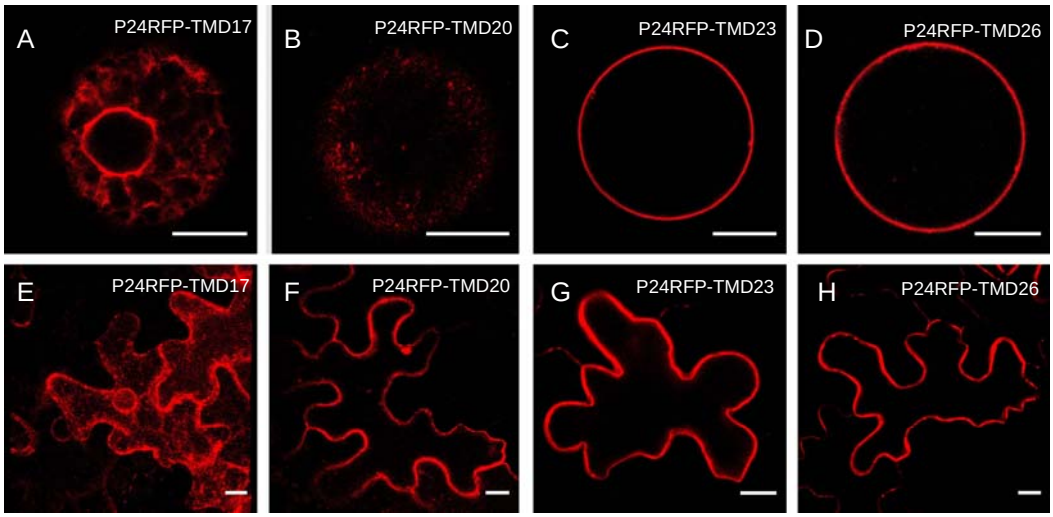


Figure 13. Intracellular localisation of p24RFP-TMD fusion proteins in tobacco protoplasts and in agroinfiltrated leaves. Upper panels show CLSM images of RFP fluorescence in tobacco mesophyll protoplasts transformed with the different pP24RFP-TMD DNA constructs. p24RFP-TMD17 accumulated in ER (A), p24RFP-TMD20 in Golgi stacks (B) and p24RFP-TMD23 and p24RFP-TMD26 in the plasma membrane (C-D). The lower panels show epidermal cell images from *N.benthamiana* leaves infiltrated with the indicated cultures of Agrobacteria (E-H). The infiltrated leaves were harvested 3 d.p.i. and visualized with a Zeiss Axiovert LSM510 Meta microscope with excitation at 543 nm and emission at 596–636 nm (RFP). P24RFP-TMD20 also showed some plasma membrane labelling (F). The scale bars correspond to 10 µm (A-D) and to 20 µm (E-H).

As shown in Fig.13 A-D, the p24RFP-TMD17 accumulated in ER showing a reticular pattern with intense labelling of the nuclear envelope in protoplasts, the p24RFP-TMD20 accumulated in punctate structures corresponding to Golgi stacks and some in the plasma membrane, and p24RFP-TMD23 and p24RFP-TMD26 accumulated in the plasma membrane.

The sub-axial epidermis of *N.benthamiana* leaves were infiltrated with *Agrobacterium* suspensions carrying the pGp24RFP-TMD vectors and the same results were obtained 3 d.p.i. as for the protoplasts (Fig.13 E-H) but there was some plasma membrane labelling for agroinfiltrated p24RFP-TMD20 (Fig. 13 F).

Tobacco protoplasts were co-transformed with the different pP24RFP-TMD DNA constructs and the endomembrane compartment markers described before. In representative images the different p24RFP-TMD proteins are shown in red, the organelle-specific markers in green, yellow or blue, and the overlapped images in yellow, orange or purple (Fig. 14).

All the fusion proteins presented an intracellular fluorescence distribution in compartments with some obvious fluorescence of GFP in the nucleus in protoplasts transformed with p24RFP-TMD17 (Fig. 14 A). This labelling is possibly due to GFP that failed to undergo translocation into the ER because of saturation of translocons, and entered the nucleus by diffusion through the nuclear pore complex as it has been described for cytosolic GFP (Berg and Beachy 2008). The p24RFP-TMD17 fusion protein co-localised with the ER marker GFP-HDEL, apart from the above mentioned nuclear localisation of the latter. The reticular pattern was clearer with GFP than with RFP but the merged image showed an almost complete co-localisation. By contrast, the p24RFP-TMD17 protein did not colocalise with the Golgi marker ST-YFP but there was some localisation with the PVC marker GFP-BP80 (Fig. 14 B-C).

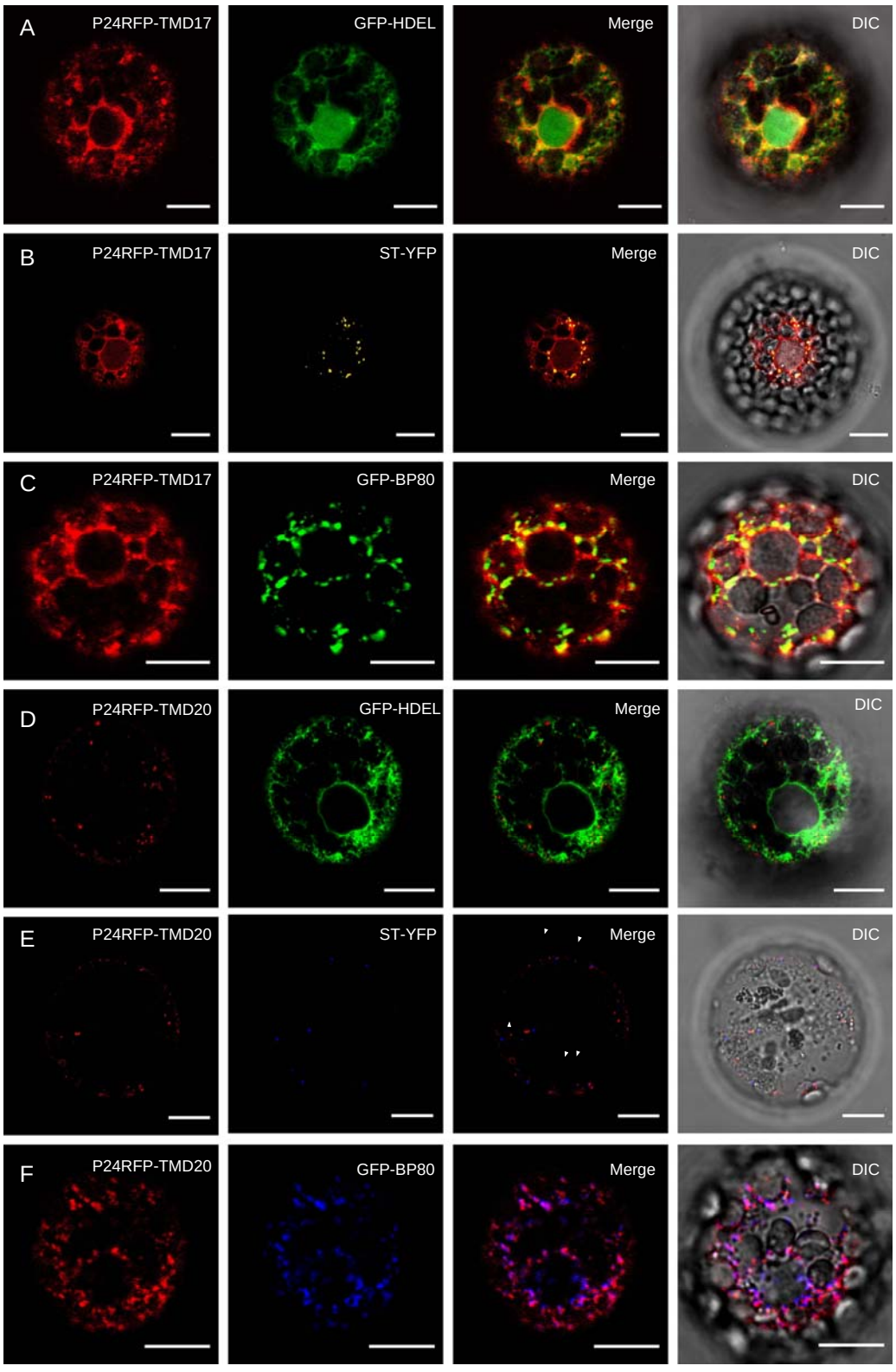


Figure 14. Fluorescent distribution of p24RFP-TMD17 and p24RFP-TMD20 fusion proteins co-transformed with different markers in tobacco protoplasts. RFP was detected in the red channel, GFP in the green channel and YFP in the yellow channel. p24 fused to the N-terminus of RFP containing the TMD17 which confers ER retention was expressed in tobacco protoplasts together with other markers (A-C). The p24RFP-TMD17 is targeted *in vivo* to the ER (red). Protoplasts transformed with p24RFP-

TMD17 and the ER marker GFP-HDEL (green) showed a perfect co-localisation of the two fusion proteins and some GFP fluorescence in the nucleus (A). The *trans*-Golgi marker (yellow) did not co-localised with p24RFP-TMD17 (B) but there was a partial labelling with the PVC marker (green) (C). By contrast, p24RFP-TMD20 (red) co-localised with the ST-YFP Golgi marker (purple) (E) but not with the ER (green) (D). As with p24RFP-TMD17, there was some localisation with the PVC marker (purple) (F). Scale bars represent 10 µm.

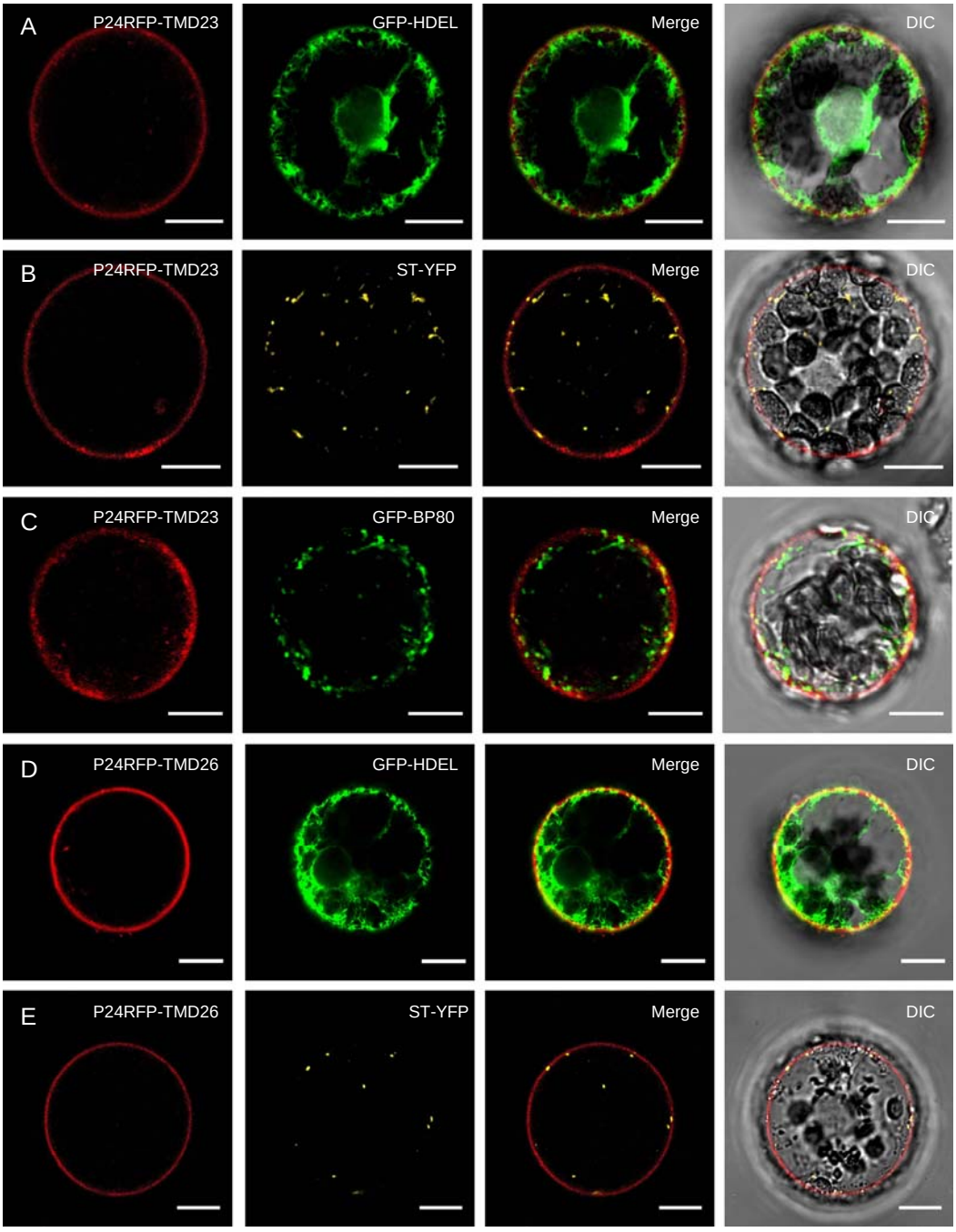


Figure 15. Intracellular localisation of p24RFP-TMD23 and p24RFP-TMD26 fusion proteins in protoplasts co-transformed with different organelle markers. p24RFP-TMD23 and p24-RFP-TMD26 were co-transformed with the different markers except the PVC marker in the p24RFP-TMD26. The recombinant p24RFP-TMD23 did not colocalise with the ER marker GFP-HDEL (A), neither with the

trans-Golgi marker ST-YFP (B) but there was some localisation with the PVC marker GFP-BP80 (C). The p24RFP-TMD26 did not show co-localisation with GFP-HDEL (D) neither with ST-YFP (E). Scale bars represent 10 µm.

The p24RFP-TMD20 accumulated within the Golgi complex as indicated by co-localisation with the *trans*-Golgi marker ST-YFP (Fig. 14 E). The merged punctate pattern showed the superposition of the two fluorescence signals in purple (arrows) which also moved synchronously (data not shown). However, some of the dots did not colocalise with the Golgi marker suggesting they correspond to another compartment. This fusion protein did not colocalise with the ER marker showing complete separate labelling of the recombinant proteins. However, it partially co-localised with the PVC marker (Fig. 14 F) suggesting some of the cytoplasmic dots corresponded to PVC. Using these two constructs, all the pictures of the co-transfection with GFP-BP80 showed the same results suggesting that the co-expression with this marker affected the production and/or folding of the recombinant proteins and higher quantities were directed to the vacuolar route.

The p24RFP-TMD23 and p24RFP-TMD26 fusion proteins showed a typical plasma membrane labelling and did not colocalise with any of the markers used (Fig. 15) except for the p24RFP-TMD23 which showed some co-localisation with the PVC marker (Fig. 15 C). In this case, the p24RFP-TMD23 showed some cytosolic punctate structures some of which corresponded to PVC as shown by the merged image while the other dots probably corresponded to Golgi stacks as some of them were mobile.
These results are consistent with the expected localisation of the recombinant proteins determined by the length of their transmembrane domain.

II.1.3 The N-terminally RFP-tagged p24 constructs are mistargeted to the vacuole

When tobacco protoplasts and leaves were transformed with pRFPp24-TMD constructs, surprisingly, no RFP labelling was detected in the expected membranes to which each recombinant protein was expected to localise.
Twenty-four hours post-transformation, tobacco protoplasts revealed strong labelling in the vacuole, which was not observed in the case of protoplasts overexpressing RFP-TMD or p24RFP-TMD from the same 35S promoter (Fig.16). Therefore, the vacuolar labelling must be related to the properties of the N-terminally RFP-tagged fusion

proteins. Despite numerous attempts, no specific single labelling of ER (pRFPp24-TMD17 construct), Golgi (pRFPp24-TMD20 construct) or plasma membrane (pRFPp24-TMD23 and pRFPp24-TMD26 constructs) was observed but for all constructs, a few protoplasts presented ER as well as vacuolar labelling (Fig. 16 C).

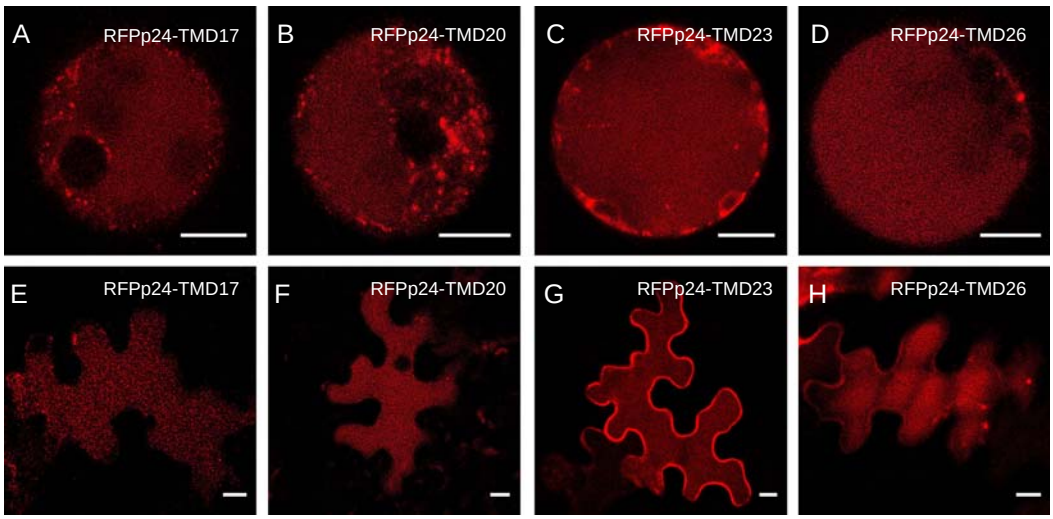


Figure 16. Subcellular localisation of transiently transformed protoplasts with RFPp24-TMD constructs and agroinfiltrated leaves. Protoplasts transformed with pRFPp24-TMD17 (A), pRFPp24-TMD20 (B), pRFPp24-TMD23 (C) and pRFPp24-TMD26 (D) showed vacuolar labelling and some punctate structures probably corresponding to PVC (A-B) and some ER labelling (C). *N.benthamiana* leaves transformed with recombinant Agrobacteria showed vacuolar labelling for RFPp24-TMD17 (E), RFPp24-TMD20 (F), RFPp24-TMD23 (G) and RFPp24-TMD26 (H). Plasma membrane labelling was observed for RFPp24-TMD23 (G) and some punctate structures in RFPp24-TMD26 (H). The leaves were observed 3 to 6 d.p.i. Scale bar for protoplasts is 10 µm and for agroinfiltrated leaves 20 µm.

This duality suggests that two populations of proteins were present within the cells possibly misfolded but with different degree of structural defects that lead either to ER retention of to vacuolar delivery for degradation. Alternatively, the recombinant protein constituted a single population directed to the vacuole for degradation by a slow process that allows detecting the protein that has not yet left the ER. Three to six d.p.i., agroinfiltrated leaves showed a clear vacuolar labelling in all the constructs but RFPp24-TMD23 also presented some plasma membrane labelling (Fig. 16 G).

The localisation of the different RFPp24-TMD fusion proteins was also investigated in tobacco protoplasts by co-expression with the GFP-HDEL, ST-YFP and GFP-BP80 markers (Fig. 17 and 18).

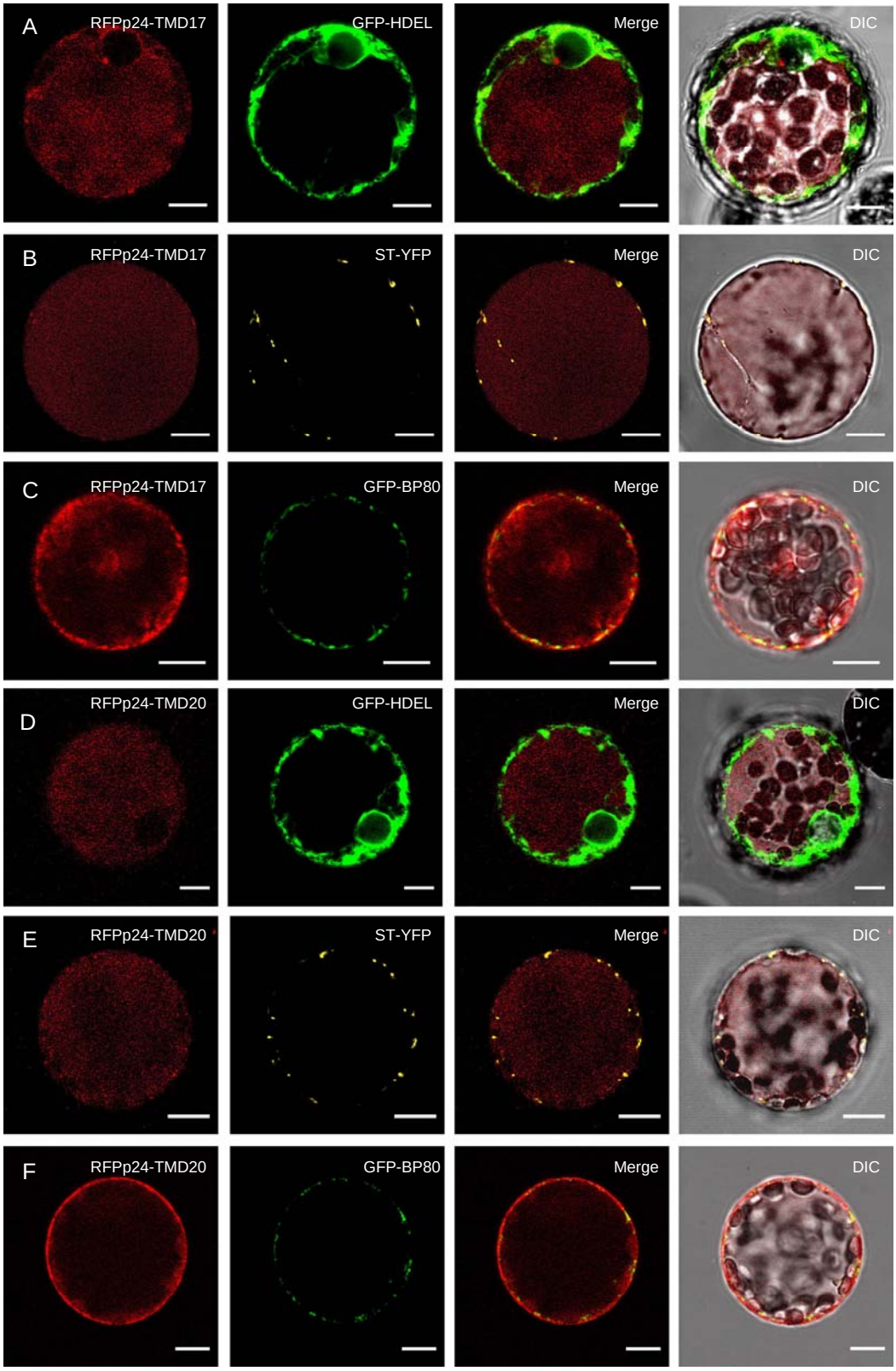


Figure 17. Confocal images of RFPp24-TMD17 and RFPp24-TMD20 localisation when co-transformed with different markers in tobacco protoplasts. pRFPp24-TMD17 and pRFPp24-TMD20 DNA constructs were transformed into tobacco protoplasts together with different organelle markers. Cells were incubated for 24 hours and then observed with a confocal microscope. RFPp24TMD17 did not show co-localisation with the ER marker GFP-HDEL (A), nor with the *trans*-Golgi marker ST-YFP (B)

or with the PVC marker GFP-BP80 (C). The RFPp24-TMD20 recombinant protein did not show co-localisation with GFP-HDEL (D), nor with ST-YFP (E) nor with the GFP-BP80 (F). Scale bars represent 10 µm.

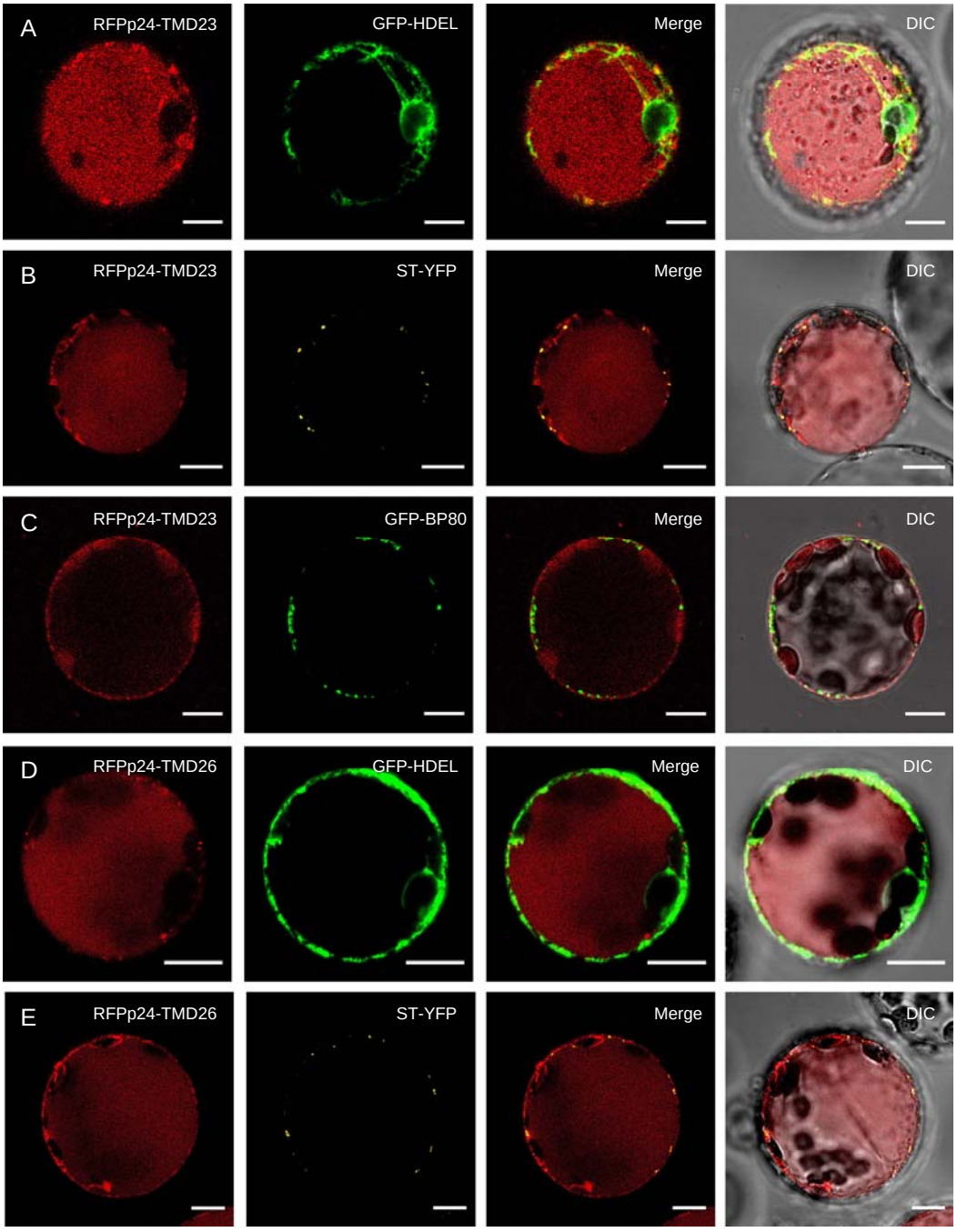


Figure 18. In vivo localisation in tobacco protoplasts of RFPp24-TMD23 and RFPp24-TMD26 fusion proteins when co-transformed with different markers. Panels A-C show the phenotypes obtained for RFPp24-TMD23. This fusion protein was localised in the vacuole and in some punctate structures when co-transformed with GFP-HDEL (panel A) and with GFP-BP80 (panel C). When co-transformed with ST-YFP, the fluorescence was observed in the vacuole and in the ER. Panels D-E show the phenotypes obtained for RFPp24-TMD26. The localisation of this recombinant protein was also in the vacuole and in some punctate structures when co-transformed with GFP-HDEL (panel D) and in the vacuole and ER when co-transformed with ST-YFP (panel E). Scale bars correspond to 10 µm.

As shown in Fig. 17 A-F, no co-localisation was observed for any of the co-transformed markers. These results are consistent with the vacuolar patterns obtained for each fusion protein expressed alone. Tobacco protoplasts transformed with pRFPp24-TMD17 and GFP-HDEL (Fig. 17 A), ST-YFP (B) and GFP-BP80 (C) showed vacuolar labelling and some ER labelling (panel A). The same results were observed for protoplasts transformed with pRFPp24-TMD20 and GFP-HDEL (Fig. 17 D), ST-YFP (E) and GFP-BP80 (F). In any case there was co-localisation with any of the markers tested.

In Figure 18, protoplasts transformed with pRFPp24-TMD23 (panels A-C) and pRFPp24-TMD26 (panels D-E) constructs and the different markers showed a clear separate fluorescence with respect to that shown by the different markers. In this case, there was some ER labelling of both fusion proteins when co-transformed with ST-YFP besides vacuolar labelling (Fig. 18 B, E). Some punctate structures were observed when the protoplasts were co-transformed with GFP-HDEL (Fig. 18 A, D).
These results suggested that the order of the sequences of interest in the different hybrid constructs tested had a significant effect on the final destination of the recombinant proteins.

II.1.4 Expression of RPF fusion proteins in transient expression of tobacco protoplasts

To confirm the expression of all the tested recombinant proteins, Western blot analysis was performed. Protoplasts transformed with the different pRFP-TMD, pP24RFP-TMD and pRFPp24-TMD constructs were extracted with extraction buffer without detergent, sonicated and a "membrane fraction" and a "soluble fraction" were separated by centrifugation (*Materials and methods* VIII.1.1). Proteins in the two fractions were then analyzed by SDS-PAGE. As these constructs contain a TMD which anchors the fusion proteins to a specific membrane, we expected to detect the fusion proteins in the "membrane fraction". Western blot analysis was performed using sheep anti-p24 antibodies (Aalto Reagents) and the majority of recombinant p24 was detected in the membrane fraction for every construct including p24 (p24RFP-TMD and RFPp24-TMD proteins) (Fig. 19, cell fraction). The recombinant proteins migrated to the expected molecular mass of 55 kDa and only a minor portion was detected in the soluble fraction (Fig. 19), suggesting that the fusion proteins were membrane-associated.

In the RFPp24-TMD fusion proteins, some putative degradation products could be detected by Western blot (Fig. 19 arrows). Moreover, recombinant RFPp24-TMD17 and RFPp24-TMD20 were also detected in the "soluble fraction" due either to the protein not being properly anchored in the expected membrane and being secreted, which it is not expected, or most probably to some pellet contamination in the supernatant during the protein extraction. No bands were detected in the RFP-TMD fusion proteins as they do not contain the p24 protein neither in the negative control in which the protein extract is from untransformed protoplasts.

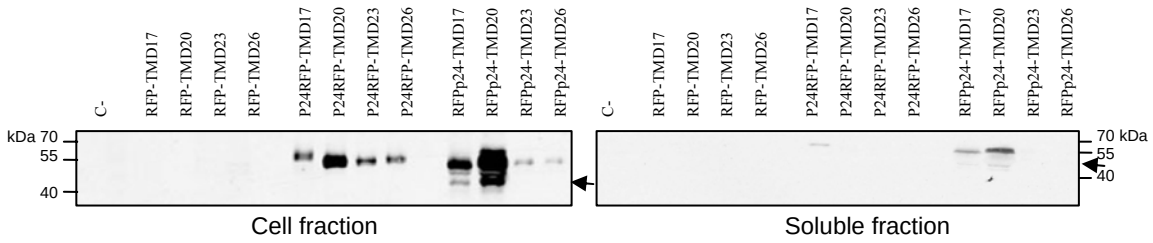


Figure 19. Western blot of the different RFP-TMD, p24RFP-TMD and RFPp24-TMD recombinant proteins from transiently transformed tobacco protoplasts. Transformed tobacco protoplasts were homogenized with extraction buffer and sonicated. Membrane and soluble fractions were separated and proteins were loaded on an SDS-gel. Western blot was performed using 1:500 sheep anti-p24 and most of the recombinant proteins were detected in the cell fraction but some in the soluble fraction. C- : wild type *Nicotiana tabacum* protoplasts extracts.

The same samples were also incubated with three different rabbit anti-RFP antibodies but the detection of the proteins was not successful with any of them. Therefore anti-RFP antibodies were not used any further.

II.1.5 The N-terminally GFP-tagged HIV-1 p24 constructs also present mistargeting

To investigate if the mistargeting of the recombinant proteins in protoplasts transformed with the pRFPp24-TMD DNA constructs was due to the specific fluorescent marker, the RFP was replaced by enhanced GFP (eGFP). We used this fluorescent protein because it contains a fluorescence-enhancing mutation.

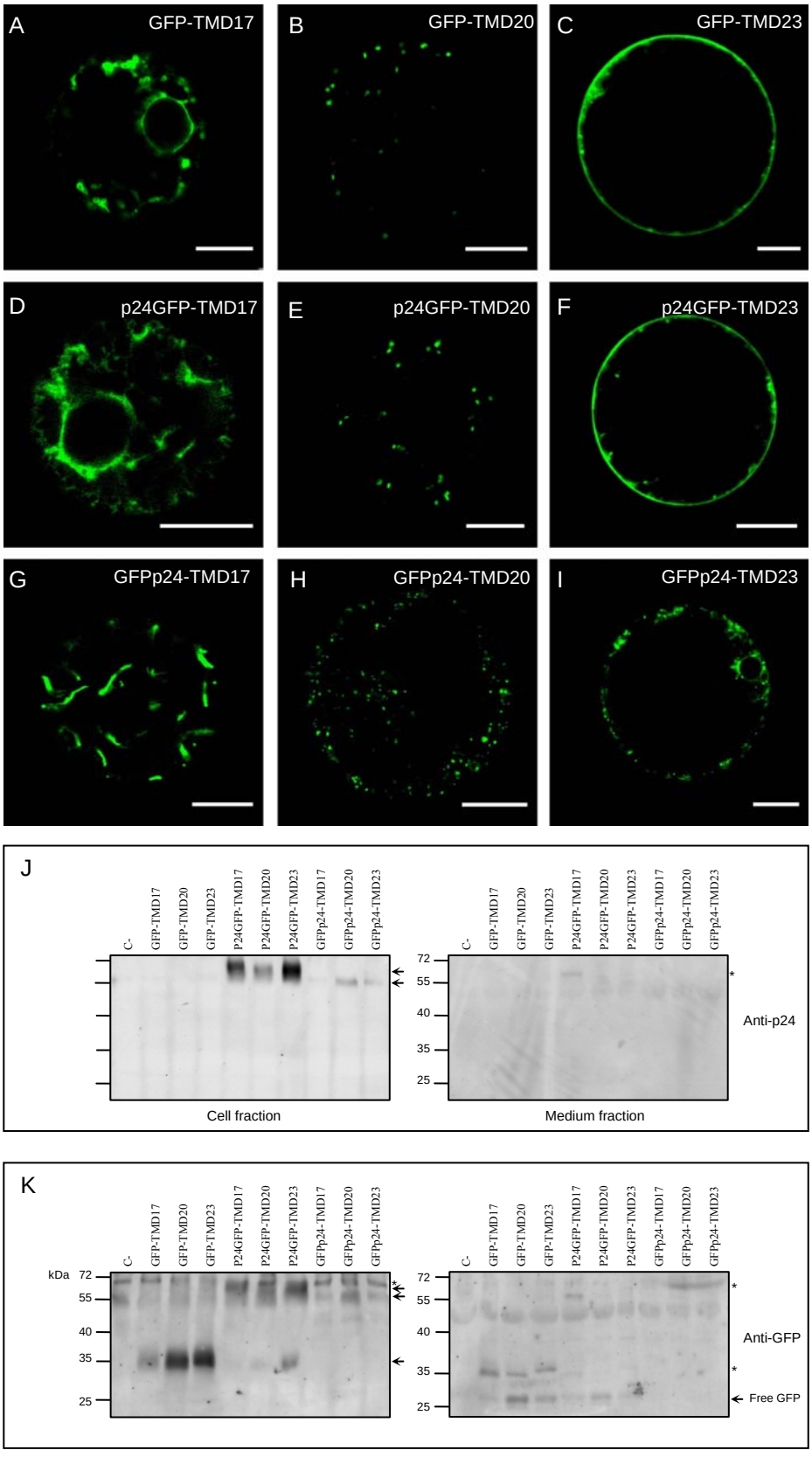


Figure 20. Localisation of GFP-containing p24 fusions in transformed tobacco protoplasts. Tobacco protoplasts were transformed with the different GFP-tagged constructs and after 24 hours the localisation of the fusion proteins was studied using a confocal microscope. GFP-TMD17 (panel A) and p24GFP-

TMD17 (panel D) showed ER labelling as expected but GFPp24-TMD17 presented ER-like aggregates (panel G). The GFP-TMD20 (panel B), p24GFP-TMD20 (panel E) and GFPp24-TMD20 (panel H) fusion proteins were observed in punctate structures. The first two correspond to Golgi stacks but the dots present for GFPp24-TMD20 were more numerous, smaller and very mobile. These structures might correspond to another organelle than Golgi stacks. The GFP-TMD23 (panel C) and p24GFP-TMD23 (panel F) fusion proteins were observed at the plasma membrane and in some punctate structures in its vicinity, probably Golgi stacks. The GFPp24-TMD23 fusion protein was observed in ER (panel I). Western blot analysis of extracts from transformed protoplasts in cell and medium fractions using anti-p24 (panel J) and anti-GFP antibodies (panel K) were performed. The arrows show the expected bands and the asterisks correspond to unspecific binding of the antibodies. C-: unstransformed tobacco protoplasts. Scale bars represent 10 μm.

When tobacco protoplasts were transformed with the different constructs, the different GFP-TMD (Fig. 20 A-C) and p24GFP-TMD (Fig. 20 D-F) proteins were targeted to the expected compartments but the GFPp24-TMD proteins were again mistargeted (Fig. 20 G-I). However, in this case they were not delivered to the vacuole.
The GFP-TMD17 and p24GFP-TMD17 contain the ER-retaining TMD. 24 hours after protoplast transformation the fluorescence was clearly in ER (Fig. 20 A, D). However, in confocal images of GFPp24-TMD17 transformed protoplasts the fluorescence was very intense in spike-like structures, most likely ER aggregates (Fig. 20 G).
In the protoplasts transformed with the constructs containing TMD20, a punctate pattern was observed in all of them (panels B, E, H). Despite the punctate pattern in protoplasts transformed with GFPp24-TMD20 (panel H), the dots were smaller, and in larger number and moved very fast when observed with the confocal microscope (data not shown) suggesting they were no real Golgi stacks. Cells transiently expressing GFP-TMD23 and p24GFP-TMD23 showed sharp fluorescent signals at the plasma membrane (Fig. 20 C, F). In both cases, fluorescence was occasionally detected in punctate structures in the vicinity of the plasma membrane probably corresponding to proteins transiting from Golgi to the plasma membrane. In contrast, GFPp24-TMD23 was not found in the plasma membrane, but in the ER (Fig. 20 I).

Protein extracts from protoplasts expressing the different GFP-tagged constructs were analyzed by SDS-PAGE and Western blot using anti-p24 or anti-GFP antibodies. In this case the samples were not fractionated between membrane and soluble fractions. Instead, total protoplasts homogenates were prepared (cell fraction) and the incubation medium was also tested to examine whether some of the proteins were secreted instead of being anchored to a membrane. Using anti-p24 antibodies (Fig. 20 J), the fusion proteins were mainly found in the protoplasts as expected. GFP-TMD fusion proteins were not detected when using this antibody as they do not contain the p24 gene.

P24GFP-TMD proteins migrated with the expected molecular mass of 56 kDa and the bands were intense. However, the GFPp24-TMD proteins presented a lower molecular mass, not corresponding to the expected 57 kDa. This correlates with the mistargeting observed for these fusion proteins by confocal microscopy, and is most probably due to some degradation of the p24 and/or TMD fusion partners. The negative control corresponding to untransformed protoplasts did not show any band. In the medium fraction a faint band in the p24GFP-TMD17 sample was detected corresponding to some contamination by protoplasts (Fig. 20 J asterisk). The same samples were also stained with an anti-GFP antibody (Fig. 20 K). In both fractions, some unspecific binding of the antibody can be observed (asterisk). The p24GFP-TMD fusion proteins were detected as previously using anti-p24 and the GFPp24-TMD proteins presented the same lower molecular weight bands. In this case, the GFP-TMD fusion proteins were detected with the expected molecular mass of 30 kDa. Moreover, some degradation products around 30 kDa were detected for p24GFP-TMD20 and p24GFP-TMD23 by the anti-GFP antibodies but not by the anti-p24 antibodies suggesting that part of the p24 protein was degraded. In the medium fraction, some free GFP was secreted (arrow) and other unspecific bands were detected (asterisks).

In summary, N-terminally GFP-tagged fusion proteins also presented mislocalisation. These results suggest that the position of the fluorescent marker markedly influence the destination of the p24 fusion proteins. Therefore, to avoid protein mislocalisation, the N-terminally XFP-tagged constructs were left aside for further experiments concerning the production of transgenic tobacco plants.

II.1.6 Cytosolic and secreted p24RFP and RFPp24

The HIV-1 p24 is originally a cytosolic protein, but we intended to produce it in the secretory lumen. While the pP24RFP-TMD constructions presented the expected localisation in preliminary experiments, the pRFPp24-TMD constructs appeared mislocalised. We wondered whether the RFP could have a position-dependent dominant effect, causing the production of misfolded recombinant proteins which would be recognized by the ER quality control system and directed to the vacuole.

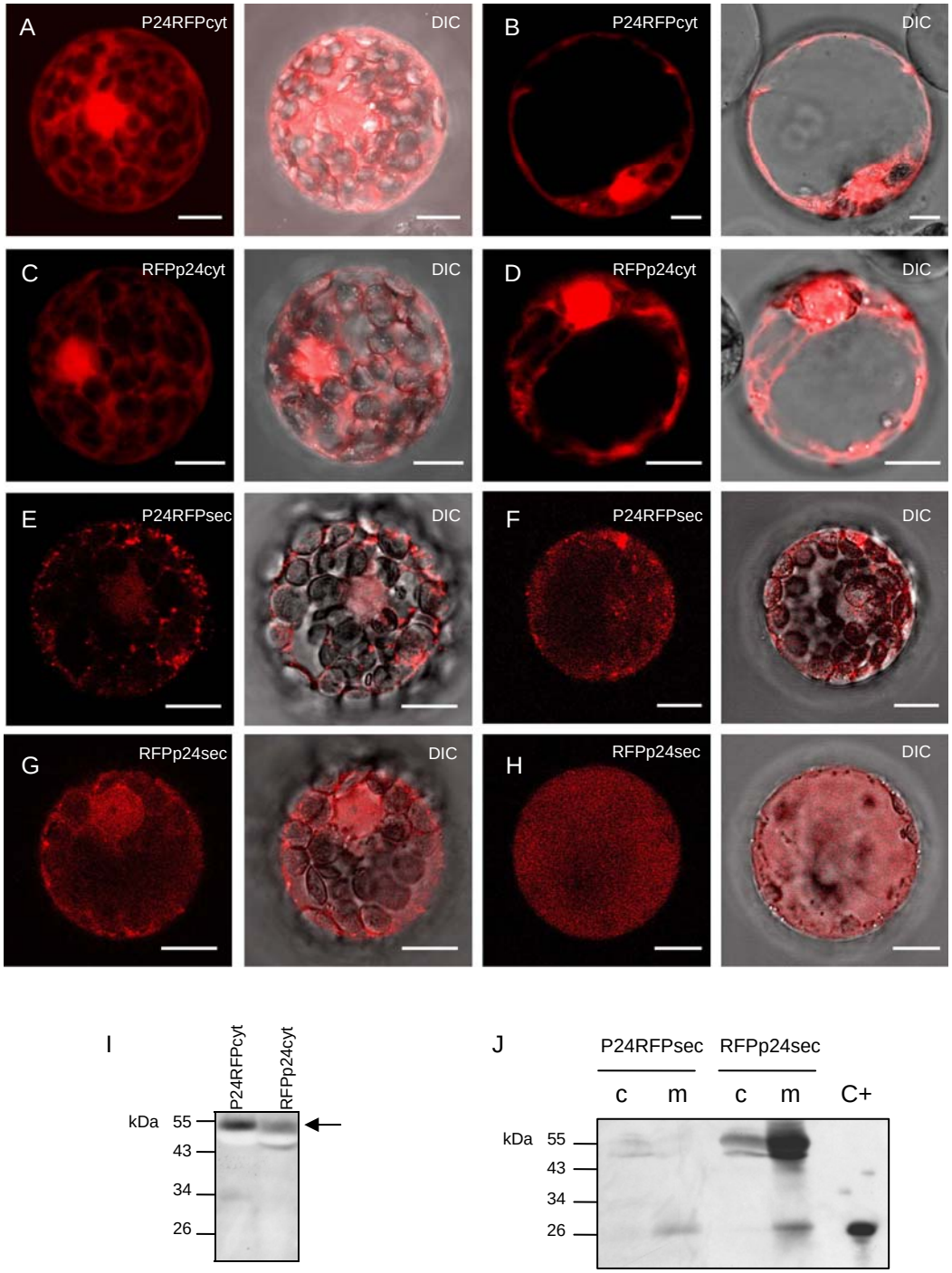


Figure 21. Localisation and Western blot analysis of cytosolic and secreted p24RFP and RFPp24 fusion proteins. Subcellular distribution of p24RFPcyt (A-B), RFPp24cyt (C-D), p24RFPsec (E-F) and RFPp24sec (G-H) in transformed tobacco protoplasts. Western blot analysis of p24RFPcyt and RFPp24cyt expressed in protoplasts (panel I) and p24RFPsec and RFPp24sec in the cells (c) and in the incubation medium (m) 24 h post-electroporation (panel J). C+: p24 protein. Scale bars correspond to 10 µm.

To investigate this, four constructs were prepared in which the fusion proteins are expected to be in the cytosol or to be secreted. The cytosolic constructs, p24RFPcyt and RFPp24cyt, do not contain a signal peptide nor a membrane-anchoring signal and after translation they are expected to remain in the cytosol. The secreted constructs, p24RFPsec and RFPp24sec, contain a signal peptide but no membrane-anchoring signal is present. These recombinant proteins are not targeted to a specific organelle and after entering the secretory pathway they are expected to travel until they are secreted to the apoplast.

Tobacco protoplasts were transformed with these different constructs and after 24 hours of incubation they were observed with a confocal microscope (Fig.21). RFP fluorescence images of protoplasts expressing p24RFPcyt and RFPp24cyt revealed a clear cytosolic labelling (Fig. 21 A-D). Images A and C show the 3D projection from z-stack images and a uniform fluorescence distribution can be seen all over the cytoplasm. Surprisingly, some nuclear fluorescence was observed in both constructs suggesting some free RFP might diffuse through the nuclear pore as it has been described for other fluorescent proteins (Berg and Beachy 2008).

As can be seen in Fig. 21 E-F, p24RFPsec showed a reticular and punctate pattern corresponding to ER and Golgi stacks. As the recombinant protein is expected to be secreted, the fluorescence can be seen in different organelles through which the fusion protein is travelling until reaching the apoplast. Some prevacuolar and vacuolar labelling was also observed in some transformed protoplasts indicating that some of the protein might be directed to this organelle. Protoplasts transformed with RFPp24sec (Fig. 21 G, H) mainly presented vacuolar labelling but ER was also visible indicating that some recombinant protein traffics through the secretory pathway as expected but is in part directed to the vacuole. Western blot analysis of the four fusion proteins using an anti-p24 antibody (Fig. 21 I) revealed that cytosolic p24RFP and RFPp24 are expressed and migrated with an expected mass of 53 kDa. No degradation products were neither detected nor free p24. In protoplasts expressing RFPp24sec, the 53 kDa fusion protein was found in higher amounts in the medium than inside the cells (Fig. 21 J). However, only a low amount p24RFPsec intact fusion protein was present inside the cells and no intact protein could be detected in the medium. These results suggest that the 53 kDa p24RFPsec fusion protein is mainly degraded. In both constructs free p24 was detected in the medium fraction and the intensity of the signal was comparable in both cases. The

experiment was repeated several times for these recombinant proteins with the same results.

II.1.7 RFPp24TMD-GFP fusion proteins are localised in the vacuole and the ER

In an effort to have more insight into the mislocalisation of the N-terminally RFP-tagged p24 constructs, GFP was fused to the C-terminus of the TMD. Protoplasts transformed with RFPp24TMD-GFP fusions showed mistargeting to the vacuole and some ER labelling with a similar pattern in the three constructs (Fig. 22). The red fluorescence was found distributed in the ER and the vacuole whereas the green fluorescence could only be seen in the ER. The merge image in which the ER labelling co-localised is shown in yellow. These results were not conclusive to understand the reason for which there was mistargeting of the recombinant proteins.

However, it was observed that some fusion proteins were not able to leave the ER as it can be seen for the red and green fluorescence but in some proportion they were degraded and only a portion was directed to the vacuole corresponding to RFPp24 or RFP alone. It appears that the full hybrid proteins do not leave the ER, but after cleavage somewhere between the RFP and the GFP, the RFP-containing part is transported to the vacuole. The 83 kDa recombinant proteins were detected in all the constructs by immunoblot using anti-p24 and anti-GFP antibodies (Fig. 22 D). The recombinant proteins migrated as a single band when immunodetected with anti-p24 but several bands were detected below the 83 kDa band when using anti-GFP. Moreover, a band around 44 kDa was detected in the three constructs indicating degradation of the fusion protein (this band most probably corresponds to the 396 amino acids fragment comprising part of the p24 together with the TMD and the full length GFP). The empty vector pGY1 did not show any band with either antibody. These results are similar to the Western blot analysis of pRFPp24-TMD in which all the constructs showed some degradation bands of the recombinant proteins. This result suggests the cleavage occurs within p24 and the membrane-anchored GFP remains in the ER while the RFP-containing part is targeted to the vacuole.

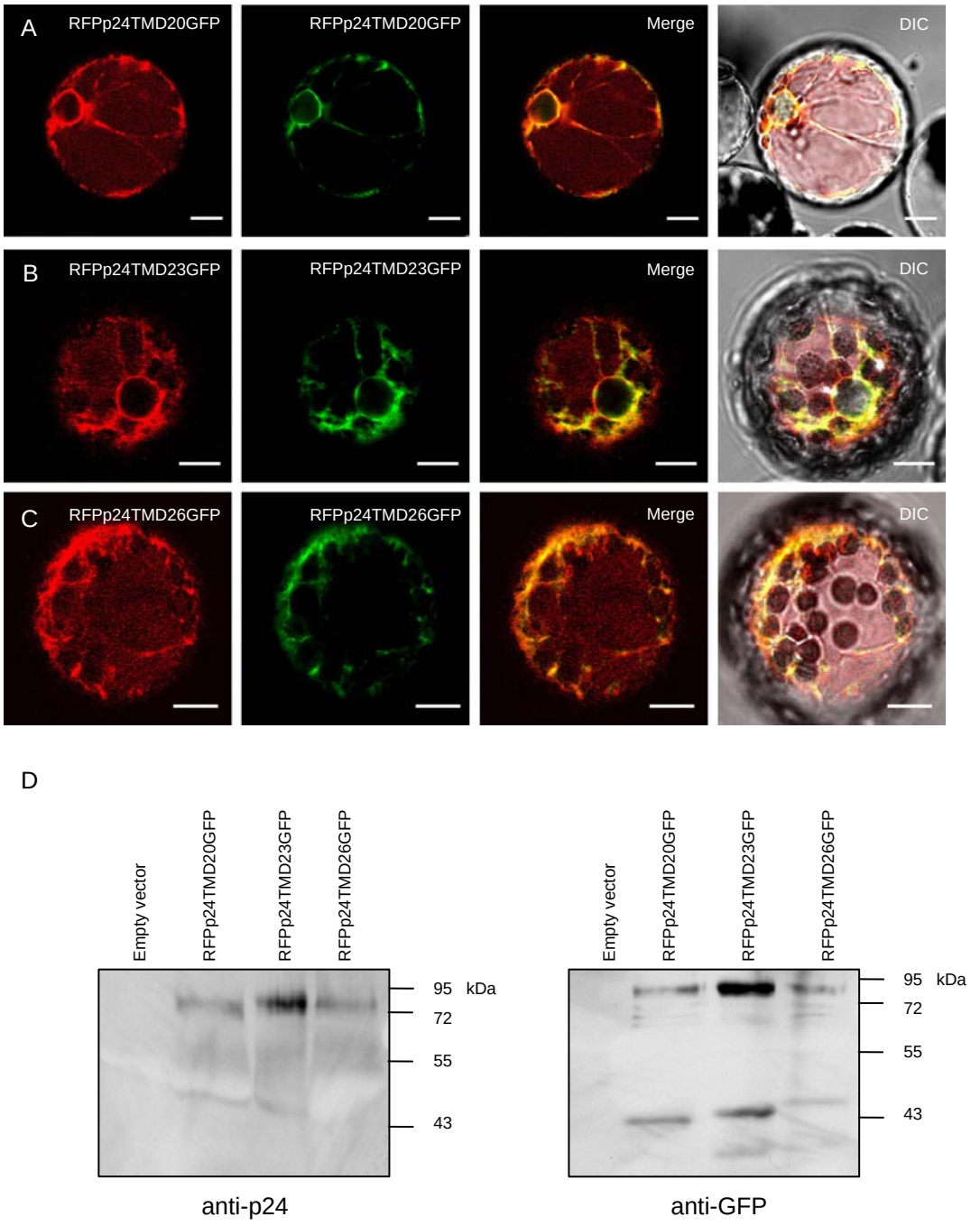


Figure 22. Subcellular localisation and Western blot analysis of RFPp24TMD-GFP fusion proteins. Subcellular distribution of RFPp24TMD20-GFP (A), RFPp24TMD23-GFP (B) and RFPp24TMD26-GFP (C) in transiently transformed tobacco protoplasts. Western blot analysis of the expressed fusion proteins and the empty vector pGY1 (EV) using anti-p24 and anti-GFP antibodies. Scale bars correspond to 10 µm.

II.1.8 Analysis of the subcellular distribution of RFP fusion proteins in other agroinfiltrated tobacco leaves

These results were obtained after the first binary vector was tested for stable plant transformation and did not work (IV.1). Then, the cassettes were cloned into pGREEN0179 and parallel to the plant stable transformation the recombinant vectors were tested by *Agrobacterium*-mediated transient expression (this experiment).

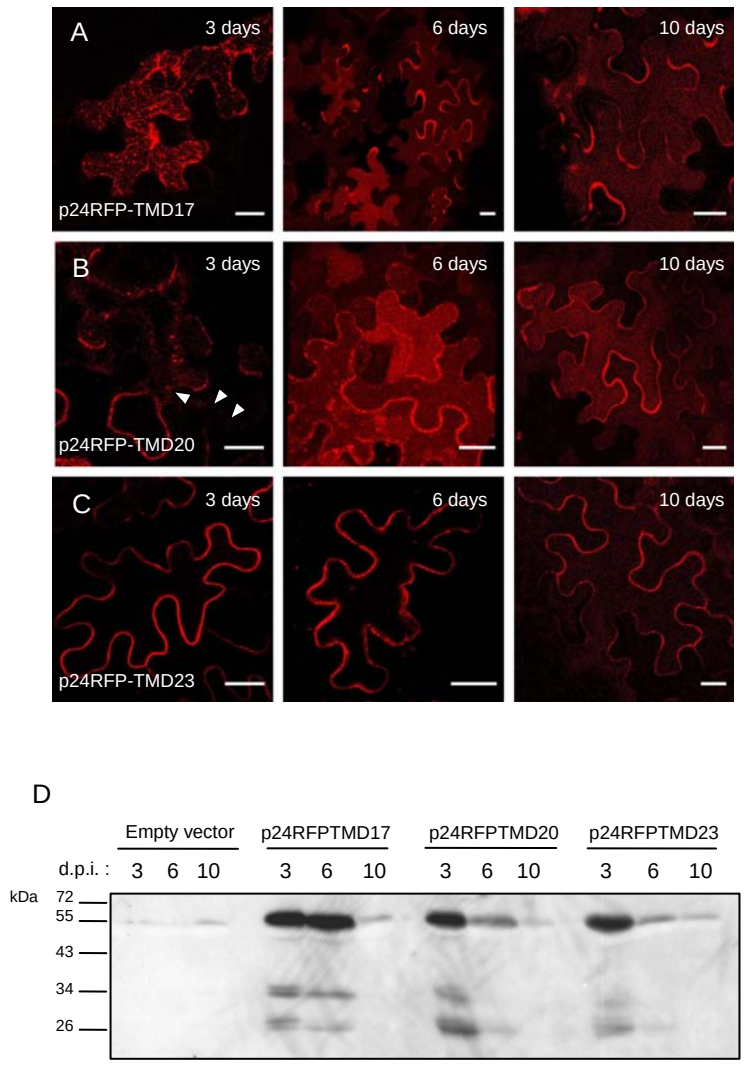


Figure 23. Efficiency of Agrobacteria delivery and gene expression of p24RFPTMD-0179 constructs. Tobacco leaves infiltrated with Agrobacteria cultures containing p24RFPTMD17-0179 (panel A), p24RFPTMD20-0179 (panel B) and p24RFPTMD23-0179 (panel C) were observed with a confocal microscope three, six and ten days post-infiltration. Western blot analysis of total soluble proteins obtained from the agroinfiltrated leaves using sheep anti-p24 antibody revealed the fusion proteins and some degradation products (panel D). Scale bars correspond to 20 µm.

Some of the cassettes cloned into the binary vector pGREEN0229 were also cloned into another binary vector, pGREEN0179. The results of *N.benthamiana* leaves agroinfiltrated with pGREEN0229-based vector have been described in II.1.1, II.1.2 and II.1.3. Recombinant Agrobacteria carrying the pGREEN0179-based constructs (table III) were agroinfiltrated in *N.benthamiana* leaves and a time-course experiment was performed to measure the time required to obtain maximal protein accumulation. Leaves from the same agroinfiltrated plant were analyzed 3, 6 and 10 days post-infiltration (d.p.i.) by confocal microscopy (Fig. 22 A-C).

Three days post-infiltration the red fluorescence corresponding to p24RFP-TMD17 was distributed in the ER showing a definite network as it can be seen in the 3D picture (Fig. 22 A). However, after 6 days, some fluorescence was observed in the vacuole and in the apoplast. At this time point, no ER labelling could be observed and the same results were found at day 10 p.i. These results indicate the fusion protein has its highest accumulation at day 3 p.i. (or before) and afterwards the protein is not able to be retained in the ER and is secreted or directed to the vacuole. Moreover, they support the hypothesis in which the normal turnover of the ER membrane occurs in the vacuole, possibly by autophagy (Tamura et al. 2004).

A punctate RFP signal was observed for p24RFP-TMD20 at day 3 p.i. (Fig. 22 B). The arrowheads indicate Golgi stacks that are mobile. Six days p.i., a punctate pattern could be still observed but also vacuolar labelling indicating some recombinant protein is directed to this organelle. At day 10 p.i., the fluorescence could be observed in the vacuole and the apoplast as for p24RFP-TMD17. The p24RFP-TMD23 was localised at the plasma membrane at days 3 and 6 p.i. (Fig. 22 C). At day 6, there were some punctae in the vicinity of the plasma membrane suggesting Golgi labelling. Ten days p.i., the fluorescence was present in the vacuole as well as at the plasma membrane. Comparing these results with the agroinfiltrated *N.benthamiana* leaves using pGREEN0229 vector (II.1.1, II.1.2 and II.1.3), in that case the leaves were only observed at one time point and no vacuolar labelling was observed for the time selected.

Protein extracts were collected at each time point to analyse by immunoblot the identity and integrity of the recombinant fusion proteins in tobacco leaves. The expressed proteins were detected using a specific anti-p24 antibody. Plants agroinfiltrated with empty vector showed a faint unspecific band (Fig. 22 D). Immunoblot analysis revealed a distinct band of approximately 55 kDa, corresponding to the molecular mass of p24RFP-TMD proteins. This band was very intense at 3 d.p.i.

in all constructs and also at 6 d.p.i. for the p24RFP-TMD17. At day 10 p.i., a massive decrease could be observed, which was similar for the three constructs. All constructs presented some degradation products of 35 and 24 kDa at 3 d.p.i. and also at 6 d.p.i. For the p24RFP-TMD17, the full length hybrid protein was detected with the same intensity at 3 and 6 d.p.i., where for the other two constructs there was marked decrease already at 6 d.p.i. This suggests that retention in the ER membrane results in a slower turnover compared to the Golgi or plasma membrane.

II.2 Analysis of RFPp24TMD in isolated vacuoles

To investigate if the fluorescence detected in the vacuoles in the RFPp24-TMD fusions corresponds to the RFP alone or to the whole recombinant protein, vacuoles were isolated from protoplasts transformed with pRFPp24-TMD17, pRFPp24-TMD20 and pRFPp24-TMD23. As negative control, protoplasts were transformed with the empty vector pGY1. Observation of the isolated vacuole fraction, which was stained with neutral red, under confocal microscope revealed that many of the neutral red-stained vacuoles observed by differential interference contrast were in fact vacuoles expressing the red fluorescent recombinant proteins when excited with the HeNe laser (Fig. 24).

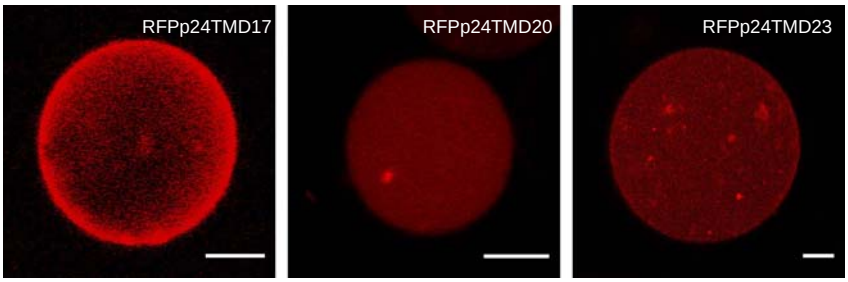


Figure 24. Isolated vacuoles from transiently transformed protoplasts. Vacuoles were isolated from protoplasts transformed with pRFPp24-TMD17, pRFPp24-TMD20 and pRFPp24-TMD23. Vacuoles presented red fluorescence in the tonoplast and in the lumen. Vacuoles isolated from protoplasts transformed with the empty vector did not show any fluorescence (not shown). Scale bars correspond to 10 µm.

Total protein extracts were concentrated with ammonium sulphate and loaded on a SDS-PAGE gel. Incubation with anti-p24 antibody revealed any band was detected (data not shown). These results demonstrate that the fluorescence seen in the vacuole for the RFPp24TMD constructs corresponds, indeed, to the part of the recombinant proteins in which p24 is not present.

III Transient expression of p24 in zein and tail-anchor fusion proteins in tobacco protoplasts

Tobacco protoplasts were transformed with pzein-p24, pP24-zein and pP24-TA constructs in order to test the expression of the recombinant proteins. These constructs did not contain any fluorescent protein as reporter and could not be studied with confocal microscopy as the other samples.

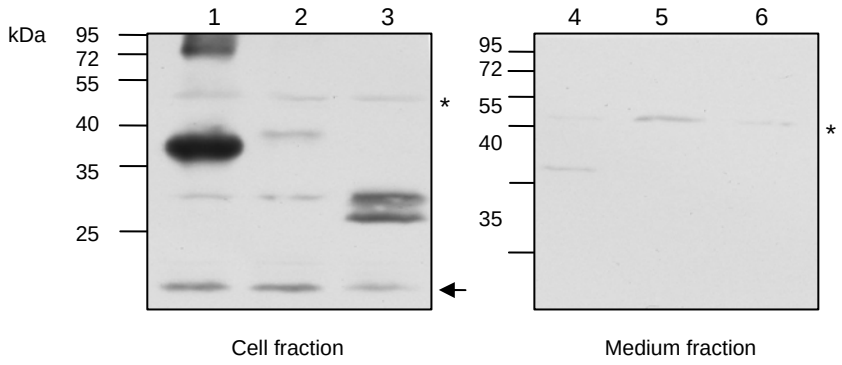


Figure 25. Expression of zein and TA fusions monitored by SDS-PAGE and Western blot analysis. Western blot analysis of p24 signals using anti-p24 on proteins extracted from tobacco protoplasts transformed with pzein-p24 (lanes 1 and 4), pP24-zein (lanes 2 and 5) and pP24-TA (lanes 3 and 6).

Protein extracts from transformed protoplasts were sonicated to obtain microsomal and soluble fractions. Protein gel blot analysis using anti-p24 antibody (Fig. 25) revealed the p24 fusions were located in the microsomal fraction, as expected. A very low amount of zein-p24 was also detected in the soluble fraction: this could either represent minor contamination by microsome (this construct accumulates to very high amounts) or minor release from the microsomal lumen, since this is the only one among the three constructs that is not membrane-anchored. (Fig. 25 right panel). The three fusion proteins migrated with the theoretically expected molecular mass. For both zein-containing proteins, the apparent molecular mass was around 41 kDa (lanes 1 and 2). Some high molecular weight proteins could be detected in the zein-p24 extracts (lane 1) suggesting protein multimers as usually formed by γ-zein (Vitale et al. 1982). This suggests that this construct is able to form protein bodies.

The p24-TA protein migrated as a doublet with apparent molecular masses of 35 and 37 kDa as theoretically expected. Because the tail anchor contains a luminal bovine opsin domain with an added *N*-glycosylation site, the doublet may be the result of incomplete

glycosylation, with the upper band corresponding to the glycosylated form, being therefore confirmed. In all three cases free p24 was also detected suggesting partial cleavage of the hybrid protein (arrow) and some unspecific bands were detected (asterisk).

In this preliminary experiment it seems that there is a strong position effect of the zein domains on the accumulation of the p24 fusion, but this will need to be confirmed analyzing transgenic plants and comparing mRNA levels, as it is described below.

IV Stable expression of rp24 in tobacco plants

It is generally accepted that the generation of stable nuclear transgenic plants is an essential step to develop an expression system for the production of the protein of interest. Plant transformation can be achieved by *Agrobacterium*-mediated transformation. After regeneration and screening of the T0 generation, the higher expressors of the recombinant proteins are selected and self- fertilized in order to obtain homozygous transgenic plants.

For the production of p24 hybrid proteins, tobacco leaf discs were infected with recombinant *A. tumefaciens* containing selected plasmids with the different cassettes tested in the previous paragraphs.

IV.1 The pGREEN0229 binary vector was not a good plasmid to stably transform tobacco plants

pGREEN plasmids were developed as optimal vectors for plant transformation using *Agrobacterium* (JIT catalogue, John Innes Centre, Norwich, UK; http://www.pgreen.ac.uk). The first binary vector used in this work to stably transform tobacco plants was pGREEN0229. This plasmid carries the *bar* gene, which confers resistance to the herbicide ammonium glufosinate (BASTA®). The different cassettes of interest under the control of the 35S CaMV promoter were inserted into the *EcoR*I site of the pGREEN0229 vector. We wanted to use this vector as the selective agent can be easily sprayed on plants grown in soil. The different constructs prepared with this plasmid are listed in table III.

After five attempts of leaf disk transformation, no viable calli could be obtained for plant regeneration. The transformation was carried under normal conditions as

described in *Materials and methods* VI.3, however when the transformed leaf disks were incubated in selective medium containing different concentrations of glufosinate, the calli did not look healthy (Fig. 26 B-C). Regeneration of calli was possible when no glufosinate was added to the medium (Fig. 26 A).

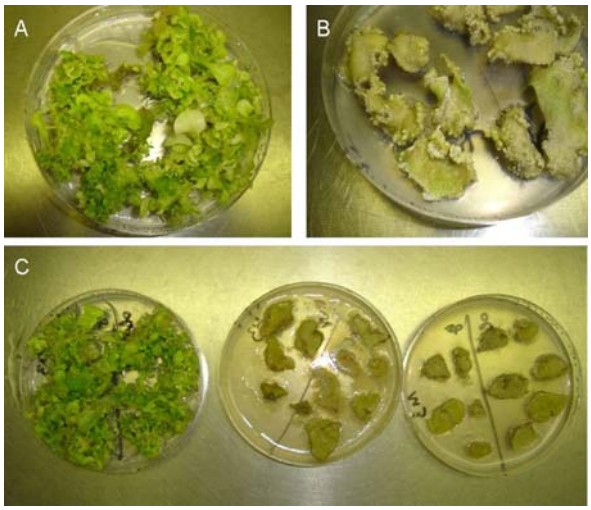


Figure 26. Callus regeneration four weeks after leaf disk transformation with recombinant pGREEN0229. A) Regenerated plantlets of tobacco leaf disks transformed with pGP24RFP-TMD17 without selection in the medium. B) Calli from pGP24-zein transformed leaf disks with 0.1 µg/ml glufosinate selection. C) Plates containing leaf disks transformed with pGP24-TA containing no selection (left), 0.1 µg/ml glufosinate (middle) and 0.5 µg/ml glufosinate (right).

However when the plantlets regenerated in non-selective medium were transferred to fresh medium containing glufosinate, they could not grow properly. Calli generated from leaf disks transformed with e.g. pGP24-zein were brown and the leaf disks were floppy and soft (Fig. 26 B). In all cases, plantlets could regenerate from the calli grown on non-selective medium, but in the plates containing glufosinate the regeneration was not possible and the leaf disks eventually died while Agrobacteria grew on the plates (Fig. 26 C).

After these unsuccessful attempts, we decided to change the vector. The different cassettes under the control of the 35S CaMV promoter were cloned into the *Eco*RI site of pGREEN0179. This binary vector confers hygromycin resistance in plants and the new constructs are also listed in table III.

The *Agrobacterium*-mediated plant transformation was successful using this binary vector.

IV.2 Growth and selection of transgenic tobacco plants

None of the different groups of transgenic plantlets showed any systematic alteration in growth or in morphology visible by eye. Twenty-five or thirty plantlets derived from calli from each transformation using the recombinant pGREEN0179 constructs were regenerated and transferred to rooting medium supplemented with 0.1 μg/ml indole acetic acid (IAA), 50 μg/ml hygromycin and 100 μg/ml carbenicillin. The different plants were named according to the recombinant protein to be expressed. After 4 weeks growing, not all the plants developed well from the 25 or 30 plantlets regenerated for each transgenic line (table IV). Some grew slowly and finally did not survive and Agrobacteria grew around them. Therefore only the healthy growing plants were analysed.

	Box 1	**Box 2**	**Box3**	**Box 4**	**Box 5**	**Box 6**	**# of plants regenerated**
p24RFP-TMD17	3/5	1/5	2/5	1/5	4/5	4/5	**15/30**
p24RFP-TMD20	4/5	2/5	3/5	3/5	1/5	-	**13/25**
p24RFP-TMD23	5/5	1/5	2/5	2/5	3/5	-	**13/25**
Zein-p24	3/5	3/5	4/5	4/5	3/5	-	**17/25**
p24-zein	3/5	2/5	4/5	4/5	3/5	-	**16/25**
p24-TA	**5/5**	**4/5**	**3/5**	**2/5**	**3/5**	**-**	**17/25**

Table IV. Number of p24-expressing transgenic tobacco plants regenerated from transformed calli. In each box the total number of plants regenerated is represented by the right number and the number of viable plants is represented by the left number. The right box shows the total number of viable plants (right number) and the total number of regenerated plants (left number).

IV.3 Verification of the insertion of the p24 hybrid genes

To verify the transgenic plants have integrated the gene of interest, genomic DNA from healthy plants derived from hygromycin-resistant calli was tested by PCR. The DNA was extracted from leaves and part of the p24 gene was amplified using primers #7 and #14. We used these primers because the sequence was common in all the constructs. Wild type plants were used as negative control and pP24RFP-TMD23 DNA construct as positive control. Twenty plants were tested corresponding to 4 plants expressing p24RFP-TMD17, 3 plants expressing p24RFP-TMD20, 2 plants expressing p24RFP-TMD23, 4 plants expressing zein-p24, 3 plants expressing p24-zein and 4

plants expressing p24-TA. The results are presented in Fig. 27 and, as expected, in the wild type plants no amplification was detected (Fig. 27 lane wt). All the 20 putative transgenic T0 plants analysed showed a band about 460 bp in an agarose gel corresponding to the expected p24 gene portion.

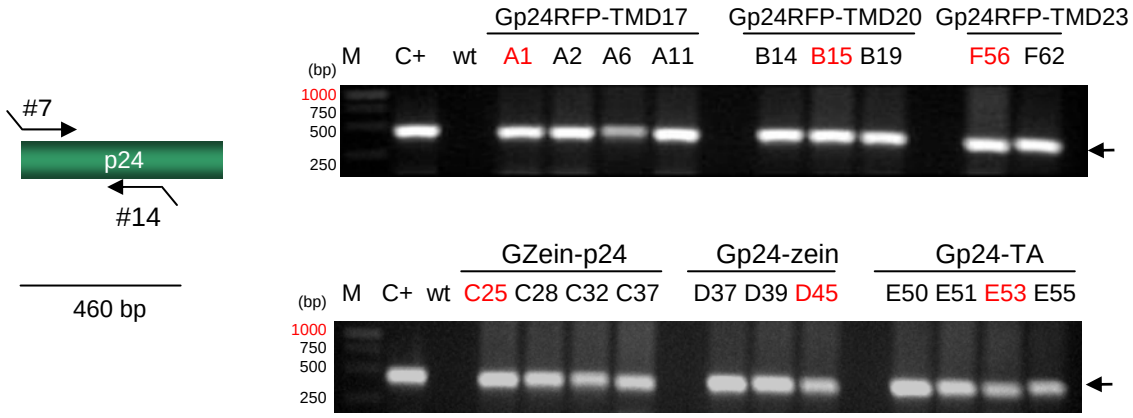


Figure 27. Agarose gel analysis of PCR products obtained after direct amplification of a fragment of the p24 gene. Genomic DNA was extracted from 20 transgenic tobacco leaf samples. Amplification of the 459 bp fragment of the p24 gene is indicated by the arrow. M: marker, wt: wild type, C+: pP24RFP-TMD23 used as positive control.

IV.4 Transgenic tobacco plants producing recombinant p24 anchored with a TMD

After 4 weeks growing in rooting medium, the different putative transgenic plants were screened for recombinant protein expression. Leaf protein extracts from all the regenerated transgenic p24RFPTMD plants were analysed by Western blot. The expressed protein was detected using sheep anti-p24 antibody. All different plant-produced p24RFP-TMD proteins had apparent molecular mass of 55 kDa as theoretically expected (Fig. 28). In the control, two polypeptides around 55 kDa were visible but the signal was weak (Fig. 28 WT asterisk). The levels of recombinant p24RFP-TMD17, p24RFP-TMD20 and p24RFP-TMD23 varied widely (arrow).

One each of the producing lines was selected for further analysis. Line A1 was chosen for p24RFP-TMD17 and B15 for p24RFP-TMD20 which are high-producing lines. F56 was chosen for p24RFP-TMD23. This is not a high-producing line but the other lines grown (F53, F62 and F63) did not developed well and died over time.

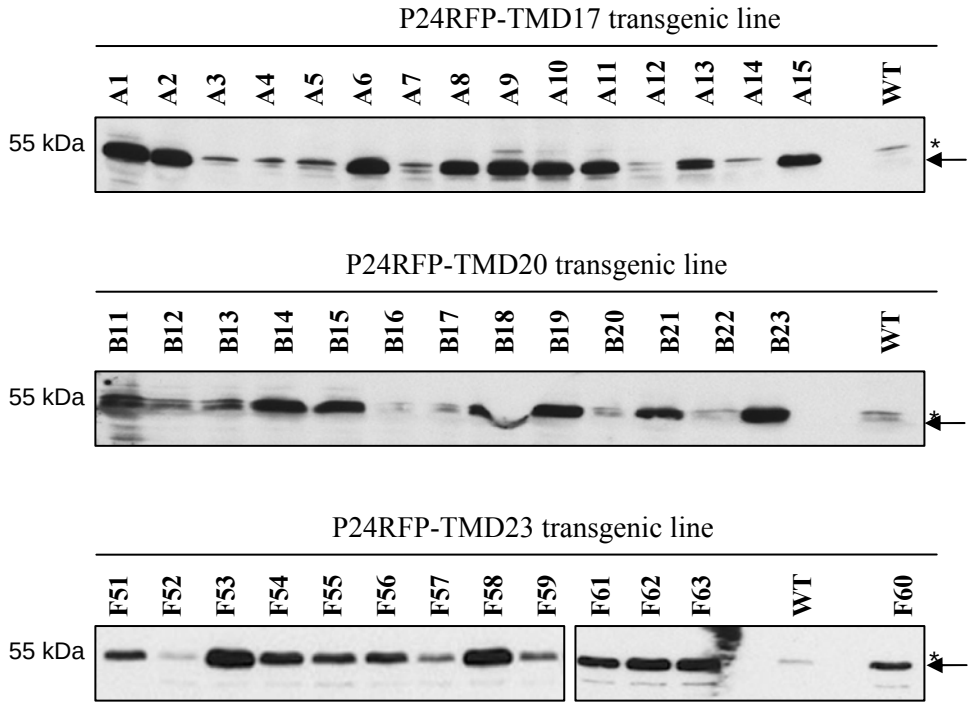


Figure 28. Screening of the transgenic p24RFP-TMD tobacco lines. Young leaves from regenerated transgenic tobacco plants were used for extraction of total soluble proteins and for analysis of the accumulation of recombinant proteins by immunoblot. 20 µg of TSP were loaded in all samples. WT (wild type) samples show two bands (asterisk) similar to the molecular mass expected for the p24RFP-TMD constructs (arrow) but weaker (around 55 kDa).

IV.5 Transgenic tobacco plants producing recombinant p24 fused to zein or tail anchor

Leaf extracts from putative transgenic *in vitro* grown T0 plants were analysed by Western blots using sheep anti-p24 antibody. The recombinant proteins were detected in all extracts. The zein-p24 and p24-zein fusion proteins migrated with apparent molecular mass of 41 kDa. A special feature of the zein-containing recombinant proteins is the presence of high molecular weight bands (Fig. 29 arrows). These probably correspond to a proportion of oligomers of the fusion proteins that were not denatured, as already observed in transient expression (Figure 25).

The p24-TA proteins were detected as a doublet in most cases. These bands most probably correspond to the glycosylated (37 kDa) and unglycosylated (35 kDa) isoforms of the fusion protein, and were also observed in transient expression (Figure 25). A band around 24 kDa was detected in all cases but in the WT sample this band was also detected and very intense suggesting unspecific binding of the antibody.

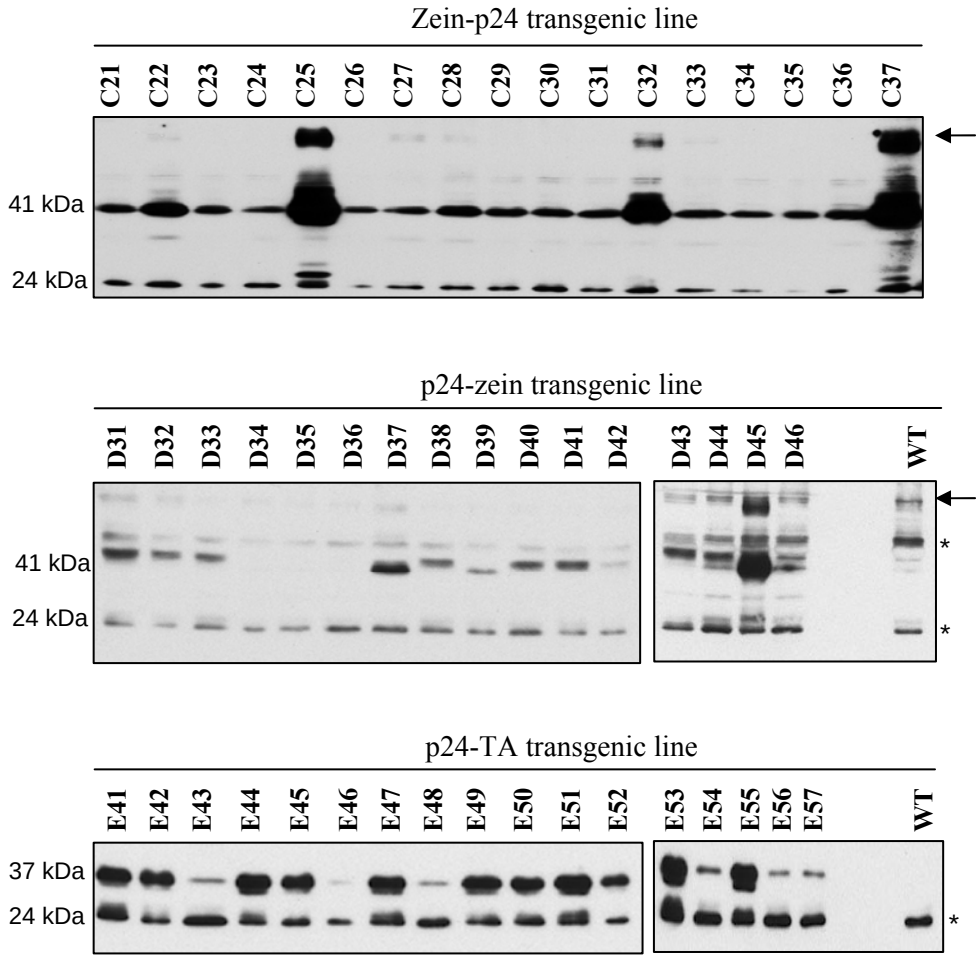


Figure 29. Western blot analysis of transgenic zein-p24, p24-zein and p24-TA tobacco plants. Protein extracts from young leaves from putative transgenic tobacco lines were analysed by Western blot using anti-p24 antibody. Zein-p24 and p24-zein migrated as a 41 kDa band and p24-TA to 37 kDa (glycosylated) and to 35 kDa (unglycosylated). High molecular weight bands were detected for zein-containing hybrid proteins (arrows). The WT sample showed some unspecific binding of the andtibody (asterisks) and although a band around 24 kDa is detected in all cases, is not clear it corresponds to free p24 as it was detected in the WT (asterisk).

A high-producing line for each transgene was chosen for further analysis: the C25 plant expressing zein-p24, the D45 plant expressing p24-zein and the E53 plant expressing p24-TA.

IV.6 Expression of the recombinant proteins in roots

As mentioned above, the cassettes coding for the different heterologous proteins were inserted under the control of the 35S CaMV promoter. This is a constitutive

promoter and the expression of the fusion proteins is expected to take place in most cells and tissues of the plant. We thus tested roots from the different selected transgenic lines for the expression of the fusion proteins (A1, B15, F56, C25, D45 and E53). About 1 gram of root material from each of the previously selected stable lines was used to extract total soluble proteins and 20 µg were loaded on a SDS-PAGE and Western blot was performed using sheep anti-p24 antibody.

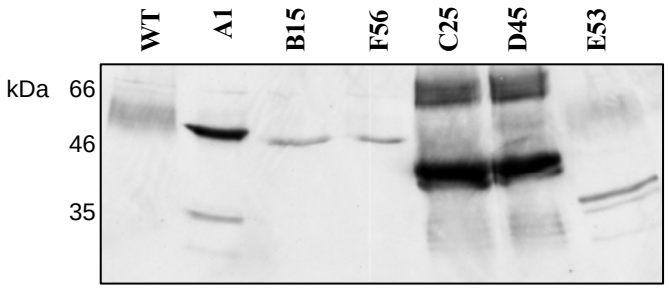


Figure 30. Expression of the fusion proteins in roots from stable transformed plants. 20 µg of the total soluble proteins from root extracts were separated by 12% SDS-PAGE, transferred to a PVDF membrane and incubated with the anti-p24 specific antibody. A1: p24RFP-TMD17, B15: p24RFP-TMD20, F56: p24RFP-TMD23, C25: zein-p24, D45: p24-zein and E53: p24-TA.

As shown in Fig. 30, Western blot analysis of total soluble proteins isolated from root tissue revealed that the p24RFP-TMD17 (A1), the p24RFP-TMD20 (B15) and the p24RFP-TMD23 (F56) proteins migrated with the expected molecular mass of 55 kDa. In A1 some degradation products were detected around 35 kDa. Zein-p24 (C25) and p24-zein (D45) presented the expected molecular mass of 41 kDa. Some high molecular weight oligomers were detected in both cases as in leaves suggesting formation of protein bodies. The p24-TA (E53) protein migrated as a doublet of 35 and 37 kDa and a little free p24 was detected. The wild type (WT) extracts showed some unspecific labelling around 60 kDa.

IV.7 Semi-quantitative analysis of p24 hybrid expression

The recombinant p24 fusion proteins were quantified in the T0 plants previously selected (A1, B15, F56, C25, D45 and E53) using semi-quantitative Western blot analysis. Representative cases from leaves expressing various levels of the recombinant proteins are shown in Figure 31. *N.tabacum* wild type leaf extracts were used as

negative control and the amounts were compared with different amounts of purified recombinant p24 produced in *Pichia pastoris* (NIBSC, EVA678).

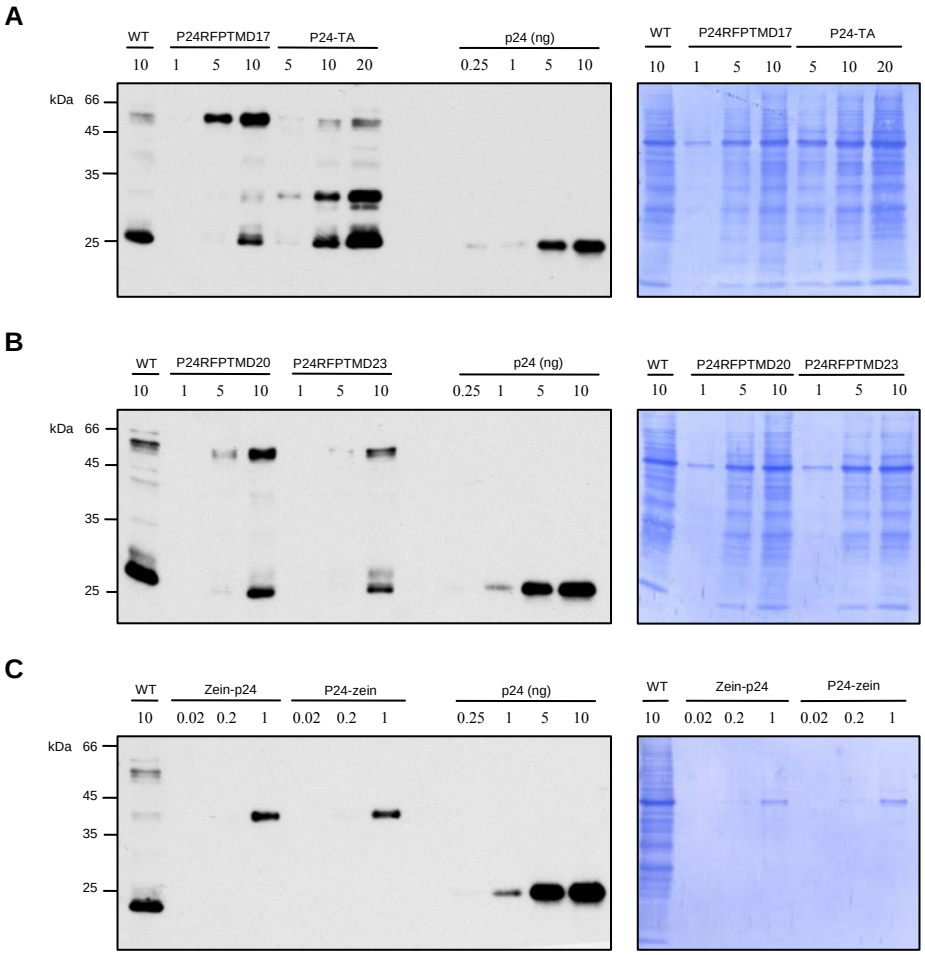


Figure 31. Semi quantitative detection of recombinant p24 by Western blot. Different amounts (expressed in micrograms) of total soluble protein from wild-type (WT) or transgenic plants were analyzed by Western blot. Increasing amounts (0.25 ng, 1 ng, 5 ng and 10 ng) of purified p24 protein expressed in *Pichia pastoris* were also analyzed as a reference for quantitation. The corresponding Coomassie blue-stained membranes are shown in each panel as protein loading controls for the plants extracts.

The highest accumulation levels were observed for plants expressing zein-p24 (C25 plant) and p24-zein (D45 plant) fusion proteins (Fig. 31 C, notice that about 1/10 the amount of TSP was loaded on the SDS gel compared with the amount of TSP loaded from the plants in panels A and B). Plants expressing p24RFP-TMD (A1, B15 and F56 plants expressing p24RFP-TMD17, p24RFP-TMD20 and p24RFP-TMD23, respectively) showed moderate accumulation levels (Fig. 31 A and B). The p24-TA expressing plant (E53 plant) gave a signal comparable to that of the p24RFP-TMD17

plant when twice as much TSP was analyzed (Fig. 31 A). The WT sample revealed some unspecific binding, especially around 60 and 20 kDa as found in other Western blots performed before (Figures 28-30). The unspecific bands are a little shifted from the expected recombinant proteins and the intensity of the recombinant proteins around 55 kDa (p24RFP-TMDs, panels A and B) suggests their specific binding. Moreover, these TMD proteins were observed by confocal microscope and fluorescence was observed so it is expected the recombinant proteins are expressed.

Semi-quantification of the different fusion proteins was done by visual comparison of the intensity of the bands obtained in the Western blot with known amounts of the standard. It was concluded that the accumulation of p24RFP-TMD17 represented about 0.3% TSP, and p24RFP-TMD20, p24RFP-TMD23 and p24-TA represented 0.15% TSP. Zein-p24 and p24-zein had the highest accumulation: about 1% TSP.

IV.8 Analysis of recombinant p24 mRNA transcription in the T0 generation

Protein accumulation depends on the combination of different parameters: the rate of transcription, mRNA stability, mRNA translation and protein stability. Northern blot analysis allows determining the steady state accumulation of mRNA, which results from the combination of transcription and stability. We first wanted to determine if the different levels of protein accumulation observed by Western blot could be related to recombinant mRNA levels. Northern blot analysis was performed on leaf extracts (Fig. 32). About 1 gram of leaf material from 6 week-old transgenic plants was divided into two halves: one was analysed by northern blot and the other half was tested by Western blot. Total RNA was extracted following the protocol as described in *Materials and methods* VII. A 700-bp DNA probe corresponding to the p24 CDS was generated by PCR and labelled with ^{32}P. Twenty independent transgenic plants plus a wild type plant as negative control were analysed. As obviously expected from the results of Western blot analysis, recombinant mRNA was present in the leaves of all of the transgenic plants but not in wild type *N.tabacum* (Fig 32, wild-type not shown). Total RNA was visualised by Ethidium bromide to confirm equal loading of the samples on the formamide gels (Fig. 32 A lower panel).

Different plants transformed with the same construct had different accumulation of recombinant mRNA. However, overall, plants expressing zein-p24 and p24-zein had higher mRNA accumulation whereas the p24-TA had lowest mRNA accumulation.

Analysis of two additional p24-TA plants (not shown) confirmed this result and also showed that when this construct was used there was very little variability in mRNA accumulation, contrarily to the other constructs. It should be noticed that the zein constructs produced two different bands: in both classes of transgenic plants, the lower bands correspond to the recombinant p24 transcript fused to the N-terminal domain of γ-zein. The upper bands (arrow in panel A) suggest secondary structures of the transcripts due to the GC-rich region of the N-terminal γ-zein domain.

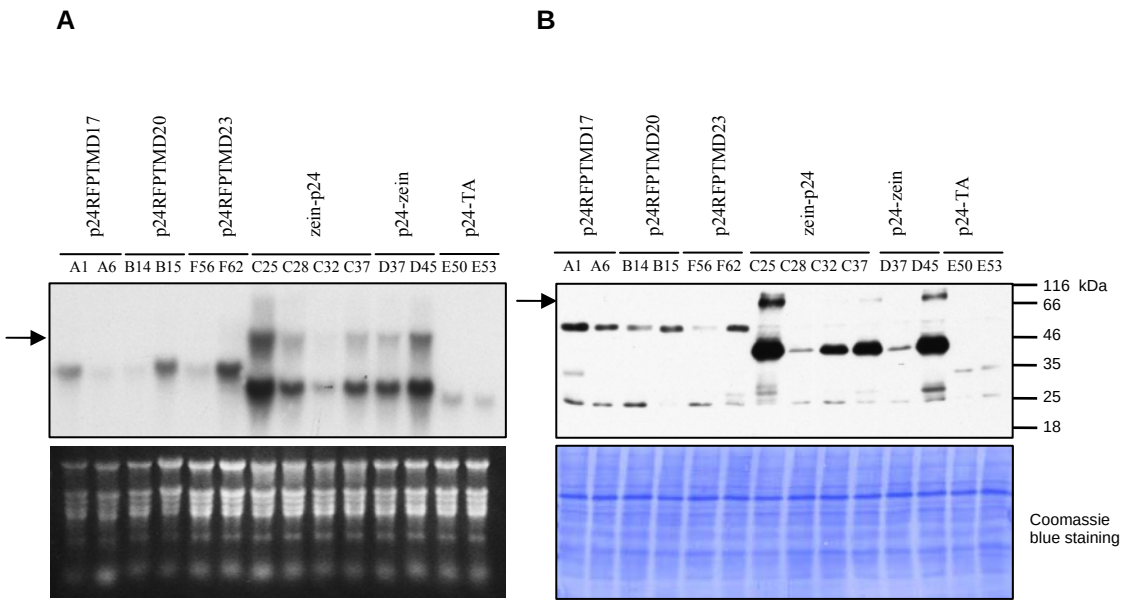


Figure 32. Northern and Western blot analysis of leaves from different transgenic plants. A) Northern blot analysis of different transgenic lines expressing p24 mRNA. P24RFP-TMDs plants expressed different levels of p24 transcript depending on the plant. Zein-p24 and p24-zein lines showed higher levels of transcription compared with the other transgenic lines. An upper band (arrow) was detected most probably corresponding to secondary structures of zein-containing transcripts. The two P24-TA lines showed lower transcription levels compared with the other lines. B) Western blot of total soluble proteins extracted from the same leaves. 10 µg of total soluble protein were separated by SDS-PAGE and immunodetected using sheep anti-p24 antibody. P24RFP-TMDs migrated at 55 kDa, zein-p24 and p24-zein at 41 kDa and p24-TA at 37 kDa. The arrow indicated oligomers of zein-p24 and p24-zein. After developing, the membrane was stained with Coomassie blue to confirm the equal loading of the samples.

Western blot analysis of the other half of material using sheep anti-p24 antibody confirmed the results of Figures 28 and 29, both qualitatively and quantitatively (Fig. 32 B). There was a correlation between mRNA and protein accumulation both within each class of transgenic plants and between classes. However it should be noticed that the

higher amount of protein with TMD17 compared to those with TMD20 and 23 is not justified by higher mRNA levels and, similarly, the very high protein accumulation of zein fusion constructs in some plants were also not entirely justified by the higher mRNA levels.

We conclude that, in general, there is a good correlation between transcription level and protein accumulation in the transgenic plants expressing the different p24 fusions, with the exception of the high protein accumulation of the zein fusions, which therefore could be due to higher translation or stability.

IV.9 Intracellular localisation of recombinant p24 in transgenic plants

In order to investigate the intracellular localisation of the recombinant proteins in the different transgenic plants, several techniques were used depending on the protein to be tested. For p24RFP-TMD plants, leaves and roots from the transgenic plants were observed with the confocal microscope taking advantage of the fluorescent protein they have fused. The plant-derived p24-TA was studied by isopycnic sucrose gradient as a preliminary result of intracellular compartment accumulation. Moreover, transgenic plants expressing p24-TA together with the zein-p24 and p24-zein were subjected to electron microscopy analysis in order to detect the putative ER localisation for p24-TA and the ER-derived PB accumulation for the zein fusions.

IV.9.1 Intracellular localisation by confocal analysis of recombinant p24RFP-TMD

Leaves and roots from two months old transgenic plants expressing recombinant p24RFP-TMD17, p24RFP-TMD20 and p24RFP-TMD23 were observed with a confocal microscope to confirm the localisation of the fusion proteins (Fig. 33). In all transgenic lines, the recombinant proteins were localised in the expected membranes in both plant tissues. In leaves, the p24RFP-TMD17 was localised in a reticular pattern corresponding to ER (arrowhead points the nuclear envelope labelling) but in this case the intensity of the fluorescence was low when compared to the fluorescence found in the apoplast indicating that an important proportion of the recombinant protein was secreted (arrow). In roots, the ER network was more visible and clear and the recombinant proteins were not detected in the apoplast.

The p24RFP-TMD20 accumulated in punctate structures in both leaves and roots (arrowheads). The labelling intensity of these structures was higher in roots than in leaves and in both cases they were mobile suggesting Golgi stacks.

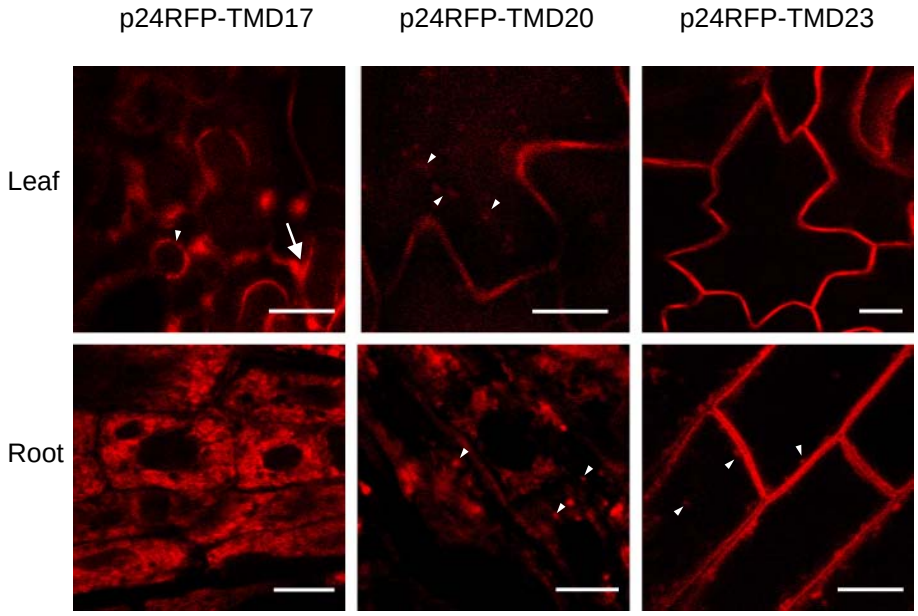


Figure 33. Intracellular localisation of fluorescent recombinant proteins in transgenic tobacco lines. Tobacco leaves and roots from p24RFP-TMD17-, p24RFP-TMD20- and p24RFP-TMD23-expressing lines (A1, B15 and F56, respectively) were observed with a confocal microscope. Scale bars correspond to 10 µm.

The p24RFP-TMD23 fusion proteins revealed a clear plasma membrane labelling in leaves and roots. In roots, some punctate structures were observed close to the plasma membrane (arrowheads), which correspond to the Golgi complex as they were mobile.

IV.9.2 Intracellular localisation of recombinant p24-TA by isopycnic sucrose gradient analysis

In order to verify the reticular localisation of p24-TA, leaves from the E53 transgenic plant were homogenized in the absence of detergent and fractionated in an isopycnic sucrose gradient by ultracentrifugation. In this case the buffer contained Mg^{2+} to maintain the ribosomes attached to the ER membrane. Leaf material was homogenized with homogenization buffer containing 0.4 M sucrose and after low speed centrifugation to remove nuclei, unbroken cells and cell wall material, the extract was

loaded on a linear sucrose gradient (20-50%, w/w) prepared in the same buffer. 500 µl fractions were collected and the proteins were concentrated by precipitation with TCA. The different fractions were loaded on a 12% SDS-PAGE and after electrophoresis and blotting proteins were detected with sheep anti-p24 antibody or rabbit anti-calreticulin antiserum (Figure 34).

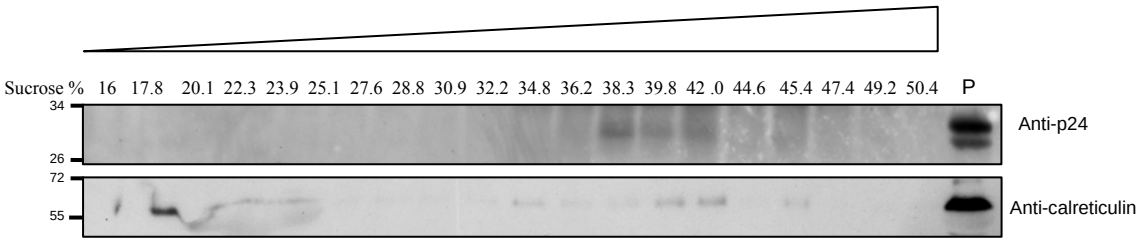


Figure 34. Isopycnic sucrose density gradient of p24-TA transgenic plant extract. Homogenates were prepared under high Mg^{2+} conditions ($MgCl_2$). The percentage refers to sucrose content. P is total pellet sample. The molecular masses are expressed in kDa.

Calreticulin is a multifunctional protein that binds Ca^{2+} and it is also a chaperone involved in protein folding in the ER. In this experiment this protein was used as an ER marker and was detected in the interval of 34% to 42% of sucrose, corresponding to the ER (Figure 34). The distribution of positive p24 signals showed that the fusion protein was detected in the narrow interval of 38.3% to 42% sucrose also corresponding to the ER. These preliminary results may hint to a localisation of the recombinant p24-TA in the ER membrane, but further experiments might be conducted to verify the ER localisation like immunofluorescence or electron microscopy.

IV.9.3 Intracellular localisation of recombinant p24-TA, zein-p24 and p24-zein by electron microscopy

An ultrastructural study was undertaken by electron microscopy in order to investigate the subcellular localisation of p24-TA, zein-p24 and p24-zein produced in transgenic plants. Fixed leaf tissue was examined by immunogold labelling incubating with anti-op3 antibody for the p24-TA sample and with anti-zein antiserum for the zein samples. The p24-TA sample resulted in a very low quality of the tissue chemical

fixation and the ER and the cytosol were not well preserved (data not shown). This experiment must be repeated in order to optimise the fixation of the sample.

The same preservation problem was observed for the zein-p24 and p24-zein samples. However, in a first gold immunolabelling experiment, putative ER-derived protein bodies were observed (Fig. 35 arrows) although no specific labelling was achieved as many gold particles were found in the vacuole (data not shown). These structures were observed in both zein-p24 and p24-zein but not in the negative control sample corresponding to wild type *N.tabacum*. These preliminary data strengthen the hypothesis that the zein-tagged p24 proteins form protein bodies however further experiments must be conducted to optimise the fixation of the samples and the specific labelling using anti-zein and/or anti-p24 antibodies.

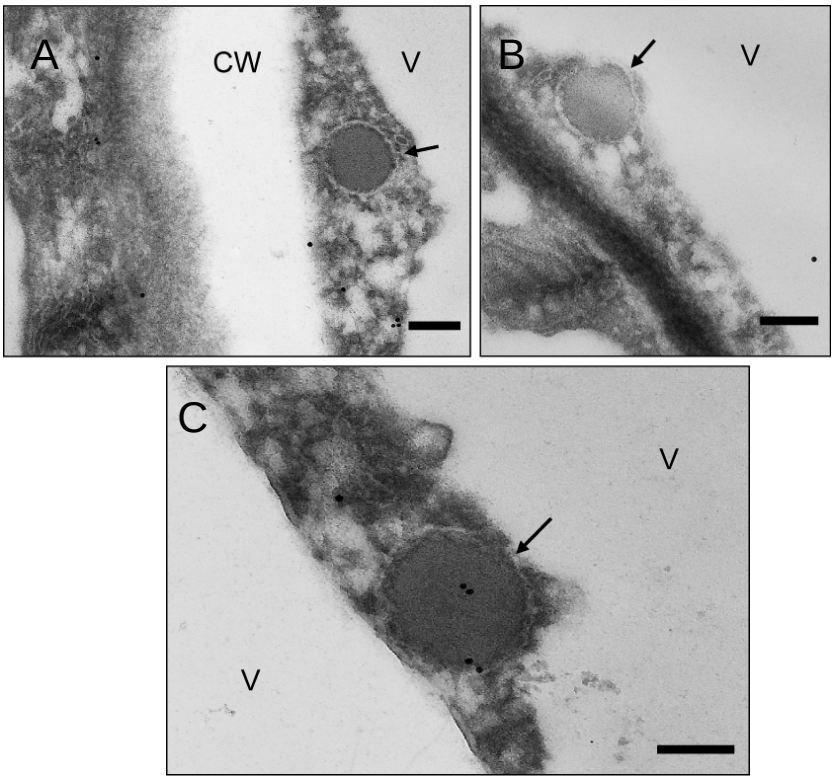


Figure 35. Electron micrographs of fixed transgenic zein-p24 (A) and p24-zein (B, C) tobacco leaves incubated with anti-zein antibody. Putative ER-derived protein bodies were observed (arrows). V: vacuole, CW: cell wall. The scale bars correspond to 180 nm.

IV.10 Stability of the recombinant p24 in transgenic tobacco plants

Pulse-chase experiments have proved to be a powerful tool to study protein folding, maturation, and degradation in different cell types (Jansens and Braakman 2003). When short pulses are applied and followed by chase, a fraction of the total protein pool can

be followed from synthesis to degradation in its natural environment. We wanted to explore by this method the stability of the different p24 recombinant proteins expressed in transgenic plants.

IV.10.1 Sheep anti-p24 antibody is not appropriate for immunoprecipitation but rabbit anti-p24 is

Immunoprecipitation allows to precipitate an antigen from a solution using a specific antibody with the help, as in our case, of Sepharose beads conjugated with Protein A or Protein G. Protein A is a protein from *Staphylococcus aureus* and Protein G a protein from *Streptococcus sp*. Both are proteins that bind to the F_c region of most classes and subclasses of immunoglobulins from several mammalian species leaving the F_{ab} region available for antigen binding.

In order to evaluate our immunoprecipitation protocol, sheep anti-p24 antibody used to immunodetect the recombinant p24 by Western blot was tested for immunoprecipitation of the p24 fusion proteins. Tobacco protoplasts were transiently transformed with the empty vector pGY1 as control and with pP24RFP-TMD17 DNA construct.

Recombinant proteins were radiolabelled with ^{35}S-Methionine/Cysteine for 1 h and chased for different times, from 0 to 8 h. The samples, corresponding to 300,000 protoplasts, were immunoprecipitated with 1:500 polyclonal sheep anti-p24 and then incubated with Protein G-Sepharose beads. As can be seen in Fig. 36 A, many bands could be detected in the precipitate from protoplast transformed with the pP24RFP-TMD17 construct or empty vector, and no specific polypeptides were precipitated from the former. This result revealed that this antibody is not suitable for immunoprecipitation.

To overcome this problem, we tested another anti p24 antibody, from rabbit. Protoplasts were again transformed with empty vector (pGY1) or pP24RFP-TMD17 and immunoprecipitated using 1:500 rabbit anti-p24 (NIBSC) (Fig. 36 B). In this case two sequential immunoprecipitation reactions were performed, the second one using the supernatant from the first one. The samples were incubated with the antibody and then with protein A-Sepharose beads. A single specific polypeptide of 55 kDa apparent molecular mass was detected in the pP24RFP-TMD17-transformed sample (lanes 3 and 4) but not in that transformed with the empty vector (lanes 5 and 6). As control, rabbit anti-BiP (BiP is an abundant chaperone of the ER) antibody was used for both samples

and a band around 75 kDa was detected in pP24RFP-TMD17 (lane 1) and in empty vector (lane 2) transformed protoplasts. The additional polypeptides co-precipitated with BiP represent newly synthesized polypeptides specifically associated to the chaperone (see for example Pedrazzini et al., 1997).

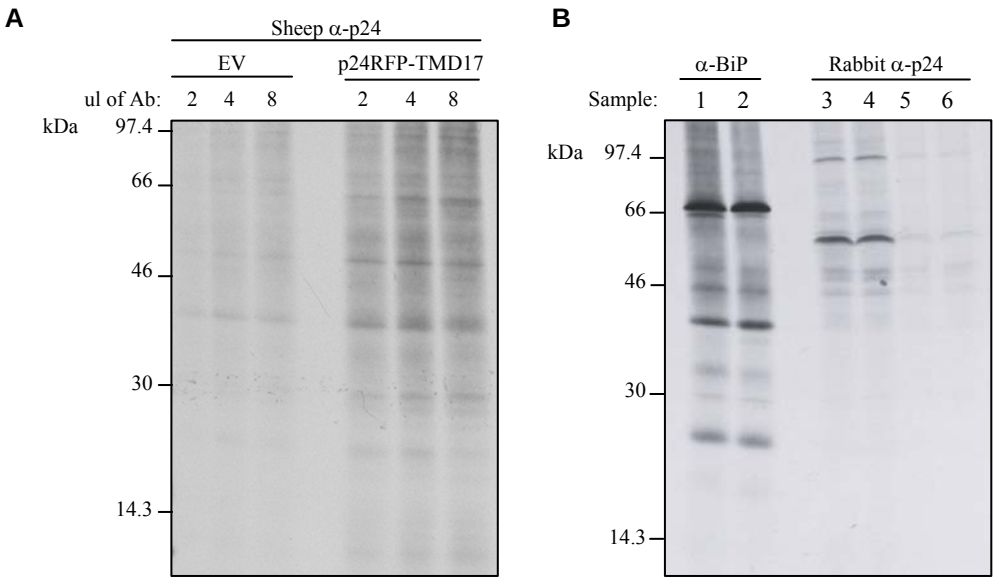


Figure 36. Evaluation of the ability of the two different antibodies to immunoprecipitate the recombinant fusion proteins in transiently transformed tobacco protoplasts. Radiolabelling and immunoprecipitation of transiently transformed tobacco protoplasts with empty vector pGY1 (EV) and pP24RFP-TMD17. A) Test of sheep anti p24 antibodies. Different amounts of sheep anti-p24 antibody were tested but the antibody could detect many bands one presumably around 55 kDa corresponding to the fusion protein. B) Test of rabbit anti p24 antibody. pP24RFP-TMD17 (lanes 1, 3 and 4) and EV (lanes 2, 5 and 6) transformed protoplasts were homogenized and immunoprecipitated with anti-BiP (lanes 1-2) or rabbit anti-p24 antibody (lanes 3-6). Lanes 4 and 6 correspond to a second round of immunoprecipitation from the supernatants of lane 3 and 5 immunoprecipitations, respectively.

The fact that the second round of immunoprecipitation still selected similar amount of the p24 recombinant polypeptide (compare lanes 3 and 4) indicated that at the concentration used the antibody is saturated by the recombinant protein when the homogenate from 100,000 protoplasts is immunoselected. The rabbit anti-p24 antibody was used for further immunoprecipitation experiments.

IV.10.2 P24RFP-TMD17 is more stable than p24RFP-TMD20 and p24RFP-TMD23

Pulse-chase radiolabelling experiments were carried out by labelling tobacco protoplasts from plants stably expressing p24RFP-TMD with ^{35}S-Methionine/Cysteine for 1 h and chasing for different times, from 0 to 8 h. Clarified cell lysates and supernatant media from equal number of cells were immunoprecipitated with rabbit anti-p24 antibody.

As shown in Fig. 37 A, p24RFP-TMD17 was readily detected within the pulse interval and migrated at the expected molecular mass of 55 kDa. The signal decreased in intensity during the chase. Comparison of the different time-points indicated a half life between 4 and 8 h. A lower band around 24 kDa (arrowhead) was also detected in the cell lysates suggesting free p24 could be already found inside the cells at time 0 h of chase. The free p24 was secreted into the medium and could be detected from 2 h to 8 h of chase in the medium fraction (Fig. 37 A arrowhead).

In our preliminary Western blot experiments using sheep anti-p24 antibody, two closely migrating bands around 35 kDa were constantly detected in stable p24RFP-TMD17 tobacco lines (Figures 30-32). These bands probably correspond to degradation products of the recombinant protein and they might contain the p24 protein as the epitopes are detected by the antibody. The two 35 kDa proteins were unstable after 8 h of chase in the cell lysates and they were also secreted after 2 h of chase (Fig. 37 A asterisks). These results suggest that there are 3 forms of the recombinant p24RFP-TMD17. One is the expected 55 kDa full length recombinant protein (arrows). Another form is the free p24 which is found inside the cells but also detected in the medium after 2 h chase (arrowheads). The intermediate form around 35 kDa and also starts to be detected in the medium at 2h chase and then disappears from the medium, possibly because converted to the 24 kDa polypeptide (asterisks).

These results partially confirm the mislocalisation of p24RFP-TMD17 observed in leaves of transgenic tobacco plants in which an important part of the fluorescence was detected in the apoplast (Fig. 33). Notice that the presence of intact p24RFP-TMD17 in the medium (circle) is probably an artefact due to contamination from protoplasts: the protein is found in the medium even at 0 h chase, a time too short to allow for secretion of newly synthesized proteins.

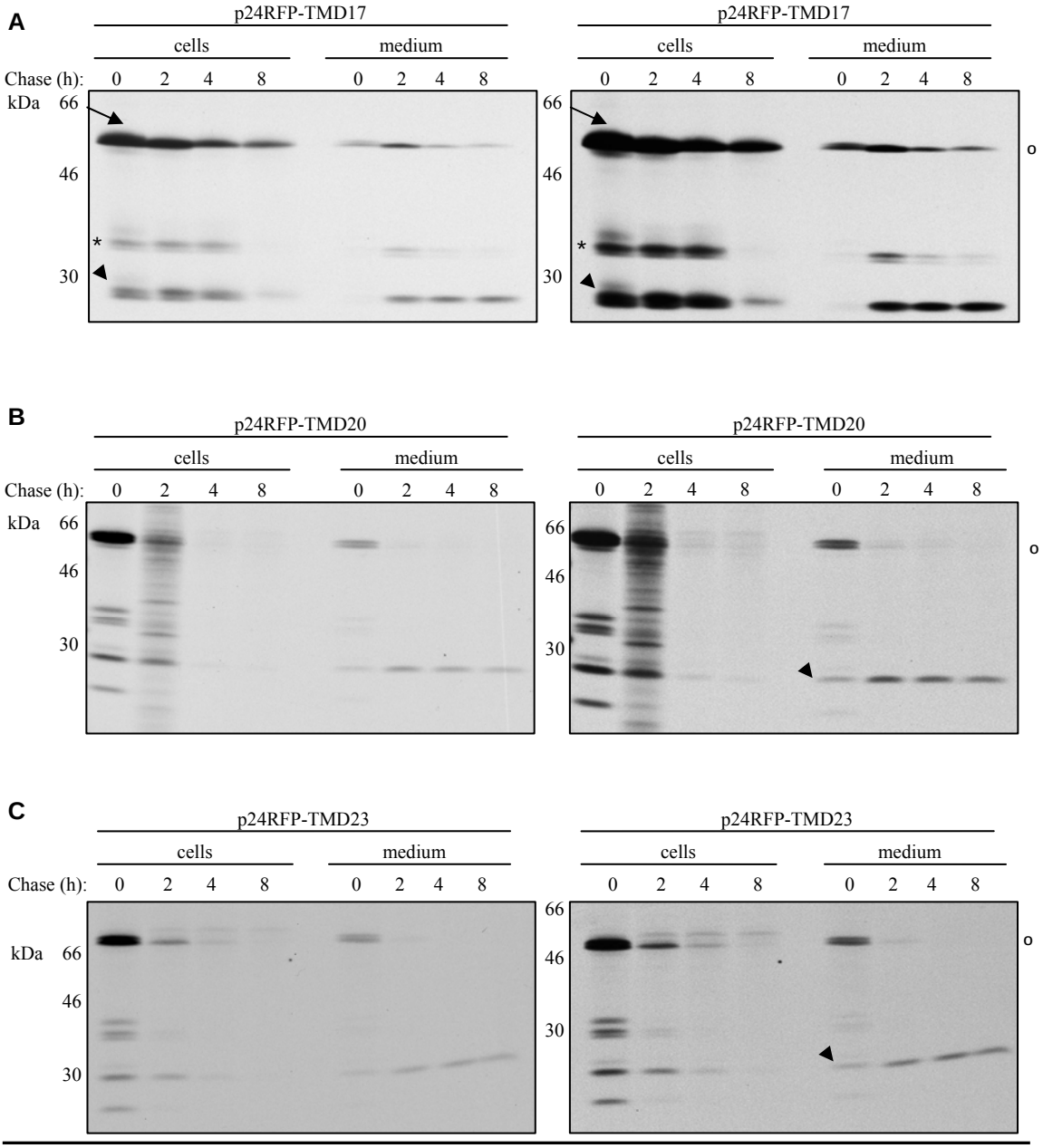


Figure 37. Plant-derived recombinant p24RFP-TMD fusions are retained inside the cells and are unstable. Tobacco protoplasts from leaves from transgenic plants were pulse labelled with ^{35}S-Methionine/Cysteine for 1 h and chased at 0, 2, 4 and 8 h as indicated above the lanes. Cell lysates (cells) and supernatants (medium) were immunoprecipitated with rabbit anti-p24 antibodies. A) p24RFP-TMD17 plant; one week exposure (left) and two weeks exposure (right). B) p24RFP-TMD20 plant; one week exposure (left) and two weeks exposure (right). C) p24RFP-TMD23 plant; one week exposure (left) and two weeks exposure (right). Each lane corresponds to 250,000 protoplasts.

The results shown in Fig. 37 B and C demonstrate that the full length p24RFP-TMD20 and p24RFP-TMD23 proteins contained in the cell fraction were detected at the end of the pulse but rapidly decayed with an apparent half-life of less than 2 hours. The

appearance of additional, smaller polypeptides at time 0 indicates that the recombinant protein starts to be degraded very soon after synthesis. Again, the presence of a minor proportion of intact recombinant polypeptides in the medium at 0 h chase is most probably due to protoplast contamination (Fig. 37 B, C circle), whereas, as was seen for p24RFP-TMD17, free p24 was detected in the medium fraction in both cases gradually increasing from 0 to 8 hours of chase, indicating secretion of the soluble polypeptide (Fig. 37 B, C arrowheads).

In conclusion, the recombinant protein with the 17 amino acid TMD is more stable than the other two membrane-bound fusions. This is consistent with the data regarding mRNA and protein levels.

IV.10.3 p24-zein fusions are stable

Wild type γ-zein is insoluble unless its disulfide bonds are reduced (Vitale et al. 1982). In this study, the fusion proteins containing the zein domain showed similar properties as an important fraction remained insoluble when no reducing agent was added to the protein extraction buffer and the immunoprecipitation was accordingly impossible (not shown).

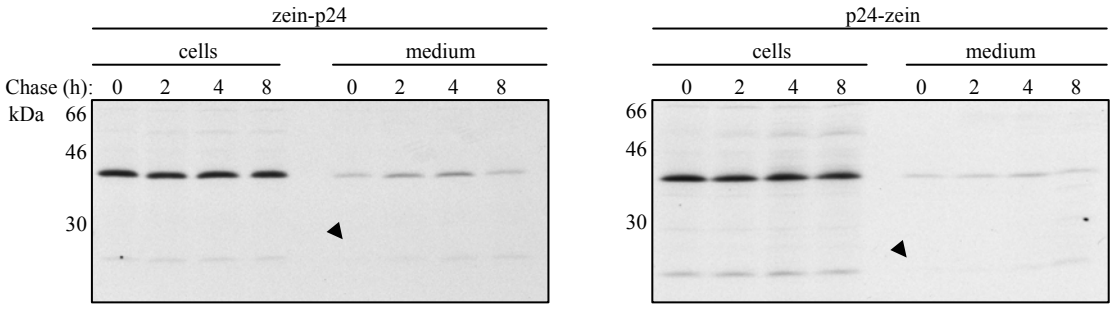


Figure 38. Pulse-chase analysis of plant-derived zein-p24 and p24-zein. Protoplasts from stably transformed plants were pulse-labelled with 35S-labelled amino acids and harvested at the indicated times. Cells were lysed in the presence of β-mercaptoethanol and the p24 proteins were immunoprecipitated using rabbit anti-p24 antibody. Radioactive proteins in the immunoprecipitates were analysed by SDS-PAGE and autoradiography. The autoradiograms were exposed for 2 weeks.

Therefore, β-mercaptoethanol was added to the immunoprecipitation involving zein fusions. Pulse-chase showed that the zein-p24 and p24-zein fusions proteins remained in the cells and the amount was not significantly reduced up to 8 h chase (Fig. 38). The

minor amounts of recombinant protein in the medium is again probably due to protoplast contamination, as it is already present at 0 h chase and does not change during subsequent time points. A small proportion free p24 can be detected in the cell lysates at 0 h chase and this is in part slowly secreted during the chase (arrowheads).
The high stability of the two polypeptides, together with the high mRNA levels, is therefore the basis for the high accumulation detected by Western blot.

IV.10.4 p24-TA is stable

Protoplasts from a p24-TA expressing plant (E53) were pulse-labelled for 1 hour and chased for different times. The p24-TA protein was immunoprecipitated using two different antibodies, rabbit anti-p24 and mouse anti-op3. The first antibody recognizes the p24 protein while the second recognizes the bovine opsin domain added to the C-terminal tail of cyt b5. Intact p24-TA was detected as a doublet of 35 and 37 kDa by either antibodies (Fig. 39 arrows).

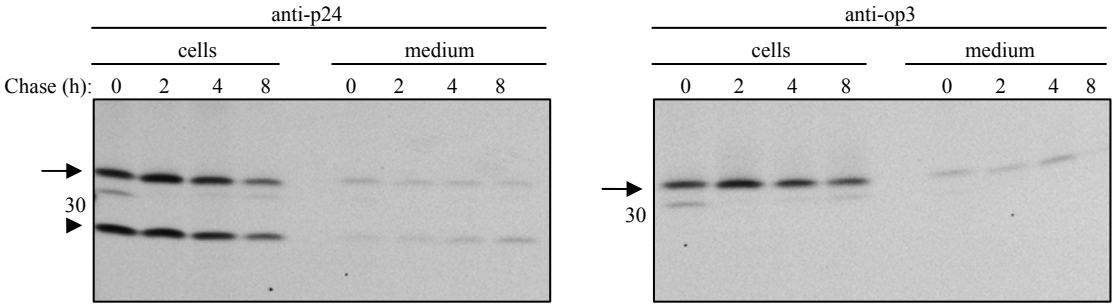


Figure 39. Analysis of the p24-TA stability from a transgenic plant. Protoplasts were pulse labelled and chased up to 8 hours. Two antibodies were used to immunoprecipitate p24-TA, rabbit anti-p24 (left panel) and mouse anti-op3 (right panel). Immunocomplexes were precipitated using protein-A Sepharose (left panel) and protein-G Sepharose (right panel). SDS-PAGE and fluorography revealed that p24-TA is stable up to 8 hours of chase.

The upper band, probably corresponding to the glycosylated isoform, was predominant in the cell fraction during the 8 hours of chase in contrast to the unglycosylated isoform, which was mainly detected at time 0. The stability of intact, probably glycosylated, p24-TA was intermediate between those of the zein construct and the TMD17 construct. A relatively high amount of free p24 was also detected at all time points inside the cells (notice that, as expected, this is not recognized by the anti-op3 antibody) and was in

small proportion secreted (Fig. 39 arrowhead). The high ration of free 24 with respect to p24-TA is in agreement with the Western blot results (see Figure 30-32).

IV.11 Purification of the rp24 from transgenic plants

The major reason for the high production cost of pharmaceutical proteins is the purification of the recombinant proteins (Twyman et al. 2003; Kirk and Webb 2005; Vitale and Pedrazzini 2005). To be commercially competitive, a recombinant protein produced in plants must have a cost-effective, efficient and potentially up-scalable-purification system.

If the protein of interest is expressed as a fusion (such as with added tags or stabilizing sequences) it could be necessary to remove it from the added sequence. This can be achieved *in vitro* by site-specific proteolysis with adequate proteases if, as in our case, specific proteolytic sites were inserted into the recombinant constructs. Thrombin is a serine protease that plays an important role in the blood coagulation cascade. This protein specifically recognizes the amino acid sequence LVPR/G and cleaves between the Arg and the Gly. Thrombin was chosen for our recombinant p24 proteins in all the constructs because the p24 sequence does not contain any specific thrombin target, the protease is not present in plant cells, and thrombin cleavage leaves only one Gly residue at the N-terminus.

To test if recombinant p24 could be released by in vitro thrombin cleavage, total soluble proteins were extracted from leaves of the different transgenic plants. The extraction was carried under non-reducing conditions except for the zein-p24 and p24-zein samples. As seen for immunoprecipitation (*Materials and methods* IV.10.3), the recombinant proteins containing zein were detected in the insoluble fraction when extracted with non-denaturing buffer. For this reason, β-mercaptoethanol was added to the extraction buffer used for these two recombinant proteins.

Immunoprecipitation was performed using three antibodies. The first antibody (anti p24) immunoprecipitated all the recombinant proteins produced in the different transgenic plants. The other antibodies bound to the zein domain (anti-zein) and to the op3 domain of the p24-TA recombinant protein (anti-op3). The immunocomplexes were incubated with protein A- or protein G-Sepharose and were subjected to thrombin digestion. The images shown in the next section correspond to Western blot analysis using rabbit anti-p24 antibody after thrombin cleavage.

IV.11.1 Purification using rabbit anti-p24

Leaf homogenates of plants expressing the zein or TMD fusions were incubated with rabbit anti-p24 antibodies and the antigen-antibody complexes were immunoprecipitated using protein A-Sepharose.

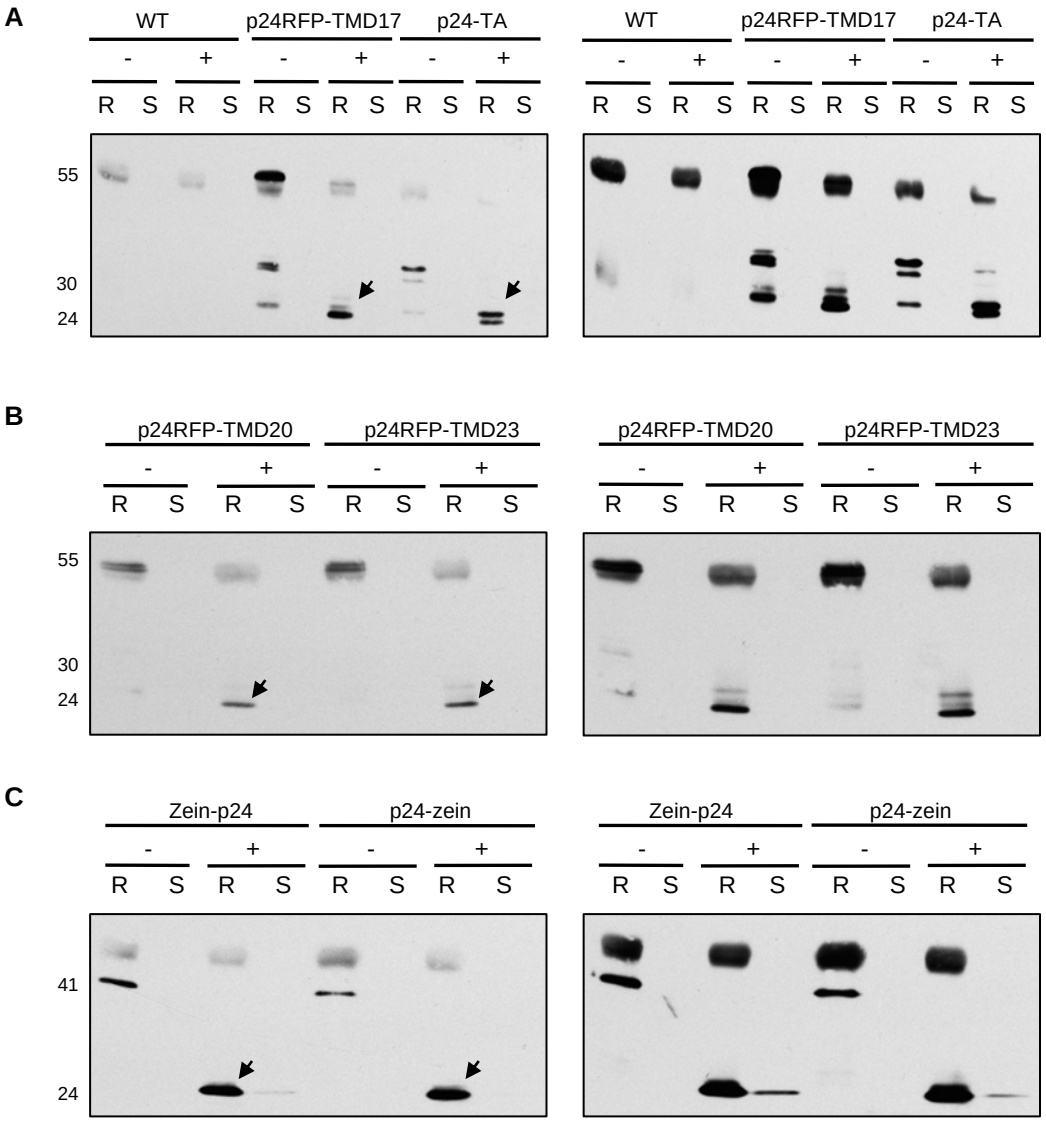


Figure 40. Immunoprecipitation using rabbit anti-p24 antibodies. Lysates from plants expressing p24 fusion proteins and wild type plants were incubated with rabbit anti-p24 antibodies. The immunocomplexes were isolated using protein A-Sepharose beads and a part of them were digested with thrombin. The samples were subjected to SDS–PAGE and after blotting the p24 was detected using a sheep anti-p24 and a donkey anti-sheep horseradish peroxidase antibodies. The arrows indicate the position of the pure p24 protein. Left panels show 1 minute exposure and right panels 5 minutes. Molecular masses are expressed in kDa. WT: wild type; R: resin; S: supernatant.

The pellets of these samples contained the antigen-antibody complexes attached to the beads and they were incubated for 16 hours with thrombin in solution or with buffer without enzyme (negative control). The suspensions were then centrifuged and the soluble fraction and the precipitate were analyzed by SDS-PAGE and Western blot using sheep anti-p24, which usually revealed a single band with the expected molecular mass of this protein.

Extracts from a wild type plant were used as negative control: the rabbit anti-p24 antibody only detected an unspecific band around 55 kDa which remained associated to the resin also after thrombin cleavage (Fig. 40).

Free p24 was released by thrombin from all recombinant constructs and remained, as expected, bound to the resin, with no or very little release in the supernatant (Fig. 40 arrows). Free p24 migrated with the expected molecular mass of 24 kDa while the different fusion proteins were detected with their corresponding molecular masses in the samples not treated with thrombin.

This purification method showed that using the rabbit anti-p24 antibody to immunoprecipitate the p24 fusions and digesting with thrombin, the fusion proteins were efficiently cleaved and the free p24 remained associated to the beads.

IV.11.2 Purification using rabbit anti-zein and mouse anti-op3

Protein lysates from zein-p24, p24-zein and wild type plants were incubated with rabbit anti-zein antibody. Protein lysates from p24-TA and wild type plants were incubated with mouse anti-op3 antibody. The antigen-antibody complexes were then immunoprecipitated using protein A-Sepharose for the rabbit anti-zein antibody samples and protein G-Sepharose for the mouse anti-op3 antibody samples. The immunocomplexes bound to the resins were incubated for 16 hours with or without thrombin. The suspensions were centrifuged and the two fractions were analyzed by SDS-PAGE and Western blot analysis using anti-p24 antibody, as in the previous thrombin cleavage experiment.

The zein-p24 and p24-zein extracts revealed a single band with the expected molecular mass of 41 kDa from the resin fraction in non-treated samples (Fig. 41 A). No fusion proteins were detected in the supernatant of non-treated samples. When the samples were digested with thrombin, free p24 was detected in the medium fraction (arrows) as expected, although some fusion proteins and free p24 were detected attached to the

resin. These results suggest the thrombin cleavage site was not exposed. Possibly, antibody binding to the zein epitopes inhibits in part access to the thrombin cleavage site. In wild type extracts only a very a faint band around 25 kDa was visible.

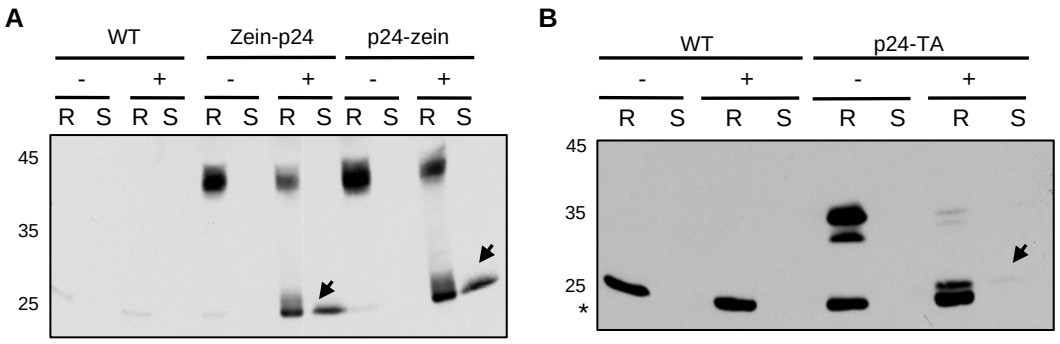


Figure 41. Immunoprecipitation analysis using rabbit anti-zein and mouse anti-op3 antibodies. A) Plant-derived zein-p24, p24-zein and wild type lysates were incubated with rabbit anti-zein antibodies. The immunocomplexes were isolated using protein A-Sepharose beads and a part were treated with thrombin (+). B) Plant-derived p24-TA and wild type lysates were incubated with mouse anti-op3 antibodies. The immunocomplexes were isolated using protein G-Sepharose beads and a part were treated with thrombin (+). All the samples were subjected to SDS–PAGE and after blotting the p24 was detected using a sheep anti-p24 and donkey anti-sheep immunoglobulin horseradish peroxidase. The arrow indicates the position of the pure p24 protein. Molecular masses are expressed in kDa; WT: wild type; R: resin; S: supernatant.

As shown in Fig 41 B, p24-TA was detected as a doublet in the resin fraction of non-treated samples corresponding to the glycosylated and unglycosylated isoforms of the protein. No fusion proteins were detected in the supernatant of non-treated samples. When the samples were digested with thrombin, free recombinant p24 was detected in the resin fraction and was only in very minor proportion released to the supernatant (arrow). This could be explained by the MHR region of p24 has been suggested to have a role in membrane affinity of GAG through hydrophobic interactions (Ogg et al. 1998). It is possible that the domain interacts with the released TMD domain of p24-TA after thrombin cleavage. Because the TA-opP3 region is associated to the resin via the antibody, this can in turn retain the removed p24. The same reasoning can be adopted for the zein-p24 and p24-zein proteins. To solve this possibility, in all the digestions of the different recombinant proteins, Triton was added to dissociate potential hydrophobic interactions. We conclude that for these experiments further investigations must be conducted to optimise the purification of the recombinant p24 like adding higher amounts of Triton in the digestion reactions.

Unspecific binding of a polypeptide around 20 kDa was observed in the wild type sample (Fig. 41 B asterisk) and the same unspecific bands were found in the different samples from the transgenic plant. In this case, their pattern was different from the pattern found when immunoprecipitating with rabbit anti-p24 or rabbit anti-zein antibodies (Fig. 40-41).

DISCUSSION

HIV-1 p24 protein as a subunit vaccine

There is a global need for an effective and affordable HIV vaccine. An ideal vaccine would induce the formation of broadly neutralizing antibodies and specific T-cell mediated immune responses (Caley et al. 1997). The p24 capsid peptide has been proposed to be a good subunit vaccine candidate as it has proved to trigger both host responses and is rather conserved among HIV clades and other immunodeficiency viruses (Matsuo et al. 1992; Hilpert et al. 1999).

Moreover, when an effective anti-HIV vaccine is developed, large-scale immunisation programs will require large amounts of vaccine (IAVI). To address this need, several approaches to produce large amounts of potential HIV subunit vaccines are under development. In this study transgenic tobacco plants have been engineered to express the main protein of the HIV-1 capsid.

Plants chosen as a production system for recombinant p24 production

The characteristics of the host for recombinant proteins expression must be considered in terms of production potential, environmental impact, food safety and human health (Sparrow et al. 2007). Plants are higher eukaryotes and thus can produce recombinant peptides and proteins similar to their counterparts that are naturally expressed in mammalian cells but cannot be produced by microbes (Ma et al. 2005). They produce a large amount of biomass and protein and their production can be increased massively by the propagation of stably transformed plants in the field. Tobacco is a well established production system and its open-field cultivation could be advantageous in terms of capacity, flexibility, scalability and production costs of recombinant pharmaceutical proteins (Boehm 2007). Tobacco is an excellent biomass and prolific seed producer and as it is not used for food or feed, and it is relatively easy to keep separate from crop products used for human or animal food chain (Sparrow et al. 2007). Tobacco is a self-pollinating crop and different techniques have already been developed for both male sterility and seed sterility of this species.

The HIV-1 p24 capsid protein has been produced in bacteria (Gupta et al. 1997; Qoronfleh 1999; Castilho et al. 2005; Bhardwaj et al. 2006; Mahboudi et al. 2006), in yeast (Vlasuk et al. 1989; Jiang et al. 2005), in insect cells (Joshua et al. 2000), in

mammalian cells (Chen et al. 2007). Although acceptable expression levels have been achieved using these expression systems, plants have been also investigated.

In plants, p24 has been transiently or stably expressed in tomato and in tobacco. Zhang et al (2000) stably transformed tomato plants using tomato bushy stunt virus but the p24 ORF was not stably maintained in the viral vector. Two years later, the same group produced p24 in transgenic tobacco plants achieving accumulation levels of 0.35% TSP in the cytosol (Zhang et al. 2002). Using Tobacco Mosaic Virus (TMV) as a vector, HIV-1 clade C p24 was inoculated into *N.benthamiana* plants obtaining yields of 100 mg per kg fresh leaf weight (Perez-Filgueira et al. 2004). Moreover, it was reported that the expression of HIV-1 clade B p24 fused to IgA in the secretory pathway of *N.tabacum* raised to accumulation levels of 1.4% TSP in transgenic tobacco plants (Obregon et al. 2006). In a recent study, HIV-1 clade C p24 and p17/p24 chimera were targeted to the ER and to the chloroplasts in tobacco plants (Meyers et al. 2008). The maximum levels of expression for p24 and p17/p24 were achieved by transient expression in tobacco chloroplasts and were about 0.3% TSP.

These studies showed that recombinant protein accumulation in tobacco leaves is relatively low when compared to the economic threshold of 1% TSP proposed by Kusnady et al. (1997). Many efforts have been put to improve transcription, translation and intracellular targeting of recombinant proteins in tobacco and other species to use them as plant-based expression systems for production of therapeutic proteins (Schillberg et al. 1999; Nuttall et al. 2005; Obregon et al. 2006; Marusic et al. 2007; Van Droogenbroeck et al. 2007; Barbante et al. 2008; de Virgilio et al. 2008; Floss et al. 2008; Irons et al. 2008; Rademacher et al. 2008; Ramessar et al. 2008; Strasser et al. 2008).

Some studies have pointed the plant secretory pathway as a suitable place to accumulate recombinant proteins (Schillberg et al. 1999; Nuttall et al. 2005; Obregon et al. 2006; Floss et al. 2008). Accumulation of the recombinant protein in a specific compartment may greatly facilitate its isolation and purification which are critical points in large scale production (Kamenarova et al. 2005).

With the aim of producing high amounts of native HIV-1 p24 (without codon usage optimisation) in tobacco plants, we compared different intracellular compartments and membrane orientations within the endomembrane system. In this work, we used 3 main strategies: p24 was fused to RFP and different TMD, to the N-terminal domain of γ-zein or to the C-terminal TA domain of cyt b5.

Localisation of recombinant p24

Each intracellular compartment of the secretory pathway has a specific proteome defined by complex sorting mechanisms that determine the final destination of proteins via anterograde or retrograde traffic. This is mediated by vesicular carriers. Coat protein complex (COP) II vesicles transport the exported cargo from ER to the Golgi complex while COPI vesicles retrieve proteins from Golgi to the ER (Bonifacino and Glick 2004). Clathrin-coated vesicles mediate traffic to endosomes, either from the Golgi complex or from the plasma membrane. A mechanism involved in the sorting of membrane proteins is the length of their TMD. The membranes of the secretory pathway have different thickness depending on the lipid-protein composition and they become thicker and more rigid from the ER towards the cell surface (Bretscher and Munro 1993; Sprong et al. 2001). The amino acid composition of a TMD determines its hydrophobicity and a correlation between increased TMD length/hydrophobicity and sorting to later compartments of the secretory pathway has been observed (Bretscher and Munro 1993; Brandizzi et al. 2002). Statistical analyses in which features of the TMD are correlated to specific subcellular localisation have suggested that retention of proteins in the ER, Golgi or plasma membrane depends in many cases on the TMD length (Munro 1995; Pedrazzini et al. 1996; Fu and Kreibich 2000; Pedrazzini et al. 2000; Brandizzi et al. 2002). However, the retention mechanism remains poorly defined as there are no general rules to accumulate proteins in a specific compartment (Twyman et al. 2003; Fischer et al. 2004; Kamenarova et al. 2005) .

Recombinant p24 fused to RFP and TMD accumulated in moderate levels in the secretory pathway

We fused recombinant p24 to TMD of different lengths to compare its accumulation in various compartments of the plant secretory pathway, as a type I membrane protein in which the HIV sequence is exposed in the compartment lumen. Transient protein expression has proved a rapid alternative methodology before stable plant transformation (Sparkes et al. 2006). Several methods are used to transiently transform cells, two widely used examples being protoplasts transformation and *Agrobacterium*-mediated leaf infiltration. Both methods achieve high levels of recombinant protein for a short time; however they are not commonly used for commercial-scale production

(Ma et al. 2003; Twyman et al. 2003; Fischer et al. 2004). These experimental procedures are relatively simple and inexpensive and their results can give a preliminary estimation of the production of the recombinant protein before preparing transgenic plants.

Tobacco protoplasts and leaf agroinfiltration were used to initially monitor the subcellular localisation of p24 fluorescent fusion proteins. The results showed that RFP-TMDs and p24RFP-TMDs recombinant proteins were targeted to the expected compartment depending on their TMD, in agreement with the results obtained by Brandizzi et al. (2002). In contrast, transient expression of RFPp24-TMDs fusion proteins resulted in major delivery of the recombinant proteins to the vacuole with some minor localisation in the ER. We hypothesize that the proteins had folding defects, and were recognised by ER protein quality control and mainly directed to the vacuole. Newly synthesised proteins that enter the ER are assisted by several chaperones and folding helper enzymes (Matlack et al. 1999). When a protein is not correctly folded, the ER quality control machinery will often target the malfoded protein to the cytosol to be degraded by the ubiquitine-proteasome system, a pathway known as ER-associated degradation (Ellgaard et al. 1999; Lord et al. 2000; Hampton 2002). However, by studying the behaviour of defective secretory proteins in yeast, it has been shown that misfolded proteins can also be directed to the vacuole for degradation as an alternative location for disposal (Spear and Ng 2003). Moreover, recent studies suggested that the ER chaperone BiP may play an active role in what could be a similar vacuolar quality control in plant cells, highlighting the possible existence of multiple quality control mechanisms in the secretory pathway (Pimpl et al. 2006; Vitale and Boston 2008). Thus our results strengthen this alternative as even recombinant proteins expected to be localised in the ER (RFPp24-TMD17) present some mistargeting to the vacuole, suggesting that this vacuolar sorting is not strongly influenced by the TMD length and most probably due to misfolding.

When the RFP was replaced by GFP, the N-terminal GFP-tagged constructs (GFPp24-TMD) also presented mistargeting in transiently transformed protoplasts. Furthermore, the RFPp24TMD-GFP recombinant proteins did not show conclusive results but they revealed that a fragment of the fusion proteins containing the lumen-facing RFP was partially located in the ER and partially in the vacuole whereas another fragment of the protein containing the cytosol-facing GFP was only observed in the ER. Western blot analysis revealed the fusion proteins were detectable by both anti-p24 and

anti-GFP antibodies but some degradation products were observed when using the latter.

The data presented regarding transient expression allow us to conclude that transmembrane domain length can modulate sorting between the ER, Golgi and plasma membrane. The strongest argument is the observation that different constructs containing transmembrane domains were differentially localised on the basis of the length of their TMDs. However, there was mislocalisation of the N-terminal XFP-tagged constructs due either to a position effect of the XFP, by which the TMD was not the only determinant to sort the recombinant protein to the expected compartment, or due to a position effect of p24. The p24 contains a hydrophobic domain at the C-terminus called major homology region (MHR) which is highly conserved and allows protein dimerization (Ogg et al. 1998). It is possible that when p24 is at the C-terminus of the construct and close to the TMD, putative interactions of this region may not allow the correct folding of the recombinant protein, independently of the XFP at its N-terminus, and the malfolded protein is recognised by protein quality control and directed to the vacuole.

For the p24RFP-TMD constructs, we also used stable nuclear transformation of tobacco plants. Transgenic tobacco plants were generated using the constructs containing the appropriate targeting signals to achieve accumulation of the recombinant p24 in ER, Golgi or plasma membrane. Leaves of transgenic tobacco plants expressing p24RFP-TMD17, p24RFP-TMD20 and p24RFP-TMD23 reached accumulation levels of about 0.3% TSP for the ER-targeted recombinant proteins (p24RFP-TMD17) and of 0.15% TSP for the Golgi-targeted (p24RFP-TMD20) and plasma membrane-targeted proteins (p24RFP-TMD23) indicating the TMD17 allows higher accumulation than the TMD20 and TMD23. For all three constructs, the variation in transgene expression was low, but high-accumulating lines could still be identified in a small group of 25 to 30 independent transformants. These results are comparable to the cytosolic p24 expression levels of 0.35% TSP achieved in transgenic tobacco plants and to the 0.3% TSP in transiently transformed tobacco chloroplasts (Zhang et al. 2002; Meyers et al. 2008).

To our knowledge, this is the first study in which a pharmaceutical recombinant protein is specifically targeted to Golgi stacks and to plasma membrane for high yield production in plants. Although anchoring a fusion protein to a membrane for accumulation purposes is not a new technique, the ER is usually the chosen organelle. The ER provides an oxidizing environment, a neutral pH and there are few proteases.

Accumulation of recombinant proteins in the ER via de H/KDEL was demonstrated to be safe in terms of stability and accumulation in comparison to other organelles (Kamenarova et al. 2005; Vitale and Pedrazzini 2005). However, the KDEL peptide must interact with its receptor, which can lead to saturation when expressing high quantities of recombinant proteins (Crofts et al. 1999). Moreover, the introduction of the H/KDEL in the amino acid sequence of p24 can affect its activity. In this study, we wanted to explore other strategies to accumulate recombinant proteins in the plant ER without using the KDEL signal and saturating the retrieval system.

Recombinant p24 fused to C-terminus cyt b5 TA

As mentioned above, HIV-1 p24 was expressed in the cytosol of transgenic tobacco plants. The first report describes accumulation levels of 0.35% TSP (Zhang et al. 2002) whereas in the second report the accumulation was very low (0.04 µg p24/kg fresh leaf weight) (Meyers et al. 2008). In a previous study, the cytosolic protein HIV-1 Nef was expressed free in the cytosol or anchored to the ER membrane by fusion to the TA of cyt b5. In transgenic tobacco plants, cytosolic soluble Nef achieved accumulation levels of 0.2% TSP whereas Nef-TA was 0.7% TSP (Marusic et al. 2007; Barbante et al. 2008). This higher accumulation, consistently observed in many independent transgenic plants, was due to increased protein stability (Barbante et al., 2008).

In an effort to improve p24 accumulation still maintaining it in its natural cytosolic folding environment, we expressed the HIV protein as a TA recombinant protein, using the same TA that was previously fused to Nef. In plant cells, the hydrophobic TA of rabbit cyt b5 leads to the recruitment of recombinant proteins to the cytosolic face of the ER, which is the natural location of the isoform of cyt b5 from which the TA was derived (Maggio et al. 2007). As in Nef-TA, the C-terminal cyt b5 TA was modified to add an *N*-glycosylation site. This site was previously used as a marker for the proper translocation of the TA to the ER membrane and therefore to verify the subcellular localisation and membrane topology of the recombinant protein (Pedrazzini et al. 2000). Our recombinant p24-TA protein was engineered to have the TA tail cleaved off and the p24, as expected to remain in the ER, will not contain any plant glycan that could potentially produce some allergic reaction.

Our p24-TA was expressed to levels of 0.15% TSP in transgenic tobacco plants. This was about half of the level of accumulation reported for cytosolic p24 by Zhang et al.

(2002), but it should be considered that we do not know whether the mRNA levels are comparable for the two constructs. The mRNA levels of p24-TA were consistently lower than the average levels of the TMD fusions that we produced. Our results however indicate that p24-TA protein accumulation is very similar to those of the TMD20 and TMD23 constructs and lower than the ER-localised TMD fusion. Our pulse-chase experiments indeed indicated that p24-TA is more stable than any of the p24RFP-TMD fusions, thus providing an explanation for similar accumulation in the presence of fewer transcripts. Therefore, by a protein stability point of view, the TA strategy seems promising also in the case of p24-TA. The negative aspect of low mRNA levels asks for more investigation.

High accumulation and stability of recombinant p24 fused to γ-zein

The ER is also naturally the starting point for the biogenesis of the other compartments of the endomembrane system, which have multiple functions in plant development and in response to the environment (Galili 2004). Some seed storage proteins belonging to the prolamin class are able to accumulate in the ER by the formation of large polymers termed protein bodies (Vitale and Ceriotti 2004). Although the mechanisms that determine the retention of prolamins in the ER are still not fully clear (Vitale and Ceriotti 2004), investigations on maize γ-zein are providing important insights. Maize γ-zein is a seed storage prolamin that polymerizes due to the formation of inter-chain disulfide bonds, and it has been shown to form PB also in vegetative tissues when constitutive expressed in *Arabidopsis* plants (Geli et al. 1994). The N-terminal PPPVHL repeat domain of γ-zein is sufficient for ER retention, in contrast to the cysteine rich C-terminal domain that is necessary to direct PB formation but is secreted when expressed alone (Geli et al. 1994). It was observed that the N-terminal half of γ-zein (PPPVHL repeat domain and the following short proline-rich non-repeated domain) has a dominant effect on the intracellular traffic of the bean vacuolar storage phaseolin. The fusion protein, termed zeolin, is able to form PB (Mainieri et al. 2004). This ability was related to the polymerization due to the Cys residues present in the zein domains (Pompa and Vitale 2006). As described above this section, the K/HDEL ER localisation signal is widely used to increase accumulation of foreign proteins in transgenic plants because the ER has low hydrolytic activity. However, a proportion of different K/HDEL fusion proteins has been found to be slowly delivered

to the vacuole by a pathway that does not seem to be mediated by the Golgi complex (Frigerio et al. 2001; Tamura et al. 2004).

We explored the possibility to accumulate recombinant p24 in ER-derived protein bodies by fusion of the protein to the N-terminal γ-zein region. Our transgenic tobacco plants expressing zein-p24 and p24-zein showed accumulation levels of about 1% TSP in the best expressor lines. This is consistent with the accumulation levels of more than 1% TSP observed for Nef fused to zeolin in tobacco plants (De Virgilio et al., 2008). In that study, Nef was also fused to the C- or N-terminus of the N-terminal domain of γ-zein but the recombinant protein was not significantly expressed, suggesting that proper folding and expression of γ-zein fusions depend on the fused protein.

Stability of recombinant proteins

A high yield production of recombinant proteins is one of the main goals in molecular farming. To achieve high levels, the construct design must optimize all stages of gene expression, from transcription to protein stability (Ma et al. 2003).

Generally, plant-derived proteins produced in transgenic tobacco are extracted from leaves. Using this expression system, recombinant proteins are produced at low levels, usually less than 0.1% of the total soluble protein. This low level of production probably reflects a combination of factors, particularly poor protein folding and stability (Ma et al. 2003). Proteases play an important role in degradation of abnormal or incorrectly processed proteins. Accordingly, the susceptibility of heterologous proteins to protease attack in plant cells could reflect their imperfect synthesis or assembly (Doran 2006). Moreover, plant and animal glycans are different and this feature could affect the final structure of foreign proteins and consequently plant-derived proteins could be more vulnerable to protease activity than their animal-derived counterparts. One example described by Stevens et al. (2000), shows that an IgG1 antibody produced by mouse hybridoma cells was more stable when incubated with tobacco leaf extract than the same antibody produced in plants. These results suggest the degradation was carried out by acidic proteases present in the plant extract and the glycans of the mouse-derived antibody conferred protection of the recombinant protein, which was not conferred by the glycans of the plant-derived antibody. This may indicate that the plant type of *N*-glycosylation contributes less to the stability of the antibody than the mouse type of *N*-glycosylation.

In our study, the half-life of all the plant-derived p24RFP-TMD recombinant proteins was determined to be between 2 and less than 8 h after chase when targeted to the plant secretory pathway. Similar results were observed for HIV-1 Nef protein directed to the tobacco secretory pathway, with levels decreasing sharply during a 5-hour chase period (Marusic et al. 2007) however our p24RFP-TMD17 was more stable (between 4 and 8 h post-chase). Moreover, recombinant full-size antibodies produced in tobacco showed that the secretory pathway is more advantageous for correct protein folding and assembly than other compartments like the cytosol (Schillberg et al. 1999).

Our strategy by which p24 attached to different TMD was accumulated in the lumen of various organelles of the plant secretory pathway did not avoid the degradation of the recombinant fusions. None of these recombinant fusions contained an *N*-glycosylation site indicating that the degradation process of the proteins could not be due to glycases between the ER and the Golgi.

P24 is a cytosolic protein which contains two cysteines. In the cytosol, cysteines are normally reduced but when nascent polypeptides are inserted into the ER, its redox balance stimulates disulphide bond formation (oxidation) while catalyzing the rearrangement of incorrect disulphides (Wilkinson and Gilbert 2004). It is possible that our plant-derived recombinant p24RFP-TMDs undergo oxidation of the two cysteines of p24 causing a degree of misfolding of the recombinant protein. However, this incorrect folding seem to be minor and anyway not efficiently recognised by ER quality control, as the recombinant fusions are still able to reach their final destination (ER, Golgi or plasma membrane) as observed by confocal microscopy. Minor misfolding could however make the recombinant proteins more susceptible to protease attack once they reach compartments that allow easy access by proteases. The fact that P24RFP-TMD17 had longer half-life than the other TMD fusions is in agreement with this hypothesis, since the ER is less hydrolytic that the other compartments of the endomembrane system.

P24RFP-TMD17 is anchored facing the lumen of the ER (type I membrane protein) whereas p24-TA is facing the cytosol (type IV membrane protein). Pulse-chase experiments indicate higher stability of the latter than the former. This may be related to more accurate p24 folding in its natural cytosolic environment. An almost equal proportion of free and membrane-bound p24 was detected in the p24-TA expressing plants indicating that some cleavage takes place. This suggests some early proteolytic

event, but the free p24 was rather stable over the chase time (8 h) again suggesting proper folding.

Another practical concern for protein expression and purification is the aggregation of the recombinant proteins. Aggregation may become a critical bottleneck and is typically addressed by protein engineering through a combination of random and designed alterations in the expressed protein sequence (Mezzasalma et al. 2007). However, promoting the formation of polymers can be a positive approach to stabilize recombinant proteins and avoid protease activity (Obregon et al. 2006; Floss et al. 2008). Mainieri et al (2004) described the fusion of bean phaseolin with the N-terminal domain of maize γ-zein resulted in a very stable recombinant protein (zeolin) that polymerizes in the ER forming PB. In this case, the mechanism of localisation in the ER may be a retention system operating through low solubility and direct interactions with lipids or prolonged association with chaperones (Li et al. 1993; Muench et al. 1997; Kogan et al. 2004).

In our strategy in which p24 was fused to the N-terminal γ-zein domain, polymerization of the recombinant zein-p24 and p24-zein proteins was suggested by the high molecular weight components observed by Western blot, very similar to those observed for zeolin (Mainieri et al. 2004), and most probably due to incomplete denaturation when samples containing high amount of the recombinant protein are analyzed. In our case, both zein-p24 and p24-zein were more stable than any of the other p24 constructs we have analyzed here, indicating the recombinant proteins were not recognized as defective by the ER quality control system and protected from protease activity.

Although the strategy to fuse a protein of interest to the N-terminal domain of γ-zein seems a promising platform to accumulate recombinant proteins, not every protein of interest accumulates as expected. De Virgilio et al. (2008) described the fusion of Nef at the C-terminus or N-terminus of the N-terminal domain of γ-zein. In both cases, the recombinant proteins were very unstable and degraded by ER quality control. Almost no protein was detected by Western blot. However, when Nef was fused to zeolin, the heterologous protein was stable and accumulated into ER-derived protein bodies.

Chloroplasts represent a good alternative to accumulate recombinant proteins. They are specifically transformed using particle bombardment leading to transplastomic rather than transgenic plants (Boehm 2007). Once stably integrated, transgenes express large amounts of proteins (up to 47% TSP) due to the high copy number (>10,000

copies) of the chloroplast genome in each plant cell (Daniell 2006). HIV-1 clade C p24 and a chimera of p17/p24 have been recently expressed in tobacco chloroplasts achieving levels of 636-2994 μg p24/kg fresh weight for p24 (the highest value ~0.2% TSP) and 5-230 μg p24/kg for p17/p24 (the highest value ~0.017% TSP) in transplastomic plants (Meyers et al. 2008). The same constructs were expressed in *Agrobacterium*-mediated transient expression in chloroplasts and they showed levels of accumulation of 937-4014 μg p24/kg of fresh weight for p24 (the highest value ~0.3% TSP), and 4800 μg p24/kg for p17/p24 (~0.36% TSP). Although the highest expression in stably transformed plants was achieved for the p24 protein, the p17/p24 chimera was chosen for a potential protein subunit vaccine as it contains more epitopes. In that study, further experiments were carried out with transient expression of p17/p24 which presented the highest accumulation with respect to the p17/p24 transplastomic plants. It must be underlined that all these constructs showed accumulation levels lower than our zein fusions.

Purification of the different plant-derived recombinant p24

The cost of drug production is most significantly influenced by the aspects of downstream processing, product purification and manufacturing under good manufacturing practices (GMP) standards. The overall technology costs arise once the rigour of current GMP is applied as required for delivering a human biological product (Kirk and Webb 2005).

One example of economic evaluation of the downstream process is the production of β-glucuronidase in transgenic corn. The annual operating costs have been calculated corresponding to 6% for milling, 40% for protein extraction and 48% for protein purification yielding a production cost of \$43 g^{-1} for an initial seed protein concentration of 0.015% dry weight, a product purity of 83%, and an annual production volume of 137 kg (Evangelista et al. 1998). A more recent study in which lactoferrin was produced in rice estimated that it would cost about \$6 to generate 1 g of recombinant lactoferrin from rice flour in a GMP facility operating at a scale of 600 kg/year (Nandi et al. 2005).

In this study, we tested the feasibility of removing p24 from the added targeting sequences by in vitro proteolysis of the affinity purified fusion proteins. Since our fusion proteins did not contain any His- or GST-tags the different heterologous proteins

were incubated with anti-p24 or anti-zein and anti-op3 polyclonal antibodies and immunoprecipitated using Sepharose-conjugated protein A or G. The immunoprecipitated complexes were then digested with the protease thrombin to release the fusion tags/anchors and maintain the free p24 soluble in the medium or attached to the beads-conjugated antibodies. The recombinant p24 fusions and the free p24 proteins obtained with the different strategies retained their ability to be recognized in their undenatured forms, as demonstrated by immunoprecipitation. Proteolytic cleavage was in most cases efficient and the pure recombinant p24 was nearly completely released from the fusion tags, demonstrating potential viability of this procedure for purification purposes.

An alternative strategy to purify our fusion proteins is gradient centrifugation. Our plant-derived recombinant proteins contain TMD, TA or domains that lead to the formation of insoluble polymers, which anchor or associate the fusion proteins to different membranes of the cell. The original idea to purify these recombinant proteins was to use purification by gradient centrifugation to separate the different organelles. Once we have the target organelle fraction in a small volume, the extraction of recombinant p24 is carried out by cleavage with thrombin releasing the p24 in the medium and the tag fusions remain associated to the membranes. Although this experimental procedure is feasible in a small scale, we did not use this purification method as scalability of centrifugation for large volumes of cell extracts is not viable (Roe 2001).

We have to consider that the introduction of a protease cleavage site, in our case thrombin, makes the whole purification process expensive. In our purification method, we used an antibody to ligate the antigen, a second step using Sepharose-conjugated protein A or G and finally digestion with thrombin. Although we have demonstrated the efficiency of our method, the high cost of the enzyme contributes to increase the final cost of the whole purification system. An alternative method to avoid protease cleavage in the purification step is to introduce a self-catalytic sequence like intein in the fusion protein (Perler 2005). The advantages offered by the self-cleavage of intein can be further optimised by fusion of the chitin binding domain (CBD) as an affinity tag. Tobacco plants expressing recombinant SMAP-29, a mammalian antimicrobial peptide of innate immunity, was fused to intein-CBD and the recombinant protein could be purified by affinity chromatography and the peptide was released by inducing intein self-cleaving with nucleophilic agents (Morassutti et al. 2002).

A recent effective and inexpensive fusion expression system consists in the fusion of the protein of interest to annexin B1-intein tag (Ding et al. 2007). The fusion protein is purified in a single-step method based on the Ca^{2+}-binding activity of annexin B1, and the annexin B1-intein fusion tag is removed based on the self-cleaving activity of the intein. Moreover, another single-process expression and purification method has been developed by New England Biolabs in which 3 different recombinant proteins can be purified from the same mixture by using a combination of a multihistidine and a modified intein as affinity tags and Ni Sepharose and chitin as affinity matrices (Porte and Chong 2008).

Other strategies have been developed to accumulate recombinant proteins taking advantage of the intracellular organelles. Oleosin has been used to trap recombinant proteins in ER-derived oil bodies (Capuano et al. 2007). Recombinant proteins accumulating in oil bodies that are expressed in different organisms can be purified using proteases (Parmenter et al. 1995) or using the intein properties described recently to avoid enzyme cleavage and high costs (Chiang et al. 2007).

Moreover, secretion to the extracellular medium has been also described as a promising approach to produce recombinant proteins (Gaume et al. 2003). However, we did not consider this strategy as our aim was to explore the accumulation of p24 within different membranes of the plant secretory pathway.

Conclusions and future perspectives

Plants offer many advantages compared to traditional systems for the production of pharmaceutical proteins. These include rapid scalability, low cost of production (plants mainteinance), absence of human pathogens and the ability to properly fold, assemble and posttranslationally modify complex proteins (Ma et al. 2003).

The subcellular targeting of the recombinant protein plays a key role in the production yield. Understanding how proteins are intracellularly targeted is important when attempting to exploit the plant secretory pathway for heterologous protein production. In this work we showed that the HIV-1 p24 protein can be specifically expressed and accumulated in different organelles in tobacco. Our experiments showed the advantages and disadvantages of associating a cytosolic protein like p24 to membranes or promoting PB formation, as suggested by the electron microscopy images. In transient expression, this membrane association resulted in some cases to be effective for the

accumulation of the recombinant fusions (p24RFP-TMD constructs) but sometimes the folding and assembly was not correct and the malfoded recombinant proteins were targeted to the vacuole (RFPp24-TMD constructs).

Most of the protein yields reported in this study were below the expected economic threshold (1% TSP) in transgenic tobacco plants expressing the different fusion proteins. However, the strategies using γ-zein fusions are the most promising for the accumulation of p24. These fusion proteins presented the highest productivity and stability whose yields were 1% TSP. Furthermore, the targeting of the p24RFP-TMD to the secretory pathway resulted in the production of unstable recombinant fusions while the zein-p24 and p24-zein fusions were very stable, presumably accumulating in ER-derived protein bodies and suggesting that polymerization protects p24. This protection can in turn be due to very efficient retention in the ER but also to masking of proteolysis sensitive sites in polymers and exclusion of PB from the normal physiological turnover of the ER. The cytosolic ER-anchored p24 (p24-TA) was also more stable than the TMD-containing constructs, but its accumulation was not higher, most probably because of the low mRNA levels. The reason for this, as well as the reason for the particularly high accumulation of the mRNAs of the zein fusions, are not known and will be interesting to investigate.

We can conclude that the strategies for accumulating recombinant proteins in plants are antigen-dependent. In tobacco, using the same fusion tags with different pharmaceutical proteins, the levels of expression, accumulation and stability varied (Marusic et al. 2007; Barbante et al. 2008; de Virgilio et al. 2008). In a previous report (Obregon et al. 2006), HIV-1 p24 containing the same p24 DNA sequence used in our constructs accumulated in the tobacco secretory pathway with levels of 1.4% TSP and the recombinant protein was stable however our plant-derived p24RFP-TMDs were unstable and expressed in low levels (0.15 - 0.3% TSP).

In this study, we have tested so far only the T0 generation of transgenic plants which are hemizygous. In homozygous plants of the T1 generation protein accumulation is expected to be higher. It has been described that tobacco seeds expressing recombinant glycoprotein B of human cytomegalovirus showed 30-fold greater specific activity in the T1 generation compared to the best T0 plants from the same transformation demonstrating enhanced levels of recombinant protein expression in a homozygous second generation plant line (Tackaberry et al. 2003). At the moment, we have collected

T1 seeds, and they will be tested to assess the 3:1 ratio and the homozygous lines will be selected for protein expression and stability.

Our p24 was not optimised for tobacco codon usage. We used the same p24 DNA sequence as Obregon et al. (2005) which was demonstrated to yield expression of 1.4% TSP when expressing the protein in the plant secretory pathway and for this reason we considered there was no need to change it. A recent study described that optimisation of HIV-1 clade C p24 for *N.tabacum* and for human codon usage did not increase the yield when transiently expressed in a non-specific compartment in tobacco plants (Meyers et al. 2008). This indicates that optimisation of the sequence does not always lead to enhancement of the expression.

In this study we used the 35S Cauliflower Mosaic Virus (CaMV) promoter in all the constructs to transiently or stably transform tobacco plants. The promoter used in the constructs to express recombinant proteins obviously plays an important role. Usually, constitutive promoters that drive the expression of the recombinant proteins in most of the plant tissues are used. The 35S CaMV promoter is the promoter of choice for dicotyledonous plant, as it is strong and constitutive (Odell et al. 1985) and newer versions have enhanced its activity (Kay et al. 1987; Dowson Day et al. 1993). However, this promoter presents a low activity in monocotyledons and other constitutive promoters are used for these plants e.g. the ubiquitin promoter for maize (Christensen and Quail 1996) and the *gos-2* promoter for rice (de Pater et al. 1992).

Another important factor is translation efficiency. The 5'-UTR (5' untranslated region) and other sequences around the AUG translation initiation codon are important for the efficient translation in plants (Kozak 1986). When this sequence is cloned before the coding sequence of interest, it can lead to an enhancement of the expression due to the increase of translation initiation. Usually 5'-UTR region from viral RNAs are used, e.g. from TMV or Tobacco Etch Virus (TEV) (Gallie et al. 1991; Niepel and Gallie 1999; Hongmin et al. 2000) but other sequences have shown the same advantages (Zou et al. 2003). In our constructs, only the zein-p24 construct contained the 5'-UTR region of the tobacco chitinase and there were no significant differences in the expression levels compared to the p24-zein which did not contain this sequence. The 5'-UTR tobacco chitinase region was chosen because it is a native tobacco sequence.

This thesis is a proof-of-concept regarding feasibility of specific targeting of a recombinant protein to different compartments of the secretory pathway for high yield production in plants. The most promising strategy is the fusion of p24 to the N-terminal

domain of maize γ-zein, either N-terminally- or C-terminally-tagged. Further improvements may be done to optimise the downstream processing as at the moment purification of the recombinant protein by using antibodies and Sepharose-A is not optimal for large scale production. For example, a His-tag could be added to the recombinant protein to facilitate its purification.

MATERIALS AND METHODS

I Water and sterilization

Double distilled water was used in all solutions. The sterilization of the solutions was carried out by autoclaving at 15 p.s.i. (1.05 Kg/cm^2) for 20 min at 121ºC or by filtration.

II Bacterial manipulations

II.1 Bacterial strains

The following *Escherichia coli* strains were used throughout this work to replicate and isolate the recombinant plasmids and for protein expression.

Strain	Genotype	Reference
XL-1 Blue	recA, endA1, gyrA96, thi, hsdR17 (rk-,mk+), supE44, relA1, λ-, lac-, F', proAB$^+$, lacI^{q}Δ(lacZ)M15, Tn10(tet)]	(Bullock 1987)
MC1061	F$^-$ *araD139* Δ*(ara-leu)7696 galE15 galK16* Δ*(lac)X74 rpsL* (Str^r) *hsdR2* ($r_K^-m_K^+$) *mcrA mcrB1*	(Raleigh et al. 1989)
DH5α	F$^-$φ80d*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 deoR *rec*A1 *end*A1 *hsd*R17(r_k^-,m_k^+) *pho*A *sup*E44 λ$^-$ *thi*-1 *gyr*A96 *rel*A1	(Woodcock et al. 1989)

Table I. *E.coli* strains used in this work.

Agrobacterium GV3101 strain was used for *Agrobacterium*-mediated gene transfer (Koncz 1986). This strain contains a chromosomal resistance to rifampicin (Rif) and the T-helper plasmid confers resistance to gentamycin (Gen). It also carries the pSOUP plasmid which is a disarmed Ti plasmid that possesses the *vir* genes needed for T-DNA transfer but has no functional T-DNA region by its own. The pSOUP plasmid confers resistance to tetracyclin (Tet) in bacteria.

II.2 Bacterial growth maintenance

All the *E.coli* bacterial cultures were grown at 37ºC in Luria-Bertani (LB) broth medium (10 g/L bacto tryptone, 10 g/L NaCl and 5 g/L bacto yeast extract) in agitation. To prepare agar plates, the broth medium was supplemented with 15 g/L of bacto-agar. The agar mixture solution was autoclaved, cooled down until 55ºC and the different antibiotics were added. The plates were poured under sterile conditions and a tip was used to remove bubbles.

Agrobacterium was grown in both LB or MGL media (MGL: 2.5 g/L bacto yeast extract, 5 g/L bacto tryptone, 5 g/L NaCl, 5 g mannitol, 1.16 g/L Na glutamate, 0.25 g/L KH_2PO_4, 0.1 g/L $MgSO_4$, 1 mg/L biotin; pH 7.0). For liquid culture, Agrobacteria were incubated at 28ºC for 36-48 h in agitation in liquid medium supplemented with Rif and kanamycin (Kan). The agar plates were prepared as for *E.coli.*

II.3 Preparation of competent E.coli bacterial cells for heat-shock transformation

II.3.1 Competent XL-1 blue *E.coli* cells using $CaCl_2$

One colony was picked from a freshly grown *E.*coli on a LB-Tet plate and incubated in 3 ml of LB broth medium containing 25 μg/ml Tet at 37ºC O/N with agitation. After inoculation of 1 ml of the O/N culture in 100 ml of LB broth medium containing Tet, the new culture was incubated at 37ºC with agitation until reaching an $O.D._{600nm}$ (optical density) of 0.4-0.6. The cells were incubated on ice for 15 min and pelleted in sterile Sorvall bottles (DuPont, Wilmington, DE, USA) at 4,000 x *g* for 15 min at 4ºC. 16 ml of cold RF-1 solution (100 mM KCl, 50 mM $MnCl_2{\cdot}4H_2O$, 30 mM KOAc pH 7.5, 10 mM $CaCl_2{\cdot}6H_2O$, 15% glycerol, pH 5.8, filter sterilized) were added, slowly resuspended and incubated for 15 min on ice. Cells were harvested at 4,000 x *g* for 15 min at 4ºC, resuspended in 4 ml of cold RF-2 solution (10 mM MOPS pH 6.8, 10 mM KCl, 50 mM $CaCl_2{\cdot}6H_2O$, 15% glycerol, pH 6.8, filter sterilized) and incubated on ice for 15 min. The cells were aliquoted with 100 μl per tube, frozen in liquid nitrogen and the final stock was kept at -80ºC.

II.3.2 Competent MC1061 and DH5α *E.coli* strains using RbCl

E.coli MC1061 and DH5α were streaked out on LB plates without any selection. One colony was picked and incubated in 3 ml 2X YT medium (16 g/L bacto tryptone, 10 g/L bacto yeast extract, 5 g/L NaCl; pH 7.0 using NaOH and autoclaved) in a 50 ml Falcon tube and incubated at 37ºC in agitation. At O.D. 550 nm 0.3, the 3 ml culture were poured into 200 ml of pre-warmed 2X YT medium and incubated at 37ºC. At O.D. 550 nm of 0.48, the culture was transferred into four 50 ml sterile Falcon tubes and placed on ice for 5 min. The tubes were centrifuged at 1560 x *g* in a swing-out rotor at 4ºC for 20 min. The supernatant was discarded and the cells were resuspended in 80 ml

of ice cold TFBI (30 mM $KC_2H_3O_2$, 100 mM RbCl, 10 mM $CaCl_2 \cdot 2H_2O$, 50 mM $MnCl_2 \cdot 4H_2O$, 15% v/v glycerol; pH 5.8 using 0.2 M CH_3COOH; filter sterilization and stored at 4ºC) and placed on ice for 5 min. The tubes were centrifuged as before and the cells resuspended in 8 ml of TFBII (10 mM MOPS, 10 mM RbCl, 75 mM $CaCl_2 \cdot 2H_2O$, 15% v/v glycerol; pH 6.6 using 5M KOH, filter sterilization and kept at 4ºC) and left on ice for 15 min. Using pre-chilled pipette tips and working in the cold room, the cells were aliquoted with 100 µl per tube in pre-chilled microcentrifuge tubes. The tubes were frozen in dry ice and stored at -80ºC.

II.4 *E.coli* transformation by heat shock

The bacterial cells were chilled at 4ºC prior to transformation. The plasmid DNA or ligation reaction were mixed with the bacteria and incubated on ice for 10 min. The mixture was subjected to heat shock by incubating the tube at 42ºC for 45 seconds and placed again on ice for 2 min. 900 µl of room temperature (RT) LB broth medium without antibiotics were added and the tube was incubated at 37ºC for 15 min with agitation. 100-150 µl of the cell suspension were plated on LB agar plates supplemented with 50 µg/ml of ampicillin (Amp) and incubated at 37ºC O/N.

For the transformation with the binary vector, the bacterial cells were incubated with the ligation mixture on ice for 10 min. The heat shock was for 90 seconds at 42ºC and then placed back on ice for 2 min. The volume added of LB without antibiotic was equal as bacteria volume (100 µl) and the cells were incubated at 37ºC for 30 min. 150 µl were plated on LB agar plates supplemented with 50 µg/ml of Kan and the plates were incubated at 37ºC O/N.

II.5 Preparation of competent *Agrobacterium tumefaciens* for electroporation

To prepare competent GV3101 *Agrobacterium* cells, 8 ml of overnight *Agrobacterium* culture were inoculated in 192 ml of LB low-salt broth medium with 50 µg/ml of Rif, 25 µg/ml Gen and 10 µg/ml Tet. The cells were incubated with agitation at 28ºC until an $O.D._{600nm}$ of about 0.5 (generally 3-6 h, very strain-dependent). After centrifugation at 2772 x *g* for 15 min at 4ºC, the pellets were pooled and resuspended in a final volume of 10 ml of ice-cold 10 mM Tris-HCl (pH 7.5). The cells were

centrifuged again at 2772 x *g* for 15 min at 4ºC and finally the pellet was resuspended in 2 ml ice-cold 10% glycerol. 100 µl aliquots were prepared and the tubes were frozen in liquid N_2 and kept at -80ºC; these cells stay competent for 6 months if kept at this temperature.

II.6 *Agrobacterium* transformation by electroporation

1 µg of recombinant pGREEN0229 and pGREEN0179 binary vectors (Kan^r) was added to 50 µl of competent GV3101 *Agrobacterium* cells. The mixture was incubated on ice for 1 min, transferred into a sterile cuvette (BioRad, Munich, Germany) and electroporated a 400Ω, 25 µF and 2.5 kV. 1 ml of LB broth was added and the cells were incubated at 28ºC for 3 h with agitation. Different quantities of the bacterial mixture (50 µl, 100 µl and the rest) were plated on LB agar plates containing the selective antibiotics (50 µg/ml Kan, 5 µg/ml Tet, 50 µg/ml Gen and 20 µg/ml Rif). The plates were incubated at 28ºC for 3 days.

II.7 Bacterial stocks maintenance

The competent bacterial stocks and the recombinant bacterial stocks were stored in 20% glycerol at –80ºC. To recover a strain, a small amount of the frozen stock was scraped with a sterile tip and grown at the convenient temperature in LB broth medium containing the specific antibiotic.

III Cloning methods

III.1 Plasmid vectors

III.1.1 pBlueScript II SK (+/-)

HIV-1 p24 sequence was inserted into pBluescript II SK (+/-) (Stratagene, GenBank: X52327). This plasmid has a size of 3 kb and contains Amp reporter marker, lacZ reporter gene which allows blue/white screening of recombinant plasmids, pBR322 promoter (pUC), *Sac*I and *Kpn*I sites flanking the multiple cloning site, T7 and T3 promoters and the origin of replication of filamentous phage f1.

III.1.2 pGEM®-T Easy

For cloning PCR products, pGEM®-T Easy vector system was used (Promega). The pGEM-T Easy vector is a linearized vector with 5'-T overhangs used for annealing the 3'-A overhangs generated by *Taq* polymerase during PCR. After a standard PCR reaction, the products might be purified to remove the *Taq* polymerase and then ligated with the pGEM®-T vectors. The ligation mix can then be transformed in bacteria and blue/white screening can be used to identify positive transformants. This vector has a size of 3015 kb and it carries the T7 and SP6 RNA polymerase promoters flanking a region of multiple cloning sites within the coding region of beta-galactosidase. It also contains the origin of replication of filamentous phage f1 and Amp resistance gene. This vector was used as an intermediate vector for cloning different PCR products used in the different clonings.

III.1.3 pDRIVE Cloning Vector

This vector (Qiagen GmbH, Hilden, Germany) was used for the same purpose as pGEM®-T Easy vector. It has a size of 3.85 kb and contains the origin of replication of the filamentous phage f1. It carries T7 and SP6 viral promoters flanking the multiple cloning site and Amp and Kan resistance genes as selectable markers. Blue/white colony screening is possible due to it contains the coding region of beta-galactosidase.

III.1.4 pGY1 vector

The pGY1 plasmid was used for the final clonings in plasmid DNA for transient expression experiments. It contains the 35S Cauliflower mosaic virus (CaMV) promoter and terminator and the signal peptide (SP) of tobacco chitinase (Neuhaus et al. 1991) and has a size of 3.5 kb. For replication in bacteria, the origin of replication pBR322 is present and it contains the gene for Amp resistance.

III.1.5 pDHA vector

The recombinant plasmid pDHA (Tabe et al. 1995) carried some sequences used for the p24-tail anchor (TA) construct cloning. It has a size of 3.6 kb and contains a flexible linker sequence, the rabbit C-terminal region of cytochrome *b*5 (cyt *b*5) which functions

as tail anchor and contains an hydrophobic transmembrane domain and the op3 synthetic sequence corresponding to the N-terminal part of the bovine opsin which is recognized by a monoclonal antibody (Maggio et al. 2007).

III.2 Binary vectors

III.2.1 pGREEN0229 and pGREEN0179

For stable expression of HIV-1 p24 in plants, the binary expression vectors pGREEN0229 and pGREEN0179 (Hellens et al. 2000) were utilized. Both plasmids have the same pGREEN backbone containing the *NptI* gene that conveys resistance to Kan (*E.coli* selection), two origins of replication in *E.coli* (pSa and ColEI) and in *Agrobacterium* (RK4). They differ in the T-DNA region they contain so pGREEN0229 plasmid has a size of 4454 kb and contain the *bar* gene for plant resistance to ammonium sulfonide under the control of the nopaline synthetase (nos) promoter. The pGREEN0179 plasmid has a size of 5144 kb and contains the hygromycin resistance gene under the control of the 35S CaMV promoter. In both T-DNAs, the multiple cloning site is located within the coding region of the beta-galactosidase gene which allows white/blue selection with X-gal and IPTG. These plasmids can only replicate in *Agrobacterium* if the strain contains the disarmed Ti plasmid pSOUP which provides replication factors *in trans*. Modified *Agrobacterium* strains provide the Vir functions necessary for excising the T-DNA fragment from the binary plasmid and promote its transfer and integration into the plant genome. The replication in *E.coli* results at high copy number with the ColEI origin of replication and at low copy number with the pSa origin of replication. The DNA cassettes were cloned into the *Eco*RI site of the T-DNA region of pGREEN0229 and pGREEN0179 vectors and the *E.coli* colonies were selected on LB agar plates supplemented with Kan and X-gal and IPTG to allow white/blue screening. The white colonies were used for miniprep plasmid isolation of the recombinant binary plasmid. The recombinant pGREEN plasmids were used for *Agrobacterium* transformation.

III.3 DNA Sequences

A derived version of BH10 strain HIV-1 p24 (Accession number M15654.1) was used in this work. The mRFP sequence was given by Dr. Pierre Pfeiffer (Strasbourg,

France) (accession number AF506027). The N-terminal domain of γ-zein, γ-zein signal sequence (accession number X56118) and $(GGGGS)_3$ were kindly donated by Dr. A. Vitale (Milan, Italy).

Tobacco chitinase signal sequence and the pGY1 plasmid were given by Prof. Dr. Jean-Marc Neuhaus (Neuchâtel, Switzerland). Tail anchored region (accession number RABB5CYTA) and opsin region were kindly donated by Dr. Emanuela Pedrazzini (Milan, Italy). Transmembrane domains were given by Dr. Nadine Paris (Rouen, France).

The p24 antigen produced in *E.coli* (NIBSC) were kindly given by Dr. Patricia Obregón and Dr. Julian Ma (London UK) and p24 produced in *Pichia pastoris* was from NIBSC.

III.4 Standard PCR amplification

Polymerase chain reaction (PCR) is a technique in which the target DNA is exponentially amplified to several orders of magnitude. The double-stranded DNA is first denatured in order to have the strands separated and in the presence of Taq polymerase, deoxyribonucleotide triphosphate (dNTPs) and two oligonucleotide primers flanking the DNA template sequence, the *Taq* DNA polymerase is directed by the two primers to synthesize the complementary strands. As the reaction comprises multiple cycles, the process allow the amplification of the DNA of interest (Saiki et al. 1988). *Taq* polymerase is a thermostable DNA polymerase isolated from the termophilic bacterium *Thermus aquaticus* (Chien et al. 1976). It possesses a 5′→3′ polymerase activity and a double-strand specific 5′→3′ exonuclease activity (Tindall and Kunkel 1988). The fragments amplified by *Taq* polymerase possess an A nucleotide chain in each 3' extreme (Clark 1988) which facilitates the direct cloning of the PCR products into pGEM®-T Easy (Promega AG, Wallisellen, Switzerland) and pDRIVE (Qiagen) vectors as they possess a T nucleotide chain at their prominent 3' extreme.

About 100 ng of template DNA were used as template in order to introduce the necessary mutations by PCR. The reactions were carried out in 2X PCR master mix (PCR Master mix kit, Promega) supplied in a proprietary reaction buffer pH 8.5 containing 50 units/ml *Taq* polymerase, 400 µM dATP, 400 µM dGTP, 400 µM dCTP, 400 µM dTTP and 3 mM $MgCl_2$; and 1-2 µl of the two primers (10 µM). The conditions of amplification are described in each construction. In all cases the annealing

temperature was adapted to the Tm value of the primers used calculated with the Primer Designer v. 2.0 software program (Scientific and Educational software).
To analyze the integrity of the amplified product, the PCR product was resolved on a 0.8-1% agarose gel and purified by gel extraction or clean-up.
In the cases where a restriction site was removed, the corresponding codifying amino acid was kept in the sequence and the new codon was optimized for *N.tabacum*.

III.5 Site-specific mutagenesis by overlap extension

Before starting any cloning, the RFP and HIV-1 p24 sequences were mutated. In the RFP sequence, there were 3 restriction sites that were not convenient for our clonings and in the HIV-1 p24 there were 4. It was not difficult to change the extreme sequences with primers to remove the restriction sites that were not useful and introduce the new restriction sites optimal for the different clonings. The difficult point was to introduce point mutations in the middle of the sequence to remove an existing restriction site. For these point mutations the amino acid sequence was not changed and the codons were optimized for *N.tabacum*. The initial site-specific mutagenesis by overlap extension of HIV-1 p24 was done to clone the different constructs of the first strategy (p24 fused to the N-terminus of RFP containing transmembrane domains (TMD)). The newly synthesized p24 was called p24N1 and it was used as PCR template for the second strategy constructs (p24 fused to the C-terminus of RFP), the zein-p24 construct (p24 fused to the C-terminus of γ-zein) and the p24-TA construct (p24 fused to the N-terminus of cytochrome b5 tail anchor) cloning.
To describe this methodology, at least four primers were needed to introduce site-specific mutations by overlap extension in the original sequences of RFP and p24. The number of primers depends on how many mutations are required to introduce. If both ends of the sequence may be changed, too, then all the primers will introduce some mutations (Higuchi et al. 1988; Ho S.N. 1989; Ho 1989; Brandizzi et al. 2002).

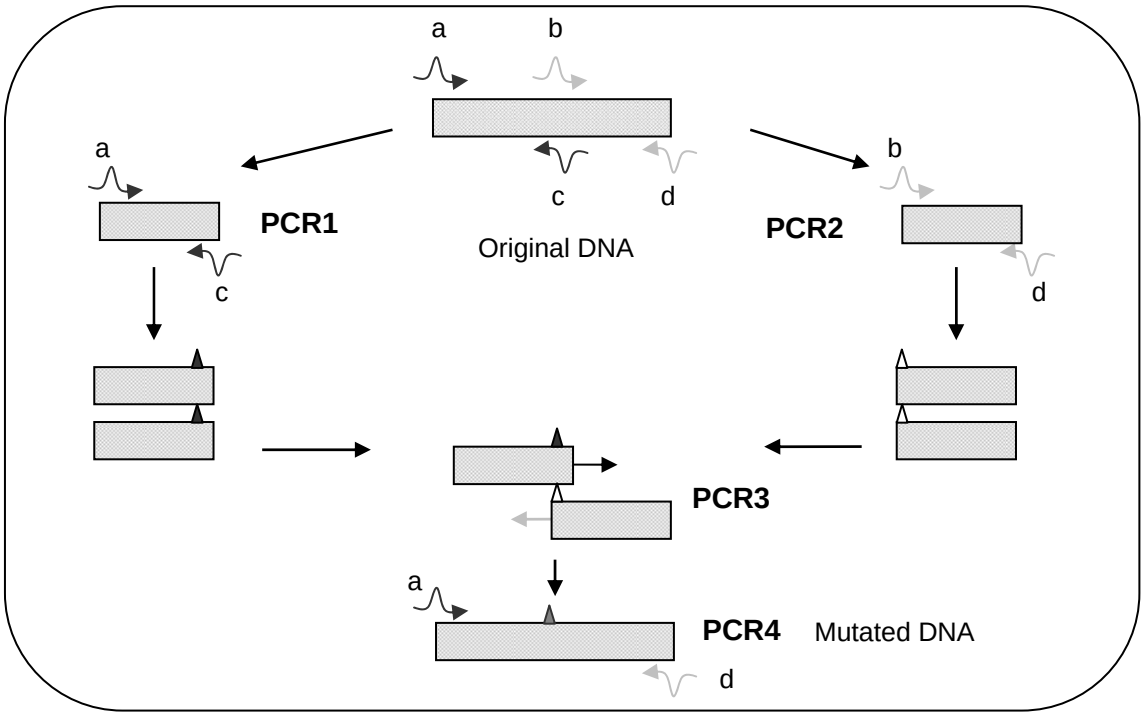


Figure 1. Diagram of site-specific mutagenesis by overlap extension. In the figure there are represented the different PCR reactions with the corresponding primers.

PCR 1		**PCR2**	
Template DNA	~100 ng	Template DNA	~100 ng
2X Master PCR Mix	12.5 µl	2X Master PCR Mix	12.5 µl
Primer (5' brown) (10 µM)	1 µl	Primer (5' blue) (10 µM)	1 µl
Primer (3' brown) (10 µM)	1 µl	Primer (3' blue) (10 µM)	1 µl
H_2O	to 25 µl	H_2O	to 25 µl

PCR3		**PCR4**	
Template DNA	~100 ng	Add to PCR3:	
(2 gel extracted bands from PCRs 1 and 2)		Primer (5' brown) (10 µM) 2	µl
2X Master PCR Mix	12.5 µl	Primer (3' blue) (10 µM)	2 µl
H_2O	to 21 µl		

Table II. Composition of the PCR reaction mixtures.

To give an overview of the protocol and as shown in figure 1, the first pair of primers, 'a' and 'd', contained the mutation(s) to be introduced into the wild type template DNA in both ends whereas the second pair of primers, 'b' and 'c', contained the mutation(s)

to be introduced in the middle of the wild type DNA template. This second pair of primers is complementary to each other. In separate amplification reactions, 1 and 2, two fragments from the original template DNA were amplified. PCR 1 reaction utilized the 'a' and 'c' primers and PCR 2 reaction utilized the 'b' and 'd' primers. The nucleic acids were amplified using the denaturation, annealing and elongation temperatures and times listed in each construct in the figure. The products of the two PCRs were analysed on a 1% agarose gel and purified. In PCR 3 reaction, the two purified products from PCR 1 and 2 were used as template. They were denatured and annealed at the region of overlap and extended as shown by the lines to form full-length double-stranded mutant DNA (reaction without primers). In PCR 4 reaction, primers 'a' and 'd' were introduced to the PCR 3 tube and the reaction was continued for 30 cycles more to amplify the full-length mutant DNA resulting in several copies of the full-length mutant DNA. The final PCR product was analysed on a 1% agarose gel and the band was purified. The PCR conditions and primers used for site-specific mutagenesis by overlap extension of RFP and p24 are described later in this chapter.

III.6 Colony PCR

To perform PCR using a bacterial colony as template, the colonies of interest were picked with a sterile yellow tip and incubated in the volume of water necessary to perform the PCR reaction in separate tubes. The water in contact with each colony was aspirated up and down with the tip for 8 to 10 times. The empty tips were placed into a 15 ml tubes containing 3 ml of LB broth supplemented with Amp or Kan and were incubated at 37ºC O/N in agitation.

The PCR kit mixture containing PCR buffer and *Taq* polymerase (Promega) and the specific primers were mixed with the water in which the colonies were mixed and the reaction was performed with the specific conditions required. The PCR products were resolved in an agarose gel and in case one colony was positive, the corresponding cultured tube was processed for miniprep plasmid DNA extraction.

III.7 Cracking gel

Some putative recombinant *E.coli* and *Agrobacterium* colonies (white) were picked with a sterile tip and were first streaked onto a new a fresh LB-Kan plate (or the

corresponding selection) to have a back-up plate of those colonies. The tips containing the different colonies were then placed into microcentrifuge tubes containing 25 µl of sterile water and resuspended. 25 µl of cracking gel solution (100m M NaOH, 10 mM EDTA pH 8.0, 1% SDS, 10% glycerol, 0,125% (w/v) bromophenol blue) were added and incubated at RT for 5 min. 20 µl of each sample were loaded on a 0.8% agarose gel and, as negative control, a blue colony screened by 0.160 mM IPTG and 32 µg/ml X-gal (plasmid without the insert) was loaded.

For each sample 3 main bands can be seen. The upper band corresponds to genomic DNA, the middle one is the plasmid and the lower one that is more diffuse correspond to RNA. If the colony contains the recombinant plasmid, the presence of the insert determines a shift in the electrophoretic mobility towards the plasmid without insert. The plate containing the different freshly transferred colonies was then incubated at the specific temperature and processed for further experiments.

IV Recombinant DNA techniques

IV.1 Agarose gel

The edges of a clean, dry plastic tray were sealed with tape or plastic devices to form a mould. Depending on the length of the DNA fragments to be separated, the gels usually contained 0.8% or 1% of agarose. Therefore, 0.8 g or 1 g of low-melting point agarose was dissolved in 100 ml of 1X TAE (40 mM Tris, 20 mM acetic acid, 1 mM EDTA) buffer in the microwave. When the agarose was dissolved, the solution was poured into a 50 ml Falcon tube to let it cool. 0.5 µg/ml of ethidium bromide were added to the solution and mixed. The mixture was poured into the mould making sure no air bubbles were formed. The comb was placed and the gel was left to solidify (30-45 min depending on the thickness). When the gel was ready, the comb and the tape or plastic devices were removed and the gel was mounted in the electrophoresis tank containing 1XTAE buffer. The DNA samples were mixed with 6X DNA loading dye solution (10 mM Tris-HCl pH 7.6, 0.15% orange G, 0.03% xylene cyanol FF, 60% glycerol and 60 mM EDTA) and loaded into the wells. The 1kb molecular DNA marker (New England Biolabs, Frankfurt am Main, Germany; and Fermentas, St.Leon-Rot, Germany) was also loaded. The DNA migrates toward the positive anode (red lead) and the gel was run at 100-120 mA. When the DNA samples or dyes migrated a sufficient

distance through the gel, the electric current was turned off a picture was taken. The DNA bands were cut if needed for gel extraction.

Range of Separation in Cells Containing Different Amounts of Standard Low-EEO Agarose

Agarose Concentration in Gel (% (w/v))	Range of Separation of Linear DNA Molecules (kb)
0.3	**5-60**
0.6	**1-20**
0.7	**0.8-10**
0.9	**0.5-7**
1.2	**0.4-6**
1.5	**0.2-3**
2.0	**0.1-2**

IV.2 Gel Extraction

The purification of PCR products and products of digestions previously loaded on an agarose gel were performed by using the PCR gel extraction/clean-up kit (Promega). The DNA was visualized on the gel and photographed using a low-wavelength UV lamp. After excising the DNA fragment of interest in a minimal volume of agarose using a clean razor blade, the slice was transferred into a microcentrifuge tube and weighed. The extraction was performed following the manufacturer's indications and the eluted DNA was stored at 4ºC or at –20ºC.

IV.3 PCR Clean-up

The target DNA was amplified using specific amplification conditions. An equal volume of Membrane Binding Solution (provided with the PCR gel extraction/clean up kit from Promega) to the PCR reaction was added and the next steps performed following the manufacturer's instructions.

IV.4 DNA Digestion

The different restriction enzymes were purchased from Fermentas, New England Biolabs and Promega. The reactions were carried out with the buffer supplied with each enzyme in accordance to the supplier's recommendations for temperature of the reaction. 1 µg of DNA was mixed with 1-2 units of the appropriate enzyme and

incubated from 40 min to 1 h at the enzyme's optimal temperature. For double digestions, the enzymes were incubated in the most suitable buffer for their combination. If combination of the two enzymes was not possible, the DNA was digested with one enzyme and purified with the PCR purification kit (Promega) prior to the digestion with the second enzyme. At the end of all the reactions, the mixture was purified using the PCR clean up kit (Promega).

IV.5 Dephosphorylation of the digested vector

The dephosphorylation was carried out in a final volume of 20 µl containing 1X dephosphorylation buffer (0.1 M Tris-HCl pH 7.5 at 37°C, 0.1 M $MgCl_2$) provided with the kit (Fermentas), the previously digested vector (about 1 µg) and 10 units of calf intestinal alkaline phosphatase (CIAP). The mixture was incubated at 37ºC for 30 minutes and cleaned up following the PCR clean up protocol.

IV.6 Ligation

To perform the ligations using pGEM-T Easy (Promega) and pDRIVE (Qiagen) vectors, the T4 DNA ligase used was the one provided with each kit. For the ligations using pGY1, pDHA and pGREEN vectors, T4 DNA ligase (Promega) was used. The insert:vector molar ratio was adjusted to 3:1 in 1X T4 ligase buffer (30 mM Tris-HCl pH 7.8, 10 mM $MgCl_2$, 10 mM DTT and 1 mM ATP) and 0.1 U of T4 DNA ligase in a final volume of 10 - 20 µl. The reaction was incubated at RT for 1 h and the mixture was directly used to transform *E.coli*.

IV.7 Small scale plasmid isolation from bacteria (Miniprep)

With the help of a sterile yellow tip, an isolated colony was transferred into 3 ml of LB broth medium containing Amp (100 µg/ml). After incubating the culture at 37ºC O/N with agitation, 1.5 ml of the culture was transferred into a microcentrifuge tube and the cells were pelleted at 16,813 x *g* for 2 min. The culture left was kept at 4ºC. The supernatant was discarded completely and the pellet was resuspended in 100 µl of ice-cold Solution I (1 M Tris-HCl pH 8.0, 1 M glucose and 0.5 M EDTA pH 8.0) and incubated at RT for 3 min. 200 µl of freshly prepared solution II (10 N NaOH and 20%

SDS) were added and the suspension was mixed by inversion until obtaining a transparent homogenous suspension. After incubating on ice for 5 min, 150 µl of ice-cold solution III (5 M KOAc and glacial acetic acid) were added and mixed with the vortex at maximum speed. The tube was incubated for 5 min on ice and centrifuged at 16,813 x *g* for 5 min at RT to pellet the cell debris and chromosomal DNA. The supernatant was transferred into clean microcentrifuge tube and the same volume of chloroform was added (400 µl) and mixed vigorously. After centrifugation for 5 min at 16,813 x *g,* the aqueous phase (superior) was recovered and subjected to a second chloroform treatment if the phase looked not very clean. Finally, the aqueous phase was transferred into a new tube and 900 µl of ice-cold 100% ethanol were added in order to precipitate the DNA. The mixture was mixed by inversion and incubated at room temperature for 10 min. The pellet was washed with 500 µl of 70% ethanol at RT. The tube was centrifuged again at 16,813 x *g* for 5 min at room temperature and the supernatant was discarded. After drying the pellet for 5 min at 65ºC, it was resuspended in 40 µl of TE buffer (10 mM Tris-HCl pH 8 and 1 M EDTA pH 8.0) containing RNAse (20 µg/ml). The DNA yield was tested on an agarose gel. This protocol was modified from the alkaline lysis protocol described by Sambrook et al. (Sambrook 1989).

IV.8 Medium scale plasmid isolation from bacteria (Midiprep)

As described in the protocol above, with the help of a sterile yellow tip, an isolated colony was transferred, but in this case, into 100 ml of LB broth medium containing Amp (100 µg/ml). The culture was incubated at 37ºC O/N with agitation. It is also possible to start from a glycerol stock (inoculating with a tip) or from a saturated culture (inoculating 50 µl).

The culture was centrifuged at 2372 x *g* for 5 min at 4ºC and the pellet was resuspended in 1 ml of ice-cold Solution I. The solution was resuspended with a pipette several times until a homogenous suspension was visualized. 2 ml of fresh Solution II were added and the suspension was mixed by inversion and incubated on ice for 10 min. 1 ml of cold Solution III was added and mixed vigorously by inversion. After incubating for 10 min on ice, the cells were pelleted (10 min at 2372 x *g* at 4ºC) and the supernatant was transferred into a 50 ml conical tube and filtered through cheesecloth. The DNA was precipitated by adding 4 ml of RT isopropanol. The solution was mixed by inversion and incubated at RT for 10 min. The pellet was resuspended in 350 µl of TE buffer and

350 µl of 5 M LiCl and centrifuged again. The supernatant was transferred into a clean 1.5 ml microcentrifuge tube and 700 µl of isopropanol were added. The solution was mixed and incubated for 10 min at RT. The pellet was washed with 1.5 ml of 70% ethanol, dried very well and resuspended in 80 µl of TE buffer containing RNAse (20 µg/ml). The tube was incubated at 37ºC for 2 h to dry the pellet.

IV.9 Large scale plasmid isolation from bacteria (Maxiprep)

A pre-culture was prepared starting from a single colony and incubated in 2 ml of LB supplemented with 50 µg/ml Amp. The ideal is to start from a freshly streak out plate prepared the day before. After 3-4 h incubation at 37ºC in agitation, the pre-culture was slightly turbid and was poured into 500 ml pre-warmed LB without Amp. This culture was grown for 24 h at 37ºC in agitation. The culture was centrifuged in 500 ml bottles in a Sorvall SLA-3000 rotor (DuPont) at 2268 x *g* at 4ºC for 60 min. The supernatant was discarded and the pellet was completely resuspended in 8 ml of ice cold TE 50/1 buffer (50 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0). The cell suspension was transferred into clean SS34 tubes pre-chilled on ice. 2.5 ml of freshly prepared lysozyme solution (10 mg/ml) was added and the tubes were mixed by up side down and back for 10 times. They were incubated 5 min on ice and 2 ml of 0.5 M EDTA pH 8.0 were added and mixed as before. The suspension was incubated another 5 min on ice and it was more viscous at this point. In a separate tube, 100 µl of ribonuclease A solution (20 mg/ml) were mixed with 150 µl of 10% Triton X-100 and 800 µl of TE 50/1 for each sample. 1 ml of this mixture was added to the cell suspension and incubated on ice for 30 min. The tubes were then centrifuged in a Sorvall SS34 rotor (DuPont) at 39,000 x *g* for 60 min at 4ºC. The supernatants were transferred into 50 ml Falcon tubes and the same amount (~11 ml) of equilibrated phenol (phenol pH 8.0 with 0.1% 8-hydroxyquinoline) was added and mixed vigorously. The tubes were centrifuged at 2772 x *g* for 20 min without break. The water phase was recovered and transferred into clean 50 ml Falcon tubes. The same amount (~11 ml) of chloroform was added and mixed and the tubes were centrifuged at 2772 x *g* for 15 min without break. The water phase was recovered (~9 ml) and transferred into clean 30 ml Corex tubes and the samples were filled up to 10 ml with TE 50/1 buffer. Then, 1 ml of 5 M $NaClO_4$ was added and the tubes were sealed with parafilm in order to turn them and mix the

solutions. 8 ml of isopropanol were added and again the solutions were mixed as before. The tubes were centrifuged in a Sorvall HB6 rotor (DuPont) at 19,000 x *g* for 15 min at 4ºC. The supernatant was discarded and the pellet was dried in a vacuum pump and resuspended in 500 µl of TE buffer.

IV.10 Genomic DNA extraction from transgenic plants and PCR

200 mg of leaf tissue from transgenic plants were homogenized with a pestle in 200 µl of extraction buffer (500 mM NaCl, 100 mM Tris pH 7.5, 50 mM EDTA). 20 µl of 20% SDS were added, vortexed and incubated at 65ºC for 10 min. 235 µl of phenol/chloroform/isoamylalcohol (25:24:1) were added and vortexed. The mixture was centrifuged at 12,000 x *g* for 4 min at 4ºC. 200 µl were added to a new eppendorf and the DNA was precipitated before performing the PCR.
The PCR was performed using primers #7 (for) and #14 (rev) and the program was 4 min of initial denaturation at 94ºC followed by 40 cycles of 1 min at 94ºC, 1 min at 63ºC and 1 min and 15 sec at 72ºC. A final extension of 2 min at 72ºC was performed.

IV.11 DNA precipitation

About 1 µg of DNA was diluted to a final volume of 50 µl with water. 5 µl of 3 M Na acetate (pH 5.7) were added (1/10 volume) and 125 µl 100% ethanol (2.5 volumes). The mixture was incubated on ice for 30 min and centrifuged at 12,000 x *g* for 30 min at 4°C. The supernatant was discarded and the pellet was washed with 70% ethanol and centrifuged as before for 10 min. The supernatant was discarded and the pellet was air dried.

IV.12 Determination of the DNA concentration

To determine the concentration of DNA, the absorbance was measured by spectrophotometry at 260 nm wavelength and the DNA concentration calculated. An absorbance of 1 at 260 nm corresponds to 50 µg/ml of DNA (Sambrook 2001).

IV.13 *Agrobacterium* miniprep preparation

Putative recombinant *Agrobacterium* colonies were picked and inoculated in 5 ml of LB supplemented with Kan (50 µg/ml), Tet (5 µg/ml), Gen (50 µg/ml) and Rif (20 µg/ml) and incubated at 28°C for 36-48 hours. 1.5 ml of the culture were centrifuged at 12,000 x *g* for 10 min. The bacterial pellets were resuspended in 100 µl of suspension solution (50 mM glucose, 25 mM Tris-HCl, 10 mM EDTA (pH 8.0)). 20 µl of 20 mg/ml lysozyme solution were added and after mixing well the tubes were incubated at 37°C for 15 minutes. To lyse the cells, 200 µl of lysis solution (0.2 M NaOH, 1% SDS) were added and the tubes were inverted several times to mix well the solution. After adding 50 µl of equilibrated phenol to the 2 volumes of cell lysis solution, the tubes were mixed with the vortex. To neutralize the reaction, 200 µl of neutralization solution (3 M Na acetate, pH 5.2) were added and mixed completely by repeated inversion of the tubes. The mixtures were centrifuged at 12,000 x *g* for 5 min and the aqueous phases were transferred into microcentrifuge tubes. 2.5 volumes of 95% ethanol were added and the tubes were incubated on ice for 10 min. To spin down the DNA, the tubes were centrifuged at 12,000 x *g* for 5 min and the pellet was air dried and resuspended in TE buffer.

IV.14 Sequencing

DNA sequencing was carried out using the dideoxy nucleotide chain-termination method (Sanger et al. 1977) with Big Dye terminator v3.1 labelling kit (Applied Biosystems, Darmstadt, Germany). The DNA samples were purified plasmids or PCR products ligated to pGEM®-T Easy or pDRIVE vectors. The primers used were the universal M13 forward (for) and reverse (rev) or specific forward and reverse primers. The DNA samples pellet was resuspended in sterile water and the ABI PRISM Big Dye Terminator v3.1 Cycle Sequencing Kit was used to set up the sequencing reactions. Sequences were analysed on ABI 3100 16 capillars automatic sequencer.

V Construction of the expression constructs

V.1 Cloning of RFP-TMD constructions

V.1.1 Site-specific mutagenesis of RFP by overlap extension

All the primers used in the different PCRs in the different clonings are listed in page 179. In order to mutate the original RFP gene, four primers were designed with Primer Designer software. With primer #1 (for) a *Nhe*I site was introduced and a *Hind*III site was removed. With primers #2 (for) and #3 (rev) a *Pst*I site was removed and with primer #4 (rev) a *Pst*I site was introduced. Primers #2 and #3 are complementary.

For the mutagenesis, the monomeric RFP (mRFP) gene was used as template DNA and the different reactions were carried out with the PCR Master Mix kit (Promega) which includes *Taq* polymerase. The conditions of the different PCRs are described in table III.

	Template	**Name of the PCR product**	**Initial denaturing step**	**Denaturing step**	**Annealing step**	**Elongation step**	**# Cycles**	**Final Extension**	**5' primer**	**3' primer**
PCR 1	Original RFP	Fragment 1	No	45'' at 94ºC	45'' at 60ºC	1' at 72ºC	30	5' at 72ºC 1 cycle	#1	#3
PCR 2	Original RFP	Fragment 2	No	45'' at 94ºC	45'' at 63ºC	1' at 72ºC	30	5' at 72ºC 1 cycle	#2	#4
PCR 3	Fragment 1 and 2	-	1' at 94ºC	45'' at 94ºC	45'' at 62ºC	1' at 72ºC	10	No	No	No
	-	Final RFP	No	45'' at 94ºC	45'' at 62ºC	1' at 72ºC	30	5' at 72ºC	#1	#4

Table III. PCR conditions of the different reactions to mutagenize the mRFP gene.

In the PCR 1 the N-terminal fragment of the original RFP was mutated with primers #1 and #3 resulting in a fragment of 340 bp. In the PCR 2 the C-terminal part was mutated with primers #2 and #4 and the band size was 363 bp. In the PCR 3, the template DNA was a mixture of the 2 PCR products (fragment 1 and fragment 2) and the reaction was run for 10 cycles. After adding the primers the final product was the new RFP (~700 bp).

The new mutated RFP was directly ligated into the pGEM-T Easy® plasmid (Promega). XL-1 *E. coli* competent cells were transformed with the pGEM-T Easy® plasmid

ligation product and ten colonies were chosen for PCR screening using primers #1 and #4. The PCR program was 2 min at 94ºC followed by 25 cycles of 45 sec at 94ºC, 45 sec at 62ºC and 1 min at 72ºC including a final extension of 5 min at 72ºC. One colony was positive and small miniprep was prepared. The pGEM®::RFP plasmid was analysed by restriction analysis cutting with *Nhe*I-*Sal*I and *Eco*RI-*Pst*I to ascertain the plasmid contained the RFP gene and also sequenced to assure there were no mutations in the new RFP sequence.

V.1.2 Construction of RFP-TMD23

The new RFP was cut with *Nhe*I and *Sal*I restriction enzymes and a triple ligation was performed. The backbone pGY1 vector, which contains the 35S CaMV promoter and terminator, was cut with *Nhe*I and *Hind*III restriction enzymes and the original TMD23 was cut with *Sal*I and *Hind*III. These cut fragments together with the previously cut RFP fragment were ligated and transformed in XL-1 *E.coli* competent cells. Only 3 colonies grew and were screened by colony PCR (table IV, pC+23NheI conditions). Only one colony was positive, and midiprep was prepared and the plasmid was called pC+23NheI.

The putative positive pC+23NheI was confirmed by restriction analysis, digesting with *Bam*HI/*Pst*I, *Nhe*I/*Pst*I and *Bam*HI/*Sal*I, however, after further clonings, the sequence contained an *Nhe*I site that was not convenient for the final clonings and the construct was recloned (see V.1.6).

Colonies	5' primer	3'primer	Initial denaturing step (1 cycle)	Denaturing step	Annealing step	Elongation step	Cycles (#)	Final elongation
pGEM®::RFP	#1	#4	94ºC 2 min	94ºC 45 sec	63ºC 45 sec	72ºC 1 min	25	72ºC 1 min
pC+23NheI	#1	#4	94ºC 4 min	94ºC 45 sec	62ºC 45 sec	72ºC 1 min	40	72ºC 5 min
pC+20	#5	#6	94ºC 4 min	94ºC 30 sec	56ºC 30 sec	72ºC 40 sec	25	72ºC 1 min
pG017, pG020 pG023, pG026	#6	#27	94ºC 4 min	94ºC 1 min	56ºC 45 sec	72ºC 2 min	25	72ºC 2 min

Table IV. Conditions for colony PCR

V.1.3 Construction of RFP-TMD17

The RFP-TMD17 construct was obtained by cloning the original TMD17 into the *Sal*I and *Pst*I sites of the pC+23NheI construct. The ligation mixture was transformed into XL-1 *E.coli* and 2 colonies were picked for miniprep. The two DNA minipreps were digested with *Nhe*I/*Pst*I to ascertain the RFP gene was present. Both colonies contained the gene and one of them was chosen for midiprep. Sequencing analysis of the construct demonstrated the sequence of the TMD corresponded to the TMD17. The new construct with the TMD17 was called pC+17NheI (after RFP-TMD17, see V.1.6).

V.1.4 Construction of RFP-TMD26

To clone the TMD26, the original TMD26 was cut with *Sal*I and *Pst*I and cloned into the *Sal*I and *Pst*I sites of the pC+23NheI construct. After the ligation and transformation, many white colonies grew. Two of them were chosen for miniprep and the DNA was digested with *Nhe*I/*Pst*I to ascertain the RFP gene was present. One of them was chosen for midiprep and the construct was sequenced to assure the TMD corresponded to the TMD26 which indeed did. The new construct with the TMD26 was called pC+26NheI (after RFP-TMD26, see V.1.6).

V.1.5 Construction of RFP-TMD20

In the pC+20NheI construct, the original TMD20 was cloned into the *Sal*I and *Hind*III sites of pC+23NheI construct. The ligation mixture was transformed into XL-1 *E.coli* and positive colonies were confirmed by colony PCR (see conditions in table IV, pC+20). All the colonies screened were positive and one was picked for miniprep. The DNA was sequenced to verify the TMD20 sequence and it was called pC+20NheI, after, RFP-TMD20 (see V.1.6).

V.1.6 Recloning of the RFP-TMDs constructs

The different TMDs contained a *Nhe*I site that was not convenient for the p24RFP-TMDs cloning strategy where the p24 might be cloned into the *Nhe*I and *Nco*I sites of the different RFP-TMDs constructs. The different TMD had a part of the sequence that

was common in all of them and a part that was different (depending on the TMD, some contained extra VLI motifs).

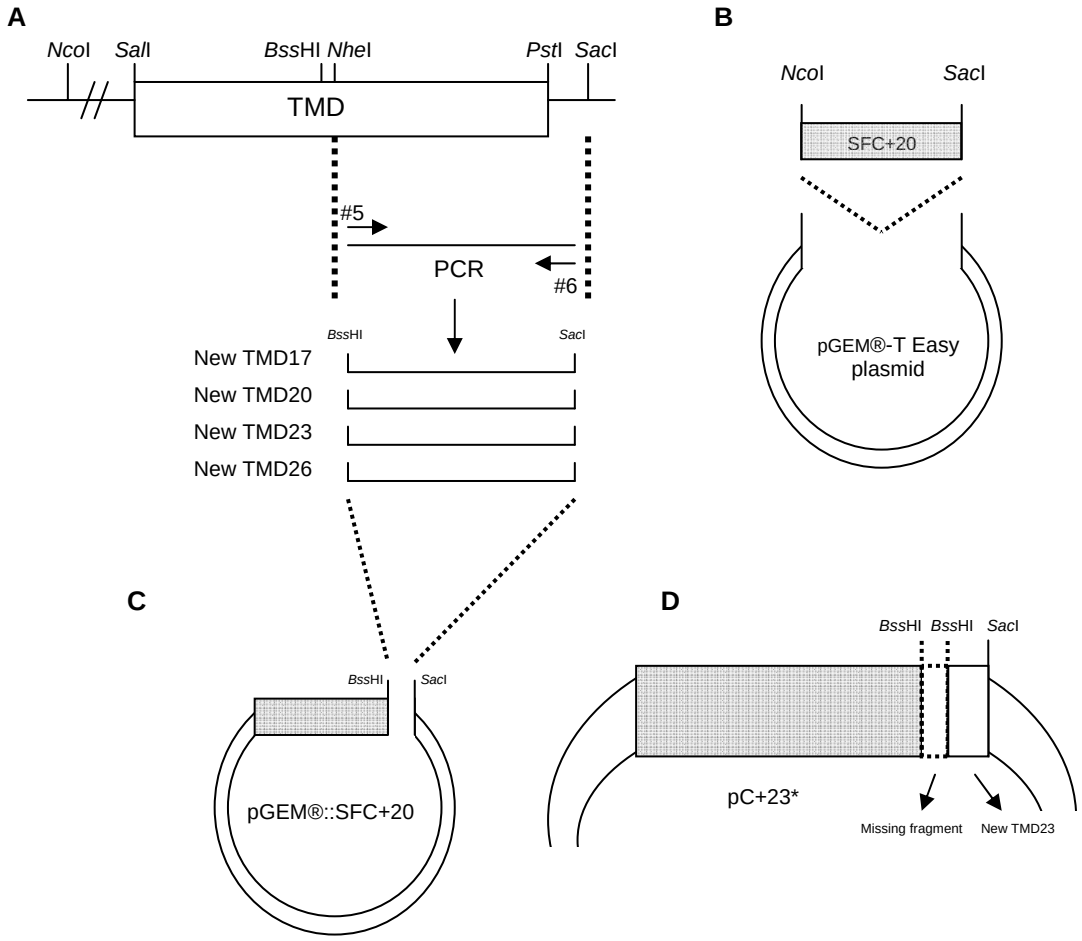


Figure 2. Schematic representation of the recloning of RFP-TMD constructs. A) PCR was performed on the different TMDs in the *Bss*HI and *Sac*I sites using primers #5 and #6. B) Representation of the pGEM®::SFC+20 construct in which a short fragment of pC+20NheI was cloned into the *Nco*I and *Sac*I sites of pGEM®-T Easy vector. C) Diagram of the cloning of the new TMDs obtained by PCR cloned into the *Bss*HI and *Sac*I sites of pGEM®::SFC+20 construct. D) Diagram of the final cloning pC+23. During the cloning a 92 bp fragment between two *Bss*HI sites was removed and with further cloning it was replaced in the RFP-TMD23 construct only.

The *Nhe*I site was in the fragment that was different and the cloning strategy consisted in changing the sequence between the *Bss*HII and the *Sac*I sites (Figure 2, panel A). The *Nhe*I site was very close to the *Bss*HII site and with the designed primer #5 (for), the *Nhe*I site was removed and a *Bss*HII site was introduced in the original TMDs for the cloning. At 5' of the *Sac*I site, there was the differential sequence of all the TMDs that was kept using primer #6 (rev) in the PCR reaction. The aim was to

change the four TMDs and therefore four different PCR reactions were performed to change partially the TMDs. For these reactions, *Taq* polymerase was used, the primers mentioned above and as template the four different constructs prepared with RFP containing the original TMDs mentioned above (pC+17NheI, pC+20NheI, pC+23NheI and pC+26NheI). The samples were denaturated for 4 min at 94ºC followed by 35 cycles of amplification which included 15 seconds of denaturation at 94ºC, 15 seconds of hybridisation at 56ºC and 15 seconds of elongation at 72ºC. A final extension at 72ºC for 1 min was the final step. The four PCR products were digested with *Bss*HII/*Sac*I restriction enzymes.

Previously to ligate the new partial TMDs in the final vectors, an intermediate cloning was performed using pGEM® vector (Figure 2, panel B). The pC+20NheI construct was cut with *Nco*I and *Sac*I restriction enzymes and the short fragment of 1012 bp was cloned into the *Nco*I and *Sac*I sites of pGEM® vector. This construct was called pGEM®::SFC+20. The ligation mixture was transformed in XL-1 *E.coli* cells and after blue/white screening with X-gal and IPTG, 10 white colonies were picked for miniprep plasmid isolation. The new construct was confirmed by restriction analysis by digesting with *Nco*I/*Sac*I and *Bss*HII/*Sac*I.

The different new partial TMDs PCR products previously cut with *Bss*HII/*Sac*I were cloned into these sites of the pGEM®::SFC+20 construct (Figure 2, panel C). The ligation mixture was transformed in XL-1 *E.coli* cells and four colonies of the pGEM®::SFnewTMD17, pGEM®::SFnewTMD23 and pGEM®::SFnewTMD26, and one colony of the pGEM®::SFnewTMD20 were selected for miniprep. The different minipreps were confirmed by restriction analysis digesting with *Nhe*I/*Sac*I. The new different recombinant vectors were cut with *Nco*I and *Sac*I in order to replace the new complete TMDs into the *Nco*I and *Sac*I sites of the previously cut pC+20NheI. All the TMDs were cloned into the same backbone plasmid (pC+20NheI) as the difference of the TMDs remains in the new sequence that was cloned. Miniprep plasmid isolation was performed and the constructs were confirmed by restriction analysis digesting with *Nhe*I. The constructs were called pC+17*, pC+20*, pC+23* and pC+26* and they were prepared for midiprep plasmid isolation.

All the constructs seemed to be properly cloned but some restriction analysis with *Nco*I and *Sac*I revealed that during the intermediate cloning, when the different partial TMDs were cloned into the pGEM::SFC+20 in the *Bss*HII and *Sac*I sites, *Bss*HII cut twice and not once and a fragment of 92 bp was lost during the cloning (Figure 2, panel D). To

reintroduce this fragment, the pC+23NheI was cut with *Bss*HII and the 92 bp fragment were cloned into the *Bss*HII site of all of the pC+17*, pC+20*, pC+23* and pC+26*. The sequence was confirmed by DNA sequencing and the only successful recombinant construct was the RFP-TMD23. The pC+17NheI, the pC+20NheI and the pC+26NheI previously cloned constructs did not need to be changed as for the cloning of p24RFP-TMD strategy the RFP-TMD23 could be used and from the new construct replace the other TMDs. Therefore, and due to the difficulty of these clonings, these plasmids were called RFP-TMD17, RFP-TMD20 and RFP-TMD26 and maxipreps were prepared.

V.1.7 Cloning of RFP-TMDs into pGREEN0229

The RFP-TMDs constructs were cut with *Eco*RI and the cassettes were cloned into the *Eco*RI site of pGREEN0229 binary vector previously dephosphorilated. After transformation in *E.coli* with the conditions for binary vector transformation, the colonies were screened by blue/white. Different white colonies were screened by colony PCR (conditions described in table IV). After miniprep plasmid isolation, the constructs were confirmed by restriction analysis with *Eco*RI and the new clones were called pG017 (containing the cassette of the RFP-TMD17), pG020, pG023 and pG026.

V.2 Cloning of p24 fused to the N-terminus of RFP

V.2.1 Mutagenesis of p24 (p24N1) and construction of pP24RFP-TMD23

Several changes were introduced in the original HIV-1 p24 and eight primers were designed to perform these mutations using site-specific mutagenesis by overlap extension.

With primer #7 (for) an *Xho*I restriction site was removed and a *Nhe*I and a *Sal*I sites were introduced. With primers #8 (for) and #9 (rev) a *Sph*I and a *Pst*I sites were removed. A *Hin*dIII site was removed with primers #8 (for) and #11 (rev). With primer #12 (rev) a thrombin cleavage site (Th) and an *Nco*I site were introduced whereas an *EcoR*I site was removed. After performing all these PCRs, a mutation was introduced by the *Taq* polymerase and it was removed with primers #13 (for) and #14 (rev).

	5' primer	3' primer	Initial denaturing step	Denaturing step	Annealing step	Elongation step	# Cycles	Final Extension
PCR1	#7	#9	No	45'' at 94ºC	45'' at 60ºC	1' at 72ºC	30	5' at 72ºC 1 cycle
PCR2	#8	#11	No	45'' at 94ºC	45'' at 55ºC	1' at 72ºC	30	5' at 72ºC 1 cycle
PCR3	#10	#12	No	45'' at 94ºC	45'' at 58ºC	45'' at 72ºC	30	5' at 72ºC 1 cycle
PCR fragments 2-3	No	No	1' at 94ºC	45'' at 94ºC	45'' at 57ºC	1' at 72ºC	10	No
	#8	#12	No	45'' at 94ºC	45'' at 57ºC	1' at 72ºC	30	5' at 72ºC
Final PCR fragments 2-3 + 1 = p24N1	No	No	1' at 94ºC	45'' at 94ºC	45'' at 61ºC	1' at 72ºC	10	No
	#7	#12	No	45'' at 94ºC	45'' at 61ºC	1' at 72ºC	25	2' at 72ºC

Table V. Site-specific mutagenesis by overlap extension of p24N1. Conditions of the different PCR reactions to mutagenize the p24 gene.

The HIV-1 p24 was mutated using the PCR Master Mix kit (Promega) and the new p24 was called p24N1. The conditions of the different PCRs are described below.

The new mutated p24, p24N1, was cloned directly into the *Nhe*I/*Nco*I sites of the pRFP-TMD23 and XL-1 *E.coli* were transformed. Ten colonies were screened by colony PCR using primers #7 and #12. The PCR program was 4 min at 94ºC followed by 35 cycles of 45 sec at 94ºC, 45 sec at 61ºC and 1 min at 72ºC including a final extension of 2 min at 72ºC. The PCR products showed that four colonies were positive. One was chosen for miniprep and the construct was digested with *Nhe*I/*Nco*I and *Sal*I/*Hin*dIII to confirm the cloning. The new construct was called pN123*.

	Initial denat.	Denaturing step	Annealing step	Elongation step	# Cycles	Final Extension	5' primer	3' primer
PCR1	2' at 94ºC	30'' at 94ºC	30'' at 63ºC	45'' at 72ºC	35	1' at 72ºC 1 cycle	#7	#14
PCR2	2' at 94ºC	30'' at 94ºC	30'' at 60ºC	45'' at 72ºC	35	1' at 72ºC 1 cycle	#13	#12
PCR3=**newp24N1**	2' at 94ºC	45'' at 94ºC	45'' at 61ºC	1' at 72ºC	10	No	No	No
	No	45'' at 94ºC	45'' at 61ºC	1' at 72ºC	25	2' at 72ºC 1 cycle	#7	#12

Table VI. Second site-specific mutagenesis by overlap extension of p24N1

The p24N1 sequence was confirmed by sequencing and the sequence had 4 point mutations, not related with any restriction site, that were kept after several PCR. Three

out of the four mutations did not change the encoded amino acid and it was assumed they were substitutions done for proper previous cloning. However, there was one point mutation that changed the amino acid and had to be removed using site-specific mutagenesis by overlap extension. Two new primers were designed, #13 and #14, and also the primers #7 and #12 were used. The conditions of the different PCRs are listed in table VI.

The new p24N1 was digested with *Nhe*I and *Nco*I, cloned into the previously cut pN123* construct and XL-1 *E.coli* were transformed. Three colonies were screened by colony PCR using primers #7 and #12. The PCR conditions were the same used for PCR screening of N123*. Three colonies were positive and one was chosen for miniprep which was digested with *Sal*I/*Hin*dIII and the construct was sequenced. The new sequence determined the point mutation was successfully removed and the construct was called pP24RFP-TMD23.

V.2.2 Construction of pP24RFP-TMD17, pP24RFP-TMD20 and pP24RFP-TMD26

The p24RFP-TMD23 construct was used to obtain the other transmembrane constructs by replacing the transmembrane domains in the *Sal*I and *Hin*dIII sites. The pN123 construct was digested with *Sal*I/*Hin*dIII and the RFP-TMD17, RFP-TMD20 and RFP-TMD26 as well. After ligation and transformation into XL-1 *E.coli*, five colonies from each construct were screened by colony PCR using primers #7 (for) and #6 (rev). The PCR program was 4 min at 94ºC followed by 25 cycles of 45 sec at 94ºC, 45 sec at 60ºC and 1 min at 72ºC including a final extension of 2 min at 72ºC. The PCR products demonstrated all the colonies were positive. One positive colony of each construct was chosen for miniprep and the DNA was digested with *Eco*RI/*Bsp*HI to confirm the clonings. The new constructs were called pP24RFP-TMD17, pP24RFP-TMD20 and pP24RFP-TMD26 depending on the TMD they contain which were also confirmed by sequencing.

V.2.3 Cloning of the p24RFP-TMDs cassettes into pGREEN0229

The different p24RFP-TMDs constructs were digested with *Eco*RI and *Bsp*HI and the *Eco*RI-*Eco*RI bands corresponding to the cassettes were cut for gel extraction. They were ligated into the *Eco*RI site of pGREEN0229 vector, previously dephosphorilated

and transformed into XL-1 *E.coli*. Four colonies of each construct were screened by colony PCR using primers #27 (for) and #6 (rev). The PCR program was 4 min at 94ºC followed by 25 cycles of 1 min at 94ºC, 45 sec at 56ºC and 2 min at 72ºC including a final extension of 2 min at 72ºC. One positive colony of each construct was chosen for miniprep and the constructs were digested with *Eco*RI to confirm the clonings. The new constructs were called pG117, pG120, pG123 and pG126.

V.2.4 Cloning of the p24RFP-TMDs cassettes into pGREEN0179

The same cassettes from p24RFP-TMD17, p24RFPTMD20 and p24RFP-TMD23 were cloned into the dephosphorilated *Eco*RI site of pGREEN0179. The ligation was transformed into MC1061 *E.coli* cells for the new p24RFP17G0179 and into DH5α *E.coli* cells for p24RFP20G0179 and p24RFP23G0179. The colonies p24RFP17G0179 were screened by colony PCR using primers #7 (for) and #12 (rev). The program was 4 min of initial denaturation at 94ºC followed by 30 cycles of 45 sec at 94ºC, 45 sec at 61ºC and 1 min at 72ºC. A final extension of 2 min at 72ºC was also performed. Only 2 colonies of were positive and one was chosen for miniprep and re-confirmed by restriction analysis. The new construct was called p24RFP17G0179.
The transformations of p24RFP20G0179 and p24RFP23G0179 were screened by blue/white selection using X-gal and IPTG. White colonies were confirmed by cracking gel analysis. One positive colony of each construct was picked for miniprep and re-confirmed by restriction analysis. The new constructs were called p24RFP20G10279 and p24RFP23G0179.

V.3 Cloning of pRFPp24-TMD constructions

V.3.1 Mutagenesis of p24 (p24N2)

Two primers were necessary to mutate the p24 gene to prepare these constructs. With primer #15 (for) an *Xho*I and a thrombin cleavage sites were introduced whereas with the primer #16 (rev) a *Sal*I site was introduced. The template DNA for the standard PCR was the p24N1 fragment and the new p24 for this strategy was called p24N2. The PCR program was 2 min at 94ºC followed by 30 cycles of 45 sec at 94ºC, 45 sec at 66ºC and 1 min at 72ºC. A final extension of 2 min was performed. The new p24N2 was cloned into pDRIVE and the ligation mixture was incubated at 4ºC for 30 min. XL-1

*E.*coli cells were transformed and eight white colonies were screened by colony PCR using primers #15 and #6. The PCR program was 4 min at 94ºC followed by 35 cycles of 30 sec at 94ºC, 30 sec at 62ºC and 1 min at 72ºC including a final extension of 2 min at 72ºC. Two colonies were positive and one of them was chosen for miniprep. The DNA was digested with *Xho*I/*Sal*I and also sequenced. The positive sequence was then processed to prepare the different constructs.

V.3.2 Construction of pRFPp24-TMD26

The pDRIVE:p24N2 construct was digested with *Xho*I and *Sal*I and the lower band around 700 bp corresponding to p24N2 was kept for ligation. The RFP-TMD26 construct was digested with *Sal*I and dephosphorilated. The fragments were ligated and the mixture was transformed in *E.*coli XL-1. The p24N2 contains an *Xho*I and a *Sal*I sites and both are compatible with the *Sal*I site of the vector. Therefore, the insert could have been inserted in both orientations and the correct orientation was identified by colony PCR using primers #2 and #16. Ten colonies were screened and the PCR program was 4 min at 94ºC followed by 25 cycles of 45 sec at 94ºC, 45 sec at 60ºC and 1 min 10 sec at 72ºC with a final extension of 2 min at 72ºC. Only one colony was positive and it was chosen for miniprep and digested with *Hin*dIII/*Sal*I and with *Bam*HI and *Sal*I. The construct was called pRFPp24-TMD26.

V.3.3 Construction of pRFPp24-TMD17, pRFPp24-TMD20 and pRFPp24-TMD23

The TMD of the RFPp24-TMD26 was substituted by the different TMD to prepare the RFPp24-TMD17, RFPp24-TMD20 and RFPp24-TMD23 constructs. The RFP-TMD17, RFP-TMD20 and RFP-TMD23 constructs were digested with *Sal*I and *Hin*dIII and the TMD17, 20 and 23, respectively, were released from the vectors showing a shorter fragment in a 1% agarose gel. These fragments were cut and purified from the gel. The RFPp24-TMD26 was also digested with *Sal*I and *Hin*dIII and the fragments were ligated and transformed into *E.coli* XL-1. Six colonies of each construct were screened by colony PCR using primers #15 and #6. The PCR program was 4 min at 94ºC followed by 25 cycles of 45 sec at 94ºC, 45 sec at 62ºC and 1 min 10 sec at 72ºC with a final extension of 2 min at 72ºC. All the colonies were positive and one of each

construct was chosen for miniprep and the DNA was sequenced. The constructs were called pRFPp24-TMD17, pRFPp24-TMD20 and pRFPp24-TMD23.

V.3.4 Cloning of the RFPp24-TMDs constructs into pGREEN0229

The RFPp24-TMDs constructs were cut with *Eco*RI and *Bsp*HI and the bands corresponding to the *Eco*RI-*Eco*RI cassettes containing the genes of interest were cloned into the *Eco*RI site of pGREEN0229 previously dephosphorilated. The ligation mixture was transformed into XL-1 *E.coli* following the protocol for binary vectors. One white colony of pG217, pG220 and pG223 constructs was picked for miniprep and the DNA was digested with *Eco*RI to assure the correct cloning of the cassettes. For the pG226, colony PCR was performed using primers #27 (for) and #6 (rev). The PCR program was 4 min at 94ºC followed by 25 cycles of 1 min at 94ºC, 45 sec at 56ºC and 2.5 min at 72ºC with a final extension of 2 min at 72ºC. One colony was positive and chosen for miniprep. The DNA was digested with *Eco*RI. The new clones were called pG217, pG220, pG223 and pG226. All these steps are represented in the figure below.

V.4 Cloning of p24 fused to the C-terminus of γ-zein domain

For this strategy six primers were designed, two to mutate the γ-zein sequence, two to mutate p24 and another two to mutate the $(GGGGS)_3$ linker sequence. In order to introduce γ-zein, the pDHA-zeolin plasmid (Mainieri et al. 2004) was kindly provided by Dr. A.Vitale.

V.4.1 Mutagenesis of zeinM1

The new mutated γ-zein sequence, called zeinM1, contains a *Bam*HI site and the 5'-UTR sequence of tobacco chitinase signal sequence introduced by the forward primer #17 at the 5' of zein signal sequence; at the same time a *Bam*HI site was removed. With the reverse primer #18 an *Nco*I site was introduced and a *Pst*I site was removed in the γ-zein region. The original pDHA::zeolin sequence was digested with *Bam*HI and *Hin*dIII to release the zein coding sequence. This DNA was used as template for PCR using primers #17 and #18. The mixture was denatured for 1 min at 94ºC, then the PCR reaction continued for 45 seconds at 94ºC, 45 seconds at 66ºC and 1 min at 72ºC for 30

cycles. A final elongation step at 72ºC was performed for 5 min. The PCR product was ligated into pGEM® vector and transformed into XL-1 *E.coli*. Five white colonies were screened by colony PCR using primers #17 and #18. The PCR program was 4 min at 94ºC, and 30 cycles of 45 seconds at 94ºC, 45 seconds at 66ºC and 1 min at 72ºC with a final elongation of 5 min at 72ºC. One colony was picked for miniprep and the DNA sequence was confirmed by sequencing.

V.4.2 Mutagenesis of $(GGGGS)_3$ linker

The linker $(GGGGS)_3$ was introduced to allow independent protein folding of p24 and γ-zein. The encoding sequence was amplified from the original pDHA::zeolin plasmid previously digested with *Bam*HI and *Pst*I. With the forward primer #19 an *Nco*I site was introduced and a *Bam*HI site was removed and with the reverse primer #20 a *Sal*I site was introduced. The $(GGGGS)_3$ sequence was amplified with an initial denaturation step at 94ºC for 1 min followed by 35 cycles of 5 seconds at 94ºC, 5 seconds at 59ºC and 5 seconds at 72ºC. A final extension at 72ºC for 20 seconds was the final step. Prior to the reaction, the primers were boiled for 5 min at 95ºC to avoid secondary structures. The PCR product was digested with *Nco*I and *Sal*I and was directly cloned into p24-zein construct in which the correct sequence was confirmed by DNA sequencing.

V.4.3 Mutagenesis of p24M1 and cloning of pzein-p24

To mutate the p24 gene, the forward primer was the #21, almost the same as primer #15 used in the RFPp24-TMD strategy but with a base less to keep the correct open reading frame. The reverse primer #22 introduced a stop codon and a *Pst*I. The DNA template was p24N1 and the new p24 sequence was called p24M1. The PCR program was 1 min of initial denaturation at 94ºC and 30 cycles of 45 seconds at 94ºC, 45 seconds at 58ºC and 1 min at 72ºC. The final elongation at 72ºC was for 2 min.

To perform the final cloning of this construct, the zeinM1 sequence was first cloned into the *Bam*HI and *Nco*I sites of pGY1 vector. The ligation mixture was transformed into XL-1 *E.coli* cells. The different pGY1::zeinM1 colonies were screened by colony PCR (II.3.6) and the PCR conditions are described in table VII. One positive colony was chosen for miniprep and the DNA was digested with *Bam*HI/*Nco*I.

Colonies	5' primer	3'primer	Initial denaturing step (1 cycle)	Denaturing step	Annealing step	Elongation step	Cycles (#)	Final elongation
pGY1::zeinM1	#17	#18	94ºC 4 min	94ºC 45 sec	66ºC 45 sec	72ºC 1 min	35	72ºC 1 min
pGY1::zeinM1::$(GGGGS)_3$	#17	#20	94ºC 4 min	94ºC 45 sec	62ºC 45 sec	72ºC 1 min	30	72ºC 2 min
pzein-p24	#19	#22	94ºC 4 min	94ºC 45 sec	57ºC 45 sec	72ºC 1 min	25	72ºC 2 min

Table VII. Colony PCR conditions used in zein-p24 cloning.

The $(GGGGS)_3$ sequence, which was cloned into the *Nco*I and *Sal*I sites of the pP24-zein construct described below, was excised from this construct cutting with *Nco*I and *Pst*I enzymes and the fragment was ligated into the same sites of the pGY1::zeinM1 construct. The ligation mixture was transformed into XL-1 *E.coli* cells and six colonies pGY1::zeinM1::$(GGGGS)_3$ were screened by colony PCR (II.3.6). The conditions are described in table VII. One positive colony was chosen for miniprep and it was called pGY1::zeinM1::$(GGGGS)_3$.

The p24M1 sequence was cut with *Xho*I and *Pst*I restriction enzymes and cloned into the *Xho*I and *Pst*I sites of the previously cut pGY1::zeinM1::$(GGGGS)_3$ miniprep. The ligation was transformed into XL-1 *E.coli* cells and five colonies were screened by colony PCR. The PCR program is described in table VII. One colony was chosen for miniprep and the DNA was digested with *Bam*HI/*Sal*I, *Bam*HI/*Pst*I and *Eco*RI/*Bsp*HI. The resulting final construct was called pzein-p24 and was re-confirmed by sequencing which showed that the different sequences were correct and any unwanted mutation was introduced by PCR.

V.4.4 Cloning of zein-p24 cassette into pGREEN0229

The pzein-p24 was digested with *Eco*RI and *Bsp*HI and the band corresponding to the *Eco*RI-*Eco*RI cassette of interest was cloned into the *Eco*RI site of the digested and dephosphorilated pGREEN0229 vector. One white colony was chosen for miniprep and

the DNA was digested with *Eco*RI to confirm the correct cloning. The new construct was called pGzeinp24.

V.4.5 Cloning of zein-p24 cassette into pGREEN0179

The same cassette was cloned into the dephosphorilated *Eco*RI site of pGREEN0179. The ligation was transformed into DH5α *E.coli* cells and after blue/white selection with X-gal and IPTG, some white colonies were confirmed by cracking gel analysis. One positive colony was picked for miniprep and re-confirmed by restriction analysis. The new construct was called pzeinp24G0179.

V.5 Cloning of p24-zein: p24 fused to the N-terminal region of γ-zein

The γ-zein region had to be mutated and two primers were designed. The zein signal sequence was removed from the original γ-zein region and a *Sal*I site was introduced with the forward primer #23. With the reverse primer #24 a stop codon and a *Pst*I site were introduced. The PCR program was 1 min of initial denaturation at 94ºC and 45 seconds at 94ºC, 45 seconds at 63ºC and 1 min at 72ºC for 40 cycles. 5 min of final extension were performed at 72ºC. The new γ-zein was called zeinM2 and was cloned into pGEM® vector and the ligation mixture was transformed into XL-1 *E.coli* cells. Five white colonies were screened by colony PCR using primers #23 and #24 and the same PCR program described above for 30 cycles. One positive colony was chosen for miniprep and midiprep and the DNA was sequenced to verify the zeinM2 sequence.

The backbone plasmid was the p24RFP-TMD26 construct as the signal sequence and the p24 gene were the same. As described in the pzein-p24 construct, the linker $(GGGGS)_3$ was first cloned in this construct before cloning into the zein-p24 construct. A first $(GGGGS)_3$ PCR not sequenced, that after resulted to not have the correct DNA sequence, was ligated into the *Nco*I and *Sal*I sites of p24RFP-TMD26. Also, the p24 of the p24RFP-TMD26 construct contained a single mutation and at the end it was changed. The ligation mixture was transformed into XL-1 *E.coli* and ten colonies were screened by colony PCR. The conditions are described in table VIII (see pP24M2Δ::G3Δ). Eight colonies were positive and one was chosen for miniprep. This clone was called pP24M2Δ::G3Δ and it was digested with *Bam*HI/*Sal*I and *Sal*I/*Pst*I. The bands corresponded and the zeinM2 sequence was cloned into the *Sal*I and *Pst*I

sites of pP24M2Δ::G3Δ construct and transformed into XL-1 *E.coli* cells. Five colonies were confirmed by colony PCR and the PCR program is described in table VIII (pP24M2Δ::G3Δ::zeinM2 conditions). The positive clone was chosen for miniprep plasmid isolation. At this point the p24M2 (in fact is p24N1 from p24RFP-TMD26) contained a single mutation and needed to be changed. The new p24N1 gene was cloned into the *Nhe*I and *Nco*I sites of the pP24M2Δ::G3Δ::zeinM2 construct by removing the old p24N1 gene and replacing it for the new one. Five colonies were screened by colony PCR using the conditions for pP24M2::G3Δ:zeinM2 described in table VIII. All the colonies were positive and one was chosen for miniprep plasmid isolation. To confirm the cloning, the miniprep was digested with *Nhe*I/*Nco*I and *Nco*I/*Sal*I. The positive cloned was called pP24M2::G3Δ:zeinM2 as the $(GGGGS)_3$ linker was still not properly cloned.

The linker $(GGGGS)_3$ PCR was repeated as described before. The new linker was cloned into the *Nco*I and *Sal*I sites of the pP24M2::G3Δ:zeinM2 construct. The ligation mixture was transformed into XL-1 *E.coli* cells and four colonies were screened by colony PCR. The conditions are described in table VIII (pM2 conditions). Two colonies were positive and one was chosen for miniprep. The DNA was digested with *Nco*I/*Pst*I and *Sal*I/*Pst*I to assure the correct cloning and the final construct was called pP24-zein. It was sequenced to confirm the $(GGGGS)_3$ had no mutations.

Colonies	5' primer	3'primer	Initial denaturing step (1 cycle)	Denaturing step	Annealing step	Elongation step	Cycles (#)	Final elongation
pP24M2Δ::G3Δ	#7	#20	94°C 4 min	94°C 45 sec	61°C 45 sec	72°C 45 sec	30	72°C 1 min
pP24M2Δ::G3Δ::zeinM2	#7	#24	94ºC 4 min	94ºC 1 min	61ºC 1 min	72ºC 1 min 10 sec	30	72ºC 2 min
pP24M2::G3Δ:zeinM2	#7	#12	94ºC 4 min	94ºC 45 sec	61ºC 45 sec	72ºC 1 min	25	72ºC 2 min
pM2 = pP24-zein	#19	#20	94ºC 4 min	94ºC 45 sec	59ºC 45 sec	72ºC 45 sec	30	72ºC 1 min

Table VIII. Colony PCR for pP24-zein construction.

V.5.1 Cloning of p24-zein cassette into pGREEN0229

The p24-zein construct was digested with *Eco*RI and *Bsp*HI to release the *Eco*RI-*Eco*RI cassette. This fragment was cloned into the *Eco*RI site of pGREEN0229 previously dephosphorilated. XL-1 *E.coli* cells were transformed and one white colony was chosen for miniprep. The DNA was digested with *Eco*RI to confirm the correct cloning and the new construct was called pGp24zein.

V.5.2 Cloning of p24-zein cassette into pGREEN0179

The same cassette was cloned into the *Eco*RI site of pGREEN0179 previously dephosphorilated. The ligation was transformed into DH5α *E.coli* cells and after blue/white selection with X-gal and IPTG, the white colonies were confirmed by cracking gel analysis. One positive colony was picked for miniprep and re-confirmed by restriction analysis. The new construct was called p24zeinG0179.

V.6 Cloning of p24-TA: p24 fused to the C-terminus of cyt b5 tail anchor

To mutate the p24 for this construct, the new p24N1 was mutated with primers #25 (for) and #26 (rev). With the forward primer a *Bam*HI site was introduced and with the reverse primer a *Sal*I site was introduced. The conditions of the PCR reaction were 2 min of initial denaturation at 94ºC followed by 30 cycles of 45 seconds of denaturation at 94ºC, 45 seconds of hybridisation at 57ºC and 1 min of elongation at 72ºC. A final extension at 72ºC for 2 min was the last step. The new p24 was called p24linker and was cloned into the *Bam*HI and *Sal*I sites of pDHA::zeolin and transformed in XL-1 *E.coli* cells.

Thirteen colonies were screened by colony PCR using primers #25 and #26. The PCR program was one cycle at 94°C for 4 min, 35 cycles of 45 sec at 94°C, 45 sec at 57°C and 1 min at 72°C. A final extension of 2 min at 72°C was also performed. One colony was chosen for miniprep and it was digested with *Bam*HI/*Nde*I, *Bam*HI/*Sal*I and *Nco*I/*Sal*I. The p24linker was sequenced to assure the correct mutagenesis. The pDHA::cytb5 plasmid contained the linker GAGSGG, the C-terminal end of cyt b5 and the N-terminal sequence of bovine opsin containing an N-glycosylation consensus site (op-3). The final clone in pDHA vector was called p24-TA.

V.6.1 Cloning of p24-TA into pGREEN0229

The p24-TA construct was digested with *Eco*RI and *Bsp*HI and the cassette of interest flanked by *Eco*RI sites was cloned into the *Eco*RI site of pGREEN0229 vector, previously dephosphorilated. The ligation mixture was transformed into XL-1 *E.coli* cells and one white colony was chosen for miniprep. The construct was digested with *Eco*RI to confirm the correct cloning and it was called pGp24-TA.

V.6.2 Cloning of p24-TA into pGREEN0229

The same cassette was cloned into the dephosphorilated *Eco*RI site of pGREEN0179. The ligation was transformed into DH5α *E.coli* cells and after blue/white selection with X-gal and IPTG, some white colonies were screened by cracking gel analysis. One positive colony was picked for miniprep and re-confirmed by restriction analysis. The new construct was called p24TAG0179.

V.7 Construction of GFP-TMD constructs

The enhanced GFP (eGFP) gene was excised from pCK(X/S)LTEV-EGFP vector (Ritzenthaler et al. 2002) by digesting with *Nco*I and *Xba*I. Two primers were designed in order to mutate both extremes and introduce a *Nhe*I site at 5' (primer #28, forward) and a *Sal*I site at 3' (primer #29, reverse). The PCR was performed at 94ºC for 2 min, followed by 30 cycles at 94ºC for 30 seconds, at 60ºC for 45 seconds and at 72ºC for 1 min. A final extension at 72ºC for 2 min was also performed. Three tubes were marked T0, T20 and T40 and the PCR product was digested with *Nhe*I and *Sal*I. Starting from the time before incubation, every 20 min 1µl of the mixture was transferred to one of these tubes (T20 sample), each containing 5µl DNA loading dye. After the last sample at 40 min (T40) was collected the three samples were ran on a 1% agarose gel. This was done to estimate the time needed for complete digestion with the enzymes. The pRFP-TMD23 construct was also digested and dephosphorilated to prevent re-ligation and after the digestion of the PCR product (new GFP) the two fragments were ligated. The final construct was called pS023 and the GFP sequence was confirmed by sequencing.
The construct containing GFP and the TMD20 was called pS020. The pC+20 construct (with RFP) was cut and dephosphorilated. The large fragment was ligated with GFP previously cut with *Nco*I and *Sal*I and *E.coli* were transformed as described previously.

Twenty colonies were picked for miniprep and screened by restriction analysis. The first digestion was with *Hin*dIII and *Cfr*10I to control if the exchange of RFP for GFP had worked. If the constructs contained GFP, the digested plasmid would give rise to three bands whereas if the constructs contain RFP (re-ligation) the gel should show only two bands. The minipreps that showed a positive pattern were then cut with *Nco*I and *Sal*I to see if the restriction sites were restored.

To prepare the pS017 construct, the GFP was cloned into the pC+17 construct in which the RFP was removed previously by cutting with *Nco*I and *Sal*I. Ten colonies were screened by colony PCR using primers #28 and #29. The PCR was performed at 94ºC for 4 min, followed by 30 cycles at 94ºC for 45 seconds, at 60ºC for 45 seconds and at 72ºC for 1 min. A final extension at 72ºC for 2 min was also performed. Midiprep was prepared and it was digested *Nco*I and *Sal*I to confirm the proper cloning. The new construct was called pS017.

V.8 Construction of p24 fused at N-terminus of GFP and TMD

The pS023 construct was digested with *Nco*I and *Sal*I restriction enzymes and dephosporilated. The pP24RFP-TMD23 construct was also digested with *Sal*I and *Nco*I and dephosphorilated. A time digestion study was performed as described as above. The GFP was ligated into *Nco*I and *Sal*I sites. The ligation mixture was transformed into *E. coli* MC1061 and twenty colonies were picked and incubated for culture miniprep. The minipreps were digested with *Hind*III and *Cfr*10I to control if the exchange of RFP for GFP had worked. If the constructs contain GFP the digested plasmid will give rise to four bands whereas if the constructs contain RFP (re-ligation) the gel should show only three bands.

To prepare the pS120 construct, the pP24RFP-TMD20 construct was digested with *Nco*I and *Sal*I and a time digestion study was performed as described as above. After 60 min digestion another 1µl of *Sal*I and 1µl of *Nco*I were added and the reaction was continued for another 30 min. After dephosphorylation of the pN120 digestion, the large fragment was ligated to GFP cut with *Nco*I and *Sal*I and transformed in competent *E. coli* MC1061 as described previously. Twenty colonies were picked and incubated for culture miniprep. The minipreps were digested as before.

The pS117 construct was prepared as the pS017 but in the ligation reaction, the GFP was cloned into the previously cut pN117.

V.9 Construction of p24 fused at C-terminus of GFP and TMD

To prepare the pS223 construct, the pS023 construct was digested with *Sal*I and dephosphorilated. pDRIVE:p24N2 was also digested with *Sal*I and *Xho*I. In both cases, a time digestion study was performed as described as before. The lower band of the digested pDRIVE:p24N2 corresponding to p24 (~700bp) was ligated to the digested pS023. The ligation mixture was transformed in competent *E. coli* MC1061. Fifteen colonies were screened by colony PCR using primers #28 and #16. The PCR was performed at 94ºC for 4 min, followed by 35 cycles at 94ºC for 45 seconds, at 59ºC for 45 seconds and at 72ºC for 1 min 45 sec. A final extension at 72ºC for 2 min was also performed. Three positive colonies were chosen for culture miniprep and the minipreps were digested with *Sal*I/*Hin*dIII and *Nhe*I/*Sal*I. To clone pS217 and pS220, the pS223 construct was cut with *Sal*I and *Hin*dIII. The pC+17 and pC+20 were digested with *Sal*I and *Hin*dIII and the short fragments were ligated to the large fragment of pS223 digested with *Sal*I and *Hin*dIII. The ligation mixture was transformed into competent *E.coli* MC1061 and one colony of each new construct was picked for midiprep. The DNA was sequenced and it was confirmed that the pS217 construct contained the TMD17 and the pS220 the TMD20.

V.10 Construction of p24RFPcyt

In order to control that the original p24 was correctly localised in the cytosol, two constructs without transmembrane domain were prepared.

The first construct was the p24RFP in which the p24 is fused to the N-terminus of RFP. The pP24RFP-TMD23 plasmid was digested with *Nhe*I and *Sal*I restriction enzymes and the short fragment, corresponding to p24 gene, was used as DNA template for PCR. The reaction was carried with primers #30 (for) and #31 (rev) and the program was 4 min at 94ºC, followed by 35 cycles at 94ºC for 45 seconds, at 64ºC for 45 seconds and at 72ºC for 1 min 45 sec. Two minutes at 72ºC was the final extension step. The fragment obtained purified and digested with *BamH*I and *Pst*I. The cut fragment was ligated with the large fragment of the previously digested (with *BamH*I and *Pst*I) and dephosphorilated pP24RFP-TMD26 vector in a final volume of 10 µl. 5 µl of the ligation were transformed into competent *E.coli* MC1061. Ten colonies were screened by colony PCR using primers #30 and #31. The PCR was performed at 94ºC for 4 min,

followed by 35 cycles at 94ºC for 45 seconds, at 64ºC for 45 seconds and at 72ºC for 1 min 45 sec. A final extension at 72ºC for 2 min was also performed. One positive colony was chosen for culture midiprep and it was digested with *BamH*I/*Pst*I to ensure the p24 gene was excised. The new construct was also confirmed by sequencing.

V.11 Construction of RFPp24cyt

The second cytosolic construct was the RFPp24 in which the p24 was fused to the C-terminus of RFP. The PCR template was the short fragment, corresponding to p24 gene, obtained by digesting pRFPp24-TMD23 plasmid with *Nhe*I and *Sal*I. The PCR program was 2 min at 94ºC, followed by 30 cycles at 94ºC for 45 seconds, at 60ºC for 45 seconds and at 72ºC for 1 min 50 sec. Two minutes at 72ºC was the final extension step. The reaction was carried with primers #32 (for) and #33 (rev). The fragment obtained was purified and digested with *BamH*I and *Pst*I. The cut fragment was ligated with the large fragment of the previously digested (with *BamH*I and *Pst*I) and dephosphorilated pP24RFP-TMD26 vector in a final volume of 10 µl. 5 µl of the ligation were transformed into competent *E.coli* MC1061. Ten colonies were screened by colony PCR using primers #30 and #31. The PCR was performed at 94ºC for 4 min, followed by 35 cycles at 94ºC for 45 seconds, at 64ºC for 45 seconds and at 72ºC for 1 min 45 sec. A final extension at 72ºC for 2 min was also performed. One positive colony was chosen for culture midiprep and it was digested with *BamH*I/*Pst*I to ensure the p24 gene was excised. The new construct was also confirmed by sequencing.

V.12 Construction of p24RFPsec

A construct containing the p24 and the RFP genes and any TMD was prepared. The new cassette was amplified from the short fragment released from the digestion of pP24RFP-TMD23 with *Nhe*I and *Sal*I. In the PCR reaction primers #7 and #34 were used and the program was 2 min at 94ºC, followed by 30 cycles at 94ºC for 45 seconds, at 63ºC for 45 seconds and at 72ºC for 1 min 50 sec. A final extension at 72ºC for 5 min was also performed. The amplified fragment was digested with *Nhe*I and *Sph*I enzymes and the pC+23 as well. The digested PCR product was ligated into the large fragment of the digested pC+23 and transformed into *E.coli* DH5α. Twenty-five colonies were screened by colony PCR using primers #7 and #12 and the program was 4 min at 94ºC followed by 30 cycles of 45 sec at 94ºC, 45 sec at 61ºC and 1 min at 72ºC with a final

extension of 2 min at 72ºC. Only one colony was positive and miniprep was prepared and digested with *Nhe*I/*Nco*I and *Nhe*I/*Sal*I. The construct sequence was confirmed by DNA sequencing.

V.13 Construction of RFPp24sec

A construct containing the RFP and the p24 genes and any TMD was also prepared. The PCR reaction was prepared using the short fragment released from the digestion of pRFPp24-TMD23 with *Nhe*I and *Sal*I as DNA template. The primers #1 and #35 were used and the program was 2 min at 94ºC, followed by 30 cycles at 94ºC for 45 seconds, at 63ºC for 45 seconds and at 72ºC for 1 min 50 sec with a final extension at 72ºC of 5 min. The new amplified fragment was digested with *Nhe*I and *Sph*I enzymes and it was ligated into the large fragment of the digested pC+23. The ligation product was transformed into *E.coli* DH5α and ten colonies were screened by colony PCR. For this reaction, primers #7 and #12 were used and the program was 4 min at 94ºC followed by 30 cycles of 45 sec at 94ºC, 45 sec at 61ºC and 1 min at 72ºC with a final extension of 2 min at 72ºC. Only one colony was positive and miniprep was prepared and digested with *Eco*RI/*Nco*I and *Nhe*I/*Sal*I. The construct sequence was confirmed by DNA sequencing.

V.14 Construction of N220-GFP, N223-GFP and N226-GFP

A new GFP sequence was prepared using primers #36 and #37 in which a *Bgl*II and a *Sph*I restriction sites were added. The PCR program was 2 min at 94ºC followed by 35 cycles of 30 sec at 94ºC, 30 sec at 61ºC and 45 sec at 72ºC with a final extension of 2 min at 72ºC. The new GFP sequence, called GFPTM, was inserted into the *Bgl*II and *Sph*I sites of previously digested pRFPp24-TMD20, pRFPp24-TMD23 and pRFPp24-TMD26. Ten colonies of pN220GFPTM and 5 colonies of pN223GFPTM and pN226GFPTM were screened by colony PCR using primers #36 and #37. The PCR was performed with the same conditions as for GFPTM amplification except that the initial denaturation was for 4 min instead of 2 min. One colony of each positive construct was picked and incubated for culture midiprep.
All the constructs are listed in the next table.

Construct name	p24 gene	Other genes	LAMP1 TMD	Bac res	Plant res
pRFP-TMD17	No	mRFP	17	Amp	None
pGRFP-TMD17	No	mRFP	17	Kan	BASTA
pRFP-TMD20	No	mRFP	20	Amp	None
pGRFP-TMD20	No	mRFP	20	Kan	BASTA
pRFP-TMD23	No	mRFP	23	Amp	None
pGRFP-TMD23	No	mRFP	23	Kan	BASTA
pRFP-TMD26	No	mRFP	26	Amp	None
pGRFP-TMD26	No	mRFP	26	Kan	BASTA
P24RFP-TMD17	Yes	mRFP	17	Amp	None
pG117	Yes	mRFP	17	Kan	BASTA
P24RFP17G0179	Yes	mRFP	17	Kan	Hyg
P24RFP-TMD20	Yes	mRFP	20	Amp	None
pG120	Yes	mRFP	20	Kan	BASTA
P24RFP20G0179	Yes	mRFP	20	Kan	Hyg
P24RFP-TMD23	Yes	mRFP	23	Amp	None
pG123	Yes	mRFP	23	Kan	BASTA
P24RFP23G0179	Yes	mRFP	23	Kan	Hyg
P24RFP-TMD26	Yes	mRFP	26	Amp	None
pG126	Yes	mRFP	26	Kan	BASTA
RFPp24-TMD17	Yes	mRFP	17	Amp	None
pG217	Yes	mRFP	17	Kan	BASTA
RFPp24-TMD20	Yes	mRFP	20	Amp	None
pG220	Yes	mRFP	20	Kan	BASTA
RFPp24-TMD23	Yes	mRFP	23	Amp	None
pG223	Yes	mRFP	23	Kan	BASTA
RFPp24-TMD26	Yes	mRFP	26	Amp	None
pG226	Yes	mRFP	26	Kan	BASTA
Zein-p24	Yes	N-term γ-zein	None	Amp	None
pGM1	Yes	N-term γ-zein	None	Kan	BASTA
zeinp24G0179	Yes	N-term γ-zein	None	Kan	Hyg
P24-zein	Yes	N-term γ-zein	None	Amp	None
pGM2	Yes	N-term γ-zein	None	Kan	BASTA
P24zeinG0179	Yes	N-term γ-zein	None	Kan	Hyg
P24-TA	Yes	C-term cyt b5	None	Amp	None
pGM3	Yes	C-term cyt b5	None	Kan	BASTA
p24TAG0179	Yes	C-term cyt b5	None	Kan	Hyg
GFP-TMD17	No	eGFP	17	Amp	None
GFP-TMD20	No	eGFP	20	Amp	None
GFP-TMD23	No	eGFP	23	Amp	None
P24GFP-TMD17	Yes	eGFP	17	Amp	None
P24GFP-TMD20	Yes	eGFP	20	Amp	None
P24GFP-TMD23	Yes	eGFP	23	Amp	None
GFPp24-TMD17	Yes	eGFP	17	Amp	None
GFPp24-TMD20	Yes	eGFP	20	Amp	None
GFPp24-TMD23	Yes	eGFP	23	Amp	None
P24cyt	Yes	mRFP	None	Amp	None
p24RFPΔTMD	Yes	mRFP	None	Amp	None
RFPp24ΔTMD	Yes	mRFP	None	Amp	None
pN220GFP	Yes	mRFP and eGFP	20	Amp	None
pN223GFP	Yes	mRFP and eGFP	23	Amp	None
pN226GFP	Yes	mRFP and eGFP	26	Amp	None

VI Transformation of plants

VI.1 Growth and maintenance of *Nicotiana tabacum* and *Nicotiana benthamiana*

VI.1.1 Under sterile conditions

Nicotiana tabacum cv Petit Havana SR1 and derived stable transformed plants were cultivated in MS medium (4,30 g/l MS medium basal salt mixture (Duchefa Biochemie, Harlem, The Netherlands); 20 g/l saccharose; 0,2 g/l MES; pH 5,7 with KOH; 0,8 g/l Agar) in sterile Magenta boxes or jars in the growth chamber. The conditions were 24°C constant temperature, 70% relative humidity and 16 hours light / 8 hours dark regime.

VI.1.2 In soil

Wild type *Nicotiana benthamiana* and transgenic *Nicotiana tabacum* plants were grown in the green house in soil with 60-65% humidity and 16 h photoperiod at 24°C constant temperature. When the plants reached a height of 40-60 cm they were used for *Agrobacterium* infiltration.

VI.2 Transient expression

VI.2.1 In tobacco leaves by agroinfiltration

Nicotiana benthamiana plants were grown in the green house as described. The recombinant *Agrobacterium* containing the pGREEN plasmids with the different DNA cassettes were streaked out on LB agar plates supplemented with Rif (25 μg/ml) and Kan (25 μg/ml) and incubated at 28ºC for 2 days. One colony was picked and incubated in 5 ml of LB broth medium supplemented with the same antibiotics as above and incubated O/N at 28ºC in agitation. The day after the culture was saturated and another culture was started from the saturated one. The final volume of the new culture was 3 ml of LB broth medium supplemented with Rif and Kan, and the amount of *Agrobacterium* culture added was the amount necessary to have a final $O.D._{600nm}$ of the new culture of 0.3, so the dilution factor was calculated. This new culture was incubated at 28ºC for about 4 h or until the $O.D._{600nm}$ was of 0.5. Then, the *Agrobacterium* culture was centrifuged at 2772 x *g* for 5 min at RT. The supernatant was discarded and the pellet was washed with 1 ml of agroinfiltration solution (9.76 g/L MES, 0.552 g/L

Na_2HPO_4, 0.48 g/L NaH_2PO_4, 5 g/L glucose, 100 µM acetosyringone). The cells were centrifuged as before and washed again. The pellets were resuspended in the appropriate volume of agroinfiltration medium in order to have an $O.D._{600nm}$ of 0.5. The cells were incubated at RT for 30 min and were infiltrated with a syringe without needle. The tip of the syringe was pressed against the underside of a leaf while a counterpressure to the other side of the leaf was simultaneously done with a finger. The recombinant *Agrobacterium* solution was then injected into the airspaces inside the leaf through stomata and could be visually appreciated. The agroinfiltrated area was delimited with a marker pen in order to remember it as some days later the agroinfiltrated area cannot be distinguished from a non agroinfiltrated one. After 48-72 h the agroinfiltrated parts of the leaves were visualized with the CLSM using the filter for RFP.

VI.2.2 In tobacco protoplasts by DNA electroporation

Nicotiana tabacum cv Petit Havana plants were grown under sterile conditions as described previously. Leaves were cut removing the middle nerve and with a special device some little holes were made all over the abaxial surface. The leaf pieces were transferred into a Petri dish containing 7 ml of digestion mix (0.2% macerocyme R10, 0.4% cellulose R10 in TEX buffer: 0.4 M sucrose, 750 mg/l $CaCl_2 \cdot 2H_2O$, 500 mg/l MES, 250 mg/l NH_4NO_3; pH 5.7 with KOH) with the abaxial side in contact with the medium. Usually, one Petri dish (about 5 x 10^6 protoplasts) is enough for 2 electroporations. The plates were incubated O/N at 25ºC in the dark. Still under sterile conditions, the plates were gently shacked to release protoplasts from the cuticula. The mixture was filtered through a 100 µm sieve previously rinsed with electroporation buffer (0.4 M sucrose, 2.4 g/L HEPES, 6 g/L KCl, 600 mg/ml $CaCl_2$; pH 7.2 with KOH). This rinse is important in order to release further protoplasts from the tissue remnants and provide a first step in the adaptation to the new medium. The filtrated protoplasts were transferred into a 50 ml tube and centrifuged at 85 x *g* for 15 min at RT without break. The living protoplasts remain floating on the top of the solution whereas death protoplasts and cell debris form a pellet or stay in solution. The pellet was removed with a fine Pasteur pipette connected to a peristaltic pump. In order to avoid that many protoplasts stick to the pipette, a window is created to place the pipette to the bottom of the tube. Electroporation buffer was added up to 30 ml and the mixture was centrifuged at 85 x *g* for 10 min. The underlying solution was removed as described

above. Electroporation buffer was added up to 15 ml and the mixture was centrifuged as before. The protoplasts were resuspended in the appropriate volume in order to have about 2 - 2.5 x 10^6 protoplasts per millilitre. 500 µl of protoplasts (~1x10^6 protoplasts) were placed into disposable 1 ml plastic cuvettes with a wide opening blue tip. About 40 µg of DNA were mixed in a final volume of 100 µl of electroporation buffer and mixed with the protoplasts. The cells were incubated for 5 min and then electroporated with a home made electroporation device. The electroporation was performed at 1000 µF and 160 V with a Gene Pulser II (BioRad) and the cuvettes were incubated for 30 min at RT under sterile conditions. The transformed protoplasts were placed into small Petri dishes containing 2 ml of TEX buffer. The dishes were incubated O/N at 25ºC in the dark. The day after, the samples were placed into a 15-ml tube and centrifuged at 85 x *g* for 5 min without break. The pellet was removed and the suspension was ready to look with a Confocal Laser Scanning Microscope (CLSM) (Carl Zeiss Inc., Oberkochen, Germany) or the samples were prepared for protein extraction.

VI.2.3 In tobacco protoplasts by PEG- mediated transformation

The wild type *Nicotiana tabacum* cv Petit Havana plants used were grown under the same sterile conditions as described above. The protocol used was described in (Pedrazzini et al. 1994) and some changes were introduced. The leaf pieces were transferred into a Petri dish containing 7 ml of digestion mix (0.2% macerocyme R10, 0.4% cellulose R10 in K3 buffer: 3.78 g/L Gamborg's B5 basal medium with vitamins (Sigma-Aldrich-Aldrich Co., St. Louis, MO, USA), 136.2 g/L sucrose, 250 mg/L xylose, 250 mg/L NH_4NO_3, 750 mg/l $CaCl_2$, 1 mg/L 6-bencylaminopurine (BAP), 1 mg/L naphtalenacetic acid (NAA); pH 5.5 with KOH) with the abaxial side in contact with the medium. The plates were incubated O/N at 25ºC in the dark. Still under sterile conditions, the plates were gently shacked to release protoplasts from the cuticula. The mixture was filtered through a 100 µm sieve previously rinsed with W5 buffer (0.9 g/L glucose, 9 g/L NaCl, 0.37 g/L KCl, 18.37 g/L $CaCl_2$). The filtrated protoplasts were transferred into a 50 ml tube and centrifuged at 85 x *g* for 15 min at RT without break. The pellet was removed and the floating living protoplasts were washed by adding 4 volumes of W5 buffer and centrifuged at 85 x *g* for 10 min. This step was performed twice. The protoplasts were resuspended in 10 ml of W5 buffer and incubated 30 min in the dark. The cells were counted by diluting 50 µl of the cells in 450 µl of K3-FDA

buffer (10µl of 5 mg/ml fluorescein diacetate in 5 ml of K3 buffer). 10 µl were loaded into a grid slide and fluorescent protoplasts were counted using a microscope. The protoplasts were then centrifuged at 85 x *g* and the cells were resuspended in MaCa buffer (0.5 M mannitol, 20 mM $CaCl_2$, and 0.1% MES pH 5.7) to a final concentration of 1 x 10^6 cells/ml. The cells were subjected to heat shock for 5 min at 45°C and cooled down at RT for 5 min. 40 µg of the DNA to be transformed were placed into 15-ml sterile tubes and 1 ml of the cells were incorporated and mixed. 1 ml of 40% polyethylene glycol (PEG) was added drop wise and mixed very slowly. The mixture was incubated at RT for 30 min. To wash the samples, the tubes were filled with W5 buffer and centrifuged at 85 x *g* for 10 min. The cells were resuspended in 1 ml of K3 buffer (keeping the concentration 1 x 10^6 cells/ml) and incubated at 25°C O/N in the dark.

VI.3 Stable transformation of tobacco plants

Transgenic *N.tabacum* cv. Petite Havana SR1 plants were generated by leaf disc transformation using recombinant *A. tumefaciens* and transgenic T0 plants were regenerated from transformed calli (protocol derived from Fraley et al. (1983) and Horsch (1985). Briefly, wild type plants were grown on Murashige and Skoog (MS) medium supplemented with vitamins (Duchefa, Cesano Boscone, Italy) in sterile Magenta boxes and the young leaves were used for *Agrobacterium*-mediated transformation. The recombinant Agrobacteria suspension was prepared as described before and the OD_{600nm} was adjusted to 0.6 - 0.7 after dilution in LB broth medium supplemented with Kan, Rif, Gen and Tet and incubated at 28°C until the OD_{600nm} was 0.8 – 0.9. The leaves were cut into 8-10 pieces of 1 cm^2 (without the central vein) and transferred into Petri dishes containing 10 ml of recombinant Agrobacteria suspension and incubated for 5 min at RT. The leaf pieces were then dried with sterile Whatman filters to eliminate the excess of bacterial suspension. They were transferred into Petri dishes containing MS saccharose agar medium (4.4 g/L MS supplemented with vitamins, 30 g/L saccharose, 8 g/L plant agar (Duchefa), pH 5.6 - 5.8 and autoclaved) supplemented with 0.5 µ/ml BAP and 0.1 µg/ml NAA lacking antibiotic. The plates were incubated at 26°C in the light for two days. The leaf pieces were transferred onto MS saccharose plates containing 0.5 µ/ml BAP, 0.1 µg/ml NAA, 100 µg/ml carbanicillin and 50 µg/ml hygromycin and incubated at 26°C in the growth chamber

with 16 h photoperiod for two weeks until the calli were visible and optimal to cut. Twenty-five calli were transferred into Magenta boxes (5 calli per box) containing the regeneration medium (MS saccharose medium supplemented with 100 µg/ml carbanicillin, 50 µg/ml hygromycin and 0.1 µg/ml indole-3-acetic acid (IAA). The boxes were incubated at 26ºC with a 16 h photoperiod for 4 weeks until roots developed. The small plants were transferred into Magenta boxes containing regeneration medium and incubated at 26°C in 16 h light period until used for experiments. Young leaves from regenerated transgenic plants were used for extraction of total soluble proteins, for analysis of recombinant protein expression by Western blot, extraction of total RNA and expression of mRNA by Northern blot and stability of the recombinant proteins by immunoprecipitation.

VII RNA manipulations

VII.1 Total RNA extraction

The protocol to extract total RNA was following the Trizol® protocol from Invitrogen. About 125 mg of transgenic tobacco leaves were split into two samples, keeping one half for protein analysis and the other half for RNA extraction. Both samples were frozen in liquid N_2 and kept at -80ºC until used. For RNA extraction, the frozen samples were placed into a clean RNAse-free microcentrifuge tube and the tissue was mashed with a pestle. 400 µl of Trizol® (Invitrogen, San Giuliano Milanese, Italy) were added to each samples and after some mashing 600 µl of Trizol® were added (total of 1 ml). The samples were incubated at RT for 2-3 min. The tubes were centrifuged at 16,000 x *g* at 4°C for 10 min. The supernatant (~900 µl) was transferred into a new tube and 200 µl of RNAse-free chloroform were added. The tube was vortexed and incubated at RT for 2-3 min. After centrifuging at 16,000 x *g* at 4°C for 15 min, the supernatant (~500 µl) was transferred into a new tube and 500 µl of isopropanol were added. The tubes were briefly mixed and centrifuged at 16,000 x *g* at 4°C for 10 min. The supernatant was discarded and 1.2 ml of 75% RNAse-free ethanol was added to the pellet. The tubes were centrifuged at 16,000 x *g* at 4°C for 5 min and the pellet was air dried under the hood. The pellet was resuspended in 30 µl of DEPC water and incubated at 55°C for 10 min. The total RNA was frozen in liquid N_2 and kept at -80°C.

VII.2 DNAse I treatment

One of the aliquots of total RNA was thawed for DNAse I treatment. 15 µl of total RNA were incubated with 4 µl 10X reaction buffer (200 mM Tris-HCl pH 8.3, 500 mM KCl, 10 mM $MnCl_2$), 1 µl DNAse I (Roche, Monza, Italy) and 20 µl H_2O at 25°C for 30 min. To increase the volume, 160 µl of DEPC water were added (total volume of 200 µl). 200 µl of Buffer-saturated phenol (GibcoBRL, San Giuliano Milanese, Italy) were added and after vortexing, the tubes were centrifuged at 16,000 x *g* at RT for 4 min. The supernatant (~180 µl) was transferred into a new microcentrifuge tube and kept on ice. To the phenol samples, 180 µl of 600 mM Na acetate were added. The tubes were centrifuged at 16,000 x *g* at RT for 2 min and the supernatant was placed into the corresponding tubes kept on ice. To precipitate the RNA, 860 µl of 100% RNAse-free ethanol were added and the tubes were incubated on ice for 1 h. The tubes were centrifuged at 16,000 x *g* at 4°C for 30 min and the pellet was washed with 200 µl of 75% RNAse-free ethanol. The tubes were centrifuged at 16,000 x *g* at 4°C for 25 min and the pellet was air dried and resuspended in 15 µl of DEPC water.

VII.3 Determination of the RNA concentration

To prepare RNAse-free quartz cuvettes, the 100 µl cuvette (Hellma, Müllheim, Germany) was washed with 0.1 M NaOH and 1 mM EDTA. The final washings were performed with DEPC water. To determine the concentration of the RNA, the DNAseI-treated RNA samples were diluted 1:100 in distilled water and loaded into the quartz cuvettes. The samples were read at 260 nm in a RNA/DNA calculator GeneQuantII spectrophotometer (Pharmacia Biotech, Milan, Italy). The concentration was calculated by knowing that 40 µg/ml of single stranded RNA gives an absorbance of 1 at 260 nm in water (Sambrook 2001).

To determine the purity of the RNA, the ratio of the readings (A260/A280) provides an estimate value with respect to contaminants such as proteins. This ratio is influenced by pH and therefore the readings may be performed in 10 mM Tris-HCl pH 7.5 buffer. If the readings are performed in water, the pH can vary greatly as water is not buffered and the sensitivity to protein contaminants can be reduced.

VII.4 Determination of the RNA quality

A 1% agarose gel was prepared without ethidium bromide. To prepare the RNA samples, 1 µg of DNAse I-treated RNA was incubated with 1 µl of 200 µg/ml ethidium bromide, 1 µl of 37% formaldehyde (Merck, Rome, Italy) and 5 µl of formamide (Sigma-Aldrich) in a total volume of 12 µl. The mixture was incubated at 55°C for 20 min. As control, 1 µg, 0.5 µg and 0.25 µg of λ DNA/HindIII plus marker (Fermentas) were loaded together with 1 µl 200 µg/ml ethidium bromide. The samples were mixed with 6X loading dye (Fermentas), loaded and the gel was run until the dye has first entered the gel by about 1 cm. A picture was taken to determine the quality of the RNA. The gel was run for longer time and another picture was taken.

VII.5 Northern blot

The Northern blot was carried out as described in Viotti et al. (1982) but some changes were introduced. Briefly, the RNA samples were denatured in 50% formamide, 2.2 M formaldehyde, 20 mM MOPS (3-(N-morpholino)propanesulfonic acid) pH 7.0; 5 mM Na acetate and 1 mM EDTA and resolved in a 1% agarose gel containing 2.2 M formaldehyde for 3 h at 50 V. After electrophoresis, the RNA in the gel was transferred into a nylon Hybond-N+ membrane (Amersham Bioscience Biosciences, Milan, Italy) by capillarity. In a cuvette, in this order, at the bottom a glass was placed which holds a piece of 3 MM Whatman paper that is in contact with the bottom of the cuvette, followed by the agarose gel up side down, the Hybond-N+ membrane previously wetted in 2X SSC (300 mM NaCl and 30 mM sodium citrate pH 7.0), 4 pieces of 3 MM Whatman paper previously wetted in 2X SSC, 3 packs of tissue paper, another glass and a weight of 1.1 kg. This structure was wrapped in Saran wrap in order to avoid evaporation during the incubation at RT O/N. In the cuvette there is 10 X SSC buffer (1.5 M NaCl and 15 mM sodium citrate, pH 7.0) which is in contact with the Whatman paper below the agarose gel and this buffer will rise by capillarity, getting the RNA into the membrane where it will remain.

The DNA probe was labelled with the DecaLabel™ DNA labelling kit (Fermentas) using α-^{32}P-dCTP (Perkin-Elmer, Monza, Italy) following the manufacturer instructions. 100 ng of p24 gene DNA probe was mixed with free-nuclease H_2O to a final volume of 10 µl. 10 µl of 5X reaction buffer (0.25 M Tris-HCl buffer, pH 8.0 at

20°C) containing 25 mM $MgCl_2$, 5 mM dithiotreitol and random decamer primer (12.5 o.u./ml) were mixed, vortexed and centrifuged for 5 sec. The tube was incubated in a boiling water bath for 7 minutes and kept on ice for 4 min. 3 µl of Mix C (0.33 mM dGTP, 0.33 mM dATP and 0.33 mM dTTP) were added followed by the addition of 5 µl of α-^{32}P-dCTP (50 µCi) and 1 µl of Klenow fragment. The mixture was mixed and centrifuged for 5 sec. The tube was incubated at 37°C for 5 min. 4 µl of dNTP mix (0.25 mM dGTP, 0.25 mM dATP, 0.25 mM dTTP and 0.25 mM dCTP) were added and the tube was incubated at 37°C for 5 min. To stop the reaction, 1 µl of 0.5 M EDTA pH 8.0 was added to the mixture. The mixture was cleaned using QIAquick PCR purification kit (Qiagen). Five volumes of PB buffer (provided with the kit) were added and the mixture was placed into a column and centrifuged for 1 min at maximum speed. The flow through was discarded and column was washed with 600 µl of PE buffer. After centrifuging for 1 min at maximum speed, the column was centrifuged again as before. The DNA was eluted in 50 µl of elution buffer EB. The quality of the radiolabelled probe was determined with a β-Liquid Scintillation Analyser 1600 TR (Packard, Canberra Company, Cassina De' Pecchi, Italy).

The different membranes were pre-hybridised in 1X hybridisation mix (10% (w/v) dextran sulphate, 6X SCP buffer (20X is 0.265 M Na_2HPO_4, 334.6 M NaH_2PO_4, 2 M NaCl, 20 mM EDTA, pH 6.5), 10% sodium lauryl sarcosine, 0.5 mg/ml heparin) at 65ºC for 2.5 h. Before starting the hybridisation, 100 µg/ml of sonified single stranded salmon sperm DNA were incubated at 100ºC for 10 min, and the radiolabelled probe was boiled for 5 min. Both solutions were added to the pre-hybridisation mix and incubated for 16 h at 65ºC. The membranes were washed 3 times as followed. A quick wash with 2X SSC buffer containing 0.1% SDS shacking by hand at RT. The second wash was performed with 2X SSC buffer containing 0.1% SDS at 65ºC for 40 min. The third wash was with 0.2X SSC buffer with 0.1% SDS at 65ºC for 30 min raising the astringency with the diluted SSC. The membrane was wrapped in Whatman paper and Saran wrap and exposed to a Xar film (Kodak, Cinisello Balsamo, Italy) at -80°C for 4 h.

VIII Protein analysis

VIII.1 Total protein purification

VIII.1.1 From transiently transformed protoplasts

Protoplasts were transformed and on the third day the tubes were filled with 250 mM NaCl and centrifuged at 180 *x g* with break. The supernatant was removed and the pellet was resuspended in α-amylase extraction buffer (6.705 g/L malic acid, 3.5 g/L NaOH, 2.92 g/L NaCl, 0.295 g/L $CaCl_2$, 0.05 g/L Na azide, pH 5.2 supplemented with protein inhibitors: 5 mM DTT, 10 μg/ml leupeptin, 1 μg/ml aprotinin, 1 μg/ml pepstatin, 1 μg/ml E64 and 0.2 μg/ml phenanthrolin). The pellets were placed in microcentrifuge tubes to a final weight of 200 mg. Extra drops of α-amylase extraction buffer were added to reach the weight. The samples were sonicated for 3 seconds at 50% burst with a sonicator (Bandelin Sonoplus GM70, Bandelin, Berlin, Germany) and kept on ice. They were centrifuged at 16,813 x *g* for 10 min at 4ºC. 150 μl of the supernatant were placed in new microcentrifuge tubes which were labelled as 'Soluble Fraction'. The rest of the supernatant was removed and the same quantity of the total supernatant removed was added of α-amylase extraction buffer. The samples were sonicated as before and 150 μl were placed in new microcentrifuge tubes. These samples were labelled as 'Membrane Fraction'.

For protoplasts in which Phaseolin extraction buffer was used (12.11 g/L Tris pH 7.8, 11.68 g/L NaCl, 0.372 g/L EDTA, 2% β-mercaptoethanol, 0.2% Triton) no sonication was performed.

VIII.1.2 From stable transformed plants

The stable transformed plants were used for total protein extraction. 200 mg of leaf previously frozen at -80ºC were homogenized in an ice-cold mortar with ice-cold homogenation buffer (0.1 M Tris-HCl pH 7.8, 0.2 M NaCl, 1 mM EDTA, 2% β-mercaptoethanol and 0.2% Triton X-100) supplemented with Complete Protease Inhibitor Cocktail (Boehringer, Milan, Italy). The ratio of buffer/weight was 5:1. The mixture was transferred into microcentrifuge tubes and centrifuged at 5000 x *g* for 10 min at 4ºC. The supernatant was transferred into a pre-chilled microcentrifuge tube and frozen at -20ºC.

VIII.2 Quantification of total soluble proteins

VIII.2.1 From transiently transformed protoplasts

To quantify the protein concentration, the Bradford assay was used (Bradford 1976). From transiently transformed protoplasts, 10 µl of protein samples were diluted in 10 µl of α-amylase extraction buffer (1:2 dilution) and added to the plastic disposable cuvettes (Sarstedt, Numbrecht, Germany). The blank was 20 µl of α-amylase extraction buffer and 1 ml of Bradford reagent (Sigma-Aldrich) was added to all the cuvettes and incubated at RT for at least 5 min. The standard curve was done with known concentrations of bovine serum albumin (BSA) and the O.D. measured at 595 nm and the concentrations were calculated with the O.D. values and the standard equation.

VIII.2.2 From stable transformed plants

In this case, ovoalbumin was used to perform the standard curve as the homogenation buffer in which the proteins were purified contains β-mercaptoethanol and interferes with BSA. The different protein samples to be tested were diluted 1:6, 1:10 or 1:15 in homogenation buffer to a final volume of 30 µl. As blank 30 µl of homogenation buffer were used. The diluted samples and the blank were placed into disposable plastic cuvettes and 1 ml of Bradford reagent (Sigma-Aldrich) was added to all samples and incubated at least for 5 min at RT. The samples were read at 595 nm and the protein concentration was determined by the formula given by the standard curve and the dilution used.

VIII.3 Western blot

VIII.3.1 From transiently transformed protoplasts and agroinfiltrated tobacco leaves

In all cases 12% separating acrylamide gels were prepared. For a big gel (30 wells): 10.05 ml dH_2O, 3.15 ml 3 M Tris-HCl pH 8.8, 9 ml Protogel (30% w/v acrylamide, 0.8% w/v bis-acrylamide), 225 µl 10% SDS, 12.5 µl TEMED, 75 µl 10% ammonium persulfate (APS). The stacking gel was prepared with 11.3 ml 20% sucrose, 1 ml 1 M Tris-HCl pH 6.8, 2.5 ml Protogel, 150 µl 10% SDS, 25 µl TEMED, 40 µl 10% APS. The amount of the samples was adjusted to load 2.5 µg of total protein per well with sample buffer mix (0.1% bromophenol blue, 5 mM EDTA, 200 mM Tris pH 8.8, 1 M

sucrose, 25% of 10% SDS, 0.015% 1 M DTT). The gel was run in running buffer (6 g/L Tris, 28.8 g/L glycine, 10 ml 10% SDS) at 50 mA for about 2.5 h. The nitrocellulose membrane was blotted in Trans-Blot® SD semi-dry electrophoretic transfer cell (BioRad) and run at 200 mA for 2 h with electroblotting buffer (25 mM Tris, 150 mM glycine, 10% (v/v) methanol). After blotting, the membrane was stained with Ponceau solution (0.1% Ponceau S in 5% acetic acid). After one minute, the Ponceau solution was poured off and the membrane rinsed with tap water. The protein bands were visible after some washings it was determined the quantity and the quality of the protein samples loading. The membrane was further washed with PBS-T (Phosphate Buffered Saline with Tween: 8.7 g/L NaCl, 2.25 g/L $Na_2HPO_4 \cdot 2H_2O$, 0.2 g/L KH_2PO_4; pH 7.4 supplemented with 1 ml Tween-20). The blocking solution (1% BSA and 5% milk in PBS) was incubated for 1 h at RT on a rocker. The membrane was washed several times with PBS-T and incubated with the primary antibody. To detect p24, 1:500 polyclonal sheep anti-p24 (Aalto BioReagents, Dublin, Ireland) was used and, to detect RFP, 1:500 monoclonal mouse anti-DsRed (Becton Dickinson, Franklin Lakes, NJ, USA) and 1:500 polyclonal rabbit anti-RFP (Biozol Diagnostica Vertrieb GmbH, Eching, Germany). The antibodies were diluted in 1% BSA, 0.02% NaN_3 in PBS and incubated with the membrane O/N at 4ºC. The membrane was washed and the diluted antibodies were kept at 4ºC. As secondary antibody, 1:20,000 donkey anti-sheep-HPO (Horseradish Peroxidase) (Sigma-Aldrich-Aldrich Co.) was used to detect anti-p24, and 1:20,000 goat anti-mouse-HPO (Sigma-Aldrich) and 1:20,000 goat anti-rabbit-HPO (Sigma-Aldrich) to detect anti-RFP. The incubation was for 1 h at RT on a rocker. The membrane was washed several times with PBS-T and developed. To develop, two ECL solutions were prepared. Solution A contains 5 ml 100 mM Tris-HCl pH 8.5, 50 µl of 250 mM Luminol and 25 µl p-Coumaric acid. Solution B is 5 ml 100 mM Tris-HCl pH 8.5 and 3 µl 30% H_2O_2. The two solutions were mixed and placed on the membrane for 1 min. The solution was discarded and the excess of solution was absorbed with absorbing paper. The membrane was then placed on transparent film, wrapped and placed on a film cassette. An Amersham Bioscience ECL film was inserted and the cassette was closed to expose for the appropriate time, usually for 1, 5 and 30 min. The film was developed with developer solution, fixed, rinsed with tap water and air dried.

VIII.3.2 From stable transformed plants

In this case 15% separating acrylamide gels were prepared. For a big gel (30 wells): 7.5 ml dH_2O, 7.5 ml of separating buffer (1.5 M Tris-HCl pH 8.8 and 0.4% SDS), 15 ml separating acrylamide (30% w/v acrylamide, 0.15% w/v bis-acrylamide), 15 µl TEMED, 150 µl 10% APS. The 4.5% stacking gel was prepared with 9 ml water, 3.75 ml stacking buffer (0.5 M Tris-HCl pH 6.8 and 0.4% SDS), 2.25 ml stacking acrylamide (30% w/v acrylamide, 0.8% w/v bis-acrylamide), 15 µl TEMED, 45 µl 10% APS. The gel was run in running buffer (3 g/L Tris, 14.4 g/L glycine, 1 g SDS) at 35 mA for about 5 h or at 10 mA at 4ºC O/N. The polyvinylidene difluoride (PVDF) membrane was blotted in Trans-Blot® cell electrophoretic transfer system (BioRad) and run at 180 mA 60 V max at 4ºC O/N or at 400 mA 100 V max at 4ºC for 4 hours with electroblotting buffer (25 mM Tris, 150 mM glycine, 10% (v/v) methanol). After blotting, the membrane was stained with Ponceau solution (5 mg/ml Ponceau S in 5% trichloroacetic acid (TCA)). After twenty minutes, the Ponceau solution was poured off and the membrane rinsed with tap water. The incubation with the antibodies was carried out as described before. To develop, West-pico supersignal kit (Pierce, Pero, Italy) was used following the manufacturer instructions. The same amount of the two solutions were mixed and incubated with the membrane for 5 minutes. The solution was discarded as before and the membrane was then placed on transparent film, wrapped and placed on a film cassette. A Xar film was inserted (Kodak) and the cassette was closed to expose for the appropriate time, usually for 1, 5 or 30 min. The film was developed with the developer solution, fixed, rinsed with tap water and air dried.

VIII.4 Staining of protein gels and membranes

VIII.4.1 Coomassie brilliant blue staining

Proteins separated on SDS-PAGE gels were visualized by Coomassie brilliant blue staining. Proteins were detected after incubating the gel or the PVDF membrane for 30 min in Coomassie staining solution (Coomassie brillant blue G-250 0.25% (w/v), methanol 50% (v/v), glacial acetic acid 9% (v/v)) at RT with agitation. The staining was removed with destaining solution (30% (v/v) isopropanol and 7% acetic acid) until the protein bands were clearly visible. The gel or the membrane was air dried.

VIII.4.2 Ponceau S staining

To ascertain the transfer of the proteins was correct, after the blotting the membrane was incubated in Ponceau S solution (0.1% Ponceau S in 5% acetic acid or 5 mg/ml Ponceau S in 5% TCA). The membrane was incubated for one minute when the solution was in acetic acid and for 20 minutes when was in TCA. The Ponceau solution was poured back and the membrane was rinsed with tap water. After the bands could be visualized, the quality of the transfer was evaluated.

VIII.5 Vacuole isolation

Tobacco protoplasts were transformed and about 20 h after electroporation, the protoplasts were centrifuged at RT at 85 x *g* without break. The pelleted death cells were removed with a fine Pasteur pipette and the floating living protoplasts were washed and pelleted in 50 ml of 250 mM NaCl. The supernatant was discarded and the pellet kept on ice. 5 ml of preheated at 42ºC lysis medium (0.2 M mannitol, 10% Ficoll-400, 20 mM EDTA, 2 mM DTT, 5 mM HEPES, 10 μg/ml neutral red, 150 μg/ml BSA, pH 8.0) was added and mixed gently with the pipette. The solution was incubated at 42ºC for 1 min. The vacuole buffer (0.6 M betaine, 10 mM HEPES, 150 μg/ml BSA, 0.1 μg/ml pepstatin; pH 7.5) was mixed with lysis medium at a ratio of 1:1. 3 ml of the 1:1 lysis medium /vacuole buffer was slowly layered on the sample and 1 ml of vacuole buffer was added on top. The mixture was centrifuged at 3000 x *g* for 15 min at 4ºC without break and the vacuole layer could be seen in red due to the neutral red dye. The vacuoles were removed by "hoovering" and placed on ice. At this moment a sample was be taken and observed with the CLSM. The rest of the sample was sonicated for 5 seconds and centrifuged at 1500 x *g* for 3 min to pellet the vacuoles. 100 μl of protein sample buffer was added to 100 μl of sample and the tubes were frozen at -80ºC.

VIII.6 Radioactive labelling and immunoprecipitation

Labelling was performed at 25°C in the dark using 100 μCi/ml of Pro-Mix (a mixture of ^{35}S-Met and ^{35}S-Cys; Amersham Bioscience Biosciences) in the presence or absence of 150 μg/ml BSA. The chase was performed by adding unlabelled Methionine and Cysteine to a final concentration of 10 mM and 5 mM respectively. The samples were

washed with 3 volumes of W5 buffer and centrifuged for 10 min at 60 x *g*. The supernatant was labelled as "medium sample" and the pellet as "cell sample".

The "cell samples" were homogenized by adding 2 volumes of ice-cold homogenation buffer (150 mM Tris-HCl pH 7.5, 150 mM NaCl, 1.5 mM EDTA, 1.5% (w/v) Triton X-100) supplemented just before use with Complete Protease Inhibitor Cocktail (Boehringer). After vortexing, the samples were centrifuged at 11,700 x *g* for 5 min at 4°C. The supernatant was washed with 1 ml of NET-gel buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.1% Nonidet P-40, 0.25% gelatin, 0.02% sodium azide) and centrifuged at 11,700 x *g* for 5 min at 4°C. The supernatant was transferred into a new tube.

The "medium samples" were centrifuged at 11,700 x *g* for 5 min at 4°C. 850 µl were transferred into a new tube and 150 µl of supplement buffer (2% gelatin, 120 µl of 1 M Tris-HCl pH 7.6, 24 µl 5 M NaCl, 2.4 µl 0.5 M EDTA, 12 µl Triton X-100 and 48 µl of 25X Complete Protease Inhibitor Cocktail) were added to a final volume of 1200 µl. After vortex, the samples were centrifuged at 11,700 x *g* for 5 min at 4°C and the supernatant was transferred into a clean tube.

The last supernatants from both, cell and medium samples, were subjected to immunoprecipitation using 1:500 rabbit polyclonal anti-p24 (ARP432, Medical Research Council AIDS Directed Programme). The samples were incubated at 4°C for 1.5 h in rotation. 100 µl of Protein A-Sepharose (GE healthcare, Milan, Italy) were added and incubated for 2 h at 4°C rotating. The beads were pelleted by centrifugation at 11,700 x *g* for 1 min at 4°C and washed 3 times with NET-gel buffer. An equal volume of 2X SDS-PAGE loading buffer (40 mM Tris-HCl pH 8.6, 2% SDS, 0.66% β-mercaptoethanol, 20% glycerol, 0.03% Bromophenol blue) was added and the samples were boiled at 95°C for 5 min. The samples were loaded on a 15% SDS-PAGE using Rainbow ^{14}C-methylated proteins (Sigma-Aldrich-Aldrich Co.) as molecular mass markers. After electrophoresis, the gel was fixed with 30% isopropanol and 7% of acetic acid and washed 3 times with dimethyl sulfoxide (DMSO) (Mallinckrodt Baker, Milan, Italy). Then, the gel was treated with 2,5-diphenyloxazole (PPO) (Merck) dissolved in DMSO for 2 h with agitation and washed 3 times with water. The gel was dried on a filter paper and Saran wrap paper for 2 h at 70°C. After removing the Saran wrap paper, the dried gel was exposed with a radiography Xar film (Kodak) in an appropriate cassette and stored at -80°C until developed.

VIII.7 Purification of the recombinant p24 protein

Five hundred mg from leaves of transgenic plants were homogenized as described before but in this case the ratio of buffer/weight was 6:1. The protein concentration was determined and 900 µg were used for the experiment. The different samples were adjusted to have 900 µg of total soluble protein (TSP) in 300 µl of final volume. To immunoprecipitate the different fusion proteins, 700 µl of Net-gel buffer were added and in the samples with the zein fusions, the Net-gel buffer contained 8% β-ME. Then the antibody to immunoprecipitate one of the proteins from the fusion was added. When the p24 was the protein to be immunoprecipitated, 1:500 of rabbit anti-p24 (ARP432) was used. To immunoprecipitate zein and tail-anchor, 1:200 polyclonal rabbit anti-zein (Bellucci et al. 2000) and 1:300 monoclonal mouse anti-op3 (Adamus et al. 1991) were used. The mixtures were incubated at 4ºC for 2 h in agitation. 100 µl of protein A sepharose for rabbit antibodies or 30 µl of protein G sepharose (Pierce) for mouse antibodies were added and incubated for 1.5 h as before. The samples were washed three times with Net-gel buffer and one time with PBS. The beads were resuspended in 294 µl of PBS and 6 µl 10% Triton and separated into treated and non-treated samples. 4 µl (4 units) of thrombin (Amersham Bioscience) were added to the samples labelled as "treated" and 4 µl of PBS to the "non-treated" and incubated at 22ºC O/N in agitation (this corresponds to 8.8 units of thrombin per mg of protein). The day after 110 µl were taken to separate the resin and the supernatant by centrifuging at 11,700 x *g* for 2 min. 110 µl of the supernatant were labelled as "medium sample" and mixed wit 55 µl of 3X denaturation buffer. The "resin sample" was resuspended in 20 µl of PBS and mixed with 25 µl of 3X denaturation buffer. The samples were boiled for 5 min and loaded in a 15% SDS-PAGE. The Western blot was performed as described before.

VIII.8 Confocal microscopy

Cells were observed under a Zeiss Axiovert LSM510 Meta microscope using the Plan-Neofluar 25X/0.8 corr DIC and the C-Apochromat 63x/1.2 W corr water immersion objectives. Special settings were designed for observing single-, double- and triple- expression with different XFP-tagged constructs. Fluorescence was detected by the Metadetector using the main beam splitters HFT 405/514, HFT 458/514 and HFT 488/543. Fluorophores were excited by line switching in the multi-tracking mode of the

microscope. The GFP was excited at 488 nm and emission at 496-518 nm and YFP by excitation at 514 nm and emission at 550-571 nm, both with the Argon laser. The RFP was excited with the HeNe (helium-neon) laser at 543 nm and emission at 603-646 nm. The double detection of GFP/RFP was performed at excitation 488/543 nm and emission at 496-529 nm and 593-646 nm whereas the double detection of YFP/RFP was performed at excitation 514/543 and emission at 530-600 nm and 560-615 nm. Pinholes were adjusted to 1 Airy Unit for each wavelength. Images were post-processed using the Zeiss LSM Image Browser (Version 3.5.0.376).

VIII.9 Electron microscopy

Small pieces of young leaves were fixed in 1.6% (w/v) paraformaldehyde mixed with 1.5% (v/v) glutaraldehyde in 0.1 M phosphate buffer pH 6.9 for 1 h at RT. After washing with 0.1 M phosphate buffer and double-distilled water, the samples were incubated with 2% uranyl acetate for 30 min at RT followed by 4 washings with double-distilled water. The samples were dehydrated in ethanol and embedded overnight in LR White resin at 60ºC. Ultrathin sections were cut using a Leica Microsystems Ultracut UCT (Leica Microsystems Nussloch GmbH, Nussloch, Germany), mounted on 300-mesh nickel grids and immunogold labeled. Grids were floated on drops 3% bovine serum albumin (BSA) in PBS for 15 min. They were then incubated with anti-zein antiserum (1:200 dilution) or anti-op3 antiserum (1:10 dilution) in 1% BSA in PBS for 1 h at RT. Controls were incubated following the same protocol. After washing with 1% BSA in PBS, the sections were incubated in the same buffer with goat anti-rabbit secondary antibody for anti-zein and with goat anti-mouse secondary antibody for anti-op3 (1:50 dilution) conjugated with 15 nm gold particles (BBInternational, Cardiff, UK). The grids were washed in drops of 1% BSA in PBS, and double-distilled water, poststained in 3% uranyl acetate in water, and examined under an electron microscope (EM CM10; Philips, Eindhoven, The Netherlands).

IX Bioinformatic tools

IX.1 Primer design

The different primers were designed using Primer Designer (version 2.0, Scientific and Educational software, 1990-91). This software permits to design oligonucleotides to amplify specific DNA fragments by PCR.

IX.2 GentLe and VectorNti

Before starting any DNA cloning, the data sequences of interest were first introduced into the Gentle Cloning Program (Magnus Manske, V 1.8.4, 2003) and Vector NTI (Invitrogen) in order to create, virtually, all the DNA constructs. These programs allowed to double check if the expected bands of the subclonings were correct when performing the different experiments on the bench (PCR, digestions, ligations, etc.). Both programs are used for DNA and amino acid editing, database management, plasmid maps, restriction analysis and ligation reactions, DNA alignments, sequencer data import, calculators, gel image display, PCR and other features.

IX.3 SignalP

This program predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms. The method incorporates a prediction of cleavage sites and a signal peptide/non-signal peptide prediction based on a combination of several artificial neural networks and hidden Markov models.
The hidden Markov model (HMM) calculates the probability of whether the submitted sequence contains a signal peptide or not. The eukaryotic HMM model also reports the probability of a signal anchor, previously named uncleaved signal peptides. It contains submodels for the N-terminal part (n-region) of the signal peptide, the hydrophobic region (h-region) and the region around the cleavage site (c-region). Furthermore, when a signal peptide is found, the cleavage site is assigned by a probability score together with scores for the three regions (Bendtsen et al. 2004).

BIBLIOGRAPHY

Adamus, G., et al. (1991). "Genetic control of antibody response to bovine rhodopsin in mice: epitope mapping of rhodopsin structure." J Neuroimmunol 34(2-3): 89-97.
Agarwal, S., et al. (2008). "Expression of modified gene encoding functional human alpha-1-antitrypsin protein in transgenic tomato plants." Transgenic Res.
Alberts, B., et al. (2002). "Molecular biology of the cell."
Arcalis, E., et al. (2004). "Unexpected deposition patterns of recombinant proteins in post-endoplasmic reticulum compartments of wheat endosperm." Plant Physiol 136(3): 3457-66.
Avesani, L., et al. (2003). "Improved in planta expression of the human islet autoantigen glutamic acid decarboxylase (GAD65)." Transgenic Res 12(2): 203-12.
Bagga, S., et al. (1997). "Co-expression of the maize delta-zein and beta-zein genes results in stable accumulation of delta-zein in endoplasmic reticulum-derived protein bodies formed by beta-zein." Plant Cell 9(9): 1683-96.
Bakker, H., et al. (2001). "Galactose-extended glycans of antibodies produced by transgenic plants." Proc Natl Acad Sci U S A 98(5): 2899-904.
Barbante, A., et al. (2008). "Anchorage to the cytosolic face of the endoplasmic reticulum membrane: a new strategy to stabilize a cytosolic recombinant antigen in plants." Plant Biotechnol J.
Bardor, M., et al. (2003). "Immunoreactivity in mammals of two typical plant glyco-epitopes, core alpha(1,3)-fucose and core xylose." Glycobiology 13(6): 427-34.
Bardor, M., et al. (2003). "Monoclonal C5-1 antibody produced in transgenic alfalfa plants exhibits a N-glycosylation that is homogenous and suitable for glyco-engineering into human-compatible structures." Plant Biotechnol J 1(6): 451-62.
Barta, A., et al. (1986). "The expression of a nopaline synthase-human growth hormone chimeric gene in transformed tobacco and sunflower callus tissue." Plant Molec Biol 6: 347-357.
Belanger, H., et al. (2000). "Human respiratory syncytial virus vaccine antigen produced in plants." Faseb J 14(14): 2323-8.
Bellucci, M., et al. (2000). "Accumulation of maize γ-zein and γ-zein: KDEL to high levels in tobacco leaves and differential increase of BiP synthesis in transformants." Theor Appl Genet 101(5-6): 796-804.
Bendtsen, J. D., et al. (2004). "Improved prediction of signal peptides: SignalP 3.0." J Mol Biol 340(4): 783-95.
Berg, R. H. and R. N. Beachy (2008). "Fluorescent protein applications in plants." Methods Cell Biol 85: 153-77.
Bhardwaj, D., et al. (2006). "Recombinant HIV-1 p24 protein: cloning, expression, purification and use in the development of ELISA kits." Current Science 91(7): 913-917.
Birch-Machin, I., et al. (2004). "Accumulation of rotavirus VP6 protein in chloroplasts of transplastomic tobacco is limited by protein stability." Plant Biotechnol J 2(3): 261-270.
Boehm, R. (2007). "Bioproduction of therapeutic proteins in the 21st century and the role of plants and plant cells as production platforms." Ann N Y Acad Sci 1102: 121-34.
Bonifacino, J. S. and B. S. Glick (2004). "The mechanisms of vesicle budding and fusion." Cell 116(2): 153-66.
Borgese, N., et al. (2003). "Biogenesis of tail-anchored proteins." Biochem Soc Trans 31(Pt 6): 1238-42.
Borgese, N., et al. (1993). "NADH-cytochrome b5 reductase and cytochrome b5 isoforms as models for the study of post-translational targeting to the endoplasmic reticulum." FEBS Lett 325(1-2): 70-5.
Borgese, N., et al. (2001). "Targeting of a tail-anchored protein to endoplasmic reticulum and mitochondrial outer membrane by independent but competing pathways." Mol Biol Cell 12(8): 2482-96.
Bouche, F. B., et al. (2005). "Induction of broadly neutralizing antibodies against measles virus mutants using a polyepitope vaccine strategy." Vaccine 23(17-18): 2074-7.
Bradford, M. M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." Anal Biochem 72: 248-54.
Brandizzi, F., et al. (2002). "The destination for single-pass membrane proteins is influenced markedly by the length of the hydrophobic domain." Plant Cell 14(5): 1077-92.
Breitler, J. C., et al. (2002). "Efficient microprojectile bombardment-mediated transformation of rice using gene cassettes." Theor Appl Genet 104(4): 709-719.
Bretscher, M. S. and S. Munro (1993). "Cholesterol and the Golgi apparatus." Science 261(5126): 1280-1.
Broothaerts, W., et al. (2005). "Gene transfer to plants by diverse species of bacteria." Nature 433(7026): 629-33.
Bullock, W., Fernandez, J.M., and Short, J.M. (1987). "XL1-Blue, high efficiency plasmid transforming recA Escherichia coli strain with b-galactosidase section." Biotechniques 5: 376-9.
Caley, I. J., et al. (1997). "Humoral, mucosal, and cellular immunity in response to a human immunodeficiency virus type 1 immunogen expressed by a Venezuelan equine encephalitis virus vaccine vector." J Virol 71(4): 3031-8.
Capuano, F., et al. (2007). "Properties and exploitation of oleosins." Biotechnol Adv 25(2): 203-6.
Carrillo, C., et al. (2001). "Induction of a virus-specific antibody response to foot and mouth disease virus using the structural protein VP1 expressed in transgenic potato plants." Viral Immunol 14(1): 49-57.
Castilho, P. V., et al. (2005). "Heterologous expression, characterization and structural studies of a hydrophobic peptide from the HIV-1 p24 protein." Peptides 26(2): 243-9.
Chea, A., et al. (2007). "In vitro Antimicrobial Activity of Plants used in Cambodian Traditional Medicine." Am J Chin Med 35(5): 867-73.
Chen, J., et al. (2007). "Enhanced production of p24 Gag protein in HIV-1-infected rat cells fused with uninfected human cells." Exp Mol Pathol 83(1): 125-30.
Chiang, C. J., et al. (2007). "One-step purification of insoluble hydantoinase overproduced in Escherichia coli." Protein Expr Purif 52(1): 14-8.
Chien, A., et al. (1976). "Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus." J Bacteriol 127(3): 1550-7.
Chikwamba, R., et al. (2002). "A functional antigen in a practical crop: LT-B producing maize protects mice against Escherichia coli heat labile enterotoxin (LT) and cholera toxin (CT)." Transgenic Res 11(5): 479-93.
Chikwamba, R. K., et al. (2003). "Localisation of a bacterial protein in starch granules of transgenic maize kernels." Proc Natl Acad Sci U S A 100(19): 11127-32.
Choi, N. W., et al. (2005). "Synthesis and assembly of a cholera toxin B subunit-rotavirus VP7 fusion protein in transgenic potato." Mol Biotechnol 31(3): 193-202.
Chong, D. K. and W. H. Langridge (2000). "Expression of full-length bioactive antimicrobial human lactoferrin in potato plants." Transgenic Res 9(1): 71-8.
Chong, D. K., et al. (1997). "Expression of the human milk protein beta-casein in transgenic potato plants." Transgenic Res 6(4): 289-96.
Chou, K. C. (2001). "Prediction of protein cellular attributes using pseudo-amino acid composition." Proteins 43(3): 246-55.

Christensen, A. H. and P. H. Quail (1996). "Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants." Transgenic Res 5(3): 213-8.
Christou, P. (1996). "Transformation technology." Trends Plant Sci 1: 423-431.
Chugh, P. (2003). "Cloning of gag gene of HIV-1 subtype c (Indian strain) into a mammalian expression vector and in vitro expression studies." Indian J Med Microbiol 21(2): 77-81.
Clark, J. M. (1988). "Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases." Nucleic Acids Res 16(20): 9677-86.
Coleman, C. E., et al. (1996). "The maize gamma-zein sequesters alpha-zein and stabilizes its accumulation in protein bodies of transgenic tobacco endosperm." Plant Cell 8(12): 2335-45.
Cook, N. R., et al. (2004). "Lysosome associated membrane protein 1 (Lamp1) traffics directly from the TGN to early endosomes." Traffic 5(9): 685-99.
Crofts, A. J., et al. (1999). "Saturation of the endoplasmic reticulum retention machinery reveals anterograde bulk flow." Plant Cell 11(11): 2233-48.
Dai, Z., et al. (2000). "Improved plant-based production of E1 endoglucanase using potato: expression optimization and tissue targeting." Mol Breed 6: 277-285.
Daniell, H. (2006). "Production of biopharmaceuticals and vaccines in plants via the chloroplast genome." Biotechnol J 1(10): 1071-9.
Datta, S. A., et al. (2007). "Conformation of the HIV-1 Gag protein in solution." J Mol Biol 365(3): 812-24.
Davey, M. R., et al. (2005). "Plant protoplasts: status and biotechnological perspectives." Biotechnol Adv 23(2): 131-71.
De Cosa, B., et al. (2001). "Overexpression of the Bt cry2Aa2 operon in chloroplasts leads to formation of insecticidal crystals." Nat Biotechnol 19(1): 71-4.
de Pater, B. S., et al. (1992). "The promoter of the rice gene GOS2 is active in various different monocot tissues and binds rice nuclear factor ASF-1." Plant J 2(6): 837-44.
de Virgilio, M., et al. (2008). "The human immunodeficiency virus antigen Nef forms protein bodies in leaves of transgenic tobacco when fused to zeolin." J Exp Bot.
Ding, F. X., et al. (2007). "A novel, cheap and effective fusion expression system for the production of recombinant proteins." Appl Microbiol Biotechnol 77(2): 483-8.
Dong, J. L., et al. (2005). "Oral immunization with pBsVP6-transgenic alfalfa protects mice against rotavirus infection." Virology 339(2): 153-63.
Doran, P. M. (2006). "Foreign protein degradation and instability in plants and plant tissue cultures." Trends Biotechnol 24(9): 426-32.
Dorokhov, Y. L., et al. (2007). "Superexpression of tuberculosis antigens in plant leaves." Tuberculosis (Edinb) 87(3): 218-24.
Dowson Day, M. J., et al. (1993). "Plant viral leaders influence expression of a reporter gene in tobacco." Plant Mol Biol 23(1): 97-109.
Drake, P. M., et al. (2003). "Rhizosecretion of a monoclonal antibody protein complex from transgenic tobacco roots." Plant Mol Biol 52(1): 233-41.
Dus Santos, M. J., et al. (2005). "Development of transgenic alfalfa plants containing the foot and mouth disease virus structural polyprotein gene P1 and its utilization as an experimental immunogen." Vaccine 23(15): 1838-43.
Ellgaard, L., et al. (1999). "Setting the standards: quality control in the secretory pathway." Science 286(5446): 1882-8.
Evangelista, R. L., et al. (1998). "Process and economic evaluation of the extraction and purification of recombinant beta-glucuronidase from transgenic corn." Biotechnol Prog 14(4): 607-14.
Fernandez-San Millan, A., et al. (2008). "Human papillomavirus L1 protein expressed in tobacco chloroplasts self-assembles into virus-like particles that are highly immunogenic." Plant Biotechnol J 6(5): 427-41.
Fischer, R. (2005). "Qualitative and quantitative comparison of protein expression systems." Oral presentation held at the Conference of plant-made pharmaceuticals.
Fischer, R., et al. (1999). "Expression and characterization of bispecific single-chain Fv fragments produced in transgenic plants." Eur J Biochem 262(3): 810-6.
Fischer, R., et al. (2004). "Plant-based production of biopharmaceuticals." Curr Opin Plant Biol 7(2): 152-8.
Fisk, H. J. and A. M. Dandekar (2005). "Electroporation: introduction and expression of transgenes in plant protoplasts." Methods Mol Biol 286: 79-90.
Floss, D. M., et al. (2008). "Biochemical and functional characterization of anti-HIV antibody-ELP fusion proteins from transgenic plants." Plant Biotechnol J 6(4): 379-91.
Franke, E. K. and J. Luban (1996). "Inhibition of HIV-1 replication by cyclosporine A or related compounds correlates with the ability to disrupt the Gag-cyclophilin A interaction." Virology 222(1): 279-82.
Franke, E. K., et al. (1994). "Specific incorporation of cyclophilin A into HIV-1 virions." Nature 372(6504): 359-62.
Frankel, A. D. and J. A. Young (1998). "HIV-1: fifteen proteins and an RNA." Annu Rev Biochem 67: 1-25.
Frigerio, L., et al. (2008). "Multiple Vacuoles in Plant Cells: Rule or Exception?" Traffic.
Frigerio, L., et al. (2001). "Influence of KDEL on the fate of trimeric or assembly-defective phaseolin: selective use of an alternative route to vacuoles." Plant Cell 13(5): 1109-26.
Fu, J. and G. Kreibich (2000). "Retention of subunits of the oligosaccharyltransferase complex in the endoplasmic reticulum." J Biol Chem 275(6): 3984-90.
Fujiyama, K., et al. (2006). "In Planta production of immunogenic poliovirus peptide using tobacco mosaic virus-based vector system." J Biosci Bioeng 101(5): 398-402.
Galili, G. (2004). "ER-derived compartments are formed by highly regulated processes and have special functions in plants." Plant Physiol 136(3): 3411-3.
Gallie, D. R., et al. (1991). "Post-transcriptional regulation in higher eukaryotes: the role of the reporter gene in controlling expression." Mol Gen Genet 228(1-2): 258-64.
Ganser-Pornillos, B. K., et al. (2008). "The structural biology of HIV assembly." Curr Opin Struct Biol 18(2): 203-17.
Gao, Y., et al. (2003). "Oral immunization of animals with transgenic cherry tomatillo expressing HBsAg." World J Gastroenterol 9(5): 996-1002.
Gaume, A., et al. (2003). "Rhizosecretion of recombinant proteins from plant hairy roots." Plant Cell Rep 21(12): 1188-93.
Geli, M. I., et al. (1994). "Two Structural Domains Mediate Two Sequential Events in [gamma]-Zein Targeting: Protein Endoplasmic Reticulum Retention and Protein Body Formation." Plant Cell 6(12): 1911-1922.
Girard, L. S., et al. (2006). "Expression of a human anti-rabies virus monoclonal antibody in tobacco cell culture." Biochem Biophys Res Commun 345(2): 602-7.
Gleba, Y., et al. (2005). "Magnifection--a new platform for expressing recombinant vaccines in plants." Vaccine 23(17-18): 2042-8.

Golden rice "www.goldenrice.org."
Golovkin, M., et al. (2007). "Smallpox subunit vaccine produced in Planta confers protection in mice." Proc Natl Acad Sci U S A 104(16): 6864-9.
Gomez, N., et al. (2000). "Oral immunogenicity of the plant derived spike protein from swine-transmissible gastroenteritis coronavirus." Arch Virol 145(8): 1725-32.
Gomord, V., et al. (2005). "Biopharmaceutical production in plants: problems, solutions and opportunities." Trends Biotechnol 23(11): 559-65.
Greco, R., et al. (2007). "Production of recombinant HIV-1/HBV virus-like particles in Nicotiana tabacum and Arabidopsis thaliana plants for a bivalent plant-based vaccine." Vaccine 25(49): 8228-40.
Guerrero-Andrade, O., et al. (2006). "Expression of the Newcastle disease virus fusion protein in transgenic maize and immunological studies." Transgenic Res 15(4): 455-63.
Guetard, D., et al. (2008). "Immunogenicity and tolerance following HIV-1/HBV plant-based oral vaccine administration." Vaccine.
Gupta, S. K., et al. (1997). "Human immunodeficiency virus type-1 p24 sequence from an Indian strain: expression in Escherichia coli and implications in diagnostics." Gene 190(1): 27-30.
Hampton, R. Y. (2002). "ER-associated degradation in protein quality control and cellular regulation." Curr Opin Cell Biol 14(4): 476-82.
Hanke, T., et al. (2002). "Development of a DNA-MVA/HIVA vaccine for Kenya." Vaccine 20(15): 1995-8.
Hanson, B., et al. (1999). "A simple method to enrich an Agrobacterium-transformed population for plants containing only T-DNA sequences." Plant J 19(6): 727-34.
Hellens, R. P., et al. (2000). "pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation." Plant Mol Biol 42(6): 819-32.
Hernandez, M., et al. (2007). "A new highly effective anticysticercosis vaccine expressed in transgenic papaya." Vaccine 25(21): 4252-60.
Hiatt, A., et al. (1989). "Production of antibodies in transgenic plants." Nature 342(6245): 76-8.
Higuchi, R., et al. (1988). "A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions." Nucleic Acids Res 16(15): 7351-67.
Hilpert, K., et al. (1999). "Interaction of the capsid protein p24 (HIV-1) with sequence-derived peptides: influence on p24 dimerization." Virology 254(1): 6-10.
Ho S.N., H. H. D., Horton R.M., Pullen J.K., and Pease L.R. (1989). "Site-directed mutagenesis by overlap extension using the polymerase chain reaction." Gene 77: 51-59.
Ho, S. N., Hunt, H.D., Horton, R.M., Pullen, J.K., and Pease, L.R. (1989). "Site-directed mutagenesis by overlap extension using the polymerase chain reaction." Gene 77: 51-59.
Hongmin, Q., et al. (2000). "Effect of UTRs from TMV-RNA on the expression of foreign gene in transgenic plants." Chinese Science Bulletin 45(18).
Hood, E. E., et al. (1997). "Commercial production of avidin from transgenic maize: characterization of transformant, production, processing, extraction and purification." Mol Breed 3: 291-306.
Horn, M. E., et al. (2004). "Plant molecular farming: systems and products." Plant Cell Rep 22(10): 711-20.
Huang, L. K., et al. (2006). "Expression of avian reovirus sigmaC protein in transgenic plants." J Virol Methods 134(1-2): 217-22.
Huang, Y., et al. (2005). "Immunogenicity of the epitope of the foot-and-mouth disease virus fused with a hepatitis B core protein as expressed in transgenic tobacco." Viral Immunol 18(4): 668-77.
Huang, Z., et al. (2001). "Plant-derived measles virus hemagglutinin protein induces neutralizing antibodies in mice." Vaccine 19(15-16): 2163-71.
Huang, Z., et al. (2005). "Virus-like particle expression and assembly in plants: hepatitis B and Norwalk viruses." Vaccine 23(15): 1851-8.
Hwang, Y. T., et al. (2004). "Novel targeting signals mediate the sorting of different isoforms of the tail-anchored membrane protein cytochrome b5 to either endoplasmic reticulum or mitochondria." Plant Cell 16(11): 3002-19.
IAVI "www.iavi.org."
Im, E. J., et al. (2007). "Vaccine platform for prevention of tuberculosis and mother-to-child transmission of human immunodeficiency virus type 1 through breastfeeding." J Virol 81(17): 9408-18.
Irons, S. L., et al. (2008). "Fluorescent protein fusions to a human immunodeficiency virus monoclonal antibody reveal its intracellular transport through the plant endomembrane system." Plant Biotechnol J.
Jaffé, E. R. and D. E. Hultquist (1995). "Cytochrome b5 reductase deficiency and enzymatic hereditary methemoglobinemia." The Metabolic and Molecular Bases of Inherited Disease 7th ed.: 3399–3415.
Jani, D., et al. (2004). "Studies on the immunogenic potential of plant-expressed cholera toxin B subunit." Plant Cell Rep 22(7): 471-7.
Jansens, A. and I. Braakman (2003). "Pulse-chase labeling techniques for the analysis of protein maturation and degradation." Methods Mol Biol 232: 133-45.
Jensen, L. G., et al. (1996). "Transgenic barley expressing a protein-engineered, thermostable (1,3-1,4)-beta-glucanase during germination." Proc Natl Acad Sci U S A 93(8): 3487-91.
Jiang, W. Z., et al. (2005). "Expression and characterization of Gag protein of HIV-1(CN) in Pichia pastoris." J Virol Methods 123(1): 35-40.
Jin, C., et al. (2008). "A plant-derived human monoclonal antibody induces an anti-carbohydrate immune response in rabbits." Glycobiology 18(3): 235-41.
Joh, L. D., et al. (2005). "High-level transient expression of recombinant protein in lettuce." Biotechnol Bioeng 91(7): 861-71.
Johansen, L. K. and J. C. Carrington (2001). "Silencing on the spot. Induction and suppression of RNA silencing in the Agrobacterium-mediated transient expression system." Plant Physiol 126(3): 930-8.
Joshua, M. N., et al. (2000). "Comparison of the biological activities of human immunodeficiency virus 1 P24 and GP41 expressed in Spodoptera frugiperda cells by use of bac-to-bac system." Acta Virol 44(3): 125-30.
Joung, Y. H., et al. (2004). "Expression of the hepatitis B surface S and preS2 antigens in tubers of Solanum tuberosum." Plant Cell Rep 22(12): 925-30.
Kamenarova, K., et al. (2005). "Molecular farming in plants: An approach of agricultural biotechnology." J Cell Molec Biol 4: 77-86.
Kapila, J., et al. (1997). "An Agrobacterium-mediated transient gene expression system for intact leaves." Plant Sci 122(1): 101-108.
Kapusta, J., et al. (1999). "A plant-derived edible vaccine against hepatitis B virus." Faseb J 13(13): 1796-9.
Karasev, A. V., et al. (2005). "Plant based HIV-1 vaccine candidate: Tat protein produced in spinach." Vaccine 23(15): 1875-80.
Karnoup, A. S., et al. (2005). "O-linked glycosylation in maize-expressed human IgA1." Glycobiology 15(10): 965-81.

Kay, R., et al. (1987). "Duplication of CaMV 35S Promoter Sequences Creates a Strong Enhancer for Plant Genes." Science 236(4806): 1299-1302.
Khoudi, H., et al. (1999). "Production of a diagnostic monoclonal antibody in perennial alfalfa plants." Biotechnol Bioeng 64(2): 135-43.
Kim, T. G., et al. (2004). "Synthesis and assembly of SIVmac Gag p27 capsid protein cholera toxin B subunit fusion protein in transgenic potato." Mol Biotechnol 28(1): 33-40.
Kim, T. G. and W. H. Langridge (2004). "Synthesis of an HIV-1 Tat transduction domain-rotavirus enterotoxin fusion protein in transgenic potato." Plant Cell Rep 22(6): 382-7.
Kim, T. G., et al. (2004). "SIVmac Gag p27 capsid protein gene expression in potato." Protein Expr Purif 36(2): 312-7.
Kirk, D. D. and S. R. Webb (2005). "The next 15 years: taking plant-made vaccines beyond proof of concept." Immunol Cell Biol 83(3): 248-56.
Kogan, M. J., et al. (2004). "Exploring the interaction of the surfactant N-terminal domain of gamma-Zein with soybean phosphatidylcholine liposomes." Biopolymers 73(2): 258-68.
Koncz, C. a. S., J (1986). "The promoter TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of Agrobacterium binary vector." Mol.Gen.Genet. 284: 383-396.
Korsmeyer, S. J., et al. (2000). "Pro-apoptotic cascade activates BID, which oligomerizes BAK or BAX into pores that result in the release of cytochrome c." Cell Death Differ 7(12): 1166-73.
Kozak, M. (1986). "Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes." Cell 44(2): 283-92.
Kumar, G. B., et al. (2005). "Expression of hepatitis B surface antigen in transgenic banana plants." Planta 222(3): 484-93.
Kuroda, R., et al. (1998). "Charged amino acids at the carboxyl-terminal portions determine the intracellular locations of two isoforms of cytochrome b5." J Biol Chem 273(47): 31097-102.
Kusnadi, A. R., et al. (1998). "Processing of transgenic corn seed and its effect on the recovery of recombinant beta-glucuronidase." Biotechnol Bioeng 60(1): 44-52.
Kusnadi, A. R., et al. (1997). "Production of recombinant proteins in transgenic plants: Practical considerations." Biotech Bioeng 56: 473-484.
Lamphear, B. J., et al. (2004). "A corn-based delivery system for animal vaccines: an oral transmissible gastroenteritis virus vaccine boosts lactogenic immunity in swine." Vaccine 22(19): 2420-4.
Lenzi, P., et al. (2008). "Translational fusion of chloroplast-expressed human papillomavirus type 16 L1 capsid protein enhances antigen accumulation in transplastomic tobacco." Transgenic Res.
Li, H. Y., et al. (2006). "Accumulation of recombinant SARS-CoV spike protein in plant cytosol and chloroplasts indicate potential for development of plant-derived oral vaccines." Exp Biol Med (Maywood) 231(8): 1346-52.
Li, J. T., et al. (2006). "Immunogenicity of a plant-derived edible rotavirus subunit vaccine transformed over fifty generations." Virology 356(1-2): 171-8.
Li, X., et al. (1993). "Rice prolamine protein body biogenesis: a BiP-mediated process." Science 262(5136): 1054-6.
Li, Y., et al. (2006). "High expression of foot-and-mouth disease virus structural protein VP1 in tobacco chloroplasts." Plant Cell Rep 25(4): 329-33.
Lord, J. M., et al. (2000). "Endoplasmic reticulum-associated protein degradation." Semin Cell Dev Biol 11(3): 159-64.
Lou, X. M., et al. (2007). "Expression of the human hepatitis B virus large surface antigen gene in transgenic tomato plants." Clin Vaccine Immunol 14(4): 464-9.
Loza-Rubio, E., et al. (2008). "Development of an edible rabies vaccine in maize using the Vnukovo strain." Dev Biol (Basel) 131: 477-82.
Ma, J. K., et al. (2005). "Molecular farming for new drugs and vaccines. Current perspectives on the production of pharmaceuticals in transgenic plants." EMBO Rep 6(7): 593-9.
Ma, J. K., et al. (2003). "The production of recombinant pharmaceutical proteins in plants." Nat Rev Genet 4(10): 794-805.
Maggio, C., et al. (2007). "Intracellular sorting of the tail-anchored protein cytochrome b5 in plants: a comparative study using different isoforms from rabbit and Arabidopsis." J Exp Bot 58(6): 1365-79.
Mahboudi, F., et al. (2006). "A serological screening assay of human immunodeficiency virus type 1 antibodies based on recombinant protein p24-gp41 as a fusion protein expressed in Escherichia coli." J Biotechnol 125(2): 295-303.
Mainieri, D., et al. (2004). "Zeolin. A new recombinant storage protein constructed using maize gamma-zein and bean phaseolin." Plant Physiol 136(3): 3447-56.
Marquet-Blouin, E., et al. (2003). "Neutralizing immunogenicity of transgenic carrot (Daucus carota L.)-derived measles virus hemagglutinin." Plant Mol Biol 51(4): 459-69.
Martineau, B., et al. (1994). "On Defining T-DNA." Plant Cell 6(8): 1032-1033.
Marusic, C., et al. (2007). "Expression, intracellular targeting and purification of HIV Nef variants in tobacco cells." BMC Biotechnol 7: 12.
Massa, S., et al. (2007). "Anti-cancer activity of plant-produced HPV16 E7 vaccine." Vaccine 25(16): 3018-21.
Matlack, K. E., et al. (1999). "BiP acts as a molecular ratchet during posttranslational transport of prepro-alpha factor across the ER membrane." Cell 97(5): 553-64.
Matsuo, K., et al. (1992). "Highly conserved epitope domain in major core protein p24 is structurally similar among human, simian and feline immunodeficiency viruses." J Gen Virol 73 (Pt 9): 2445-50.
McCormick, A. A., et al. (2008). "Plant-produced idiotype vaccines for the treatment of non-Hodgkin's lymphoma: safety and immunogenicity in a phase I clinical study." Proc Natl Acad Sci U S A 105(29): 10131-6.
McGarvey, P. B., et al. (1995). "Expression of the rabies virus glycoprotein in transgenic tomatoes." Biotechnology (N Y) 13(13): 1484-7.
Menassa, R., et al. (2004). "Spider dragline silk proteins in transgenic tobacco leaves: accumulation and field production." Plant Biotechnol J 2(5): 431-8.
Menkhaus, T. J., et al. (2004). "Considerations for the recovery of recombinant proteins from plants." Biotechnol Prog 20(4): 1001-14.
Meyers, A., et al. (2008). "Expression of HIV-1 antigens in plants as potential subunit vaccines." BMC Biotechnol 8(1): 53.
Mezzasalma, T. M., et al. (2007). "Enhancing recombinant protein quality and yield by protein stability profiling." J Biomol Screen 12(3): 418-28.
Mishra, S., et al. (2006). "Ubiquitin fusion enhances cholera toxin B subunit expression in transgenic plants and the plant-expressed protein binds GM1 receptors more efficiently." J Biotechnol 127(1): 95-108.
Misumi, S., et al. (2007). "Proteomics: Biomedical and Pharmaceutical Applications. Book Chapter: Proteomics of HIV-1 Virion." 339-365.

Mitoma, J. and A. Ito (1992). "The carboxy-terminal 10 amino acid residues of cytochrome b5 are necessary for its targeting to the endoplasmic reticulum." Embo J 11(11): 4197-203.
Miyawaki, A., et al. (2003). "Lighting up cells: labelling proteins with fluorophores." Nat Cell Biol Suppl: S1-7.
Modelska, A., et al. (1998). "Immunization against rabies with plant-derived antigen." Proc Natl Acad Sci U S A 95(5): 2481-5.
Morassutti, C., et al. (2002). "Production of a recombinant antimicrobial peptide in transgenic plants using a modified VMA intein expression system." FEBS Lett 519(1-3): 141-6.
Muench, D. G., et al. (1997). "Molecular cloning, expression and subcellular localisation of a BiP homolog from rice endosperm tissue." Plant Cell Physiol 38(4): 404-12.
Munro, S. (1995). "An investigation of the role of transmembrane domains in Golgi protein retention." Embo J 14(19): 4695-704.
Müntz, K. (1998). "Deposition of storage proteins." Plant Mol Biol 38(1-2): 77-99.
Muthaura, C. N., et al. (2007). "Traditional antimalarial phytotherapy remedies used by the Kwale community of the Kenyan Coast." J Ethnopharmacol 114(3): 377-86.
Nandi, S., et al. (2005). "Process development and economic evaluation of recombinant human lactoferrin expressed in rice grain." Transgenic Res 14(3): 237-49.
Navarre, C., et al. (2006). "Expression and secretion of recombinant outer-surface protein A from the Lyme disease agent, Borrelia burgdorferi, in Nicotiana tabacum suspension cells." Transgenic Res 15(3): 325-35.
Neuhaus, J. M., et al. (1991). "A short C-terminal sequence is necessary and sufficient for the targeting of chitinases to the plant vacuole." Proc Natl Acad Sci U S A 88(22): 10362-6.
Nicholson, L., et al. (2005). "A recombinant multimeric immunoglobulin expressed in rice shows assembly-dependent subcellular localisation in endosperm cells." Plant Biotechnol J 3(1): 115-27.
Niepel, M. and D. R. Gallie (1999). "Identification and characterization of the functional elements within the tobacco etch virus 5' leader required for cap-independent translation." J Virol 73(11): 9080-8.
Nuttall, J., et al. (2005). "A functional antibody lacking N-linked glycans is efficiently folded, assembled and secreted by tobacco mesophyll protoplasts." Plant Biotechnol J 3(5): 497-504.
Obregon, P., et al. (2006). "HIV-1 p24-immunoglobulin fusion molecule: a new strategy for plant-based protein production." Plant Biotechnol J 4(2): 195-207.
Odell, J. T., et al. (1985). "Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter." Nature 313(6005): 810-2.
Ogg, G. S., et al. (1998). "Four novel cytotoxic T-lymphocyte epitopes in the highly conserved major homology region of HIV-1 Gag, restricted through B*4402, B*1801, A*2601, B*70 (B*1509)." Aids 12(12): 1561-3.
Oparka, K. J. and N. Harris (1982). "Rice protein-body formation: all types are initiated by dilation of the endoplasmic reticulum." Planta 154(2): 184-188.
Oszvald, M., et al. (2008). "Expression of Cholera Toxin B Subunit in Transgenic Rice Endosperm." Mol Biotechnol.
Oszvald, M., et al. (2007). "Expression of a synthetic neutralizing epitope of porcine epidemic diarrhea virus fused with synthetic B subunit of Escherichia coli heat labile enterotoxin in rice endosperm." Mol Biotechnol 35(3): 215-23.
Parmenter, D. L., et al. (1995). "Production of biologically active hirudin in plant seeds using oleosin partitioning." Plant Mol Biol 29(6): 1167-80.
Pedrazzini, E., et al. (1994). "Binding of BiP to an assembly-defective protein in plant cells." Plant J 5: 103-110.
Pedrazzini, E., et al. (1996). "A mutant cytochrome b5 with a lengthened membrane anchor escapes from the endoplasmic reticulum and reaches the plasma membrane." Proc Natl Acad Sci U S A 93(9): 4207-12.
Pedrazzini, E., et al. (2000). "Mechanism of residence of cytochrome b(5), a tail-anchored protein, in the endoplasmic reticulum." J Cell Biol 148(5): 899-914.
Perez-Filgueira, D. M., et al. (2004). "Preserved antigenicity of HIV-1 p24 produced and purified in high yields from plants inoculated with a tobacco mosaic virus (TMV)-derived vector." J Virol Methods 121(2): 201-8.
Perez Filgueira, D. M., et al. (2004). "Passive protection to bovine rotavirus (BRV) infection induced by a BRV VP8* produced in plants using a TMV-based vector." Arch Virol 149(12): 2337-48.
Perler, F. B. (2005). "Protein splicing mechanisms and applications." IUBMB Life 57(7): 469-76.
Petruccelli, S., et al. (2006). "A KDEL-tagged monoclonal antibody is efficiently retained in the endoplasmic reticulum in leaves, but is both partially secreted and sorted to protein storage vacuoles in seeds." Plant Biotechnol J 4(5): 511-27.
Pimpl, P., et al. (2006). "Golgi-mediated vacuolar sorting of the endoplasmic reticulum chaperone BiP may play an active role in quality control within the secretory pathway." Plant Cell 18(1): 198-211.
Porceddu, A., et al. (1999). "Transgenic plants expressing human glutamic acid decarboxylase (GAD65), a major autoantigen in insulin-dependent diabetes mellitus." Mol Breed 5: 553-560.
Porta, C. and G. P. Lomonossoff (1996). "Use of viral replicons for the expression of genes in plants." Mol Biotechnol 5(3): 209-21.
Porte, M. and S. Chong (2008). "Single-process expression and purification of multiple recombinant proteins through cocultivation and affinity purification." Anal Biochem.
Prat, S., et al. (1985). "Nucleic acid (cDNA) and amino acid sequences of the maize endosperm protein glutelin-2." Nucleic Acids Res 13(5): 1493-504.
Qian, B., et al. (2008). "Immunogenicity of recombinant hepatitis B virus surface antigen fused with preS1 epitopes expressed in rice seeds." Transgenic Res 17(4): 621-31.
Qoronfleh, M. W. (1999). "Dissolved oxygen concentration affects the accumulation of HIV-1 recombinant proteins in Escherichia coli." Appl Biochem Biotechnol 80(2): 107-20.
Rademacher, T., et al. (2008). "Recombinant antibody 2G12 produced in maize endosperm efficiently neutralizes HIV-1 and contains predominantly single-GlcNAc N-glycans." Plant Biotechnol J 6(2): 189-201.
Raleigh, E. A., et al. (1989). "Genetic and physical mapping of the mcrA (rglA) and mcrB (rglB) loci of Escherichia coli K-12." Genetics 122(2): 279-96.
Ramessar, K., et al. (2008). "Cost-effective production of a vaginal protein microbicide to prevent HIV transmission." Proc Natl Acad Sci U S A 105(10): 3727-32.
Ramirez, N., et al. (2003). "Expression and characterization of an anti-(hepatitis B surface antigen) glycosylated mouse antibody in transgenic tobacco (Nicotiana tabacum) plants and its use in the immunopurification of its target antigen." Biotechnol Appl Biochem 38(Pt 3): 223-30.
Ramirez, Y. J., et al. (2007). "Fruit-specific expression of the human immunodeficiency virus type 1 tat gene in tomato plants and its immunogenic potential in mice." Clin Vaccine Immunol 14(6): 685-92.
Richter, L. J., et al. (2000). "Production of hepatitis B surface antigen in transgenic plants for oral immunization." Nat Biotechnol 18(11): 1167-71.

Ritzenthaler, C., et al. (2002). "Grapevine fanleaf virus replication occurs on endoplasmic reticulum-derived membranes." J Virol 76(17): 8808-19.
Robinson, D. G., et al. (2005). "Protein sorting to the storage vacuoles of plants: a critical appraisal." Traffic 6(8): 615-25.
Roe, S. (2001). "Protein purification techniques: a practical approach."
Rosales-Mendoza, S., et al. (2007). "Expression of Escherichia coli heat-labile enterotoxin b subunit (LTB) in carrot (Daucus carota L.)." Plant Cell Rep 26(7): 969-76.
Rosales-Mendoza, S., et al. (2008). "Ingestion of transgenic carrots expressing the Escherichia coli heat-labile enterotoxin B subunit protects mice against cholera toxin challenge." Plant Cell Rep 27(1): 79-84.
Ruhlman, T., et al. (2007). "Expression of cholera toxin B-proinsulin fusion protein in lettuce and tobacco chloroplasts--oral administration protects against development of insulitis in non-obese diabetic mice." Plant Biotechnol J 5(4): 495-510.
Sack, M., et al. (2007). "Functional analysis of the broadly neutralizing human anti-HIV-1 antibody 2F5 produced in transgenic BY-2 suspension cultures." Faseb J 21(8): 1655-64.
Saiki, R. K., et al. (1988). "Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase." Science 239(4839): 487-91.
Salmon, V., et al. (1998). "Production of human lactoferrin in transgenic tobacco plants." Protein Expr Purif 13(1): 127-35.
Sambrook, J., and Russell, D.W. (2001). "Molecular Cloning a Laboratory Manual." (3rd Edition).
Sambrook, J., Fritsch, E.E., and Maniatis T. (1989). "Molecular cloning: A laboratory manual." (New York: Cold Spring Harbor Laboratory Press).
Sanger, F., et al. (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A 74(12): 5463-7.
Santi, L., et al. (2008). "An efficient plant viral expression system generating orally immunogenic Norwalk virus-like particles." Vaccine 26(15): 1846-54.
Schahs, M., et al. (2007). "Production of a monoclonal antibody in plants with a humanized N-glycosylation pattern." Plant Biotechnol J 5(5): 657-63.
Schillberg, S., et al. (2003). "Molecular farming of recombinant antibodies in plants." Cell Mol Life Sci 60(3): 433-45.
Schillberg, S., et al. (1999). "Apoplastic and cytosolic expression of full-size antibodies and antibody fragments in Nicotiana tabacum." Transgenic Res 8(4): 255-63.
Schunmann, P. H. D., et al. (2002). "Biopharming the SimpliRED™ HIV diagnostic reagent in barley, potato and tobacco." Mol Breed 9: 113-121.
Sheen, J. (2001). "Signal transduction in maize and Arabidopsis mesophyll protoplasts." Plant Physiol 127(4): 1466-75.
Shin, E. A., et al. (2006). "Synthesis and assembly of an adjuvanted Porphyromonas gingivalis fimbrial antigen fusion protein in plants." Protein Expr Purif 47(1): 99-109.
Shirabe, K., et al. (1995). "A novel point mutation in a 3' splice site of the NADH-cytochrome b5 reductase gene results in immunologically undetectable enzyme and impaired NADH-dependent ascorbate regeneration in cultured fibroblasts of a patient with type II hereditary methemoglobinemia." Am J Hum Genet 57(2): 302-10.
Shulga, N. Y., et al. (2004). "Expression and characterization of hepatitis B surface antigen in transgenic potato plants." Biochemistry (Mosc) 69(10): 1158-64.
Song, R., et al. (2006). "Enhancing efficacy of HIV gag DNA vaccine by local delivery of GM-CSF in murine and macaque models." J Interferon Cytokine Res 26(6): 380-9.
Sparkes, I. A., et al. (2006). "Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants." Nat Protoc 1(4): 2019-25.
Sparrow, P. A., et al. (2007). "Pharma-Planta: road testing the developing regulatory guidelines for plant-made pharmaceuticals." Transgenic Res 16(2): 147-61.
Spear, E. D. and D. T. Ng (2003). "Stress tolerance of misfolded carboxypeptidase Y requires maintenance of protein trafficking and degradative pathways." Mol Biol Cell 14(7): 2756-67.
Sprong, H., et al. (2001). "How proteins move lipids and lipids move proteins." Nat Rev Mol Cell Biol 2(7): 504-13.
Stoger, E., et al. (2002). "Practical considerations for pharmaceutical antibody production in different crop systems." Mol Breed 9(3): 149-158.
Stoger, E., et al. (2000). "Cereal crops as viable production and storage systems for pharmaceutical scFv antibodies." Plant Mol Biol 42(4): 583-90.
Strasser, R., et al. (2008). "Generation of glyco-engineered Nicotiana benthamiana for the production of monoclonal antibodies with a homogeneous human-like N-glycan structure." Plant Biotechnol J 6(4): 392-402.
Streatfield, S. J., et al. (2003). "Corn as a production system for human and animal vaccines." Vaccine 21(7-8): 812-5.
Sun, H. J., et al. (2006). "Functional expression of the taste-modifying protein, miraculin, in transgenic lettuce." FEBS Lett 580(2): 620-6.
Svab, Z. and P. Maliga (1993). "High-frequency plastid transformation in tobacco by selection for a chimeric aadA gene." Proc Natl Acad Sci U S A 90(3): 913-7.
Tabe, L. M., et al. (1995). "A biotechnological approach to improving the nutritive value of alfalfa." J Anim Sci 73(9): 2752-9.
Tackaberry, E. S., et al. (2003). "Increased yield of heterologous viral glycoprotein in the seeds of homozygous transgenic tobacco plants cultivated underground." Genome 46(3): 521-6.
Tacket, C. O., et al. (2004). "Immunogenicity of recombinant LT-B delivered orally to humans in transgenic corn." Vaccine 22(31-32): 4385-9.
Tamura, K., et al. (2004). "Endoplasmic reticulum-resident proteins are constitutively transported to vacuoles for degradation." Plant J 39(3): 393-402.
Taylor, C. B. (1997). "Comprehending cosuppression." Plant Cell 9: 1245-1249.
Tindall, K. R. and T. A. Kunkel (1988). "Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase." Biochemistry 27(16): 6008-13.
Tsuboi, T., et al. (2008). "Wheat germ cell-free system-based production of malaria proteins for discovery of novel vaccine candidates." Infect Immun 76(4): 1702-8.
Turner, B. G. and M. F. Summers (1999). "Structural biology of HIV." J Mol Biol 285(1): 1-32.
Twyman, R. M., et al. (2003). "Molecular farming in plants: host systems and expression technology." Trends Biotechnol 21(12): 570-8.
Ullah, A. H., et al. (2002). "Cloned and expressed fungal phyA gene in alfalfa produces a stable phytase." Biochem Biophys Res Commun 290(4): 1343-8.
UNAIDS "www.unaids.org."
Vain, P., et al. (1996). "Intron-mediated enhancement of gene expression in maize (Zea mays L.) and bluegrass (Poa pratensis L.)." Plant Cell Rep 15(7): 489-494.

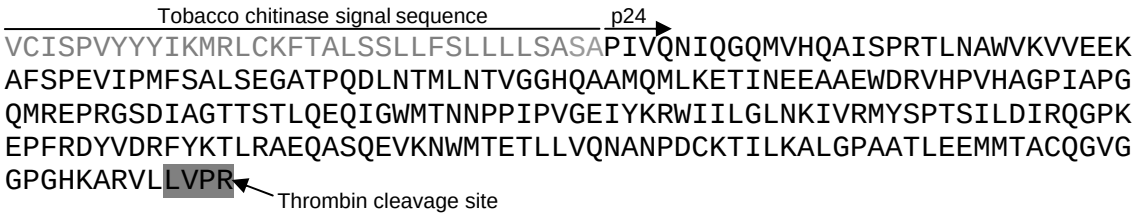
Van Droogenbroeck, B., et al. (2007). "Aberrant localisation and underglycosylation of highly accumulating single-chain Fv-Fc antibodies in transgenic Arabidopsis seeds." Proc Natl Acad Sci U S A 104(4): 1430-5.
Vidugiriene, J. and A. K. Menon (1994). "The GPI anchor of cell-surface proteins is synthesized on the cytoplasmic face of the endoplasmic reticulum." J Cell Biol 127(2): 333-41.
Vitale, A. and R. S. Boston (2008). "Endoplasmic Reticulum Quality Control and the Unfolded Protein Response: Insights from Plants." Traffic.
Vitale, A. and A. Ceriotti (2004). "Protein quality control mechanisms and protein storage in the endoplasmic reticulum. A conflict of interests?" Plant Physiol 136(3): 3420-6.
Vitale, A. and G. Hinz (2005). "Sorting of proteins to storage vacuoles: how many mechanisms?" Trends Plant Sci 10(7): 316-23.
Vitale, A. and E. Pedrazzini (2005). "Recombinant pharmaceuticals from plants: the plant endomembrane system as bioreactor." Mol Interv 5(4): 216-25.
Vitale, A., et al. (1982). "Reduced soluble proteins associated with maize endosperm protein bodies." J Exp Bot 33: 439-448.
Vlasuk, G. P., et al. (1989). "Purification and characterization of human immunodeficiency virus (HIV) core precursor (p55) expressed in Saccharomyces cerevisiae." J Biol Chem 264(20): 12106-12.
Wang, L. J., et al. (2001). "The expression of Mycobacterium tuberculosis MPT64 protein in carrots." Acta Bot Sin 43: 132-137.
Webster, D. E., et al. (2006). "Measles virus hemagglutinin protein expressed in transgenic lettuce induces neutralising antibodies in mice following mucosal vaccination." Vaccine 24(17): 3538-44.
Wen, S. X., et al. (2006). "A plant-based oral vaccine to protect against systemic intoxication by Shiga toxin type 2." Proc Natl Acad Sci U S A 103(18): 7082-7.
Wigdorovitz, A., et al. (1999). "Induction of a protective antibody response to foot and mouth disease virus in mice following oral or parenteral immunization with alfalfa transgenic plants expressing the viral structural protein VP1." Virology 255(2): 347-53.
Wigdorovitz, A., et al. (2004). "Protective lactogenic immunity conferred by an edible peptide vaccine to bovine rotavirus produced in transgenic plants." J Gen Virol 85(Pt 7): 1825-32.
Wigdorovitz, A., et al. (1999). "Protection of mice against challenge with foot and mouth disease virus (FMDV) by immunization with foliar extracts from plants infected with recombinant tobacco mosaic virus expressing the FMDV structural protein VP1." Virology 264(1): 85-91.
Wilkinson, B. and H. F. Gilbert (2004). "Protein disulfide isomerase." Biochim Biophys Acta 1699(1-2): 35-44.
Wink, M. (1988). "Plant breeding - Importance of plant secondary metabolites for protection against pathogens and herbivores." Theor. Appl. Genet. 75: 225-233.
Witcher, D. R., et al. (1998). "Commercial production of b-glucuronidase (GUS): a model system for the production of proteins in plants." Mol Breed 4: 301-312.
Woodard, S. L., et al. (2003). "Maize (Zea mays)-derived bovine trypsin: characterization of the first large-scale, commercial protein product from transgenic plants." Biotechnol Appl Biochem 38(Pt 2): 123-30.
Woodcock, D. M., et al. (1989). "Quantitative evaluation of Escherichia coli host strains for tolerance to cytosine methylation in plasmid and phage recombinants." Nucleic Acids Res 17(9): 3469-78.
Wu, L., et al. (2003). "Expression of foot-and-mouth disease virus epitopes in tobacco by a tobacco mosaic virus-based vector." Vaccine 21(27-30): 4390-8.
Wu, Y. Z., et al. (2003). "Oral immunization with rotavirus VP7 expressed in transgenic potatoes induced high titers of mucosal neutralizing IgA." Virology 313(2): 337-42.
Xin, K. Q., et al. (2007). "Chimeric adenovirus 5/35 vector containing the clade C HIV gag gene induces a cross-reactive immune response against HIV." Vaccine 25(19): 3809-15.
Xu, F., et al. (2003). "Immunogenicity of an HIV-1 gag DNA vaccine carried by attenuated Shigella." Vaccine 21(7-8): 644-8.
Yabal, M., et al. (2003). "Translocation of the C terminus of a tail-anchored protein across the endoplasmic reticulum membrane in yeast mutants defective in signal peptide-driven translocation." J Biol Chem 278(5): 3489-96.
Yang, Z. Q., et al. (2007). "Expression of the fusion glycoprotein of Newcastle disease virus in transgenic rice and its immunogenicity in mice." Vaccine 25(4): 591-8.
Yang, Z. Q., et al. (2005). "[Expression and immunization testing of fusion protein of Newcastle disease virus in leaf tissue of transgenic rice]." Yi Chuan Xue Bao 32(12): 1305-11.
Yasuda, H., et al. (2005). "Expression of the small peptide GLP-1 in transgenic plants." Transgenic Res 14(5): 677-84.
Yusibov, V., et al. (2002). "Expression in plants and immunogenicity of plant virus-based experimental rabies vaccine." Vaccine 20(25-26): 3155-64.
Yusibov, V., et al. (1997). "Antigens produced in plants by infection with chimeric plant viruses immunize against rabies virus and HIV-1." Proc Natl Acad Sci U S A 94(11): 5784-8.
Zeitlin, L., et al. (1999). "Using monoclonal antibodies to prevent mucosal transmission of epidemic infectious diseases." Emerg Infect Dis 5(1): 54-64.
Zelada, A. M., et al. (2006). "Expression of tuberculosis antigen ESAT-6 in Nicotiana tabacum using a potato virus X-based vector." Tuberculosis (Edinb) 86(3-4): 263-7.
Zhang, G., et al. (2000). "In planta expression of HIV-1 p24 protein using an RNA plant virus-based expression vector." Mol Biotechnol 14(2): 99-107.
Zhang, G. G., et al. (2002). "Production of HIV-1 p24 protein in transgenic tobacco plants." Mol Biotechnol 20(2): 131-6.
Zhao, Y. and R. W. Hammond (2005). "Development of a candidate vaccine for Newcastle disease virus by epitope display in the Cucumber mosaic virus capsid protein." Biotechnol Lett 27(6): 375-82.
Zou, Z., et al. (2003). "The stem-loop region of the tobacco psbA 5'UTR is an important determinant of mRNA stability and translation efficiency." Mol Genet Genomics 269(3): 340-9.
Zuo, J. and N. H. Chua (2000). "Chemical-inducible systems for regulated expression of plant genes." Curr Opin Biotechnol 11(2): 146-51.

APPENDIX

List of primers used for the different PCR

#	Primer	Primer sequence in 5' – 3' orientation	Site removed	Site introduced
1	**RFP1**	TT**GCTAGCGCCATGG**CCTCCTCC	*Hin*dIII	*Nhe*I/*Nco*I
2	**RFP2**	CCCTGCAAGACGGCGAGTTCATCT	*Pst*I	-
3	**RevRFP2**	TCGCCGTCTTGCAGGG	*Pst*I	-
4	**RevRFP**	AA**CTGCAG**TTA**GTCGAC**CCGGGT	-	*Pst*I/*Sal*I
5	**NewTM**	CAAGCGCGCTTGCGGGGCTGGTCCTCAT	*Nhe*I	-
6	**M13**	GTAAAACGACGGCCAGTG	-	-
7	**NewP1P24**	**GTCGACgctagc**GCCCCTATAGTGCAGAACAT	*Xho*I	*Sal*I, *Nhe*I
8	**P2P24**	GCTGAATGGGATAGAGTGCATCCAGTTCATG	*Pst*I, *Sph*I	-
9	**RevP2P24**	CATGAACTGGATG-CACTCTATCCCATTCAGC	*Pst*I, *Sph*I	-
10	**P3P24**	AGCAGGCTTCACAGG	*Hin*dIII	-
11	**RevP3P24**	CCTGTGAAGCCTGCT	*Hin*dIII	-
12	**P4P24**	**CCATGGCTCCTCTTGGAACAAG**CAAAACTCTTGCCTTATG	*Eco*RI	*Nco*I, thrombin
13	**P7P24**	CAATATAGCCCTACC**A**GCATTCTGGAC	Point mutation	
14	**Rev2P7P24**	CTTAATGC**T**GGTAGGGCTATACATTCTTACT	Point mutation	
15	**P5P24**	CAA**CTCGAG**tc**ttgttccaagagg**ACCTA	*Nhe*I	*Xho*I, thrombin
16	**RevP5P24**	CAA**GTCGAC**CCCATGGCTCCTCTTG	-	*Sal*I
17	**Z1**	CAA**GGATCCgtttgcatttcaccagtttactactacattaaa**ATGAGGGTGTT GCTCGT	*Bam*HI internal	*Bam*HI, 5'UTR tob. chit.
18	**RevZ1**	CAA**CCATGG**GCTGGCACGGGCTT	*Pst*I	*Nco*I
19	**G3F**	CAA**CCATGG**CTGGAGGTGGGGGAT	*Bam*HI	*Nco*I
20	**RevG3**	CAA**GTCGAC**TGAGCCACCTCCG	-	*Sal*I
21	**NewP5P24**	CAA**CTCGAGcttgttccaagagg**ACCTA	*Nhe*I	XhoI, thrombin
22	**RevP6P24**	CAA**CTGCAGtta**CAAAACTCTTGCCTTATG	Thrombin	Stop, *Pst*I
23	**Z3**	CAA**GTCGAC**GGCGGCTGCGG	Signal seq zein	*Sal*I
24	**RevZ3**	CAA**CTGCAGtta**CTGGCACGGGCTT	-	Stop, *Pst*I
25	**Lp24F**	CAA**GGATCC**CATGCCTATAGTGCAGAACATC	*Nhe*I, *Sal*I	*Bam*HI
26	**Lp24Rev**	CAA**GTCGAC**TCCTCTTGGAACAAGC	*Nco*I	*Sal*I
27	**M13Rev**	CAGGAAACAGCTATGACC	-	-
28	**NeweGFPF**	CAA**GCTAGC**CCATGGTGAGCAAGG		*Nhe*I
29	**eGFPRev**	CAA**GTCGAC**CCCTTGTACAGCTCGTCCA	*Xba*I	*Sal*I
30	**Forp24RFPcyt**	CAA**GGATCC**ATGCCTATAGTGCAGAACATCCA	*Nhe*I	*Bam*HI, ATG
31	**NewRevp24cyt**	CAA**CTGCAG**TTAAGCATAATCAGGAACATCATATGGATAGC CG	*Sal*I	*Pst*I, stop
32	**ForRFPp24cyt**	CAA**GGATCC**ACCATGGCCTCCTCC	*Nhe*I	*Bam*HI, ATG
33	**RevRFPp24cyt**	CAA**CTGCAG**TTATCCTCTTGGAACAAGC	*Sal*I	*Pst*I, stop
34	**Revp24SP**			
35	**RevRFPSP**			
36	**GFPTMF**	CAA**AGATCT**ATGGTGAGCAAGGGC	*Nco*I	*Bgl*II
37	**GFPTMRev**	CAA**GCATGC**TTACTTGTACAGCTCGTCC	*Xba*I	*Sph*I

HIV-1 p24 sequence:



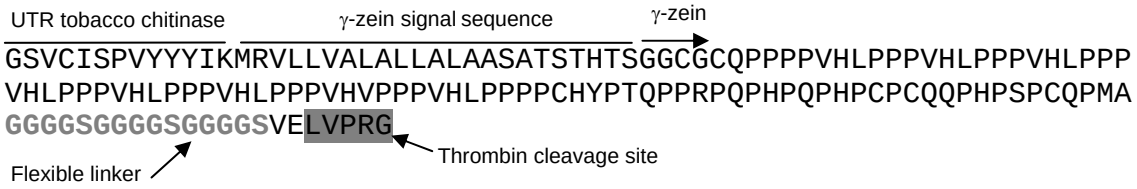
mRFP sequence:

```
MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQ
FQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTN
FPSDGPVMQKKTMGWEASTERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTTYMAKKPVQLPGA
YKTDIKLDITSHNEDYTIVEQYERAEGRHSTGAPG
```

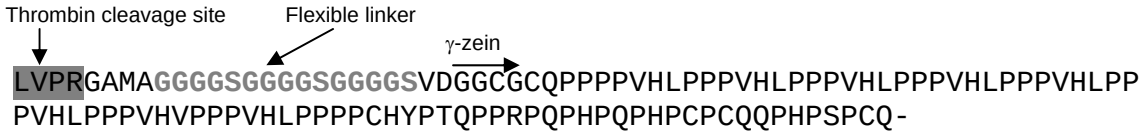
eGFP sequence:

```
MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTT
LTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKG
IDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDG
PVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGST
```

N-terminal γ-zein domain sequence in zein-p24 construct:



N-terminal γ-zein domain in p24-zein construct:



Cyt b5 tail anhor in p24-TA construct: