UNIVERSITÀ DEGLI STUDI DELLA TUSCIA DI VITERBO

DIPARTIMENTO DI AGROBIOLOGIA E AGROCHIMICA

CORSO DI DOTTORATO DI RICERCA

IN

BIOTECNOLOGIE VEGETALI - XX CICLO

CHIMERIC OIL BODIES AS AN ALTERNATIVE DELIVERY SYSTEM FOR PEPTIDE VACCINES

Settore scientifico-disciplinare:AGR/07 e BIO/04

Coordinatore: Prof.ssa Stefania Masci

Tutor: Dott.ssa Selene Baschieri
(ENEA C.R. Casaccia)

Tutor: Dott. Eugenio Benvenuto
(ENEA C.R. Casaccia)

Dottorando: Dott.ssa Floriana Capuano

'Who are you*?' said the Caterpillar.*

This was not an encouraging opening for a conversation. Alice replied, rather shyly, 'I--I hardly know, sir, just at present-- at least I know who I WAS when I got up this morning, but I think I must have been changed several times since then'.

Lewis Carroll-
Alice's adventures in
Wonderland-

A mia madre Angela e a mio padre Luigi.

PUBLICATIONS

Capuano F, Beaudoin F, Napier JA, Shewry PR.
Properties and exploitation of oleosins.
Biotechnol. Adv. 25: 203-206, 2007.

Lico C, **Capuano F**, Renzone G, Donini M, Marusic C, Scaloni A, Benvenuto E, Baschieri S.
Peptide display on Potato virus X: molecular features of the coat protein-fused peptide affecting cell-to-cell and phloem movement of chimeric virus particles.
J. Gen. Virol. 87: 3103-3112, 2006.

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Vaccines and the immune system

Prevention of infectious diseases by vaccination is one of the greatest medical achievements for mankind. During the past century, the average human life span has increased by approximately 30 years, a significant portion of which has been attributed directly to vaccination (Cooper and O'Reilly, 2002).The ability to elicit broadly neutralizing antibody and long-lasting immune responses against pathogens is a crucial goal for prophylactic and therapeutic vaccines. Prophylactic vaccination consists in eliciting immune responses able to prevent infections, while therapeutic vaccination is aimed to overcome a pre-existing pathology. Protection can be achieved by delivering antibodies (passive immunisation) or by training the immune system to actively face the pathogen (active immunisation). Today, despite the awareness that many diseases caused by viruses and bacteria can be prevented by vaccination, infections are still a major cause of morbidity and mortality. The need to develop new or improved vaccines is pointed out by the increase in the frequency of antibiotic-resistant bacteria and the risk of re-emergence of eradicated pathogens, due to the threat of bioterrorism.

1.1.1 A short outline of pathogen-biased immune responses

In the past 200 years vaccine manufacturing has undergone many changes to obtain safer and more effective formulations (Ulmer et al., 2006). A better understanding of the molecular basis of immune responses has started a new era in vaccine research as the basic goal of vaccination is to mimick an infection in order to stimulate a complete immune response (innate and adaptive, antibody- and cell -mediated) without inducing morbidity.

The first contact with a pathogen/antigen triggers the innate immune response. Innate immunity is a primitive mechanism of defence that involves cells endowed with phagocytic and cytotoxic activity (i.e. polymorphonucleate granulocytes, mastocytes, Natural Killer cells, B lymphocytes, dendritic cells (DC) and macrophages) situated close to the physical barriers of the organism (mainly epithelia and mucosae) (Medzhitov and Janoway, 1997). The activation of innate immunity is induced by the recognition by these cells through the so called Toll-like Receptors (TLR) of Pathogen-Associated Molecular Patterns (PAMP), small molecular motifs consistently conserved among similar pathogens (i.e lypopolysaccharide,

peptidoglycan, DNA CpG motifs, flagellin) (Krishnan et al., 2007). In parallel to innate responses, pathogens activate specific long-lasting adaptive responses mediated by cells (B and T lymphocytes) expressing on their membranes highly specific antigen receptors. These receptors mainly bind to proteins or peptides (Garcia et al., 1999). The antigen-specific receptors of B lymphocytes are immunoglobulins. These composite glycoproteins block pathogens usually by interacting with portions of folded proteins (conformational epitopes - groups of amino acids of the antigen not arranged as a straight chain in the primary structure-). The antigen-immunoglobulin interaction activates the B cell and induces its proliferation and differentiation in a plasma cell, able to secrete immunoglobulins. As an alternative, the B cell can differentiate into a memory B cell able to persists for a long time in the lymph and blood streams, ensuring a quicker response in the case of a second pathogen attack (immunological memory).

The antigen-specific receptors of T lymphocytes (T cell receptor, TCR) are structurally similar to immunoglobulins, but exist only in the membrane-associated form (Bentley and Mariuzza, 1996). Differently from immunoglobulins, TCR recognizes antigens only after they have been processed by specialized cells (Antigen Presenting Cells, APC) or by pathogen-infected cells. The TCR recognizes a complex made of an antigen-derived peptide (linear epitope –group of amino acids arranged as a straight chain in the primary structure-) associated to a Major Histocompatibility Complex (MHC) molecule on the plasma membrane of these cells. APC take up antigen from the extracellular environment (for this reason named exogenous antigen), process and present derived fragments in association to MHC class II molecules. T lymphocytes that recognize epitopes associated to this class of MHC are defined as helper T lymphocytes, characterized by the presence of the co-receptor CD4 on the plasma membrane. These T cells show the distinctive feature of secreting soluble factors (cytokines) with regulatory functions on the immune response. The CD4+ helper T lymphocytes can be further divided in T helper 1 (Th1) and T helper 2 (Th2) on the basis of the pattern of cytokines they secrete.When epitopes are presented by cells that have been infected by the pathogen (endogenous antigens), they associate to MHC class I molecules. T lymphocytes that recognize epitopes associated to this class of MHC molecules are cytotoxic T lymphocytes (CTL), phenotypically characterized by the presence of the co-receptor CD8 on the plasma membrane. These cells are functionally characterized by the ability to kill pathogen infected cells through the secretion of cytolytic factors (perforins and granzymes). It is now becoming clear that the MHC class I and II pathways of antigen processing and

presentation are not so clearly disconnected and that some specialized APC are able to cross-present exogenous antigens to CTL (Yewdell et al., 1999; Houde et al., 2003; Roy, 2003).

1.1.2 Factors influencing vaccine efficiency

Ideally, vaccines should elicit specific and long-lasting immune responses as much similar to those induced during pathogen infection both at systemic and mucosal level. Indeed, mucosae usually are access sites for pathogens and play a major role in protection due to the presence of a specialized lymphoid tissue compartment (MALT, Mucosal Associated Lymphoid Tissue) involved in the early steps of pathogen recognition and clearance (mucosal immunity). The activation of both systemic and mucosal compartments of the immune system by vaccination depends on several factors, but primarily on the delivery route and antigen immunogenicity. It is common notion that different compartments of the immune system are activated according to the vaccine delivery route. While delivery through the parenteral route elicit only systemic responses and is extremely effective in stimulating long-lasting immunity, delivery through mucosal routes (intranasal, oral, rectal, or vaginal) can elicit both systemic and mucosal immunity (Ryan et al., 2001; Holmgren and Czerkinsky, 2005). Nevertheless, different mucosal districts do not respond in the same way to vaccination. For example, intranasal is much more effective than oral immunisation in eliciting the activation of the different districts of the immune system, including those of distal mucosae (Mowat, 2003).

Important determinants of the immunogenicity of a given antigen are primarily the molecular dimensions, that might become crucial in activating and addressing the elicited immune response. A “big” antigen for example, may contain more conformational and linear epitopes as compared to a “small” one, and has more chances to activate specific B and T cell clones. One of the strategies commonly used to increase the efficiency of a vaccine consists in coupling the antigen to a “carrier” (usually a protein) and/or in co-administering an adjuvant. Adjuvants are substances able to enhance the *in vivo* immunogenicity of antigens by helping to simulate the signalling cascade that promote the maturation of the APC and the optimal activation of adaptive immunity. They can be divided in two main groups on the basis of the mechanism of action (Pashine et al., 2005). A first group is composed of adjuvants able to activate innate immunity, the so called “immune potentiators”. A second group includes all those substances behaving as delivery devices by displaying antigens in repetitive patterns and inducing their targeting to APC. So far, pre-clinical studies have described several diverse compounds exhibiting adjuvant properties, but only a few have been approved for human use: i) alum (aluminium salt particles) highly effective in enhancing antibody responses, through

the elicitation of Th2-biased immune responses, but not in inducing the activation of cell-mediated immunity; ii) MF59, an oil-in-water squalene emulsion; iii) AS04, an alum formulation containing the Toll-like receptor 4 (TLR4) agonist monophosphoryl lipid A (Aguilar and Rodriguez, 2007). The efficiency of new generation adjuvants, based for example on liposomes (formulated with monophosphoryl lipid A) or on immuno-stimulating complexes incorporating the antigen (ISCOMS) is under evaluation in clinical trials.

1.1.3 Vaccine types

There are different types of vaccines licensed for human use. These include vaccines containing live-attenuated or inactivated/killed organisms and few subunit vaccines (Ulmer et al., 2006).

1.1.3.1 Vaccines containing live, attenuated organisms

The success of this approach depends on decoupling virulence from induction of protective immunity by the use for vaccination of live pathogen mutants less virulent than the wild type (w.t.) counterpart. The main feature of these vaccines is their ability to induce potent immune responses because of the preserved pathogen invasiveness and ability to replicate within host cells and also to the fact that pathogen surface molecules activate innate immunity. This vaccination approach has been adopted against the infection of several bacterial and viral pathogens, such as *Vibrio cholerae*, *Mycobacterium tuberculosis*, *Salmonella typhi*, yellow fever virus, measles virus, mumps virus, poliovirus, rubella virus, varicella-zoster virus, adenovirus and rotavirus. However, the use of vaccines containing live organisms have major limitations due to the fact that it is not always possible to identify sufficiently attenuated strains and there is a finite, underlying risk of reversion to virulence.

1.1.3.2 Vaccine containing inactivated/killed organisms

Vaccines containing inactivated or killed organisms despite including all the antigens of a pathogen able to induce the activation of the immune response, are generally much less efficient than live vaccines and require adjuvants co-delivery. This is due to the fact that pathogens are killed by chemical treatments (formaldehyde, β-propiolactone), that render them unable to access host cells and replicate. In this way they are active only as exogenous antigens, hence not competent to activate the whole-spectrum of immune responses. Several vaccines such as those against influenza virus, hepatitis A virus, rabies virus, and *Bordetella pertussis* are produced by this approach.

1.1.3.3 Purified or recombinant subunit vaccines

Subunit vaccines contain microrganism-derived components selected on the basis of immunogenic properties and safety concerns. This vaccination strategy while being extremely valuable in terms of biosafety, is less advantageous in terms of effectiveness, mainly because isolated proteins by themselves are much less immunogenic than pathogens. Moreover, purified antigens may not retain their native conformation eliciting the production of antibodies that bind only weakly the native counterpart. These are the reasons why subunit vaccines require almost unavoidably the co-administration of adjuvants to be efficient. At the moment, subunit vaccines present on the market are those raised against tetanus, Hepatitis B (HBV) and Human Papillomavirus (HPV). While tetanus vaccine is produced by inactivating tetanus toxoid purified from the pathogen, HBV and HPV vaccines are manufactured by ectopically expressing the viral surface proteins in yeast and insect cells and purifying the virus like particles (VLP) generated by their self-assembly.

1.1.4 Trends in vaccinology

Despite considerable success of traditionally designed vaccines, sometimes these approaches are uneffective in preventing or treating infections induced by classical and emerging pathogens. The attempt to overcome these limitations, for example by identifying key antigens through 'reverse vaccinology' (a bioinformatics-based approach), is paving the way to next-generation vaccines (Rappuoli, 2004).

1.1.4.1 Gene-based vaccines

Gene-based vaccines have been mainly developed with the aim of eliciting the activation of CTL able to eliminate infected cells. By this approach, the sequence coding the antigen is delivered to cells in order to be expressed and further processed and presented by the immune system in association to MHC class I molecules, thus mimicking the events of a normal infection process. Gene-based vaccines can be delivered as plasmid DNA through gene gun or using recombinant viruses as vectors to enhance the potency of vaccination. Viral vectors based on adenovirus, poxviruses and alphaviruses, are more efficient in stimulating the immune response activation, probably because able to reproduce an infection-like environment. Despite their higher efficiency, manufacturing recombinant viral vectors is far more complicated than simple plasmid DNA. Moreover, pre-existing or vector-induced immunity that may limit the effectiveness of vaccination is a potential drawback.

1.1.4.2 Virus Like Particles (VLP)

VLP are particles that structurally mimic viruses, generated by self-assembly of viral capsid proteins when these are expressed in heterologous systems. VLP have several advantages over other vaccine types. They are particulate antigens and have the advantage to simulate the structure of the viral surface without carrying the viral genome. Therefore they are unable to induce infection while retaining all conformational epitopes usually targeted by neutralizing antibodies. The capsid proteins of several different viruses, including Human Immunodeficiency Virus type-1 (HIV-1), Norwalk virus and HPV, have been expressed to produce VLP that have been tested in human clinical trials. An anti-HPV VLP-based vaccine has been recently licensed for human use.

1.1.4.3 Dendritic cell-based vaccines

DC-based vaccines are individualized treatments that aim to exploit the central role of these cells in directing the adaptive immune responses against pathogens through antigen presentation. Manufacturing these vaccines implies recovery of a blood sample from an individual and isolation of DC from this sample. Cells are cultured, pulsed with the antigen of interest and finally re-injected to the original donor. At the present time, DC-based vaccines are under clinical trial to evaluate their efficacy in preventing and curing different pathologies and have been shown to induce both antigen-specific CD8+ and CD4+ T cell responses *in vivo* (Walsh et al., 2003).

1.1.4.4 Peptide vaccines

Molecular insights on the mechanisms of activation of adaptive immune responses indicate that peptides rather than whole proteins are responsible for the activation of T and B lymphocytes (even if at a lower extent). By using peptides it is possible to stimulate specific clones and bias the immune response towards an antibody or a cytotoxic response eliminating safety concerns. The selection of peptides to be included in vaccine formulations require an epitope mapping of the antigen of interest to identify immunodominant sequences. The identification of B cell activating peptides is complicated by the fact that often immunoglobulin recognize conformational epitopes. In the attempt to mimic conformational structures (particularly those of viral envelopes) cyclic peptides, branched peptides, peptomers (cross-linked peptide polymers) and other complex multimeric structures, as well as peptides conjugated to other molecules have been developed. Mix of peptides or

"mimotopes" identified in combinatorial peptide libraries that could mimic relevant structures and potentially inducing broad responses have been successfully used in some pre-clinical vaccination trials (respiratory syncytial virus, measles, hepatitis C, etc.).

When the aim is to identify peptides activating CTL, the task is complicated by the fact that immunodominant peptides vary as a function of the individual immunogenetic background (MHC haplotype). Indeed, MHC molecules are encoded by a polygenic locus and each gene, particularly in humans, is characterized by an astounding allelic diversity. The differences among the MHC class I molecules encoded by each gene and by each allele of a gene, map in the region encoding the peptide binding pocket, so that different MHC molecules bind different peptides. To overcome the problem that different individuals respond to different peptides of a same antigen, the possibility exists that a "cocktail" of peptides containing the immunodominant epitope for each haplotype can be used. The attempt to regroup the different haplotypes in a smaller number of supertypes, in a way to reduce the number of peptides to be included in the "cocktail", has been performed (Sette and Sidney, 1999). Another major obstacle limiting the usefulness of peptide vaccines is the poor immunogenicity and the requirement of appropriate carriers and/or adjuvants to get a successful activation of the immune response.

1.1.5 HIV-1 and influenza virus: open challenges to vaccine technologies

Vaccination is the best strategy to prevent infections. Nonetheless, for several pathogens the development of efficient vaccination strategies has not yet succeeded. Among these pathogens are influenza virus and HIV-1. Indeed, despite the evident differences in their biology, starting from the fact that influenza causes an acute, while HIV-1, a chronic infection, both these pathogens are always "one step ahead" host neutralizing responses (Karlsson Hedestam et al., 2008). Although vaccination against influenza virus is effective as a single predominant variant circulates globally each year, the development of a vaccine against pandemic influenza represents a challenge because of the high number of variants deriving from the animal reservoir (usually birds) potentially able to generate the pandemic viral strain. This high variability somehow parallels the challenge of inducing broadly neutralizing antibody responses against HIV-1 by vaccination. However, the persistence of influenza virus occurs in the entire human population thanks to the infection of individuals with no or low protective immunity against new variants, while HIV-1 generates an enormous number of variants in each infected individual and these are simultaneously seeded and circulate in the population. Besides, HIV-1 has evolved many immune escape mechanisms that together with its high

variability represent the biggest challenge for vaccine development (Kwong et al., 2002; Wey et al., 2003). Because the main target of the neutralizing antibody response against both viruses are envelope proteins, the attempt to develop protective vaccines are mainly focused on these antigens and on the identification of conserved neutralizing elements within them. The precursor of the components of influenza virus envelope, haemagglutinin 0 (HA0), after translocation to the Endoplasmic Reticulum (ER) and trimerization, is targeted to the plasma membrane through the Golgi and is cleaved in HA1 and HA2 on released virus particles (Doherty et al., 2006). HIV-1 envelope (Env) is also composed by two glycoproteins, gp120 and gp41 but, differently from influenza virus envelope components, they are both generated by proteolytic cleavage of a heavily glycosylated precursor (gp160) into the Golgi and the subsequent targeting to the plasma membrane for particles budding (Frankel and Young, 1998). HA1 mediates the binding to the terminal sialic acid of glycoproteins and glycolipids on host cell surface while HA2 induce fusion. However, HA2 needs a low pH environment to get the permissive functional conformation and for this reason the fusion to the membrane occurs once the virus has been endocytosed by the target cell. Similarly, HIV-1 gp120 confers tropism towards CD4+ T lymphocytes while gp41 (the transmembrane domain of Env) mediates viral fusion to target cells.

1.1.5.1 Current approaches to elicit protection against influenza

To develop influenza vaccines based on the production of neutralizing antibodies against the viral envelope, the "predict and produce" approach is mainly applied at present. By this approach, all variants, most likely to spread across the globe, are identified and propagated in embryonated hen eggs. After detergent inactivation trivalent vaccines consisting of two subtypes of influenza A virus (H1N1 and H3N2) and one influenza B virus go into the market for intramuscular delivery. At present the main limitation of this commercialized influenza vaccine formulation is the narrow neutralizing activity of the elicited antibodies. One of the possible solutions to this problem and to the need of improving the efficiency of a pandemic influenza vaccine would be to enlarge the repertoire of cross-neutralizing antibodies by identifying conserved and exposed antigens among virus variants. It should also be envisaged to co-deliver adjuvants able not only to provide dose-sparing effects, but also to expand sub-dominant B and T cell clones, otherwise remaining below the threshold of activation. Recently, a live-attenuated cold-adapted influenza vaccine based on viruses able to replicate efficiently at 25°C in the nasal passages but not at the high temperature occurring in the deeper respiratory tract, has also been licensed in USA. This vaccine, when administered

intranasally, induces a protection superior than that offered by the inactivated vaccine, possibly because the replicating virus express antigens longer and activates more compartments of the immune system. The identification of conserved antigens/epitopes is a complementing approach to the classic "predict and produce" methods for the development of a protective influenza vaccine as cross-neutralizing antigens could provide broader protection even to avian flu. In fact, it is extremely difficult to predict which flu strain might acquire the ability to infect and disseminate throughout the human population seasonally. Moreover, the existence of the avian flu reservoir represents a big challenge for vaccine development as it drives the "antigen shift" that has been responsible of pandemics in the past. Recently, a comprehensive analysis of all the epitopes of influenza A proteins described in the literature has been undertaken to inventory data available and to identify novel epitopes of interest for vaccine development (Bui et al., 2007). According to this analysis, the proteins forming the envelope (HA, Neuraminidase (NA), Matrix) are the most attractive as the antibody response seems to be crucial to control viremia. NA is interesting also because partially shared between human and avian variants, possibly providing cross-protection to viruses derived from the avian flu reservoir. Although the variability of the sequence of the epitopes recognized by antibodies is higher if compared to that of T-cell epitopes, appreciable conservation has been observed.

Another problem relevant to influenza vaccine development concerns manufacturing. In fact, classic production systems based on the use of hen eggs are not valuable for the propagation of avian flu viral strains and alternative procedures must be envisaged.

1.1.5.2 Current approaches to elicit protection against HIV-1

As for the development of a vaccine inducing protective immunity against HIV-1 through the elicitation of neutralizing antibodies, experimental evidences obtained by analysing serum antibodies of infected individuals indicate that only a handful of sites on HIV-1 Env are conserved and vulnerable to antibody-mediated neutralization. These observations are confirmed by the fact that very few human monoclonal antibodies are able to neutralize a broad range of primary isolates of HIV-1 (Moore et al., 2001). The paucity of neutralizing antibodies elicited against HIV depends on the high mutation rate and on the shielding of the viral surface through large N-linked glycans of host origin, able to render Env immunologically silent. The monoclonal antibodies b12 (Zhou et al., 2007) and 2G12 (Trkola et al., 1996), isolated in HIV-1 infected individuals, show specificity for different regions of gp120: b12 recognize the CD4 receptor binding site while 2G12 binds to a cluster of

conserved glycans on the outer domain, representing an exception to the masking role of glycans on HIV-1 particles. Two more neutralizing antibodies specific for the transmembrane glycoprotein gp41 (2F5 and 4E10; Muster et al., 1993; Stiegler et al., 2001) and a few others are under investigation to design effective immunogens (Zolla-Pazner et al., 2007). The epitopes recognized by 2F5 and 4E10 monoclonal antibodies are conserved and linear and this renders the region in which they are included on gp41 (the membrane proximal external region, MPER) an attractive target for the development of anti-HIV-1 vaccines (Zwick et al., 2005). Due to the specificities of the described antibodies, many efforts aimed to develop an effective vaccine are focalized on the production of a form of gp120 conformationally closer to the natural counterpart therefore able to elicit similar antibody responses. Because the protein is extremely flexible and useless to elicit protective responses as a single isolated subunit, design efforts have been shifted to the engineering of soluble versions of oligomeric Env (containing full-length gp120 covalently linked to different versions of gp41 ectodomain) to form spikes resembling in their properties to the functional trimer. Recent data obtained by immunization with constructs encoding this complex (the gp140 molecule) showed a response with a wider cross-neutralizing spectrum (Phogat and Wyatt, 2007). Similarly, the attempt to structurally stabilize gp120 has been approached through the complexing to CD4 receptor, with the idea to obtain the release of the immunogenic site recognized by the b12 antibody. As b12 recognizes the CD4 binding site (BS), this epitope must be conserved, therefore representing a good candidate vaccine target. In the attempt to focus the antibody response on the CD4 BS, gp120 has been mutagenized by the introduction of new glycosylation sites, with the aim of focusing the immune response towards the region recognized by b12. It has been shown that, by this strategy, the binding of non-neutralizing and poorly neutralizing antibodies is inhibited, while leaving unmodified the broadly neutralizing properties of b12 antibody (Pantophlet et al., 2003). Recently, the attempt to stabilize gp41-derived peptides through their insertion into a lipid context has also been described and a number of trials are ongoing to evaluate B- and T-cell responses against MPER-derived immunogens arrayed on carrier molecules (HB surface antigen) (Phogat et al., 2007) and heterologous lipid membrane. Approaches to develop vaccines based on the use of chemically inactivated virions or on the use of VLP are also under development (McBurney et al., 2006).

1.1.5.3 Influenza and HIV vaccines: a possible role of CTL

Since pathogen infection activate CTL responses beside eliciting neutralizing antibodies, the research in the field of vaccinology is trying to reproduce such a whole spectrum response. As

for the development of anti-HIV-1 vaccines, the elicitation of the CTL response is thought to be crucial for the clearance of the viral reservoir (infected cells) responsible for the chronic infection. Several studies have demonstrated that CTL play a central role in controlling HIV viremia in infected individuals. The appearance of HIV-specific CD8+ T cells is concomitant with the suppression of viral load during primary infection (Borrow et al., 1994; Koup et al., 1994) and the loss of HIV-specific CD8+ T-cell activity is often associated with rapid progression to Acquired Immunodeficiency Syndrome (AIDS) (Klein et al., 1995). For the development of vaccines aiming to elicit cytotoxic response, several targets are under investigation. Nef (Negative factor)-, Vpr (Viral protein r)-, RT (Reverse Transcriptase)-, Tat (transactivator)- and Gag (Group-specific antigen)-based peptide vaccines have been tested in various forms. Deeper insights onto the mechanisms of CTL activation have demonstrated that just few peptides are responsible for virus clearance. These peptides are generally known as "immunodominant". However, recent studies have reassessed the importance of CTL responses also against "subdominant" peptides (Zhong et al., 2003). These peptides, being not submitted to the immune response pressure, are less subjected to the onset of mutations and are therefore more conserved, ensuring the elimination of virus variants escaping CTL surveillance, as in the case of HIV-1. Moreover, evidences from studies on the murine model seem to indicate that the response to some immunodominant epitopes can be detrimental and sometimes does not correlate with virus clearance (Crowe et al., 2006). In general, it is assumed that T cell epitopes selected naturally by the immune system are good candidates for peptide-based vaccines. However, some of these epitopes may have limited efficacy in terms of anti-viral control and may even lead to delayed viral clearance. Two major problems have to be faced when trying to formulate efficient peptide-based vaccines able to elicit CTL responses. The first is that both immunodominant and subdominant peptides are different among individuals as a function of their haplotypes. The second is that the delivery system must be able to target peptides on MHC class I molecules. Very good results have been obtained using complex, conjugated peptides that include multiple CTL and helper epitopes (Alexander et al., 2002).

1.2 Plants as vaccine factories

Nowadays, subunit vaccines may be obtained either by purification from pathogen's cultures or by expression in heterologous systems. A wide range of systems (bacteria, yeast, insect, mammalian) have been successfully developed to this aim (Baneyx and Mujacic, 2004;

Gerngross, 2004; Wurm, 2004; Kost et al., 2005). Bacteria and fungi represent robust production systems while mammalian cells are the favourite as in their environment even complex molecules such as antibodies fold properly and undergo post-translational modifications essential for their activity and their immunological and biochemical properties (glycosylation, methylation, acetylation, phophorylation, myroystilation). However, because yields and quality can be low due to product complexity, costs can be high due to the expensive and sensitive culture conditions and contamination from pathogens represents a sizeable risk for human health, it is sensible to consider alternative heterologous expression systems. Among the possible alternatives, plants have been considered as an efficient system to guarantee availability of bio-active, safe recombinant pharmaceuticals, in particular vaccine antigens that can be manufactured in a cost-effective manner (Ma et al., 2003). In fact, despite the major emphasis as to the relevance of transgenic plants was initially given to the possibility to confer traits of agronomic importance, such as disease resistance (Vaeck et al., 1987; Broglie et al., 1991; Tavladoraki et al., 1993), in the past decade plants have also emerged as "factories" for the commercial production of valuable recombinant proteins and vaccines. The application of plants for the production of therapeutic proteins offers several advantages such as absence of mammalian pathogens, cost effectiveness, large-scale production and relative ease in expression and purification. In spite of the evident advantages of using plants as biofactories, this production platform has some limitations. In particular, the different glycosylation pattern as compared to mammals (Chrispeels and Faye, 1996). This is of particular importance for the production of human therapeutics, as many human proteins are glycosylated and their efficacy depends on their intrinsic biochemical features. Strategies aimed at addressing plant glycosylation have been developed to express proteins similar or identical to the cognate molecule (Cox et al., 2004; Bakker et al., 2006).

1.2.1 Methods for antigen production in plants

1.2.1.1 Stable transformation

Different strategies and techniques have been deployed to construct stably transformed (transgenic) plants. Stable nuclear transformation is generally achieved by *Agrobacterium*-mediated gene transfer (Graves and Goldman, 1986; Feldmann and Marks, 1987; Clough and Bent, 1998) and sometimes also by the less efficient direct-gene delivery methods (e.g., particle bombardment, electroporation) used especially to transform plant species recalcitrant

to *Agrobacterium*-mediated gene introduction (such as many monocotyledonous species) (Birch, 1997). Several antigens have been expressed in plants by this method (Tab. 1).

Because expression levels obtained through nuclear transformation are not always satisfactory, due to interfering factors at different stages of the procedure, some basic shrewdness can be taken. The first forethought to prevent problems due to intrinsic features of the heterologous sequence consists in codon usage optimization and removal of cryptic signals that can negatively affect post-translational steps (Sullivan and Green, 1993; Koziel et al., 1996). Among basic interventions, the selection of suitable 5'- and 3'- regulatory elements able to ensure reproducible high-level and tissue-specific transgene expression is also important. In fact, it has been demonstrated that the activity of nominally constitutive promoters can vary as a function of the developmental stage, organ and plant species (Williamson et al., 1989; Malik et al., 2002; Samac et al., 2004).

In the attempt to boost transgene expression in plants, viral promoters (Guilley et al., 1982), tissue-specific promoters (De Jaeger et al., 2002) and promoters of highly transcribed genes (Outchkourov et al., 2003) have been used. In a similar direction, efforts are oriented in attempting to adjust gene insertion into the genome by homologous recombination in order to induce the integration of the heterologous gene into a pre-determined genomic location in a single copy (Hanin and Paszowski, 2003; Srivastava and Ow, 2004). It has indeed been reported that the integration of several direct and inverted copies of the heterologous gene (De Neve et al., 1997; Tzfira et al., 2004), a fairly common event, can induce the accidental inactivation of genes fundamental for plant metabolism but, above all, transgene silencing (Hobbs et al., 1993; Jorgensen et al., 1996; Muskens et al., 2000; De Buck et al., 2001). Although single-copy transgenes can induce silencing as well (Elmayan and Vaucheret, 1996; Day et al., 2000), they generally have a uniform and stable expression (De Buck et al., 2004).

In the attempt to enhance production efficiency of transgenic lines, plants stably encoding replicating viral vectors, the so-called replicons or amplicons, have also been developed, expressing uniformly and synchronously a foreign gene, thanks to the efficiency of viral replicases (Palmer et al., 1999; Mallory et al., 2002; Gleba et al., 2004; Zhang and Mason, 2005).

Alternative to nuclear transformation is the integration of the gene of interest into the plastid genome to produce stably genetically modified plants (Daniell, 2006). By this approach the foreign gene is inserted by homologous recombination into a specific site of the plastid genome, eliminating the "position effect" occurring in nuclear transformation. Despite the high levels of transgene transcription, no gene silencing has been described in plastid

transformed plant lines, and differences in the levels of expression occur only as a consequence of the activity of chloroplast proteases. Although this transformation technique is somehow laborious due to difficulties in obtaining homoplasmic plants (where the integration of the transgene occurs in every plastid genome), chloroplasts remain ideal "bioreactors" although closer to prokaryotic expression systems. Chloroplast genomes defy the laws of Mendelian inheritance in that they are maternally inherited and therefore minimize the outcross of transgenes via pollen, resulting in an environmentally friendly approach for the production of recombinant proteins. By the construction of transplastomic plants, having up to 10 000 plastid genomes per leaf cell, very high levels of expression have been obtained (Tregoning et al., 2003).

As to the production of vaccines, a wishful thinking of plant biotechnologists is to use stably transformed plants not only as a measure to reduce production costs and improve product safety and stability, but also as a means to efficiently deliver vaccines at mucosal level, eliminating the purification steps and the use of needles (Kapusta et al., 1999; Tacket et al., 2000; Tacket et al., 2004; Thanavala et al., 2005). Unfortunately, this dream is still far to come true mainly because antigen expression in plants through nuclear stable transformation still suffer limitations due to the low and variable expression levels of the heterologous gene among independent lines. This is a major drawback when direct oral delivery of the plant tissue is foreseen as this impairs a precise *a priori* definition of the vaccine dose. From this point of view transplastomic plants offer more realistic chances of success. Preliminary data have been recently published on the oral delivery of *Nicotiana tabacum* transplastomic tissues expressing two plague–derived antigens. However, data indicate that this oral boost, following a subcutaneous immunization, is efficacious in eliciting a systemic antibody response but induces only low IgA amounts at mucosal level (Arlen et al., 2008). Moreover, chloroplast transformation techniques of edible crops still need to be improved (Kanamoto et al., 2003; Zhou et al., 2008).

Table 1. Representative antigens stably expressed in plants

Pathogen	Antigen	Plant platform	References
Hepatitis B Virus	Surface antigen	*Lactuca sativa* *Lupinus luteus* *Nicotiana tabacum* *Solanum tuberosum*	Mason et al., 1992; Thanavala et al., 1995 ; Ehsani et al., 1997 ; **Kapusta et al., 1999**; Richter et al., 2000 ; **Thanavala et al. 2005**
Norwalk Virus	Capsid protein	*Nicotiana tabacum* *Solanum tuberosum*	Mason et al., 1996; **Tacket et al., 2000**
Enterotoxic *E. coli*	Heat-labile enterotoxin B subunit Fimbrial subunit protein FaeG	*Nicotiana tabacum* *Solanum tuberosum* *Zea mays*	Haq et al., 1995; Mason et al., 1998 ; **Tacket et al., 1998**; **Tacket et al., 2004;** Joensuu et al., 2004
Rabies Virus	Glycoprotein	*Lycopersicon esculentum*	McGarvey et al., 1995
Cytomegalovirus	Glycoprotein B	*Nicotiana tabacum*	Tackaberry et al., 1999
Vibrio cholerae	Cholera toxin B subunit	*Solanum tuberosum* *Oryza sativa*	Arakawa et al., 1997; Arakawa et al.,1998; Oszvald et al., 2008
Foot-and-mouth disease virus	VP1 Structural protein	*Arabidopsis thaliana*	Carrillo et al., 1998
Porcine transmissible gastroenteritis coronavirus	Glycoprotein S	*Arabidopsis thaliana* *Nicotiana tabacum* *Zea mays*	Gomez et al., 1998; Tuboly et al., 2000; Streatfield et al., 2001
Mannheimia haemolytica	Mannheimia haemolytica GS60	*Medicago sativa*	Lee et al., 2008
Rabbit haemorrhagic disease virus	VP60 Structural protein	*Solanum tuberosum*	Martìn-Alonso et al., 2003
Canine parvovirus	VP2 protein	*Arabidopsis thaliana*	Gil et al., 2007
SARS-CoV	spike protein (S1)	*Nicotiana tabacum* *Lactuca sativa* *Lycopersicon esculentum*	Li et al., 2006; Pogrebnyak et al.,2005
HIV-1	Nef	*Nicotiana tabacum*	Marusic et al., 2007; Barbante et al., 2008; de Virgilio et al., 2008

In **bold characters** are indicated references describing clinical trials of the plant-expressed antigens.

1.2.1.2 Transient antigen production in plants

As an alternative to stable genetic transformation, plant viruses, mainly ss(+)RNA viruses (e.g. Tobacco Mosaic Virus (TMV), Cowpea Mosaic Virus (CPMV), Tomato Bushy Stunt Virus (TBSV), Plum Pox Virus (PPV), Potato Virus X (PVX), Alfalfa mosaic virus (AlMV)), (Pogue et al., 2000) or *Agrobacterium tumefaciens* (Kapila et al., 1997), have been exploited as vectors for transfer heterologous sequences destined to transient expression in plants. By this approach the transgene is not transmitted from the transformed plant to the progeny and, for this reason, repeated infection/transformation cycles are necessary. Nonetheless, the main advantage of this strategy of gene delivery to plant cells is that high yields of the product of interest can be obtained in relatively short periods of time. Several methods have been developed to transiently express heterologous proteins in plants using viral vectors. The foreign sequence can be integrated into the viral genome as an additional coding sequence or used to replace a viral gene (Canizares et al, 2005). A further strategy has been developed with the goal of using plant viruses not only as vehicles of the heterologous sequence but also as carriers of the encoded antigen in vaccine formulations from plants. This approach consists in inserting into the viral genome the sequence of interest so that in infected plant tissues chimeric virus particles are produced displaying on their surface as fusion to coat protein (CP) subunits the encoded polypeptide (Johnson et al., 1997). These modified virions that can be produced in large quantities using plants as biofactories, are attractive because, due to their structure, they can be used to safely and efficiently deliver antigens, with a significant increase in immunogenicity. According to the structure of the virus particles, some restrictions are imposed as to the size of the polypeptide, in that large sequences can create steric hindrance and interfere with virus particle assembly. The site of the insertion must be also selected on a structure-based analysis. For example, TMV tolerates peptide insertion at the C-terminus of the CP of a maximum of 23 amino acids. In addition to that, factors other than length have been identified as interfering with chimeric virus particles production/assembly (Bendhamane et al., 1999; Porta et al., 2003; Lico et al., 2006). When used in animal models, plant-derived chimeric virus particles have been demonstrated to confer antibody mediated immunity against a number of diseases. In most of the cases, this was obtained by parenteral immunization, but there are encouraging indications that also mucosal immunization is effective (Marusic et al., 2001). Recently, some data have been produced indicating that chimeric plant virus like particles produced in *E. coli* or plant virus particles chemically modified through the *in vitro* linking to antigens, can be effective in activating cytotoxic T cell immune responses (McCormick et al., 2006; Lacasse et al., 2008).

As for transient expression by Agro-infiltration, despite this technique has been mainly used to predict expression efficiency of constructs before nuclear stable transformation, it is currently considered also as an efficient means for large scale production. Recently, Icon Genetics Inc. has developed a system (magnICON) for the over-expression of foreign genes that conjugates the efficiency of *Agrobacterium*-mediated transformation with viral rate of expression. (Marillonet et al., 2005) To pursue this goal, TMV genome has been optimized, by the removal of cryptic sequences that, if recognized by the plant machineries, could affect viral replication/spreading, and splitted into 5' and 3' modules. The 5' module includes the viral polymerase and the movement proteins genes while the 3' module carries the gene of interest in place of the viral CP gene. The two modules, inserted into binary vectors are used to transform *A. tumefaciens*, that are then used to infiltrate plant tissues together with a third strain of *A. tumefaciens* carrying a vector encoding a recombinase. When the three different sequences are transferred into plant cells, the recombinase recognizes specific target-sequences artificially added to each module inducing their recombination, hence the expression of the heterologous gene. This "deconstructed" virus system is unable to spread systemically as TMV needs the CP to move throughout the plant and in the environment. Moreover, it is not influenced by the dimension of the inserted gene due to the absence of the packaging process, while it drives the plant cell machinery to the production of heterologous protein. Using this strategy, high yields of several *Yersinia pestis* antigens have been expressed in *Nicotiana benthamiana* plants, and used to efficiently induce protective immune responses against the pathogen, when administered subcutaneously to guinea pigs (Santi et al., 2006).

1.2.2 Considerations on antigen purification from plant tissues

Plants are a cheap and reliable system for the production of recombinant molecules. Nonetheless, the downstream processing of plant material may involve expensive and time-consuming methods, heavily affecting the final cost of good (Menkhaus et al., 2004). The intended final use of the recombinant product dictates special processing requirements. For example, pharmaceutical applications, including vaccine formulations, require an accurate downstream processing with a final product purity of at least 95-98%. This could change in the future if the transformed plant tissues will be used not only as a production but also as a delivery system, turning the processing minimal or unnecessary. Many research efforts are devoted to develop and improve downstream processing methods with the aim to reduce associated costs and to improve the feasibility for commercial application. In general,

processing and purification procedures consist in plant tissue harvesting and fractionation along with particle size reduction, extraction of the target protein into an aqueous medium, clarification of the crude extract and final purification (Menkhaus et al., 2004). The first issue to be considered to achieve high yield of the product of interest is the stability of the recombinant protein, keeping the notion that protein degradation does not occur only in the extraction buffer but also within cells. For this reason it is important to target the protein to proper cellular compartments as this may facilitate high yields of recovery. Different subcellular compartments may be eligible to accumulate different recombinant proteins. For example, the highest accumulation of complete antibodies has been reported by targeting into the apoplastic space while, the stability of single-chain Fv antibody fragments (scFv) is increased by the retention in the lumen of the ER (Conrad and Fiedler, 1998). Another example of the importance of protein targeting to the proper compartment for stabilization has been provided in the case of HIV-1-derived protein Nef (Marusic et al., 2007). The expression of Nef in the cytoplasm or in the ER lumen of tobacco cells showed that the targeting to the cytoplasm has a stabilizing effect on the protein that accumulates up to 0.4% Total Soluble Protein (TSP), while ER targeting and introduction into the secretory pathway has a detrimental effect leading to protein degradation. The stability of Nef in the cytoplasm has been further increased by its fusion to a sequence derived from the C5 cytochrome (Tail Anchor, TA), promoting targeting to the cytosolic side of the ER membrane (Barbante et al., 2008). By this approach Nef accumulated in tobacco cells to over 0.7% TSP. However, alternative approaches as targeting to *Nicotiana tabacum* protein bodies through zeolin fusion (de Virgilio et al., 2008) or plastid expression (Zhou et al., 2008) resulted in higher accumulation levels of Nef (1% and 40% TSP, respectively). Targeting recombinant proteins accumulation in various subcellular compartments or tissues involves the development of appropriate extraction protocols. A recent work compares the efficiency of different procedures in extracting from plant tissues a monoclonal antibody accumulating in different cell compartments (apoplast, ER or plasma membrane) (Hassan et al., 2008). The data indicate that harsher techniques (mechanical shearing or manual grinding) and presence of detergents in the extraction buffer are required to purify the membrane-bound antibody, while gentler procedures are suitable to extract the antibody from apoplast or ER. From the point of view of costs and purification, protein accumulation in the apoplast by fusion to ER targeting signals is the strategy of choice in that the scale-up would be simplified. Recombinant proteins accumulating in this extracellular space can indeed be purified without the need of tissues fractionation and by simply recovering the liquids from the extracellular spaces

through buffer infiltration. In the case of expression in cultured plant cells, apoplast targeting results in protein secretion directly into the culture medium. Secretion through the roots (rhizosecretion) transformed with secretory versions of heterologous sequences under the control of either the constitutive (CaMV) 35S, or the modified *Agrobacterium* mannopine synthase (mas2') promoter, preferentially active in roots (Ni et al., 1996) allows an easy recovery of large amounts of recombinant protein directly from the medium of hydroponically grown plants (Borisjuk et al., 1999). Several approaches have been suggested for the stabilisation of the recombinant protein either by the addition of gelatine (Drake et al., 2003) or the co-secretion of protease inhibitors (Komarnytsky et al., 2006). An advantage of this approach is that the protein can be continually collected along over the plant lifetime, reducing time, costs and labour.

The purification procedure to be followed to separate the recombinant protein from a complex plant extract is of course dictated by its targeting to a specific subcellular compartment/tissue but also by unique physicochemical properties of proteins, such as the isoelectric point or affinity to a particular ligand. Reports published on recovery and purification of recombinant products from plant tissues are typically based on chromatographic approaches that typically exploit affinity. The easiest way to render the protein affine to some molecule is to fuse the sequence to a molecular tag (Kusnadi et al., 1998; Terpe, 2003). Recombinant proteins with tags bind to columns packed with the affinity ligand so that contaminants in the crude plant extract are removed. Protein tags are generally represented by small peptides (His-, poly-Arg-, FLAG-, c-myc-, S-, and Strep II-, ELP-tag) or also by whole proteins (glutathione S-transferase, maltose-binding protein). It has been shown that the use of tags may have a positive effect on the biochemical properties of the target protein by improving yields, preventing proteolysis, facilitating refolding and increasing solubility (Arnau et al., 2005). When the recombinant protein is an antibody or an antibody fragment the purification can be performed by exploiting the high specificity of the interaction with the antigen or by using Protein A or Protein G (Chargelegue et al., 2000). Although chromatographic methods are really efficient in obtaining highly pure proteins, they usually correlate with low yield of the final product especially because to prevent column clogging preliminary purification steps are necessary. This increases the possibility to waste the protein in intermediate steps. Recently, a very efficient alternative purification method, requiring the fusion of the protein of interest to elastin-like peptides (ELP), has been used to purify an antibody expressed in *Nicotiana tabacum* plants (Floss et al., 2008). The purification of proteins fused to ELPs can indeed be achieved by a simple procedure called 'inverse transition cycling' as these peptides are highly

soluble in aqueous solutions below their transition temperature (*T*t) while collapsing, aggregating and altering protein solubility, when the temperature is raised above *T*t (Meyer and Chilkoti, 1999). Therefore, this approach may be suitable to significantly enhance the accumulation of recombinant proteins in plant (Patel et al., 2007).

1.2.3 Seed-based production platforms

Seeds are an eligible plant tissue for the large-scale expression of recombinant proteins. The main reason is that they have evolved to constitute a stable environment for the accumulation of storage proteins in a reduced volume. Recombinant protein targeting to mature seeds results in its accumulation in a sort of natural cellular 'warehouse' where it can be stockpiled for several years until recovery (Stoger at al., 2000). This provides an evident advantage over targeting to sporophytic tissues that require, after harvesting, an immediate processing or storage at very low temperature to prevent proteolysis. Heterologous protein restriction to the seed do not normally interfere with vegetative plant growth and limits adventitious contact with non-target organisms such as microbes in the biosphere and leaf-eating herbivores. Moreover, the small complexity of the seed proteome facilitates the purification and downstream processing of the protein of interest. Cereals, legumes and oilseeds have been largely investigated to be used for heterologous protein production (Stoger et al., 2005). The cereal crops are all potentially good production systems as their grains have high protein content and are produced in very large amounts each year. Among them, maize, despite being a cross-pollinating species, seems to be the most interesting for several reasons: i) it gives the highest annual grain yield; ii) it has the shortest generation interval; iii) it benefits of a relatively easy *in vitro* manipulation and transformation procedures and of several seed-specific and seed-restricted promoters that can be used to drive transgene expression. Maize has been used to produce the first two commercial plant-derived recombinant proteins, avidin and β-glucuronidase (Hood et al., 1997; Witcher et al., 1998) and represents, at the moment, a good candidate as a commercial platform for the production of pharmaceutical and technical proteins, including recombinant antibodies, vaccines and enzymes. Also rice, a self-pollinating species, is an interesting crop as protein content per grain is high. Barley and wheat have lower producer price as compared to maize and rice but it is not yet clear why they do not accumulate recombinant proteins at high levels. Legumes have also been explored as potential production platforms due to the high protein content (>40% in soybean). However, transformation procedures are time-consuming compared to maize and in some cases, such as soybean, the high oil content of the grains might interfere with downstream

processing steps. Pea has a similar annual grain yield and seed protein content to soybean. However, the producer price is about 50% higher than that of soybean.

The most obvious approach to get protein accumulation in seeds is to put the expression of the sequences under the control of one of the seed-specific promoters discovered in association to seed storage protein genes. One of the promoters most frequently used in monocots is the rice glutelin promoter Gt-1, which confers endosperm-specific expression both in rice and maize (Yang et al., 2003). Several other monocot seed-specific promoters have been identified, including the embryo-specific maize globulin-1 promoter (Belanger and Kriz, 1991), the maize endosperm-specific 27-kDa zein promoter (Russell and Fromm, 1997) and the barley D hordein (Horvath et al., 2000) and aleurone-specific promoters (Hwang et al., 2003). The expression levels achieved in monocots seed tissues using these promoters have exceeded those obtained using constitutive promoters (Hood et al., 2003). For heterologous genes expression in dicots, the novel seed-specific promoter bean arc5-I from the common bean (*Phaseolus vulgaris*) has been characterized. The expression in *Arabidopsis thaliana* of a ScFv antibody under the control of this promoter resulted in antibody levels higher than 36% of TSP in homozygous seeds (De Jaeger et al., 2002).

Another approach followed to accumulate proteins in seeds is the targeting to the protein storage organelles (i.e. the protein bodies and protein storage vacuoles), as these compartments have evolved to facilitate stable protein accumulation. Several recent studies have shown that conventional targeting strategies might not work as expected when applied in seeds. In fact, recombinant proteins tagged with the secretory pathway targeting peptide KDEL in rice endosperm were mainly found in protein bodies rather than in the apoplast (Nicholson et al., 2005), probably because, to reach the apoplast, proteins have to pass through the endomembrane system, which is generally well articulated in storage cells and "specialised" at targeting proteins to storage organelles. This abnormal behaviour has effects on the glycosylation pattern of proteins expressed in seeds (Stoger et al., 2005).

1.2.4 Organelle coupling-based purification

An innovative approach to perform a preliminary cost-effective purification of the recombinant protein from plant tissues, removing the bulk of native proteins, consists in its coupling to plant cell organelles (in particular, chloroplasts and oil bodies) that can be easily separated from the other cellular components by centrifugation on sucrose density gradients.

1.2.4.1 Chloroplast targeting

Protein coupling to chloroplasts can be obtained by the fusion with PAP-fibrillins (plastoglobulins), a family of proteins coating the low density lipoprotein particles attached to the thylakoid membranes called plastoglobules. The PAP/fibrillin family proteins are highly conserved in the plant kingdom and plastoglobules are ubiquitously found in all types of plastids, mainly in leaf chloroplasts. In leaf crops producing high biomass, plastoglobule targeting would allow accumulation of large amount of the recombinant proteins and easy downstream processing of leaf material. As for chloroplast targeting of recombinant proteins, this approach may limit the toxicity of foreign proteins to plant cells and facilitate the downstream processing. In particular, sequestration of the recombinant proteins in particles such as plastoglobules (Fig. 1) minimizes the deleterious effects on photosynthetic light and dark reactions due to the accumulation of the foreign protein into the chloroplasts.

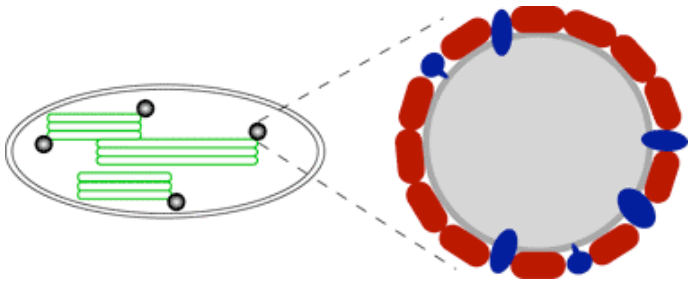

Figure 1. Plastoglobules in chloroplasts.
Schematic representation of a chloroplast (left side) highlitghing the thylakoid-associated plastoglobules lipoprotein particles (Kessler and Vidi, 2007).

The expression of the Yellow Fluorescent Protein (YFP) as fusion to these proteins demonstrated the advantages of this approach in terms of recombinant protein purification from leafy crop plants such as tobacco (Vidi et al., 2007). In fact, chloroplasts were previously fractionated from other cellular components and later plastoglobules were easily separated from thylakoid membranes and from the chloroplast envelope by floatation centrifugation on density sucrose gradient. This approach provided a 250 fold enrichment of the recombinant protein.

1.2.4.2 Oil body targeting

Oil body targeting of recombinant proteins is an oilseed-based production and recovery technology successfully developed at SemBioSys Genetics Inc. (*http://www.sembiosys.com/*) (Fig. 2).

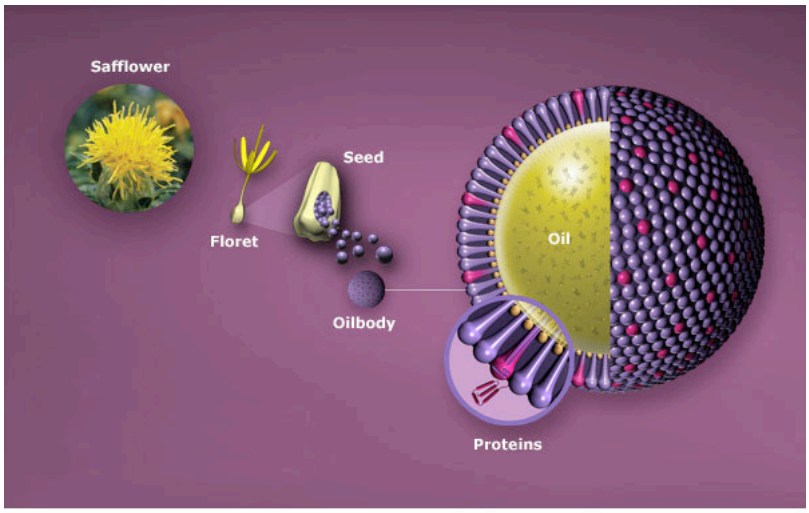

Figure 2. The oleosin-fusion technology.
Schematic representation of heterologous proteins targeting to oil bodies through oleosin sequence fusion (envisaged by SemBioSys, *http://www.sembiosys.com/*).

Owing to their physicochemical properties, oil body coupling can be exploited to simplify downstream processing of transgenic seeds accumulating recombinant proteins. In fact, oil bodies easily separate from other seed components by liquid–liquid phase separation as they tend to float. This mild process eliminates the need for refolding and can reduce the number of chromatography steps required to obtain a purified product, thereby reducing consistently the overall cost of good (Deckers et al., 2004). Exploratory studies performed by fusing *Arabidopsis* oleosin to β-glucuronidase (GUS) have shown that the enzymatic activity of the recombinant protein is preserved (van Rooijen and Moloney, 1995a). Similarly, by oleosin fusion to Green Fluroescent Protein (GFP) it has been demonstrated that the chimeric oleosin is correctly targeted to and tightly associated with oil bodies as shown in tobacco leaf cells (Wahlroos et al., 2003). Few pharmaceuticals have been produced in plants using the oleosin fusion technology. As a proof of concept, the recombinant human precursor insulin has been expressed and recovered as an active molecule from transgenic *A. thaliana* seeds (Nykiforuk

et al., 2006). When analysed by reverse phase high-performance liquid chromatography (RP-HPLC), oil body-targeted insulin showed a peak with a retention time almost identical to that of the standard insulin (yeast recombinant human insulin). The expression of insulin in transgenic seeds through the fusion to *A. thaliana* oleosin was approximately six times higher that obtained in potato tubers by the fusion to cholera toxin B (CTB) (Arakawa et al., 1998). The fusion accounted for 0.1% TSP present in the tubers, corresponding for mature insulin to 0.022% TSP. A further example is provided by the expression of hirudin in *B. napus* seeds. Hirudin is an anticoagulant peptide used to treat thrombosis, originally isolated from the medicinal leech *Hirudo medicinalis*. Its expression in plants and targeting to the oil bodies through fusion to *Arabidopsis* oleosin guarantees high levels accumulation and provide a mean for the easy purification (Parmenter et al., 1995). Both insulin and hirudin bound to oil bodies undergo a further processing to free the active molecule from the organelle. This is achieved by inserting an endoprotease cleavage site between the oleosin and the recombinant product sequences. Beside the use of recombinant protein coupling to oil bodies aimed to improve the purification procedure, in the patent of SemBioSys, the use of this method to produce innovative vaccine formulations is envisaged. This application considers the possibility of using the purified chimeric oil bodies as carriers to efficiently deliver antigens in innovative vaccine formulations. The immunogenic properties of oil bodies, mainly due to their corpuscular nature, could be potentiated owing to their lipid nature. The patent describes also different procedures that can be used to couple antigens to oil bodies surface, other than fusion to a heterologous oleosin, such as cross-linking by biotin-streptavidine interaction. Moreover, the patent describes the possibility to deliver chimeric oil bodies not only through "classical" immunization routes but also transdermally, an easy non-invasive delivery strategy that provides a prolonged stimulation of the immune system and may lead to self administration by reducing health care system costs. Results concerning this kind of potential applications have never been published up to now.

1.3 Oil body protein composition and biogenesis

Oilseeds accumulate lipids to supply the energy requirements for the growth of the seedling after germination. Such lipids are generally stored as triacylglycerols (TAG) in spherical compartments referred to as spherosomes (Frey-Wyssling et al., 1963), oleosomes (Murphy, 1990) or, most frequently, oil bodies. Oil bodies are found in the endosperm and embryo cells of oilseeds and pollen. They have variable diameters with a narrow range between 0.5 and 2.0

μm and a hydrophobic TAG core surrounded by a half-unit phospholipid (PL) membrane and an outer shell of specialized proteins (Tzen et al., 1993). Chemical analyses reveal that 1 to 4% (by weight) of seed oil bodies is composed of proteins (Tzen and Huang, 1992). Oil bodies have been analysed biochemically due to the economic importance of oilseeds. Among the oil body proteins, oleosins are the most abundant (Jolivet et al., 2004). They are relatively low molecular weight proteins of about 16-24 kDa that accumulate on the surfaces of lipid bodies in desiccation tolerant seeds. Their distribution is consistent with an active role in oil mobilization as it has been proposed they could act as binding site for lipase (Huang, 1996). Oleosins form a family with similar structural properties. They include a long hydrophobic core of about 70 residues organized around a unique 12–amino acid motif (proline knot) and two polar N- and C-terminal domains flanking the hydrophobic region (Abell et al., 1997). A high degree of similarity is located in, but not restricted to, the hydrophobic domain and proline knot motif, both of which are essential for the correct targeting to oil bodies (van Rooijen and Moloney, 1995b; Beaudoin and Napier, 2000). Sixteen oleosin genes have been characterized in the *A. thaliana* genome (Kim et al., 2002). According to their tissue specific expression, these genes have been divided into three groups: those specifically expressed in maturing seeds (5), those expressed in both maturing seeds and floral microspores (3) and those specifically expressed in the floral tapetum (8). Oleosins are responsible for the stabilisation of oil bodies, preventing these organelles from coalescing during seed desiccation. Several observations suggest that they are also involved in ensuring the maintenance of the appropriate size and surface volume/ratio to balance the conflicting needs for efficient storage and ease of TAG mobilization (Napier et al., 1996). Higher TAG-to-oleosin ratios result in larger oil bodies. To this purpose it has indeed been observed that: i) species containing higher amounts of oleosins (i.e. *Brassica napus*) have smaller oil bodies compared with those with lower oleosins content; ii) oleosins are not present in fruit oil bodies, such as avocado (*Persea americana*) and olive (*Olea europaea*) (Ross et al., 1993) typically containing larger oil bodies; iii) manipulating oleosin expression and accumulation levels modulates the size of oil bodies (Siloto et al., 2006).

How and when oleosin accumulates on oil bodies is a matter of debate. The finding that oil body biogenesis can occur in the absence of oleosins (Ross et al., 1993) supports a model in which TAG droplets surrounded by a PL monolayer accumulate in ER subdomains later acquiring their oleosin coat. An alternative model suggests that accumulation of TAG, oleosin and PL is concomitant during oil body assembly (Peng and Tzen, 1998). Besides abundant oleosins, other three minor proteins of 27, 37 and 39 kDa have been identified in sesame oil

bodies (Chen et al., 1998). The gene encoding 27 kDa protein has been cloned. Sequence analysis revealed the presence in this protein, named caleosin, of a putative calcium-binding motif (Chen et al., 1999). Similarly to oleosin structure, caleosin contains three structural domains: an N-terminal hydrophilic domain including the calcium-binding motif, a central hydrophobic anchoring domain with a proline knot and a C-terminal hydrophilic domain containing a potential phosphorylation site (Chen et al., 1999). Caleosin function is presumably modulated by calcium binding and phosphorylation state. It has been suggested that it could be involved in calcium-mediated fusion of oil bodies (Frandsen et al., 2001). Also the gene encoding the 37 kDa polypeptide has been cloned (Lin et al., 2002) revealing that the encoded protein, named steroleosin, possesses an N-terminal hydrophobic segment anchoring to the oil body surface a soluble domain homologous to sterol-binding dehydrogenases/reductases, normally involved in signal transduction in diverse organisms. Steroleosin could be involved in diverse signal transduction (Jolivet et al., 2004).

1.4 Organelle proteome analysis by mass spectrometry approach

Recently, the development of fine mass-spectrometry (MS) techniques has lead to more comprehensive studies on eukaryotic cells allowing to address many biological questions. The availability of an ever-increasing number of sequenced genomes and advances in MS technology enables acquisition of information on protein identity, abundance, subcellular localization, post-translational modifications, interactions, and ultimately, function. The aim of organelle proteomics is first to establish the protein repertoire of various subcellular compartments. At its simplest, organelle proteomics involves isolating the organelle of interest and producing a catalogue of the proteins present in that organelle by some form of separation of proteins or their proteolytic fragments followed by identification utilizing MS. Subcellular fractionation and biochemical enrichment of these proteins are required to overcome the abundance issues and detect even low abundance proteins. The majority of studies on plant organelle proteomic have been focused on the chloroplast and mitochondria, which can be isolated rather easily. However, plant research is nowadays focusing on developing procedures to purify new plant organelles and characterize them through the MS approach. Using these sensitive techniques few catalogues of proteins have been obtained for vacuoles (Jaquinod et al., 2007), plastids (Jain et al., 2008) and oil bodies (Jolivet et al., 2004). However, organelle protein catalogues may contain a high number of false positives because of the difficulty or near impossibility of isolating cellular organelles to 100%

homogeneity. Moreover, the high sensitivity of the new generation mass-spectrometer means that even low abundance contaminants may be detected. The problem with contaminants in such catalogues is that they cannot be distinguished from true residents, hence calling into question the reliability of the entire catalogue. One approach to address the problem of contaminants is known as subtractive proteomics. In this approach, a preparation of the organelle is compared against a "control" sample to correct for background contamination. Manipulating or interfering with the accumulation of a true resident protein to its subcellular localization and analysing the sample by quantitative MS, true resident proteins would modify, meanwhile contaminants would stay at same level as the control (Borner et al., 2006). By this method a set of 241 true lipid raft residents has been identified by comparing membrane protein sets derived from control and cholesterol-disrupting drugs treated samples. Membrane associated proteins decreased in treated samples, meanwhile no variations were observed for non-specific proteins (Foster et al., 2003). On the other hand, dynamic studies may be performed on organelles as proteins that are stably associated may change their distribution according to the cellular growth conditions (Andersen et al., 2005). While the subtractive approach can be useful for studying a single organelle, it still requires high level enrichment and this is not always achievable. Moreover, analysis of dynamic cell compartments as ER and Golgi apparatus and characterization of proteins trafficking between them and their budding vesicles require a more global survey. In fact, many proteins are shared between these compartments and recently, several methods have been developed to survey protein localization in multiple organelles. It has been suggested that a true resident has a characteristic distribution pattern within a gradient and that it was not restricted to a single fraction. Proteins which are co-resident in organelles will co-fractionate upon biochemical separation (De Duve, 1971). In this way, the location of a protein of unknown residency can be identified by matching to specific marker proteins. The quantitative profiling of protein levels in the different fractions is carried out and statistical analysis is used to identify co-varying proteins. The first of these methods, Protein Correlation Profiling, was initially developed to identify contaminants in a purified preparation of centrosomes (Andersen et al., 2003). Centrosomes were first purified on a sucrose gradient and peptides derived from five fractions of the gradient were analysed by consecutive Liquid Chromatography (LC)-MS runs by a label-free approach. The peak intensity of the peptides were utilised to profile the relative abundance of proteins in different fractions. The centrosomal proteins showed a different distribution profile compared to that of contaminants along the gradient. A recently developed method which works on similar principles is the

Localization of Organelle Proteins by Isotope Tagging (LOPIT). LOPIT was used to map proteins to multiple organelles in *A. thaliana*. Organelles were first partially separated on a density gradient, and gradient fractions were labeled for relative quantitation of protein levels in order to profile each protein distribution in the gradient. Stable isotope tags (Isotope-Coded Affinity Tag (ICAT) and iTRAQ reagents) were used to distinguish proteins from the different fractions (Dunkley et al., 2004; Dunkley et al., 2006). In particular the use of iTRAQ reagents for labelling increased considerably the proteome coverage. The multivariate statistical methods (Principal Components Analysis (PCA), Partial Least Squares-Discriminant Analysis (PLS-DA)) applied to this quantitative MS technique allows proteins to be clustered based on their co-variation. A given protein which shows a similar distribution to a known organelle marker clusters with it on a PCA scatter plot and is likely to reside in the same organelle. Proteins from different organelles form separate clusters. In particular, using this approach, 527 proteins have been localised confidently to the ER, Golgi apparatus, vacuolar membrane, plasma membrane, mitochondria and plastids (Dunkley et al., 2006). Localizations were confirmed by immunofluorescence microscopy of GFP-fused proteins. The increasing need of developing procedures to prepare highly purified organelles and sensitive techniques to detect either low abundant true organelle residents or contaminants find an answer in organelle proteomics investigation.

1.4.1 Mass spectrometry-based proteomics workflow

The typical workflow for an MS-based proteomics experiment involves a preliminary protein extraction step followed by a pre-fractionation phase (biochemical fractionation, affinity purification, two-dimensional or one-dimensional gel separation) aimed to reduce sample complexity. Afterwards, proteins are digested to peptides with a sequence specific protease, usually trypsin, to generate samples easier to handle and solubilize and with mass spectra easier to interpret. After the digestion peptides are separated by one or more steps of high-pressure liquid chromatography (HPLC) to be then identified. When peptide mixtures derive from less then 2 or 3 proteins, peptide mass fingerprinting (PMF) can be easily and quickly achieved without sequencing or fragmentation of peptide ions using a tandem MS approach. During PMF, peptides generated by the tryptic digest of an isolated protein are analyzed by MS giving a set of peptide masses which are unique to that protein. By matching the tryptic peptide masses to a database of proteins that has been digested *in silico*, the protein can be identified. When more complex peptide mixtures have to be analyzed, tandem MS (MS/MS) must be performed. Column eluted peptides are ionised at the ion source by the process of

ElectroSpray Ionization (ESI). At this stage, peptides ionize under the application of a strong electric potential and the co-eluting solvent evaporates. Alternatively, peptides can be ionized by matrix-assisted laser desorption ionization (MALDI), in which analyte is mixed with a solid crystalline matrix, typically an organic acid, and bombarded with a laser to bring about ionization. A mass spectrum of the peptides is recorded, followed by parent ions fragmentation in order to yield peptide sequence information. Fragmentation can be induced in several ways: by collision with an inert gas in a process known as collision-induced dissociation (CID) or by radical directed cleavage in electron transfer dissociation (ETD) or electron capture dissociation (ECD). Resulting MS/MS spectra contain fragment ions giving information on peptide sequences and allowing protein identification.

The final stage in a proteomics workflow is protein identification. With MS/MS data, it is possible to interpret spectra to obtain peptide sequences using the mass differences between consecutive fragment ions. This is known as *de novo* sequencing, but it depends highly on manual interpretation and is usually used only if the genome sequence of the study organism is not available. The most common method of identifying proteins in high-throughput studies is by database searching. The MS/MS spectra are used to generate peak lists which are searched by matching against theoretical MS/MS spectra derived from *in silico* digestion of protein databases. The MS/MS spectra are assigned to peptides with probability scores to indicate how likely the match is to be true. The identified peptides are then 'assembled' into a list of proteins that are likely to have been present in the sample. In this way, a large number of proteins can be identified from a complex sample in a single experiment. Several commercial algorithms exist for database searching, such as SEQUEST (Yates et al., 1995) and Mascot (Perkins et al., 1999).

1.4.2 Principles for Mass Spectra interpretation

In LC-MS/MS the peptide mixture is first loaded onto the reversed-phase HPLC (RP-HPLC) columns and peptides are separated according to their hydrophobicity. The most hydrophilic peptides elute first. The eluted peptides are directly coupled to the ion source and a mass spectrum of the eluting peptides is recorded (MS survey scan). The records of the intensity of the peptide ions are displayed on the spectra as related to their mass to charge ratio (m/z). Peptides are often protonated during electrospray ionization, giving rise to $(M+nH)^{n+}$ ions, where M is the mass of the peptide and n is usually 2 or 3. The MS survey scan can be used to determine the charge state of a peptide using the isotope envelope. Due to the natural abundance of ^{13}C, a peptide gives rise to several isotopic peaks consisting of no ^{13}C atoms,

one ^{13}C atom, two ^{13}C atoms and so forth. These isotopic peptides are 1 Da apart from each other, but because the mass spectrometer measures the ratio m/z, the corresponding peaks are separated by an m/z value dependent on the charge state. Hence, a doubly charged peptide will have isotope peaks 0.5 Da apart. This can therefore be used to determine the peptide charge state. Knowing the charge state and m/z value in turn allows one to calculate the mass of the peptide. Following the survey scan, ions are selected for fragmentation to yield sequence information. In CID, peptide ions of a particular m/z are isolated (this is known as the precursor ion), and bombarded with inert gas molecules which transfer their kinetic energy and induce peptide bond cleavage. The mass spectrum of the fragment or product ions is recorded and this forms the MS/MS spectrum. The selection of ions for MS/MS can be controlled by the user if there is a particular interest in certain species. This is known as Selected Ion Monitoring (SIM). More commonly though, the instrument is operated in data-dependent acquisition mode, whereby the 2 or 3 (or more) most abundant peptide species eluting at a given time are automatically selected for MS/MS. The instrument then returns to a survey scan and selects the next ions for MS/MS. Usually a dynamic exclusion is set up so that once a peptide of a certain mass has been selected and sequenced, it is excluded from re-selection to allow for other peptides to be sampled. Cycling between MS and MS/MS modes in this way is known as the duty cycle and allows one to maximize the number of peptides that are sampled, even including those of lower abundance.

1.4.3 Quantitative mass spectrometry

Recently, several high throughput methods have emerged involving quantitative strategies that allow comparative or quantitative analyses of proteomes. Quantitative data come in two forms, the absolute amount of protein in the sample or the relative change in protein amount between two states. The traditional approach to protein quantitation has been to first separate proteins by two-dimensional gel electrophoresis, stain the gels (or label the proteins with fluorophores before gel separation) to obtain a measure of spot intensity and hence protein level, then excise differentially expressed spots of interest for identification by MS. This is both labour intensive and restricted by the limitations of 2-D gel electrophoresis such as bias against proteins that are hydrophobic or are too basic or acidic (Lilley et al., 2002). Hence in recent years the community has shifted to MS-based approaches which overcome these shortcomings. However, MS is not inherently quantitative in that the signal intensity of a peptide is not a direct read-out of the protein abundance. This lack of a direct relationship is because other factors such as size, charge and hydrophobicity of the peptide affect its

ionization and mass-spectrometric response. Hence, two peptides derived from the same protein, though present at the same stoichiometry, can give rise to different signal intensities in MS. MS-based quantitation must therefore compare chemically identical peptides. This can be achieved in two main ways: using a label-free approach or by differential labelling of peptides with stable isotopes. Label-free quantification approaches have the advantage that they can be performed with any type of sample. Clear disadvantages are the multiple occasions for experimental variation to occur during sample processing and analysis. Some quantitative information can be extracted from lists of protein identifications (qualitative experiments). Although protein and peptide identification scores provide a poor estimate of protein abundance as scores are based on the number of matching sequence fragments with the database rather than their absolute intensities, more accurate indices may be utilised for quantification as spectral counting and exponentially modified Protein Abundance Index (emPAI). The spectral counting is based on the idea that the number of MS/MS spectra assigned to a protein is a reflection of the protein abundance in the sample (Liu et al., 2004). Hence, by comparing the number of assigned spectra per protein in two different sample runs, one can obtain an estimate of the relative protein abundance. The emPAI value-based quantification relies on the observation that the number of peptides observed for a protein increases in relation to its abundance. The PAI value is obtained by normalizing the number of observed peptides to the number of observable peptides for the protein under consideration. Ishihama et al. (2005) viewed that the number of peptides observed correlated logarithmically with the protein amount. This observation led to the concept of an exponentially modified PAI (emPAI). An additional method for free-label quantification considers the peak area in the chromatogram describing the intensity of the signal related to the eluting peptides. This value is known as extracted ion current (XIC) and although the MS detector response to different peptides of the same protein is also very different because of the peptide extraction and ionization properties, the peak area is linearly related to the amount for the same peptide between two different samples over several orders of magnitude (typically 3). The intensities of the same peptides between two states of the proteome can be compared to determine their relative abundance.

Besides label-free approaches, other methods have been developed for stable isotope-based quantification in proteomics which are mainly distinguished by the way stable isotopes are introduced into the peptide or the protein. Methods for labelling fall into these categories: i) spiking in an isotopically labelled analogue; ii) having cells incorporating the label metabolically; iii) incorporating through an enzymatic reaction during protein digestion; iv)

introducing a chemical, isotopically labelled tag into the peptide or protein. Labelling is carried out using non radioactive isotopes such as ^{13}C, ^{15}N, ^{2}H. Although they can be equivalently incorporated in peptides, they can modify peptide properties. For example, deuterated peptides are cheaper to synthesise than ^{13}C, ^{15}N-labelled peptides, but they behave differently in RP-chromatography as compared to the unlabelled counterpart. The simplest approach to introduce isotope-labelled peptides in a quantitative experiment is to chemically synthesise them and spike a known amount into the protein sample as internal standard. The AQUA (Absolute Quantification) internal standard peptides are synthesized with an amino acid sequence that exactly mimics the corresponding native peptide produced during proteolysis. The heavy stable isotopes incorporate ^{13}C- and/or ^{15}N- labelled amino acids as they do not show chromatographic retention shifts. The synthetic and the native peptides share the same physicochemical properties including chromatographic elution, ionization efficiency, and relative distributions of fragment ions, but are differentially detected in a mass spectrometer due to their mass difference. An absolute quantification of the analyte is achieved by calculating the amount of the native peptide, hence the protein it came from, by direct comparison with the peak intensity of the internal standard spiked at a known concentration (Kirkpatrick et al., 2005). To reduce interference from the background ions, quantification can be performed on specific fragments of the peptide generated in the mass-spectrometer using selected- or multiple-reaction monitoring (SRM or MRM, respectively). In these methods the MS is set to detect pre-programmed precursor-fragment combination with very high sensitivity and specificity. The internal peptide standard is introduced at a late stage in the sample processing workflow during or after protein digestion. Early incorporation of the label is favorable as it allows different samples to be combined earlier in the process, minimizing variation resulting from sample preparation procedures. The earliest stage at which differential incorporation of stable isotopes can be achieved is during growth. Cells can be metabolically labeled by growing them in culture with isotopically enriched ^{15}N medium. ^{15}N is incorporated into every amino acid in the protein in this approach, originally pioneered in bacteria (Oda et al., 1999) but also applicable to higher organisms by feeding them on a diet of ^{15}N-labeled bacteria, yeast (Krijgsveld et al., 2003) or algae (Wu et al., 2004). However, since peptides contain variable numbers of nitrogen atoms, matching light and heavy peptide pairs for relative quantitation is a challenge. An alternative technique achieves metabolic labeling of proteins by Stable Isotope Labeling of Amino Acids in Cell culture (SILAC) (Ong et al., 2002). In the most common approach, lysine and arginine are replaced with $^{13}C_6$-lysine and $^{13}C_6$-arginine respectively, so that every tryptic peptide is labeled and

peptide pairs always differ in mass by the same amount, making matching very easy. Alternatively, labelling through an enzymatic reaction, consists in digesting protein in the presence of $H_2{}^{18}O$ resulting in peptides labeled with ^{18}O at their C-termini.

Finally, proteins or peptides can be chemically labeled with isotopically coded tags. There is a wide variety of commercially available tags with a number of specificities for labeling different functional groups, hence allowing selective tagging of proteins or peptides of a specific class of interest. For example, ICAT are biotinylated tags that label cysteine residues in proteins (Gygi et al., 1999), while isotope-coded protein labels (ICPL) label free amino groups in proteins and peptides (Schmidt et al., 2005). With most of these tags, the relative abundance of the labeled peptides is measured in MS. However, the iTRAQ reagent (Ross et al., 2004) works on quite a different principle. In fact, the iTRAQ reagents label peptides at their N-termini and lysine residues. They are a set of four tags that are isobaric in MS mode, but produce isotopically distinct reporter fragment ions in MS/MS, hence allowing relative quantitation during MS/MS. Chemical labeling approaches can be used on any type of sample and some tags, such as iTRAQ, are useful for simultaneous analysis of multiple samples.

2. THE AIM OF THE PROJECT

One of the main goal of research in the vaccine field is the identification of new delivery strategies are able to elicit not only humoral but also innate and cytotoxic responses. Although antibody responses are those preventing pathogen from entering the host, the cytotoxic responses are crucial for the clearance of infected cells. Vaccine preparations based on pathogen subunits and peptides, despite being safe, do not normally elicit the immune system activation without the co-delivery of immune-potentiators as they are poorly immunogenic antigens by themselves. Moreover, their administration preferentially elicit antibody responses as they behave as exogenous antigens being presented in association to MHC class II molecules and activating CD4+ T cells. This is a major limitation as vaccine-induced responses should resemble as much as possible to pathogen-induced responses therefore be able to evoke CTL responses. Among delivery systems capable to overcome this limitation, lipid-based antigen formulations have been identified. By this approach the effective activation of naïve CTL is induced by exogenous antigens uptake and processing via the phagosome-cytosolic pathway.

The present PhD project is aimed at devising *Arabidopsis thaliana* plants stably expressing antigenic peptides in association to oil bodies as oleosin fusions and successive characterization of these organelles for biochemical and immunological properties. The exploitation of oil bodies for the production and easy purification of pharmaceuticals from plants has been largely documented in literature (Parmenter et al., 1995; Nykiforuk et al., 2006). The lipid core and the particulate nature of oil bodies have suggested that these organelles may be exploited also in vaccine formulations. After extraction and purification from seeds oil bodies, chimeric and wild type organelles are going to be fine characterized by mass spectrometry before being tested as antigen carriers.

3. MATERIALS AND METHODS

3.1 Biological Materials

3.1.1 Bacterial Strains

Escherichia Coli XL10-Gold (Stratagene): Tet^rΔ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac The [F' proAB $lacI^qZ\Delta M15$ Tn10 (Tet^r) Amy Cam^r]

Agrobacterium tumefaciens GV3101: $VirG^+$ Rif^r pMP90 (Gm^r)

3.1.2 Plasmids

pET3d (Novagen): is an expression vector derived from pBR322 and contains a T7 transcription cassette.

pKMS2: is a cloning vector derived from pBlue Script KS that contains the promoter and terminator sequences of *Glycine max* cv Century 84 p24 oleosin isoform A gene and the *Not* I restriction site between the two regulatory sequences.

pTOPO TA vector (Invitrogen) is a cloning vector that is provided linearized with a single 3'-thymidine (T) overhangs and 3'-covalently bound topoisomerase I. This vector is really efficient for PCR products cloning as *Taq* polymerase adds a single deoxy-adenosine (dA) to the ends of the PCR products because of its non-template dependent terminal transferase activity, providing the insert with a compatible ends for cloning. The presence of topoisomerase I promotes the ligation between the 3' ends of the vector and 5' end of the insert.

pOBP: is cloning vector derived from pET3d and containing the sunflower oleosin coding sequence (Oil body Binding Protein, OBP).

pBIN19: is a binary vector that can replicate in both *E. coli* and *A. tumefaciens*. In this vector the Transferred DNA (T-DNA) region is replaced by a multiple cloning site (MCS). As the

cloning site is inserted within the T-DNA border sequences (Left Border, LB, and Rigth Border, RB) the transfer of a foreign DNA into the nuclear genome of dicotyledonous plants is allowed. The vector contains the neomycin phosphotransferase II (nptII) gene conferring kanamycin resistance (Km^r) for bacterial selection as well as the β–galactosidase gene (LacZ') for *E. coli* screening.

pBIN19-35S: is a binary vector derived from pBIN19. It contains the 35S promoter sequence of the Cauliflower Mosaic Virus (CaMV) upstream the MCS; this promoter allows the constitutive expression of the foreign DNA in plant tissues.

3.1.3 Oligonucleotides

Oligonucleotide primers were designed to anneal to the required target sequence template (Tab. 1). The primers were used to generate the DNA inserts for the cloning procedures, for PCR screening of positive bacteria clones and plants and for sequencing. Suitable restriction sites were introduced into the 5' and 3' ends of the primers utilised for PCR amplification of DNA inserts subsequently cloned into several vectors. The GC dinucleotide was added to both primer ends to allow efficient cleavage by restriction enzymes. All primers were purchased from Genosys Biotechnologies LTD (London).

3.1.4 Plant material

Arabidopsis thaliana plants, cv Columbia, were grown in large pots filled with damp Fison F2 and covered with polythene. Individual plants were grown for 4-6 weeks in green house with a photoperiod of 16 hours light at 22°C and 8 hours night at 15°C.

3.2 Non-Biological Materials

3.2.1 Bacterial Media

2YT 1.6% w/v Bactotryptone (Oxoid Ltd., Basingstoke, UK), 1% Bacto Yeast Extract (Oxoid), 0.5% w/v NaCl, pH 7.0

SOC 2% Bactotryptone (Oxoid), 0.5% w/v Bacto Yeast extract (Oxoid), 5 mM NaCl, 2.5 mM KCl, 20 mM glucose, 10 mM MgCl, pH 7.0

In the case of cultures on solid medium in Petri dishes, 1.6% w/v Bactoagar was also included. For vectors carrying ampicillin resistance gene as selection marker, the working concentration of ampicillin was 100 μg/ml for both solid and liquid media. *E. coli* transformed with the binary vectors were selected on solid medium and grown in liquid culture containing 50 μg/ml kanamycin. Meanwhile, *A. tumefaciens* GV3101 transformed with the binary vectors were selected on solid medium and grown in liquid culture containing 50 μg/ml kanamycin, 100 μg/ml rifampicin, 25 μg/ml gentamycin.

3.2.2 Buffer and solutions

3.2.2.1 Tris-Tricine buffers

10X Anode Buffer	2 M Trizma Base pH 8.9
10X Cathode Buffer	1 M Tris, 1 M Tricine, 1% SDS pH 8.25
3X Gel Buffer	3 M Tris-HCl, 0.3 % (w/v) SDS pH 8.45
2X TSB (Tricine Sample Buffer)	100 mM Tris-HCl pH 6.8 , 24 % (w/v) Glycerol, 8 % (w/v) SDS, 5 % (v/v) Mercaptoethanol, 0.02 % (w/v) Bromophenol Blue

3.2.2.2 Tris-glycine buffers

1X Separating buffer	0.4 M Tris-HCl pH8.8
1X Stacking buffer	0.125 M Tris-HCl pH6.8
5X SDS running buffer	0.95 M glycine, 0.1 M Trizma Base, 5% SDS, pH8.3
3X Protein loading buffer	24% glycerol, 0.15 M Tris-HCl pH6.8, 0.1% BPB, 1.6% SDS, 3% β-mercaptoetanolo

3.2.2.3 Western blot buffers

1X Transfer buffer	20% v/v methanol, 0.2 M glycine, 25 mM Trizma Base and 0.02% SDS.
Blocking solution	1% BSA (Bovin Serum Albumin) in TBS or 5% milk/PBS
1X TBS (Tris buffered saline)	150 mM NaCl, 10 mM Trizma Base, pH 8.0.
1X T-TBS (Tween-Tris buffered saline)	0.05% v/v Tween 20, 150 mM NaCl, 10 mM Trizma Base, pH 8.0
10X PBS (Phosphate buffered saline)	1.5 M NaCl, 0.015 M $Na_2HPO_4 \cdot H_2O$, 0.09M NaH_2PO_4X $12H_2O$, pH 7.2
1X T-PBS	PBS 1X, 0.1% Tween
1X Alkaline phosphate buffer	100 mM NaCl, 5 mM $MgCl_2$, 100 mM Tris-HCl, pH 9.5

3.2.2.4 Southern blot solutions and buffers

Depurination buffer	0.25 M HCl
Denaturation solution	1.5 M NaCl, 0.5 M NaOH
Neutralising buffer	1.5 M NaCl, 0.5M Trizma base, pH7.2
20X SSC	3 M NaCl, 0.3 M tri-sodium citrate pH7.0
50X Denhardts	1% w/v ficoll, 1% w/v PVP, 1% BSA
hybridization solution	6X SSC, 50% formamide, 5X Denhardts, 0.5% SDS

3.2.3 Plant transformation and selection media

Dipping medium: 5% sucrose solution (autoclaved or prepared fresh), 0.05% Silwett L-77.

Selection medium: 4.4 g/l MS salts, 0.5 g/l MES, 1% w/v Agar (Technical Nb 3, Sigma) and 50 μg/ml kanamycin.

3.3 Enzymes

Enzymes were obtained from Promega and New England Biolabs Corporations.

3.4 Immunochemicals

The oleosin polyclonal antiserum was raised at Rothamsted Research Institute immunizing rabbits with urea washed oil bodies isolated by Dr. D. Lacey (IACR-Long Ashton Research Station).The monoclonal antibody (mAb) 2F5 was provided by the National Institute of Health (NIH) AIDS Research and Reference Reagent Program; catalog no.1475. Secondary antibodies were provided by Sigma.

3.5 Sterilisation

All liquids, glassware and heat stable plastics were sterilised by autoclaving at 15 psi and 120°C for 20 minutes. Solutions intolerant to these conditions were sterilised by filtration through 2 μm Acrodisks (Gelman Sciences, Ann Arbour, MI, USA).

3.6 Methods

3.6.1 Epitope selection

Peptide sequences including or perfectly matching to T and B cell epitopes for expression on oil bodies surface as oleosin-fusions were selected using available epitope maps and referring to literature. The on-line algorithm SYFPEITHI for epitope prediction developed by the Department of Immunology of the University of Tubingen (Rammensee et al., 1999; *http://www.syfpeithi.de/)* was additionally utilised. The SYFPEITHI database is based on published T-cell epitopes and MHC ligands and predicts putative T-cell epitopes within a specified polypeptide sequence according to the MHC aplotype. The score values for each peptide are predicted considering the amino acids in the anchor and auxiliary anchor positions of a selected MHC molecule (as well as other frequent amino acids) and evaluating the frequency of the respective amino acid in natural ligands, T-cell epitopes, or binding peptides.

The score is calculated according to the following rules: amino acids have a specific value depending on whether they are anchor, auxiliary anchor or preferred residue in the selected peptide. Ideal anchors are given 10 points, unusual anchors 6-8 points, auxiliary anchors 4-6 and preferred residues 1-4 points. Amino acids supposed to have a negative effect on the binding ability are given values between -1 and -3.

3.6.2 Small scale purification of plasmid DNA

A single colony was used to inoculate 3-5 ml of 2YT media and grown overnight with appropriate antibiotic selection. Bacteria were then transferred to 1.5 ml Eppendorf tubes and pelleted at 20000g, 4°C, 1 minute. Plasmid DNA extractions were carried out using Qiagen Plasmid Mini kit.

3.6.3 Large scale purification of plasmid DNA

A single colony was used o inoculate 3 ml of 2YT media and grown for 8 hrs with appropriate antibiotic selection (preculture). One ml of the preculture was inoculated in a 2 L flask containing 500 ml growing media and appropriate antibiotic selection. The culture was grown overnight at 37°C (*E. coli*) or 28°C (*Agrobacterium tumefaciens*), 250 rpm. Then, bacterial cultures were transferred into Oak Ridge tubes and pelleted at 6000Xg, 4°C for 10 minutes. Plasmid DNA extractions were carried out using the Qiagen Plasmid Maxi kit. The final DNA pellet was resuspended in water and analysed on 0.8% agarose gel.

3.6.4 Plasmid DNA restriction

One μg plasmid DNA was digested according to the supplier instructions using 5 units of suitable enzymes. The linearization of the plasmid was assessed by analysis on 0.8% w/v agarose gel.

3.6.5 Oligonucleotide annealing

Ten μl of 5' oligonucleotide (1μg/μl) mixed with 10 μl of 3' oligonucleotide (1μg/μl) have been denatured at 95°C for 2 minutes and then let to anneal at 25°C for 1 hr.

3.6.6 Ligation reactions

The ligation of purified PCR products or purified DNA fragments obtained after restriction was carried out using the ligase supplied by Promega according to the suppliers instructions. Finally, ligase was inactivated at 65°C for 15'.

3.6.7 DNA electrophoresis

DNA samples were analysed on 1-2% agarose gel prepared in TAE 1X (TAE 50X: 2 M Trizma Base, 0.05 M $EDTANa_2$, glacial acetic acid, pH8.2). DNA samples were added with loading buffer 5X (25.5% glycerol, 0.05% bromophenol blue (BPB), 0.05 M $EDTANa_2$ pH8.0) prior to electrophoresis. DNA separation was carried out at a constant voltage about 10 V/cm (gel 5x7.5 cm) o 6 V/cm (gel 15x20 cm). DNA molecules were finally visualized on gel using the fluorescent dye ethidium bromide, at a final concentration 0.5 μg/ml.

3.6.8 Polymerase chain reaction

Amplification of specific DNA fragments was performed by Polymerase chain reaction (PCR). This method employed a heat-stable polymerase derived from *Termophilus aquaticus* (*Taq*) to exponentially amplify the DNA template of interest through thermal cycling of denaturing and synthesis steps. PCR was performed under the following conditions: an initiation step at 94°C for 2' for DNA double strands denaturation, 25-30 thermal cycles (involving a denaturation step (94°C for 45''), a primers annealing step (54°C for 45'') and an elongation step (72°C for 1')) and a final elongation step at 72°C for 10'. The thermal cycles were repeated several times (25-30 cycles) in order to amplify the DNA sequence between the primers to 2n, where n indicates the number of cycles. The reaction mix was prepared as following: 3 μl *Taq* Buffer 10X (160 mM $(NH_4)2SO_4$, 670 mM Tris-HCl pH8.8, 0.1% Tween-20) (Euroclone), 200 μM dATP, 200 μM dCTP, 200 μM dGTP, 200 μM dTTP, forward and reverse primers at a final concentration of 6.6 ng/μl each, 1.6 mM MgCl2, 15-20 ng DNA template, 1 unit (U) DNA *Taq* polymerase (EuroTaq, Euroclone), sterile bdH2O in a final volume of 30 μl. The amplification of DNA was performed using a Peltier Thermal Cycler-200 (PTC-200). PCR method has also been applied to screen bacterial colonies transformed with ligation products (Sandhu *et al.*, 1989). Briefly, each colony was picked with a sterile tip and the tip was then immerged into a 0.2 ml PCR tube containing the PCR reaction mix. The remaining cells were then spread onto a plate containing the appropriate antibiotic selection in order to keep the clone under screening. PCR was performed under the

following conditions: 94°C for 7 minutes for cell disruption, 33 thermal cycles involving a denaturation step (94°C for 45''), an annealing step (54°C for 45''), an elongation step (72°C for 1') and a final elongation step at 72°C for 10'.

3.6.9 PCR site-directed mutagenesis

Site-directed mutagenesis was used to create a new restriction site into the pOBP vector. This method uses "mutagenic" primers containing the desired mutation that bind to the template DNA and represent the starting point for DNA amplification performed by the high fidelity, non-strand displacing *Pfu* turbo DNA polymerase. The two primers 5'-*Avr*II and 3'-*Avr*II (Tab.1) were designed for the mutagenesis encoding the *Avr*II restriction site and annealing at the 5' end of the sunflower oleosin sequence in the pOBP vector. The forward primer was flanked by sequences homologous to the 3'end of OBP (5'-*Avr*II, Tab.1) while the reverse primer was flanked by sequences homologous to the vector portion downstream OBP (3'-*Avr*II, Tab.1). The amplification reactions were carried out using the QuikChange XL Site-Directed Mutagenesis Kit (Stratagene) and the Peltier Thermal Cycler-200 (PTC-200). A reaction mixture was prepared as follows: 1X reaction buffer, 1μl dNTP mixture (Stratagene), 300 ng of each primer, 50 ng template DNA, 2.5 U *Pfu* Turbo polymerase (Stratagene) in 50 μl final volume. Three different annealing temperatures (45°C, 50°C or 55°C) were tested for amplification. Then, the PCR method was as follows: initial denaturing step at 95°C for 1', 18 thermal cycles (involving a denaturing step at 95°C for 30'', an annealing step performed at 45°C, 50°C or 55°C for 5'', an elongation step at 68°C for 30'') and a final elongation step at 68°C for 7'. Finally, the methylated, non-mutated parental DNA was removed from the amplification products by digestion with *Dpn*I restriction enzyme. Ten units of *Dpn*I were added to each reaction and samples were incubated at 37°C for 1 h. Five μl of the PCR reaction were then analysed on 1% agarose gel. Two μl of amplification products were used to transform Stratagene XL-10 Gold competent cells following the suppliers instructions.

3.6.10 Sequencing and sequence analysis

Sequencing service has been provided by Geneservice DNA Sequencing Facility, (Oxford, UK). DNA for sequencing was prepared adding specific primers and lyophilising the DNA samples using a Speedvac (Savant) (30°C for 15'). The analysis of the sequences was performed on-line using the T-COFFEE server (*http://www.igs.cnrs-mrs.fr/Tcoffee/tcoffee_cgi/index.cgi*).

3.6.11 Transformation of bacteria competent cells

Bacterial competent cells were transformed either by the heat-shock or the electroporation method.

Aliquots of *E. coli* competent cells were added with 5% β-Mercaptoethanol (β-ME) and DNA (either *Dpn* I treated-DNA (2 μl), ligation products (7.5 μl) or plasmid DNA (2 μl)) and were chilled on ice for 30'. Afterwards, cells were incubated in a 42°C water bath for 30'' (heat-pulse) and immediately transferred to ice for 2'. Five hundred μl of preheated SOC medium were added to the tubes. The transformed cells were incubated at 37°C with shaking at 250 rpm for 1 h. Finally, they were spread onto selective media and incubated overnight at 37°C. As an alternative method for competent cells transformation electroporation was performed. By this method an aliquot of competent cells was mixed either with the ligation products (7.5 μl) or plasmid DNA (2 μl) and placed in a pre-cooled cuvette (0.1 mm electrode). The ligation products were de-salted by spin column chromatography through Sephadex CL-6B before electroporation. An electric pulse was applied to the mix using a GenePulser (BioRad Laboratories) set to 200 Ω, 1.8 kV, 2.5 μF. Immediately after the pulse, 0.5 ml of SOC media were added and cells gently resuspended and incubated at the appropriate temperature for 45 minutes. The cells were then plated on selective media and incubated overnight at the required temperature.

3.6.12 Preparation of *Agrobacterium* competent cells

A single colony of the *Agrobacterium* GV3101 was picked to inoculate 5 ml of 2YT containing 50μg/ml rifampicin and 25μg/ml gentamycin in a 15 ml snap-cap tube (Falcon tube). The culture was grown overnight at 30°C with shaking at 250 rpm. The day after, two 1 L flasks, each containing 250 ml of 2YT, were inoculated with 2.5 ml (1/100 volume) of the overnight culture and grown at 30°C with vigorous shaking until mid-log (O.D.600 =1-1.5). Cells were transferred to two pre-cooled 250 ml Oak ridge bottles and placed on ice for 20'. The bottles were centrifuged at 9000Xg for 10' at 4°C. The supernatant was poured off and the cells were resuspended in 150 ml of cold sterile water and centrifuged again. The cells were washed four times more with cold sterile water. Finally, cells in each tube were resuspended with 15 ml of cold sterile 10% glycerol, pooled together in a cold 50 mL polypropylene Falcon centrifuge tube and centrifuged at 3000Xg for 10' at 4°C. The supernatant was poured off and cells were resuspended very gently in 1 ml of cold sterile 10%

glycerol. Fifty μl aliquots of cells were prepared and quick-freezed in liquid nitrogen. *Agrobacterium* competent cells were finally stored at -80°C.

3.6.13 Transformation of *A. thaliana* plants by floral dip

Arabidopsis plants were grown as described (see section 3.1.4) for 4-6 weeks until the main bolts were about 10 cm. The main bolts were then clipped leaving about 2.5 cm of bare stem. The clipping procedure removed apical dominance encouraging growth of multiple secondary bolts. After a week the plants were ready for *Agrobacterium*-mediated transformation of the multiple bolts. A single colony of *Agrobacterium* carrying the appropriate construct was grown in 200 ml of selective medium (see section 2.2.2) overnight (16-20 hours) to a stationary phase (OD_{550} =1,3) on a rotary shaker (250-300 rpm). After centrifugation cells were resuspended in 250 ml of dipping medium (see section 3.2.3) and placed into a glass bowl. One pot was inverted into the solution making sure that all flowers and buds were immersed into the bacterial solution for 15''. After dipping transformed plants were kept in dark for two days and then put back into greenhouse with regular watering for 4-6 weeks. Finally, plants were dried and seeds were harvested.

3.6.14 Selection of T_0 plants integrating the transgene and isolation of T_1 isogenic lines

Seeds were harvested separately for each construct and weighted (1g= about 50,000 seeds). Four hundred mg of seeds per construct were soaked in 70% v/v ethanol for 1' and then in a solution prepared with 10% v/v Sodium Hypocloride and 0.05% v/v Tween 20 for 15' for sterilization. Seeds were rinsed 6 times with sterile water and resuspended in 0.8% melted sterile agarose (50°C) to spread evenly onto selective medium (see section 2.2.2) in 20 cm Petri dishes (0.1 g of seed per dish). Dishes were left to dry under the flow hood and kept in dark in the cold room for 2-3 days for seed vernalization. Then plates were moved to the growth chamber set at 22°C with 24 hours light. After 1-2 weeks T_0 transformants were distinguished as healthy green plants compared to pale yellow seedlings (non positive). Once T_0 transformants were identified, individual plants were transplanted to pots and grown for 4-6 weeks until seeds could be harvested. For each T_0 plant a determined number of seeds was plated onto selective medium and grown as before, for 1-2 weeks. As transformants T_0 plants obtained with floral dip are usually hemyzigous for the T-DNA insertion (Clough & Bent, 1998) and being *Arabidopsis* an autogamous plant, a segregation ratio of 3 resistant vs 1

sensitive plants in T_1 generation indicated that a single copy of the T-DNA was present in the corresponding T_0 plant (higher ratios indicated more than one copy of the T-DNA was inserted). T_1 plants generated from a single copy T_0 plant (3/4 resistant) were transplanted and grown as described. Seeds were collected and plated on selective medium, as described previously, to assess homozygous plants (integrating the transgene in the same locus on both chromosomes). The isogenic T_1 plants were identified by 100% resistant plants in their T_2 lineage and were maintained as "homozygous" lines bearing each transgene construct.

3.6.15 Plant Genomic DNA miniprep with CTAB/chloroform/isopropanol

One hundred mg of leaf tissue were grinded in 200 μl 2X CTAB extraction buffer (0.1M Trizma Base pH 9.0, 2% Cetyl Trimethyl Ammonium Bromide (CTAB), 1.4 M NaCl, 20 mM EDTA pH.8.0) supplemented with RNase A (0.5 μg/100μl extraction buffer). Samples were incubated at 65°C for 20 minutes to favour disruption of the plant cell wall. Then, 200 μl chloroform ($CHCl_3$) were added. Samples were mixed vigorously and centrifuged for 2 minutes, 14000 rpm at room temperature. The supernatant was transferred to a new tube and added with 100 μl 100% isopropanol. After incubating the samples for 1-2 minutes at room temperature, DNA was pelleted by centrifuging for 5 minutes, 14000 rpm at room temperature. An additional wash of the pellet was performed using 70% ethanol. Dry DNA pellets were finally resuspended in 50 μl water.

3.6.16 Southern blot

This technique allows the identification of a specific DNA sequence separated by gel electrophoresis within a genome using a specific radiolabeled probe (Southern 1975). On day one, 1 μg genomic DNA from leaves was digested using the restriction enzyme *BamH* I (10U). Restriction reactions were performed at 37°C for 4 hrs, in a 50 μl final volume. Digestions were run on 0.8% agarose gel, at 25V overnight. On day two, DNA molecules separated on gel were first depurinated (formation of apurinic sites) and then denatured (in a strongly alkaline solution the apurinic sites in DNA molecules are hydrolysed and double strand hydrogen bonds are removed) by soaking the gel for 15' with gentle agitation in *depurination buffer* and *denaturation buffer,* respectively. The denaturation step was repeated twice to achieve a better transfer of small, single-stranded DNA molecules to nylon membranes. A neutralisation step was carried out to bring the gel pH down to neutral and improve the DNA transfer to the membrane by incubating the gel twice in the *neutralising*

buffer for 15'. For the blotting, a positively charged Nylon membrane (Hy-bond, Amersham) and 4 pieces of 3MM Whatman paper were cut to size of the gel. One piece of 3MM Whatman paper was cut longer to act as a wick. Both ends of the wick dipped into *20X SSC* and two sheets of *2X SSC* pre-wetted 3MM paper were placed centrally onto the wick. The gel was rinsed with distilled water and placed onto the 2 pieces of 3MM paper. Parafilm was used to mask the edges and prevent SSC by-passing the gel. Pre-wetted nylon membrane was placed on gel and covered with the other two sheets of 3MM pre-wetted paper. A 5-10 cm stack of absorbent paper towels was placed on top followed by a gel tray and a weight. The DNA transfer took place overnight by capillarity. On the third day, the membrane was briefly rinsed in 2X SSC and the DNA molecules transferred were cross-linked to the membrane using the Stratalinker UV crosslinker for 1' following the suppliers instructions. A pre-hybridization step of the membrane was followed by hybridization with a specific probe incubating the membrane at 65°C in a rotisserie oven overnight. The probe was prepared by labelling 1 μg pOBP linearized plasmid using the "Prime it II" Random Primer kit (Stratagene). This labelling procedure relies on the ability of random hexanucleotides to anneal to multiple sites along the length of a DNA template. The primer–template complexes formed represent a substrate for the Klenow fragment of DNA polymerase I. The newly synthesized DNA is made radioactive by adding a radiolabeled nucleotide ([α–^{32}P]dATP or [α−32P]dCTP) in the reaction mixture. To remove unbound radiolabeled nucleotides from the probe mix "NucTrap" columns (Stratagene) were used, according to the suppliers instructions. The purified probe was denatured in a boiling water bath for 10' before adding to the *hybridization solution* (6X SSC, 50% formamide, 5X Denhardts, 0.5% SDS). On the fourth day, the membrane was washed twice (0.5% SDS, 2X SSC) from the probe solution in a 42°C shaking bath and let to dry on a 3MM Whatman paper. The membrane was then placed in a developing cassette with an X-ray film. The film was exposed overnight at -20°C and developed on the day after using an automatic developer.

3.6.17 Total protein extraction from leaves and siliques

Fifty mg of leaves tissue or 50 mg mature siliques (mature seeds before desiccation) were grinded in liquid nitrogen using mortar and pestle. The homogenous powder was transferred to a 1.5 ml sterile Eppendorf tube and resuspended in 200 μl of Tricine Sample Buffer (Bio-Rad) added with 0.1 M DTT. Cell debris were pelleted by centrifuging the samples for 5' at maximum speed in a table top microcentrifuge. The supernatants were transferred to fresh

tubes and heated at 65°C for 15' to unfold proteins and break protein secondary structures. Twenty µl each sample were analysed by Coomassie staining and Western blot.

3.6.18 Protein quantification

Protein from plant crude extracts or oil body total soluble protein (TSP) were quantified by Bradford assay (Bradford, 1976) that relies on the modification of the absorbance maximum for an acidic solution of *Coomassie Brilliant Blue G-250* dye (from 465, red to 595, blue) after binding to proteins. The Micro assay kit (Bio-Rad) and the Bovine Serum Albumin (BSA) standard curve as reference were used to determine the protein concentration.

3.6.19 SDS-PAGE

Proteins and polypeptides were separated on the basis of their molecular weight in 1mm denaturing polyacrylamide gels consisting of a stacking phase (4% polyacrilammide (acrilamide/bisacrilammide 29:1), 0.1% Sodium Dodecyl Sulfate (SDS), 0.06% Ammonium persulfate (APS), 0.2% TEMED in *1X stacking buffer* or *3X Gel buffer*) and a separating phase (12% polyacrilammide, 0.1% SDS, 0.05% APS, 0.1% TEMED in *1X separating buffer* or *3X gel buffer*). Both Tris-Glycine and Tris-Tricine buffer systems were used. Gels were immerged in *1X running buffer* or *10X Anode and Cathode buffer* and run vertically using the Mini-Protean 3 (Bio-Rad) apparatus. Protein samples were prepared for running adding *3X protein loading buffer* or *2X TSB* and denaturing at 100°C for 5'. Runs were performed at a constant voltage of 130V until the outcome of the dye.

3.6.20 Coomassie staining

Polypeptides separated on polyacrilammide gels were visualized by staining with *Coomassie brilliant blue G250* dye. Gels were fixed for at least 20 minutes in methanol: acetic acid: water (45:1:54), then stained for 12-18 hours using 17% (w/v) ammonium sulphate, 34% methanol, 0.5% acetic acid, 0.1% (w/v) Coomassie G250. The day after gels were destained in distilled water until desired contrast.

3.6.21 Western blot

This technique allows the identification of a specific polypeptide blotted onto a nitrocellulose membrane using an antibody. After separation on polyacrilammide gel polypeptides were transferred onto 0.2 µm Hybond nitrocellulose membrane (Whatman), pre-soaked in *1X*,

transfer buffer using a "Trans-Blot SD" semi-dry blotting apparatus (BioRad) set at 13V for 30'. After blotting the membrane was immerged in 5% skimmed milk (SIMILAC formula PLUS, Abbott) in *1X PBS* or 1% BSA (Bovin Serum Albumin) in *TBS* overnight at 4°C for blocking. The day after the membrane was rinsed three times with *1X T-PBS* and twice with *1X PBS* and then incubated with a solution of 2% milk in PBS containing the primary antibody (1:5000 dilution of rabbit serum raised against sunflower urea-washed oil bodies; 1:100 dilution of mAb 2F5 (EVA3063) for 2 hrs. Rinses in T-PBS and PBS were repeated. The membrane was then incubated with an alkaline phosphatase-conjugated secondary antibody (1:5000 dilution of anti-rabbit whole IgG; 1:200 anti-human IgG) resuspended in 2% milk in PBS for 1 h. Finally, the membrane was rinsed and the immobilized antibody detected by the alkaline phosphatase reaction to the substrate (NBT/BCIP purple solution (Sigma)).

3.6.22 Oil body purification according to Tzen (Tzen et al., 1997)

One gram of *A. thaliana* seeds were grinded in 40 ml *grinding medium* (0.6 M Sucrose, 10 mM sodium phosphate buffer pH 7.5, 1mM EGTA, 1 mM PMSF) at 4°C. Cell debris were pelleted by centrifuging the sample at 5000Xg for 15 min and the supernatant was placed at the bottom of a 38 ml centrifuge tube. The centrifuge tube was topped up with 15 ml *flotation medium* (0.4 M Sucrose, 10 mM sodium phosphate buffer pH7.5) and spun at 10000Xg for 20 min in a swinging-bucket rotor. The fat pad (floating unwashed oil bodies, Step 1, S1) was skimmed from the top (~5 ml*)* and resuspended in 20 ml *detergent washing solution* (0.1% Tween 20, 0.2 M Sucrose, 5 mM sodium phosphate buffer pH 7.5). The sample was topped up with 10 ml of 10 mM sodium phosphate buffer pH 7.5 and spun at 10000Xg for 20 min. The fat pad (detergent washed oil bodies, Step 2 Tzen, S2T) collected on top (~5 ml) was resuspended in 15 ml *ionic elution buffer* (0.6 M Sucrose, 10 mM sodium phosphate buffer pH 7.5, 2 M NaCl) and topped up with 15 ml *floating medium* (0.25 M Sucrose, 10 mM sodium phosphate buffer pH 7.5, 2 M NaCl). The sample was centrifuged at 10000Xg for 20 min. The fat pad (salt washed oil bodies, Step 3 Tzen, S3T) was resuspended in 1-2 ml 10 mM sodium phosphate buffer pH 7.5 and mixed with 10 ml 8 M urea. Urea washing of the oil bodies took place for 10', at room temperature at 60 rpm. Ten ml *grinding medium* was mixed with urea washed oil bodies and the mix placed at the bottom of a 35 ml centrifuge tube and layered on the top with 15 ml of 10 mM sodium phosphate buffer pH 7.5. The sample was centrifuged again and urea washed oil bodies (~10 ml) skimmed from the top. Oil bodies were finally resuspended in 1.5 ml 10 mM sodium phosphate buffer pH 7.5 and stored at 4°C (Step 4 Tzen, S4T).

3.6.23 Oil body purification according to Deckers (Deckers et al., 2004)

One gram of *A. thaliana* seeds were grinded in 5 volumes cold *grinding buffer* (50 mM Tris-HCl pH 7.5, 0.6 M Sucrose, 0.5 M NaCl) and the mix was centrifuged 5000Xg for 30' to remove solids. The oil body fat pad was skimmed from the top of the tube with a metal spatula and resuspended in one volume of *grinding buffer*. The oil bodies were then layered underneath 5 volumes of cold 50 mM Tris-HCl pH 7.5 in a centrifuge tube and centrifuged. The washing procedure was repeated twice (Step 2 and 3 Deckers, S2D and S3D). A further washing to remove seed protein contaminants was carried out incubating oil bodies with 0.1 M Na_2CO_3 pH 11 for 30', at room temperature with shaking at 60 rpm. Later, the mixture was layered underneath 5 volumes of water and centrifuged. Finally, the washed oil bodies were resuspended in cold 50 mM Tris-HCl pH 7.5 (Step 4 Deckers, S4D).

3.6.24 Treatment to remove lipid contaminants from oil body preparation (de-lipidation)

Protein samples that were analysed by SDS-PAGE and Mass Spectrometry (MS) were previously treated to remove lipid contaminants forming the oil bodies core. One hundred µl chloroform/methanol (2:1 v/v) and 100 µl 10% TCA/Acetone were added to the samples (Wang at al., 2004). Samples were incubated on ice for 5' and centrifuged at maximum speed for 1'. Two washes were performed with 500 µl cold 10% TCA and one using 1ml cold water. Proteins were finally precipitated by adding 1ml cold acetone and incubating at -20°C for at least 1hr. After precipitation protein pellet was resuspended in 1% SDS buffer (1% SDS, 0.1 M sodium phosphate buffer pH 7.5) containing or not containing 6 M urea.

3.6.25 Treatment with trypsin of oil body associated proteins for MS analysis

Twenty to fifty µg of a pellet containing oil body total soluble protein (TSP) were resuspended in 50 mM ammonium bicarbonate pH 7.5 to a final concentration of 0.1-0.5 µg/µl. Acetonitrile was added to a final concentration of 5% to promote protein resuspension. Denaturation of proteins was carried out by reduction with 10 mM DTT and incubation of the sample for 1hr at room temperature and alkylation by adding 10 mM iodoacetamide, to prevent proteins to fold back. Incubation with iodoacetamide was carried out in the dark for 15' at room temperature. Trypsin (Sequence Grade Modified Trypsin, Promega) protease was

added to the protein sample to a final ratio substrate to enzyme of 20:1 (w/w) and the reaction was incubated overnight at 37°C. Samples were centrifuged twice at maximum speed to remove the non-digested pellet. Formic acid was added to the supernatant to a final concentration 0.1% to improve the protonation of the tryptic peptides at the ion source.

3.6.26 LC-MS/MS analysis of oil body total proteins

The qualitative and quantitative analysis of oil body TSP Liquid Chromatography Mass Spectrometry (LC-MS) analysis was carried out on in-gel tryptic digests (GeLC). Tryptic digestions of proteins separated on polyacrilammide gels were carried out on protein bands excised from Coomassie-stained 1D-gels (50 μg of oil body total proteins loaded on a 12% SDS-PAGE) and stored in a 96 well plate. The gel pieces were washed in 25 mM ammonium bicarbonate in 50% acetonitrile, subjected to reduction (25 mM DTT) and alkylation (12.5 mM iodoacetamide) in order to break disulphide bridges between cysteine residues and prevent reforming. Fifty μl trypsin (Sequence Grade Modified Trypsin, Promega) solution (15 ng mL^{-1} in 25 mM ammonium bicarbonate) was added to the dehydrated gel piece and allowed to digest overnight at 37°C. Salts and buffers in the digest mixtures were removed using ZipTips (Millipore Corporation) composed of a plug of C_{18} packing material. The samples were then loaded from a 96 well plate onto a pre-column for 3 minutes (0.1% formic acid, 30 mL min^{-1}) to concentrate. After 3 minutes, the peptides were eluted from the pre-column by 5-40% acetonitrile gradient flow over 40 minutes and bound to the PicoFrit column (75 mm, 50mm, New Objective Inc) for separation. Peptides were finally sprayed directly from the column into the source region on an LTQ XL Linear Ion Trap Mass Spectrometer (Thermo Finnigan). Tandem MS (MS/MS) spectra were searched using the Mascot search engine version 2.2 and a modified *A. thaliana* FASTA database downloaded from the National Center for Biotechnology Information FTP site including the sunflower oleosin and the chimeric sunflower oleosins sequences.

3.6.27 Enzyme-Linked ImmunoSorbent Assay (ELISA) for epitope detection and quantification

This technique allows the detection of a specific protein in a mixture by adhesion of the protein to a polypropylene surface and antibody detection. The presence and quantification of the chimeric oleosin on the oil body surface by detecting 2F5e or the sunflower oleosin was determined by direct ELISA, coating Nunc immunomodule plates (Maxisorb) with serial

dilutions (5, 4, 3, 2, 1 μl) of oil body preparations in a final volume of 50 μl overnight at + 4°C. As for quantification of 2F5e on chimeric oil body surface a group of wells was coated with different amounts (0.5, 1, 2, 4 and 8 ng) of a synthetic polypeptide containing 2F5e (QTQQEKNEQELLELDKWASL; NIH AIDS Research and Reference Reagent Program, catalog no.2030) as positive control/internal standard. As negative control, a group of wells was coated with oil bodies extracted from the seeds of plants transgenic for the non-chimeric sunflower oleosin. As for quantification of the sunflower portion of the chimeric oleosin a group of wells was coated with serial dilutions (5, 4, 3, 2, 1 μl) of oil bodies carrying the non chimeric sunflower oleosin as a standard. After coating, wells were washed three times with PBS containing 0.1% Tween 20 (PBS-T) and two times with PBS. After 2 h of blocking at 37°C with PBS containing 5% skimmed milk and washing as above, wells were incubated with 2F5 monoclonal antibody (EVA3063; Buchacher et al., 1994) (1:500 dilution in PBS 2% milk at 37°C for 2 h) or sunflower oleosin-specific polyclonal antibody (1:5000 dilution in PBS 2% milk at 37°C for 2 h). Wells were then washed, and the presence of bound antibodies was detected by the application of a peroxidase-conjugated anti-human IgG antibody (1:200) or a biotin-conjugated anti-rabbit IgG antibody (1:5000) mixed with streptavidin-peroxidase (1:2000) (37°C for 2 h). Afterwards the chromogenic substrate 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) (ABTS) (Kirkegaard & Perry Laboratories, Gaithersburg, MD) was added and the colorimetric reaction measured with an automated ELISA reader (ELISA SUNRISE (Tecan)) at 405 nm after 1h.

3.6.28 Animals and Immunization

Six week-old C57BL/6J female mice in groups of 4 (Charles River Laboratories Italia Spa, Lecco, Italy) were maintained under standard housing conditions in the ENEA Casaccia Animal Care Unit. The experimental procedures were approved by the ENEA ethical committee and performed according to the current Italian law.

3.6.29 ELISPOT assay

A mouse IFN-γ ELISPOT kit was used according to manufacturer's instructions (eBioscience, San Diego, CA). Briefly, 19 days after the last immunization, spleens were harvested from mice and single splenocyte suspensions prepared by mechanical disruption. After red blood cells lysis, 0.5 x 10^6 splenocytes were distributed in triplicate wells of PVDF-bottomed 96 well plates (Millipore, Billerica, MA) previously coated overnight at 4°C with a rat anti-mouse IFN-γ antibody. Cells were left unstimulated (negative control) or stimulated with

either ASNENMETM or TYQRTRALV synthetic peptides (0.4 μg/well). As positive internal control, 0.4 x 10^4 splenocytes were stimulated with 1 ng Phorbol Myristic Acetate (PMA) - 0.5 mg ionomycin /ml (Sigma, St. Louis, MO). After 20 h at 37°C in atmosphere with 0.5 % CO_2 and 95% relative humidity, plates were washed and incubated with a biotinylated rat anti-mouse IFN-γ antibody followed by avidin-Horse Radish Peroxidase (HPR). The binding of the antibody was then visualised by adding the substrate 3-amino-9-ethyl carbazole (AEC). Spots were counted using an automated ELISPOT reader (Aelvis, Hannover, Germany). The frequency of peptide-specific T cells present in the responding population was calculated by subtracting the mean number of spots obtained in unstimulated wells from the mean number of spots obtained in stimulated wells, and reported as number of specific spot-forming cells (SFC)/spleen. Statistical analysis was performed using the unpaired Student's *t*-test. Differences were considered statistically significant when $P<0.05$.

3.6.30 Antibody titration in animal sera

Individual mouse serum samples, obtained by tail bleeds, were analyzed by enzyme-linked immunosorbent assay (ELISA) to determine the concentrations of anti-2F5e antibodies. Briefly, individual wells of 96-well microtiter Maxisorp ELISA plates (Nunc, Roskilde, Denmark) were coated overnight at 4°C with 50 μl PBS containing 4 ng/μl synthetic polypeptide containing 2F5e (QTQQEKNEQELLELDKWASL; NIH AIDS Research and Reference Reagent Program, catalog no.2030). After coating, wells were washed three times with PBS containing 0.1% Tween 20 (PBS-T) and two times with PBS. After 2 h of blocking at 37°C with PBS containing 5% skimmed milk and washing as above, 50 μl of individual mouse serum dilutions (in PBS with 2% milk) were added to triplicate wells, and incubated overnight at 4°C. Wells were then washed, and the presence of bound antibodies was detected using an Horse Radish Peroxidase (HRP)-labelled sheep anti-mouse IgG polyclonal antibody (1:2500) (Amersham Bioscience, General Electrics, New Brunswick, NJ) and 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid) (ABTS) (Kirkegaard & Perry Laboratories, Gaithersburg, MD) as substrate. The colorimetric reaction was measured with an automated ELISA reader (ELISA SUNRISE (Tecan)) at 405 nm, and antibody levels were expressed as optical density values at 405 nm (OD_{405}). Endpoint titers were defined as the reciprocal of the highest serum dilution giving an absorbance ≥0.1 OD unit above the blank (absorbance of the pre-immune sera). Geometric mean titers (GMT) were determined for each group, including non-responder mice.

4. RESULTS

4.1. Epitope selection

To direct the expression on oil bodies surface of amino acid sequences including or corresponding to T and B cell epitopes derived from HIV-1 (Nef and Env) and influenza virus proteins (Nucleoprotein, NP) we have adopted an oleosin-fusion strategy. Nef is a phosphorylated and N-terminally myristoylated protein of about 27 kDa (206 aa) fundamental for HIV-1 pathogenesis (Arora et al., 2002). As with many other HIV-1 regulatory proteins, it is considered as a possible target of an anti-HIV-1 vaccine formulation for the activation of CTL mediated responses. Nef peptides to be expressed as oleosin fusions have been selected referring to Nef CTL experimentally-based epitope-mapping available on the HIV-1-dedicated NIH website (National Institute of Health of Bethesda, USA; *http://www.hiv.lanl.gov/content/immunology/maps/ctl/Nef.html*) and considering the haplotypes of the murine models available to test the immunological potential of chimeric oil bodies (Balb/c, C57BL/6 and H2 knock-out/HLA-A2 transgenic mice). After the two polypeptides were identified on this basis, Nef125 (aa sequence 125-QNYTPGPGIRYPLTFGWCYKLVPVEPE-151) (Fig. 3) was further analysed using the on-line algorithm SYFPEITHI for epitope prediction (Rammensee et al., 1999; *http://www.syfpeithi.de/*). By this approach 33 peptides, either 9mer or 10mer, fitting the binding site on MHC class I molecules in animal models (HLA-A2, H2-d, H2-b) were identified (Tab. 2).

As Env protein of HIV-1 is the only structural protein able to induce neutralizing antibodies (Moore J.P., 2001), a highly conserved and immunogenic 6 amino acid spanning gp41-derived epitope (2F5e) is the other HIV-1-derived sequence selected for fusion to oleosin. This peptide (aa position 662-668, sequence ELDKWA) is recognized by the human monoclonal antibody 2F5 and represents one of the few epitopes able to induce neutralizing antibodies in humans (Muster et al., 1993).

NP is a major structural protein in the ribonucleoprotein complex of influenza virus and is a major target structure for cross-reactive CTL. The epitope for oleosin fusion was derived from NP of the influenza virus strain A/PR/8/34. The selected 9 amino acid spanning peptide (aa position 366-374, ASNENMETM) is the immunodominant epitope in C57BL/6J mice (H2-Db haplotype) able to elicit CTL responses responsible for virus clearance in this murine model (Flynn et al., 1998).

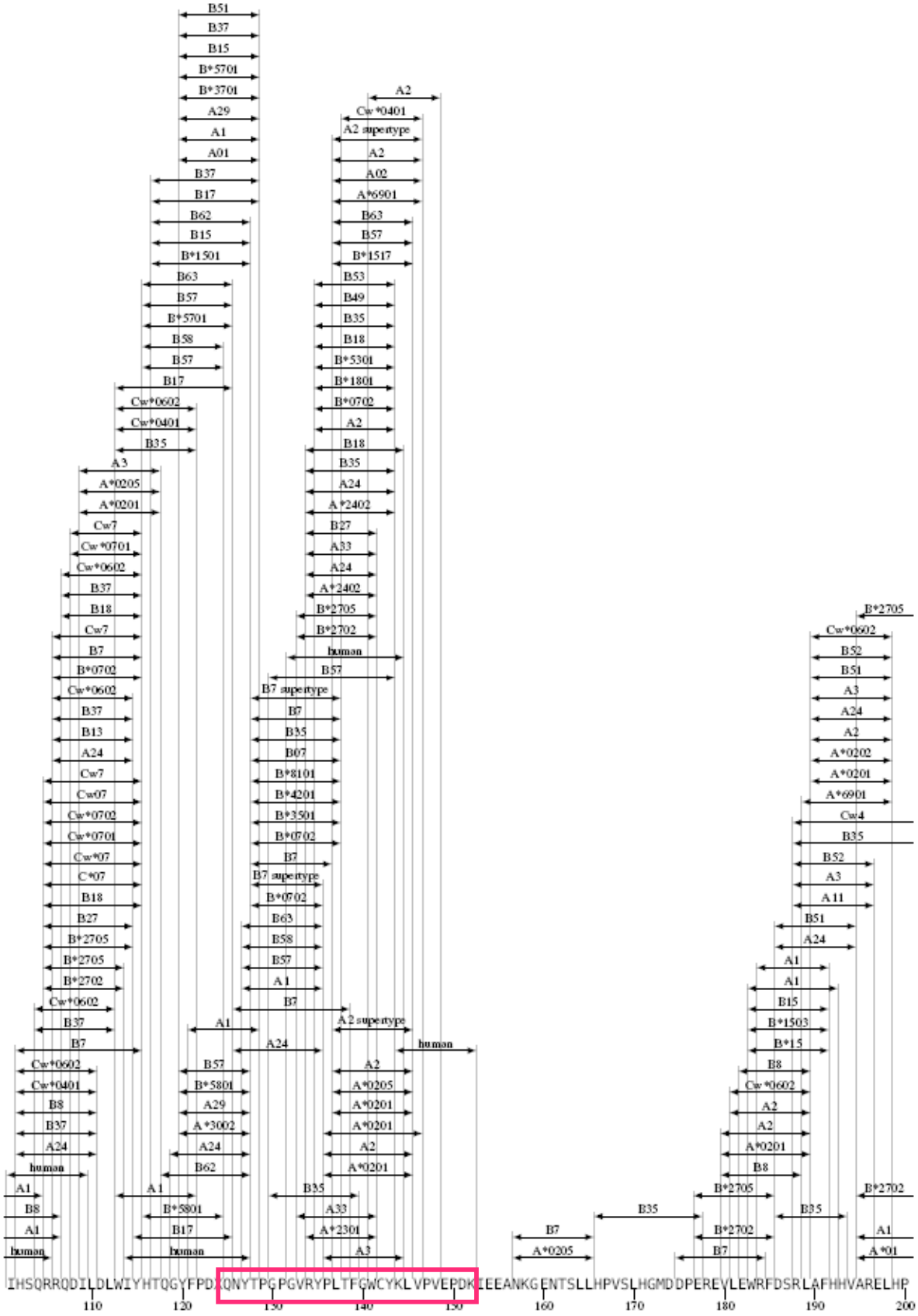

Figure 3. Nef CTL epitope map of the region including the Nef125 polypeptide (pink box) selected to be expressed as oleosin fusion.

Table 2. List of the epitopes binding to MHC/HLA class I molecules (H2-d, H2-b, HLA-A2) predicted by the SYFPEITHI algorithm as included in the Nef125 polypeptide sequence

Peptide sequence																												Haplotype	Score*
Q	N	Y	T	P	G	P	G	I	R	Y	P	L	T	F	G	W	C	Y	K	L	V	P	V	E	P	E			
				P	G	P	G	I	R	Y	P	L															9mer	Db	14
						P	G	I	R	Y	P	L	T	F													9mer	Db	14
Q	N	Y	T	P	G	P	G	I																			9mer	Db	11
												L	T	F	G	W	C	Y	K	L							9mer	Db	10
			T	P	G	P	G	I	R	Y	P																9mer	Ld	16
				P	G	P	G	I	R	Y	P	L															9mer	Ld	12
						P	G	I	R	Y	P	L	T	F													9mer	Ld	12
					G	P	G	I	R	Y	P	L	T														9mer	Ld	11
										Y	P	L	T	F	G	W	C	Y									9mer	Ld	11
												L	T	F	G	W	C	Y	K	L							9mer	Ld	11
				P	G	P	G	V	R	Y	P	L															9mer	Dd	31
													T	F	G	W	C	Y	K	L	V						9mer	Kd	19
Q	N	Y	T	P	G	P	G	I																			9mer	Kd	19
			T	P	G	P	G	I	R	Y	P																9mer	Kd	16
			T	P	G	P	G	I	R	Y	P	L															10mer	Kd	12
																	C	Y	K	L	V	P	V	E	P	E	10mer	Kd	12
									R	Y	P	L	T	F	G	W	C	Y									10mer	Kd	11
												L	T	F	G	W	C	Y	K	L	V						10mer	Kd	11

Table 2. Continued

Peptide sequence																															
	N	Y	T	P	G	P	G	I	R	Y																		10mer	Kd	10
											P	L	T	F	G	W	C	Y	K	L								10mer	Kd	10
			T	P	G	P	G	I	R	Y	P	L																10mer	Db	12
					G	P	G	I	R	Y	P	L	T	F														10mer	Db	11
														F	G	W	C	Y	K	L	V	P	V				10mer	Db	11
											P	L	T	F	G	W	C	Y	K	L								10mer	Db	8
												L	T	F	G	W	C	Y	K	L								9mer	HLA-A*0201	22
															G	W	C	Y	K	L	V	P	V					9mer	HLA-A*0201	17
							G	I	R	Y	P	L	T	F	G														9mer	HLA-A*0201	13
Q	N	Y	T	P	G	P	G	I																					9mer	HLA-A*0201	12
				P	G	P	G	I	R	Y	P	L																	9mer	HLA-A*0201	10
											P	L	T	F	G	W	C	Y	K	L								10mer	HLA-A*0201	18
												L	T	F	G	W	C	Y	K	L	V							10mer	HLA-A*0201	18
														F	G	W	C	Y	K	L	V	P	V					10mer	HLA-A*0201	18
			T	P	G	P	G	I	R	Y	P	L																	10mer	HLA-A*0201	13
										Y	P	L	T	F	G	W	C	Y	K									10mer	HLA-A*0201	12

* For score definition refer to Materials and Methods section.

4.2 Gene engineering for plant expression of HIV-1 and influenza-derived peptides as oleosin fusions

4.2.1 pOBP vector modification

The sunflower oleosin cDNA sequence (oil body binding protein, OBP) inserted into a pET-3d vector (pOBP) (Fig. 4a) was modified by site-directed mutagenesis PCR to introduce an additional restriction site at the 3' end. Primers encoding the *Avr* II restriction site were designed. The forward primer was flanked by sequences homologous to the 3'end of OBP (5'-*Avr* II oligonucleotide) while the reverse primer was flanked by sequences homologous to the vector portion upstream OBP (3'-*Avr* II oligonucleotide) (Tab.3). The PCR reactions were performed in a temperature gradient to assess the best annealing conditions. PCR products analysis on 0.8% agarose gel showed that the amplification of the vector occured only at an annealing temperature of 55°C. XL10 GOLD competent cells (Stratagene) were transformed using this PCR product and correctly mutagenized plasmids (pOBP/*Avr* II) (Fig. 4b) were identified by *Avr* II restriction and sequence analysis.

4.2.2 Cloning of sequences encoding Nef, Env or NP epitopes into pOBP/*Avr* II vector

The double-stranded DNA fragments encoding the peptides Nef 125, 2F5e and NP were obtained by *in vitro* annealing of synthetic oligonucleotides (5'-Nef125 and 3'-Nef125; 5'-2F5 and 3'-2F5; 5'-NP and 3'-NP) (Tab. 3). Prior to annealing, oligonucleotides were phosphorylated using T4 Polynucleotide Kinase. The DNA fragments were then ligated to the *Avr* II-linearized and dephosphorylated pOBP/*Avr* II vector. Cells transformed with the ligation reaction products were screened by PCR using T7 and epitope-specific (3'-Nef125PCR, 3'-2F5 or 3'-NP) primers. By this approach, only clones carrying plasmids where sequences encoding the epitope were inserted in the correct orientation, were identified. The plasmids pOBP/*Avr* II-Nef125, pOBP/AvrII-2F5e and pOBP/*Avr* II-NP (Fig. 4c) were further analysed by *Avr* II-restriction and sequencing.

4.2.3 Transferring of the chimeric oleosin coding cassettes into pKMS2 vector

To generate DNA fragments encoding chimeric oleosins and having *Not* I-compatible ends, the chimeric oleosin sequences into pOBP/*Avr* II-Nef125, pOBP/*Avr* II-2F5e or pOBP/*Avr* II-NP vectors were amplified with appropriate pairs of primers (5'*Not*OleoF and 3'-*Not*125R, 3'-*Not*2F5R or 3'-*Not*NPR; Tab. 3). PCR products were cloned into Topo-TA vectors for sequencing prior to transfer into the vector pKMS2. The vector pKMS2 (kindly provided by Dr. E.Cahoon) carrying the *Glycine max* cv Century p24 oleosin promoter and terminator sequences and between them a multicloning site including *Not* I was utilised to add regulatory sequences to the chimeric oleosin coding cassettes. Colony PCR screening, performed using appropriate up- and downstream primers (5'-pKMS2F and 3'-*Not*125R, 3'-*Not*2F5R or 3'-*Not*NPR; Tab. 3) identified the clones carrying the chimeric oleosin inserted in the correct orientation.

4.2.4 Transfer of the chimeric oleosin coding cassettes with the oleosin specific regulatory sequences to the binary vector pBIN19

The chimeric oleosin coding cassettes, flanked by the oleosin promoter and terminator sequences into pKMS2-derived plasmids (pKMS2-Nef125, pKMS2-2F5e and pKSM2-NP) (Fig. 4d), were excised using the restriction enzymes *Kpn* I and *Sac* I. To skip gel purification procedures, a further digestion with the restriction enzyme *Pvu I* was also performed, to modify pKSM2 backbone ends prior to the set-up of a ligation reaction with *Kpn* I and *Sac* I digested pBIN19. After *E. coli* transformation, positive clones were identified by PCR using 5'-pKMS2F and 3'-pKMS2R primers.

4.2.5 Cloning of Nef125-, 2F5e- or NP-oleosin fusions into the binary vector pBIN19-35S

The chimeric oleosin coding cassettes were excised from Topo-TA vectors using *Not* I and their sticky ends were filled-in with DNA Polymerase I (Klenow fragment) prior to the insertion into *Sma* I-linearized pBIN19-35S binary vector. Transformed *E. coli* cells were screened by PCR using 5'-35S and 3'-*Not*125R, 3'-*Not*2F5R or 3'-*Not*NPR primers (Tab. 3). Positive clones with the insert in the correct orientation were used for plasmids extraction and restriction analysis.

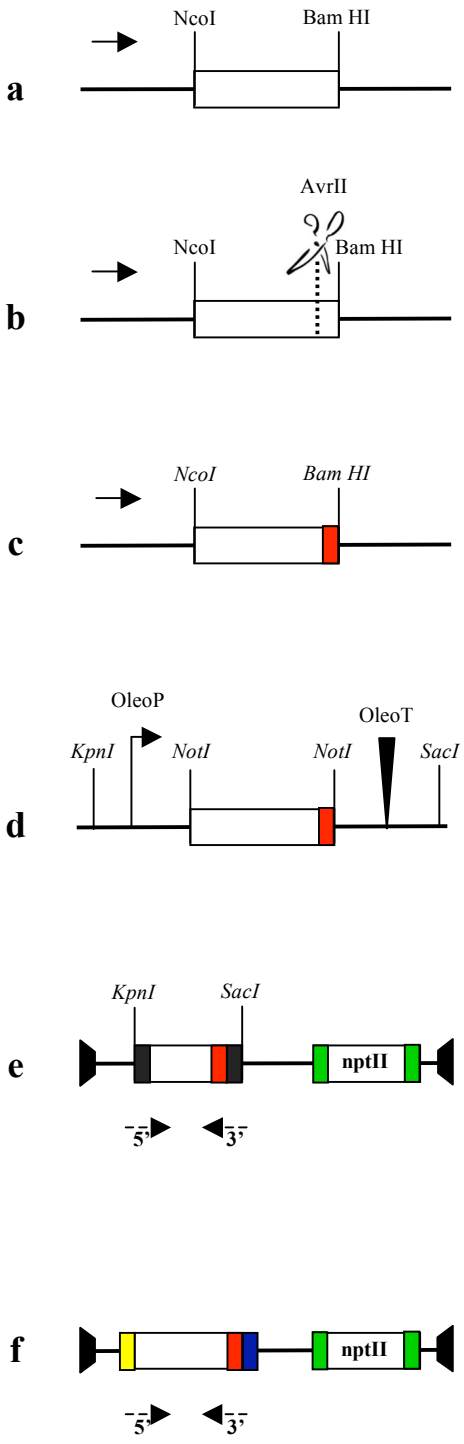

Figure 4. Constructs.

Schematic representation of (a) pOBP), (b) pOBP/*Avr*II, (c) pOBP/*Avr*II-Nef125, -2F5e, -NP, (d) pKMS2-Nef125,-2F5e,-NP, (e) pBIN19-GM-Nef125, -2F5e, -NP, (f) pBIN19-35S-Nef125, -2F5e, -NP constructs.

White box: sunflower oleosin c-DNA (OBP); black arrow: T7 promoter; dashed arrow: 5'-NotOleoF and 3'-Not125R or 3'-Not2F5R or 3'-NotNPR primers; OleoP: *Glycine max* p24 oleosin promoter; OleoT: *Glycine max* p24 oleosin terminator; black trapeziums: T-DNA left (LB) and right (RB) borders; green boxes: *Pnos* (nopaline synthase gene promoter) and *nosA ter* (nopaline synthase gene terminator); yellow box: CaMV 35S constitutive promoter; blue box: *nosA ter*; nptII: neomycin phosphotransferase II gene; red box: Nef125-, 2F5e- or NP-peptide encoding sequence; black boxes: *G. max* p24 oleosin promoter and terminator.

Table 3. Oligonucleotide sequences

5'-AvrII	5'-GAAGGGGGGAAACCTAGGTAAGGTGAGAAA-3'
3'-AvrII	5'-TTTCTCACCTTACCTAGGTTTCCCCCCTTC-3'
5'-Nef 125	5-CTAGGCAAAATTATACTCCTGGACCTGGAATTCGTTATCCTCTTACTTTTGGATGGTGTTA TAAGCTTGTTCCTGTTGAACCTGAATAACTCGAGC-3'
3'-Nef125	5'-CTAGGCTCGAGTTATTCAGGTTCAACAGGAACAAGCTTATAACACCATCCAAAAGTAAGA GGATAACGAATTCCGGATCCAGGAGTATAATTTTGC-3'
5'-NP	5'-CTAGCGCTTCTAACGAGAACATGGAGACTATGTAAC-3'
3'-NP	5'-CTAGCTTACATAGTCTCCATGTTCTCGTTAGAAGCC-3'
5'-2F5	5'-CTAGCGAACTTGATAAGTGGGCTTAAC-3'
3'-2F5	5'-CTAGGTTAAGCCCACTTATCAAGTTCG-3'
3'-Nef125PCR	5'-CTAGGCTCGAGTTATTCAGG-3'
T7 Promoter	5'-TAATACGACTCACTAT-3'
5'-pKMS2 F	5'-ATATAACAACACCCCCGT-3'
5'-pKMS2 R	5'-CCTACATAGGCAAGAGGT-3'
3'-*Not*125R	5'-GGCCGCGGCCGCTTATTCAGGTTCAACAGGAA-3'
3'-*Not*NPR	5'-GGCCGCGGCCGCTTACATAGTCTCCATGTTCT-3'
3'-*Not*2F5R	5'-GGCCGCGGCCGCTTAAGCCCACTTATCAAGTT-3'
5'-*Not*OleoF	5'-AATTGCGGCCGCATGGCCACCACAACCTACGA-3'
5'-35S	5'-CTATCCTTCGCAAGACCCTTC-3'

4.3 *A. thaliana* transformation and characterization of transgenic plants

4.3.1 *A. thaliana* plants transformation by floral dip

Competent *A. tumefaciens* cells (strain GV3101) were transformed by electroporation with the binary vectors pBIN19- or pBIN19-35S carrying the chimeric oleosin sequences (pBIN19-GM-Nef125, pBIN19-GM-2F5e, pBIN19-GM-NP, pBIN19-35S-Nef125, pBIN19-35S-2F5e or pBIN19-35S-NP) (Fig. 4e and f). Transformed *Agrobacterium* cultures were used to transfer gene constructs to 5 weeks old *A. thaliana* plants by floral dip (Clough and Bent, 1998). About three weeks later, seeds were collected and sowed on selective kanamycin solid medium (100 mg seeds /plate) to identify T_0 plants.

4.3.2 Analysis of T_0 resistant plants to verify transgene expression

Expression of the fusion protein was verified on crude leaves and mature siliques extracts by Western blot using a polyclonal antibody specific to sunflower oleosin, without cross-reaction to endogenous *A. thaliana* oleosins. By this approach three major bands were evidenced on the blotted membrane together with minor bands probably resulting from the aspecific binding of the antibody. The major band with the predicted molecular mass (20-23 kDa) was clearly visible in the leaves and siliques extracts of plants transformed with pBIN19-35S vectors (constitutive promoter) (Fig. 5) and only in the siliques extracts of the plants transformed with pBIN19-GM vectors (seed-specific promoter) (Fig. 6 and Fig. 7). The other two major bands were 15 and 19 kDa for plants expressing the Nef125 chimeric oleosin (Fig. 5 and Fig. 6), 12 and 16 kDa for plants expressing the 2F5e and NP peptide chimeric oleosin (Fig. 7). The analysis of the expression of the chimeric oleosins in leaves and siliques of both plants transformed with pBIN19-35S or pBIN19-GM demonstrated that the *Glycine max* oleosin derived regulatory sequences induced tissue-specific expression of foreign oleosin at higher levels than those obtained by the 35S promoter in mature siliques. Moreover, for plants expressing the chimeric oleosins under the control of the 35S promoter a higher accumulation was observed in siliques than in leaves (Fig. 8). The presence of the peptide as oleosin fusion was indicated in the extracts of the plants transformed with Nef125 fusion by a shift in the migration of the oleosin band as compared to the band identified in the extracts of transgenic plants transformed with non-chimeric sunflower oleosin. The presence of the 2F5e peptide

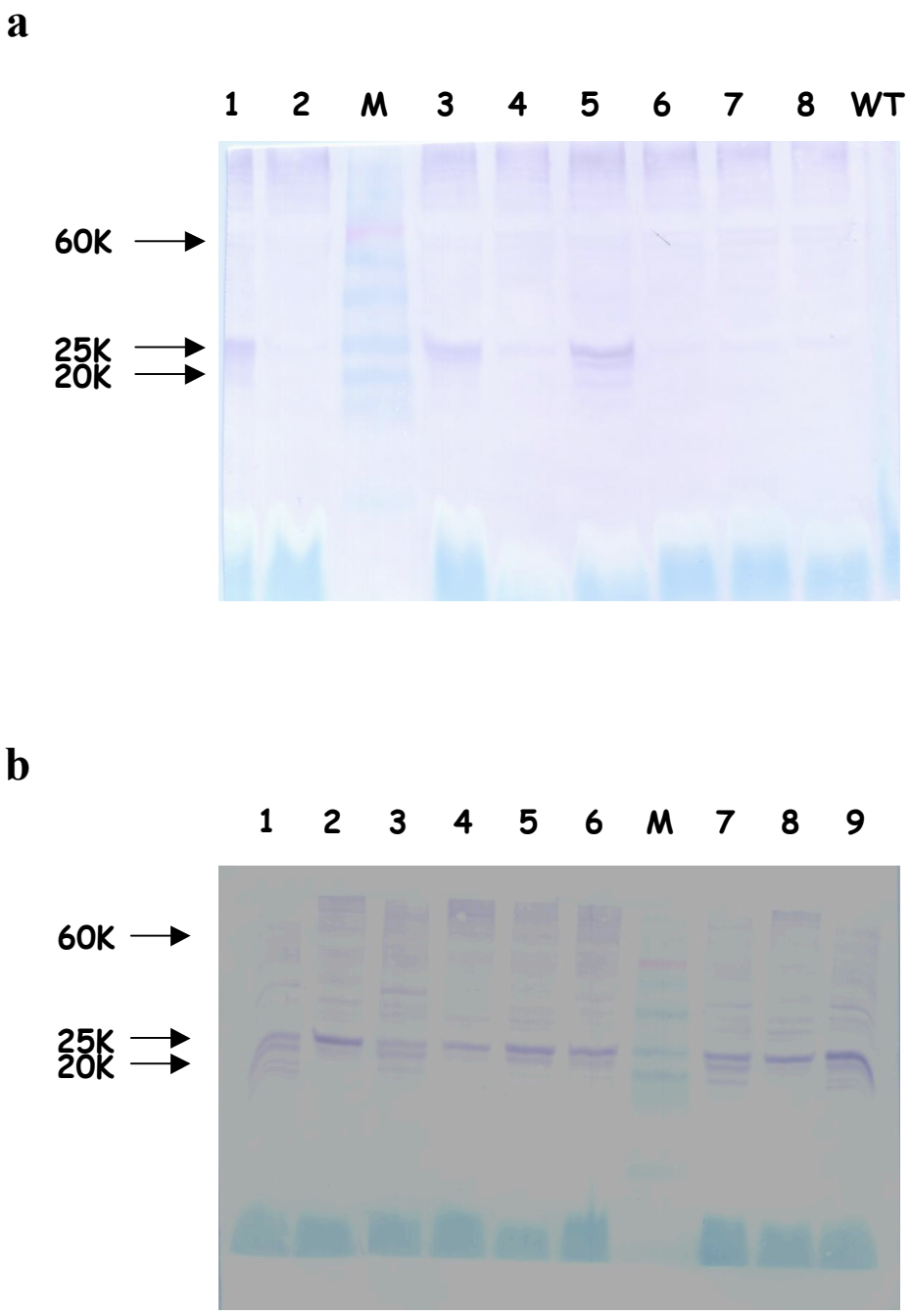

Figure 5. Analysis of the 35S-driven expression of chimeric oleosin in a representative group of *A. thaliana* transgenic plants (T_0).

Western blot analyses of crude extracts of leaves (panel a: lane 1-8) and siliques (panel b: lane 1-9) from T_0 plants transformed with pBIN19-35S-Nef125 performed with an antibody specific for the sunflower oleosin.
M: Pre-stained Molecular Weight Marker (the arrows indicate the molecular weight of the marker bands of interest); WT: leaf protein extracts of w.t. plants.

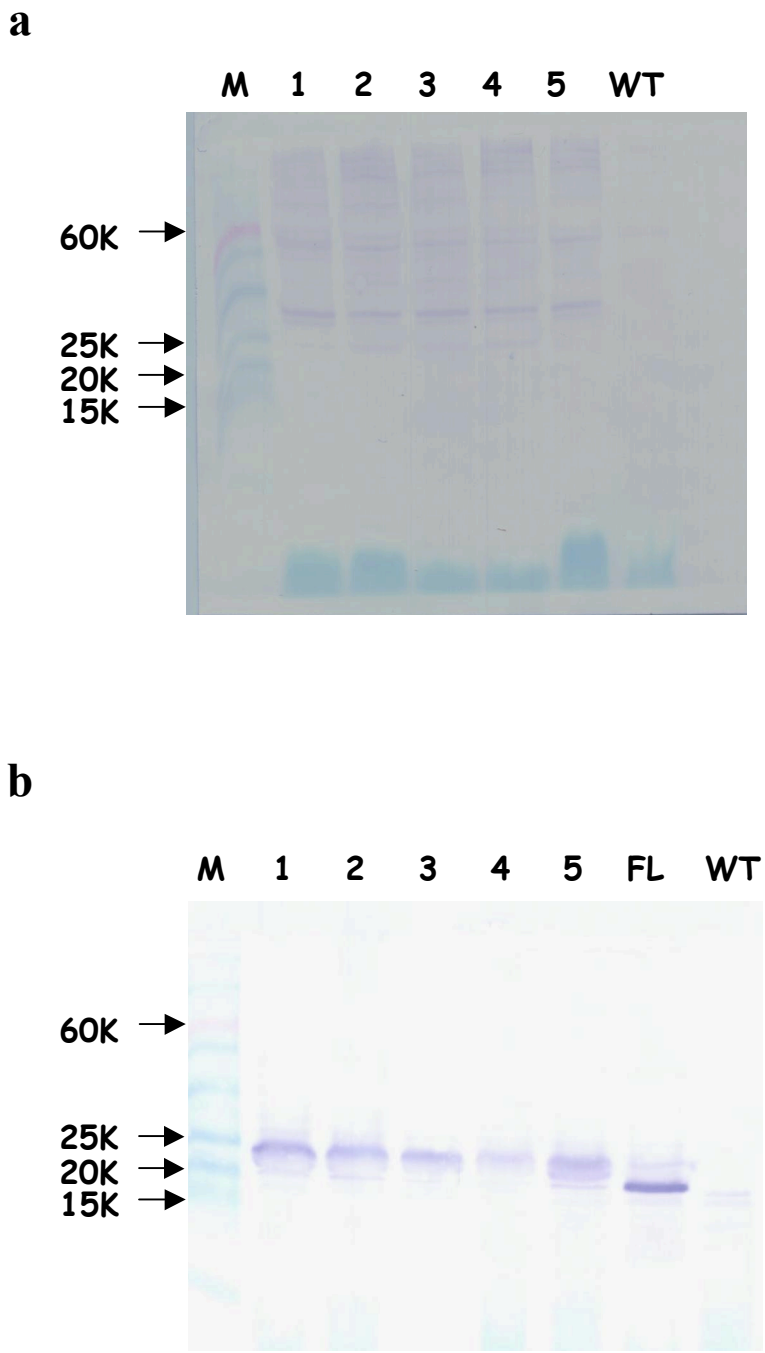

Figure 6. Analysis of the *G. max* oleosin promoter-driven expression of chimeric oleosin in a representative group of *A. thaliana* transgenic plants (T_0).

Western blot analyses of crude extracts of leaves (panel a: lane 1-5) and siliques (panel b: lane 1-9) from T_0 plants transformed with pBIN19-GM-Nef125 performed with an antibody specific for the sunflower oleosin.
M: Pre-stained Molecular Weight Marker (the arrows indicate the molecular weight of the marker bands of interest); WT: leaf protein extracts of w.t. plants.

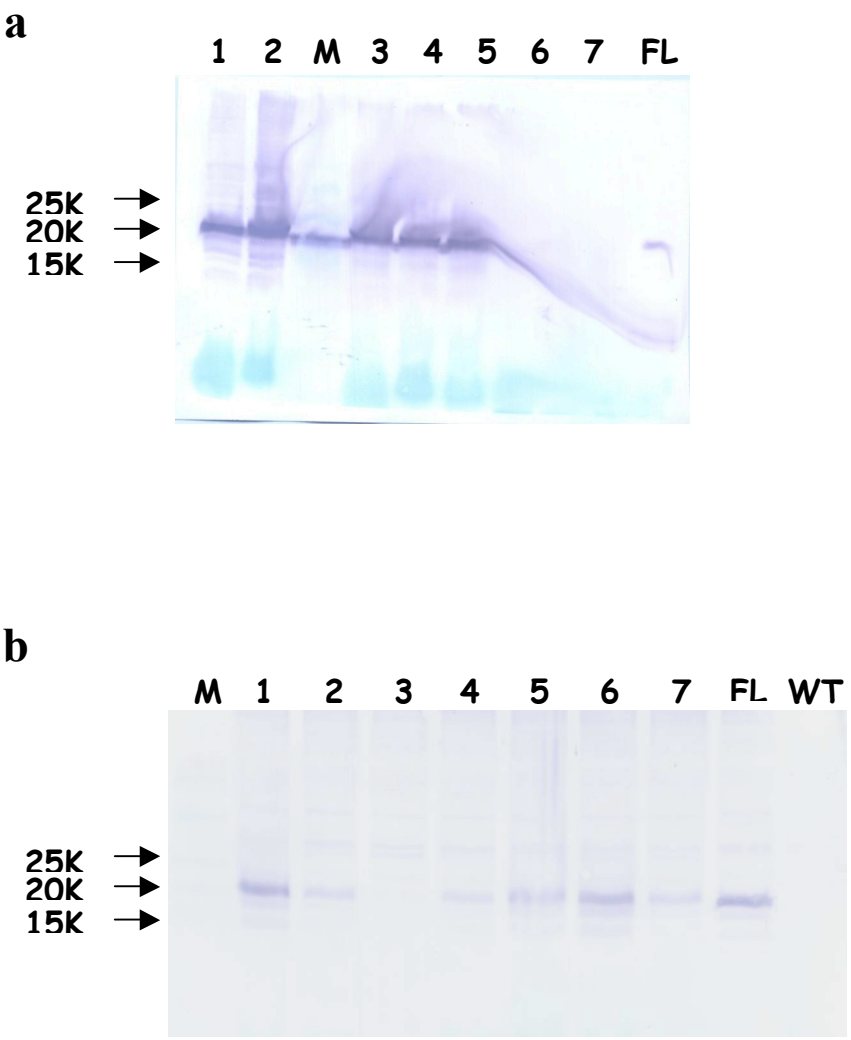

Figure 7. Analysis of the *G. max* oleosin promoter-driven expression of chimeric oleosin in a representative group of *A. thaliana* transgenic plants (T_0).

Western blot analyses of crude extracts of siliques from T_0 plants transformed with pBIN19-GM-NP (panel a: lanes 1-7) or with pBIN19-GM-2F5e (panel b: lanes 1-7) performed with an antibody specific for the sunflower oleosin.

M: Pre-stained Molecular Weight Marker (the arrows indicate the molecular weight of the marker bands of interest); FL: protein extracts from *A. thaliana* plants transformed with the non-chimeric sunflower oleosin gene driven by 35S promoter; WT: siliques protein extracts of w.t. plants.

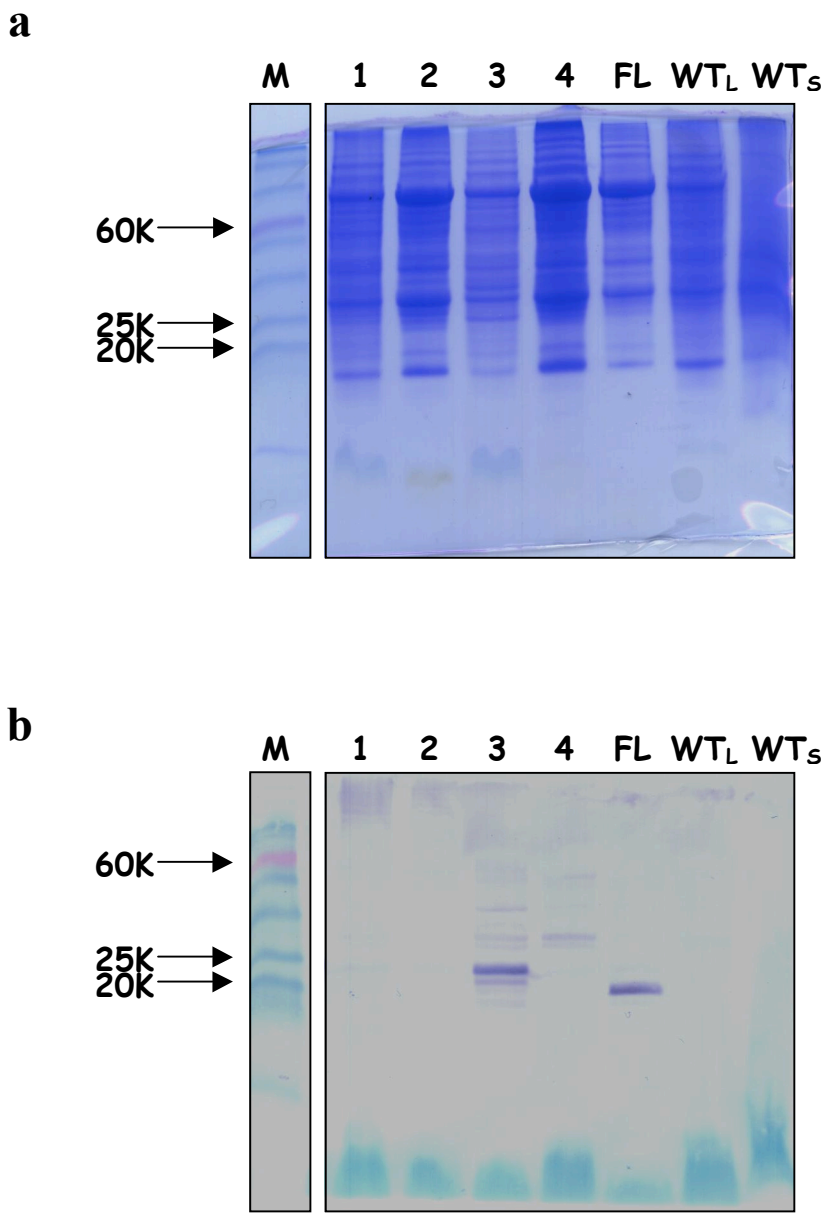

Figure 8. Analysis of the accumulation of chimeric oleosins driven by 35S or *G. max* oleosin promoter in leaves and siliques of a representative group of *A. thaliana* transgenic plants (T_0).

SDS-PAGE (panel a) and Western blot analyses (panel b) of crude extracts of leaves (lanes 1 and 3) and siliques (lanes 2 and 4) from T_0 plants transformed with pBIN19-35S-Nef125 (lanes 1 and 2) or pBIN19-GM-Nef125 (lanes 3 and 4).

M: Pre-stained Molecular Weight Marker (the arrows indicate the molecular weight of the marker bands of interest); FL: protein extracts from *A. thaliana* plants transformed with the non-chimeric sunflower oleosin gene under the control of 35S promoter; WT_L and WT_S: leaf and siliques protein extracts of w.t. plants, respectively.

was instead verified by the use of the monoclonal antibody 2F5 (Buchacher et al., 1994) (Fig. 9). No reagents were available to reveal the expression of the NP peptide.

4.3.3 Genetic analysis of T_0 plants to verify transgene integration

Leaves from T_0 plants were used to extract genomic DNA and perform PCR analysis using transgene specific primers (5'-*Not*OleoF and 3'-*Not*125R, 3'-*Not*2F5R or 3'-*Not*NPR primers, Tab. 3). The integration of the transgene was confirmed by the presence of a 800 base pair (bp) band (Fig. 10a). Transgene integration was verified also through Southern blot analysis of BamHI-digested genomic DNA using as probe the linearized pOBP plasmid. After digestion the genomic DNA samples were run on 0.8% agarose gel. Then, the gel was blotted and the membrane hybridised. The hybridization with the probe showed the presence of a single band as expected for the plants integrating a single copy of the transgene (Fig. 10b). The segregation test performed sowing the seeds collected from T_0 plants confirmed the results obtained with southern blot. By these analyses 3 independent T_1 lines of *A. thaliana* plants for each construct integrating a single copy of the transgene were identified (not shown) and grown to T_2 generation.

4.4 Wild-type and chimeric oil bodies characterization

4.4.1 Comparison of the efficiency of different oil body purification procedures

Oil bodies were purified from the seeds of w.t. and transgenic T_2 plants. To this aim the two protocols available in the literature were compared to define the most effective in removing seed protein contaminants (Tzen et al.,1997; Deckers et al., 2004). Oil bodies were initially extracted from seeds by grinding with pestle and mortar in a cold 0.6 M Sucrose, 10mM sodium phosphate buffer (Step 1, S1) and purified by centrifugation through a 0.4 M sucrose cushion. This is a step common to both protocols and is used to separate the buoyant oil bodies from other cellular components. The extracted oil bodies present in a white fat pad on the top of the cushion were then sequentially washed with 0.1%Tween (Step 2 Tzen, S2T) , 2 M NaCl (Step 3 Tzen, S3T), and 8M urea (Step 4 Tzen, S4T) (Tzen et al, 1997); or following Deckers' protocol twice with Tris-HCl pH 7.5 (S2D and S3D) and once with 0.1 M Na_2CO_3 pH 11 (S4D). The quality of the two preparations was analysed by SDS-PAGE. The results clearly showed that the protocol developed by Tzen was much more effective in removing

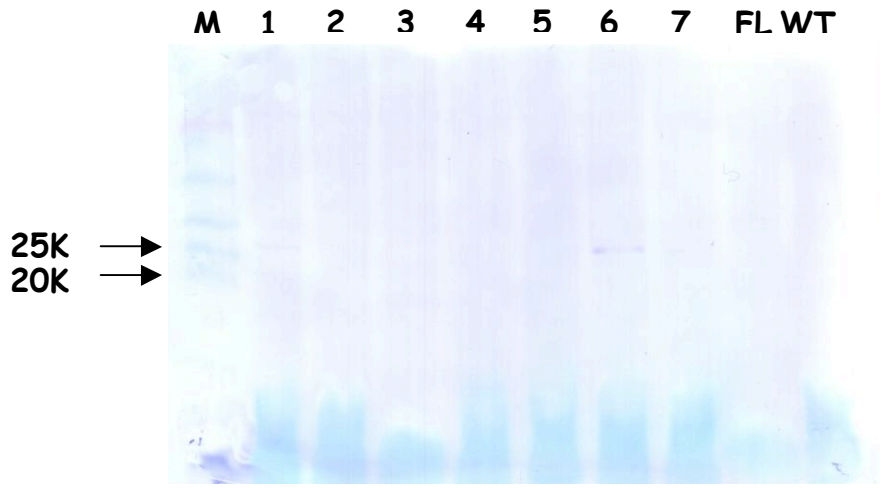

Figure 9. **Analysis of the expression of the 2F5e-chimeric oleosins in *A. thaliana* transgenic plants (T_0).**
Western blot analysis of crude extract of siliques from T_0 plants transformed with pBIN19-GM-2F5e (lanes 1-7) performed with the 2F5 monoclonal antibody.
M: Pre-stained Molecular Weight Marker (the arrows indicate the molecular weight of the marker bands of interest); FL: protein extracts from *A. thaliana* plants transformed with the non-chimeric sunflower oleosin gene under the control of 35S promoter; WT: siliques protein extracts of w.t. plants.

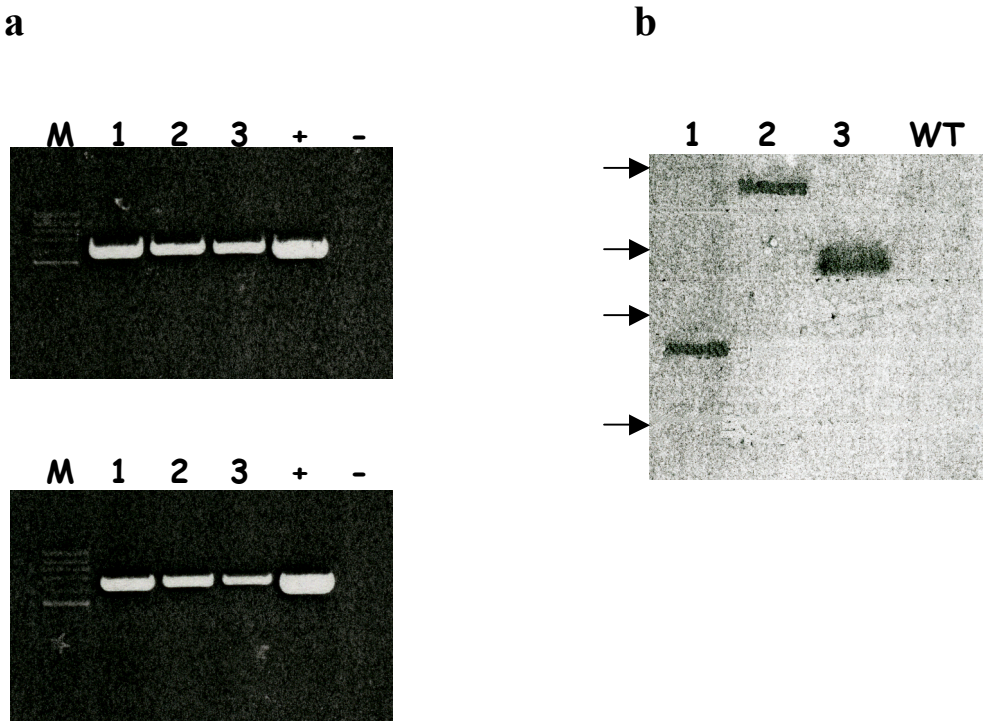

Figure 10. Genetic analysis of representative T_0 lines of transformed *A. thaliana* plants.
a) PCR analysis performed on genomic DNA extracted from the leaves of Nef125 and 2F5e T_0 plants using transgene specific primers (lanes 1-3; top: 5'-NotOleoF and 3'-Not125R; bottom: 5'-NotOleoF and 3'-Not2F5R)(M1:100bp DNA ladder Fermentas (10 fragments: 1000, 900, 800, 700, 600, **500**, 400, 300, 200, 100 bp); +:pKMS2-Nef125 (top) or pKMS2-2F5e (bottom); -: PCR performed without template.
b) Southern blot analysis performed on *BamH*I-digested genomic DNA extracted from the leaves of Nef 125 (lanes 1 and 2) and 2F5e (lane 3) lines of PCR-positive plants or w.t. *A. thaliana cv* Columbia plants (w.t.), using as probe the linearized pOBP plasmid carrying the cDNA sequence of the sunflower oleosin. The arrows indicate the molecular size of the marker bands (4361, 2322, 2027, 564 bp).

contaminants compared to the protocol of Deckers that gave a protein profile similar to the total seed extract (S1) (Fig. 11 and 12). The SDS-PAGE analysis of the intermediate steps of purification obtained using the protocol of Tzen showed that the 8 M urea treatment was crucial to remove proteins migrating between 21 and 50 kDa and that only six bands were clearly visible at the final step within the 6–55 kDa range (Fig. 12).

4.4.2 Characterization of protein composition of w.t. and chimeric oil bodies by MS

The identification of the proteins associated with w.t. oil bodies using the protocol of Tzen (sample S4T) was initially performed by MS analysis of the proteins separated through SDS-PAGE. Before SDS-PAGE protein samples were treated to remove the lipid contaminant that would interfere with protein migration. The gel was sliced as shown in Fig. 13 and the proteins within each slice, along each lane, digested with trypsin. Each peptide mixture was then submitted to LC–MS/MS. The results obtained by the analysis of most evident Coomassie-stained bands led essentially to the identification of six isoforms of oleosins (slices 12 and 13, S4T) with molecular mass ranging from 14.8 to 21.2 kDa (Tab. 4) together with proteins previously annotated as associated to oil bodies (Jolivet et al., 2004) (two *A. thaliana* seed gene (ATS) isoforms (At4g26740; At5g55240) (slice 9, lane S4T); 11-β-hydroxysteroid dehydrogenase-like protein (At5g50600) (slice 6, lane S4T); predicted GPI (Glycosylphosphatidylinositol)-anchored protein (At1g54860) (slice 12, lane S4T); β-Tonoplast Intrinsic Protein (TIP) (At1g17810) (slice 11, lane S4T)). However, a more comprehensive study revealed that the sample was more complex than expected when also slices not visibly stained by Coomassie were analysed. The merging of all the results obtained by Mascot searching (*www.matrixscience.com*) of the spectra corresponding to the tryptic peptides generated across S4T lane identified 75 proteins with a score sufficient for protein identification (at least two unique peptides with an ion score >35 for each protein) (Tab. 5). Among these proteins structural components of the ribosome as well as proteins involved in embryonic development (Late Embryogenesis Abundant protein isoforms, LEA; *Arabidopsis thaliana* Seed protein isoforms, ATS), enzymatic activity (oxidoreductase, dehydrogenase, disulfide isomerase activity), transmembrane transporter activity (β-TIP; Pore Homologue protein; Adenylate translocator) and GTP-binding proteins were identified. A comparison between the list of the total proteins identified in S1 and S4T was carried out to monitor proteins that were removed or enriched across the purification procedure. The number of total

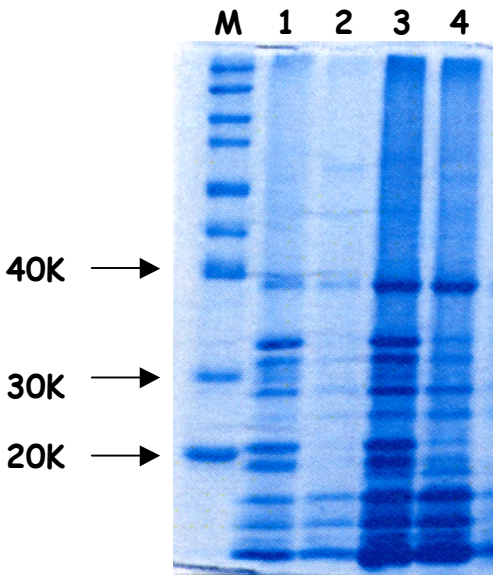

Figure 11. SDS-PAGE analysis of wild type and chimeric oil bodies prepared according to the protocols developed by Deckers and Tzen.

Protein analysis by SDS-PAGE and Coomassie-staining of w.t. (lanes 1 and 2) or chimeric (lanes 3 and 4) oil bodies prepared according to Deckers (lane 1 and 3) or Tzen (lane 2 and 4).

M: Novex Sharp Protein Standard (Invitrogen).

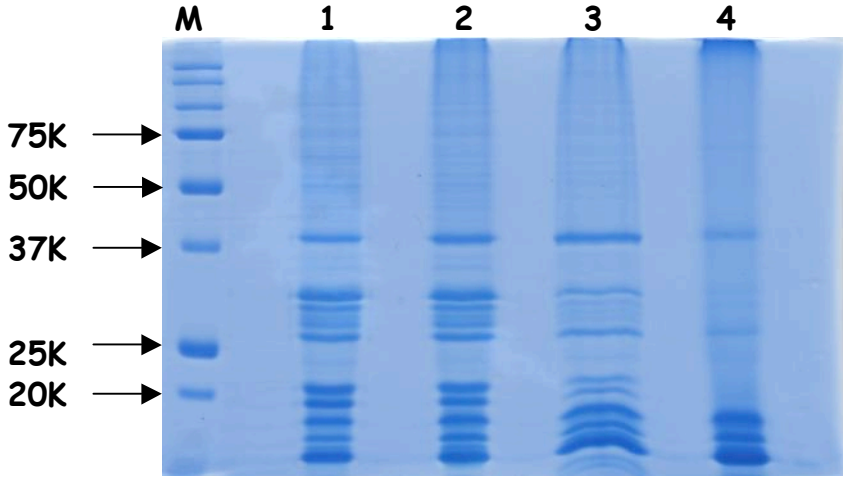

Figure 12. SDS-PAGE analysis of proteins associated to w.t. oil bodies purified according to the procedure developed by Tzen.

Total seed extracts (S1) (lane 1) were centrifuged through a sucrose cushion to isolate oil bodies according to the protocol developed by Tzen. These were sequentially washed with 0.1% Tween (S2T) (lane 2), 2M NaCl (S3T) (lane 3) and finally 8M urea (S4T) (lane 4).

M: Precision Plus All blue Protein Standard (Bio-Rad).

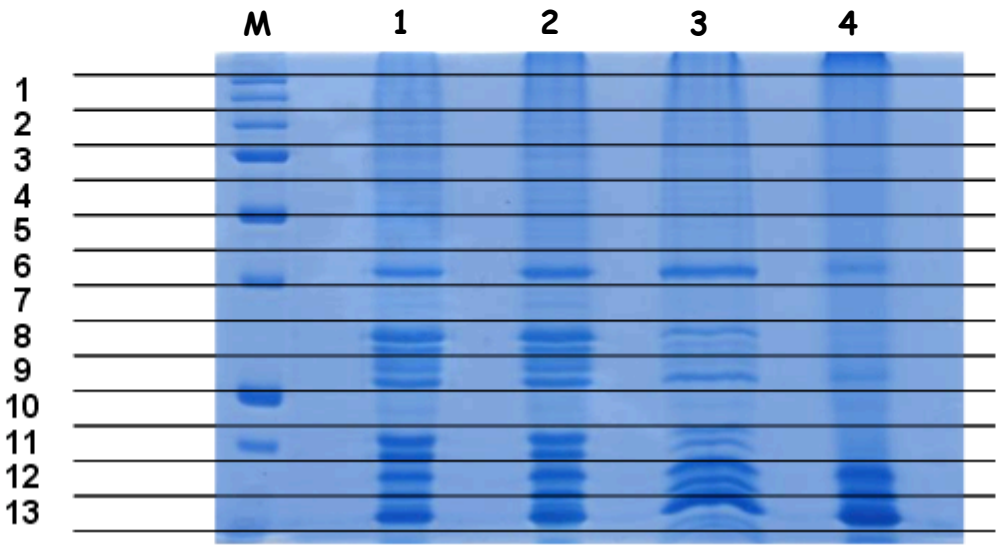

Figure 13. Shematic representation of SDS-PAGE slicing for MS analysis of w.t oil bodies protein composition.

Table 4. List of oleosin isoforms identified by MS analysis in the w.t. oil body preparation

Accession Number	Protein description	Protein MW	Protein pI	Rename[a]
At3g27660	oleosin isoform	20300	6.92	S2
At4g25140	oleosin 18.5K	18558	9.43	S3
At3g01570	putative oleosin	19742	7.11	S1
At5g40420	oleosin	21266	9.36	S4
At5g51210	oleosin-like	14843	9.86	n.i.
At1g48990	oleosin like protein	18432	9.89	n.i.

MW: Molecular Weight; pI: Isoelectric point; n.i.: not identified by Jolivet et al., 2004.
[a] Nomenclature used by Kim et. al, 2002 and Jolivet et al., 2004.

proteins decreased from 361 (S1) to 75 (S4T), with 44 proteins shared between the two samples (Tab. 5). The shared proteins were grouped in 6 categories (structural constituents of ribosome; enzyme activity; nutrient reservoir; embryonic development; oleosins; other function) referring to the Gene Ontology (GO) database (*http://www.ncbi.nlm.nih.gov/sites/entrez*). To quantify the changes in the profiles of common proteins the "exponentially modified Protein Abundance Index" (emPAI) was utilised (Ishihama et al., 2005). The emPAI value, calculated as the ratio between the number of experimentally observed peptides for a protein to the number of peptides generated by its *in silico* digestion, is a valuable index for protein quantification in a complex sample. Indeed, it has been established that the number of unique peptides of a protein detected by MS correlate with the abundance of that protein. Protein abundance, expressed as molar fraction percentages (% mol), was calculated as the ratio between the emPAI value for each protein and the summation of emPAI values for all the identified proteins within each sample ((emPAI/ΣemPAI) X 100). From this analysis 12 proteins out of 44 at least doubled their content (Tab. 6). These proteins included oleosins and the embryo-specific protein 3. The majority of the proteins decreasing in abundance were the seed-storage proteins (12S cruciferins α and β sub-units) and the structural constituents of the ribosome. The remaining proteins, unmodified between the two samples were mainly those classified as enzymes and those involved in embryonic development. A schematic representation of protein content variation is reported in Fig. 14. The 31 proteins identified only in S4T, apart for the 14.8 kDa (At5g51210) oleosin isoform, were mainly derived from the nucleus (histone 2 and histone 4), from the endomembrane system (cytochrome b5, GDSL-motif lipase/hydrolase-like protein, calnexin-like protein) and from plastids. Besides the characterization of the total protein content in w.t. oil body preparations, LC-MS/MS analysis was also performed to sequence the sunflower oleosin present in the oil bodies preparations obtained from transgenic seeds. Lipids were removed from the protein samples and proteins were separated by SDS-PAGE. The gel was Coomassie-stained and sliced across the whole lane. Tryptic digestion of the proteins in each band was then performed and the peptide mixture was sprayed into the mass analyser. The spectra obtained were searched with Mascot using a modified *A. thaliana* FASTA database in which the sunflower oleosin and the chimeric sunflower oleosins sequences were included. The presence of the epitope sequence was verified in one gel band for the Nef125 polypeptide and for NP. In particular, for Nef 125 2 tryptic peptides mapping in the Nef-derived polypeptide were identified, as well as 6 peptides from the sunflower oleosin sequence with a protein coverage of 38% (Fig.15). The NP peptide was also detected

Table 5. List of proteins identified by MS analysis in purified oil bodies (S4T)

Protein description	Protein Mass	Protein score	Protein pI	emPAI
11-beta-hydroxysteroid dehydrogenase-like	39518	4471	5.92	25.57
oleosin isoform	20300	4082	6.92	11.79
oleosin, 18.5K	18558	659	9.43	9.54
GTP binding / phospholipase activator/ protein binding	20667	294	6.43	3.68
putative oleosin	19742	1110	7.11	4.01
oleosin	21266	2209	9.36	3.47
embryo-specific protein 1 (ATS1)	28134	630	5.81	2.63
oleosin-like	14843	100	9.86	1.88
CRU3 (CRUCIFERIN 3)	58541	2614	6.53	1.84
ATHSD5 (HYDROXYSTEROID DEHYDROGENASE 5)	43528	343	6.41	1.79
histone H4	11402	97	11.48	1.77
40S ribosomal protein S14	16304	131	10.60	1.63
40S ribosomal protein s14 like	16320	124	10.60	1.63
ribosomal protein S14 like protein	16285	111	10.60	1.63
CRA1 (CRUCIFERINA)	52905	917	7.68	1.51
40S ribosomal like protein (S18)	17591	165	10.54	1.47
ribosomal protein L9, putative	22118	55	9.48	1.47
embryo-specific protein 1; Ca2+-binding EF-hand protein-like	27972	635	5.62	1.37
cupin family protein	55200	485	6.64	0.94
oleosin like protein	18432	95	9.89	0.91
pore protein homolog	18574	250	6.97	0.90
hypothetical protein, 3~ partial	27687	146	4.53	0.79
Putative 40S ribosomal protein S15A	14852	61	9.90	0.70
cytochrome b5	15132	53	5.11	0.68
embryo-specific protein 3 (ATS3)	23410	201	6.27	0.67
putative ribosomal protein s19 or s24	15363	66	10.70	0.67
oxidoreductase	31596	111	6.11	0.66
putative cruciferin 12S seed storage protein	50869	525	6.52	1.05
histone H2B like protein	15723	148	10.05	0.65
LEA domain-containing protein	32653	82	5.28	0.64
ras-related small GTP-binding protein RAB1c	22532	145	5.27	1.03
putative 40S ribosomal protein S15	17118	117	10.34	0.59
40S ribosomal protein S3	27612	117	9.57	0.55
putative aspartic protease	56397	1013	5.97	0.54
AWPM-19-like membrane family protein	19947	59	9.74	0.49
benzodiazepine receptor-related	21277	87	9.34	0.45
predicted GPI-anchored protein	21897	57	6.08	0.44
GTP-binding protein, ara-5	22805	61	5.02	0.42
40S ribosomal protein - like	23079	83	10.17	0.41

Table 5. Continued

Protein description	Protein Mass	Protein score	Protein pI	emPAI
GTP-binding RAB2A like protein	23378	79	6.96	0.41
GAPC-2	37004	112	6.67	0.39
LEA76 homologue type2	24172	77	8.83	0.39
GTPase AtRAB8	24095	61	8.35	0.39
putative aspartic proteinase	55320	362	5.37	0.34
ribosomal protein S3a homolog	27503	113	9.57	0.34
RNA recognition motif (RRM)-containing protein	27218	80	5.46	0.34
translocon-associated protein alpha (TRAP alpha) family protein	28264	136	5.08	0.33
globulin-like protein	56983	131	5.44	0.33
β-Tonoplast Intrinsic Protein	28222	90	6.54	0.33
Photosystem II chlorophyll-binding protein PsbS	27990	77	9.25	0.33
calnexin - like protein	60790	137	4.81	0.31
putative histone H2B	15072	146	10.05	0.30
histone H2B - like protein	15215	146	10.08	0.30
prohibitin, putative	30619	81	6.93	0.30
FLA15	48157	96	6.22	0.29
FLA16	49184	87	6.35	0.28
putative histone H2B	16519	146	10.00	0.27
hypothetical protein (Chloroplast, mitochondrion, membrane)	17932	80	8.54	0.25
AtHVA22b-like protein (regulatory protein)	18730	84	6.40	0.24
GDSL-motif lipase/hydrolase-like protein	37039	72	5.83	0.24
putative vacuolar sorting receptor	71762	79	5.30	0.19
40S ribosomal protein S5	23090	65	9.69	0.19
putative 40S ribosomal protein S5	23021	65	9.66	0.19
Soul-1	23700	71	7.74	0.18
unknown protein (endoplasmic/plasma membrane)	24549	61	9.32	0.18
putative seed maturation protein	23869	58	6.97	0.18
FLA18	50988	81	6.72	0.17
FLA17	50491	75	6.21	0.17
manganese superoxide dismutase-like protein	26932	84	6.25	0.16
putative seed storage protein (vicilin-like)	55898	67	5.83	0.16
putative disulfide isomerase	55852	67	4.81	0.16
putative prohibitin	31617	54	9.65	0.14
late embryogenesis abundant domain-containing protein	67155	76	5.78	0.13
putative non-green plastid inner envelope membrane protein	36735	82	7.00	0.12
-chloroplast genome- cytochrome f	35449	61	8.34	0.12
Putative vicilin storage protein (globulin-like)	84041	53	6.39	0.10

pI: Isoelectric point; emPAI: exponentially modified Protein Abundance Index; in **bold characters** are indicated S1 and S4T shared proteins

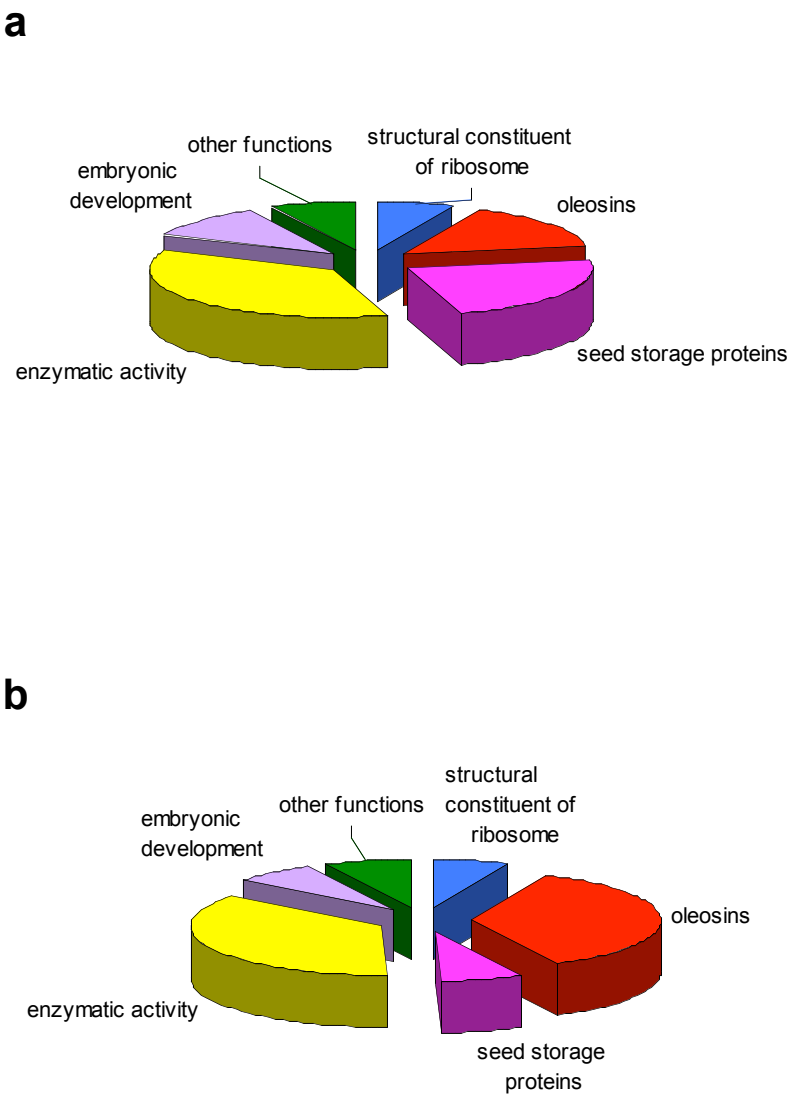

Figure 14. Characterization of protein content of oil bodies before and after Tzen purification procedure. Variation of the abundance of protein groups (referring to Gene Ontology database) associated to oil bodies before (a) and after (b) purification.

Table 6. List of the proteins shared between the two set of data S1 and S4T and doubling their content

Accession N°	Protein description	Protein MW	S1 Protein content (% mol)	S4T Protein content (% mol)
At3g52580	ribosomal protein S14 like protein	16285	0.8084	1.9759
At1g48990	oleosin like protein	18432	0.1190	1.103
At4g25140	oleosin, 18.5K	18558	5.9815	11.5650
At3g01570	putative oleosin	19742	2.5394	4.8611
At3g27660	oleosin isoform	20300	4.716	14.2926
At5g40420	oleosin	21266	2.1773	4.2065
At2g47770	benzodiazepine receptor-related	21277	0.1041	0.5455
At5g07190	embryo-specific protein 3 (ATS3)	23410	0.4860	0.8122
At1g17810	β-TIP	28222	0.1636	0.4
At2g21160	alpha subunit precursor TRAP complex/Signal sequence receptor alpha subunit (SSR-alpha)	28264	0.1636	0.4
At4g10020	ATHSD5 (HYDROXYSTEROID DEHYDROGENASE 5)	43528	1.334	2.16
At1g62290	putative aspartic protease	56397	0.3273	0.6546

MW: Molecular Weight; % mol: Protein content expressed as (emPAI/Σ emPAI) X 100

a

Peptide sequence	Monoisotopic ion [M]	[M+H]+	[M+H]++
EGGKPR	642.34	643.34	322.17
LTHPQR	750.41	751.41	376.205
LVPVEPE	781.41	782.41	391.705
QQQGPSTGK	929.45	930.45	465.725
MATTTYDR	957.42	958.42	479.71
HDQHTGDR	964.4	965.4	483.2
STMSVPVQR	1003.5	1004.5	502.75
QNYTPGPGIR	1101.55	1102.55	551.775
LQDVGEYTGQK	1236.59	1237.59	619.295
YPLTFGWCYK	1276.59	1277.59	639.295
HHVTTTQPQYR	1366.67	1367.67	684.335
IQHTAHEMCDQGQGQGQGGGK	2165.93	2166.93	1083.965

b

Protein description	Peptide sequence	Ion score
Sunflower oleosin	LQDVGEYTGQK	59
	GKLQDVGEYTGQK	60
	HHVTTTQPQYR	32
	STMSVPVQR	44
	HDQHTGDR	57
	IQHTAHEMCDQGQGQGQGGGK	75
Nef125 polypeptide	QNYTPGPGIR	36
	YPLTFGWCYK	49

c

OBP-Nef125

MATTTYDRHHVTTTQPQYRHDQHTGDRLTHPQRQQQGPSTGKIMVIMALLPITGILFGLAGITLVGTVI
GLALATPLFVIFSPVIVPAMIAIGLAVTGFLTSGTFGLTGLSSLSYLFNMVRRSTMSVPVQRDYVKGKL
QDVGEYTGQKTKDLGQKIQHTAHEMCDQGQGQGQGGGKEGRKEGGKPRQNYTPGGIRYPLTFGWCYKLV
PVEPE

Figure 15. Identification by LC-MS/MS analysis of peptides generated by tryptic digestion of Nef125 chimeric oleosin associated to oil bodies.
a) List of the peptides generated by *in silico* digestion with trypsin of the Nef125 chimeric sunflower oleosin. **b)** List of tryptic peptides experimentally identified by LC-MS/MS. **c)** Amino acid sequence of Nef125 chimeric oleosin showing in red sunflower oleosin-derived and in blue Nef125 polypeptide-derived fragment peptides identified by LC-MS/MS. M: Monoisotopic ion; [M+H]+, [M+H]++: protonated and doubly protonated ions.

(although it contained one or both the methionines in an oxidised form), as well as 5 peptides bearing into the sunflower oleosin sequence (Fig.16). The sequence coverage for this chimeric protein was 34%. No fragment peptides including 2F5e were found by MS/MS analysis (Fig. 17). However, 5 peptides derived from the sunflower oleosin sequence were identified giving a protein coverage of 28%. The *in silico* analysis of the tryptic peptides that can be generated by digestion with trypsin of 2F5e chimeric sunflower oleosin was also performed confirming that the fragment peptides bearing the epitope sequence are quite small and probably undetectable by the MS analyser.

4.5 Evaluation of the immune responses elicited by immunisation with chimeric oil bodies

4.5.1 Preparation and characterization of oil bodies for immunisation

As urea is toxic for mice, oil bodies preparations for *in vivo* experiments were carried out using a protocol similar to that developed by Deckers, using sterile 0.9% NaCl for washing instead of sodium carbonate buffer. The protein profile of the oil body preparations, analysed by SDS-PAGE, showed a pattern of bands similar to that obtained with Deckers protocol (Fig. 18). To prevent bacterial contamination all the procedure was performed under sterile conditions and oil body preparations were aliquoted and flash-frezeed. Quantification of the total proteins in oil body preparations was performed after lipid removal by using the Bradford assay solution and BSA (Bovine Serum Albumin) as a standard. Due to the unavailability of reagents able to detect the presence of the peptide in the preparation, it was not possible to determine the NP peptide doses injected with these oil bodies quantities.

The attempt to detect and quantify 2F5e associated to purified chimeric oil bodies was carried out by direct ELISA performed using the Mab 2F5 for the detection. Four different dilutions of the chimeric and non chimeric oil bodies preparation were distributed in triplicates into the wells of Maxisorp plates in parallel to different dilutions of a synthetic polypeptide (QTQQEKNEQELLELDKWASL) that embodies 2F5e in the sequence. From this analysis a positive signal could be detected only in the chimeric oil bodies sample. However, looking at the two regression lines obtained by titrating the chimeric oil body sample and the HIV-1 Env-derived synthetic peptide, it was evident that they were not parallel (Fig. 19a). In order to evaluate if this could be due to the interference of some components of the oil body preparation, a second ELISA was performed by titrating chimeric and non chimeric oil bodies

a

Peptide sequence	Monoisotopic ion [M]	[M+H]+	[M+H]++
LTHPQR	750.41	751.41	376.205
QQQGPSTGK	929.45	930.45	465.725
MATTTYDR	957.42	958.42	479.71
HDQHTGDR	964.4	965.4	483.2
STMSVPVQR	1003.5	1004.5	502.75
LQDVGEYTGQK	1236.59	1237.59	619.295
HHVTTTQPQYR	1366.67	1367.67	684.335
EGGKPSASNENMETM	1580.64	1581.64	791.32
IQHTAHEMCDQGQGQGQGGGK	2165.93	2166.93	1083.965

b

Protein description	Peptide sequence	Ion score
Sunflower oleosin	STMSVPVQR	48
	LQDVGEYTGQK	49
	HHVTTTQPQYR	37
	IQHTAHEMCDQGQGQGQGGGK	80
	DLGQKIQHTAHEMCDQGQGQGQGGGK	25
NP peptide	EGGKPSASNENMETM	61
	EGGKPSASNENMETM	35

c

OBP-NP

MATTTYDRHHVTTTQPQYRHDQHTGDRLTHPQRQQQGPSTGKIMVIMALLPITGILFGLAGITLVGTVIGL
ALATPLFVIFSPVIVPAMIAIGLAVTGFLTSGTFGLTGLSSLSYLFNMVRRSTMSVPVQRDYVKGKLQDVG
EYTGQKTKDLGQKIQHTAHEMCDQGQGQGQGGGKEGRKEGGKPSASNENMETM

Figure 16. Identification by LC-MS/MS analysis of peptides generated by tryptic digestion of NP chimeric oleosin associated to oil bodies.

of the peptides generated by *in silico* digestion with trypsin of the NP chimeric sunflower oleosin. **b)** List of tryptic peptides experimentally identified by LC-MS/MS. **c)** Amino acid sequence of NP chimeric oleosin showing in red sunflower oleosin-derived and in blue NP polypeptide-derived fragment peptides identified by LC-MS/MS. M: Monoisotopic ion; [M+H]+, [M+H]++: protonated and doubly protonated ions.

a

Peptide sequence	Monoisotopic ion [M]	[M+H]+	[M+H]++
ASELDK	661.32	662.32	331.66
LTHPQR	750.41	751.41	376.205
QQQGPSTGK	929.45	930.45	465.725
MATTTYDR	957.42	958.42	479.71
HDQHTGDR	964.4	965.4	483.2
STMSVPVQR	1003.5	1004.5	502.75
LQDVGEYTGQK	1236.59	1237.59	619.295
HHVTTTQPQYR	1366.67	1367.67	684.335
IQHTAHEMCDQGQGQGQGGGK	2165.93	2166.93	1083.965

b

Protein description	Peptide sequence	Ion score
Sunflower oleosin	STMSVPVQR	36
	LQDVGEYTGQK	53
	HHVTTTQPQYR	26
	IQHTAHEMCDQGQGQGQGGGK	68

c

OBP-2F5e

MATTTYDRHHVTTTQPQYRHDQHTGDRLTHPQRQQQGPSTGKIMVIMALLPITGILFGLAGITLVGTV
IGLALATPLFVIFSPVIVPAMIAIGLAVTGFLTSGTFGLTGLSSLSYLFNMVRRSTMSVPVQRDYVKG
KLQDVGEYTGQKTKDLGQKIQHTAHEMCDQGQGQGQGGGKEGRKEGGKASELDKWA

Figure 17. Identification by LC-MS/MS analysis of peptides generated by tryptic digestion of 2F5e chimeric oleosin associated to oil bodies.
a) List of the peptides generated by *in silico* digestion with trypsin of the 2F5e chimeric sunflower oleosin. **b)** List of tryptic peptides experimentally identified by LC-MS/MS. **c)** Amino acid sequence of 2F5e chimeric oleosin showing in red sunflower oleosin-derived fragment peptides identified by LC-MS/MS. M: Monoisotopic ion; [M+H]+, [M+H]++: protonated and doubly protonated ions.

using the sunflower oleosin-specific polyclonal antibody. As expected, the two regression lines were parallel in this case (Fig. 19b). However, this result just indicated that the two samples contained the same amount of sunflower oleosins. Unfortunately, up to now it has not been possible to estimate the non-chimeric sunflower content.

4.5.2 Enhanced frequency of ASNENMTEM-specific IFN-γ secreting cells after *in vivo* administration of plant-produced NP peptide chimeric oil bodies

Six week-old C57BL/6J female mice in groups of 4 were immunized subcutaneously (s.c.) into the tail base at day 0 and 14 with a single dose of chimeric oil bodies carrying the sunflower oleosin fused to the NP-derived peptide ASNENMETM containing a total of 75 μg of proteins. Control mice were immunized with sterile saline alone, with the synthetic peptide ASNENMETM (50 μg/inoculum) (>95% purity; Sigma, St. Louis, MO), or with oil bodies preparations carrying the non chimeric sunflower oleosin containing a total of 75 μg of proteins. Immonogens were emulsified in Incomplete Freund's Adjuvant (IFA), and administered in a final volume of 50 μl. An additional group of mice was immunized with the same dose of chimeric oil bodies in 50 μl sterile saline, to evaluate if oil bodies have intrinsic adjuvant properties. Nineteen days after the second immunisation mice were sacrificed, single cell suspensions prepared from spleens, and IFN-γ ELISPOT performed to measure the responses to the NP peptides ASNENMETM and TYQRTRALV. The TYQRTRALV peptide is the CTL immunodominant peptide of NP in Balb/c mice (H2-Dd haplotype) and is not able to induce the activation of CTL-mediated immune responses in C57BL/6J mice. Figure 20 shows results demonstrating that a good peptide-specific response was induced in mice immunized with the oil bodies preparation carrying the NP chimeric oleosin. The number of spot forming cells (SFC) is comparable to that obtained by classical immunization with a high peptide dose in IFA ($P<0.05$). No peptide-specific responses were obtained by immunizing mice with the chimeric oil bodies preparation without adjuvant and by immunizing mice with the non-chimeric preparation.

4.5.3 Detection of 2F5e-specific antibodies elicited by immunization with chimeric oil bodies in mice sera

Eight week-old C57BL/6J female mice in groups of 4 were immunized subcutaneously (s.c.) into the tail base at day 0 and 14 with a single dose of chimeric oil bodies carrying the

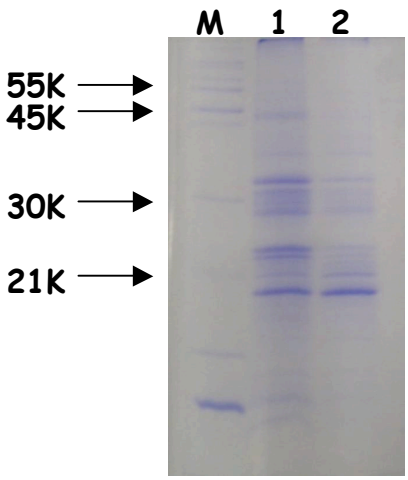

Figure 18. SDS-PAGE analysis of the oil body preparations used to immunize mice.

SDS-PAGE analysis of the oil body preparations washed with 0.9% NaCl used for *in vivo* experiments (lane 1: non chimeric oil bodies; lane 2: chimeric oil bodies). M:Protein Standard RPN5800 (Amersham).

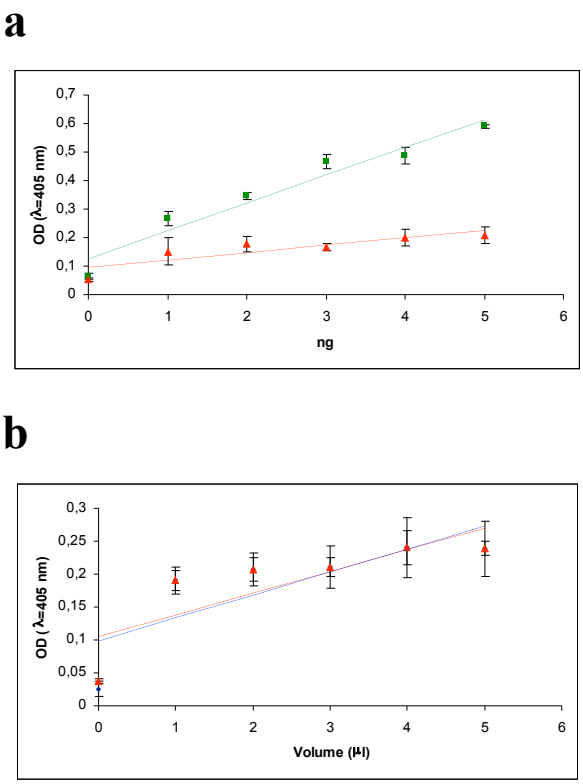

Figure 19. Detection of 2F5e and of sunflower oleosin associated to oil body preparations by ELISA.

a) Regression lines obtained by titrating HIV-1 Env-derived synthetic peptide (green squares) and the 2F5e chimeric oil bodies preparations (red triangles). b) Regression lines obtained by titrating the non chimeric (blue diamonds) and 2F5e chimeric oil bodes (red triangles).

sunflower oleosin fused to 2F5e ELDKWA corresponding to a total of 75 μg proteins with or without IFA. Control mice were immunized with non-chimeric oil bodies containing a total of 75 μg proteins with or without IFA in a final volume of 50 μl. Blood samples were collected from each mouse at day 0 (pre-immune mice), 14, 21 and 42. Direct ELISA coating wells with the HIV-1 Env-derived synthetic peptide was performed to titrate 2F5e specific antibodies in the sera. However, no responder mice could be identified.

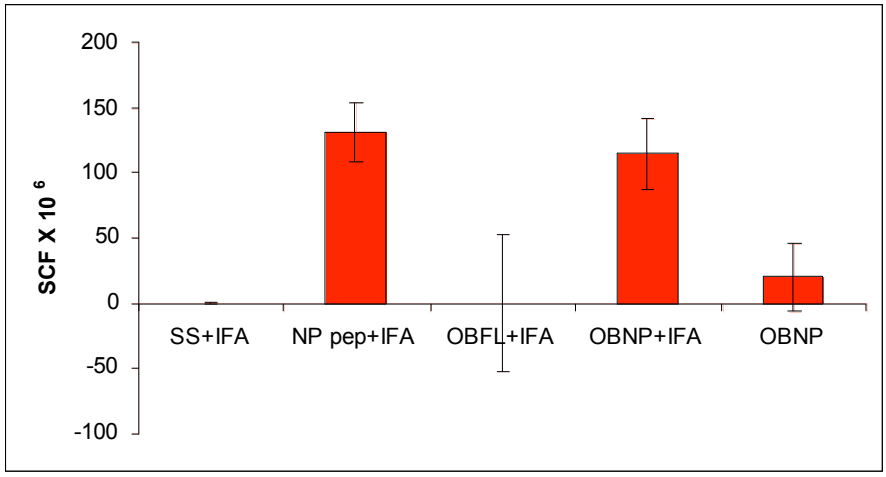

Figure 20. Enhanced frequency of ASNENMTEM-specific IFN-g secreting cells after *in vivo* administration of plant-produced NP peptide chimeric oil bodies.

Evaluation by IFNg ELISPOT assay of the *in vivo* induction of ASNENMETM specific T cell responses by the s.c. immunization with IFA emulsions of apyrogen saline, ASNENMETM peptide (pep), non chimeric or chimeric oil body preprations. An extra-group of mice was immunized with the chimeric oil bodies without IFA (no IFA). Results are expressed in terms of spot forming cells (SFC)/106 of splenocytes in culture on the y-axis. Each value represents the mean of triplicate wells after subtraction of the SFC counted in the respective negative control (unstimulated cells) ± SD . Cells were considered responsive only when the number of SFC in the wells stimulated with the peptide was at least two times that counted in the corresponding unstimulated control. The *in vitro* response to the H-2d-restricted peptide TYQRTRALV is not shown because for all the groups the number of SFC counted in the stimulated wells was equal to that counted in negative controls. Mice immunized with saline, with non chimeric oil bodies or with chimeric oil bodies without IFA do not show a peptide-specific response.

*t-student $p<0.05$

5. DISCUSSION

Several proteins of pharmaceutical interest have been successfully expressed in plant cells through the covalent coupling to proteins within the oil-storing organelles known as oil bodies (Van Roijen and Moloney, 1995a; Parmenter et al., 1995; Nykiforuk et al., 2006). Owing to their physico-chemical properties, these organelles facilitate the purification which leads to good yields of the final product and allows seed-targeted expression of the pharmaceutically important proteins. A similar strategy has been recently developed that is based on the use of plastoglobules i.e. lipo-proteic complexes found into choloroplasts bound to thylakoids. At present, this technology is still in its infancy but could represent a valuable alternative to the use of oil bodies especially when the recombinant protein has to be produced in leaf crops (Vidi et al., 2007).

Studies devoted to the characterization of oil bodies identified in oleosins the main protein component of these organelles. These proteins form a shield around the TAG lipid core. Different oleosin isoforms have been identified in plants and their expression is usually time- and tissue-specific (Kim et al., 2002). Oleosins accumulate at very high levels on the oil body surface and thus to achieve successful targetting of the protein to oil bodies, the gene sequence encoding the heterologous protein/peptide must be fused to the gene sequence encoding an oleosin. Investigations of oleosin topology and mechanisms of targeting to oil bodies through the ER suggest that the hydrophobic central domain of these proteins is highly conserved among different species (Abell et al., 1997; Beaudoin and Napier, 2000). The surface-exposed C-terminus, as well as the N-terminus however, is highly variable and thus amenable to fusion with heterologous protein (Van Roijen and Moloney, 1995b; Beaudoin and Napier, 2000).

The primary goal of the present PhD thesis was to establish *Arabidopsis thaliana* (cv. Columbia) lines genetically modified to independently express three immunogenic peptides derived from human immunodeficiency and influenza viruses (Nef125 polypeptide, 2F5e peptide, NP peptide) as C-terminus fusions with the 19 kDa isoform of the sunflower oleosin. The selected oleosin sequence is one of the two major isoforms present in sunflower, reported to be expressed earlier during seed development and oil deposition (Thoyts et. al, 1995). The selected epitopes are known to be inducers of B cell- or T cell- mediated immune responses in humans and/or mice (Yedwell et al., 1985; Muster et al., 1993; Billaut-Mulot et al., 2001; Crowe et al., 2006).

A. thaliana plants were transformed with pBIN19-based binary vectors in which the chimeric oleosin sequences were under the control of regulatory sequences promoting the expression in all plant tissues (pBIN19-35S) or only in the seeds (pBIN19-GM). The reason for using both vectors was to allow comparison of the efficacy of the regulatory sequences both in terms of expression levels and tissue-specificity. Western analysis with a polyclonal antibody specific to the sunflower oleosin of crude protein extracts of leaves and siliques from plants transformed with pBIN19-35S-based constructs showed that, despite the fact that these constructs should guarantee transgene expression in the entire plant, the expression was more pronounced in siliques. Previous studies on transgenic *A. thaliana* plants ectopically expressing the wild type sunflower oleosin demonstrated that varying levels of protein accumulation in different tissues do not correlate with the corresponding levels of transcript, but do correlate with the diverse accumulation of TAG (Beaudoin and Napier, 2000). On these premises, it may be postulated that in leaves oleosins accumulate in the ER but fail to be transferred to oil bodies that are formed to a lower extent in this non-accumulating oil tissues.
Western analysis with the sunflower oleosin specific polyclonal antibody on extracts of leaves and siliques from plants transformed with the pBIN19-GM-based constructs demonstrated that the oleosin regulatory sequences derived from *Glycine max* tightly controlled the expression that was detectable only in siliques. Moreover, these sequences seem to be more efficient than the 35S promoter as expression levels of the sunflower oleosin in the siliques of plants transformed with pBIN19-GM were generally higher then those obtained from plants transformed with the constructs pBIN19-35S. Classical methods such as 1D PAGE followed by densitometry were found not to be appropriate to determine the precise expression levels of the chimeric sunflower oleosins in T_2 plants transformed with pBIN19-GM-based vectors because of co-migration of the chimeric oleosins with endogenous oleosins and the lipidic nature of the organelles. Little difference in band intensities was observed on Coomassie-stained SDS-PAGE gels between wild type and transgenic extracts. The *G. max*-derived regulatory sequences have been previously demonstrated to induce the expression of the 24 kDa soybean oleosin isoforms A and B in rapeseed plants at Coomassie-staining sensitivity levels without affecting targeting to oil bodies of the endogenous counterparts (Sarmiento et al., 1997). The low expression levels observed could therefore be due to: i) low stability of the sunflower oleosin in *A. thaliana*; ii) different efficacy of the promoter in controlling a soybean or a sunflower sequence or iii) different efficacy of the promoter in different plant species.

Beside the differences in expression profiles in pBIN19-35S and pBIN19-GM lines, a common feature to the two groups of transgenic plants was the identification of three main bands, when crude protein extracts from leaves and/or siliques were separated by SDS-PAGE and analysed by Western blot. In the plants transformed with the sunflower oleosin fused to the Nef125 polypetide, these bands migrated with an apparent molecular weight of 23, 19 and 15 kDa, whereas three polypeptides in the molecular range of approximately 20, 16 and 12 kDa were visualized in the plants transformed with the sunflower oleosin fused to 2F5e and NP peptides. Analysis of the nucleotide sequence of the sunflower oleosin indicated that the three bands may be derived from three different ATG codons in frame. The use of these alternatives ATG would indeed generate polypeptides with expected molecular weights perfectly matching those experimentally visualized.

The molecular masses of the observed bands in Nef 125 plants extracts were the first indication that the sunflower oleosin was indeed fused to the HIV-1-derived sequence. The same conclusion was reached after Western analysis of crude extracts from siliques of plants transformed with the 2F5e oleosin fusion (too small to induce a detectable migration shift on SDS-PAGE) with the monoclonal antibody (MAb) 2F5 (Buchacher et al., 1994). In the case of the NP peptide, due to the lack of specific antibodies, the presence was confirmed by MS/MS analysis of peptide fragments obtained by digesting with trypsin the oleosin bands identified by Coomassie-staining when proteins associated with oil bodies were separated on SDS-PAGE. Tandem mass spectrometry derived fragmentation data perfectly matched those predicted by *in silico* digestion. On the contrary, the tandem mass spectrometry approach did not reveal the 2F5e sequence in association with oil bodies extracted from the plants encoding this chimeric sunflower oleosin. Probably, as confirmed by the *in silico* prediction, this was due to the fact that trypsin digestion generates fragments bearing the epitope that, when doubly charged have m/z below the mass analyzer ion selection threshold ($400<m/z<1600$). Nonetheless, the MS/MS analysis of this sample identified the presence of five peptides of the sunflower oleosin sequence. Nef125 polypeptide presence was also further confirmed by MS/MS of oil bodies, although the C-terminal fragment of this polypeptide could not be detected. Similarly to the 2F5 peptide when this fragment is doubly charged its m/z is below the mass analyzer ion selection threshold.

The MS approach was an excellent method not only to verify the presence of the polypeptides fused to the sunflower oleosin but also to get a detailed catalogue of the proteins associated to wild type oil bodies. Beside the interest in widening the knowledge, the possible use of oil bodies to the vaccine field imperatively requires the definition of the complete list of the

proteins associated to this organelle both to improve antigen delivery and/or to prevent side effects. The efficiency in removing seed contaminants from oil-bodies preparations of two different protocols of purification was verified by comparing the protein complement of each oil body preparation after each purification procedure (Tzen et al., 1997; Deckers et al., 2004). This experiment demonstrated that contaminants removal was more effective when performed by the final 8 M urea wash following the protocol devised by Tzen rather than a protocol devised by Deckers in which a final wash of 0.1 M Na_2CO_3 was employed. Although urea is a strong chaotropic agent, it has been demonstrated, by light microscopy observations, that oil bodies keep their integrity and dimensions after washing using solutions which contain urea (Tzen et al., 1997). After the purification, lipids were removed from oil bodies preparations to improve protein separation by SDS-PAGE and the subsequent MS analysis.

The list of the proteins found through MS in the oil bodies preparations was compared to a published reference list obtained purifying oil bodies from *A. thaliana* cv. Wassilewskija seeds with the Tzen protocol (Jolivet et al., 2004). Both lists have been obtained by GeLC-MS/MS. Indeed the separation of proteins through SDS-PAGE before MS reduces sample complexity (allowing the identification of low abundant proteins) and allows the use of high SDS concentration, therefore a better solubilization of hydrophobic proteins, such as oleosins, often under-represented in proteomics studies (Santoni et al., 2000). However, while, in our case, MS analysis was performed on slices cut across the whole lane of a gel loaded with oil bodies-associated proteins, the reference list was obtained by analysing only slices sampled in correspondence of bands visible after Coomassie-staining. As a consequence of this different approach, the reference list identifies as associated to oil bodies only a limited number of proteins (mainly structural) (i.e. four oleosin isoforms; a protein homologous to calcium binding protein (ATS 1); a 11-β-hydroxysteroid dehydrogenase-like protein; a probable aquaporin; a glycosylphosphatidylinositol-anchored protein with unknown function) while our list (in which natural components were discriminated from contaminants referring to the Gene Ontology database) indicates that oil bodies protein-composition is far more complex. In our complete list, the proteins with the highest emPAI value, therefore the most abundant, are a 11-β-hydroxysteroid dehydrogenase-like protein (identified as a genuine resident of oil bodies and also in the reference list) and five oleosin isoforms (out of a total of six oleosins found in our analysis). Comparing the relative abundance of the different oleosin isoforms we found that the most abundant isoform was the 20.3 kDa, despite the fact that the prevalent isoform was considered the 18.5 kDa, in the previously published list. The analysis of oleosin gene expression in different *A. thaliana* tissues (florets, siliques, leaves, root, stem) indicates

that they are exclusively expressed in the floral tapetum and in siliques (Kim et al., 2002). Moreover, five genes have been identified that are exclusively expressed in siliques and at different times during embryo development (Kim et al., 2002). The oleosin isoforms reported in the reference list were the three intermediate- (18.5 kDa (At4g25140), 19.5 kDa (At3g01570), 21.3 kDa At5g40420) and the late- (20.3 kDa (At3g27660)) expressed isoforms (Jolivet et al, 2004). Beside these isoforms, we have identified also the early-expressed 14.3 kDa oleosin isoform (At5g51210), previously identified as an mRNA transcript (Kirik et al., 1996), and the 18.4 kDa isoform (At1g48990) never identified before in plant tissues, neither as transcript nor protein. Among all oleosins identified, the 19.5 kDa and the 20.3 kDa isoforms exhibit a neutral to sligthly acidic isoelectric point (pI) disproving the theory that the alkaline pI is a common feature of oleosins (Kim et al., 2002).

The comprehensive comparison of the MS data obtained by analysing "crude" oil bodies preparations (total seed extract; S1) and the final purified sample obtained with the Tzen protocol (S4T) indicated that the total number of proteins steeply decreases during the purification procedure. The preliminary analysis of the data on the basis of the emPAI value of the proteins shared between sample S1 and S4T, then classified and regrouped referring to the Gene Ontology database, allowed the identification of a group of proteins that were enriched and a group of proteins that were depleted during purification. Altogether these results confirmed what was observed on the Coomassie-stained gel where protein bands with an apparent molecular weight between 21 and 50 kDa (identified by the MS analysis as the α and β sub-units of 12S cruciferins) tended to disappear while the bands corresponding to the oleosin isoforms were enriched. Unexpectedly, from this analysis oleosins represent 30% of total oil body-associated proteins in purified oil bodies rather than 80% as previously reported by band intensity (Jolivet et al., 2004). However, it must be considered that emPAI values-based analyses implies that proteins are equally susceptible to digestion and the generated peptides equally detectable by LC-MS/MS. In the case of easily soluble proteins this assumption is probably valid but, if proteins include sections with no arginine or lysines, large tryptic peptides are generated which either are: i) poorly extracted from the gel; ii) not eluted from the column effectively during reverse phase chromatography; iii) ionised poorly or the m/z of the resulting peptide ions are so large that they are out of the mass range. In the case of hydrophobic proteins all that above described is more likely to occur as hydrophobic proteins usually have stretches with no charged residues and/or the resulting peptides are very hydrophobic and bind very strongly to the RP columns. Looking at the peptide coverage of the oleosins and comparing this with 11-β-dehydrogenase it was indeed verified that the

protein coverage was about 35% and 80%, respectively. In this way demonstrating that oleosin are detected less efficiently, giving some indications as to whether sections of sequence from the oleosin are missing.

Among the proteins that are enriched throughout the purification steps, beside oleosins and β-TIP, present also in the reference list (Jolivet et al., 2004), unexpected proteins with enzymatic activity were also enriched, while, seed storage proteins and the structural constituents of ribosomes were identified among depleted proteins.

A further group of proteins, identified exclusively in the final purified fraction, was detected by mass spectrometric analysis only when the high abundant contaminants were removed. The analysis of these proteins on the basis of Gene Ontology, identified nucleus- and plastid-associated proteins, that can be considered as residual contaminants, and membrane-associated proteins. These latter together with the ribosome structural components may associate with oil bodies during their biogenesis, and immature oil bodies with synthetic machinery attached may also have partitioned with mature oil bodies during the purification protocol. Several studies have shown that the protein composition of plant oil bodies and of the analogous lipid bodies found in fungi and animal cells is very different from plant oil bodies which have the simplest protein composition (Kamisaka et al., 1997; Athenstaedt et al., 1999; Ohashi et al., 2003; Liu et al., 2004). By our approach two new oleosins (14 kDa oleosin isoform (At5g51210) and the 18.4 kDa oleosin isoform (At1g48990)) and 3 membrane associated-proteins (benzodiazepine receptor-related (At2g47770), embryo-specific protein 3 (ATS3) (At5g07190), alpha subunit precursor TRAP complex/Signal sequence receptor alpha subunit (SSR-alpha) (At2g21160)) have been identified as oil body-associated proteins. If the data presented here will be further confirmed using alternative approaches, such as subtractive proteomics (Borner et al., 2006), the concept that plant oil-bodies are “simple” organelles will be put into question.

Oil bodies purified from transgenic T_2 seeds which has been analysed by Western and MS to assess the presence on their surface of the immunogenic peptides as sunflower oleosin fusions were used to induce peptide-specific immune responses in mice. In the preliminary experiments mice were immunized only with oil bodies carrying the antibody epitope 2F5e or with those carrying the MHC class I-restricted NP peptide. To avoid urea contamination, known to be toxic for mice, oil bodies for immunisation were prepared from transgenic seeds using a protocol similar to that developed by Deckers, using sterile 0.9% NaCl for washing instead of sodium carbonate buffer. The attempts to quantify the dose of chimeric sunflower

into the purified oil bodies preparations by ELISA using peptide- (MAb 2F5) or sunflower oleosin-specific antibodies were unsuccessful. In fact, the titration curves obtained for the synthetic peptide and for the 2F5e chimeric oil bodies were inconsistent suggesting that the standard curve was not the appropriate reference to extrapolate peptide concentration in the oil body preparation. Probably, this is due to the interference of the lipid components. Indeed, an ELISA performed titrating 2F5e and non chimeric sunflower oleosin oil bodies using for detection the polyclonal antibody specific to the sunflower oleosin gave two parallel curves. This result indicate that non chimeric sunflower oleosin oil bodies could be the proper standard knowing the amount of heterologous oleosin associated to their surface. This last information will be acquired through the SRM and MRM quantitative MS-based approaches using AQUA peptides spiked in the sample mix at known concentration as standard.

The immunological properties of the purified chimeric oil bodies carrying the NP peptide as sunflower oleosin fusion, were tested by subcutaneously injecting at day 0 and 14 C57BL/6J mice. Control mice were immunized with sterile saline alone, with the synthetic peptide ASNENMETM, or with oil bodies purified from *A. thaliana* plants expressing the wild type sunflower oleosin sequence. Immonogens were emulsified in Incomplete Freund's Adjuvant (IFA). An additional group of mice was immunized with chimeric oil bodies in sterile saline to examine the intrinsic adjuvant capacities. Nineteen days after the second immunisation mice were sacrificed to set up IFN-γ ELISPOT assays aimed to assess the number of cytotoxic T cells specific to the NP-derived synthetic peptides ASNENMETM and TYQRTRALV (negative control) in the spleen of each group of mice. The results of these experiments clearly demonstrated that chimeric oil bodies activate ASNENMTEM-specific CD8+ T cells. No response was obtained without adjuvant co-delivery.

The immunological properties of the purified chimeric oil bodies carrying the sunflower oleosin fusion the 2F5e peptide were also assessed following the same immunization schedule used for NP chimeric oil bodies. However, in this case the response was evaluated by titrating 2F5e-specific antibodies in the sera collected at day 0 (pre-immune mice), 14, 21 and 42 in 2F5e coated wells. The antibody response obtained using the prime-boost protocol was somewhat uninformative, suggesting that a further immunisation may be required.

It is common knowledge that the efficacy of vaccination depends both on the antigen dose and on the delivery system (Rosenthal and Zimmerman, 2006). In subunit- and peptide-based vaccines, the nature of the carrier affects the type of activated immune responses. Exogenous

antigens such as pathogen subunits mainly activate humoral responses and vaccinations performed with these antigens usually fail in eliciting CTL responses. This is due to the fact that they cannot enter the processing pathway ending in MHC class I molecules-associated presentation. This is true also when these antigens are co-administered with safe commercial adjuvants. Indeed, at present, adjuvants approved for human use (aluminum hydroxide and MF59) are effective only in inducing humoral immunity. Also IFA, used herein, is known as an adjuvant mainly effective in eliciting humoral responses, but studies performed using the combination of IFA with other adjuvant compounds widened the spectrum also to cell-mediated immune responses (Chang et al., 1998).

Some delivery strategies are able to overcome the barriers due to the mechanism of exogenous antigen processing as they exploit the so-called "cross-presentation" pathway of processing, culminating in the activation of naïve CTL by APC that are able to address exogenous antigens to the phagosome-cytosol pathway (Albert et al., 1988). An example of this is represented by liposome-based vaccines. Both liposome encapsulated and surface-conjugated peptides are indeed able to elicit CTL responses when administered in combination with innate immune activators (Nagata et al., 2007), although a stronger CTL response is obtained when antigens are conjugated to the surface. The results obtained by immunising mice with the NP chimeric oil bodies demonstrate that this delivery strategy is effective in activating CTL responses. The response of the mice immunised with the synthetic NP peptide was higher than that observed in mice receiving the chimeric oil bodies, but the difference in the response is much less pronounced than the difference in the dose of antigen delivered. Although the amount of the NP peptide administered through oil bodies was certainly significantly lower than 50 μg, the system was still effective in addressing/delivering the antigen to proper APC for processing and presentation to MHC class I- restricted cytotoxic T lymphocytes, no matter if the epitope is not inserted in the natural context. Moreover, it is known that synthetic peptides do not require processing to activate CTL, rather they directly bind to MHC class I molecules displayed on cell surfaces. The use of plant oil bodies prepared by a "milder" purification protocol to render oil bodies compatible with mice immunization, results in oil body preparations with a greater degree of contamination with non-oil body protein. These contaminating proteins may mask the epitope, as suggested by results obtained with 2F5e chimeric oil bodies and may also lead to sensitization and detrimental allergic responses. It has been shown however, that allergic reactions to sesame and peanut proteins often targets oleosins isoforms (Pons et al. 2002; Leduc et al., 2006). Alternative approaches aimed to

reduce the risk of eliciting allergic responses should therefore be devised, in order to engineer chimeric oil bodies free of allergenic properties.

6. ACKNOWLEDGEMENTS

My PhD studies have been an exciting time of my life with so much learning and personal development. Developing this project as a collaboration among three laboratories has been a great experience and I've got the chance to meet some special people I would like to thank.

I would firstly like to thank Dr. Eugenio Benvenuto, for giving me the opportunity to work on this project and my supervisor Dr. Selene Baschieri for her support and advices through the years. I would especially like to thank Dr. Selene Baschieri for the enlightening discussions about the many directions questions may be addressed and for showing me that improvements are always possible.

Special thanks must go to Dr. Kathryn Lilley and Prof. Johnathan Napier that keenly supervised my work for the time I spent working in their laboratories.

I am also very grateful to Dr. Frederic Beaudoin not only for the technical supervision of my work but also for the expert advices that were crucial for the development of this project. I would especially like to thank Dr. Frederic Beaudoin for his promptness at providing materials, for the many interesting discussions and his friendly support and encouragement through the years.

I also acknowledge the people I have collaborated with on some aspect of the project. Nick Bond worked with me on the study of the oil body proteome and the protein sequencing through the Mass Spectrometry approach. Camillo Mancini carried out the injections in mice.

I also thank everyone in the three labs for interesting discussion, practical help and friendly chats. In particular, I acknowledge the following people: Dr. Monica Venegas Caleron for her advices and help with plant transformation, Dr. Johanna Rees and Svenja Hester for help with Mass Spectrometry, Dr. Giampaolo Buriani and Dr. Chiara Lico for technical advices and help with immunizations.

On a personal note, I thank the many special friends I have met in the course of this journey that shared with me their culture, knowledge and feelings and eventually their home.

Finally, I especially thank my parents, Angela and Luigi, for their support and encouragement. Through these years they have always been there for me, I could not have done all this without them.

The PhD work was supported by the EC Integrated Project “Pharma-Planta” (LSHBCT-2003-503565). Three laboratories were involved: Rothamsted Research Institute (tutors: Prof. J.A. Napier, Dr. F. Beaudoin); Centre for Proteomics, University of Cambridge (tutor: Prof. K.S. Lilley); ENEA, Department of Biotechnology (tutor: Dr. S. Baschieri). The whole project was supervised by Dr. S. Baschieri & Dr. E. Benvenuto, ENEA.

7. REFERENCES

Abell BM, Holbrook LA, Abenes M, Murphy DJ, Hills MJ, Moloney MM.
Role of the proline knot motif in oleosin endoplasmic reticulum topology and oil body targeting.
Plant Cell 9:1481-1493, 1997.

Aguilar JC, Rodríguez EG.
Vaccine adjuvants revisited.
Vaccine 25: 3752-3762, 2007.

Albert ML, Sauter B, Bhardwaj N.
Dendritic cells acquire antigen from apoptotic cells and induce class I-restricted CTLs.
Nature 392: 86–89, 1988.

Alexander J, Oseroff C, Dahlberg C, Qin M, Ishioka G, Beebe M, Fikes J, Newman M, Chesnut RW, Morton PA, Fok K, Appella E, Sette A.
A decaepitope polypeptide primes for multiple CD8 IFN-γ and Th lymphocyte responses: evaluation of multiepitope polypeptides as a mode for vaccine delivery.
J. Immunol. 168: 6189–6198, 2002.

Andersen JS, Wilkinson CJ, Mayor T, Mortensen P, Nigg EA, Mann M.
Proteomic characterization of the human centrosome by protein correlation profiling. Nature 426: 570-574, 2003.

Andersen JS, Lam YW, Leung AKL, Ong S-E, Lyon CE, Lamond AI, MannM.
Nucleolar proteome dynamics.
Nature 433: 77-83, 2005.

Arakawa T, Chong DK, Merritt JL, Langridge WH.
Expression of cholera toxin B subunit oligomers in transgenic potato plants. Transgenic Res. 6:403-413, 1997.

Arakawa T, Yu J, Chong DKX, Hough J, Engen PC, Langridge WHR.
A plant-based cholera toxin B subunit insulin fusion protein protects against the development of autoimmune diabetes.
Nat. Biotechnol. 16: 934–938, 1998.

Arlen PA, Singleton M, Adamovicz JJ, Ding Y, Davoodi-Semiromi A, Daniell H.
Effective plague vaccination via oral delivery of plant cells expressing F1-V antigens in chloroplasts.
Infection and Immunity 76: 3640–3650, 2008.

Arnau J, Lauritzen C, Petersen GE, Pedersen J.
Current strategies for the use of affnity tags and tag removal for the purification of recombinant proteins.
Protein Expression and Purification 48: 1–13, 2006.

Arora VK, Fredericksen BL, Garcia JV.
Nef: agent of cell subversion.
Microbes Infect..4:189-199, 2002.

Athenstaedt K, Zweytick D, Jandrositz A, Kohlwein SD, Daum G.
Identification and characterization of major lipid particle proteins of the yeast *Saccharomyces cerevisiae*.
J. Bacteriol. 181: 6441–6448, 1999.

Bakker H, Rouwendal GJA, Karnoup AS, Florack DE, Stoopen GM, Helsper JPFG, van Ree R, van Die I, Bosch D.
An antibody produced in tobacco expressing a hybrid β-1,4-galactosylatransferase is essentially devoid of plant carbohydrate epitopes.
Proc. Natl. Acad. Sci. USA 103: 7577-7582, 2006.

Baneyx F, Mujacic M.
Recombinant protein folding and misfolding in *Escherichia coli*.
Nat. Biotechnol. 22: 1399-1408, 2004.

Barbante A, Irons S, Hawes C, Frigerio L, Vitale A, Pedrazzini E.
Anchorage to the cytosolic face of the endoplasmic reticulum membrane: a new strategy to stabilize a cytosolic recombinant antigen in plants.
Plant Biotechnol. J. 6:560-575, 2008.

Beaudoin F, Napier JA.
The targeting and accumulation of ectopically expressed oleosin in non-seed tissues of *Arabidopsis thaliana*.
Planta 210: 439-445, 2000.

Belanger FC, Kriz AL.
Molecular basis for allelic polymorphism of the maize globulin-1 gene.
Genetics 129: 863–872, 1991.

Bendahmane M, Koo M, Karrer E, Beachy RN
Display of epitopes on the surface of Tobacco Mosaic Virus: impact of charge and isoelectric point of the epitope on virus-host interaction.
J. Mol. Biol. 290: 9-20, 1999.

Bentley GA, Mariuzza RA.
The structure of the T cell antigen receptor.
Annu. Rev. Immunol. 14: 563-590, 1996.

Billaut-Mulot O, Idziorek T, Loyens M, Capron A, Bahr GM.
Modulation of cellular and humoral immune responses to a multiepitopic HIV-1 DNA vaccine by interleukin-18 DNA immunization/viral protein boost.
Vaccine 19:2803–2811, 2001.

Birch RG.
Plant transformation: problems and strategies for practical application.
Annu. Rev. Plant Physiol. Plant Mol. Biol. 48: 297-326, 1997.

Borisjuk NV, Borisjuk LG, Logendra S, Petersen F, Gleba Y, Raskin I.
Production of recombinant proteins in plant root exudates.

Nat. Biotechnol. 17: 466 – 469, 1999.

Borner GHH, Harbour M, Hester S, Lilley KS, Robinson MS.
Comparative proteomics of clathrin-coated vesicles.
J. Cell Biol. 175: 571-578, 2006.

Bradford MM.
A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.
Ann. Biochem. 72: 248-254, 1976.

Borrow P, Lewicki H, Hahn BH, Shaw GM, Oldstone MB.
Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection.
J. Virol. 68:6103–6110, 1994.

Broglie K, Chet I, Holliday M, Cressman R, Briddle P, Knowlton S, Mauvais CJ, Broglie R.
Transgenic plants with enhanced resistance to the fungal pathogen *Rhizoctonia solani*.
Science 254: 1194-1197, 1991.

Buchacher A, Predl R, Strutzenberger K, Steinfellner W, Trkola A, Purtscher M, Gruber G, Tauer C, Steindl F, Jungbauer A, et al.
Generation of human monoclonal antibodies against HIV-1 proteins; electrofusion and Epstein-Barr virus transformation for peripheral blood lymphocyte immortalization.
AIDS Res. Hum. Retroviruses 10: 359-69, 1994.

Bui HH, Peters B, Assarsson E, Mbawuike I, Sette A.
Ab and T cell epitopes of influenza A virus, knowledge and opportunities.
Proc. Natl. Acad. Sci. USA 104: 246-251, 2007.

Canizares MC, Nicholson L, Lomonossoff GP.
Use of viral vectors for vaccine production in plants.
Immunol. Cell Biol. 83: 263–270, 2005.

Carrillo C, Wigdorovitz A, Oliveros JC, Zamorano PI, Sadir AM, Gomez N, Salinas J, Escribano JM, Borca MV.
Protective immune response to foot-and-mouth disease virus with VP1 expressed in transgenic plants.
J. Virol. 72: 1688-1690, 1998.

Chang JCC, Diveley JP, Savary JR, Jensen FC.
Adjuvant activity of incomplete Freund's adjuvant.
Advanced Drug Delivery Reviews 32: 173–186, 1998.

Chargelegue D, Vine ND, van Dolleweerd CJ, Drake P, Ma JKC.
A murine monoclonal antibody produced in transgenic plants with plant-specific glycans is not immunogenic in mice.
Transgenic Res. 9: 187-194, 2000.

Chen ECF, Tai SSK, Peng CC, Tzen JTC.
Identification of three novel unique proteins in seed oil bodies of sesame.
Plant Cell Physiol. 39: 935–941, 1998.

Chen JCF, Tsai CCY, Tzen JTC.
Cloning and secondary structure analysis of caleosin, a unique calcium-binding protein in oil bodies of plant seeds.
Plant Cell Physiol. 40: 1079–1086, 1999.

Chrispeels MJ, Faye L.
The production of recombinant glycoproteins with defined nonimmunogenic glycans. In Owen MRL, Pen J (eds): Transgenic Plants: A Production System for Industrial and Pharmaceutical Proteins. Chichester, UK: John Wiley & Sons, pp 99–113, 1996.

Clough SJ, Bent AF.
Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*.
Plant J. 16: 735-743, 1998.

Conrad U, Fiedler U.
Compartment-specific accumulation of recombinant immunoglobulins in plant cells: an essential tool for antibody production and immunomodulation of physiological functions and pathogen activity.
Plant Mol. Biol. 38: 101–109, 1998.

Cooper E, O'Reilly M.
A novel staff vaccination strategy.
Infect. Control. Hosp. Epidemiol. 23: 232-3, 2002.

Cox KM, Sterling JD, Regan JT, Gasdaska JR, Frantz KK, Peele CG, Black A, Passmore D, Moldovan-Loomis C, Srinivasan M, Cuison S, Cardarelli PM, Dickey LF.
Glycan optimization of a human monoclonal antibody in the acquatic planta *Lemna minor*.
Nat. Biotech. 24: 1591-97, 2004

Crowe SR, Miller SC, Woodland DL.
Identification of protective and non-protective T cell epitopes in influenza.
Vaccine 24 : 452–456, 2006.

Daniell H.
Production of biopharmaceuticals and vaccines in plants via the chloroplast genome.
Biotechnol. J. 1: 1071–1079, 2006.

Day CD, Lee E, Kobayashi J, Holappa LD, Albert H, Ow DW.
Transgene integration into the same chromosome location can produce alleles that express at a predictable level, or alleles that are differentially silenced.
Genes Dev. 14: 2869–2880, 2000.

De Buck S, Van Montagu M, Depicker A.
Transgene silencing of invertedly repeated transgenes is released upon deletion of one of the transgenes involved.
Plant Mol. Biol. 46: 433-445, 2001.

De Buck S, Windels P, De Loose M, Depicker A.

Single copy T-DNAs integrated at different positions in the *Arabidopsis* genome display uniform and comparable β-glucuronidase accumulation levels.
Cell Mol. Life Sci. 61: 2632-2645, 2004.

De Duve C.
Tissue fractionation. Past and present.
J. Cell Biol. 50: 20d-55d, 1971.

De Jaeger G, Scheffer S, Jacobs A, Zambre M, Zobell O, Goossens A, Depicker A, Angenon G.
Boosting heterologous protein production in transgenic dicotyledonous seeds using *Phaseolus vulgaris* regulatory sequences.
Nat. Biotechnol. 20: 1265–1268, 2002.

De Neve M, De Buck S, Jacobs A, Van Montagu M, Depicker A.
T-DNA integration patterns in co-transformed plant cells suggest that T-DNA repeats originate from co-integration of separate T-DNAs.
Plant J. 11: 15-29, 1997.

de Virgilio M, De Marchis F, Bellocci M, Mainieri D, Rossi M, Benvenuto E, Arcioni S, Vitale A.
The human immunodeficiency virus antigen Nef forms protein bodies in leaves of transgenic tobacco when fused to zeolin.
J. Exp. Bot. 59: 2815–2829, 2008.

Deckers HM, van Rooijen G, Boothe J, Moloney MM, Schryvers AB, Hutchins WA.
Immunogenic formulations comprising oil bodies.
Patent US6761914 B2, 2004.

Doherty PC, Turner SJ, Webby RG, Thomas PG.
Influenza and the challenge for immunology.
Nat. Immunol. 7: 449-455, 2006.

Drake PMW, Chargelegue DM, Vine ND, van Dolleweerd CJ, Obregon P, Ma JKC.

Rhizosecretion of a monoclonal antibody protein complex from transgenic tobacco roots.
Plant Mol. Biol. 52: 233–241, 2003.

Dunkley TPJ, Watson R, Griffin JL, Dupree P, Lilley KS.
Localization of Organelle Proteins by Isotope Tagging (LOPIT).
Molecular & Cellular Proteomics 3:1128-1134, 2004.

Dunkley TPJ, Hester S, Shadforth IP, Runions J, Weimar T, Hanton SL, Griffin JL, Bessant C, Brandizzi F, Hawes C, Watsonn RB, Cupree P, Lilley KS.
Mapping the *Arabidopsis* organelle proteome.
Proc. Natl. Acad. Sci. USA 103: 6518–6523, 2006.

Ehsani P, Khabiri A, Domansky NN.
Polypeptides of hepatitis B surface antigen produced in transgenic potato.
Gene 190:107-111, 1997.

Elmayan T, Vaucheret H.
Expression of single copies of a strongly expressed *35S* transgene can be silenced post-transcriptionally.
Plant J. 9: 787–797, 1996.

Feldmann K, Marks M.
Agrobacterium-mediated transformation of germinating seeds of *Arabidopsis thaliana*: a non-tissue culture approach.
Mol. Gen. Genet. 208: 1-9, 1987.

Flynn KJ, Belz GT, Altman JD, Ahmed R, Woodland DL, Doherty PC.
Virus-specific CD8+ T cells in primary and secondary influenza pneumonia.
Immunity 8: 683-691, 1998.

Floss DM, Sack M, Stadlmann J, Rademacher T, Scheller J, Stöger E, Fischer R, Conrad U.
Biochemical and functional characterization of anti-HIV antibody-ELP fusion protins from transgenic plants.
Plant Biotechnol. J. 6: 379-391, 2008.

Foster LJ, de Hoog CL, Mann M.
Unbiased quantitative proteomics of lipid rafts reveals high specificity for signalling factors.
Proc. Natl. Acad. Sci. USA 100: 5813-5818, 2003.

Frandsen GI, Mundy J, Tzen JTC.
Oil bodies and their associated proteins, oleosin and caleosin.
Physiol. Plant. 112: 301–307, 2001.

Frankel AD, Young JAT.
HIV-1: fifteen proteins and an RNA.
Annu. Rev. Biochem. 67: 1-25, 1998.

Frey-Wyssling A., Grieshaber E, Muhlethaler K.
Origin of spherosomes in plant cells.
J. Ultrastruct. Res. 8: 506–516, 1963.

Garcia KC, Teyton L, Wilson IA.
Structural basis of T cell recognition.
Annu. Rev. Immunol. 17: 369-397, 1999.

Gerngross TU.
Advances in the production of human therapeutic proteins in yeasts and filamentous fungi.
Nat. Biotechnol. 22: 1409-1414, 2004.

Gil F, Reytor E, Pérez-Filgueira DM, Escribano JM.
Multimerization of peptide antigens for production of stable immunogens in transgenic plants.
J. Biotechnol. 128: 512-518, 2007.

Gleba Y, Klimyuk V, Marillonnet S.
Magnifection – a new platform for expressing recombinant vaccines in plants.
Vaccine 23: 2042-2048, 2005.

Gomez N, Carrillo C, Salinas J, Parra F, Borca MV, Escribano JM.

Expression of immunogenic glycoprotein S polypeptides from transmissible gastroenteritis coronavirus in transgenic plants.
Virology 249: 352-358, 1998.

Graves A, Goldman S.
Agrobacterium tumefaciens-mediated transformation of the monocot genus *Gladiolus*: detection of expression of T-DNA-encoded genes.
J. Bacteriol. 169: 1745-1746, 1987.

Guilley H, Dudley RK, Jonard G, Balazs E, Richards KE.
Transcription of Cauliflower Mosaic Virus DNA: detection of promoter sequences, and characterization of transcripts.
Cell 30: 763-773, 1982.

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R.
Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotech.17: 994-999, 1999.

Haq TA, Mason HS, Clements JD, Arntzen CJ.
Oral immunization with a recombinant bacterial antigen produced in transgenic plants.
Science 268:714-716, 1995.

Hanin M, Paszkowski J.
Plant genome modification by homologous recombination.
Curr. Opin. Plant Biol. 6: 157-162, 2003.

Hassan S, van Dolleweerd CJ, Ioakeimidis F, Keshavarz-Moore E, Ma JKC.
Considerations for extraction of monoclonal antibodies targeted to different subcellular compartments in transgenic tobacco plants.
Plant Biotechnol. J. 6: 733-748, 2008.

Hobbs SL, Warkentin TD, DeLong CM.
Transgene copy number can be positively or negatively associated with transgene expression.
Plant Mol. Biol. 21: 17-26, 1993.

Holmgren J, Czerkinsky C.
Mucosal immunity and vaccines.
Nat. Med. 11: S45-S53, 2005.

Hood EE, Witcher DR, Maddock S, Meyer T, Baszczynski C, Bailey M, Flynn P, Register J, Marshall L, Bond D et al.
Commercial production of avidin from transgenic maize: characterization of transformant, production, processing, extraction and purification.
Mol. Breed. 3:291-306, 1997.

Hood EE, Bailey MR, Beifuss K, Magallanes-Lundback M, Horn ME, Callaway E, Drees C, Delaney DE, Clough R, Howard JA.
Criteria for high-level expression of a fungal laccase gene in transgenic maize.
Plant Biotechnol. J. 1:129-140, 2003.

Horvath H, Huang J, Wong O, Kohl E, Okita T, Kannangara CG, von Wettstein D.
The production of recombinant proteins in transgenic barley grains.
Proc. Natl. Acad. Sci. USA 97:1914–1919, 2000.

Houde M, Bertholet S, Gagnon E, Brunet S, Goyette G, Laplante A, Princiotta MF, Thibault P, Sacks D, Desjardins M.
Phagosomes are competent organelles for antigen cross-presentation.
Nature 425: 402-406, 2003.

Huang AHC.
Oleosins and oil bodies in seeds and other organs.
Plant Physiol. 110:1055-1061, 1996.

Hwang YS, Nichol S, Nandi S, Jernstedt JA, Huang N.
Aleurone and embryo-specific expression of the β-glucuronidase gene controlled by the barley Chi26 and Ltp1 promoters in transgenic rice.
Plant Cell Rep. 20: 647-654, 2001.

Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M.
Exponentially modified Protein Abundance Index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein.
Molecular & Cellular Proteomics 4:1265–1272, 2005.

Jain R, Katavic V, Agrawal GC, Guzov VM, Thelen JJ.
Purification and proteomic characterization of plastids from *Brassica napus* developing embryos.
Proteomics 8: 3397–3405, 2008.

Jaquinod M, Villiers F, Kieffer-Jaquinod S, Hugouvieux V, Bruley C, Garin J, Bourguignon J.
A proteomics dissection of *Arabidopsis thaliana* vacuoles isolated from cell culture.
Molecular & Cellular Proteomics 6: 394-412, 2007.

Joensuu JJ, Kotiaho M, Riipi T, Snoeck V, Palva ET, Teeri TH, Lång H, Cox E, Goddeeris BM, Niklander-Teeri V.
Fimbrial subunit protein FaeG expressed in transgenic tobacco inhibits the binding of F4ac enterotoxigenic Escherichia coli to porcine enterocytes.
Transgenic Res. 13: 295-298, 2004.

Johnson J, Lin T, Lomonossoff G.
Presentation of heterologous peptides on plant viruses: genetics, structure, and function.
Annu. Rev. Phytopathol. 35: 67-86, 1997.

Jolivet P, Roux E, d'Andrea S, Davanture M, Negroni L, Zivy M, Chardot T.
Protein composition of oilbodies in *Arabidopsis thaliana* ecotype WS.
Plant Physiol. Biochem. 42: 501–509, 2004.

Jorgensen RA, Cluster PD, English J, Que Q, Napoli CA.
Chalcone synthase cosuppression phenotypes in petunia flowers: comparison of sense vs. antisense constructs and single-copy vs. complex T-DNA sequences.
Plant Mol. Biol. 31: 957–973, 1996.

Kamisaka Y, Mishra S, Nakahara T.
Purification and characterization of diacylglycerol acyltransferase from the lipid body fraction of an oleaginous fungus.
J. Biochem. 121: 1107–1114, 1997.

Kanamoto H, Yamashita A, Asao H, Okumura S, Takase H, Hattori M, Yokota A, Tomizawa K.
Efficient and stable transformation of *Lactuca sativa* L. cv. Cisco (lettuce) plastids.
Transgenic Res. 15: 205-217, 2006.

Kapila J, De Rycke R, Van Montagu M, Angenon G.
An *Agrobacterium* -mediated transient gene expression system for intact leaves.
Plant Sci. 122: 101-108, 1997.

Kapusta J, Modelska A, Figlerowicz M, Pniewski T, Letellier M, Lisowa O, Yusibov V, Koprowski H, Plucienniczak A, Legocki AB.
A plant-derived edible vaccine against hepatitis B virus.
FASEB J. 13:1796-1799, 1999.

Karlsson Hedestam GB, Fouchier RAM, Phogat S, Burton DR, Sodroski J, Wyatt RT.
The challenges of eliciting neutralizing antibodies to HIV-1 and to influenza virus.
Nat. Rev. Microbiology 6: 143-155, 2008.

Kessler F, Vidi P-A.
Plastoglobule-lipid bodies: their function in chloroplast and their potential for application.
Adv. Biochem. Engin./Biotechnol. 107: 153-172, 2007.

Kim HU, Hsieh K, Ratnayake C, Huang AH.
A novel group of oleosins is present inside the pollen of *Arabidopsis*.
J. Biol. Chem. 277: 22677–22684, 2002.

Kirik V, Kölle K, Balzer HJ, Bäumlein H.
Two new oleosin isoforms with altered expression patterns in seeds of the Arabidopsis mutant fus3.

Plant Mol. Biol. 31: 413-417, 1996.

Kirkpatrick DS, Gerber SA, Gygi SP.
The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications.
Methods 35:265–273, 2005.

Klein MR, van Baalen CA, Holwerda AM, Kerkhof-Garde SR, Bende RJ, Keet IPM, Eeftinck-Schattenkerk JKM, Osterhaus ADME, Schuitemaker H, Miedema F.
Kinetics of gag-specific cytotoxic T-lymphocyte responses during the clinical course of HIV-1 infection: a longitudinal analysis of rapid progressors and long-term asymptomatics.
J. Exp.Med. 181:1365–1372, 1995.

Komarnytsky S, Borisjuk N, Yakoby N, Garvey A, Raskin I.
Cosecretion of Protease Inhibitor Stabilizes Antibodies Produced by Plant Roots. Plant Physiol. 141: 1185–1193, 2006.

Kost TA, Condreay JP, Jarvis DL.
Baculovirus as versatile vectors for protein expression in insect and mammalian cells.
Nat. Biotechnol. 23: 567-575, 2005.

Koup RA, Safrit JT, Cao Y, Andrews CA, Wu Y, McLeod G, Borkowsky W, Farthing C, Ho DD.
Temporal association of cellular immune response with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome.
J. Virol. 68:4650–4655, 1994.

Koziel MG, Carozzi NB, Desai N.
Optimizing expression of transgenes with an emphasis on plant-transcriptional events. Plant Mol. Biol. 32: 393-405, 1996.

Krijgsveld J, Ketting RF, Mahmoudi T, Johansen J, Artal-Sanz M, Verrijzer CP, Plasterk RH, Heck AJ.
Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics.

Nat. Biotechnol. 21: 927-931, 2003.

Krishnan J, Selvarajoo K, Tsuchiya M, Lee G, Choi S.
Toll-like receptor signal transduction.
Exp. Mol. Med. 39: 421-38, 2007.

Kusnadi AR, Hood EE, Witcher DR, Howard JA, Nikolov ZN.
Production and purification of two recombinant proteins from transgenic corn. Biotechnol. Prog. 14:149-155, 1998.

Kwong PD, Doyle ML, Casper DJ, Cicala C, Leavitt SA, Majeed S, Steenbeke TD, Venturi M, Chaiken I, Fung M, Katinger H, Parren PW, Robinson J, Van Ryk D, Wang L, Burton DR, Freire E, Wyatt R, Sodroski J, Hendrickson WA, Arthos J.
HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites.
Nature 420: 678–682, 2002.

Lacasse P, Denis J, Lapointe R, Leclerc D, Lamarre A.
Novel plant virus-based vaccine induces protective cytotoxic T-lymphocyte-mediated antiviral immunity through dendritic cell maturation.
J. Virol. 82: 785-794, 2008.

Lee RW, Cornelisse M, Ziauddin A, Slack PJ, Hodgins DC, Strommer JN, Shewen PE, Lo RY.
Expression of a modified Mannheimia haemolytica GS60 outer membrane lipoprotein in transgenic alfalfa for the development of an edible vaccine against bovine pneumonic pasteurellosis.
J. Biotechnol. 135:224-231, 2008.

Leduc V, Moneret-Vautrin DA, Tzen JTC, Morisset M, Guerin L, Kanny G.
Identification of oleosins as major allergens in sesame seed allergic patients.
Allergy 61: 349–356, 2006.

Li HY, Ramalingam S, Chye ML.

Accumulation of recombinant SARS-CoV spike protein in plant cytosol and chloroplasts indicate potential for development of plant-derived oral vaccines.
Exp. Biol. Med. 231: 1346-1352, 2006.

Lico C, Capuano F, Renzone G, Donini M, Marusic C, Scaloni A, Benvenuto E, Baschieri S.
Peptide display on *Potato virus X*: molecular features of the coat protein-fused peptide affecting cell-to-cell and phloem movement of chimeric virus particles.
J. Gen. Virol. 87: 3103-3112, 2006.

Lilley KS, Razzaq A, Dupree P.
Two-dimensional gel electrophoresis: recent advances in sample preparation, detection and quantitation.
Curr. Opin. Chem. Biol. 6: 46-50, 2002.

Lin L-J, Tai SSK, Peng C-C, Tzen JTC.
Steroleosin, a sterolbinding dehydrogenase in seed oil bodies.
Plant Physiol. 128: 1200–1211, 2002.

Liu H, Sadygov RG, Yates JR.
A model for random sampling and estimation of relative protein abundance in shotgun proteomics.
Anal. Chem. 76: 4193-4201, 2004.

Ma JK-C, Drake PMW, Christou P.
The production of recombinant pharmaceutical proteins in plants.
Nat. Rev. Gen. 4: 794-805, 2003.

Malik K, Wu K, Li X-Q, Martin-Heller T, Hu M, Foster E, Tian L, Wang C, Ward K, Jordan M, Brown D, Gleddie S, Simmonds D, Zheng S, Simmonds J, Miki B.
A constitutive gene expression system derived from the tCUP cryptic promoter elements.
Theor. Appl. Genet. 105: 505-514, 2002.

Mallory AC, Parks G, Endres MW, Baulcombe D, Bowman LH, Pruss GJ, Vance VB.
The amplicon-plus system for high-level expression of transgenic plants.

Nat. Biotechnol. 20: 622-625, 2002.

Marillonnet S, Thoeringer C, Kandzia M, Klimyuk V, Gleba Y.
Systemic *Agrobacterium tumefaciens*-mediated transfection of viral replicons for efficient transient expression in plants.
Nat. Biotechnol. 23:718-723, 2005.

Martín-Alonso JM, Castañón S, Alonso P, Parra F, Ordás R.
Oral immunization using tuber extracts from transgenic potato plants expressing rabbit hemorrhagic disease virus capsid protein.
Transgenic Res. 12: 127-130, 2003.

Marusic C, Rizza P, Lattanzi L, Mancini C, Spada M, Belardelli F, Benvenuto E, Capone I.
Chimeric plant virus particles as immunogens for inducing murine and human immune responses against human immunodeficiency virus type 1.
J. Virol. 75:8434-8439, 2001.

Marusic C, Nuttall J, Buriani G, Lico C, Lombardi R, Baschieri S, Benvenuto E, Frigerio L.
Expression, intracellular targeting and purification of HIV Nef variants in tobacco cells.
BMC Biotechnology 7: 12-24, 2007.

Mason HS, Lam DM, Arntzen CJ .
Expression of hepatitis B surface antigen in transgenic plants.
Proc. Natl. Acad. Sci. USA 89:11745-11749, 1992.

Mason HS, Ball JM, Shi JJ, Jiang X, Estes MK, Arntzen CJ.
Expression of Norwalk capsid protein in transgenic tobacco and potato and its oral immunogenicity in mice.
Proc. Natl. Acad. Sci. USA 93:5335-5340, 1996.

Mason HS, Haq TA, Clements JD, Arntzen CJ.
Edible vaccine protects mice against *Escherichia coli* heat-labile enterotoxin (LT): potatoes expressing a synthetic LT-B gene.
Vaccine 16:1336-1343, 1998.

McBurney SP, Young KR, Ross TM.
Membrane embedded HIV-1 envelope on the surface of a virus-like particle elicits broader immune responses than soluble envelope.
Virology 358: 334-346, 2006.

McCormick AA, Corbo TA, Wykoff-Clary S, Palmer KE, Pogue GP.
Chemical conjugate TMV-peptide bivalent fusion vaccines improve cellular immunity and tumor protection.
Bioconjug. Chem. 17: 1330-8, 2006.

McGarvey PB, Hammond J, Dienelt MM, Hooper DC, Fu ZF, Dietzschold B, Koprowski H, Michaels FH.
Expression of the rabies virus glycoprotein in transgenic tomatoes.
Bio/Technology 13:1484-1487, 1995.

Medzhitov R, Janeway CA Jr.
Innate immunity: the virtues of a non clonal system of recognition.
Cell 91: 295-298, 1997.

Menkhaus TJ, Bai Y, Zhang C, Nikolov ZL, Glatz CE.
Considerations for the Recovery of Recombinant Proteins from Plants.
Biotechnol. Prog. 20: 1001-1014, 2004.

Meyer DE, Chilkoti A.
Purification of recombinant proteins by fusion with thermally-responsive polypeptides.
Nat. Biotechnol. 17:1112-1115, 1997.

Moore JP, Parren PW, Burton DR.
Genetic subtypes, humoral immunity and human immunodeficiency virus type I vaccine development.
J. Virol. 75: 5721-5729, 2001.

Mowat AM.

Anatomical basis of tolerance and immunity to intestinal antigens.
Nat. Rev. Immunol. 3: 331-341, 2003.

Murphy DJ.
Storage lipid bodies in plants and other organisms.
Prog. Lipid. Res. 29:299-324, 1990.

Muskens MW, Vissers AP, Mol JN, Kooter JM.
Role of inverted DNA repeats in transcriptional and post-transcriptional gene silencing.
Plant Mol. Biol. 43: 243-260, 2000.

Muster T, Steindl F, Purtscher M, Trkola A, Klima A, Himmler G, Ruker F, Katinger H.
A conserved neutralizing epitope on gp41 of human immunodeficiency virus type 1. J. Virol. 67:6642–6647, 1993.

Nagata T, Toyota T, Ishigaki H, Ichihashi T, Kajino K, Kashima Y, Itoh Y, Mori M, Odac H, Yamamurac H, Taneichi M, Uchida T, Ogasawara K.
Peptides coupled to the surface of a kind of liposome protect infection of influenza viruses.
Vaccine 25: 4914–4921, 2007.

Napier JA, Stobart AK, Shewry PR.
The structure and biogenesis of plant oil bodies: the role of the ER membrane and the oleosin class of protein.
Plant Mol. Biol. 31: 945-956, 1996.

Ni M, Cui D, Gelvin SB.
Sequence-specific interactions of wound-inducible nuclear factors with mannopine synthase 2' promoter wound responsive elements.
Plant Mol. Biol. 30**:** 77–96, 1996.

Nicholson L, Gonzalez-Melendi P, van Dolleweerd C, Tuck H, Perrin Y, Ma JK-C, Fischer R, Christou P, Stoger E.
A recombinant multimeric immunoglobulin expressed in rice shows assembly-dependent subcellular localization in endosperm cells.

Plant Biotech. J. 3: 115-127, 2005.

Nykiforuk LC, Boothe JG, Murray EW, Keon RG, Goren HJ, Markley NA, Moloney MM.
Transgenic expression and recovery of biologically active recombinant human insulin from *Arabidopsis thaliana* seeds.
Plant Biotechnol. J. 4: 77–85, 2006.

Oda Y, Huang K, Cross FR, Cowburn D, Chait BT.
Accurate quantitation of protein expression and site-specific phosphorylation.
Proc. Natl. Acad. Sci. USA 96: 6591-6596, 1999.

Ohashi M, Mizushima N, Kabeya Y, Yoshimori T.
Localization of mammalian NAD(P)H steroid dehydrogenase-like protein on lipid droplets.
J. Biol. Chem. 278: 36819–36829, 2003.

Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M.
Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.
Mol. Cell. Proteomics 1: 376-386, 2002.

Oszvald M, Kang TJ, Tomoskozi S, Jenes B, Kim TG, Cha YS, Tamas L, Yang MS.
Expression of cholera toxin B subunit in transgenic rice endosperm.
Mol. Biotechnol. 40: 261-268, 2008.

Outchkourov NS, Peters J, de Jong J, Rademarkers W, Jongsma MA.
The promoter-terminator of chrysanthemum *rbcS1* directs very high expression levels in plants.
Planta 216:1003-1012, 2003.

Palmer KE, Thomson JA, Rybicki EP.
Generation of maize cell lines containing autonomously replicating maize streak virus-based gene vectors.
Arch. Virol. 144:1345-1360, 1999.

Pantophlet R, Wilson IA, Burton DR.
Hyperglycosylated mutants of human immunodeficiency virus (HIV) type 1 monomeric gp120 as novel antigens for HIV vaccine design.
J. Virol. 77: 5889-5901, 2003.

Parmenter DL Boothe JG, van Rooijen GJH, Yeung EC, Moloney MM.
Production of biologically active hirudin in plant seeds using oleosin partitioning. Plant Mol. Biol. 29: 1167-1180,1995.

Pashine A, Valiante NM, Ulmer JB.
Targeting the innate immune response with improved vaccine adjuvants.
Nat. Med. 11: S63-S68, 2005.

Patel J, Zhu H, Menassa R, Gyenis L, Richman A, Brandle J.
Elastin-like polypeptide fusions enhance the accumulation of recombinant proteins in tobacco leaves.
Transgenic Res. 16:239-249, 2007.

Peng C-C, Tzen JTC.
Analysis of the three essential constituents of oil bodies in developing sesame seeds. Plant Cell Physiol. 39: 35–42, 1998.

Perkins DN, Pappin DJ, Creasy DM, Cottrell JS.
Probability-based protein identification by searching sequence databases using mass spectrometry data.
Electrophoresis, 20, 3551-3567, 1999.

Phogat S, Svehla K, Tang M, Spadaccini A, Muller J, Mascola J, Berkower I, Wyatt R.
Analysis of the human immunodeficiency virus type 1 gp41 membrane proximal external region arrayed on hepatitis B surface antigen particles.
Virology 373:72-84, 2007.

Phogat S, Wyatt R.
Rational modification of HIV-1 envelope glycoproteins for immunogen design.

Curr. Pharm. Des. 13: 213–227, 2007.

Pogrebnyak N, Golovkin M, Andrianov V, Spitsin S, Smirnov Y, Egolf R, Koprowski H.
Severe acute respiratory syndrome (SARS) S protein production in plants: development of recombinant vaccine.
Proc. Natl. Acad. Sci. USA 102: 9062-9067, 2005.

Pogue GP, Lindbo JA, Garger SJ, Fitzmaurice WP.
Making an ally from an enemy: plant virology and the new agriculture.
Annu. Rev. Phytopathol. 40:45-74, 2002.

Pons L, Chery C, Romano A, Namour F, Artesani MC, Guéant J-L.
The 18 kDa peanut oleosin is a candidate allergen for IgE-mediated reactions to peanuts.
Allergy 57 (Suppl. 72): 88–93, 2002.

Porta C, Spall VE, Findlay KC, Gergerich RC, Farrance CE, Lomonossoff GP
Cowpea mosaic virus-based chimaeras. Effects of inserted peptides on the phenotype, host range, and transmissibility of the modified viruses.
Virology 310: 50-63, 2003.

Rammensee HG, Bachmann J, Emmerich NN, Bachor OA, Stevanovic S. SYFPEITHI: database for MHC ligands and peptide motifs.
Immunogenetics 50: 213-219, 1999.

Rappuoli R.
From Pasteur to genomics: progress and challenges to infectious diseases.
Nat. Med. 10: 1117-1185, 2004.

Richter LJ, Thanavala Y, Arntzen CJ, Mason HS.
Production of hepatitis B surface antigen in transgenic plants for oral immunization. Nat. Biotechnol. 18:1167-1171, 2000.

Rosenthal KS, Zimmerman DH.
Vaccines: all things considered.

Clin. Vaccine Immunol. 13: 821-829, 2006.

Ross JHE, Sanchez J, Millan F, Murphy DJ.
Differential presence of oleosins in oleogenic seed and mesocarp tissues in olive (*Olea europea*) and avocado (*Persea americana*).
Plant Sci. 93: 247–280, 1993.

Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, Pappin DJ.
Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.
Mol. Cell. Proteomics 3: 1154-1169, 2004.

Roy CR.
Immunology: professional secrets.
Nature 425: 351-352, 2003.

Russell DA, Fromm ME.
Tissue-specific expression in transgenic maize of four endosperm promoters from maize and rice.
Transgenic Res. 6: 157–168, 1997.

Ryan EJ, Daly LM, Mills KH.
Immunomodulators and delivery systems for vaccination by mucosal routes.
Trends Biotechnol. 19: 293-304, 2001.

Samac DA, Tesfaye M, Dornbusch M, Saruul P, Temple SJ.
A comparison of constitutive promoters for expression of transgenes in alfalafa (*Medicago sativa*).
Transgenic Res. 13: 349-361, 2004.

Santi L, Giritch A, Roy CJ, Marillonnet S, Klimyuk V, Gleba Y, Webb R, Arntzen CJ, Mason HS.

Protection conferred by recombinant *Yersinia pestis* antigens produced by a rapid and highly scalable plant expression system.
Proc. Natl. Acad. Sci. USA 24:861-866, 2006.

Santoni V, Molloy M, Rabilloud T.
Membrane proteins and proteomics: un amour impossible?
Electrophoresis 21: 1054–1070, 2000.

Sarmiento C, Ross JH, Herman E, Murphy DJ.
Expression and subcellular targeting of a soybean oleosin in transgenic rapeseed. Implications for the mechanism of oil-body formation in seeds.
Plant J. 11:783-796, 1997.

Schmidt A, Kellermann J, Lottspeich F.
A novel strategy for quantitative proteomics using isotope-coded protein labels.
Proteomics 5: 4-15, 2005.

Sette A, Sidney J.
Nine major HLA class I supertypes account for the vast preponderance of HLA-A and –B polymorphism.
Immunogenetics 50: 201-212, 1999.

Siloto RMP, Findlay K, Lopez-Villalobos A, Yeung EC, Nykiforuk CL, Moloney MM.
The accumulation of oleosins determines the size of seed oilbodies in *Arabidopsis*. The Plant Cell 18: 1961–1974, 2006.

Southern EM.
Detection of specific sequences among DNA fragments separated by gel electrophoresis.
J. Mol. Biol. 98: 503-517, 1975.

Srivastava V, Ow DW.
Marker-free site-specific gene integration in plants.
Trends Biotechnol. 22: 627-629, 2004.

Stiegler G, Kunert R, Purtscher M, Wolbank S, Voglauer R, Steindl F, Katinger H.
A potent cross-clade neutralizing human monoclonal antibody against a novel epitope on gp41 of human immunodeficiency virus type 1.
AIDS Res. Hum. Retroviruses 17:1757–1765, 2001.

Stoger E, Sack M, Perrin Y, Vaquero C, Torres E, Twyman RM, Christou P, Fischer R.
Practical considerations for pharmaceutical antibody production in different crop systems.
Mol. Breed. 9:149-158, 2000.

Stoger E, Ma JKC, Fischer R, Christou P.
Sowing the seeds of success: pharmaceutical proteins from plants. Current Opinion in Biotechnology 16:167–173, 2005.

Streatfield SJ, Jilka JM, Hood EE, Turner DD, Bailey MR, Mayor JM, Woodard SL, Beifuss KK, Horn ME, Delaney DE, Tizard IR, Howard JA.
Plant based vaccines: unique advantages.
Vaccine 19: 2742-2748, 2001.

Sullivan ML, Green PJ.
Post-transcriptional regulation of nuclear-encoded genes in higher plants: the roles of mRNA stability and translation.
Plant Mol. Biol. 23: 1091-1104, 1993.

Tackaberry ES, Dudani AK, Prior F, Tocchi M, Sardana R, Altosaar I, Ganz PR. Development of biopharmaceuticals in plant expression systems: cloning, expression and immunological rectivity of human cytomegalovirus glycoprotein B (UL55) in seeds of transgenic tobacco.
Vaccine 17:3020-3029, 1999.

Tacket CO, Mason HS, Losonsky G, Clements JD, Levine MM, Arntzen JC.
Immunogenicity in humans of a recombinant bacterial antigen delivered in transgenic potato.
Nat. Med. 4:607-609, 1998.

Tacket CO, Mason HS, Losonsky G, Estes MK, Levine MM, Arntzen JC.

Human immune responses to a novel Norwalk virus vaccine delivered in transgenic potatoes. J.Infect.Dis. 182: 302-305, 2000.

Tacket CO, Pasetti M, Edelman R, Howard JA, Streatfield S.
Immunogenicity of recombinant LT-B delivered orally to humans in transgenic corn. Vaccine 22:4385-4389, 2004.

Tavladoraki P, Benvenuto E, Trinca S, De Martinis D, Cattaneo A, Galeffi P.
Transgenic plants expressing a functional single-chain Fv antibody are specifically protected from virus attack.
Nature 366: 469-472, 1993.

Terpe K.
Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems.
Appl. Microbiol. Biotechnol. 60:523–533, 2003.

Thanavala Y, Yang Y, Lyons P, Mason HS, Arntzen C.
Immunogenicity of transgenic plant-derived hepatitis B surface antigen.
Proc. Natl. Acad. Sci. USA 92:3358-3361, 1995.

Thanavala Y, Mahoney M, Pal S, Scott A, Richter L, Natarajan N, Goodwin P, Arntzen CJ, Mason HS.
Immunogenicity in humans of an edible vaccine for hepatitis B.
Proc. Natl. Acad. Sci. USA 102:3378-3382, 2005.

Thoyts PJE., Millichip MI, Stobart AK, Griffiths WT, Shewry PR, Napier JA.
Expression and *in vitro* targeting of a sunflower oleosin.
Plant Mol. Biol. 29: 403-410, 1995.

Tregoning JS, Nixon P, Kuroda H, Svab Z, Clare S, Bowe F, Fairweather N, Ytterberg J, Wijk KJ, Dougan G, Maliga P.
Expression of tetanus toxin Fragment C in tobacco chloroplasts.
Nucleic Acid Res. 31: 1174-1179, 2003.

Trkola A, Purtscher M, Muster T, Ballaun C, Buchacher A, Sullivan N, Srinivasan K, Sodroski J, Moore JP, Katinger H.
Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1.
J. Virol. 70: 1100–1118, 1996.

Tzen JTC, Huang AHC.
Surface structure and properties of plant seed oil bodies.
J. Cell. Biol. 117: 327-335, 1992.

Tzen JTC, Cao Y, Laurent P, Ratnayake C, Huang AHC.
Lipids, proteins, and structure of seed oil bodies from diverse species.
Plant Physiol. 101: 267–276, 1993.

Tzen JTC, Peng C, Cheng D, Chen ECF, Chiu JMH.
A New Method for Seed Oil Body Purification and Examination of Oil Body Integrity Following Germination.
J. Biochem. 121: 762-768, 1997.

Tzfira T, Li J, Lacroix B, Citovsky V.
Agrobacterium T-DNA integration: molecules and models.
Trends Genet. 20: 375-83, 2004.

Tuboly T, Yu W, Bailey A, Degrandis S, Du S, Erickson L, Nagy E.
Immunogenicity of porcine transmissible gastroenteritis virus spike protein expressed in plants.
Vaccine 18:2023-2028, 2000.

Ulmer JB, Valley U, Rappuoli R.
Vaccine manufacturing: challenges and solutions.
Nature Biotechnology 24: 1377-1383, 2006.

Vaeck M, Reynaerts A, Höfte H, Jansens S, De Beuckeleer M, Dean C, Zabeau M, Van Montagu M, Leemans J.
Transgenic plants protected from insect attack.
Nature 328: 33–37, 1987.

van Rooijen GJ, Moloney MM.
Plant seed oil-bodies as carriers for foreign proteins.
Biotechnology 13:72-77, 1995a.

van Rooijen GJ, Moloney MM.
Structural requirements of oleosin domains for subcellular targeting to the oilbody. Plant Physiol. 109: 1353–1361, 1995b.

Vidi PA, Kessler F, Bréhélin C.
Plastoglobules: a new address for targeting recombinant proteins in the chloroplast. BMC Biotechnology 7:4-16, 2007.

Wahlroos T, Soukka J, Denesyuk A, Wahlroos R, Korpela T, Kilby NJ.
Oleosin expression and trafficking during oil body biogenesis in tobacco leaf cells.
Genesis 35: 125–132, 2003.

Walsh SR, Bhardwaj N, Gandhil RT.
Dendritic cells and the promise of therapeutic vaccines for human immunodeficiency virus (HIV)-1.
Curr. HIV Res. 1:205-16, 2003.

Wang W, Vignani R, Scali M, Sensi E, Tiberi P, Cresti M.
Removal of lipid contaminants by organic solvents from oilseed protein extract prior to electrophoresis.
Analytical Biochemistry 329: 139–141, 2004.

Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, Salazar-Gonzalez JF, Salazar MG, Kilby JM, Saag MS, Komarova NL, Nowak MA, Hahn BH, Kwong PD, Shaw GM.
Antibody neutralization and escape by HIV-1.

Nature 422: 307–312, 2003.

Williamson JD, Hirsch-Wyncott ME, Larkins BA, Gelvin SB.
Differential accumulation of a transcript driven by the CAMV 35S promoter in transgenic tobacco.
Plant Physiol. 90: 1570-1576, 1989.

Witcher DR, Hood EE, Peterson D, Bailey M, Bond D, Kusnadi A, Evangelista R, Nikolov Z, Wooge C, Mehigh R et al.:
Commercial production of β-glucuronidase (GUS): a model system for the production of proteins in plants.
Mol. Breed. 4:301-312, 1998.

Wu CC, MacCoss MJ, Howell KE, Matthews DE, Yates JR.
Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis.
Anal. Chem. 76: 4951-4959, 2004.

Wurm FM.
Production of recombinant proteins therapeutics in cultivated mammalian cells.
Nat. Biotech. 22: 1393-1398, 2004.

Yang DC, Guo FL, Liu B, Huang N, Watkins SC.
Expression and localization of human lysozyme in the endosperm of transgenic rice. Planta 216: 597-603, 2003.

Yates JR, Eng JK, McCormack AL.
Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases.
Anal. Chem. 67: 3202-3210, 1995.

Yewdell J, Antón LC, Bacik I, Schubert U, Snyder HL, Bennink JR.
Generating MHC class I ligands from viral gene products.
Immunol. Rev. 172: 97-108, 1999.

Zhang X, Mason H .
Bean Yellow Dwarf Virus replicons for high-level transgene expression in transgenic plants and cell cultures.
Biotechnol. Bioengineering 93: 271-279, 2005.

Zhong W, Reche PA, Lai CC, Reinhold B, Reinherz EL.
Genome wide characterization of a viral cytotoxic T lymphocyte epitope repertoire.
J. Biol. Chem. 278: 45135–44, 2003.

Zhou F, Badillo-Corona JA, Karcher D, Gonzalez-Rabade N, Piepenburg K, Borchers AM, Maloney AP, Kavanagh TA, Gray JC, Bock R.
High-level expression of human immunodeficiency virus antigens from the tobacco and tomato plastid genomes.
Plant Biotechnol J. 2008 Epub ahead of print

Zhou T, Xu L, Dey B, Hessell AJ, Van Ryk D, Xiang SH, Yang X, Zhang MY, Zwick MB, Arthos J, Burton DR, Dimitrov DS, Sodroski J, Wyatt R, Nabel GJ, Kwong PD.
Structural definition of a conserved neutralization epitope on HIV-1 gp120.
Nature 445: 732–37, 2007.

Zolla-Pazner S, Cohen SS, Krachmarov C, Wang S, Pinter A, Lu S.
Focusing the immune response on the V3 loop, a neutralizing epitope of HIV-1 gp120 envelope.
Virology 372: 233-246, 2008.

Zwick MB, Jensen R, Church S, Wang M, Stiegler G, Kunert R, Katinger H, Burton DR.
Anti-human immunodeficiency virus type 1 (HIV-1) antibodies 2F5 and 4E10 require surprisingly few crucial residues in the membrane-proximal external region of glycoprotein gp41 to neutralize HIV-1.
J. Virology 79: 1252–1261, 2005.