Developing a novel hepatitis B core – based antigen presentation system

by

Hadrien Peyret

This thesis is submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy at the University of East Anglia.

John Innes Centre, Norwich, UK

September 2014

Declaration

I hereby certify that the work contained within this thesis is my own original work, except where due reference is made to other contributing authors. This thesis is submitted for the degree of Doctor of Philosophy at the University of East Anglia and has not been submitted to this or any other university for any other qualification.

Hadrien Peyret

Abstract

Plant-produced proteins of pharmaceutical interest are beginning to reach the market, and the advantages of transient plant expression systems are gaining increasing recognition. In parallel, the use of virus-like particles (VLPs) has become standard in vaccine design. The pEAQ-*HT* vector derived from the cowpea mosaic virus – based CPMV-*HT* expression system has been shown to allow the production of large amounts of recombinant proteins, including VLPs, in *Nicotiana benthamiana*. Moreover, previous work demonstrated that a tandem fusion of the core antigen (HBcAg) of hepatitis B virus (HBV) could direct the formation of core-like particles (CLPs) in plants.

The work presented here demonstrated that the tandem core system is better suited for the plant-based production of CLPs presenting foreign antigens than SplitCore technology. It was shown that tandem core technology allows the plant-based production of CLPs which are suitable for the presentation of antigens either via chemical coupling or through antibody-antigen interactions. Of particular significance was the successful display of single domain antibody fragments of camelid origin (nanobodies, or VHH). The resulting "tandibody" particles, as they are named here, can bind to their cognate antigen to yield CLPs covered in the antigen of interest. Furthermore, it was shown that target antigens can be attached to CLPs via a fusion partner, raising the possibility of the development of a universal generic antigen-display platform.

In addition to the transient expression work on HBcAg, lines of stably transformed *N. benthamiana* were created which constitutively expressed human gastric lipase (hGL), an enzyme of medical importance. In the future, this approach could be used to create transgenic plants constitutively expressing a universal tandibody, thus obviating the need for infiltration.

Acknowledgements

First and foremost, I would like to thank my supervisor, George P. Lomonossoff, for being much more than a supervisor. I could not have done any of this without his encouragement, insight, enthusiasm, and advice. George has given me the freedom to explore new paths, try new things, gain new skills, and put myself forward in the scientific community. Of equal importance, he has given me the freedom to fail, which has given me invaluable experience as a scientist. I hope I never forget the lessons that I learned from George; from the importance of planning experiments properly, to the importance of drinking at conferences. I would also like to thank my secondary supervisor, Alison Smith, for her helpful insight and reassurance. Pooja Saxena also deserves a big thank you for being the first person to tell me about all the cool things that happen in the Lomonossoff lab, and for supervising me during my rotation. And speaking of rotations, I would like to thank Nick Brewin, Jonathan Jones, Alexandre Robert-Seilaniantz, Michael McArthur, and all of the rotation students for giving me the opportunity to make an informed decision about what to do with my life.

I have been hugely fortunate to work in such a fun lab, and I want to thank all Lomonossoff group members past and present, as well as the members of the Biological Chemistry department, for creating such a lovely environment, both during coffee breaks and at the bench. Of particular importance in this regard is Keith Saunders, who was, on a daily basis, a friendly, supportive and helpful guide as I walked in the shadow of the valley of data. I am also hugely grateful to Eva Thuenemann, for all of her help and advice, as well as for writing the bible. I also wish to thank all of my collaborators, Dave Rowlands, Nicola Stonehouse, Clare Stevenson, and Robert Gilbert. Of course, I also acknowledge all those professionals without whom my job would have been much harder: Andrew Davis, Gerhard Saalbach, Sarah Tolland, Kim Findlay, Elaine Barclay, and the staff in the media kitchen and the glasshouse.

Last but not least, special thanks go to my family and Lucie for all of their love and support.

Table of Contents

List of Figures

Abbreviations

aa	Amino acid
bp	Base pairs
BPV	Bovine papillomavirus
BTV	Bluetongue virus
CaMV	Cauliflower mosaic virus
CDR	Complementarity-determining region
CLP	Core-like particle
CPMV	Cowpea mosaic virus
Cryo-EM	Cryo-electron microscopy
C-terminus	Carboxy terminus
delRNA-2	Deleted RNA-2 expression system (CPMV)
DGM	Dynamic Gastric Model
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
dpi	Days post infiltration
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
ER	Endoplasmic reticulum
EV	Empty vector control. This refers to plants agroinfiltrated with bacteria carrying the pEAQ-*HT* plasmid which does not contain a GOI in the MCS.
FMDV	Foot and mouth disease virus
FWT	Fresh weight tissue
GFP	Green fluorescent protein
GOI	Gene of interest
HBV	Hepatitis B virus
HBcAg	Hepatitis B core antigen
HBsAg	Hepatitis B surface antigen
hGL	Human gastric lipase
HIV	Human immunodeficiency virus
HPV	Human papillomavirus
HRP	Horseradish peroxidase
HT	HyperTrans
Ig	Immunoglobulin
INPACT	In-plant activation
IRES	Internal ribosome entry site
KD	Lysine-Aspartate linker designed to facilitate chemical conjugation. The peptide sequence is GDDDKKKKKKDDDS
kDa	kilodalton
LB	Left border of T-DNA
LB (media)	Luria-Bertani broth
m-α-phOx	Monomeric HBcAg construct based on m-EL with the anti-phOx scFv sequence inserted in the e1 loop
MCS	Multiple cloning site
m-EL	Monomeric HBcAg construct based on t-EL with no protein inserted in the e1 loop (empty-loop)
m-KD	Monomeric core HBcAg construct based on m-EL with KD peptide inserted in the e1 loop
mRNA	Messenger RNA
MS	Murashige and Skoog
MWCO	Molecular weight cut-off

N-terminus	Amino terminus
OD$_{600}$	Optical density measured at 600 nm wavelength
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pEAQ	Easy and quick plasmid
PEG	Polyethylene glycol
phOx	phenyl-oxazolone
PVX	Potato virus X
RB	Right Border
RdRp	RNA-dependent RNA polymerase
RNA	Ribonucleic acid
scFv	Signle-chain variable fragment
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
sCore-sGFP	SplitCore HBcAg construct with sGFP used in a split GFP context. Composed of CoreN-sGFP1-10 and CoreN-sGFP11.
sGFP	Solubility-enhanced GFP
SPR	Surface plasmon resonance
TBE	Tris/Borate/EDTA buffer
TBSV	Tomato bushy stunt virus
t-α-phOx	A tandem HBcAg construct based on t-EL with the anti-phOx scFv sequence inserted in the C-terminal e1 loop
T-DNA	Transfer DNA
t-EL	Hetero-tandem core HBcAg construct with no protein inserted in the e1 loop (empty loop)
TEM	Transmission electron microscopy
Ti	Tumour-inducing
t-KD	Tandem core HBcAg construct based on t-EL with KD peptide inserted in C-terminal e1 loop
TMV	Tobacco mosaic virus
t-sGFP	Tandem core HBcAg construct based on t-EL with sGFP inserted in the C-terminal e1 loop
TYDV	Tobacco yellow dwarf mastrevirus
τ-GFP	Anti-GFP tandibody. A tandem core HBcAg construct based on t-EL with an anti-GFP nanobody sequence inserted in the C-terminal e1 loop
τ-gp120	Anti-gp120 tandibody. A tandem core HBcAg construct based on t-EL with an anti-gp120 nanobody sequence inserted in the C-terminal e1 loop
τ-p24	Anti-p24 tandibody. A tandem core HBcAg construct based on t-EL with an anti-p24 nanobody sequence inserted in the C-terminal e1 loop
TVCV	Turnip vein-clearing virus
UV	Ultraviolet
UTR	Untranslated region
VHH	Variable region of the heavy chain of a heavy-chain antibody (synonym: nanobody)
VLP	Virus-like particle
wt	Wild-type

Chapter 1 : General Introduction

This thesis describes the development of a novel technique for the display of antigens of interest on the surface of a virus-like particle (VLP) using a plant expression system. The following introduction will first introduce the field of plant molecular farming and the use of plants to produce proteins of interest. The CPMV-*HT* plant expression system which was used throughout this thesis will be described in detail. The concept of a VLP will be introduced, with particular emphasis on the production of VLPs intended for use as vaccines. The hepatitis B virus (HBV) core antigen (HBcAg) protein will then be introduced, and the history of HBcAg manipulation for antigen-display purposes will be reviewed. Finally, the main objectives and hypotheses of the thesis will be presented.

I - Plant molecular farming

Plant molecular farming refers to the science of producing large quantities of a molecule of interest in plants, through the use of genetic manipulation. The use of plants as expression systems for the production of recombinant proteins has emerged as an attractive alternative to systems based on bacteria, yeast, or animal cells. Producing proteins in plants potentially has advantages in terms of cost as well as safety, since plant pathogens do not infect mammals, and mammalian pathogens do not infect plants, thus reducing contamination risks when pharmaceutical proteins are produced (Kusnadi *et al.*, 1997; Lico *et al.*, 2008). Moreover, certain plants such as *Nicotiana benthamiana* can be grown in high density and still produce large amounts of biomass in a matter of weeks. There are essentially two approaches to producing heterologous proteins in plants – stable transformation (either nuclear or plastid) and transient expression. Stable transformation involves the production of true-breeding lines

of genetically transformed plants. This approach has advantages in terms of reproducibility and potential large-scale production; however it is often very time-consuming and is unsuitable for the rapid screening of a wide variety of different constructs (Kusnadi *et al.*, 1997). Transient expression in plants allows for production of high titres of recombinant proteins in a matter of days in what is essentially a batch process (Pogue *et al.*, 2002). Furthermore, techniques have been developed for scaling-up production of transiently-expressed proteins in plants (D'Aoust *et al.*, 2010).

This introduction, like this thesis, will focus on the production of recombinant proteins, which are historically the most widely-produced type of molecule in modern plant molecular farming (as opposed to secondary metabolites). The history of this field of science dates back to the 1970's, when it was discovered that the gram-negative bacterium *Agrobacterium tumefaciens* contains a special plasmid, named Ti (tumour-inducing), which is capable of directing the transfer of part of the plasmid (the transfer DNA, or T-DNA), to the plant cell, and integrating the T-DNA into the plant genome (Chilton *et al.*, 1977). This led to different research groups replicating this natural process *in vitro*, by infecting tobacco cells (Marton *et al.*, 1979) or protoplasts (Wullems *et al.*, 1981b) with *A. tumefaciens* and regenerating genetically modified shoots from these cells. At the time, the only way to obtain complete plants from such modifications was to graft regenerated shoots on to healthy tobacco plant root stock, and this allowed the plant to flower, set seed, and produce transgenic offspring (Wullems *et al.*, 1981a). However, it was quickly shown that disabling the tumour-controlling genes of the Ti plasmid allowed plants to be regenerated directly after transformation with *A. tumefaciens* (Greve *et al.*, 1982; Barton *et al.*, 1983). Such mutated varieties of *Agrobacterium* were called "disarmed strains". Up to this point, plants were only transformed with a native *A. tumefaciens* plasmid, so the introduced transgenes were restricted to bacterial opine synthesis and tumour-inducing genes.

The leap from research to biotechnology was made in 1983. Heterologous genes were inserted into the T-DNA: yeast alcohol dehydrogenase and bacterial neomycin phosphotransferase (which confers kanamycin resistance) were transferred to plants alongside the native nopaline synthase (*nos*) gene. While the genes were shown to be present, only nopaline (the product of the *nos* gene) was shown to be produced (Barton *et al.*, 1983). However, it was very quickly realised that this could be a promoter issue, so antibiotic resistance genes were placed downstream of the *nos* gene promoter, allowing the expression of antibiotic resistance genes in transgenic plants (Herrera-Estrella *et al.*, 1983). Another significant breakthrough came when it was discovered that the *vir* (virulence) genes on the Ti plasmid which direct the transfer of the T-DNA into the plant cell and its integration into the plant genome could be dissociated from the T-DNA itself. That is to say, the *vir* genes could be placed on one plasmid where it could act as genetic background, while the T-DNA could be placed on a separate small plasmid that would be easy to manipulate in *Escherichia coli*. The T-DNA on this "binary vector", as it became known, could be modified to contain a selection marker (such as a kanamycin resistance gene) and a multiple cloning site for insertion of heterologous genes of interest (de Framond *et al.*, 1983; Hoekema *et al.*, 1983; Bevan, 1984).

These binary vectors have been continually improved ever since as a result of new research into the T-DNA transfer process. For example, it was discovered that it was more effective to place the gene of interest (GOI) near the right border (RB) of the T-DNA, and the selection marker near the left border (LB), because the RB is the first to be integrated into the plant genome (Wang *et al.*, 1984), and because there can sometimes be incomplete transfer of the T-DNA at the LB (Rossi *et al.*, 1996). This ensures that transformants displaying antibiotic resistance will also contain the GOI. This paradigm shift in biotechnology did not go unnoticed by industry: indeed one of the first binary vectors was created by a team from Monsanto (Fraley *et al.*, 1983), who considered that the field was now wide open for any gene, from any species, to be expressed in plants. The primary result of this was the rapid development of

industrial-scale genetic engineering technology for agricultural purposes (Marshall, 2012), but at the same time it was also the birth of plant molecular farming.

The next big development came when leaf-disc transformation was developed: this allowed transgenic tobacco plants to be regenerated from calli obtained from transformed leaf discs, as opposed to protoplasts, which made the production of transgenics much simpler (Horsch and Klee, 1986). At the same time, the field of plant virology merged with the nascent field of plant biotechnology when it was shown that inserting tandem copies of the cauliflower mosaic virus (CaMV) genome into a T-DNA allowed a viral infection to be initiated on a grown plant via *Agrobacterium*-mediated gene transfer (Grimsley *et al.*, 1986). This technique, called "agroinfection", and later, "agroinoculation", involved pipetting a small volume of a suspension of modified bacteria onto an abraded plant leaf. From there, the bacteria infect the plant with the virus, which then spreads systemically. At the same time, genetic components of viral origins were being deployed to improve the efficacy of binary vectors. The most obvious example is the constitutive 35S promoter from CaMV, which was first introduced as a feature of a binary vector by Pietrzak *et al.* (1986).

Interestingly, while the developments described above had mostly come from research on solanaceous species, and *Nicotiana* in particular, the next important development came from research in *Arabidopsis thaliana*. Stable transgenic plants could be produced *in planta* (i.e. without the need for tissue culture) with this species ever since the description of agrobacterium-mediated transformation of germinating seeds (Feldmann and Marks, 1987), but this was not deemed to be a very efficient technique in terms of frequency of transformants in the progeny. To improve the transformation efficiency of *Arabidopsis* and to obtain large numbers of transgenic progeny from the original transformant, agroinfiltration was invented (Bechtold *et al.*, 1993). This technique, described in detail by the original team in a subsequent methods book chapter (Bechtold and Pelletier, 1998), involves submerging an

entire *Arabidopsis* plant in a suspension of *Agrobacterium* and applying a vacuum, which is then gently released to allow the bacteria to infiltrate the intercellular spaces of the plant. While this technique was originally aimed at producing transgenic *Arabidopsis* (and is still known as "floral dipping" for this purpose), it soon became adapted to the transient production of heterologous proteins such as antibodies in the excised leaves of bean and tobacco plants (Kapila *et al.*, 1997; Vaquero *et al.*, 1999). The first example of agroinfiltration on leaves of whole plants, however, seems to come from Schob *et al.* (1997), in a paper aimed at studying transgene silencing in tobacco: this is the first description of what is now known as syringe-infiltration, an alternative to vacuum infiltration. Bizarrely, the first description of vacuum infiltration of whole plants, and not just excised leaves, for transient expression purposes, was not published until almost a decade later (Marillonnet *et al.*, 2005), with detailed methods described after that (Gleba *et al.*, 2007; Garabagi *et al.*, 2012). Until then, it seems transient expression on whole plants was either done by syringe infiltration (which is more practical on a very small scale) or agroinfection with replicating viral vectors, as in Liu and Lomonossoff (2002). A diagram summarising transient expression via agroinfiltration is provided in figure 1.1.
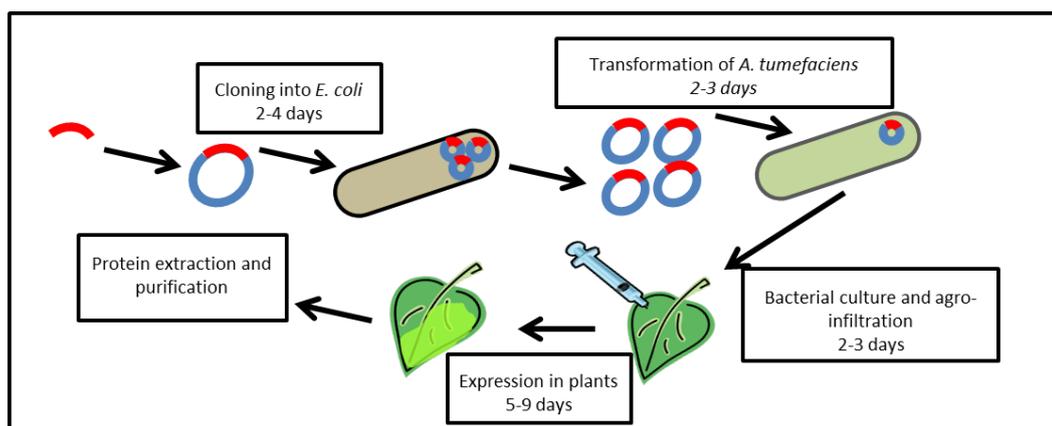
Figure 1.1 Transient expression via agroinfiltration. The gene of interest is cloned into an expression vector and propagated in *E. coli*. After sequence verification it is transformed into *A. tumefaciens.* These are used to infiltrate the intercellular space of *N. benthamiana* leaves, where the gene is transferred to the plant cells. After 5-9 days the agroinfiltrated leaves are harvested for protein extraction and purification.

Replicating viral vectors, in which a GOI is integrated in a viral genome, became inherently linked to agroinfiltration techniques when it was discovered that combining agrobacterial transfection efficiency with viral replication could allow "deconstructed" viral vectors to be developed. This involved removing viral genes which were not strictly necessary to the production of the recombinant protein, such as those coding for the viral coat protein, and positioning different viral functions on separate binary plasmids. Probably the most famous example of this is the magnICON system, in which a hybrid tobamovirus genome with elements from tobacco mosaic virus (TMV) and turnip vein-clearing virus (TVCV) is split into three components which are co-infiltrated into the same plants leaves from a mix of three *Agrobacterium* cultures: the 3' module contains the gene of interest in place of the sequence of the viral coat protein along with the *nos* terminator, the 5' module contains the *Arabidopsis* actin 2 (ACT2) promoter along with the tobamovirus polymerase and movement protein genes (the latter of which contains a subgenomic promoter), while the recombinase module containing the PhiC31 integrase gene (from *Streptomyces* phage C31) acts to fuse the 3' and 5' modules together in the nucleus to allow the formation of a complete tobamovirus-based replicon (Fig. 1.2). Upon transcription, the RNA is able to replicate, spread from cell to cell within the leaf (though not throughout the entire plant), and produce high levels of the protein of interest (Marillonnet *et al.*, 2004; Marillonnet *et al.*, 2005). This system has been used to produce a wide variety of proteins in plants, including antigens (Webster *et al.*, 2009), antibodies (Grohs *et al.*, 2010), and VLPs (Matić *et al.*, 2012).

Deconstructed viral vectors are now also being used to make stable transgenics, as with the INPACT system (Dugdale *et al.*, 2013). This is based on a geminivirus expression system (from tobacco yellow dwarf mastrevirus, or TYDV), in which a simplified version of the viral genome including the GOI is integrated stably into transgenic plants, in a way that is analogous to lysogenic bacteriophages (Fig. 1.2). Upon induction with ethanol, the viral genome is excised from the chromosome, circularises, and this newly formed replicon replicates as the GOI is

transcribed and translated. Another deconstructed viral vector, named delRNA-2 (Sainsbury *et al.*, 2008), is based on the cowpea mosaic virus (CPMV) RNA-2 (Fig. 1.2). The genesis of this system, along with its successor, are described in the following section, which details the development of a non-replicating system based on cowpea mosaic virus (CPMV), called the CPMV-*HT* system, which gave rise to the pEAQ series of expression vectors, which were used throughout this thesis. Part of the following section has already been published in a review article (Peyret and Lomonossoff, 2013), available in appendix 2.
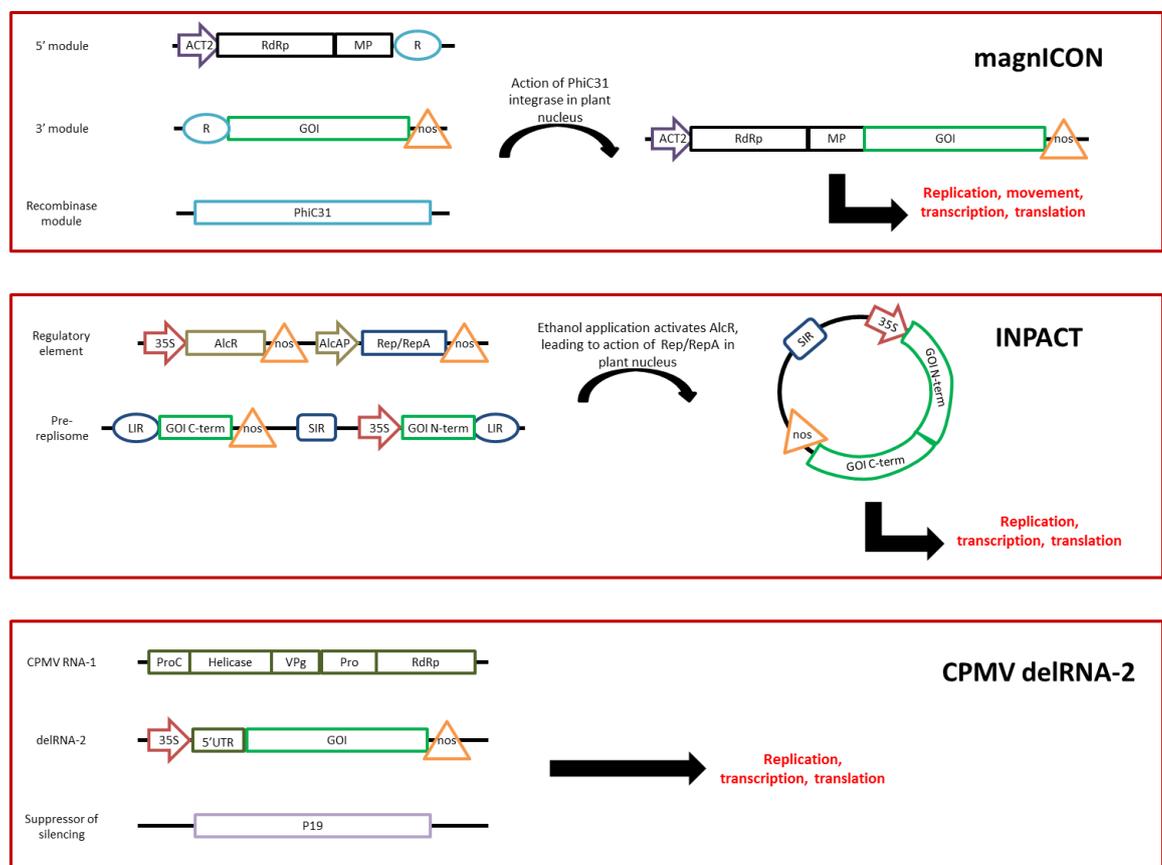
Figure 1.2 Simplified diagrams of three different deconstructed viral vector systems. Top: the magnICON expression system is a tobamovirus-based deconstructed viral vector composed of three modules. The recombinase module fuses the 3' module containing the GOI to the 5' module containing the promoter, viral polymerase and movement protein. This allows replication, cell-to-cell movement, transcription and translation of the GOI-containing replicon. Middle: the INPACT expression system is a TYDV-based deconstructed viral vector composed of two elements which are stably integrated (as opposed to transiently expressed) into the plant genome. Upon induction by ethanol, the regulatory element excises the GOI-containing pre-replisome from the plant genome and circularises it to form a functional replisome. This allows replication, transcription and translation in every cell of the plant. Bottom: the CPMV delRNA-2 system is based on the bipartite genome of CPMV. In the RNA-2 component, everything downstream of the 5'UTR is replaced with the GOI, and this component is replicated thanks to the genes present on RNA-1. Transcription and translation take place, with the suppressor of silencing acting to increase transcript availability.

II - The CPMV-*HT* system and the pEAQ vectors

CPMV is a bipartite virus in the family Comoviridae, with a genome consisting of separately-encapsidated RNA-1 (6.0kb) and RNA-2 (3.5kb) molecules, each of which has a single long open reading frame encoding a polyprotein. RNA-1 encodes the viral replication machinery and the 24K protease, which is responsible for processing the polyproteins encoded by both RNAs at a number of specific sites. This activity by 24K includes releasing itself from the RNA-1 polyprotein (Goldbach and Wellink, 1996). RNA-2, which is entirely dependent on RNA-1 for its replication, encodes both viral coat proteins and the viral movement protein. It has been shown that most of the RNA-2 polyprotein can be deleted and replaced with a foreign sequence without abolishing the ability of the molecule to be replicated by RNA-1 (Rohll *et al.*, 1993; Cañizares *et al.*, 2006), leading to the development of replication-competent deleted RNA-2 (delRNA-2) constructs. These had advantages over the previous vectors based on full-length RNA-2 molecules both in terms of biocontainment and the size of insert tolerated. However, because of their inability to spread from cell-to-cell and the absence of the small (S) coat protein, which acts as the native suppressor of gene silencing (Liu *et al.*, 2004), these delRNA-2 constructs had to be introduced into leaves by agro-infiltration (Liu and Lomonossoff, 2002) in the presence of RNA-1 and a heterologous suppressor (Cañizares *et al.*, 2006). The delRNA-2 approach was successfully used to express a number of proteins (Sainsbury *et al.*, 2009a). However, to preserve the ability of delRNA-2 constructs to be replicated by RNA-1, it was essential to retain the sequence of the first 512 nucleotides at the 5' end of RNA-2, including two in-frame AUGs at positions 161 and 512, as well as the entire RNA-2 3'UTR (Rohll *et al.*, 1993). This required the precise in-frame positioning of the heterologous sequence between AUG512 and the 3' UTR which made insertion of

heterologous sequences into delRNA-2 constructs a cumbersome two-step process (Sainsbury *et al.*, 2009a).

During these studies, it was noticed that co-inoculation of delRNA-2 constructs with the powerful P19 suppressor of gene silencing from tomato bushy stunt virus (TBSV) stabilised the mRNA transcribed from the incoming T-DNA to such an extent as to make replication by RNA-1 unnecessary to achieve high levels of protein expression. Through studies originally intended to simplify cloning, it was found that removal of AUG 161 caused a 10-fold increase in yield of an inserted GFP gene through increased translational efficiency of the mRNA transcribed from the T-DNA (Sainsbury and Lomonossoff, 2008). A further enhancement was obtained when an additional out-of-frame upstream AUG at 115 was also removed, despite the fact that removing AUG 115 alone decreases translation compared to wild-type. These unexpected results indicate that AUG 161 inhibits translation initiation at AUG 512. This may be explained by the need to promote replication of RNA-2, a process which requires the activity of RNA-dependent RNA polymerase on a region of an RNA relatively unoccupied by ribosomes. This could also explain the observed effect of removing AUG 115, which initiates an open reading frame that extends beyond AUG 161: AUG 115 might allow ribosomes to bypass AUG 161 and re-initiate at AUG 512 (Sainsbury and Lomonossoff, 2008). Such regulatory activity provided by upstream open reading frames is known to occur in other eukaryotic systems (Meijer and Thomas, 2002).

The new expression system, based on a deleted version of CPMV RNA-2 with a mutated 5'UTR, was shown to enhance expression of GFP, DsRed, the Hepatitis B virus core antigen HBcAg, and human anti-HIV antibody 2G12 (Sainsbury and Lomonossoff, 2008; Sainsbury *et al.*, 2010). Because high level expression is based on enhanced translation rather than replication, the system was termed CPMV-hypertrans or CPMV-*HT*. This early *HT* expression system, which is the direct precursor to the pEAQ vector series, facilitated *in planta* production of Influenza

virus-like particles (VLPs): particles that have the appearance of a virus, but do not contain infectious viral genetic material. Influenza hemagglutinin (HA) is highly immunogenic and, although it is highly variable, it is a good target for narrow host-range VLP vaccines. Influenza HA was produced in *N. benthamiana* using CPMV-*HT* and this protein budded from plant-cell membranes to form vesicular VLPs (D'Aoust *et al.*, 2009). In fact, CPMV-*HT* was determined to provide higher yields of VLPs than the alfalfa plastocyanin gene-based expression system which was previously used. For this reason, CPMV-*HT* was chosen for industrial-scale production of HA VLPs which are currently undergoing clinical trials in a commercial context (D'Aoust *et al.*, 2010).

The pEAQ series of transient expression vectors (Sainsbury *et al.*, 2009b) are designed to make use of the CPMV-*HT* effect to achieve high levels of expression while at the same time facilitate the insertion of heterologous sequences between the modified (*HT*) 5' UTR and the 3' UTR. They are small binary vectors carrying genes essential for replication of the plasmid in *E. coli* and *A. tumefaciens* on the backbone, along with a T-DNA region. The 5.2 kilobase backbone of all pEAQ plasmids carries the *oriV* and *colE1* origins of replication, the *trfA* gene required for plasmid replication, and the *nptIII* prokaryotic neomycin phosphotransferase gene which confers resistance to kanamycin. The T-DNA region varies between pEAQ plasmids to allow for maximum flexibility in cloning and expression, although only the pEAQ-*HT* vector (published on GenBank under accession number GQ497234.1) was used for the work described in this thesis (Fig. 1.3). This specific plasmid contains the P19 suppressor of gene silencing and NPTII eukaryotic kanamycin resistance gene near the LB of the T-DNA, and a multiple cloning site within the CPMV-*HT* cassette near the RB. The design of all pEAQ vectors, including pEAQ-*HT*, is modular, allowing the insertion of multiple CPMV-*HT* cassettes on the same T-DNA. This is important to maximise the efficiency with which multiple proteins can be expressed within the same cell (Montague *et al.*, 2011). Once a gene or genes of interest have been inserted into the appropriate pEAQ vector, the plasmid is transformed into *A. tumefaciens* prior to

inoculation into *N. benthamiana*. The *Agrobacterium* strain of choice for transfer of pEAQ vectors to plants has usually been LBA4404 (Sainsbury *et al.*, 2009b), although strains C58C1, EHA105 and the cysteine auxotroph C58::pEHA105/Cys32 have also been used with success (Sun *et al.*, 2011; Larsen and Curtis, 2012). While the pEAQ vectors have been used for the production of stable transgenic plants, their main use has been for transient expression via agroinfiltration of *N. benthamiana*.
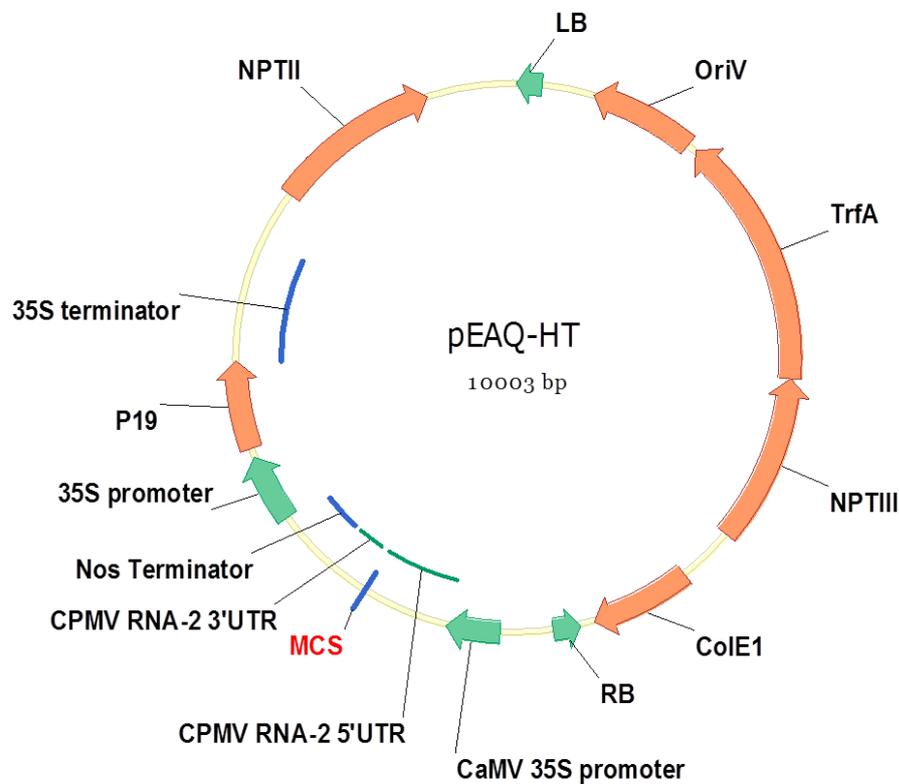
Multiple Cloning Site (MCS): 5' - AgeI – ATG – 6XHis – XmaI – 6XHis – TAG – XhoI – 3'

Figure 1.3 Plasmid map of the pEAQ-*HT* expression vector. Essential elements for plasmid replication in *E. coli* and *A. tumefaciens* are located on the plasmid backbone. In the T-DNA, NPTII provides kanamycin resistance to transformed plant cells, and the P19 suppressor of silencing increases the availability of transgene transcript. The GOI is inserted in the multiple cloning site (MCS) between the modified CPMV RNA-2 5'UTR bearing the *HT* mutations, and the native CPMV RNA-2 3'UTR. These UTRs are themselves flanked by the CaMV 35S promoter and the *nos* terminator.

Though the pEAQ system could be used to produce transgenic suspension cultures (Sun *et al.*, 2011), it proved impossible to regenerate lines of transgenic *N. benthamiana* with constructs expressing wild-type P19 as this interferes with plant regeneration (Saxena *et al.*, 2011). This was addressed through the deployment of a version of the P19 suppressor of gene silencing (P19/R43W) that is not developmentally toxic to plants and therefore compatible with the production of stable transgenic lines (Saxena *et al.*, 2011). Through a point mutation in the P19 gene that substitutes arginine 43 with tryptophan, the suppressor activity of P19 is reduced and ceases to prevent regeneration of fully-grown *N. benthamiana* plants from callus after transformation with *Agrobacterium*. It was found that plants transiently expressing P19/R43W along with GFP produce half as much recombinant GFP as plants transiently expressing wild-type P19 and GFP, but this reduced yield still represents a 7-fold increase compared to transient expression of GFP in the absence of a suppressor of gene silencing. Because the difference between wild-type P19 and P19/R43W is a single nucleotide substitution, the pEAQ-*HT* vector can be used to rapidly test a construct transiently, and if yield is considered adequate, a straightforward site-directed mutagenesis (SDM) reaction on the same plasmid can yield a vector which can itself be tested transiently, but which is also ready for the production of transgenic plants. This is advantageous as the production of lines of stably transformed plants is far more time-consuming than transient expression. It is thought that the R43W mutation preserves P19's capacity to bind siRNAs (crucial for suppression of silencing), but prevents it from binding to microRNAs, which play a role in early plant development. This development has allowed pEAQ to be used for the production of high-yielding, fertile, stable, homozygous transgenic lines that constitutively produce GFP (Saxena *et al.*, 2011) and human gastric lipase (see chapter 7).

The pEAQ vectors have been used to produce a variety of different proteins through several different approaches. Along with the transgenic plants described above, pEAQ has also been used to produce recombinant proteins in cell suspension cultures (Sun *et al.*, 2011; Larsen and

Curtis, 2012). Moreover, numerous groups have produced active enzymes transiently through agroinfiltration of pEAQ vectors. The most relevant for this thesis (see chapter 7) is human gastric Lipase (hGL), which was produced through transient expression in *N. benthamiana*. This enzyme, responsible for degradation of lipids in the human stomach, has uses in in vitro digestion models as well as a potential medical use for patients suffering from pancreatic insufficiency. Previous efforts to produce recombinant hGL, in insect cells or yeast, have suffered from very low yields and poor recovery (Crabbe *et al.*, 1996; Canaan *et al.*, 1998). By using pEAQ, hGL has been produced at levels of about 0.5 g/kg FWT, representing a maximum yield of enzymatic activity of 193 U/g FWT (Vardakou *et al.*, 2012). This corresponds to an activity of about 310 U/mg of protein, which is about three-fold lower than native human-produced hGL. The reasons for this discrepancy are unknown as of yet, but it is possible that the purification process could be further optimised to better preserve enzymatic activity. The recombinant hGL from plants shares important characteristics with its native human-produced version: it is stable at 40 °C as well as at low pH, it exhibits higher affinity for short-chain lipids over long-chain lipids, and crucially, it appears to be resistant to digestion by pepsin, the protease present in the stomach. This resistance is thought to be caused by the glycosylation of the protein, and although the glycosylation profile is predicted to be different for plant-produced hGL from native hGL in terms of the glycans used, this does not appear to compromise resistance to digestion by pepsin.

While hGL was produced in plants with a view to use the enzyme as a reagent or even a medical therapeutic post-purification, pEAQ has also been used to produce enzymes for fundamental research. The expression in *N. benthamiana* of hexahistidine (His) -tagged recombinant OsChia4a, a rice chitinase, allowed research to be carried out on this enzyme's anti-fungal activity as well as on the regulation of the gene coding for this enzyme by the plant hormone jasmonic acid (Miyamoto *et al.*, 2012). While this is the only currently published

example of the pEAQ system being used to study gene function, it is currently being deployed for many such studies worldwide.

It has also been shown that the pEAQ vectors are well suited to the manipulation of entire synthetic pathways through the co-expression of multiple enzymes. The coding sequences for two sesquiterpene synthases from *Artemisia annua*, amorpha-3,11-diene synthase (ADS) and epi-cedrol synthase (ECS), were cloned into pEAQ and expressed in *N. benthamiana* to yields of 90 and 96 mg/kg FWT respectively (Kanagarajan *et al.*, 2012). Moreover, these two enzymes converted the product farnesyl diphosphate into (respectively) amorpha-4,11-diene and epi-cedrol, which are intermediates in the biosynthetic pathway leading to the production of artemisinin, an important antimalarial drug. This indicates that pEAQ allowed relatively high-level expression of these enzymes and that these function *in planta* with appropriate product specificity. Two enzymes from the monocot crop species *Avena sativa* (oat) have also been expressed in order to study their mode of action (Sainsbury *et al.*, 2012b). The oxidosqualene cyclase (OSC) and the cytochrome P450 (CYP450) enzymes were expressed both separately and on the same T-DNA in pEAQ vectors, and both were found to be active *in planta*. The product β-Amyrin was produced by OSC, and when both enzymes were co-expressed, a novel compound was produced which was later identified as an intermediate in the pathway towards avenacin, an important antimicrobial compound produced by *Avena* species (Geisler *et al.*, 2013). This indicates that the pEAQ vectors can be used not only to produce large quantities of a particular enzyme, but also to study their modes of action alone or with other enzymes involved in the same biosynthetic pathway, as well as potentially allowing metabolic engineering: the introduction of entire biosynthetic pathways in plants.

This is made possible, or at least much easier, by the fact that the pEAQ vectors can carry more than one GOI in the T-DNA, which improves co-expression in the same cell as compared to co-infiltration of different constructs carried by separate *Agrobacterium* clones (Montague *et al.*,

2011). This is done by having multiple CPMV-*HT* cassettes on the same T-DNA, so each GOI benefits from the yield-enhancing effects of the expression system. However, when that is undesirable (either because proteins of interest must be present in a specific ratio, or due to protein toxicity), one of the GOI on the T-DNA can be down-regulated by removing the *HT* mutation on its 5'UTR, thus leading to efficient co-expression in the same cell, but not at the same levels (Thuenemann *et al.*, 2013a). This technique is not only useful for the expression of enzymes, but first and foremost for the production of VLPs, the production of which has been the main use of the pEAQ vectors so far, and is of particular relevance to this thesis.

III - Virus-like particles (VLPs)

A virus-like particle (VLP) is a viral particle, usually produced in a recombinant system, which is devoid of viral genome (Roy and Noad, 2008; Plummer and Manchester, 2010; Kushnir *et al.*, 2012). Typically, the viral coat protein(s) or surface protein(s) are expressed in a heterologous system, and these structural proteins self-assemble into nanoparticles that strongly resemble (and sometimes are indistinguishable from) native viral particles. Crucially, the fact that only a small portion of the viral genome is used in their production (namely, the gene(s) coding for the structural proteins) means that VLPs are entirely non-infectious, making them inherently safe to produce, handle, and use as vaccines, without any risk of reversion or recombination with wild virus strains (Minor *et al.*, 1986; Martin *et al.*, 2004; Kushnir *et al.*, 2012; Lee *et al.*, 2012). This is in stark contrast to the currently used non-recombinant vaccines, which require cultures of pathogens to be grown before the pathogen itself is inactivated (killed vaccines), attenuated (live attenuated vaccines), or broken apart for purification of specific protein immunogens (subunit vaccines).

The first ever example of a recombinant vaccine is a 22 nm particle formed in the yeast *Saccharomyces cerevisiae* through heterologous expression of HBV surface antigen HBsAg (Valenzuela *et al.*, 1982; Hilleman, 1987; McAleer *et al.*, 1992). These VLPs mimicked particles derived from the plasma of HBV-infected individuals very closely (Ellis and Gerety, 1989), but obtaining abundant quantities of clean material was much simpler when using recombinant technology. HBsAg VLPs were developed and ultimately licensed for use as human vaccines under the trade names Recombivax®, produced by Merck (Krugman and Davidson, 1987); or later Engerix-B®, produced by GlaxoSmithKline (Crovari *et al.*, 1987). It has since become obvious that these products represented a revolution in vaccine design due to their efficacy and safety (Hilleman, 2011), and other VLP vaccines have since been brought to market, such as the two human papillomavirus (HPV) vaccines Gardasil®, made in yeast by Merck (Shank-Retzlaff *et al.*, 2006); and Cervarix®, made in insect cells by GlaxoSmithKline (Monie *et al.*, 2008).

While all recombinant vaccines have the inherent safety profile associated with producing just one or a few of a pathogen's proteins, VLPs are also much more immunogenic than simple soluble proteins. This is due to the size, geometry, and content of VLPs (Bachmann and Jennings, 2010). Firstly, the size of VLPs (typically between 10 and 100 nm) is ideal for direct traffic into the lymphatic system, which contains the highest concentration of B cells. Antigens larger than 200 nm can only access the lymphatic system by being transported there by antigen-presenting cells (APCs) after phagocytosis, which results in antigen breakdown and presentation of antigen fragments on the surface of APCs. Antigens smaller than 200 nm however, can access the lymphatic system directly in native form, meaning they reach follicular dendritic cells (FDCs) intact, and can be presented on the surface of these in immune complexes for recognition by follicular B cells. At this point, the geometry of VLPs becomes a critical advantage: the symmetrical, quasi-crystalline structure of VLPs (which is absent in soluble protein subunit antigens) acts as a pathogen-associated molecular pattern (PAMP),

which allows the VLPs to be very efficiently cross-linked on the surface of B cells by B cell receptors (membrane-bound IgGs). This cross-linking allows activation of the B cells, causing a germinal centre response: proliferation of B cells.

At the same time as the VLPs stimulate this B cell response, they also stimulate the T cell response. VLPs that penetrate the lymphatic system in a native state can be cross-presented on the surface of CD8$^+$ dendritic cells (which are typically only found in the lymphoid organs), and this cross-presentation allows cytotoxic T lymphocytes to be activated. Moreover, the repetitive geometry of the VLPs favours binding to IgM in the circulatory system through multivalent, high-avidity interactions. This allows recruitment of the complement system: C1q binds the IgM, which in turn bind C3b, and this opsonisation causes complement-mediated phagocytosis by APCs. Once the particle is taken up, it is broken down in the endosomal-lysosomal compartments, and the resulting peptides are presented on the dendritic cell surface on the tips of major histocompatibility complex (MHC) class II molecules, where they stimulate CD4$^+$ T helper cells, causing them to activate. Activated T cells and B cells then "collaborate", allowing differentiation of B cells into plasma cells (that produce circulatory IgGs) and memory B cells, which are essential to long-lasting immunity. Upon challenge after immunisation with a VLP, the memory B cells can quickly differentiate into plasma cells while primary B cells are activated and differentiate into new memory B cells (Zabel *et al.*, 2013). The nucleic acid which is often encapsidated into VLPs also plays a beneficial role in stimulating an immune response. VLPs often encapsidate random host RNA in lieu of the viral genome, and this RNA is, in fact, a ligand for toll-like receptors (TLR) 7 and 8. These come into play at the point where the VLP is taken up by APCs: TLR 7 and 8 are located on the luminal side of endosomes, where they are stimulated by RNA taken up by the endosomal-lysosomal pathway, so the immune system recognises endosomal RNA as a sign of infection (Zabel *et al.*, 2013). These TLR stimulate both the T cell response as well as antibody production. The fact that the TLR ligand is encapsidated within the VLP is thought to reduce the risk of side effects

from VLP vaccines: only the cells which have taken up the antigen are activated by the TLR ligand, which focuses the immune system in a way that cannot happen if the antigen and TLR ligand are simply mixed together and taken up by different cells (Bachmann and Jennings, 2010).

These advantages permit the use of VLPs directly as vaccines, or as presentation systems for heterologous antigens through fusion or conjugation of an antigen on the surface of a VLP. In this context, the ability of VLPs to stimulate a strong B cell response was at one point thought to be potentially problematic, as it was thought that it might lead to "carrier suppression": the suppression of an efficient T cell response against antigens presented on the surface of a VLP when antibodies against the VLP carrier already exist from prior vaccination. This was found to be unlikely to be a major issue by Ruedl *et al.* (2005), who showed that the presence of antibodies to a VLP carrier does not significantly affect overall induction of immunity. It is clear that VLPs combine three crucial advantages (size, repetitive structure, and RNA encapsidation) which make them ideally suited for use as vaccines. This explains why so far, the only recombinant vaccines approved by the American Food and Drug Administration (FDA) are VLP vaccines: all other recombinant subunit vaccine candidates have so far failed to obtain regulatory approval. It is also worth noting that subunit vaccines on the market today typically require the use of adjuvants such as aluminium hydroxide (alhydrogel), which can in fact be seen as crude ways of making antigen subunits take on a particulate form. Bachmann and Jennings (2010) refer to this as the "manufacturer's dirty little secret", by analogy with the "immunologist's dirty little secret", which refers to the use of adjuvants to make innocuous proteins immunogenic (Janeway, 1989). Although all commercially available VLP vaccine formulations contain adjuvant, there is evidence to suggest that in some cases, an antigen fused onto the surface of a VLP is just as protective as the antigen mixed with adjuvant, meaning the use of VLPs could be a way to reduce or eliminate the need for adjuvantation (Wang *et al.*, 2012).

The production of VLPs in plants is part of the wider field of molecular "pharming", or producing proteins of pharmaceutical interest in plants. One of the most common types of protein produced in this regard is candidate vaccines (the other is probably antibodies). This includes subunit vaccine candidates, in which a pathogen-derived protein or protein fragment is produced in plants either stably (Haq *et al.*, 1995; Mortimer *et al.*, 2012; Perez Aguirreburualde *et al.*, 2013; Gorantala *et al.*, 2014) or transiently (Mortimer *et al.*, 2012; Gomez *et al.*, 2013), sometimes linked to a carrier to act as a protein adjuvant. But increasingly, candidate vaccine production in plants has focused on the production of VLPs (Thuenemann *et al.*, 2013b), for the reasons described above. While the HBsAg VLP vaccines were licensed in the mid- 1980's, the field of plant-produced VLPs was not far behind: HBsAg VLPs were the first example of heterologous VLPs produced in plants (Mason *et al.*, 1992). These particles were produced in stable transgenic tobacco plants, and their properties are very similar to those produced in yeast. Other enveloped virus particles have been produced in plants, most notably human immunodeficiency virus (HIV) VLPs, through the co-expression of the Gag polyprotein as well as the gp41 cytosolic and transmembrane component (Kessans *et al.*, 2013); and influenza A VLPs, through the expression (using the CPMV-*HT* system) of haemagglutinin (HA) transmembrane protein (D'Aoust *et al.*, 2009; D'Aoust *et al.*, 2010). Moreover, human papillomavirus (HPV) has been produced in potato tubers (Warzecha *et al.*, 2003) and in tobacco leaves using pEAQ (Matić *et al.*, 2012).

Using a similar approach, bovine papillomavirus (BPV) VLPs were also produced in plants (Love *et al.*, 2012). More impressive is the production of bluetongue virus (BTV) VLPs using pEAQ, which involves the co-expression of four different viral structural proteins in the correct ratios in order to obtain large quantities of correctly-assembled VLPs. These proved to be effective at protecting sheep from challenge (Thuenemann *et al.*, 2013a). Finally, hepatitis B virus (HBV) core antigen (HBcAg) has been produced in plants on numerous occasions (Tsuda *et al.*, 1998; Huang *et al.*, 2006; Mechtcheriakova *et al.*, 2006; Thuenemann *et al.*, 2013b), and it has been

shown to form VLPs, which in this case are more accurately called core-like particles (CLPs), since HBV is an enveloped virus. The production and potential pharmaceutical uses of plant-produced VLPs have been recently reviewed (Scotti and Rybicki, 2013; Thuenemann *et al.*, 2013b). Figure 1.4 shows three different types of VLP produced in plants using the pEAQ vectors. Note that plant-produced HBcAg (the first image), just like wild-type HBcAg, takes on both T=3 and T=4 conformations, although T=4 dominates (Crowther *et al.*, 1994). The most relevant type of VLP for the purposes of this thesis is particles used for antigen-display. HBcAg in particular has played a large role in this field.
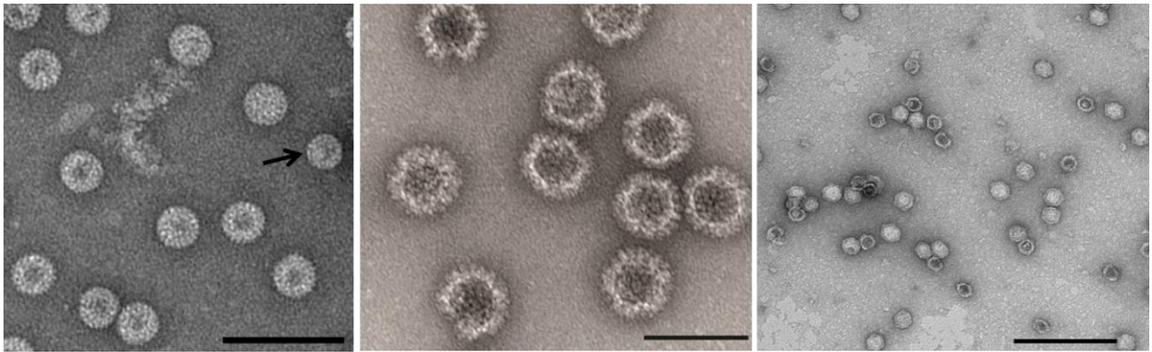
Figure 1.4 Electron micrographs of virus-like particles produced in plants using pEAQ. Left: HBcAg CLPs. The arrow indicates a T=3 particle, whereas the others are T=4. Middle: BTV VLPs, image courtesy of Eva Thuenemann. Right: CPMV VLPs, image courtesy of Alaa Aljabali. Scale bars are 100 nm.

IV - Antigen display

Antigen display refers to the presentation of heterologous antigens on the surface of a VLP. The purpose is to enhance the immunogenicity of the antigen by using the carrier VLP to present a quasi-crystalline ordered repeat antigen structure to the immune system: the VLP moiety acts as both a scaffold and an adjuvant for the antigen of interest (see above and Plummer and Manchester, 2011). Because the point of an antigen-carrier VLP is that the VLP moiety is incapable of causing an infection in the animal host, infectious plant viruses used for

antigen display can be considered antigen-carrier VLPs, even though they are not technically VLPs. There are numerous examples of replicating plant viruses used to display immunogenic epitopes of animal pathogens for vaccine purposes. Chimaeric CPMV particles displaying a foot and mouth disease virus (FMDV) epitope, human rhinovirus epitope, or HIV epitope fused to the S coat protein have been produced in cowpea plants, and these particles could stimulate an immune response against the target epitope in test animals (Usha *et al.*, 1993; Porta *et al.*, 1994; McLain *et al.*, 1995). Moreover, this technique also permitted the development of a vaccine against mink enteritis virus that provides protective immunity in target animals (Dalsgaard *et al.*, 1997). TMV particles have also been used to display heterologous epitopes: a leaky stop codon strategy was used to produce chimaeric TMV particles in which the C-terminus of some of the coat proteins in the rod are fused to epitopes from the malaria parasite (Turpen *et al.*, 1995), influenza virus, or HIV (Sugiyama *et al.*, 1995).

Moreover, engineered antibody fragments called single chain variable fragments (scFv) have been displayed on the surface of potato virus X (PVX) by fusion of the heterologous protein to the viral coat protein via the FMDV 2A peptide, which causes cleavage of some of the polypeptides through a ribosomal stutter mechanism (Smolenska *et al.*, 1998). While this is not *sensu stricto* antigen-presentation, the display of a heterologous protein on the surface of a viral particle is technically very similar whether the heterologous protein in question is an antigen or an antibody fragment. Moreover, the display of antibody fragments on the surface of VLPs is extremely relevant to this thesis. It should therefore be pointed out that while the example cited above is the only one, to my knowledge, of antibody display in plants; examples of antibody display in other systems are legion: scFv have been displayed on the surface of hamster polyomavirus (HaPyV) particles produced in yeast (Pleckaityte *et al.*, 2011); on the surface of measles virus particles (Peng *et al.*, 2003) and murine leukaemia virus particles (Urban *et al.*, 2005) produced in mammalian cells; and on the surface of baculovirus particles produced in insect cells (Kitidee *et al.*, 2010). The display of scFv on the surface of

bacteriophage viruses in bacteria is known as "phage display" technology, and it has been used for years to identify specific antibodies which bind to an antigen of interest (Smith and Petrenko, 1997; Willats, 2002). It is not particularly surprising that scFv display was found to be possible in plants, since the production of unfused scFv and other types of antibodies, both native and engineered (sometimes called "plantibodies"), have been produced in plants numerous times through both stable (Hiatt *et al.*, 1989; Ma *et al.*, 1995; Schouten *et al.*, 1997) and transient expression (Vaquero *et al.*, 1999; Monger *et al.*, 2006; Sainsbury *et al.*, 2009b; Rosenberg *et al.*, 2013).

The more classic antigen display, however, involves fusing an antigen or epitope of interest to a non-replicating VLP in a heterologous expression system. Early successes with this sort of technique showed that the antigenicity of a protein or epitope of interest can be dramatically increased by presenting it on the surface of a VLP carrier: one of the earliest experiments involved fusing an FMDV epitope to the N-terminus of HBcAg CLPs, which led to the formation of chimaeric CLPs which were almost as antigenic as FMDV in animal models (Clarke *et al.*, 1987). It then became widely accepted as fact that any epitope or antigen fused to a VLP carrier would be much more immunogenic than if it were alone, to such an extent that numerous groups have actually produced and tested antigens displayed on VLP carriers without ever comparing them to the antigen or epitope on its own (Birkett *et al.*, 2002; De Filette *et al.*, 2005; Sominskaya *et al.*, 2010; Pastori *et al.*, 2012; Ravin *et al.*, 2012).

This assumption, however, is not foolproof, as there are some examples of antigens being less immunogenic (Arora *et al.*, 2012, 2013), or at least no more immunogenic (Middelberg *et al.*, 2011; Mazeike *et al.*, 2012) when displayed on the surface of a VLP than when administered alone. That having been said, it is true that in general, display on the surface of a VLP clearly enhances the immunogenicity of the target antigen or epitope. This normally holds true when comparing the chimaeric VLP to the antigen alone (Eriksson *et al.*, 2011; Yin *et al.*, 2011; Wang

22

et al., 2012; Skrastina *et al.*, 2013), the antigen mixed with the native VLP (Jegerlehner *et al.*, 2002; Denis *et al.*, 2008; Dhanasooraj *et al.*, 2013), or the chimaeric VLP after denaturation (Chackerian *et al.*, 1999). This demonstrates that it is the particulate structure on which the antigen is displayed that confers the increase in immunogenicity, and not simply the VLP or individual capsid protein acting as a standard adjuvant. This shows that antigen display is a potentially powerful technique which allows the inherent immunogenicity of a given antigen or epitope to be enhanced for vaccine development purposes. The main limitation of this technique is the risk of abolishing capsid formation by inserting a foreign sequence which is either too long or too complex, and thus interferes with carrier protein folding or capsid assembly. For this reason, most of the examples of antigen display, including those cited above, tend to be of short linear epitopes. However, one carrier in particular has the potential to overcome this limitation: the CLP formed by HBcAg.

V - The HBcAg CLP as an antigen carrier

HBcAg is a 20 kDa protein made up of 183-5 amino acids (depending on the strain of HBV). Its secondary structure is rich in α-helices, with the longest ones forming a hairpin structure, with the immunodominant c/e1 loop bridging the ascending and descending helices (Fig. 1.5). Two HBcAg monomers dimerise spontaneously with the hairpin structure forming the dimer interface, and these dimers self-assemble to form either T=3 (90 dimers) or T=4 (120 dimers) core-like particles with the hairpin structures forming characteristic protruding spikes on the surface of the particle (Crowther *et al.*, 1994; Conway *et al.*, 1998; Wynne *et al.*, 1999). The CLP structure in general, and the dimer structure in particular, are very stable. This could be explained by disulphide bridges between monomers, with C61 from each monomer forming a disulphide bridge between the hairpin helices, and C48 from each monomer sometimes forming a bridge at the base of the spike (Zheng *et al.*, 1992; Wynne *et al.*, 1999). These are

not strictly necessary for dimer or capsid formation, since Jegerlehner *et al.* (2002) managed to obtain CLPs after having removed all of the cysteine residues from the sequence, suggesting that these disulphide bridges probably play more of a stabilising role. The C-terminus is very arginine-rich, and truncations at position 176, 149, or 144 are tolerated, as they do not interfere with capsid assembly but do drastically reduce nucleic acid encapsidation (Gallina *et al.*, 1989; Zheng *et al.*, 1992).
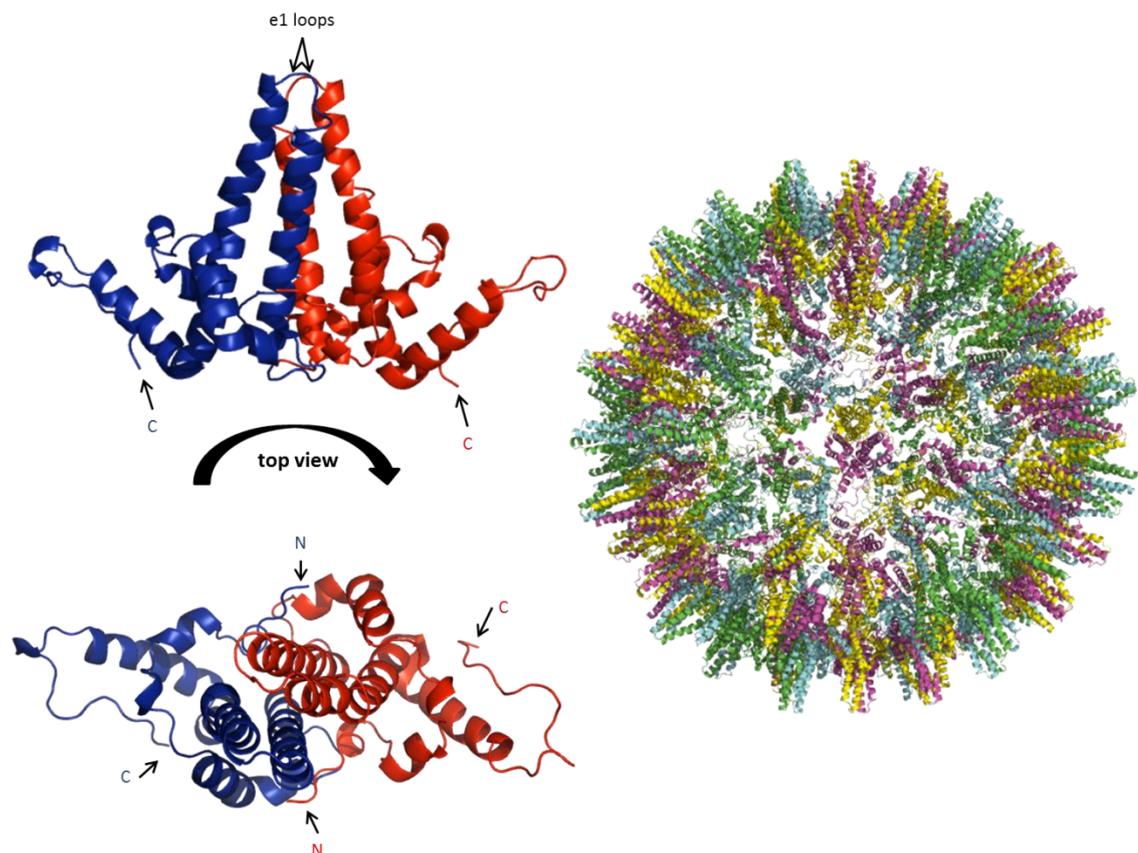
Figure 1.5 The structure of HBcAg and the HBcAg CLP. Left: two HBcAg monomers (in blue and red) assembled to form a dimer, with the e1 loops exposed. The amino (N) and carboxy (C) termini are indicated with arrows. These dimers form the building blocks of the core-like particle (CLP) of HBV (right). Structures were obtained from PDB entry 1QGT (Wynne *et al.*, 1999). Images were generated using PyMOL.

HBcAg is known to be inherently immunogenic, thanks to both B-cell and T-cell epitopes (Francis *et al.*, 1990; Pumpens and Grens, 2001; Cooper *et al.*, 2005). It is a remarkably tolerant

protein in terms of the insertions and fusions that can be made without disrupting protein folding or capsid assembly. There have been numerous studies looking into how fusions and insertions are tolerated in HBcAg, and they have revealed the advantages and disadvantages of inserting sequences at the N-terminus, C-terminus, or the immunodominant e1 loop, located at the tip of the spikes which protrude outwards from the surface of the HBcAg CLP. It is now possible to draw some general conclusions about the possibilities as well as the limits to such modifications, and to devise different strategies to attempt to overcome such limits. Many groups have inserted identical sequences at different positions within HBcAg in order to assess the feasibility and immunogenicity of such insertions (Schödel *et al.*, 1992; Yon *et al.*, 1992; Koletzki *et al.*, 1999; Lachmann *et al.*, 1999; Koletzki *et al.*, 2000). They all conclude that insertions in the e1 loop allow for greater yields of chimeric particles as well as better surface exposure than N- or C-terminal fusions, regardless of the truncation state of the C-terminus. Moreover, e1 insertions reduce the antigenicity of the HBcAg carrier, which is often logs higher than the target insert in N- or C-terminal fusions: because the e1 loop is the immunodominant region, insertions in this area tend to transfer the inherent antigenicity onto the insert protein or peptide (Brown *et al.*, 1991; Chambers *et al.*, 1996; Ulrich *et al.*, 1998; Pumpens and Grens, 1999). This has been extensively reviewed by Pumpens and Grens (2001), and the use of HBcAg as a vaccine platform has more recently been reviewed by Whitacre *et al.* (2009).

A) N-terminal fusions

Fusions of 5 to 120 amino acids have been attempted at the N-terminus of HBcAg, but the latter prevented the formation of particles in *E. coli* (Koletzki *et al.*, 1999; Lachmann *et al.*, 1999). The longest N-terminal fusion which allowed particle formation (in *E. coli*) was a 63 amino acid-long epitope corresponding to three M2e peptides from Influenza virus fused together (De Filette *et al.*, 2005). It is noteworthy that the surface accessibility of N-terminal insertions seems to be sequence-specific rather than size-specific. Indeed, a pentapeptide was

found to be exposed on HBcAg particles (Lachmann *et al.*, 1999), whereas a much longer 30 amino acid-long oligopeptide necessitated the use of a four amino acid – long linker between it and HBcAg to provide surface accessibility (Schödel *et al.*, 1992). However, accessibility of the insert as determined by in vitro methods such as sandwich ELISA, is not always a perfect correlate for antigenicity in vivo. Indeed, an N-terminal 45 amino acid-long Hantavirus epitope was found to be less exposed than identical C-terminal or e1 insertions when assayed by ELISA, but it stimulated greater production of antibodies in animals than the C-terminal insertion (Koletzki *et al.*, 2000). Moreover, the N-terminus was found to be the optimal fusion site (better than the e1 loop) in the case of the M2e epitope: fusing 3 copies of M2e (with cysteines replaced by alanine or serine) at the N-terminus of HBcAg was found to provide a superior immune response in animals than an M2e insertion in the e1 loop (though a single copy was still more immunogenic in the e1 loop than at the N-terminus). Adding a copy of M2e in the e1 loop to the construct already bearing 3 copies of M2e on the N-terminus did not improve immunogenicity, and the particles carrying 2 or 3 copies of M2e at the N-terminus elicited higher antibody titres against M2e than against the HBcAg carrier (De Filette *et al.*, 2005). This is thought to be because the immunogenic potential of the M2e epitope requires its N-terminus to be free and exposed, which is not the case when it is inserted into the e1 loop. These data indicate that N-terminal insertions are normally well tolerated in terms of capsid assembly up to about 65 amino acids, and are most often surface exposed. While they do not normally provide a stronger immune response than e1 insertions, the N-terminus can be the fusion site of choice for epitopes that require a free N-terminus, if multiple repeats of the epitope can be used.

B) C-terminal fusions

The C-terminus of HBcAg has been the subject of multiple experiments, with conflicting results on whether C-terminal inserts are surface-exposed on the assembled capsids. The truncation

state of the HBcAg C-terminus seems to play a crucial role in deciding whether C-terminal inserts are surface exposed: fusions to full-length (183 amino acids) cores or cores truncated after position 155 seem to be buried (Yon *et al.*, 1992; Beterams *et al.*, 2000), while fusions to cores truncated after position 144 or 149 seem to be exposed (Stahl and Murray, 1989; Schödel *et al.*, 1992; Ulrich *et al.*, 1992; Yoshikawa *et al.*, 1993; Grene *et al.*, 1997; Koletzki *et al.*, 2000). This seems to hold true unless the insert is particularly small: the pentapeptide DPAFR was found to be inaccessible when fused after HBcAg position 149 (Lachmann *et al.*, 1999). Interestingly, larger (37-55 amino acids) sequences inserted after position 144 seem to be surface-exposed regardless of whether they are replacing or preceding the C-terminal arginine-rich tail (Borisova *et al.*, 1989). However, the specific sequence of the insert does play a role in determining surface accessibility: when fusing HIV-1 Gag epitopes to a Δ144 truncation, Grene *et al.* (1997) found that while the 90 amino acid insert was surface-exposed, certain regions of it were inaccessible, suggesting that certain sequences will have a tendency to be more buried than others. Certainly the most impressive insertion made in the C-terminus has been the fusion of the entire 167 amino acid (17 kDa) sequence of a bacterial nuclease after HBcAg position 155 (Beterams *et al.*, 2000). This resulted in the formation of capsid particles which contained properly folded, functional nuclease packaged within them. This is particularly impressive given that such a large insert seems to surpass the length of C-terminal fusions that others have tried. For example, Ulrich *et al.* (1992) found that C-terminal insertions of 90 amino acids, but not 100, 189, or 317, allowed particles to form with HBcAg truncated after position 144. Similarly, Koletzki *et al.* (1999) found that 120 amino-acid long inserts prevented the formation of particles in Δ144 truncated HBcAg. In a Δ149 truncated HBcAg, Yoshikawa *et al.* (1993) demonstrated that a C-terminal fusion of 37 amino acids of Hepatitis C Virus core protein allowed particle formation (but that this epitope was not surface accessible), as did 91 amino acids of the same core protein (which was surface accessible). The authors attempted to fuse the full 180 amino acids of the HCV core protein, as well as 2, 3, and

4 copies thereof, but they failed to prove particle formation with such large inserts. Taken together, these data indicate that there is a size limit for C-terminal fusions of about 90-100 amino acids for unstructured peptides, but that this limit can be extended for globular proteins, thus hinting that particle assembly can be limited by spurious interaction between an unstructured insert and the HBcAg polypeptide. Furthermore, C-terminal inserts tend to be surface-exposed provided that they are long enough and fused upstream of the last 34 amino acids of the HBcAg C-terminus.

C) Insertions in the e1 loop

The e1 loop has been recognised as the insertion site of choice for eliciting a strong immune response, and has therefore been the subject of great attention in recent years. It is known that small oligopeptide inserts are well tolerated in the e1 loop, and that these tend to yield higher quantities of particles than equivalent N- or C-terminal fusions (Schödel *et al.*, 1992; Yon *et al.*, 1992; Ulrich *et al.*, 1998; Koletzki *et al.*, 1999; Lachmann *et al.*, 1999; Koletzki *et al.*, 2000). Moreover such insertions disrupt the immunodominant region of HBcAg, which means that they tend to reduce the antigenicity of the HBcAg carrier, particularly if some of the HBcAg amino acids in the flexible loop (positions 74-81) around the insertion site are removed. The impressive tolerance of the e1 loop for insertions allows much larger sequences, both conformational and disordered, to be displayed in this region without abolishing particle assembly (although the yield of chimeric particles is often adversely affected to a certain extent). In a series of experiments aimed at displaying different long (~25 kDa) epitopes based on the polymorphic immunodominant molecule (PIM) protein of the parasite *Theileria parva*, it was determined that the e1 loop of HBcAg could handle such long inserts without impairing particle formation, so long as a flexible glycine-rich linker of 15 or 20 amino acids was added on either end of the insert; whereas the use of a similar linker of only 9 amino acids failed to allow particle assembly (Janssens *et al.*, 2010). The authors speculate that large inserts are

likely to be tolerated in the e1 loop so long as linkers of sufficient length allow the necessary flexibility for folding, and for the termini of the insert to be relatively close together (10-18 Å apart) at the e1 loop. Another group demonstrated that ~190 amino acids corresponding to the entire sequence of the OspC protein from *Borrelia burgdoferi*, with the exception of the signal peptide, could be fused into the e1 loop of HBcAg without impairing the folding of either HBcAg or OspC (Skamel *et al.*, 2006). Moreover, the resulting particles seemed to be coated with OspC proteins that were dimerising with inserts on neighbouring e1 spikes, thus demonstrating great flexibility both within and between the e1 spikes on assembled HBcAg particles. So far, the largest sequence directly fused into the e1 loop of HBcAg is GFP (Kratz *et al.*, 1999). The 238 amino acid –long, 27 kDa protein sequence was inserted in the e1 loop (with amino acids P79 and A80 removed) with glycine-rich linkers on either end (9 amino acids long). This construct yielded fluorescent fusion protein, some of which assembled into fluorescent particles. Cryo-electron microscopy revealed that these particles looked very much like native HBcAg capsids, but with an extra dense shell on the surface, corresponding to the position of the GFP molecules. In some cases, insertions can cause problems related to solubility of chimeric capsids, but these are not necessarily unsurmountable. An 11 kDa insertion from Dengue Virus EDIII yielded insoluble particles which had to be denatured to be purified from *E. coli*, but these were capable of being refolded into capsids displaying the insert (Arora *et al.*, 2012). This inherent insolubility was not resolved by switching to a yeast expression system (Arora *et al.*, 2013).

It is possible to use more than one insertion site on HBcAg at the same time without abolishing capsid formation. Ulrich *et al.* (1999) inserted 45 amino acids of hantavirus nucleocapsid protein in the e1 loop, and a different 44 amino acid –long epitope (from the same protein) at the C-terminus, which was truncated after HBcAg position 144. Later, Birkett *et al.* (2002) successfully inserted 20 amino acids from *Plasmodium falciparum* circumsporozoite protein in

the e1 loop, and a different 20 amino acid – long epitope at the C-terminus of a Δ149 truncation of HBcAg.

At the time of writing, there have been only two reports of antigen display on HBcAg produced in plants. Bandurska *et al.* (2008) fused the 139 amino acid – long domain-4 epitope of the anthrax protective antigen (PA) in the e1 loop and created lines of transgenic tobacco containing the sequence. While the fusion protein was expressed and was immunogenic, CLP formation was not observed. This contrasts with the findings of Huang *et al.* (2005), who displayed a 21 amino acid – long epitope from FMDV in the e1 loop of HBcAg, also produced in transgenic tobacco. This group found that CLPs were formed, and were capable of protecting mice from FMDV challenge.

D) Other insertion strategies

Numerous groups have developed alternatives to straightforward genetic fusions in order to develop easier antigen presentation strategies, or to overcome the limits to insertions described above. The most common strategy used to force HBcAg capsids to carry large inserts at the C-terminus is to produce mosaics: particles containing a mix of modified and unmodified HBcAg monomers. Koletzki *et al.* (1999) used this strategy to fuse 120 amino acids of Hantavirus protein at the C-terminus of Δ144 HBcAg: a stop codon read-through mechanism was exploited to produce modified and unmodified HBcAg from the same construct. This allowed particle formation with this insert, whereas a straightforward C-terminal fusion did not. This was seen again by Ulrich *et al.* (1999), who produced mosaics particles with HBcAg Δ144 containing 144 extra amino acids at the C-terminus. This strategy was studied in more detail by Kazaks *et al.* (2002), who found that it permitted the formation of particles with 94, 114, or 213, but not 433 amino acids fused to the C-terminus of Δ144 HBcAg, even when direct fusions failed to allow the formation of such particles (or gave prohibitively low yields). The

authors conclude that the mosaic strategy can be of use to incorporate "problematic" (i.e. long or hydrophobic) sequences into HBcAg capsids at the C-terminus. The mosaic technique clearly provides another tool for the insertion of peptides into the C-terminus of HBcAg, but it should be noted that the drawback of this strategy is that particles could contain very few copies of the insert, and some particles may not contain any insert at all.

An alternative strategy for display in the e1 loop is chemical conjugation. It is possible to genetically modify HBcAg by inserting a lysine residue in the e1 loop and removing all of the cysteines from the sequence. A peptide insert with a cysteine on one terminus can then be chemically conjugated to the e1 lysine via its cysteine (Jegerlehner *et al.*, 2002). This allows display of peptides to the e1 loop, but it should be noted that the efficiency of conjugation decreases as the size of the insert increases. Indeed, a small peptide like the FLAG tag was coupled with about 50% efficiency, M2e (23 amino acids) coupling was about 40% efficient, and coupling of a 66 amino acid- long epitope from *Toxoplasma gondii* protein GRA2 was only about 30% efficient. The coupling of a 134 amino acid-long PLA2 glycoprotein from bee venom was estimated to be even lower. Chemical coupling may therefore be an option if genetic fusion is unsuccessful, or if the insert must present a free terminus.

Another strategy is non-covalent interaction of the insert and the HBcAg carrier through a peptide tag that binds to the e1 loop. The natural interaction between HBcAg and HBsAg (Dyson and Murray, 1995) was exploited to design an oligopeptide (GSLLGRMKGA) which binds the e1 loop of HBcAg (Blokhina *et al.*, 2013). This peptide tag, when fused to M2e, allowed display of the epitope on the HBcAg particles. The strength and stability of the interaction deserves further study however, as sucrose gradient ultracentrifugation was seen to disrupt the interaction between the particles and the tagged epitope.

Yet another strategy is *in situ* post-translational cleavage of the insert in the e1 loop (Walker *et al.*, 2008). In this case, the insert is designed to contain the cleavage site for tobacco etch virus

(TEV) protease at the N-terminus. The modified HBcAg construct is then expressed with the protease, which cleaves the insert at its N-terminus without preventing capsid assembly. This was shown to allow particle assembly when a straightforward fusion would not, such as in the case of the *Borrelia burgdoferi* protein OspA, a 28 kDa protein, in which the N- and C-termini are very far apart. This strategy was simplified to give SplitCore: a strategy whereby the HBcAg sequence is split at the e1 loop into two separate open reading frames, placed together as a bicistronic operon in an *E. coli* expression system (Walker *et al.*, 2011). This allows large inserts to be fused either at the C-terminus of the upstream (N-terminal) half-monomer, or at the N-terminus of the downstream (C-terminal) half-monomer. This strategy is discussed in more detail in Chapter 3.

An interesting alternative to the above strategy is the tandem core technology, patented by Gehin *et al.* (2004) after development in *E. coli*, and tested in plants using the CPMV-*HT* system (Thuenemann, 2010). This involves genetically fusing two HBcAg monomers in order to produce the HBcAg dimer as a single polypeptide chain. This allows large insertions to be made in only one of the e1 loops in each dimer, thus reducing the risk of steric hindrance with large inserts. This technique has been shown to allow the formation of capsid particles with large inserts in the e1 loop in both bacteria and plants. Two main types of tandem core constructs were produced and tested: the homotandem core construct (CoHo) is composed of two core sequences both truncated at position 149, whereas the heterotandem core construct (CoHe) has a truncated N-terminal core but a full-length C-terminal core (more details are given in Chapter 3). While both expressed and formed particles in both plants and *E. coli*, it was found that CoHe yielded more homogenous CLPs and was more suited to antigen display in plants (Thuenemann, 2010). Of particular interest for this thesis is the fact that GFP was successfully displayed on the surface of CoHe particles in plants using the pEAQ-*HT* vector. This is the main technique which was used in this thesis to display large correctly-folded inserts on the surface of HBcAg CLPs.

E) Other antigen carriers and the advantages of HBcAg

While it is beyond the scope of this introduction to comprehensively review the use of nanoparticles for antigen display, the main differences between HBcAg and other commonly used carriers can be addressed. As mentioned previously, other viruses have been used for the surface display of epitopes. Chimaeric TMV (Sugiyama *et al.*, 1995; Turpen *et al.*, 1995) and CPMV (Usha *et al.*, 1993; Porta *et al.*, 1994; McLain *et al.*, 1995; Dalsgaard *et al.*, 1997) have both been used, but only with small epitopes (<100 amino acids). PVX, on the other hand, has proven to be much more tolerant of larger inserts, provided that the 2A peptide is used to fuse the insert protein to the N-terminus of the PVX coat protein. This ensures that a certain number of coat proteins are insert-free, which allows the assembly of chimaeric PVX rods despite potential steric hindrance caused by large protein inserts. In that sense, this 2A strategy is essentially very similar to a mosaic strategy. By using this method, whole proteins such as GFP (Cruz *et al.*, 1996), scFV (Smolenska *et al.*, 1998), and the rotavirus VP6 capsid protein (O'Brien *et al.*, 2000) have all been displayed on the surface of PVX rods. Small epitopes can be fused to the PVX coat protein directly without the need for 2A, as was demonstrated by Marusic *et al.* (2001), who displayed a 6 amino acid – long epitope from the conserved region of HIV surface glycoprotein gp41 on the surface of chimaeric PVX rods. HBcAg can accommodate larger inserts than CPMV, and unlike PVX does not necessarily require a mosaic strategy to display large folded proteins.

As mentioned above, animal viruses have also been used to display proteins of interest. Murine polyomavirus VLPs have been used to display a *Streptococcus* peptide epitope (Middelberg *et al.*, 2011) as well as the entire 34 kDa human prostate specific antigen (Eriksson *et al.*, 2011). VLPs from the related hamster polyomavirus has also been used to display a 9 amino acid - long model epitope from lymphocytic choriomeningitis virus (Mazeike *et al.*, 2012), as well as an entire scFv-Fc fusion which retained antigen-binding activity when

displayed on the surface of the VLPs (Pleckaityte *et al.*, 2011). While this is reminiscent of HBcAg in terms of the size and complexity of inserts, polyomavirus VLPs require the co-expression of at least two capsid proteins (in the latter example, the insert was fused to VP2 and co-expressed with VP1). Protein display on HBcAg does not require the co-expression of other structural proteins, which is an advantage as the stoichiometry of different capsid proteins can sometimes be a barrier to efficient VLP production (Thuenemann *et al.*, 2013a).

Other viruses have been used to display short epitopes, including bovine papillomavirus (Chackerian *et al.*, 1999), papaya mosaic virus (Denis *et al.*, 2008), and the AP205 bacteriophage (Pastori *et al.*, 2012), but the display of full-length proteins is not normally possible in such systems. As noted above, many different viruses have been used to display scFv, but most of those viruses are not typically used for antigen display *stricto sensu*. This underlines the fact that the display of a protein on the surface of a viral particle depends not only on the viral particle, but also on the nature of the protein being used as an insert. Non-viral nanoparticles have also been used: apoferritin is a ubiquitous protein complex which forms symmetrical nanoparticles reminiscent of VLPs. Apoferritin has been used to display short epitopes (Han *et al.*, 2014) as well as the 65 kDa ectodomain of influenza virus HA, which formed physiologically and immunologically relevant trimers on the surface of the apoferritin nanoparticle (Kanekiyo *et al.*, 2013).

Ultimately, HBcAg is not the only candidate for the display of heterologous proteins, and in some cases it may not be the best candidate. But it does generally have advantages over other carriers. A single protein forms the stable capsid, and there are three possible insertion sites within HBcAg including the immunodominant e1 loop which tends to boost the antigenicity of the insert. This e1 loop can accommodate large inserts, although GFP appears to be the limit in terms of size and complexity that the e1 loop can handle in the context of native HBcAg. The

use of different bionanoparticles for antigen display has been reviewed by Lee and Wang (2006).

VI - Aims of this thesis

The primary aim of this thesis is to develop plant-produced HBcAg CLPs as carriers for any antigen of interest. Specifically, the goal is to produce a generic antigen presentation system, which would easily allow any antigen to be presented on the surface of an HBcAg CLP. In order to do this, both the SplitCore and tandem core technologies mentioned above were tested to determine which might be most suitable for chimaeric CLP production in plants. Once it was determined that tandem core technology was better suited, two different generic antigen display systems were developed: the first, very simple in design, is intended to allow antigen display on a CLP via post-purification chemical conjugation. The second system, which is more complex in nature and formed the bulk of the project, investigated the possibility of displaying an antibody fragment (scFv or camelid nanobody) on the surface of the CLPs. This would allow the non-covalent binding of an antigen of interest to the chimaeric CLP via the antibody fragment. The hypothesis on which this system is founded is that all antibody fragments of a particular type (scFv or camelid nanobody) will behave in the same way when fused onto the surface of the tandem core particle. Thus, if a system could be developed whereby an antigen could be displayed on the surface of a CLP through the interaction with an antibody fragment expressed in HBcAg, it should be straightforward to switch the antibody moiety to create a CLP that will capture and display a completely different antigen.

Chapter 2 : Materials and Methods

I - DNA constructs

Numerous DNA constructs were ordered for synthesis from GeneArt (Life Technologies) in order to quickly obtain reliable, high-quality material for subsequent cloning and gene expression. Details of the specific sequences synthesised are given in the relevant chapter.

II - DNA isolation/purification

DNA was isolated from bacteria by using the QIAprep Spin MiniPrep Kit (Qiagen), from PCR reactions by using the QIAquick PCR purification kit (Qiagen), and from agarsose gels by using the QIAquick Gel Extraction kit (Qiagen). Purified DNA was stored at -20°C.

III - Restriction digests

Restriction enzymes were obtained from New England Biolabs (NEB), Roche or Invitrogen. Restriction digests were carried out in the appropriate buffer as recommended by NEB.

IV - Ligation

Ligation of DNA fragments was carried out using either Quick Ligase (NEB) or T4 Ligase (NEB) as per manufacturer's instructions. When ligations were of an insert into a plasmid, the individual fragments were quantified using a NanoDrop spectrophotometer and a 3:1 insert:plasmid molar ratio was used.

V - Media

Solution	Composition
ELISA BUFFERS	
ELISA Block Buffer	5 % (w/v) dry milk, 0.05 % (v/v) Tween-20, in PBS
ELISA sample buffer	1 % (w/v) dry milk, 0.05 % (v/v) Tween-20, in PBS
Wash buffer (ELISA)	0.05 % (v/v) Tween-20, in PBS
ELISA detection reagent	50 mM $NaHPO_4$; 25 mM citric acid; 3,3',5,5'– Tetramethylbenzidine (TMB) tablet (Sigma-Aldrich)
ELISA stop solution	1 M H_2SO_4
EXTRACTION BUFFERS	
AmBic	20 mM ammonium bicarbonate, pH 8.5
HBexB1	10 mM Tris-HCl pH 8.4; 120 mM NaCl; 1 mM EDTA; 0.75 % (w/v) sodium deoxycholate; 1 mM DTT; Complete Protease Inhibitor Cocktail tablet (Roche)
P	0.1 M sodium phosphate, pH 7.5; Complete Protease Inhibitor Cocktail tablet (Roche)
P+	20 mM sodium phosphate, pH 7.0; 1 mM EDTA; 5 mM DTT; 0.1 % (v/v) Triton X-100; Complete Protease Inhibitor Cocktail tablet (Roche)

IQUR	20 mM Tris-HCl pH 8.0; 1 mM EDTA; 5% isopropanol (v/v); 5 mM DTT; 0.1 % Triton X-100; Complete Protease Inhibitor Cocktail tablet (Roche)
AngeliTris	100 mM Tris pH 9.0; 100 mM NaCl; 1 mM EDTA; 0.75 % (w/v) sodium deoxycholate; Complete Protease Inhibitor Cocktail tablet (Roche)
AngeliBor	100 mM boric acid pH 10; 100 mM NaCl; 1 mM EDTA; 0.75 % (w/v) sodium deoxycholate; Complete Protease Inhibitor Cocktail tablet (Roche)

GROWTH MEDIA

Luria Bertani broth (LB)	10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl adjusted to pH 7.0
L-agar	as LB, with 1.5 % (w/v) Lab M agar added
SOC (Life technologies)	2 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 10 mM MgSO4, and 20 mM glucose.

WESTERN BLOT BUFFERS

Western Blot Transfer buffer	3.03 g/l Tris-HCl, 14.4 g/l glycine, 20 % (v/v) methanol
Western blot block buffer	5 % (w/v) dry milk, 0.1 % (v/v) Tween-20, in PBS

Wash buffer (western)	0.1 % (v/v) Tween-20, in PBS
Western blot detection solution	1.25 mM luminol; 200 mM paracoumaric acid; 27 mM H_2O_2; 100 mM Tris-HCl pH 8.5

OTHER

MMA	10 mM MES, pH 5.6; 10 mM $MgCl_2$; 100 µM acetosyringone
Phosphate-buffered saline (PBS)	150 mM NaCl; 2 mM NaH_2PO_4; 15 mM Na_2HPO_4; pH 7.2
Protein loading buffer (2X)	900 µl 4x LDS sample buffer (Invitrogen), 300 µl β-mercaptoethanol
TBE	10.8 g/l Tris-HCl, 5.5 g/l Boric, 2 mM EDTA

VI - Polymerase Chain Reaction (PCR)

Amplification of DNA to be used for subsequent cloning was carried out using Phusion high-fidelity polymerase (Roche) and the supplied buffer using the following protocol. Details of the primers used can be found in Appendix 1.

Recipe	50	µl	Cycles			
5X HF buffer	10	µl	98°C	30	sec	
10 mM dNTPs	1.25	µl	98°C	10	sec	
DNA template	1	µl	70°C	20	sec	30 X

10 µM forward primer	2.5	µl		72°C	20	sec/kb
10 µM reverse primer	2.5	µl		72°C	5	min
Phusion polymerase	0.5	µl				
dH$_2$O	32.25	µl				

Amplification of DNA to be used only for agarose gel electrophoresis analysis only was carried out using GoTaq Green Master Mix (Promega) using the following protocol. For colony PCR (always carried out using the C1 and C3 primers), the DNA template was replaced with a small amount of bacteria picked off an agar plate with a sterile toothpick.

Recipe	**20**	**µl**		**Cycles**		
2X Go Taq Green MM	10	µl		95°C	5	min
DNA template	1	µl		95°C	1	min
10 µM forward primer	0.5	µl		58°C	45	sec
10 µM reverse primer	0.5	µl		72°C	1	min/kb
dH$_2$O	8	µl		72°C	5	min

31 X (cycles from 95°C 1 min through 58°C 45 sec)

VII - Agarose gel electrophoresis

Agarose gel electrophoresis was used to analyse PCR reactions, or to analyse and purify restriction fragments. Gels contained 0.8- 1.2 % (w/v) agarose dissolved in TBE. The marker ladder used was HyperLadder I (Bioline).

VIII - DNA sequencing

Sequencing reactions were either carried out entirely by Eurofins Genomics; or prepared as "ready reactions" using BigDye Terminator v3.1 (Applied Biosystems) according to manufacturer's instructions, the sequences were then analysed by either Eurofins Genomics (i54 Business Park, Valiant Way, Wolverhampton, WV9 5GB, UK) or Genome Enterprise Ltd. (The Genome Analysis Centre, Norwich Research Park, Norwich, NR4 7UH, UK).

IX - Bacterial strains, growth and storage conditions

Escherichia coli strain TOP10 (Life Technologies) and *Agrobacterium tumefaciens* strain LBA4404 (Hoekema *et al.*, 1983) were used throughout. *E. coli* containing a pMA plasmid with the appropriate insert obtained from GeneArt (Life Technologies) were grown at 37 °C using LB broth or L-agar supplemented with carbenicillin (100 µg/ml) for plasmid selection. *E. coli* containing a pEAQ plasmid were grown using the same conditions, but 50 µg/ml kanamycin was used for plasmid selection instead of carbenicillin. *A. tumefaciens* containing a pEAQ plasmid were grown at 28 °C using LB broth or L-agar supplemented with kanamycin (50 µg/ml) for plasmid selection plus 50 µg/ml rifampicin to select for *A. tumefaciens* LBA4404. Once a clone of *E. coli* or *A. tumefaciens* had been identified as positive for the desired plasmid by colony PCR and sequence analysis, a glycerol stock was prepared: the clone was grown in liquid culture and an aliquot of stationary phase culture was supplemented with glycerol (final concentration 25 % [v/v]), snap-frozen in liquid nitrogen, then stored at -80°C.

X - Transformation of competent bacterial cells

Purchased *E. coli* TOP10 chemically competent cells were transformed as per manufacturer's instructions. Briefly, aliquots of cells mixed with 10-50 ng of plasmid DNA were heat-shocked at 41°C for 30 seconds then allowed to recover in SOC for one hour at 37°C before being plated on agar plates containing the appropriate antibiotic. *A. tumefaciens* LBA4404 electrocompetent cells were made in the laboratory. These were transformed by mixing 50 µl of competent cell aliquots with 10-50 ng of plasmid DNA, then electroporating at 2,500 V. The cells were allowed to recover in SOC for one hour at 28 °C before being plated on agar containing appropriate antibiotics.

XI - Plasmids used

The pMA plasmids are small simple plasmids used by GeneArt (Life Technologies) to facilitate the propagation and transportation of synthetic gene constructs. They carry a resistance gene to ampicillin, and were selected for in bacterial cultures with carbenicillin. The pEAQ-*HT* plasmid is a binary vector tailored to allow high-yielding overexpression of recombinant proteins in plants (Sainsbury *et al.*, 2009b). It contains a CPMV-*HT* cassette composed of a 35S promoter, Nos terminator, and Cowpea Mosaic Virus RNA-2 5' and 3' untranslated regions (UTRs), with two point mutations in the 5' UTR which increases translational efficiency. This is the vector which was used to express the proteins discussed in this thesis. The pFSC5 and pFSC6 plasmids are small plasmids used to insert two genes of interest (one in each plasmid) each within its own CPMV-*HT* cassette: both contain a CPMV-*HT* cassette with a multiple cloning site for ORF insertion, and the CPMV-*HT* cassette is flanked by the restriction sites PacI and SbfI for pFSC5, and SbfI and AscI for pFSC6 (Sainsbury *et al.*, 2012a). These restriction sites can be used to integrate both cassettes into a single pEAQ-*HT* vector using the PacI and AscI

restriction sites, with the shared SbfI site allowing the two cassettes to form a single DNA fragment.

XII - Agroinfiltration

For transient expression of recombinant proteins, liquid cultures of *A. tumefaciens* were prepared. For bacteria taken from glycerol stocks, the primary culture was subcultured, and this subculture was used for agroinfiltration. Agroinfiltration was carried out using a similar method to that described in (Thuenemann *et al.*, 2013a). Briefly, the cultures were centrifuged at 1,100 x g for 20 min, then resuspended in MMA buffer to an OD_{600} of 0.4 unless otherwise stated. A syringe was used to infiltrate the suspension into the extracellular space of *N. benthamiana* leaves (Schob *et al.*, 1997). Typically, for small-scale experiments designed to test expression of a recombinant protein, two leaves from each of three plants were infiltrated in order to minimise variability of expression levels.

XIII - Plant growth conditions

N. benthamiana plants were grown in glasshouses maintained at 25°C and watered daily. Supplemental lighting was provided to maintain 16 hours of daylight in the winter months (October – April, inclusive). Plants were agroinfiltrated 3 weeks after pricking out.

XIV - Small-scale protein extraction

Small-scale extractions were used to test expression and accumulation of recombinant proteins. Agroinfiltrated leaf tissue was harvested 7 days post-infiltration (dpi), and a cork-

borer (size number 7) was used to sample six leaf discs (corresponding to 90 mg of fresh-weight leaf tissue, or FWT) from infiltrated areas of the leaves. These were placed in a 2 ml screw-cap microcentrifuge tube with a ¼-inch ceramic bead (MP Biomedicals) and 270 µl of the appropriate extraction buffer (buffer P, unless otherwise stated). The leaf tissue was then homogenised using a FastPrep homogeniser (MP Biomedicals) at speed setting 5.0 for 40 seconds, or on an Omni Bead Ruptor 24 homogenizer (Camlab) at speed setting 4 for 30 seconds. The samples were then centrifuged at 16,000 g for 10 minutes, and the supernatant (soluble protein extract) was separated from the pellet (insoluble protein extract).

XV - Large-scale protein extraction

Large-scale protein extractions were used to obtain recombinant protein for experiments. Agroinfiltrated leaf tissue was harvested 6 or 7 dpi, and the infiltrated areas of the leaves were excised with a razor blade. The infiltrated leaf tissue was weighed and homogenised in a Waring blender with 3 x volume of the appropriate extraction buffer (P, unless otherwise stated). The crude extracts were filtered over one layer of Miracloth (Calbiochem) and the filtrates were clarified by centrifugation at 9,000-15,000 x g for 10-15 min (depending on volume and turbidity of the filtrate) at 4°C. The clarified extracts were then filtered over a 0.45 µm syringe filter (Sartorius). After this, the extracts were either purified by sucrose cushion (for core-like particles), or by affinity chromatography (for His-tagged proteins).

XVI - Sucrose cushions

The extracts were underlain with sucrose solutions prepared in the corresponding extraction buffer (usually 2 ml 25 % [w/v], then 0.5 ml 70 % [w/v] sucrose solutions) in UltraClear 14X89

mm ultracentrifuge tunes (Beckman Coulter) and centrifuged at 273,800 x g in a TH641 ultracentrifuge swing-out rotor (Sorvall) for 2 h 30 min at 4°C. For even larger-scale, a SureSpin 630/36 ultracentrifuge swing-out rotor (Sorvall) was used: in this case, 5 ml 25 % (w/v) sucrose and 1 ml 70 % (w/v) solutions were used in UltraClear 25X89 mm ultracentrifuge tubes. These were centrifuged at 167,000 x g for 2 h 30 min at 4 °C. Fractions were then collected by piercing the bottom of the ultracentrifuge tube with a needle and collecting the material as it dripped out.

XVII - His-tag affinity chromatography

The extracts were loaded onto a HisTrap FF crude pre-packed nickel column (GE Healthcare) using a peristaltic pump (Watson-Marlow Ltd.) as per the manufacturer's instructions.

XVIII - Photography

Photography was used to document fluorescence of GFP-containing leaves and leaf extracts as visualised under ultraviolet light. Excitation of GFP was achieved using several Blak-Ray (Ultra-Violet Product Ltd) lamps, and all photographs were taken by Mr. Andrew Davis.

XIX - Polyacrylamide gel electrophoresis

SDS-PAGE was carried out using NuPAGE Bis-Tris pre-cast gels, containing either 12 % (w/v) or 4-12 % (w/v gradient) acrylamide. Electrophoresis was carried out in NuPAGE MOPS buffer at 200 V for 50 minutes. All samples were boiled in protein loading buffer for 20-30 minutes before being loaded onto the gels. The protein marker ladder used was SeeBlue Plus 2 (Life Technologies). After electrophoresis, gels were either transferred to nitrocellulose membrane

for western blotting (see below), or stained with InstantBlue (Expedeon) for at least one hour before being destained in water for at least 16 hours. When necessary, protein bands of interest from stained gels were identified by Mass Spectrometry by Dr. Gerhard Saalbach (John Innes Centre Proteomics facility).

XX - Western blotting

After electrophoresis, gels were transferred to membrane by wet transfer in western blot transfer buffer to Amersham Hybond-ECL nitrocellulose membranes (GE Healthcare), and membranes were blocked in western blot block buffer either overnight at 8°C or for one hour at room temperature. The membranes were then blotted with the appropriate antibody (see below) in block buffer for one hour at room temperature, then washed three times in wash buffer for 15 minutes, then blotted with the appropriate horseradish peroxidase – conjugated antibody secondary antibody (if applicable, see below) before washing again. After a final wash in PBS, protein was detected using the western lot detection solution, and the chemiluminescent signal was detected either on Amersham Hyperfilm ECL (GE Healthcare), or digitally on an ImageQuant LAS 500 (GE Healthcare). For protein quantification, western blots were carried out with multiple standards of known concentrations, and the ImageQuant LAS 500 (GE Healthcare) was used to quantify the protein based on signal in the bands.

XXI - Enzyme-linked immunosorbent assay (ELISA)

ELISA was used to study binding of tandibody particles to their target antigens. Tandibody particles in PBS (50 µl per well) were used to coat the wells of a F96 Maxisorp 96-well plate (Nunc) overnight at 8 °C. For positive controls, a commercially obtained antibody against the

target antigen was coated at the same concentration and volume as the tandibodies (unless otherwise stated). The wells were then blocked in blocking buffer for 2 h at 37°C. After washing 3 times in 200 µl wash buffer, the target antigen in sample buffer was added to the wells at varying concentrations obtained via a dilution series, and the plate was incubated for 2 h at 37 °C. The GFP used was N-terminal His-tagged GFP (Millipore), the gp120 was recombinant HIV-1 IIIB gp120 produced in Baculovirus (CFAR NIBSC EVA607), and the p24 was recombinant p24 from HIV-1 HxB2 produced in *Pichia pastoris* (CFAR NIBSC ARP678). After washing again, the detection antibody (specific to the target antigen and conjugated to HRP) in sample buffer was added to the wells and the plate was incubated for 2 h at 37°C. After washing, 50 µl of detection solution was added to each well and then stopped with 50 µl of stop solution. The OD_{450} of the wells was read on a POLARstar Omega plate reader (BMG Labtech), and the average signal from control wells (octuplicate wells with PBS added instead of the antigen) was subtracted from the average signal from experimental wells (quadruplicates for each treatment) in order to obtain an average signal net of background.

A) Antibodies used in western blots and ELISA

Name	Target	Species	HRP	Use	Supplier
10E11	HBcAg	mouse	no	Western blot	Abcam
Anti-His	His-tag	mouse	no	Western blot	Qiagen
Anti-mouse	Mouse IgG	goat	yes	Western blot	Life Technologies
Ab53840	gp120	goat	yes	ELISA and western blot	Abcam
Ab6673	GFP	goat	no	ELISA	Abcam
A10260	GFP	rabbit	yes	ELISA and western blot	Life Technologies
Ab53841	p24	goat	no	ELISA	Abcam
Ab20365	p24	goat	yes	ELISA and western blot	Abcam

XXII - Dialysis

The dialysis material used for low volumes of CLPs was float-a-lyzer dialysis devices with either 100 or 1,000 kDa molecular weight cut-off (MWCO) with either 2 ml or 5 ml chamber volumes. For larger volumes or lower molecular weight proteins, SnakeSkin dialysis tubing with 10 kDa MWCO was used as per manufacturer's instructions. Buffer exchanges were carried out in either 3 l or 5 l buckets, with at least one buffer change over at least 16 hours, depending on the volume of the samples.

XXIII - Protein concentration

Protein preparations were dialysed against AmBic and concentrated by vacuum evaporation using a SpeedVac (Savant).

XXIV - Size-exclusion chromatography (SEC)

CLP preparations requiring higher purity than could be obtained from sucrose cushions, and syringe filtrations, and clarification centrifugation were purified by gel-filtration size-exclusion chromatography. Samples in PBS were purified on a FPLC column (Pharmacia) containing Sephacryl S-500 medium (GE Healthcare) with the help of an ÄKTA FPLC system (Amersham Biosciences).

XXV - Nycodenz gradient

The t-KD extract, after large-scale extraction, protein concentration, and dialysis against PBS, was underlain with nycodenz AG (Axis-Shield PoC, Oslo, Norway) solutions prepared in PBS (2 ml each of 10, 20, 30, 40, 50, and 60 %. Nycodenz solutions) in UltraClear 14X89 mm ultracentrifuge tunes (Beckman Coulter) and centrifuged at 273,800 x g in a TH641 ultracentrifuge swing-out rotor (Sorvall) for 20 h at 4 °C. Fractions were then collected by piercing the bottom of the ultracentrifuge tube with a needle and collecting the material as it dripped out.

XXVI - Apoplastic wash

In order to extract proteins located in the intercellular space (apoplast) of *N. benthamiana* leaves, apoplastic washes were carried out. These consist of infiltrating the P extraction buffer into the intercellular space of agroinfiltrated leaves at 7 dpi, then wrapping those leaves in muslin cloth and trapping the package at the top of a 50 ml centrifuge tube with the lid, before centrifugation at 2, 000 x g for 5 minutes. The buffer, along with any soluble proteins in the apoplast, is forced out of the leaf tissue and is collected at the bottom of the centrifuge tube.

XXVII - Transmission electron microscopy (TEM)

TEM was used to confirm that plant tissue expressing a construct contained CLPs, as well as to observe homogeneity and morphology of CLPs. Samples (typically in AmBic) were adsorbed onto copper-palladium grids composed of a plastic-covered hexagon mesh, washed by dripping 5-8 drops of sterile distilled water over the sample on the grid, then staining with 2 % (w/v) uranyl acetate for 1 minute. Particles were then imaged using a FEI Tecnai 20 TEM with a bottom-mounted digital camera.

XXVIII - Cryo-electron microscopy (cryo-EM)

All cryo-EM was carried out at the Division of Structural Biology (University of Oxford, Oxford, UK). Dr. Robert Gilbert provided guidance and supervision for the data collection, and carried out the data analysis. Concentrated size-exclusion chromatography fractions of τ-GFP bound to GFP were dialysed against 20 mM Tris-HCl (pH 8.5) and an aliquot was placed on a holey carbon grid and vitrified by flash-freezing. Images of the particles were obtained on a Tecnai F30 microscope at 200 kV and 80,000 x magnification and a range of defocus values from -2 to -5 µm on a Falcon 2K CCD camera. After manual selection of 2,385 particles from the obtained images using EMAN software (Ludtke *et al.*, 1999), particles were corrected for contract transfer function and classified in two dimensions before being reconstructed in 3D using RELION software (Scheres, 2012). The initial 3D model was generated with IMAGIC software from seven RELION-derived class averages (van Heel *et al.*, 1996). Using this, a reconstruction at a resolution of 25 Å was created.

XXVIV - Software

- Vector NTI Advance 11.5 (Invitrogen) was used for vector design, in silico cloning strategies, sequence analyses and alignments, as well as predictions of protein sequences, sizes and attributes.

- The BLAST function of the National Center for Biotechnology Information website (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was used both for nucleotide and protein sequence alignments.

- The GenBank database of the National Center for Biotechnology Information website (http://www.ncbi.nlm.nih.gov/genbank/) was used to find sequences of interest.

- Predicted crystallographic structures were constructed by Dr. Ellis O'Neill using Swiss-Prot (ExPASy) and imaged using the PyMOL software (DeLano Scientific).

- Microsoft Office software suite was used for analysing data and making figures.

Chapter 3 : Evaluation of tandem core and SplitCore technologies for the expression of chimaeric HBcAg particles in plants

I - Introduction

As discussed in chapter 1, there is a limit to the size of sequence that can be displayed in the e1 loop of HBcAg: the longer the epitope, the lower the yield of chimeric CLP is likely to be, and whole folded proteins are difficult to display with success. GFP was successfully displayed on HBc (Kratz *et al.*, 1999), but this was an ideal candidate: GFP forms a tight easily-folding barrel structure with N- and C-termini close together, so inserting this sequence in the e1 loop flanked by long (9 amino acid) flexible linkers on either side allowed the particles to form with GFP displayed on the outside. The reported yield of this chimaeric CLP was between 10 and 20 mg per litre of *E. coli* culture, and the authors do not specify the yield that is normally obtained for unmodified HBcAg produced in *E. coli*. For slightly more complicated inserts (either larger proteins or proteins with distant N- and C- termini), insertion in the e1 loop becomes more complicated, if not impossible.

To overcome the size limit of the proteins which can be displayed in the e1 loop, two methods stand out. The first, called SplitCore, involves splitting the HBcAg sequence into two separate open reading frames (ORFs) by inserting a stop codon in the e1 loop after P79, and a start codon before A80 (Walker *et al.*, 2011)). This technique proved successful in a bacterial expression system in which both half-cores (called CoreN and CoreC) were produced from a bicistronic operon with a ribosomal binding site (RBS) between the two ORFs. This allowed whole proteins (GFP, *Borrelia burgdorferi* protein OspA, and *Plasmodium falciparum* circumsporozoite protein CSP) to be displayed on the surface of HBc by fusing them either to the C-terminus of CoreN, or the N-terminus of CoreC, which reduces steric constraints on the

insert because in either case, one of the termini has free movement, not being fused to the rest of the HBcAg protein (Fig. 3.1).
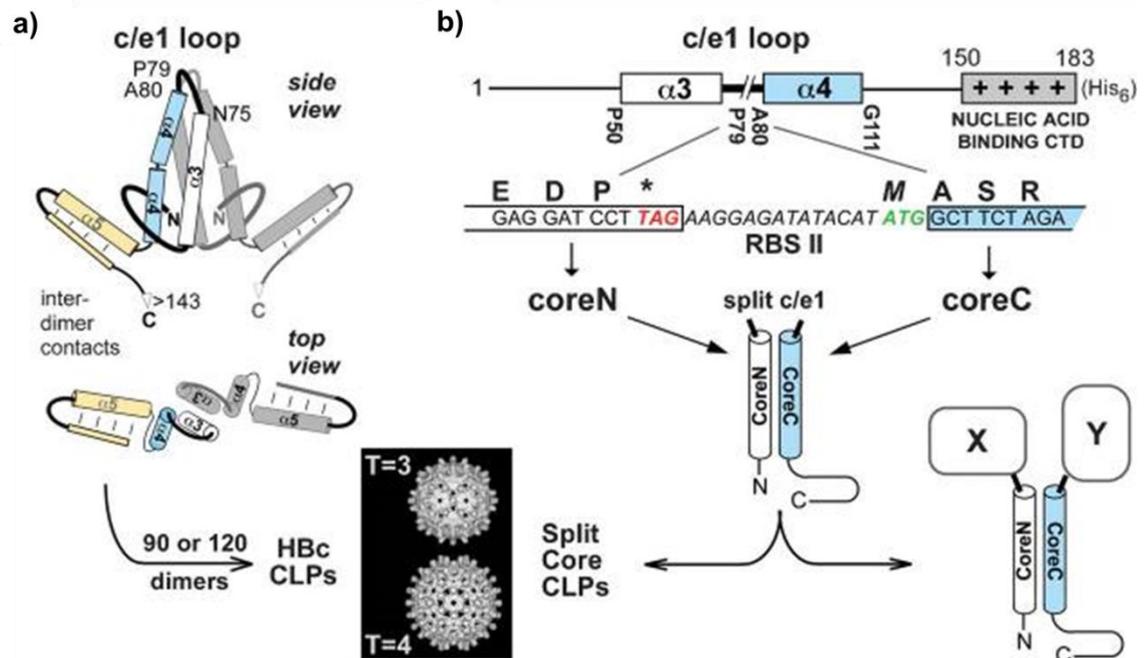
Figure 3.1 The SplitCore system as used in *E. coli*. The native HBcAg sequence (a) is split at the tip of the e1 loop by adding a stop codon after position P79, followed by a ribosome binding site (RBS), and a start codon before position A80 (b). This allows the upstream CoreN and the downstream CoreC to be produced as two separate polypeptide chains but as part of the same bicistronic operon. Sequences of interest can be inserted either at the C-terminus of CoreN or at the N-terminus of CoreC without preventing both half-cores from forming a complete core. This can then go on to form complete capsids of T=3 or T=4 conformation. Figure reprinted from Walker *et al.*, 2011.

The other technique to overcome limits to e1 insertions is called tandem core (Gehin *et al.*, 2004), and this involved fusing two HBcAg sequences together in order to produce the HBcAg dimer as a single polypeptide chain. This allows only one of the e1 loops in a dimer to be modified with an insert, thus eliminating steric hindrance of protein inserts on neighbouring e1 loops on the same dimer. This is thought to aid dimer formation which is critical for particle assembly. While the SplitCore system has only been tested in bacteria, the tandem core system has been successfully used in both bacteria and plants (Thuenemann, 2010). In this work, two different types of tandem core were studied. The first was the homo-tandem core CoHo, which is composed of a truncated N-terminal core (at amino acid position 149) fused to another similarly truncated C-terminal core. The second was the hetero-tandem core CoHe,

where the N-terminal core is truncated but the C-terminal core contains the full-length arginine-rich C-terminus (Fig. 3.2).

It was demonstrated that both types of tandem core constructs could be produced in plants using the CPMV-*HT* expression system, although CoHe gave better results in terms of CLP quality and homogeneity. Moreover, it was found that CoHe could be modified by inserting the sequence for GFP in the e1 loop of the C-terminal core, thus leading to the formation *in planta* of tandem core CLPs displaying GFP on the surface. This could be achieved with either very short (2 aa) or long (9 aa) linkers, whereas monomeric HBcAg is known to require long linkers to display GFP (Kratz *et al.*, 1999). However, it must be acknowledged that plant-produced CoHe with GFP displayed on the surface (CoHe-GFP) was largely insoluble, and a special buffer containing sodium deoxycholate (HBexB1, see Chapter 2) had to be used to obtain enough soluble CLPs for analysis to take place (Thuenemann, 2010). Moreover, the CoHe construct was in fact optimised for use in *E. coli* with bacterial expression vectors, because it was developed for use in a bacterial expression system. This meant that sequence manipulation and cloning of CoHe-based constructs was suboptimal in the pEAQ-*HT* vector used for expression in plants. The pre-existing CoHe-based vectors therefore needed to be modified in order to create comparable SplitCore and tandem core constructs, both designed to display GFP on the surface of assembled CLPs.
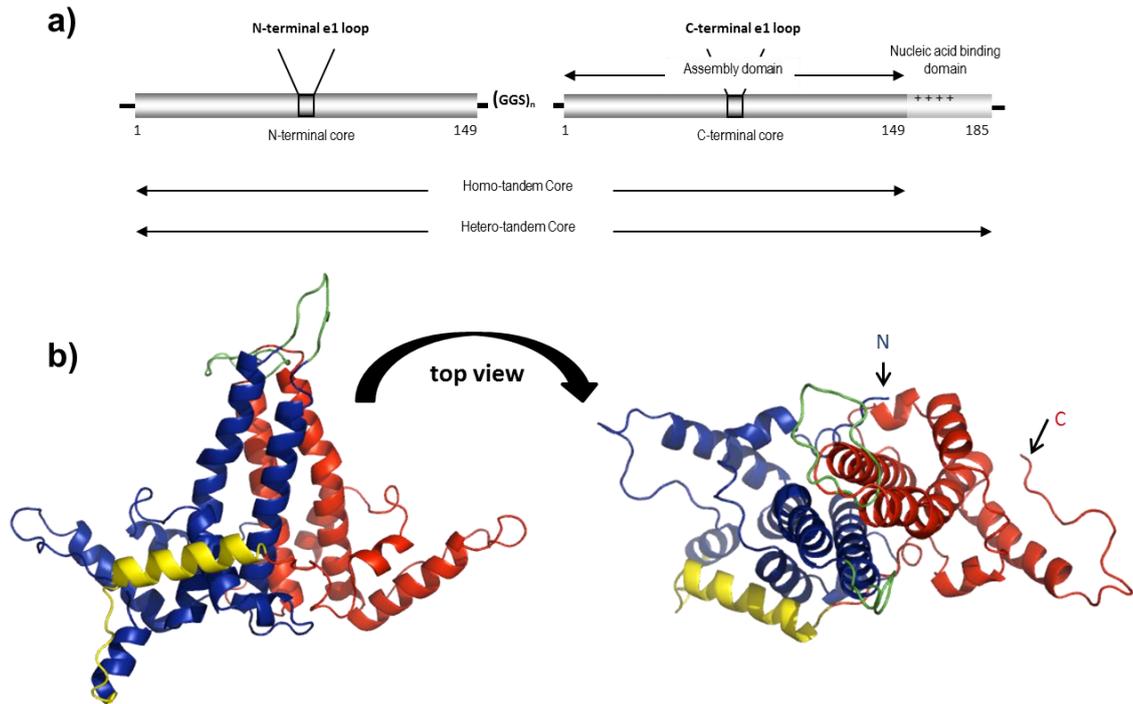
Figure 3.2 The tandem core system as used in *E. coli* and *N. benthamiana*. a) Two HBcAg sequences are fused end-to-end via a flexible glycine-serine linker (GGS) to form a tandem core. The upstream N-terminal core is truncated at position 149, whereas the downstream C-terminal core is either full-length (hetero-tandem core) or similarly truncated (homo-tandem core). The truncated C-terminus corresponds to the arginine-rich nucleic acid binding domain. b) Predicted structure of the hetero-tandem core protein (t-EL, see figure 3.3) used in this thesis: two HBcAg monomers (blue and red) are fused by a linker (yellow) to form a single polypeptide chain, with multiple cloning sites inserted in both e1 loops (green) to facilitate insertions by genetic fusion. The amino and carboxy termini are indicated in the top view image. The structure prediction is courtesy of Dr. Ellis O'Neil, and the image was generated using PyMOL.

II - Design of constructs

The CoHe construct was redesigned with the goal of simplifying its use in pEAQ-*HT*. Simultaneously, constructs were designed to evaluate the SplitCore strategy in plants. The goal was to create comparable systems to determine which was best suited for the production of an HBcAg-derived antigen display system in plants. In order to do this, a modified version of CoHe (named t-EL, for tandem HBcAg Empty Loop) was designed which contained unique restriction sites and polylinkers in strategic regions to allow easy cloning and modifications of both e1 loops in pEAQ-*HT* (construct diagrams are provided in figure 3.3). This basic sequence was the basis for a SplitCore construct aimed at testing the SplitCore system in plants. This

construct, named sCore-EL (for Split Core Empty Loop), is analogous to that described in Walker *et al.* (2011), except that it is meant to be used as two different gene cassettes instead of a bicistronic operon. GFP-carrying versions of both t-EL and sCore-EL were also created which contained a solubility-optimised GFP sequence in the e1 loop. For the tandem core construct, the sequence is in the C-terminal e1 loop (and this construct is named t-sGFP). For the SplitCore construct, the GFP sequence is split to create a split GFP system similar to that used in Walker *et al.* (2011), in which the sequence for GFP is split between the two half cores so that fluorescence only occurs if the two parts of GFP come together; this only occurs if the two halves come together. This construct, named sCore-sGFP, was intended to provide evidence for particle assembly simply by analysing fluorescence of the infiltrated leaves. These constructs were designed so that future work could be based on either tandem core or SplitCore, depending on which gave the best results in the plant expression system.

In addition, a monomeric version of HBcAg was designed that could be readily compared to the t-EL construct. This construct, named m-EL (for monomeric HBcAg Empty Loop), is identical in sequence to the C-terminal core of t-EL. This allows comparison between monomeric and tandem constructs modified for use in pEAQ-*HT*, as well as comparisons with the original tandem and monomeric constructs used in previous work (Thuenemann, 2010). Because basic sequences were ordered for synthesis from GeneArt (Life Technologies), care was taken to order sequences from which other constructs could be easily derived. This means that the initial sequence which was ordered and from which others are derived is a tandem core construct containing the sequence of an anti-phOx single chain variable fragment (scFv) in the C-terminal e1 loop (t-α-phOx, described in Chapter 4). Details are given in the following section, and construct diagrams, along with predicted protein structures of selected constructs, are shown in figure 3.3.
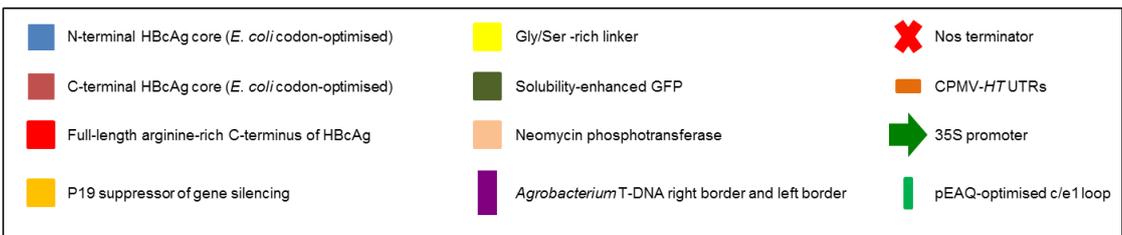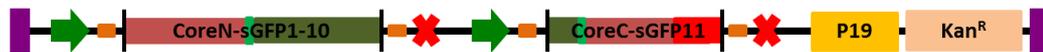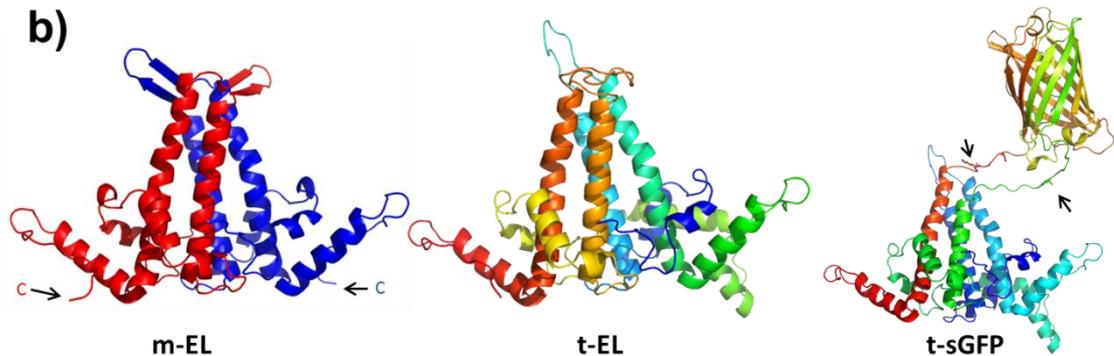
Figure 3.3 Diagrams of constructs and protein structures. a) Maps of the constructs t-EL (empty loop tandem core), t-sGFP (tandem core with sGFP sequence in the C-terminal e1 loop), m-EL (a monomeric HBcAg construct based on the C-terminal core of t-EL); and SplitCore-EL (composed of CoreN and Core-C), and sCore-sGFP (composed of CoreN-sGFP1-10 and CoreC-sGFP11). All SplitCore constructs are based on t-EL with and without split sGFP). b) Structure predictions of an m-EL dimer (left, C-termini are labelled), t-EL (middle, blue represents the N-terminus and red the C-terminus), and t-sGFP (right, blue represents the N-terminus and red the C-terminus. The first and last amino acids of sGFP are represented as sticks and indicated with arrows). Structure predictions are courtesy of Dr. Ellis O'Neil, images were generated using PyMOL.

A) Cloning

i) t-EL sequence

The features of the t-α-phOx construct described in Chapter 4 allowed the creation of different constructs from this sequence: t-EL (tandem core, empty loops) was made by digesting pEAQ-*HT*- t-α-phOx with AvrII and religating.

ii) t-sGFP sequence

The sequence for sGFP was based on eGFP used by Thuenemann (2010), with point mutations S30R, F99S, E111V, I128T, Y145F, M153T, V163A, K166T, I167V, I171V, S205T, and A206V to improve solubility and folding as described by (Cabantous *et al.*, 2005; Pedelacq *et al.*, 2006). This sequence was ordered from GeneArt (Life Technologies) without a stop codon and with 5' SalI and 3' AseI restriction sites. This construct was cloned into pEAQ-*HT*-t-EL using the SalI and AseI restriction.

iii) m-EL sequence

The m-EL construct (monomeric core with empty e1 loop) is based on a monomeric HBcAg construct containing the anti-phOx scFv sequence in e1 loop (see Chapter 4). The m-EL construct was made by digesting pEAQ-*HT*-m-α-phOx with AvrII and religating.

iv) sCore-sGFP sequences

The concept of split cores was obtained from Walker *et al.* (2011). The authors demonstrated particle formation and insert display with a split GFP experiment: The N-terminal 10 sheets of GFP (named sGFP1-10) were fused to the C-terminus of CoreN (to give CoreN-sGFP1-10), and the 11th sheet (named sGFP11) was fused to the N-terminus of CoreC (to give CoreC-sGFP11). This construct was made for use in plants using CoreN and CoreC sequences based on t-EL, and the GFP sequence was sGFP described above. The HBcAg amino acid E77 was replaced with the polylinker GG-XmaI-G-AvrII-GSGGGG to link sGFP1-10 to CoreN, and sGFP11 was linked to CoreC using the polylinker GGGGGGSGG-AvrII-GIN. An AvrII restriction site was added at the

end of the sGFP1-10 sequence immediately before the stop codon, and another was added at the start of the sGFP11 sequence immediately after the start codon. CoreN-sGFP1-10 and CoreC-sGFP11 were cloned into entry plasmids pFSC5 and pFSC6 respectively. PacI and SbfI were then used to digest pFSC5 while AscI and SbfI were used to digest pFSC6, and PacI and AscI were used to digest pEAQ-*HT*. A triple ligation then allowed the insertion of the two constructs, each now in a separate CPMV-*HT* cassette, into pEAQ, to give pEAQ-*HT*-sCore-sGFP. This single plasmid contains both genes coding for CoreN-sGFP1-10 and CoreC-sGFP11 on the same T-DNA.

v) sCore-EL

The pFSC5 and pFSC6 plasmids containing CoreN-sGFP1-10 and CoreC-sGFP11 respectively were digested with AvrII in order to remove the split GFP components. The resulting constructs, with no insert in the e1 loop, were then cloned into pEAQ-*HT* as described above to give pEAQ-*HT*-sCore-EL.

III - Comparing tandem core and SplitCore constructs

The tandem core constructs t-EL and t-sGFP, the SplitCore constructs sCore-EL and sCore-sGFP, and the monomeric m-EL control were expressed transiently in *Nicotiana benthamiana* and fluorescence of GFP-bearing constructs was inspected under ultraviolet light (Fig. 3.4). Firstly, the results show that t-sGFP fluoresces at least as much as CoHe-GFP. This demonstrates that the modifications made to the GFP sequence did not abolish fluorescence, and these modifications, along with those made to the tandem HBcAg sequence, clearly did not abolish expression. Secondly, it is clear that t-sGFP is more intensely fluorescent than sCore-sGFP. This suggests that either there is more protein accumulated in t-sGFP – expressing leaves, or that sCore-sGFP is inefficient at particle formation. It should be noted that while fluorescence from

the tandem constructs is associated with whole sGFP, fluorescence from sCore-sGFP is associated with split sGFP, and it is possible that sGFP is less fluorescent when split.
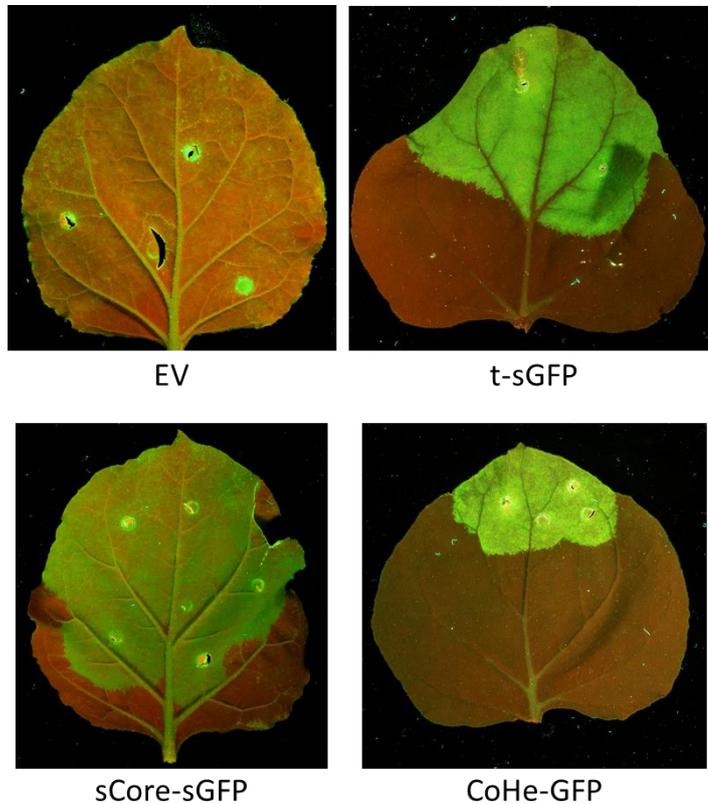
Figure 3.4 Comparing SplitCore and tandem core technologies for GFP display. Photographs of leaves expressing different constructs viewed under UV light. The caption indicates which protein(s) is being expressed by the leaf. EV: empty vector control. There is clearly more fluorescence (and therefore more correctly-folded GFP) in leaves expressing tandem constructs (right) than the SplitCore split GFP construct (bottom left). All photographs were taken 7 dpi, courtesy of Mr. Andrew Davis.

In order to compare these different constructs in more depth, small-scale protein extractions were carried out and equal volumes for each construct of the insoluble (resuspended pellet) and soluble (supernatant) fractions were analysed by western blot with the anti-HBcAg 10E11 antibody (Fig. 3.5). It should be noted that since 10E11 recognises the N-terminus of HBcAg, it is not expected to detect CoreC of the SplitCore constructs.
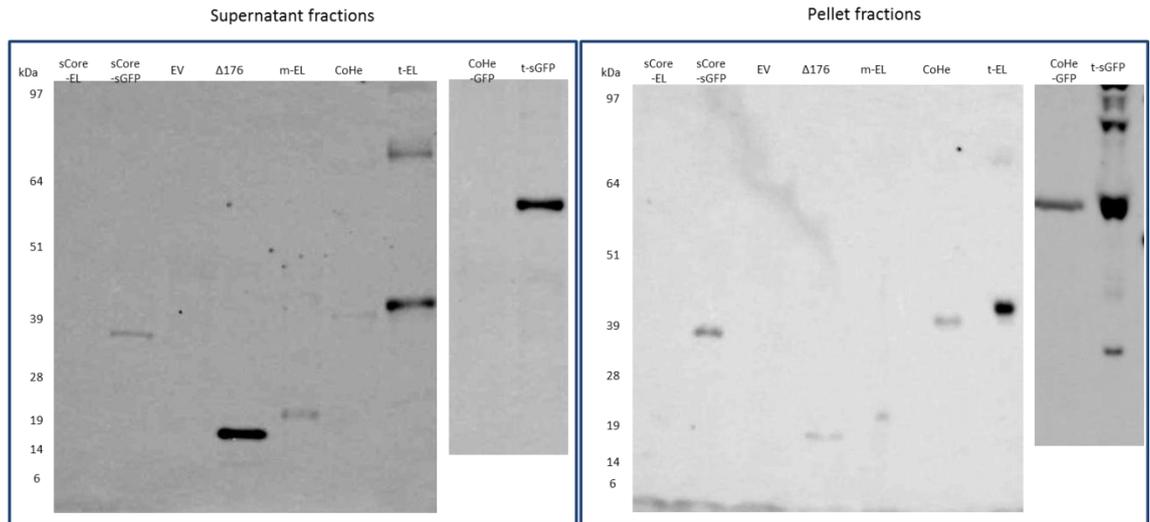
Figure 3.5 Western blots of supernatant and pellet fractions of plant extracts after expression of different construct. After a small-scale extraction, the equal volumes of supernatant fractions (left) and resuspended pellet fractions (right) were loaded onto SDS-PAGE gels for western blot analysis with the anti-HBcAg 10E11 monoclonal antibody. Monomeric constructs are expected to be about 20 kDa, while tandem core constructs are expected to be about 40 kDa, and tandem core constructs with GFP inserted in the C-terminal e1 loop are expected to be about 67 kDa. Exact sizes can be found in figure 3.3.

The results show that the SplitCore constructs did not seem to be as highly expressed. For plants infiltrated with sCore-EL, the CoreN-EL protein, with an expected size of about 9 kDa, was not detected either in the pellet or the supernatant fraction. Because of the nature of the antibody used in this western blot, we would not expect to detect CoreC. When the SplitCore system is combined with a split GFP, there is detectable accumulation of CoreN-sGFP1-10 (which is expected to a 34 kDa protein). Because the sCore-sGFP – expressing leaf tissue fluoresces, one explanation for this could be that SplitCore proteins are unstable, but the split GFP works to stabilise them by favouring the association of CoreN with CoreC. The empty vector control in the next lane shows no signal, which is to be expected. The 19 kDa monomeric wild-type HBcAg truncated at position 176 (Δ176), which has been expressed in plants before (Meshcheriakova *et al.*, 2008; Thuenemann, 2010) is expressed and mostly soluble. The 22.5 kDa m-EL construct, despite being extremely similar, does not accumulate at the same levels. There are three main differences between these two constructs. The first difference is the length: m-EL contains the full-length C-terminus of HBcAg, while Δ176 is

truncated at position 176. The second difference is the state of the e1 loop: in m-EL, the e1 loop contains a polylinker allowing easy insertion of sequences. The third main difference is that m-EL, as a distant relative of CoHe, is codon optimised for use in *E. coli*, whereas Δ176 retains the native HBV codon usage. Any of these differences, or a combination of them, could be the cause of the difference in yield.

CoHe (expected to be about 40 kDa) is seen to be expressed, though it does not seem to accumulate to the same levels as the 41.8 kDa t-EL (which was estimated to accumulate to 200-500 µg/g of fresh weight tissue, or FWT). This suggests that the modifications intended to simplify cloning in pEAQ-*HT* actually improved expression or accumulation of the protein in plants. The difference between t-sGFP and CoHe-GFP points to a similar improvement in yield: t-sGFP seems to be more abundant and more soluble than CoHe-GFP, which is consistent with the differences in fluorescence intensity shown in figure 3.4. The insolubility of CoHe-GFP seen here is in agreement with the findings of Thuenemann (2010), who developed the more complex HbexB1 extraction buffer to improve solubility of CoHe-GFP. The simpler P buffer was used for this experiment, which reveals superior yield and solubility of t-sGFP compared to CoHe-GFP. This is probably due to both the superiority of the t-EL moiety as well as the modifications made to the GFP sequence. Interestingly, the sGFP insert is inserted in the e1 loop of t-EL without any flexible linkers added on either side, relying only on the inherent flexibility of the N- and C-termini of GFP and of the e1 loop. This is thought to be impossible with monomeric HBcAg, in which long extra-long flexible linkers are thought to be necessary (Kratz *et al.*, 1999; Janssens *et al.*, 2010). Moreover, it is clear from these western blots that t-sGFP is not significantly degraded: the majority of the signal for this construct appears in the 67.8 kDa band, although there does seem to be some degradation product in the insoluble fraction. It is also noteworthy that higher molecular weight specific bands were seen for CoHe, CoHe-GFP, t-EL and t-sGFP. These are diagnostic of tandem core proteins, and were seen in every western blot for every plant-expressed tandem core construct described in this thesis,

despite the fact that samples were prepared by boiling in LDS and β-mercaptoethanol before loading on the SDS-PAGE gel. Taken together, these data clearly indicate that tandem core technology is better suited to chimeric CLP production in plants than SplitCore technology. Future work was therefore carried out using the tandem core system.

Once it had been established that m-EL, t-EL and t-sGFP are expressed in plants, it was necessary to demonstrate that they form CLPs. In order to do this, large scale extraction was carried out with a final partial purification step involving ultracentrifugation over a sucrose cushion (see Chapter 2). These partially purified fractions were then analysed by transmission electron microscopy (TEM, figure 3.6). This showed that all three constructs were capable of directing the formation of particles in plants. These particles had the distinctive appearance of CLPs, indicating that the modifications carried out to the basic HBcAg sequences, as well as insertion of GFP, did not abolish assembly. As with native wild-type HBcAg, the majority of CLPs were T=4, although some T=3 particles were observed. Moreover, the sucrose cushion interface fraction for t-sGFP fluoresced under UV light, indicating that the GFP is not cleaved from the CLP. An attempt to partially purify sCore-sGFP in a similar manner is described in chapter 4, and did not result in the detection of CLPs.
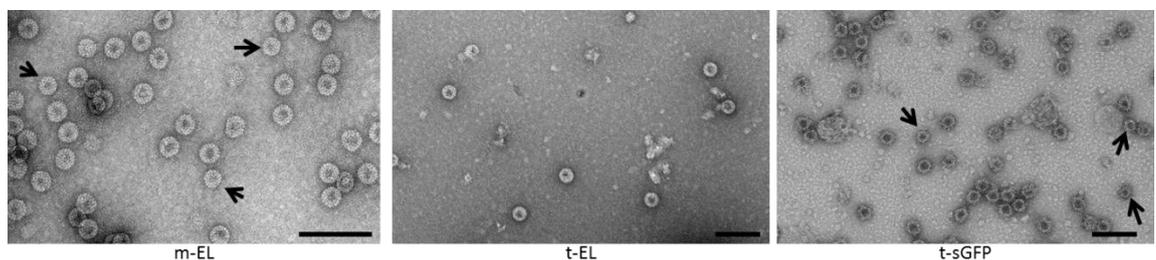
Figure 3.6 Electron micrographs of plant-produced CLPs expressed using different constructs. The caption indicates the construct used to produce the CLPs. Left: monomeric m-EL, middle: tandem core t-EL , right: tandem core with sGFP inserted in the C-terminal e1 loop. Arrows indicate T=3 particles, scale bars are 100 nm.

IV - Discussion

The data presented here describe a continuation of the work carried out by Thuenemann (2010). That work described the transfer of tandem core technology from a bacterial expression system to a plant expression system. It concluded that plants are a suitable host for tandem core technology, and that a heterotandem construct is more suitable for antigen display than a homotandem construct. The work described in this chapter built upon this previous work by attempting to make the constructs more compatible with the pEAQ-*HT* expression system, and the whole tandem core concept was compared to the more recent SplitCore system (Walker *et al.*, 2011). The results show that optimising the tandem core construct for use in pEAQ-*HT* actually improves the yield of protein obtained. This could be due to the polylinkers added in the e1 loops: these contain short glycine-rich sequences which may improve the solubility (and therefore stability) of the whole complex, thereby increasing the amount of material accumulated in plant leaves. Interestingly, the monomeric m-EL did not accumulate as much as t-EL, which suggests that the modifications made to the sequence only improve yield in a tandem construct. This is logical in the sense that the "ideal" monomeric HBcAg sequence is probably the native sequence, which has evolved to be stable and produced at high titres. Modifying the sequence to make a tandem core construct essentially creates a new protein, which might not be optimal in terms of stability or capsid assembly unless a few extra modifications are carried out. It is possible that such modifications are, serendipitously, the ones that I implemented during optimisation for use in pEAQ-*HT*. Moreover, modifying the sequence of GFP to improve its solubility as well as its suitability for split GFP experiments (Cabantous *et al.*, 2005; Pedelacq *et al.*, 2006; Walker *et al.*, 2011) clearly improved the yield of GFP-displaying tandem core compared to the previously used version of GFP. This demonstrates the importance of the insert sequence for yield and solubility of the whole chimeric CLP. The new, modified constructs were not only expressed in

plants, but also directed the production of chimeric CLPs which look indistinguishable from wild-type HBcAg CLPs in TEM analysis. This proves that the modifications made to the original sequence did not abolish capsid assembly, thus demonstrating the incredible flexibility of HBcAg as a biotechnological and research tool.

These data leave unanswered the question about the role of *E. coli* codon-optimisation in terms of recombinant protein yield in a plant expression system. It is possible that plant codon-optimisation would result in higher yields of chimeric CLPs, but this was not attempted. There are examples of codon-optimisation actually reducing yields of recombinant protein (Maclean *et al.*, 2007), so it was considered safer to continue working with *E. coli* – optimised sequences, since these yield enough material for the necessary work to be carried out. However plant codon-optimisation would be an interesting follow-up to the data presented in this thesis.

The SplitCore system was tested in plants and did not seem to be as suited to the plant expression system as it was to *E. coli* (Walker *et al.*, 2011). Indeed, expression could only be detected when split GFP was deployed, suggesting that CoreN and CoreC are rapidly degraded in plants unless they are brought together to form a complete HBcAg, which appears to be very inefficient unless the split GFP acts like a chaperone to bring the two segments together. This, in turn, means it is unlikely that inserting another protein would lead to higher yield than sCore-EL. By contrast, in *E. coli,* the use of split GFP was not necessary to obtain high expression of SplitCore CLPs (Walker *et al.*, 2011). While it is true that the conclusion about plant-produced SplitCore is based on a single set of experiments experiment, the result obtained does not bode well for display of other proteins of interest using SplitCore in plants, since insert proteins cannot always be split and assembled as readily as sGFP. It is also true that the results presented in figures 3.4 and 3.5, which show the comparison between tandem core and SplitCore technologies, involve the GFP being split in the SplitCore system but not in

the tandem core system. Had time allowed, it would have been useful to carry out this experiment with whole (non-split) GFP in either CoreN or CoreC.

The important difference between the plant and bacterial SplitCores is that in *E. coli*, a bicistronic operon was used, whereas in the plant-based system deployed here, the two halves of the protein were translated from different mRNAs, albeit transcribed from the same T-DNA. This difference may reduce the relative efficiency with which the two halves interact in the plant system. It may be possible to develop a bicistronic approach analogous to that used in *E. coli* by using an internal ribosome binding site (IRES) which is functional in plants, such as those present in the genomes of the crucifer-infecting tobamovirus (Dorokhov *et al.*, 2002) or the *Rhopalosiphum padi* virus (Groppelli *et al.*, 2007). However, even if this could be done, it would probably be inefficient at producing equimolar amounts of the CoreN and CoreC moieties. An alternative could be to use a picornavirus 2A self-cleaving peptide, which could be used to link CoreN and CoreC on the same ORF, and have them separate upon translation through a ribosome stutter mechanism (Minskaia *et al.*, 2013; Minskaia and Ryan, 2013).

Moreover, it is conceivable that tandem core technology might be combined with SplitCore technology, in an attempt to develop a system that combines the advantages of each. In such a hypothetical system, a tandem core construct could include an insert (such as GFP) in the C-terminal e1 loop flanked on either its N-terminal or C-terminal end by a 2A or 2A-like sequence. In this scenario, a single transcript would direct the production of a polypeptide chain which would be cleaved during translation at the 2A site. One terminus of the insert (depending on the position of the 2A sequence) would be free and untethered to the rest of the tandem core molecule, which may favour folding. However, it is also possible that splitting the tandem in this way may simply prevent the structural dimer from folding properly. None of these approaches were investigated because the demonstration that the tandem core technology was suitable for plant-based expression of CLPs obviated the need to optimise

SplitCore. However, the optimisations described above would be a worthy follow-up to the experiments presented in this thesis. In conclusion, the experiments described in this chapter showed that t-EL was a suitable construct on which to base future work on the display of protein inserts.

Chapter 4 : Two strategies for a generic protein presentation system

I - Introduction

The work described in the previous chapter demonstrated that the new tandem core construct, t-EL, is a good candidate to act as a scaffold for antigen display on the surface of a CLP. The ultimate goal of the work described in this thesis was to develop a generic antigen scaffold system based on HBc. A generic system would involve post-translational attachment of the antigen to the CLP scaffold, so that the same scaffold (or very similar ones) could be produced regardless of the antigen to be fixed onto its surface. Such a system might already exist, thanks to a system allowing non-covalent linking of proteins of interest to the e1 loop of HBcAg: this is through the use of a peptide tag, GSLLGRMKGA, which specifically binds the e1 loop of HBcAg and can be fused to a protein of interest (Blokhina *et al.*, 2013). However, binding of a peptide of interest (Influenza M2e) to CLPs via this tag is undone during sucrose cushion purification. This suggests that binding with this tag is likely to be relatively weak, which means that it is worth developing alternative systems. Two techniques applicable to the tandem core system fit these criteria: antibody–antigen interaction, and chemical conjugation.

In the first scenario, a version of HBc would be designed which would contain an antibody fragment, such as a single chain variable fragment (scFv), in the e1 loop (Fig. 4.1a). Such an antibody fragment would have specificity to an antigen of interest, and the antigen could then be used to coat the particle post-translationally. An advantage of such a system would be that the antigen would be presented in the same orientation at each position around the CLP, since the antibody would always bind to the same epitope. Moreover, the specificity of an antibody-antigen interaction means that the attachment of the antigen to the CLP could be carried out after purification, during purification, or even *in situ* in the expression system if both elements

are co-expressed. The disadvantages are that the antigen would not be covalently linked to the particle surface, and switching from one antigen to another would require changing the antibody fragment. This means that this system would only be generic insofar as we consider all scFv to be identical in terms of surface properties for the purposes of expression, capsid assembly, extraction and purification. This point could be addressed by using a scFv against a common tag that could be readily added to any antigen of interest, although this would require modifying the antigen.

In the second scenario, a version of HBc would be designed which would readily allow chemical conjugation of any antigen onto the e1 loops. This would be done by inserting a lysine-rich linker in the e1 loop which would allow antigens of interest to be covalently linked to the loop through bioconjugation (Fig. 4.1b). Lysine residues are commonly used for bioconjugation because the primary amine is amenable to a wide variety of different bioconjugation methods. For example, Pastori *et al.* (2012) used the bifunctional cross-linker succinimidyl-6-(β-maleimidopropionamido)-hexanoate (SMPH) to cross-link C-terminal cysteine residues of target epitopes to surface lysines of the carrier viral particle (in this case, bacteriophage AP205). Moreover, bioconjugation is not limited to proteins. Indeed, carbohydrates can be bioconjugated onto lysine residues via reductive amination: the vicinal diol of a sugar can be oxidised to two aldehydes which react with the primary amine of the lysine residue, thus forming an imine group. This imine group, which covalently links the lysine and the sugar, is unstable, but can then be stabilised by reduction to a secondary amine. The advantages of this display method are that conjugated antigens would be covalently bound to the CLP surface, and a single conjugation-ready construct would be a suitable scaffold for essentially any antigen with easily-addressable surface residues, making it truly generic in this sense. The potential disadvantages are that it would be difficult to guarantee the homogeneity of a sample of conjugated CLPs, both in terms of the number of antigens presented on a single particle, and in terms of the conformation of each conjugated antigen molecule: indeed,

depending on the conjugation method, the antigen may be attached to the CLP via one of many possible surface-exposed addressable chemical groups, so for most antigens there could be a plethora of different ways in which the molecule could be presented on the surface of the CLP. This could have negative effects on the antigenicity of the conjugated antigen: it is thought that antigens displayed on nanoparticles or virus-like particles (VLPs) have increased antigenicity because they are presented in a regular, quasi-crystalline array. Moreover, a practical consideration is that chemical conjugation is a non-specific process, which means that the conjugation reaction would need to take place with extremely pure CLP and antigen in order to avoid conjugating contaminants onto the surface of the particles.
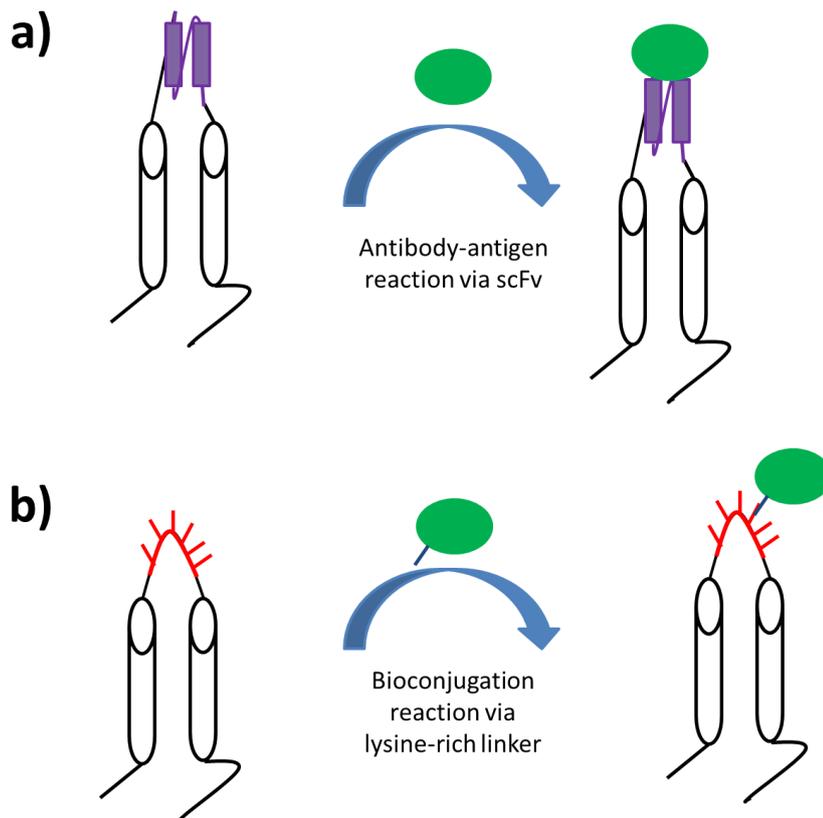
Figure 4.1 Two strategies for a generic antigen presentation system. a) A single-chain antibody (scFv, in purple) is fused to the C-terminal e1 loop (black) of a tandem core construct. When the cognate antigen (green) is placed in contact with the chimaeric core protein, the antigen binds the core through an antibody-antigen reaction. b) A lysine-rich peptide (red) is fused to the C-terminal e1 loop (black) of a tandem core construct. When an antigen of interest (green) with addressable surface residues (blue line) is placed in contact with the modified core protein, the antigen can be covalently linked to the core through a variety of bioconjugation techniques.

II - Design of constructs

In order to test both scenarios, two different constructs based on t-EL were designed (see below and chapter 3). The first, named t-α-phOx, contains an insert in the C-terminal e1 loop composed of the anti-phOx scFv flanked on either side by a glycine-rich 15 amino acid – long linker. This scFv specifically binds to the small organic hapten phenyl-oxazolone, or phOx, and has been used numerous times over many years. The anti-phOx scFv (or α-phOx) is described by Marks *et al.* (1992), and has since been displayed on the surface of bacteriophages (Hayashi *et al.*, 1995); secreted by insect cells using a baculovirus expression system (Kretzschmar *et al.*, 1996); expressed in the bacterial periplasm (Schmiedl *et al.*, 2000); displayed on the surface of *Bacillus thuringiensis* (Bt) spores (Du *et al.*, 2005); and displayed on the surface of mammalian cells (Chesnut *et al.*, 1996). In fact, a cell transfection kit based on phOx – mediated capture of α-phOx – displaying cells called Capture-Tec pHook-3 System was commercially produced by Invitrogen (now Life Technologies). This included the use of phOx-coated magnetic beads to capture cells displaying functional α-phOx. Moreover, protocols exist for "phoxylation" of proteins, notably bovine serum albumin (BSA), in order to assay recombinant α-phOx function (Chesnut *et al.*, 1996; Schmiedl *et al.*, 2000; Du *et al.*, 2005). Crucially, the sequence of α-phOx was submitted to GenBank by Hayashi *et al.* (1995), where it is publicly available (accession number X82190.1). All of these factors made α-phOx an ideal candidate to provide proof of concept for the display of scFv on the surface of HBc. The second construct, named t-KD, was designed to allow chemical conjugation and contains an insert in the C-terminal e1 loop composed of six lysine residues flanked on either side by three aspartate residues (a "KD" linker) in order to balance the overall charge. This should allow antigens to be linked directly to the e1 loop by chemical conjugation. In order to determine whether tandem core technology was specifically useful for these two scenarios, monomeric equivalents of both t-α-phOx and t-

KD, named m-α-phOx and m-KD respectively, were also produced based on the m-EL construct (see chapter 3). Diagrams of the different constructs are shown in figure 4.2, and details of the cloning are given below.

t-α-phOx : 1956 bp, 652 aa, 69.5 kDa

m-α-phOx : 1407 bp, 469 aa, 50.1 kDa

t-KD: 1155 bp, 385 aa, 42.6 kDa

m-KD: 606 bp, 202 aa, 23.3 kDa

N-terminal HBcAg core (*E. coli* codon-optimised)	Gly/Ser -rich linker	Nos terminator
C-terminal HBcAg core (*E. coli* codon-optimised)	Anti-phOx single-chain variable fragment antibody	CPMV-*HT* UTRs
Full-length arginine-rich C-terminus of HBcAg	Neomycin phosphotransferase	35S promoter
P19 suppressor of gene silencing	*Agrobacterium* T-DNA right border and left border	pEAQ-optimised c/e1 loop
KD peptide: GDDDKKKKKKDDDG		

Figure 4.2 Diagrams of construct. Maps of the constructs t-α-phOx (tandem core with anti-phOx scFv in the C-terminal e1 loop), m-α-phOx (monomeric core with anti-phOx scFv in the e1 loop), t-KD (tandem core with a lysine-rich peptide in the C-terminal e1 loop), and m-KD (monomeric core with a lysine-rich peptide in the e1 loop). The key is the same as in figure 3.3a.

A) Cloning

i) t-α-phOx and t-α-phOx-OPT sequences

In order to maximise the use of a single sequence ordered for synthesis from GeneArt (Life Technologies), the t-α-phOx construct (tandem core with anti-phOx scFv sequence in C-

terminal e1 loop) was designed with features enabling the easy cloning of t-EL, m-EL, and m-α-phOx from this one sequence. The sequence of tandem HBcAg used in this work was based on CoHe7e used by Thuenemann (2010). This is itself based on the full-length 185 amino acid sequence capsid protein sequence published Ono *et al.* (1983), available on GenBank (UniProtKB/Swiss-Prot: P03149.1). Like CoHe7e, t-α-phOx is codon-optimised for *E. coli*, the N-terminal core is truncated at position 149, a linker was used to link the N- and C-terminal cores, and the C-terminal core contains the full-length arginine-rich C-terminus. To simplify cloning and use in the pEAQ-*HT* plasmid, and to derive the aforementioned constructs from this one, a few modifications were implemented. In the N-terminal core e1 loop, the polylinker *XmaI*-Ala-Gly-Gly-Ser-Ser-Gly-*MfeI* was inserted between E77 and D78. In the C-terminal e1 loop, amino acid E77 was replaced with the polylinker *SalI*-Ala-Gly-Gly-Gly-*AvrII*-Gly-Gly-Ser-Gly-*AseI*, with the anti-phOx scFv sequence (Hayashi *et al.*, 1995), with silent mutations to remove unwanted restriction sites, inserted in the AvrII site. Between the N-terminal and C-terminal cores, the peptide linker (GGS(SGG)$_6$TGT) was used, which is very similar to that used in CoHe7e, but it contains an AgeI-like restriction site at the C-terminus (Thr-Gly: ACTGGT), which differs from an authentic AgeI restriction site (ACCGGT) by a single nucleotide. This is followed by the ACA codon (Thr) just before the starting methionine of the C-terminal core. A Kozak consensus sequence (AACAATG) was used upstream of the start codon after the AgeI restriction site, and an XhoI site was added after the stop codon. This construct was ordered for synthesis from GeneArt (Life Technologies) and cloned into pEAQ-*HT* with the AgeI and XhoI restriction sites. In later experiments, the anti-phOx scFv sequence was ordered for synthesis again with codon use optimised for *Nicotiana benthamiana* and cloned into t-EL using the SalI and AseI restriction sites: this construct was named t-α-phOx-OPT.

ii) m-α-phOx sequence

The m-α-phOx construct (monomeric core with anti-phOx scFv sequence in the e1 loop) was made by amplifying the C-terminal core of t-α-phOx with an upstream primer containing the

authentic AgeI site instead of the AgeI-like site, thus creating a monomeric construct with an upstream AgeI site followed by a favourable Kozak context. This construct was inserted into pEAQ-*HT* using AgeI and XhoI.

iii) t-KD sequence

The sequence for t-KD (tandem core with GDDDKKKKKKDDDG peptide insertion in C-terminal e1 loop) was prepared from t-EL (see chapter 3) and a set of oligonucleotides containing the KD sequence. Partially overlapping 5'-phosphorylated oligonucleotides coding for the above peptide with 5' SalI and 3' AseI restriction site overhangs were ligated into pEAQ-*HT*-t-EL which had been digested with SalI and AseI.

iv) m-KD sequence

The m-KD sequence (monomeric core with with GDDDKKKKKKDDDG peptide insertion in e1 loop) was prepared from m-EL (see chapter 3) in the same way as t-KD (see above).

III - Properties of t-α-phOx

To test expression of t-α-phOx and m-α-phOx, the constructs were individually agroinfiltrated into *N. benthamiana* leaves and small scale extractions were carried out in parallel with leaves infiltrated with m-EL, CoHe and t-EL (see chapter 3) in order to compare the expression levels of empty-loop and scFv-bearing constructs. The soluble (supernatant) and insoluble (resuspended pellet) fractions were assayed by western blotting (Fig. 4.3).
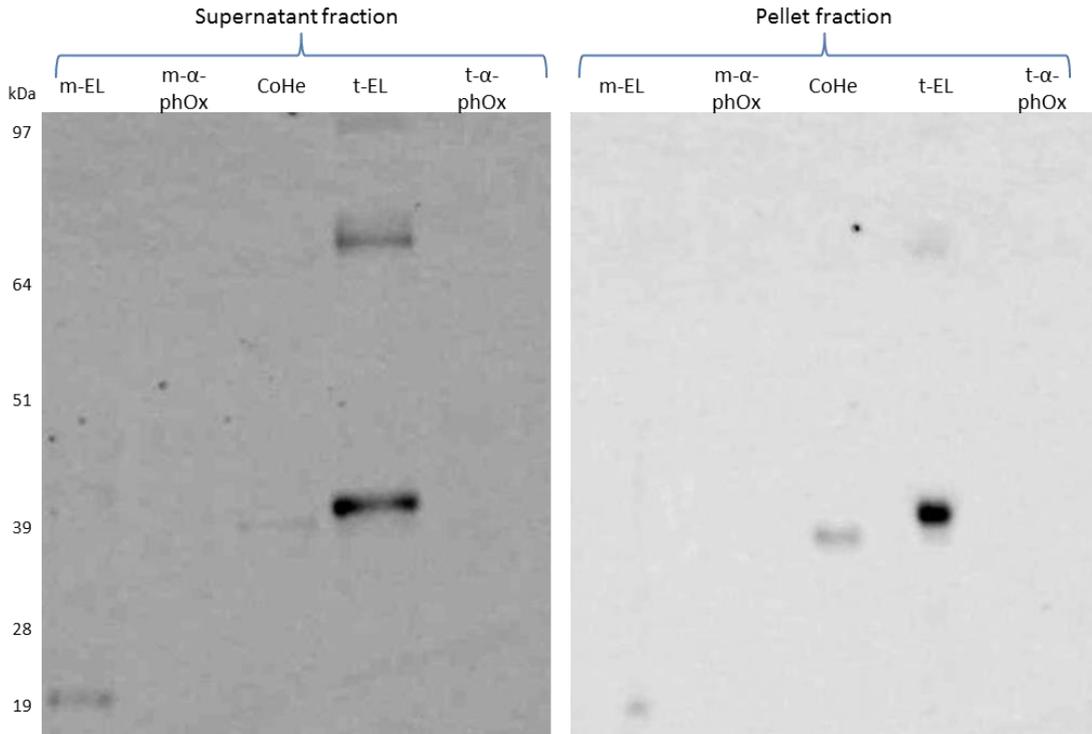
Figure 4.3 Western blots of supernatant and pellet fractions of plant extracts after expression of different constructs. After a small-scale extraction, the equal volumes of soluble fractions (left) and resuspended pellet fractions (right) were loaded onto SDS-PAGE gels for western blot analysis with the anti-HBcAg 10E11 monoclonal antibody. Monomeric constructs are expected to be about 20 kDa (47 kDa with the anti-phOx scFv inserted in the e1 loop), while tandem core constructs are expected to be about 40 kDa, (67 kDa with the anti-phOx scFv inserted in the C-terminal e1 loop). Exact sizes can be found in figure 4.2. These are the same western blots as shown in figure 3.5.

This western blot shows that there is expression of soluble m-EL (22.5 kDa) and t-EL (41.8 kDa, this is in fact the same western blot as shown in Chapter 3), with CoHe (40 kDa) accumulating at lower levels. However, there is no sign of expression from either t-α-phOx (with an expected size of about 69 kDa) or m-α-phOx (with an expected size of about 50 kDa). This disappointing result seemed worthy of further investigation since tandem cores with properly-folded, 27 kDa GFP displayed on the surface could be produced (see chapter 3). Another experiment was therefore set up in which leaf tissue infiltrated with t-α-phOx underwent a large-scale extraction with HBexB1 buffer (which yields cleaner extracts than P buffer) followed by a sucrose step gradient with increasing concentrations of sucrose (20, 30, 40, 50, and 60 % sucrose). Crucially, aliquots were taken at every step in the process in order, and these were all

analysed on the same anti-HBcAg western blot in order to determine if any material was produced, and how it behaved during the extraction process and sucrose step gradient (Fig. 4.4). No further work was carried out on m-α-phOx.
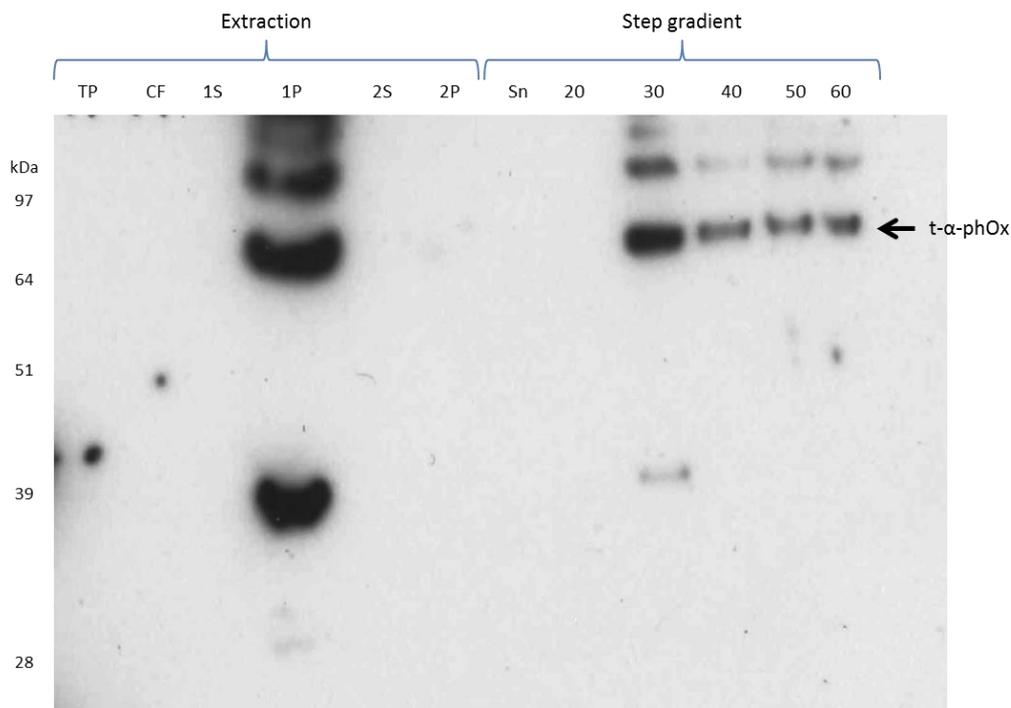
Figure 4.4 Western blot of the fractions obtained at each step of extraction and purification of t-α-phOx. TP: total protein extract. CF: crude filtrate. 1S: supernatant from the first clarification centrifugation step. 1P: pellet from the first clarification centrifugation step. 2S: supernatant from the second clarification centrifugation step. 2P: pellet from the second clarification centrifugation step. The following step gradient fractions were collected after ultracentrifugation: Sn: supernatant of the sucrose step gradient. 20: 20% sucrose fraction. 30: 30% sucrose fraction. 40: 40% sucrose fraction. 50: 50% sucrose fraction. 60: 60% sucrose fraction. The primary antibody used was the anti-HBcAg 10E11 monoclonal antibody, and t-α-phOx has a predicted mass of 69.5 kDa.

This result shows reveals that soluble t-α-phOx is difficult to detect. Indeed, nothing was detected in the total protein extract (TP, leaves just after blending in buffer), the crude filtrate (CF, just after filtration of the total protein extract over miracloth), or the supernatant of the first clarification spin (1S, just after centrifugation of the crude filtrate). There was, however, signal in the fraction corresponding to the pellet of the first clarification spin (1P), indicating that although t-α-phOx is expressed, it is mostly insoluble. There are also higher molecular

weight bands, which, as mentioned in chapter 3, are diagnostic of all tandem core – based constructs. This could correspond to stable multimers of tandem cores, or (at least in some cases), improperly folded aggregates. Moreover a band around 39 kDa in size was reproducibly seen in insoluble t-α-phOx (which is expected to be 69 kDa in size). While this could be a sign of degradation of the protein, a band of similar size was sometimes seen in western blots of plant extracts, even when these came from plants agroinfiltrated with an empty vector control (see below and chapter 6). The soluble fraction of the extract (1S) was centrifuged a second time and although t-α-phOx was not detected in either the supernatant or the pellet, this second soluble fraction (2S) was used to load the sucrose step gradient. After ultracentrifugation, t-α-phOx could be detected in the sucrose fractions, particularly the 30% sucrose fraction, this time with much less non-specific antibody binding (or degradation) being observed. Whatever that 39 kDa band is, it does not seem to be incorporated into the soluble particulate matter formed by t-α-phOx. It should be pointed out that the 30% sucrose fraction (along with the 40% fraction) is typically where t-EL particles were seen to sediment in similar experiments. While it does show that soluble t-α-phOx is forming particulate material, it is not proof that this particulate material is definitely CLPs. Moreover, the fact that t-α-phOx is detectable in the sucrose fractions but not the 2S fraction indicates that t-α-phOx is only readily detectable when it has been concentrated to a certain degree: the material detected in the pellet is essentially a very concentrated form of the insoluble proteins present in the crude filtrate, and the sucrose fractions are concentrated forms of the particulate matter present in the second soluble fraction (2S). This reveals that t-α-phOx is expressed at low levels in plants, and that the vast majority of the material is insoluble. The fraction which is soluble migrates down a sucrose step gradient, which indicates that it is particulate, but it does not seem to settle in one specific part of the gradient, indicating that the particulate matter is heterogeneous. It is possible to detect t-α-phOx in the more dilute fractions by significantly increasing the exposure time during ECL detection (data not shown).

These results suggested that it would be possible to further characterise t-α-phOx if enough material could be produced and purified. Thus, plant tissue agroinfiltrated with t-α-phOx was extracted in either HBexB1 or P buffer and concentrated over sucrose cushions or step gradients and any particles present in the fractions were then visualised by TEM: in these cases no CLPs were ever observed, despite multiple attempts. Another experiment was therefore set up to try to produce a purer, more concentrated sample of t-α-phOx. As controls, and for comparison, m-EL, t-EL, CoHe, m-α-phOx, and sCore-sGFP (see chapter 3) were also used in this experiment. Leaf material expressing the different constructs was used for large-scale extractions (see chapter 2) followed by a sucrose step gradient. The sucrose fractions in which particles were expected, or hoped, to be found (20%-60% for t-α-phOx and 30%-60% for m-α-phOx) were pooled, diluted with water, and concentrated further on YM-100 Microcon centrifugal filter devices (Millipore) as per manufacturer's instructions. These filter units allowed the removal of sucrose from the samples while concentrating them. Importantly, the molecular weight cut-off (MWCO) of these filter units is 100 kDa, so unassembled monomers of any construct would be expected to be removed from the sample. The concentrates thus obtained were analysed by anti-HBcAg western blot (Fig. 4.5).
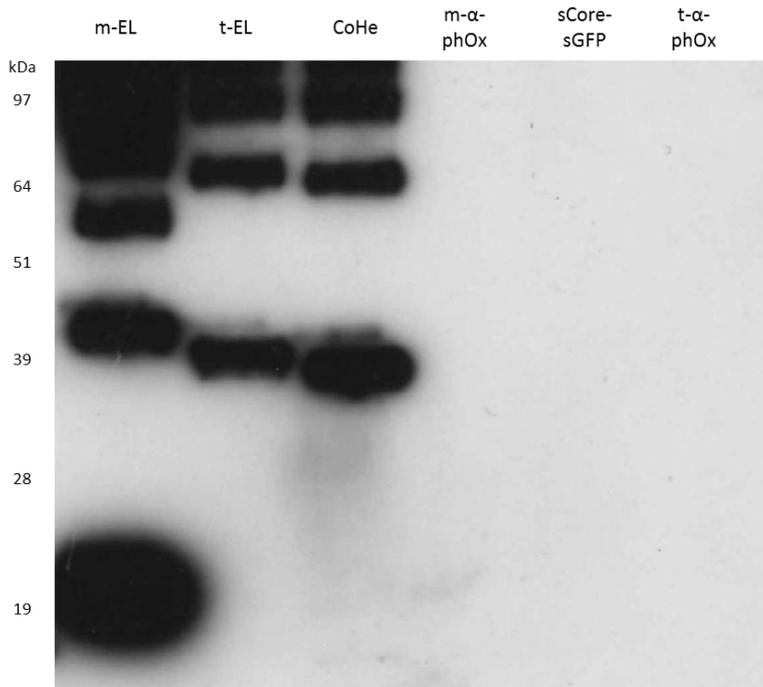
Figure 4.5 Western blot of different protein extracts after concentration. Large-scale extractions were carried out followed by a sucrose gradient, and further concentration with a Microcon centrifugal filter device. While m-EL (22.5 kDa), t-EL (41.8 KDa) and CoHe (41 kDa) are readily detected; m-α-phOx (50.1 kDa), CoreN-sGFP1-10 (34 kDa, the detectable moiety of sCore-sGFP) and t-α-phOx (69.5 kDa) are not. The primary antibody used was the anti-HBcAg 10E11 monoclonal antibody.

Again, the characteristic ladder pattern involving higher molecular weight structures was seen, this time with the monomeric core as well as the tandem core constructs. In fact the ladder pattern seen with m-EL makes it very clear that this does correspond to multimers, as the well-defined bands are clearly one (~20 kDa), two (~40 kDa) and three (~60 kDa) times the size of m-EL. This is less clear with tandem core constructs: since they are so large to begin with, multimers are always in the less-well defined higher molecular weight area of the gel. The presence of these bands revealed that while m-EL, t-EL and CoHe could be concentrated using the method described above, m-α-phOx, t-α-phOx, and sCore-sGFP were not detected. This could be because these proteins do not assemble into structures larger than 100 kDa (and were therefore lost during filtration), or because higher order structures like particles are present in such small amounts to begin with that they were not sufficiently concentrated, and any signal that there might be from these constructs is drowned out by the much stronger

signal from the other, positive samples. It is also theoretically possible that particulate matter pelleted through the gradient, but if that were the case one might expect to detect something in the 60% sucrose fraction (the bottom fraction), which was included in the pooling and concentration.

In any case, this is not conclusive proof that t-α-phOx does not form CLPs, although it clearly shows the struggle that might be necessary to obtain enough material to have a chance of observing such particles if they do in fact exist. It would in theory be possible to conclusively exclude the possibility that t-α-phOx forms CLPs, if large amounts of soluble t-α-phOx were readily detected, but never penetrated low-concentration sucrose layers during a sucrose cushion or gradient. The problem here is that low yield of soluble protein makes detection in any fraction difficult. Moreover, this low yield makes it unlikely that enough material could ever be produced for further work.

Attempts at optimising extraction and concentration of t-α-phOx were carried out with the assistance of an undergraduate student (Angelika Jaroschek). Among these experiments were attempts at developing a polyethylene glycol (PEG) precipitation protocol (which had not been attempted for other CLPs) for concentrating t-α-phOx from crude plant extract prior to sucrose gradient analysis. An initial concentration step such as this could be used to increase the amount of starting leaf material that could be used for purification. These experiments never resulted in any demonstrable improvement in obtaining concentrated soluble t-α-phOx. Another approach was an attempt to produce mosaics in order to reduce the number of scFv molecules on each assembled CLP: to this end, plants were co-infiltrated with t-α-phOx and either t-EL or m-EL in various ratios. The hope was that capsid formation might be rescued if only some of the spikes contain the scFv insert. Thus, plants were co-infiltrated with mixes of *Agrobacterium* suspensions consisting of t-α-phOx and either t-EL or m-EL in ratios of 10:1, 1:1, or 1:10. After small-scale extractions in either AngeliTris buffer or AngeliBor buffer

(respectively tris and borate – based buffers which had been previously determined by Angelika to be no worse than HBexB1 for the extraction of t-α-phOx from plant tissue), the soluble fractions were analysed by anti-HBcAg western blot (Fig. 4.6). In theory, if particle formation could be rescued by the mosaic method, more t-α-phOx should be seen here than in previous experiments: it is likely that in previous experiments, t-α-phOx was not forming particles and was therefore largely degraded in the plant, which explains why so little of it was ever detectable. If mosaics are a success, however, then t-α-phOx should be detectable in amounts comparable to t-EL.
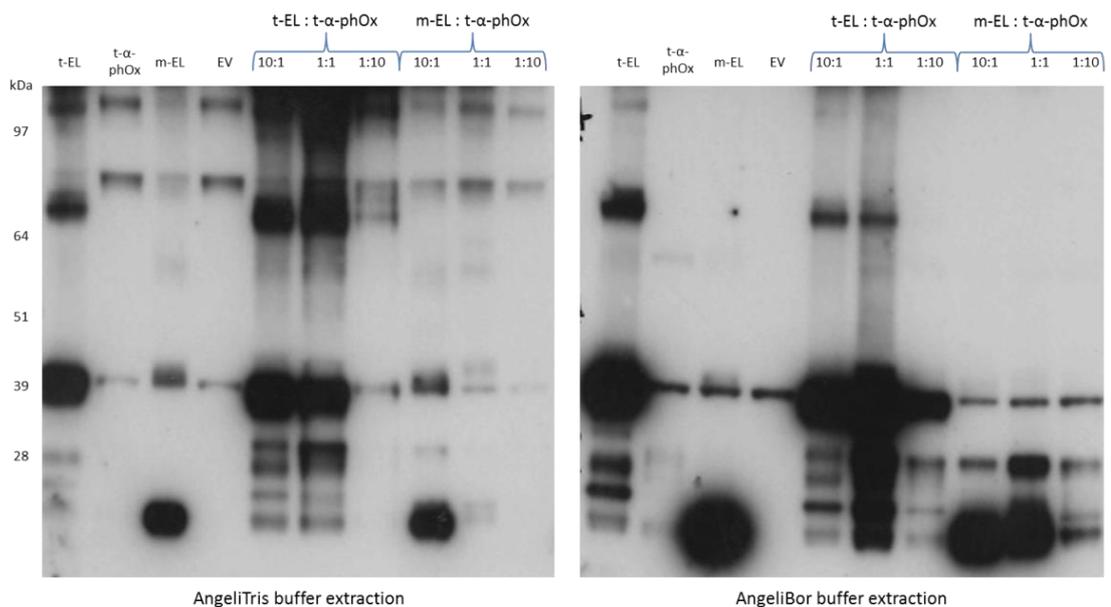
Figure 4.6 Western blots of plant extract soluble fractions after expression of mosaic HBcAg constructs. In order to rescue particle formation with t-α-phOx (69.5 kDa), Agrobacteria carrying either t-EL (41.8 kDa) or m-EL (22.5 kDa) constructs were co-infiltrated with Agrobacteria carrying t-α-phOx (right on each blot). The relative ratios of the bacteria carrying the different constructs are shown: 1:10, 1:1, or 10:1. The controls (left of each blot) are similarly-treated soluble fractions of plants expressing only t-EL, t-α-phOx, m-EL or the empty vector control (EV). All lanes represent soluble fractions obtained after small-scale extractions carried out using either AngeliTris (left) or AngeliBor (right) extraction buffers.

It is important to note the apparently non-specific bands which are present in all samples, including the empty vector negative control (EV): a 39 kDa band is visible in all samples for

both sets of extractions, and two larger bands above 64 kDa are also visible in all samples from AngeliTris buffer extractions. This means that the bands visible in t-α-phOx extractions cannot be considered to be specific. For both extraction buffers, and for both m-EL and t-EL, there is no visible influence from the presence of t-α-phOx. The co-infiltrations with t-EL or m-EL show bands of approximately 40 and 20 kDa, respectively, which correspond to these constructs and not t-α-phOx. The amount of material in the co-infiltrations correlates to the amount of t-EL or m-EL used, and the presence of *Agrobacteria* directing the expression of t-α-phOx seems to have no effect. For both extraction buffers, and for both m-EL and t-EL, there is no visible influence from the presence of t-α-phOx: mosaics could not be formed with t-α-phOx. This experiment was repeated with the same buffers used for extraction as well as HBexB1, and the ultimate result was identical. A final attempt at increasing the yield of t-α-phOx consisted of using a codon-optimised sequence of the anti-phOx scFv and using this new sequence to create a new tandem core construct with the scFv in the e1 loop. This construct, named t-α-phOx-OPT (see section II.A.i above), failed to improve accumulation of the t-α-phOx protein in plants, and failed to direct the production of detectable CLPs.

IV - Properties of t-KD

In order to test expression of t-KD and m-KD, small scale extractions were carried out on plants infiltrated with three different *Agrobacterium* clones (transformed with the same sequence-verified plasmid) for each construct. The extractions were carried out in both P and HBexB1 extraction buffers. A western blot of the soluble fraction was then carried out to assay expression of the proteins (Fig. 4.7). The results clearly show that while t-KD (with an expected size of about 42 kDa) is expressed in plants and can be extracted using either buffer, m-KD (with an expected size of about 22 kDa) either does not accumulate in leaf tissue or cannot be extracted in either buffer. A possible explanation for this is that the lysine-rich linker abolishes

HBcAg dimer formation if it is present in both e1 loops of a dimer, possibly through repulsion caused by the charged residues inserted in the e1 loop. If dimer formation is abolished, then capsid assembly is impossible, so the m-KD monomers would likely be degraded.
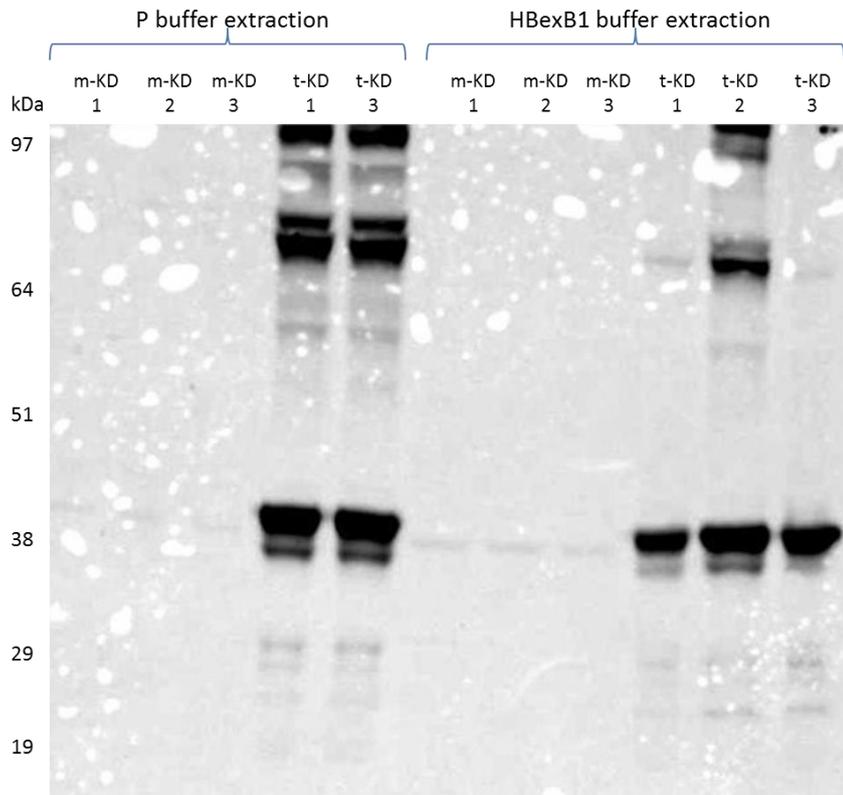
Figure 4.7 Detection of m-KD or t-KD in soluble plant extracts following agroinfiltration with different *Agrobacterium* clones. PCR-positive *Agrobacterium* clones (identified by numbers) carrying monomeric m-KD (23.3 kDa) or tandem t-KD (42.6 kDa) constructs were used to agroinfiltrate *N. benthamiana*. Small-scale extractions were performed in either P (left) or HBexB1 (right) extraction buffers, and the soluble fractions were analysed by western blot using the anti-HBcAg 10E11 monoclonal antibody. While soluble protein was detected for the tandem construct, this was not the case for the monomeric construct.

Once expression of t-KD had been established, it was necessary to demonstrate particle formation. Because both P and HBexB1 buffers seemed suitable for extracting t-KD, parallel large-scale extractions were set up in both buffers before partial purification over sucrose cushions. The samples were then observed by TEM in order to confirm the presence of CLPs (Fig. 4.8). Particles were indeed observed in samples from both extractions, but those

originally extracted in P buffer seemed much more homogeneous and less aggregated than those extracted in HBexB1. Further purification of t-KD CLPs extracted in P buffer by size-exclusion chromatography (SEC) increased the quality of the images that could be obtained by TEM.
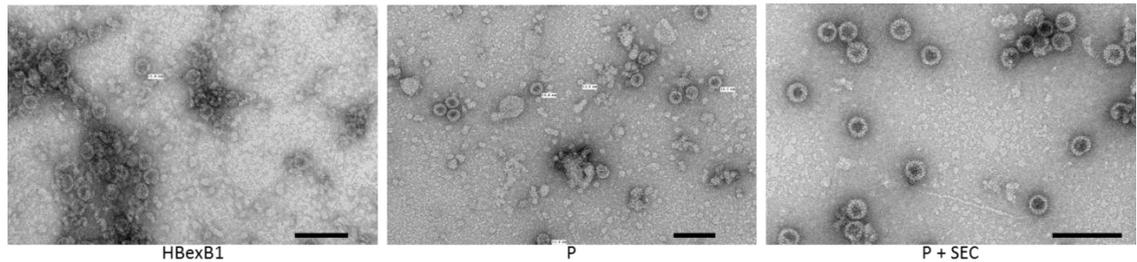
Figure 4.8 Electron micrographs of t-KD CLPs purified in different ways. Left, t-KD CLPs purified by sucrose cushion after extraction in HBexB1 buffer. Middle, t-KD CLPs purified by sucrose cushion after extraction in P buffer. Right, t-KD CLPs purified by sucrose cushion and size-exclusion chromatography after extraction in P buffer. All scale bars, 100 nm.

This success allowed large amounts of leaf material to be used to express t-KD CLPs, which could then be purified and handed over to collaborators in the Prof. Rob Field group (John Innes Centre), and Dr. Nicole Steinmetz (Case Western University, USA), for chemical conjugation and immunogenicity studies. Dr. Matthew Donaldson, of the Prof. Field group, conjugated fluorescein isothiocyanate (FITC) onto the surface of t-KD CLPs, which revealed that chemical conjugation on these particles could be done. Attempts to use t-KD to display antigens from a bacterial pathogen of the *Burkholderia* genus for vaccine development purposes are ongoing. In order to ensure a supply of highly purified particles for future studies, a protocol for the large scale purification of t-KD particles was developed. In this, particles were initially partially purified from plant extract via sucrose cushion (large-scale extraction), and the resulting extracts were dialysed against AmBic buffer, concentrated further via vacuum evaporation (see chapter 2), and then loaded onto a Nycodenz density gradient. After

ultracentrifugation, the gradient was fractionated and the fractions were assayed by western blot for presence of HBcAg (Fig. 4.9a). The two fractions with the strongest signal (5 and 6 in figure 4.9a) were pooled, dialysed against PBS, then diluted tenfold in distilled water and visualised by TEM (Fig. 4.9b). The numerous particles were homogenous in morphology and resembled native HBcAg particles with visible surface spikes and a majority of T=4 CLPs, although a few particles were seen to be misshapen and/or broken. The sample overall seemed very clean, indicating that the Nycodenz gradient is a suitable purification technique for t-KD.

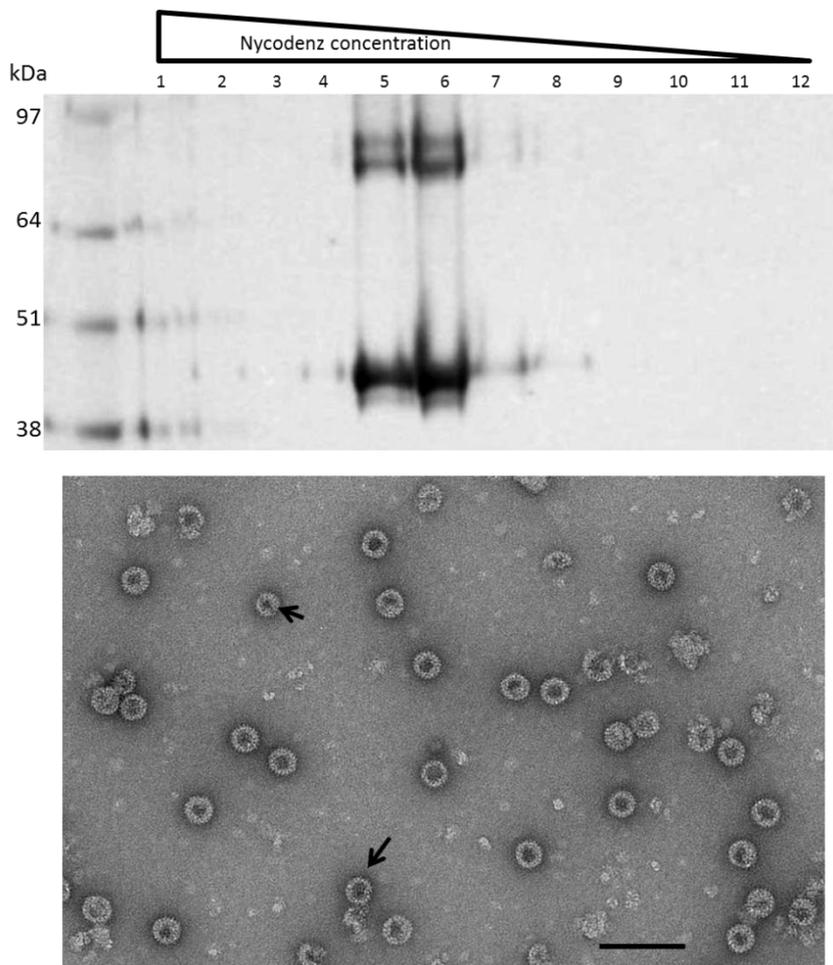

Figure 4.9 t-KD CLPs can be purified by Nycodenz gradient. Top: western blot of the different fractions of the Nycodenz gradient on which t-KD (42.6 kDa) was purified. The antibody used was anti-HBcAg 10E11 monoclonal antibody. Bottom: electron micrograph of t-KD CLPs purified by Nycodenz gradient (fractions 5 and 6 on western blot above). Arrows indicate T=3 particles. Scale bar, 100 nm.

V - Discussion

The results presented in this chapter reveal the possibilities, as well as the limitations, that come with tandem core technology. The t-α-phOx construct directed very low accumulation levels of protein, most of which was insoluble, and none of which could be shown to form CLPs. All attempts at improving solubility, concentration, and purity were unsuccessful. The creation of mosaics is sometimes used to rescue particle formation with chimeric CLPs that contain large or complex inserts. This has been carried out successfully with HBcAg CLPs, although in those cases it was used to improve particle formation of HBcAg with an insert fused to the C-terminus, and not the e1 loop (Koletzki *et al.*, 1999; Beterams *et al.*, 2000; Koletzki *et al.*, 2000). In this instance, this approach could not rescue accumulation of t-α-phOx or the formation of particles. It seems unlikely that the protein would be expressed at very low levels: the HBcAg portion of t-α-phOx is identical in sequence to t-EL, which accumulates at much higher levels, and codon-optimisation of the anti-phOx scFv failed to improve accumulation of t-α-phOx. This strongly suggests that the low level of protein accumulation is caused by rapid turn-over by the plant protein degradation machinery. This in turn suggests that t-α-phOx misfolds, a hypothesis which is supported by the failure to improve t-α-phOx accumulation through the creation of mosaics. If t-α-phOx were capable of folding properly, but could not form particles due to steric constraints, then co-infiltration with t-EL or m-EL should have rescued particle formation, and therefore caused an increase in t-α-phOx accumulation.

Taken together, these results indicate that t-α-phOx probably cannot fold properly, and most of the material is degraded in the plant. Because of this, attempts at improving accumulation are largely futile, because the amino acid sequence of t-α-phOx is not conducive to proper protein folding. This is unlikely to be because of the size of the scFv: it is almost identical in size

to GFP, which can easily be incorporated in the tandem core construct. The more likely explanation is the structure of the scFv. These proteins contain two domains, corresponding to the Light (L) and Heavy (H) chain variable regions of IgG antibodies (Ahmad *et al.*, 2012). This means that the scFv would need to fold into two distinct domains when fused to the tandem core construct despite having both the N- and C-termini anchored to the HBcAg moiety. Moreover, these termini are actually at opposite ends to the scFv, which presents further conformational issues for the presentation of this protein on the surface of a tandem core construct, despite the long linkers used. It seems likely, therefore, that the tandem core system does not provide sufficient flexibility for the scFv to fold into two distinct domains. This improper folding could lead to spurious interactions between scFv residues and HBcAg residues, thus interfering with the correct folding of the rest of the tandem core molecule, a phenomenon that has been hypothesised before with unfolded inserts (Janssens *et al.*, 2010). The result would be a misfolded protein that might well be targeted for degradation.

There was considerably more success, however, with t-KD. This construct led to production of CLPs with similar yields to t-EL and t-sGFP (see chapter 3, about 200-500 μg/g of fresh weight leaf tissue). This is perhaps not surprising given the small size of the insert, and the fact that the charge of the six lysine residues was balanced by six aspartate residues. What is perhaps more surprising is that this insert prevented accumulation of a monomeric HBcAg construct, m-KD, since much larger inserts (both folded and unfolded) have been inserted in the e1 loop of monomeric HBcAg. It could be that the charged residues in the e1 loop of one monomer interfered with those on another, thus preventing dimer formation, and therefore particle assembly. It could also be that the KD linker interacted with the rest of the HBcAg sequence during secondary structure formation, leading to improper folding of the entire protein. It might be possible to determine which of these two hypotheses is correct by using the mosaic system as was done with t-α-phOx, though this was not attempted given the success obtained with the tandem core construct. In any case, the logical conclusion to be drawn is that the

amino acid composition of the insert can play a key role in determining whether it will be successfully displayed in monomeric HBcAg CLPs. Moreover, this experiment has provided evidence that the tandem core system can allow protein accumulation and particle formation when a monomeric core approach fails.

This chapter describes an example of a case (with t-KD) where tandem core technology clearly has an advantage over the use of monomeric core. However, it also provides an example of the limitations of tandem core technology (with t-α-phOx) for the presentation of proteins with multiple domains. This has helped to guide subsequent work and will no doubt serve to guide research into applications of tandem core technology in the future.

Chapter 5 : Introducing the tandibody: a tandem core CLP displaying a functional nanobody on its surface

I - Introduction

Due to the failure to obtain CLPs with a single-chain antibody (scFv), an alternative strategy was developed in which the scFv sequence was replaced with that of a single-domain antibody: a camelid-derived heavy chain variable fragment of a heavy chain antibody (VHH, or nanobody). These synthetic antibody fragments are small (13-15 kDa), have high affinity for their cognate antigens, can function inside the cell, and are composed of only a single domain (Muyldermans and Lauwereys, 1999). They have been expressed heterologously in numerous systems and used for research and biotechnological purposes for many years (Riechmann and Muyldermans, 1999; Jobling *et al.*, 2003; Dolk *et al.*, 2005; Maass *et al.*, 2007; Teh and Kavanagh, 2010). The hypothesis was that a nanobody, being composed of only one domain, would be more likely to fold properly when fused into the C-terminal e1 loop of tandem HBcAg. However, just like an scFv, the N- and C-termini of VHH are exposed at opposite ends of the protein, so it seemed that long linkers (15 amino acids on either end) would be necessary to allow proper folding of the insert. Moreover, success with VHH insertion using this strategy would give further insight into why insertion of an scFv was unsuccessful. Figure 5.1, originally from Harmsen and De Haard (2007), shows how the structure of a nanobody compares to that of an scFv.
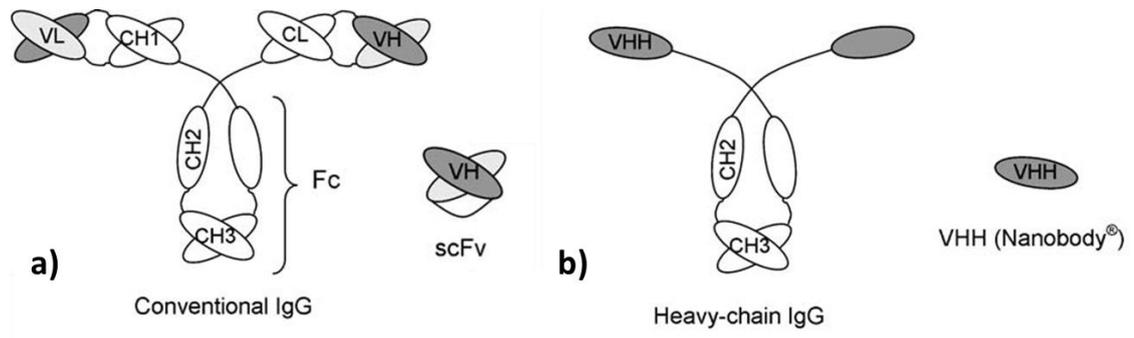
Figure 5.1 The nanobody, a single-domain antibody fragment obtained from a camelid heavy-chain IgG. Conventional single-chain scFv are two-domain constructs obtained from the fusion of the variable regions of the heavy and light chains from standard mammalian IgGs (a). Nanobodies, or VHH, are similar constructs obtained from the variable region of the heavy-chain from heavy chain IgGs of camelids (b). Nanobodies are therefore not only single-chain, but also single-domain antibody fragments. Image reprinted with permission from Harmsen and De Haard, 2007.

II - Presentation of a GFP-specific nanobody

The first nanobody used as proof of principle was the well-characterised 12 kDa "Enhancer" nanobody, originally from dromedary (Kirchhofer *et al.*, 2010). This nanobody has sub-nanomolar affinity for green fluorescent protein (GFP), and the amino acid sequence for it is publicly available on GenBank (PDB: 3K1K_C). The DNA sequence for this nanobody was ordered for synthesis from GeneArt, (Life Technologies) with stop codon removed, codon usage optimised for *N. benthamiana*, and long linkers (GGGGS)$_3$ on either side between the SalI and AseI restriction sites. This sequence was inserted into pEAQ-*HT*-t-EL between the SalI and AseI restriction sites so as to insert it into the C-terminal e1 loop of t-EL (see chapter 3). The new construct, pEAQ-*HT*-τ-GFP, is intended to direct the production of a tandem HBcAg protein displaying the anti-GFP nanobody: this fusion protein was named "anti-GFP tandibody" (abbreviated to τ-GFP), with the word "tandibody" being a contraction of "tandem core HBcAg displaying a nanobody" (Fig. 5.2).
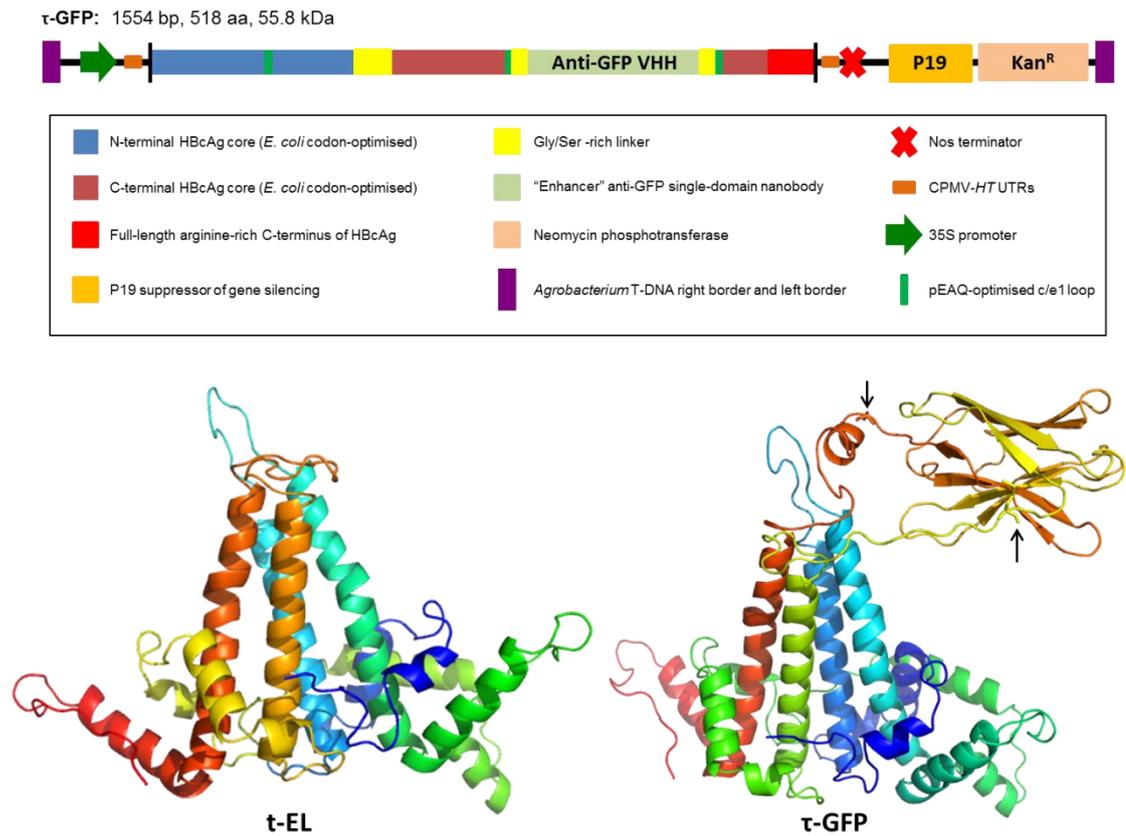
Figure 5.2 Diagram of construct and protein structures. Top: construct map for τ-GFP, the tandem core construct with the anti-GFP Enhancer nanobody inserted in the C-terminal e1 loop. The key is the same as in figure 3.3a. Bottom: structure predictions of t-EL (left) and τ-GFP (right), courtesy of Dr. Ellis O'Neil. The N-termini are in blue, and the C-termini in red. Arrows indicate the first and last amino acid of the anti-GFP VHH, which are represented in stick form. Images were generated using PyMOL.

As a control, the same VHH sequence was also inserted into the c/e1 loop of the monomeric construct, pEAQ-m-EL (see chapter 3) to give pEAQ-μ-GFP. The two constructs, together with a pEAQ-*HT* empty vector control were separately infiltrated into *N. benthamiana* leaves. Small scale extractions in P buffer were prepared 7 days post-infiltration (dpi) and the soluble fractions were subjected to western blot analysis using anti-HBcAg monoclonal antibody 10E11 (Fig. 5.3a). The results showed the accumulation of HBcAg-specific material of the expected size (55 kDa) in leaves infiltrated with pEAQ-τ-GFP (3rd lane), while no protein corresponding to the expected size (37 kDa) could be detected in extracts from leaves infiltrated with pEAQ-μ-GFP (4th lane). All samples, including those from leaves infiltrated with the empty vector, pEAQ-*HT*, showed a cross-reactive band of unknown origin at approximately 39 kDa, as was

seen in some of the other western blots in which this antibody was used (see chapter 4). These results demonstrated that a tandem construct, as opposed to a monomeric one, is required for the efficient display of VHH sequences. It should be noted, however, that when produced on its own, most of the τ-GFP is insoluble.

As a preliminary assay of whether the expressed τ-GFP retained the ability of the nanobody to bind GFP, a co-infiltration experiment was carried out, in which *Agrobacterium* cultures carrying pEAQ-*HT* containing the sequence for either GFP or τ-GFP were co-infiltrated. These leaves were treated as above but the insoluble protein fraction (pellet) was also analysed by western blot (Fig. 5.3b). This revealed that when produced on its own, τ-GFP is mainly insoluble, but when τ-GFP and GFP are co-infiltrated, most of the τ-GFP can be found in the soluble fraction. This suggests that the presence of GFP *in planta* increases the solubility of τ-GFP. This strongly hints that τ-GFP binds to GFP, and the entire complex is more soluble than τ-GFP on its own.
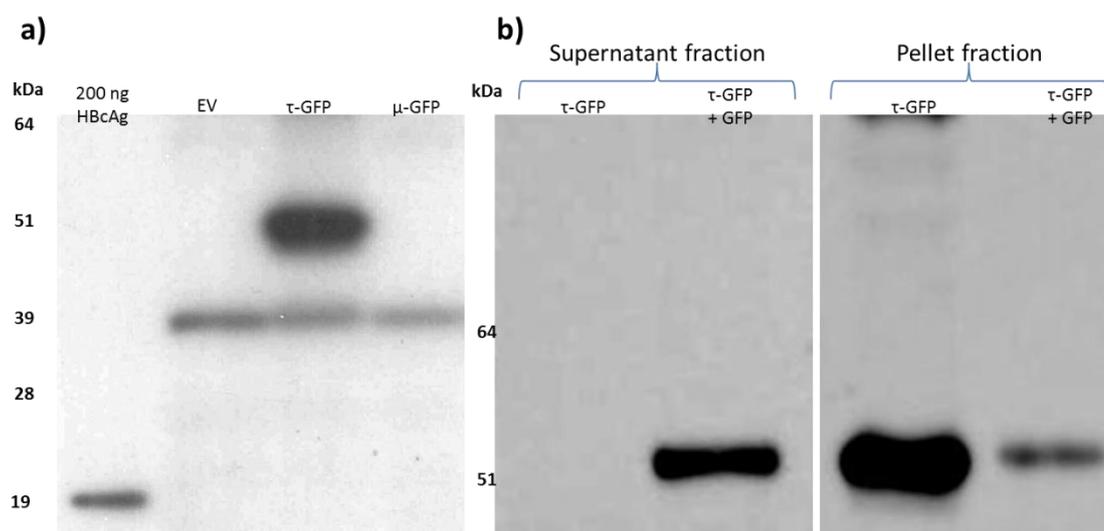
Figure 5.3 τ-GFP is expressed in plants, and its solubility is enhanced by the co-expression of GFP. a) Western blot of soluble (supernatant) plant extracts expressing different constructs. EV: empty vector control; τ-GFP: tandem core construct with anti-GFP nanobody inserted in the C-terminal e1 loop (55.8 kDa); μ-GFP: monomeric core construct with anti-GFP nanobody inserted in the e1 loop (36.5 kDa). b) Western blot of supernatant (left) and resuspended pellet (right) fractions of plant extract expressing τ-GFP alone or co-expressed with GFP.

Sucrose cushion and transmission electron microscopy (TEM) was used to determine whether τ-GFP forms particles when expressed in plants and to confirm whether the presence of GFP affects the method of purification. Large-scale extractions (see chapter 2) were carried out with leaves infiltrated with pEAQ-τ-GFP alone or co-infiltrated with pEAQ-*HT*-GFP were homogenised and the clarified lysates were loaded on to sucrose cushions consisting of two layers of sucrose (25 % and 70 %). Particulate matter of the size of assembled HBcAg particles was expected to produce a band at the interface between the two sucrose layers: intense fluorescence was seen at this interface for the tubes containing the co-expressed material of both GFP and τ-GFP. This interface was collected (for both sets of samples) and any particles within it further purified by size exclusion chromatography. The fractions from this were analysed by western blot and positive fractions were pooled, concentrated by vacuum evaporation (see chapter 2), and examined by TEM. This revealed CLPs in both samples (with and without GFP), which look extremely similar to wild-type or tandem HBcAg, but with an apparent thin layer on the surface in addition to, and partially obscuring, the characteristic spikes (Fig. 5.4). This gives them a morphology which I term "knobbly", as opposed to the wild-type "spiky" morphology. It is more difficult to distinguish T=3 particles from T=4 particles when the CLPs are knobbly, but there do appear to be a small minority of T=3 particles visible in the second image (arrows), which is typical of HBcAg. As the images show, more particles were obtained from the co-infiltration than from the infiltration of τ-GFP on its own, thus further reinforcing the previous data which suggested that the presence of GFP increases the solubility of the particles, leading to higher recovered yield. It is estimated that about 5-20 μg of τ-GFP can be obtained per gram of fresh weight leaf tissue [FWT] without GFP, and about 100-500 μg/g FWT with GFP.

Without GFP   GFP co-infiltration
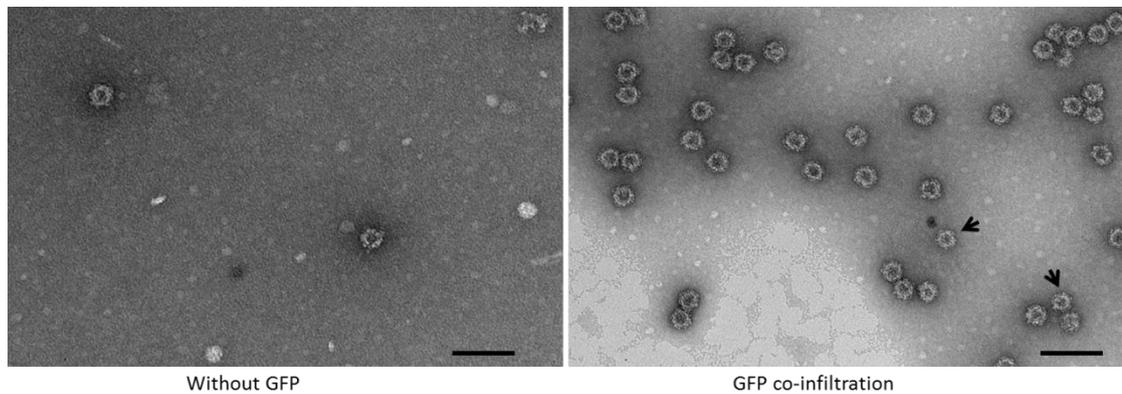
Figure 5.4 Electron micrographs of τ-GFP CLPs obtained from expression alone or with GFP. Left: τ-GFP CLPs obtained from leaves agroinfiltrated with τ-GFP alone. Right: τ-GFP CLPs obtained from leaves that were co-infiltrated with τ-GFP and GFP. Higher yields of CLPs were obtained from co-expression. Both sets of particles present "knobbly" morphology, whereby the characteristic spikes of wild-type HBcAg CLPs are less clearly defined and apparently obscured. This seems to be particularly true of the particles obtained from co-infiltration. The arrows indicate T=3 particles, both scale bars are 100 nm.

To further examine the ability of τ-GFP particles to bind GFP, a similar sucrose cushion experiment was carried out, this time with more controls and no co-infiltrations. Plant tissue expressing the τ-GFP (or t-EL) particles was mixed with plant tissue expressing GFP (or empty-vector control), and the leaves were homogenised together before the rest of the large-scale extraction and sucrose cushion were carried out as described in chapter 2 (see figure 5.5).
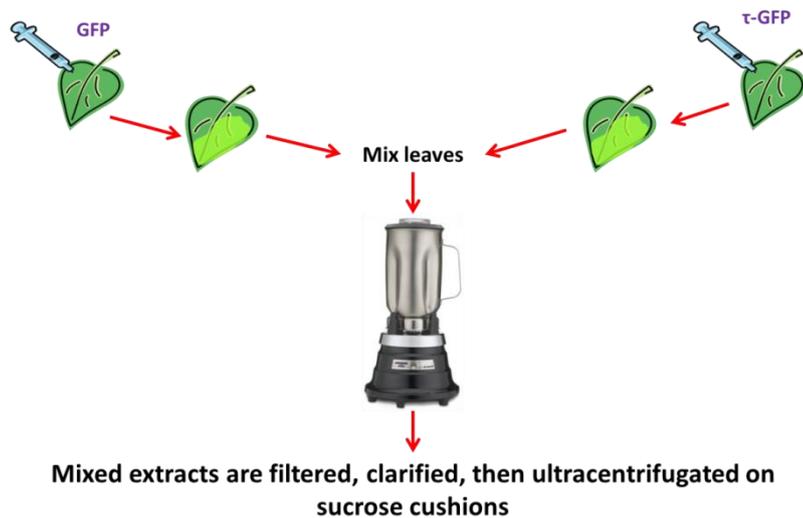
Figure 5.5 Experimental set-up to examine binding of τ-GFP to GFP. Leaves were separately agroinfiltrated with bacteria containing constructs coding for either GFP or τ-GFP. At 7 dpi, the leaves were mixed and homogenised together as part of a large-scale extraction followed by a sucrose cushion.

After ultracentrifugation of material containing both GFP and τ-GFP – expressing plant extract, the fluorescence was found to have migrated at the bottom of the sucrose cushion, co-localising with the τ-GFP particles (Fig. 5.6a). When GFP-containing plant extract was run on an identical sucrose cushion on its own or with t-EL-containing plant extract, the fluorescence was found to remain in the supernatant, indicating that it is not large enough to migrate into the sucrose layers. The co-localisation of GFP with τ-GFP in the sucrose layers was demonstrated by western blot analysis of the different fractions of the sucrose gradients using either anti-HBcAg or anti-GFP antibodies (Fig. 5.6b). This showed that while tandem HBcAg (with or without the anti-GFP nanobody) always migrates to the bottom of the cushion GFP mostly stays in the supernatant unless it is in the presence of τ-GFP. This indicates that the τ-GFP particles interact with GFP in such a way as to confer the properties of a virus-like particle. This strongly suggests that GFP binds to the nanobody moiety of assembled τ-GFP particles.
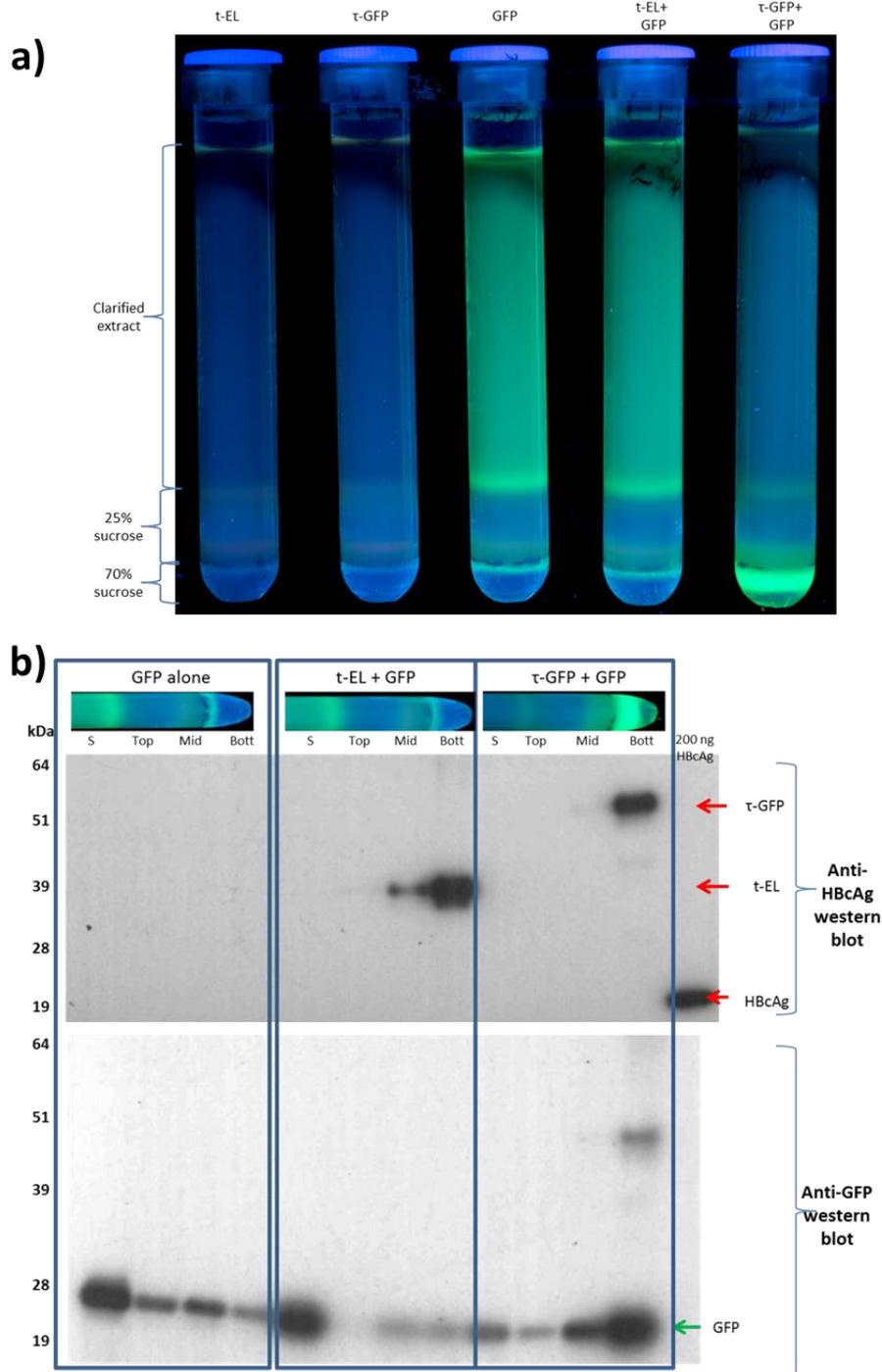
Figure 5.6 τ-GFP CLPs bind to GFP in a sucrose cushion. Plant leaves expressing different constructs were mixed as shown in figure 5.5. After large-scale extraction, sucrose cushions were loaded with the different plant extracts, and these cushions were photographed under UV light after ultracentrifugation (a). While plant extract containing only t-EL or τ-GFP do not show fluorescence, plant extracts containing GFP alone or with t-EL show fluorescence in the supernatant, above the sucrose layers. When plant extract containing GFP is mixed with extract containing τ-GFP, the fluorescence is located at the bottom of the sucrose cushion, indicating that τ-GFP CLPs have dragged GFP down through the sucrose. (b) Duplicate anti-HBcAg (top) and anti-GFP (bottom) western blots of the different fractions of the sucrose cushions above. GFP (27 kDa) is mostly confined to the supernatant when alone or mixed with t-EL (41.8 kDa), but it co-localises with τ-GFP (55.8 kDa) at the bottom of the sucrose cushion. S: supernatant, Top: top of the sucrose cushion, Mid: interface between the 25% and 70% sucrose fractions, Bott: bottom of the sucrose cushion.

The importance of the extraction buffer was tested in an attempt to increase recovered yield of particles. Leaf tissue co-infiltrated with pEAQ-*HT*-τ-GFP and pEAQ-*HT*-GFP were split into four groups and each group was homogenised in one of the following buffers: P buffer (0.1 M sodium phosphate, pH 7); P+ (20 mM sodium phosphate, pH 7, 1 mM EDTA, 5 mM DTT, 0.1% w/v Triton X-100); IQUR (20 mM Tris-HCl, pH 8, 1 mM EDTA, 5% isopropanol, 5 mM DTT, 0.1% w/v Triton X-100); or HBexB1 (10 mM Tris-HCl pH 8, 120 mM NaCl, 1mM EDTA,1 mM DTT, 0.75% w/v sodium deoxycholate). All buffers were also supplemented with protease inhibitor (Roche). The lysates were clarified and filtered over 0.45 µm syringe filters before being loaded onto sucrose cushions. Observation of the ultracentrifuge tubes under UV light after ultracentrifugation revealed that in all four cases, GFP sedimented in the sucrose layers; although the standard extraction buffer and P+ caused the GFP (and therefore presumably the τ-GFP) to be in a relatively distinct band at the interface between the two sucrose layers, as opposed to HBexB1 and IQUR buffers (Fig. 5.7a). It was also clear that the standard extraction buffer and HBexB1 had clean interfaces, unlike P+ and IQUR buffers, which gave rise to thick green contamination at the interface region. The fluorescent interfaces were collected and, after dialysis and further concentration and clarification, visualised by TEM (Fig. 5.7b). While no structurally sound particles were seen in the IQUR-extracted sample, numerous particles were seen from standard and P+ extractions, and fewer particles were seen in the HBexB1-extracted sample. This is particularly interesting given that extractions of τ-GFP alone (without GFP) in HBexB1 was also attempted but no particles were seen with TEM. It could be that the increased solubility of the particles in the presence of GFP means that there are simply more particles which are more likely to be seen as compared to τ-GFP on its own. However, it could also indicate that the presence of GFP makes the particles more resistant to the HBexB1 buffer by altering the surface properties of the particles.
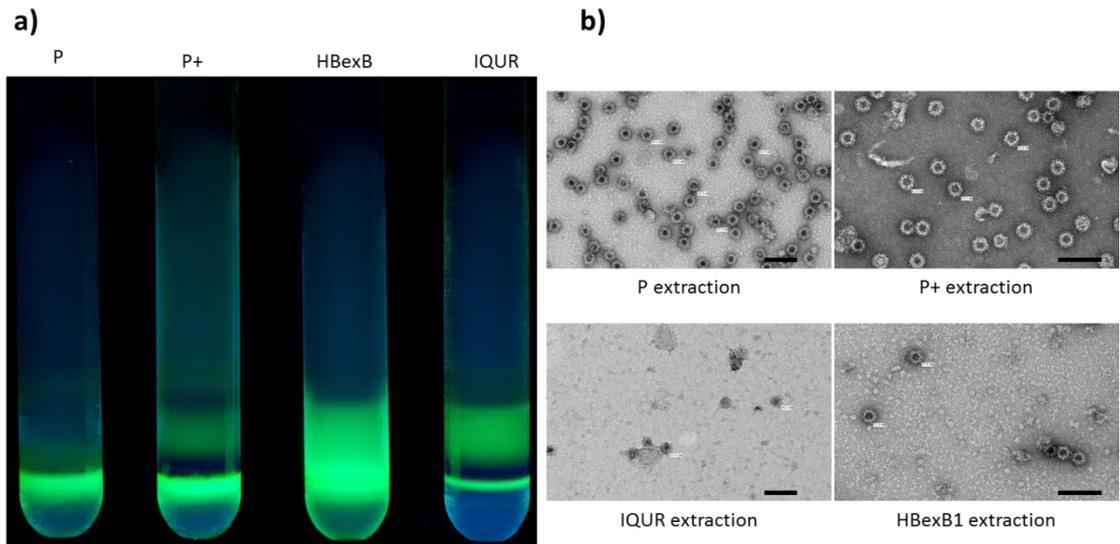
Figure 5.7 P buffer is the best buffer to extract τ-GFP CLPs that have been co-expressed with GFP. a) UV-imaged sucrose cushions of plant extracts from leaves co-infiltrated with τ-GFP and GFP, extracted in four different extraction buffers. The extract obtained with buffer P yields the least diffuse fluorescent band with the lowest amount of dark green contaminating plant material. b) The fluorescent fractions from the cushions were visualised by TEM. CLPs with knobbly phenotype were seen in fractions obtained from material extracted in P, P+, and HBexB1 buffers, but not IQUR buffer. Scale bars, 100 nm.

More detailed analysis of the binding of GFP to τ-GFP required purified preparations of the two components: GFP was obtained commercially (Millipore), while τ-GFP was produced in plant tissue on its own via a large-scale extraction followed by sucrose cushion purification. Because of solubility issues associated with producing τ-GFP without GFP, it was time-consuming to obtain sufficient τ-GFP to carry out the next experiment: analysing the binding of τ-GFP particles to GFP through a modified sandwich ELISA (Fig. 5.8a). Briefly, wells of a 96-well plate were coated with the τ-GFP particles (purified by sucrose cushion and syringe-filtration as described in chapter 2), and different amounts of GFP were added to these wells after blocking, before GFP was detected by a commercially available anti-GFP polyclonal antibody conjugated to horseradish peroxidase (HRP). As a negative control, wells were coated with the same amount of a different (but identically purified) tandibody particle, which displays a nanobody against HIV-1 surface glycoprotein gp120 (τ-gp120, see Chapter 6). The positive control was a commercially available anti-GFP polyclonal antibody different to that used for

detection and already conjugated to horseradish peroxidase (HRP). The negative and positive controls were treated identically to the experimental wells, with the sole exception of the protein used to coat the wells. Upon detection of HRP activity, it was found that the signal from wells coated with τ-GFP was very similar to the signal from wells coated with anti-GFP IgG, with signal decreasing concomitantly with the decreasing amount of GFP used (Fig. 5.8b). In contrast to this, the negative control showed no signal above background, indicating that the interaction between τ-GFP and GFP is caused specifically by the anti-GFP nanobody moiety.
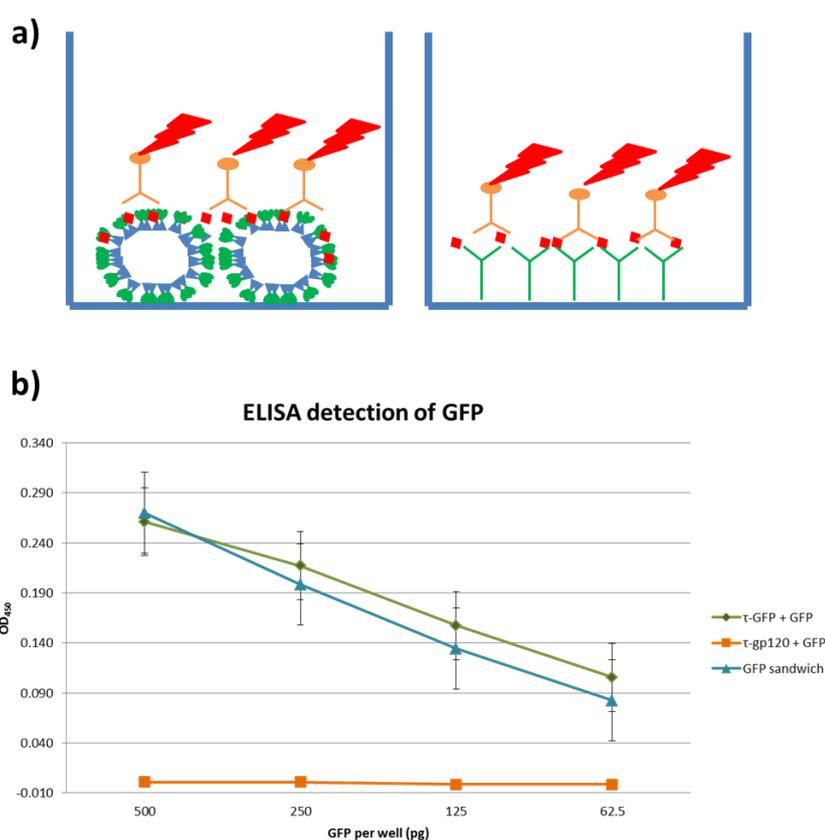
Figure 5.8 τ-GFP CLPs bind to GFP in an ELISA. Partially-purified τ-GFP CLPs (circular particles on the left) were used to coat the wells of a 96-well plate in order to act like anti-GFP antibodies in a sandwich ELISA (a). Pure GFP (red diamonds) was added at a range of concentrations to the wells and the GFP was detected with a polyclonal anti-GFP IgG conjugated to HRP (orange). As a positive control, a commercially obtained polyclonal anti-GFP IgG different to that used for detection was used (green, on the right) as part of a standard sandwich ELISA. As a negative control, a similarly-purified tandibody bearing a nanobody against HIV-1 gp120 (τ-gp120, see chapter 6) was used in the same amounts as τ-GFP. b) Absorbance measurements indicated the amount of GFP present in each well. Treatments were carried out in quadruplicates and average readings were subtracted from background to give net signal. While the τ-gp120 CLPs did not bind GFP, τ-GFP CLPs bound GFP in a similar fashion to the anti-GFP IgG. Error bars represent standard error.

In an attempt to quantify the strength of binding of τ-GFP to GFP, Surface Plasmon Resonance (SPR) was used. In this case, commercially purchased His-tagged GFP (Millipore, the same as was used in the ELISA experiment) was fixed onto a Ni-NTA SPR chip, and purified tandibody particles (either τ-GFP or τ-gp120 control) at identical concentrations were flowed over the GFP, and the interaction between the two was measured using a Biacore T200. The first SPR experiment was a simple binding test, to check for any interaction between τ-GFP or τ-gp120 and GFP (Fig. 5.9a). It revealed that while τ-gp120 did not interact with GFP, τ-GFP interacted strongly. It further indicated that there was very little dissociation of τ-GFP from GFP, or that this dissociation is very slow: this is reflected in a slow decrease in response units (RU) over time after the end of tandibody injection, which is normally indicative of dissociation between the two molecules. However, this apparent dissociation is paralleled by the decrease in RU of τ-gp120, which decreases to negative values (with 0 being the start of the flow of tandibody), indicating that this apparent dissociation in both cases is in fact dissociation of his-tagged GFP from the chip, as opposed to dissociation of the tandibody from the GFP. In this light, it is very clear that τ-GFP does not significantly dissociate from GFP during the period measured.

The second SPR experiment was a single-cycle kinetics experiment, in which increasing amounts of τ-GFP were flowed over the GFP in order to reach saturation, thus allowing the quantification of affinity of τ-GFP for GFP (Fig. 5.9b). The result was a calculated K_D of 0.175 nM, which is, in fact, lower than the published K_D of the Enhancer nanobody, which is given as 0.59 nM (Kirchhofer *et al.*, 2010). However, it should be noted that dissociation of GFP from the chip would skew the calculation of the dissociation constant in the SPR experiment. Moreover, this calculation is dependent on the concentrations of GFP and of the tandibodies, so the K_D is only as reliable as the quantification of the tandibodies used in the experiment. Because the tandibodies were quantified using western blots, and because the tandibodies

always show a diffuse ladder pattern on SDS-PAGE gels, the quantification of tandibodies is more of an estimate than an exact measurement. However, the relative quantification between the two tandibodies should be fairly accurate. So while the exact K_D is difficult to measure accurately, the difference seen between τ-GFP and τ-gp120 should be representative.
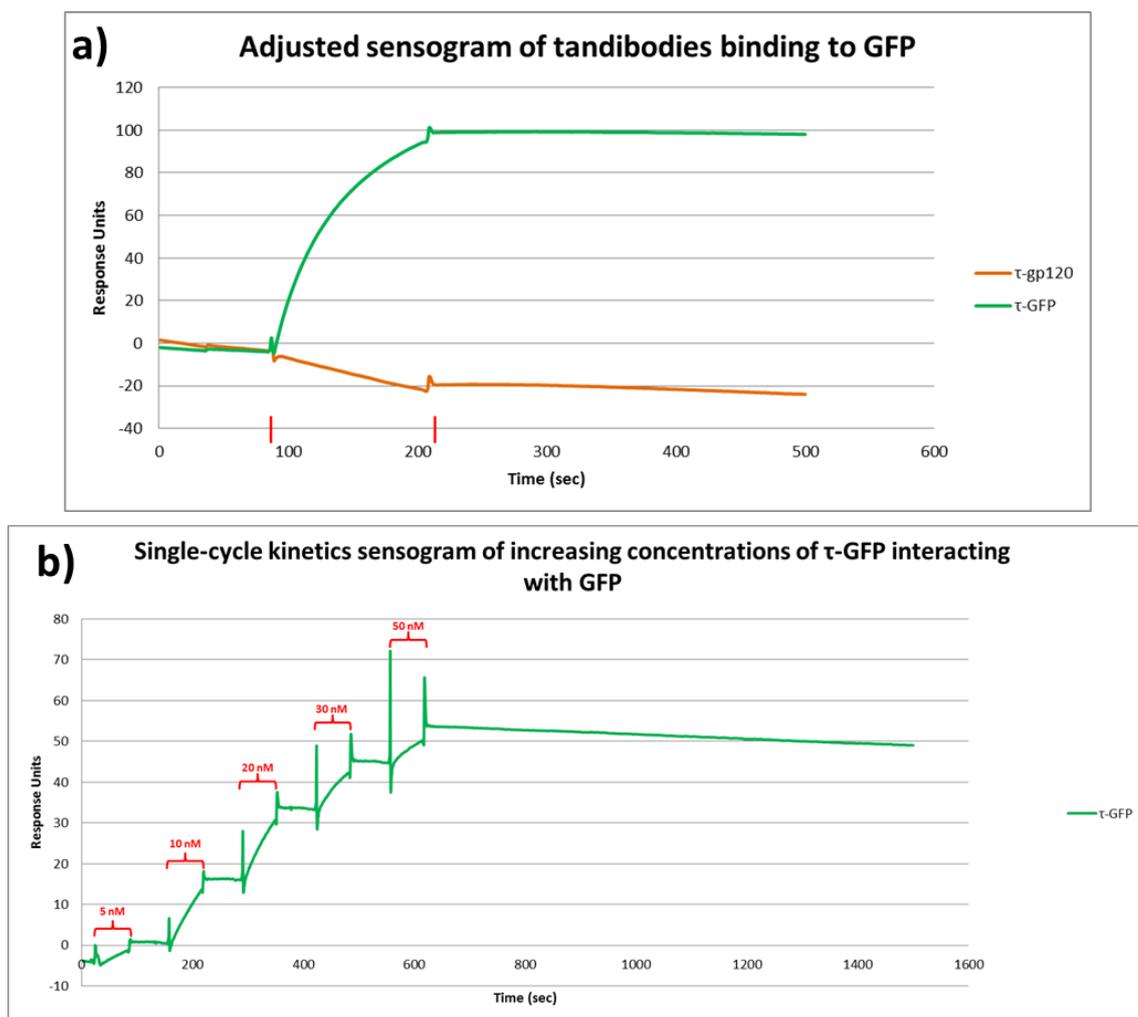
Figure 5.9 τ-GFP CLPs bind to GFP during surface plasmon resonance. Commercially obtained His-tagged GFP was bound to the surface of a Ni-NTA SPR chip, and partially purified τ-GFP CLPs were flowed over the GFP. Binding resulted in detection of Response Units. a) In a simple binding test, τ-GFP (green line) was seen to bind to GFP during tandibody injection (time between red lines) and to stay bound afterwards, while τ-gp120 (orange line) injected at the same concentration did not bind to GFP. b) A single-cycle kinetics binding experiment involved flowing increasing concentrations (indicated in red) of τ-GFP over the surface-bound GFP until saturation was reached. This allowed the quantification of τ-GFP affinity for GFP, which was calculated to be 0.175 nM.

The τ-GFP tandibody bound to GFP was also used for preliminary cryo-electron microscopy (cryo-EM). Plant material co-expressing τ-GFP and GFP was purified by sucrose cushion and size-exclusion chromatography, and the purified particles were flash-frozen in vitreous ice on a TEM grid for image collection. Visualisation of the particles confirmed the results seen with negative-stain TEM: the τ-GFP particles are intact, proving that the insertion of a nanobody sequenced has not abolished capsid assembly. The data obtained from cryo-EM image collection was analysed in collaboration with Dr. Robert Gilbert (University of Oxford) and class averages were obtained to generate a 25 Å reconstruction of the particles (Fig. 5.10). At this resolution, the map is more of a schematic of average electron density within the particles.

The particle reconstruction clearly contains spikes reminiscent of those of a standard HBcAg CLP, but long projections of extra density are formed on the outside of these particles. These projections are too few to each correspond to a nanobody-GFP complex on a single spike, since there should be one per spike, and therefore as many projections as there are spikes. What is most likely is that these projections are average positions of averaged densities of numerous nanobody-GFP complexes from a few different spikes. This is logical when one considers the length of the linkers used to fuse the anti-GFP nanobody to the spikes of the tandem core. The long linkers were intended to provide the nanobody with flexibility, and this flexibility probably means that the nanobody-GFP complex is not in a fixed orientation on the surface of the particle with respect to its tandem core partner. This makes structure reconstruction complicated, since it relies on average electron densities to determine molecular conformation and positioning. When the GFP-τ-GFP reconstruction is overlain on the preliminary map obtained for CoHe (hetero-tandem core with no insert in the e1 loop), the spikes from both maps overlay, indicating that the basic tandem core particles are similar in structure, but the GFP-τ-GFP projections are not present on the CoHe map, which indicates that these are extra density on the exterior of the CLP which is not present in CoHe, thus reinforcing the likelihood that these projections correspond to the nanobody-GFP complex.

a)

b)
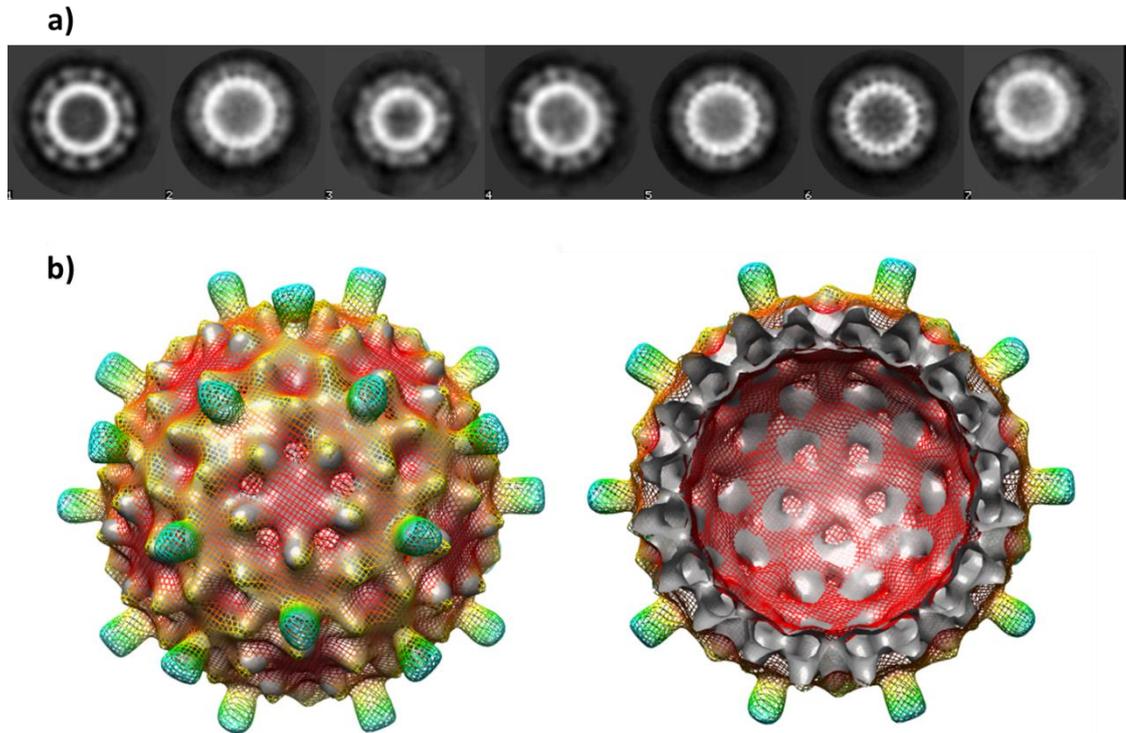
Figure 5.10 Cryo-electron microscopy of τ-GFP CLPs bound to GFP. 2,385 CLPs of τ-GFP bound to GFP were imaged and analysed with the assistance of Dr. Robert Gilbert. a) Class averages of these CLPs. b) This allowed the generation of a 3D reconstruction (resolution estimated at 25 Å) coloured by distance from the centre of the particle (red to blue). The map is shown viewed down a 5-fold axis with the bacterially-produced CoHe reconstruction on which the construct was based fitted within (grey surface). The projecting spikes represent electron density corresponding to the bound nanobody and GFP, but these do not occupy every position expected, instead appearing as an average of the density present, with the highest intensity at the 2-fold and 5-fold axes. The location of these masses may be due to the icosahedral symmetry imposed on the reconstruction. This may indicate that these projections are flexible and are not displayed in a fixed orientation on each spike.

III - Discussion

Nanobodies (or VHH) have been expressed in several expression systems including bacteria for

phage display (Dolk *et al.*, 2005; Maass *et al.*, 2007; Beekwilder *et al.*, 2008), and plants for

VHH overexpression (Ismaili *et al.*, 2007; Winichayakul *et al.*, 2009; Teh and Kavanagh, 2010).

They have found uses in fundamental antibody research (van der Linden *et al.*, 1999; Desmyter

et al., 2001; Muyldermans *et al.*, 2001) as well as applied biotechnology (Jobling *et al.*, 2003).

When attempts to display a scFv on the surface of a tandem core CLP proved unsuccessful (see

the previous chapter), presentation of a nanobody was tried as an alternative, in the hope that

the simpler structure would allow proper protein folding and CLP assembly. This proved to be highly successful -where scFv failed, VHH prevailed! The GFP-binding Enhancer nanobody (Kirchhofer *et al.*, 2010) proved to be an excellent tool for demonstrating that nanobodies displayed on the surface of HBcAg retain their binding activity. It was demonstrated that τ-GFP can be expressed in plants both with and without GFP, although co-expression with GFP increased the solubility of the particles, resulting in much higher yields of recovered protein. The particulate nature of τ-GFP with and without GFP was demonstrated with TEM as well as cryo-EM, making this the first report of an antibody fragment successfully displayed on HBcAg CLPs in any expression system, and the first time an antibody fragment of any kind has been displayed on a spherical virus capsid in a plant expression system. The only other example of antibody fragments being displayed on the surface of a virus particle in plants is the display of scFv on the surface of the filamentous PVX (Smolenska *et al.*, 1998).

The binding of GFP to τ-GFP could be readily demonstrated using a simple sucrose cushion experiment, and further analysed by ELISA and SPR. All of these techniques gave concordant results which indicate that τ-GFP CLPs bind GFP with very high affinity. This demonstrates that the fusion of a nanobody to tandem HBcAg does not have an adverse effect on the affinity of the nanobody for its target antigen. Moreover, these results show that a protein with the N- and C-termini on opposite ends can be successfully inserted in the e1 loop of HBcAg provided that a tandem construct is used: the fact that the monomeric μ-GFP construct failed to yield any detectable material demonstrates that tandem core technology can overcome the limits to insertions in the HBcAg e1 loop found with monomeric constructs. Furthermore, the data presented here offer some insight into why the attempted display of the anti-phOX scFv failed (see chapter 4). The anti-GFP nanobody also has N- and C-termini exposed on opposite ends of the protein, and yet 15 amino acid – long linkers on either side were sufficient to allow proper folding of the insert and the HBcAg scaffold, so it seems unlikely that this was the limiting factor in display of the anti-phOx scFv. The only other differences between the scFv and the

VHH are the size and the number of domains. Because the scFv is the same size as GFP, which could be displayed on the surface of a tandem core CLP (see chapter 3), the most likely limiting factor for particle formation with t-α-phOx is the number of domains on the insert. This seems to be the crucial difference between a nanobody and an scFv which determines whether the insert will allow particle formation in a tandem core construct.

The successful tandibody proof of concept paved the way for the next set of experiments: using the same methodology to create tandibodies with binding affinity for medically relevant antigens. Using GenBank as a source of available nanobody sequences, two nanobodies were chosen for the next sets of experiments: one against HIV-1 surface glycoprotein gp120, and the other against HIV-1 capsid protein p24. The results of these experiments are presented in the following chapter.

Chapter 6 : Adapting tandibody technology to display antigens of medical relevance

I - Introduction

Success with the anti-GFP tandibody opened up the possibility of displaying nanobodies on the surface of tandem HBcAg which recognise medically relevant antigens. A search on GenBank revealed the sequences for two nanobodies which recognise HIV proteins.

The first, known as A12, is described in Forsman *et al.* (2008) and Strokappe *et al.* (2012). It is a llama-derived nanobody specific to HIV-1 IIIB surface glycoprotein extracellular region gp120. HIV gp120 is part of the gp160 protein (coded for by the ENV gene), which during post-translational processing is split into gp120 (the surface-exposed glycoprotein) and gp41, which contains the transmembrane and cytoplasmic domains, along with the fusion peptide and heptad repeats which are involved in trimerisation of the gp160 complex on the surface of the viral envelope (Chakrabarti *et al.*, 2002). The A12 nanobody has been determined to bind to gp120 with a K_D of 0.1 nM. There were two reasons for choosing this nanobody to test the tandibody system: the sequence of the A12 nanobody is available on GenBank (PBD: 3R0M_A), and this nanobody was discovered by panning llama serum against an easily available antigen, recombinant gp120 from HIV-1 IIIB, obtained from CFAR, NIBSC (catalogue number EVA607). This meant that I could test binding of this nanobody with an antigen that is known to interact strongly, without introducing extraneous variables such as sequence variability in gp120 between different HIV-1 subtypes. Along with these technical considerations, I considered that it would be interesting to test the tandibody concept with a glycoprotein antigen. Indeed, the tandibody technique, because it is based on non-covalent binding of a scaffold to an antigen of interest, would, in theory, be very useful for displaying surface antigens of enveloped viruses. For these proteins to fold properly, they normally need to be trafficked through the eukaryotic

secretory pathway, which could make it difficult for them to fold properly if fused directly to the e1 loop of HBcAg or tandem HBcAg. Post-translational non-covalent linking of the antigen to the CLP scaffold has the advantage of allowing the production of each component separately in their respective optimal cellular compartment.

The second nanobody that was chosen was an alpaca-derived nanobody which binds to HIV-1 capsid protein p24 (GenBank PDB: 2XT1_B). Unlike the anti-GFP and anti-gp120 nanobodies, this nanobody is not described in the literature, but there is evidence that it binds to p24: the RCSB Protein Data Bank file linked to the GenBank sequence is, in fact, the crystal structure of this nanobody (produced in *E. coli*) in complex with the C-terminus of HIV-1 p24 (also produced in *E. coli*). It seems unlikely that such a structure could have been obtained if the nanobody did not bind to this antigen. An advantage of using this nanobody in the tandibody system is that, like gp120, the p24 protein is available to purchase as a reagent from the NIBSC (CFAR repository reference ARP678). Moreover, there was evidence that recombinant p24 could be produced in plants: HIV-1 p24 was produced transiently in *N. benthamiana* using a TMV-based vector, and the resulting p24 proved to be antigenic in animal studies (Pérez-Filgueira *et al.*, 2004). This anti-p24 nanobody was therefore selected for study in the tandibody system alongside the anti-gp120 nanobody.

II - Design of constructs

A) τ-gp120

The anti-gp120 nanobody sequence known as A12 (Forsman *et al.*, 2008) was obtained from GenBank (PBD: 3R0M_A) and synthesised by GeneArt (Life Technologies) with the stop codon removed, codon usage optimised for *N. benthamiana*, and long linkers (GGGGS)$_3$ on either side between SalI and AseI restriction sites. Between the nanobody sequence and the linkers, an AvrII site (on the N-terminal side, corresponding to amino acids Pro-Arg) and a SbfI restriction site (on the C-terminus, corresponding to amino acids Pro-Ala) were added. This construct was

inserted into pEAQ-*HT*-t-EL between the SalI and AseI restriction sites to give pEAQ-*HT*-τ-gp120 (Fig. 6.1).

B) gp120

A gp120 construct was prepared for co-expression alongside τ-gp120. The sequence for gp120 (from HIV-1 IIIB) was acquired from NIBSC, and it was found to be identical to a GenBank sequence (UniProtKB/Swiss-Prot: P03377.1). For expression in plants, a 6XHis tag was added at the C-terminus, and the N-terminal 30 amino acids were replaced with the *Arabidopsis* basic chitinase signal peptide (Samac *et al.*, 1990), which had previously been used by Sainsbury *et al.* (2008) as a strong secretion signal for recombinant protein production in plants. The native secretion signal was not used because it has been reported that it does not yield optimal secretion of gp120 in heterologous systems (Golden *et al.*, 1998). This modified gp120 sequence was codon optimised for *N. benthamiana* and ordered for synthesis. The sequence (including the 6XHis tag) was cloned into pEAQ-*HT* using the AgeI and XhoI restriction sites to give pEAQ-*HT*-gp120.

C) τ-p24

The anti-p24 nanobody sequence was obtained from GenBank (GenBank PDB: 2XT1_B) and ordered for synthesis with stop codon removed, codon usage optimised for *N. benthamiana*, and restriction sites AvrII and SbfI on the 5' and 3' ends respectively. This construct was inserted into pEAQ-*HT*- τ-gp120 between the AvrII and SbfI restriction sites to give pEAQ-*HT*-τ-p24 (Fig. 6.1).

D) p24

A p24 construct was also prepared for co-expression alongside τ-p24. The sequence for HIV-1 p24 was obtained from GenBank entry 2XT1_A, which is part of the same entry as τ-p24: this entry describes the complex of τ-p24 binding to the C-terminus of p24, so the sequences of both are given. However, because this entry only provides the sequence for the C-terminus of p24, the N-terminal part of p24 was obtained from GenBank entry AAB50258.1, which is the sequence of the entire gag polyprotein of HIV-1. So the p24 sequence to be used was taken from accession number AAB50258.1 (from position P149 to Y277) and completed with the sequence from accession number 2XT1_A (from position S1 to L86). This complete p24 sequence was codon-optimised for expression in *N. benthamiana* and ordered for synthesis. This sequence was cloned into the pEAQ-*HT* vector using the AgeI and XhoI restriction sites to give pEAQ-*HT*-p24, and with the AgeI and XhoI sites to give the C-terminally 6XHis-tagged version of p24, pEAQ-*HT*-p24His.

a)

τ-gp120: 1599 bp, 533 aa, 57.3 kDa

τ-p24: 1560 bp, 520 aa, 55.7 kDa

■ N-terminal HBcAg core (*E. coli* codon-optimised)	■ Gly/Ser -rich linker	✖ Nos terminator
■ C-terminal HBcAg core (*E. coli* codon-optimised)	■ "Anti-p24 single-domain nanobody	■ CPMV-*HT* UTRs
■ Full-length arginine-rich C-terminus of HBcAg	■ Neomycin phosphotransferase	➤ 35S promoter
■ P19 suppressor of gene silencing	■ *Agrobacterium* T-DNA right border and left border	▮ pEAQ-optimised c/e1 loop
■ "A12" anti-gp120 single-domain nanobody		

b)

Figure 6.1 Diagram of constructs and protein structures. a) Construct maps for the τ-gp120 and τ-p24 tandibodies. The key is the same is in figure 3.3a. b) Structure predictions of t-EL (top left) and τ-GFP (top right), τ-gp120 (bottom left), and τ-p24 (bottom right), courtesy of Dr. Ellis O'Neil. The N-termini are in blue, the C-termini are in red. Arrows indicate the first and last amino acid of the VHH moieties, which are represented in stick form. Images were generated using PyMOL.

III - Expression and characterisation of τ-gp120 and gp120

The pEAQ-*HT*-τ-gp120 construct was expressed in *N. benthamiana* alongside two controls, pEAQ-*HT*-EV (empty vector), and pEAQ-*HT*-τ-GFP; and small-scale extractions were subjected to western blot analysis using the 10E11 anti-HBcAg monoclonal antibody in order to assay the accumulation of recombinant protein in the soluble fractions (Fig. 6.2a). A band of about 57 kDa was seen in the plant extract expressing pEAQ-*HT*-τ-gp120, which corresponds to the expected size of τ-gp120. A band of a slightly smaller size (and stronger intensity) was seen in the extract expressing pEAQ-*HT*-τ-GFP, which corresponds to τ-GFP (with an expected size of 55.8 kDa – see figure 5.2). No specific band was seen in the lane corresponding to the empty vector control, although the non-specific 39 kDa band described in previous chapters was seen in all three lanes (see chapters 4 and 5). This western blot revealed that τ-gp120 is expressed, although not quite to the same levels as τ-GFP. This is consistent with all experiments carried out with parallel extractions of τ-GFP and τ-gp120: recovered yields of τ-GFP were always at least double those of τ-gp120.

The results shown in the previous chapter indicated that most of the τ-GFP which is expressed in *N. benthamiana* is insoluble unless GFP is co-expressed alongside it. To test whether this also holds for τ-gp120, pEAQ-*HT*-gp120, was created and its expression was tested in *N. benthamiana* with both a small-scale extraction and an apoplastic wash followed by an anti-HBcAg western blot (Fig. 6.2b). As its name suggests, gp120 is expected to run on an SDS-PAGE gel at around 120 kDa, although this is mostly due to its glycosylation: the protein itself is only 57.5 kDa. Western blot analysis revealed a very diffuse pattern of plant-produced gp120, with diffuse major bands running at 70 and 140 kDa, which suggests that the protein is glycosylated,

albeit not in the same manner as native gp120 produced in human cells. However, recombinant gp120 could not be extracted from an apoplastic wash, suggesting that it is either blocked at some stage of the secretory pathway, or that it is bound to the plant cell surface in some way.

As the gp120 construct was designed to include a C-terminal 6XHis tag, extracts from leaves agroinfiltrated with pEAQ-*HT*-gp120 were prepared 7 dpi and used for His-tag affinity chromatography. The fractions collected from the chromatography were analysed by an anti-His western blot (Fig. 6.2c). A similar pattern of bands was seen as with the previous experiment (Fig. 6.2b), but most of the signal came from the flow-through fraction (i.e. the extract that did not bind to the column) and the wash fraction (i.e. the extract that bound poorly to the column): relatively little gp120 was recovered in the elution fractions. Thus, the recombinant gp120 seemed to interact poorly with the nickel column, despite the fact that the His-tag was still present on the protein (the western blot was carried out with an anti-His antibody). This suggests that the His-tag is not properly exposed on the surface of the protein.

Finally, a co-expression experiment was carried out with gp120 and τ-gp120, alongside the co-expression of GFP and τ-GFP (see chapter 5). After small-scale extractions, anti-HBcAg western blots of the soluble (supernatant) and insoluble (resuspended pellet) fractions were carried out (Fig. 6.2d – this is in fact the same western blot as shown in figure 5.3). In both cases, bands of the appropriate sizes were seen in the insoluble fractions. However, only the GFP + τ-GFP co-expression resulted in detectable amounts of tandibody in the soluble fractions. The conclusion from this experiment is that co-expression of gp120 with τ-gp120 was not found to increase solubility or recovered yield of τ-gp120 in the way that GFP did with τ-GFP (see chapter 5). Further experiments showed that yields of τ-gp120 reach about 2-5 μg/g FWT.
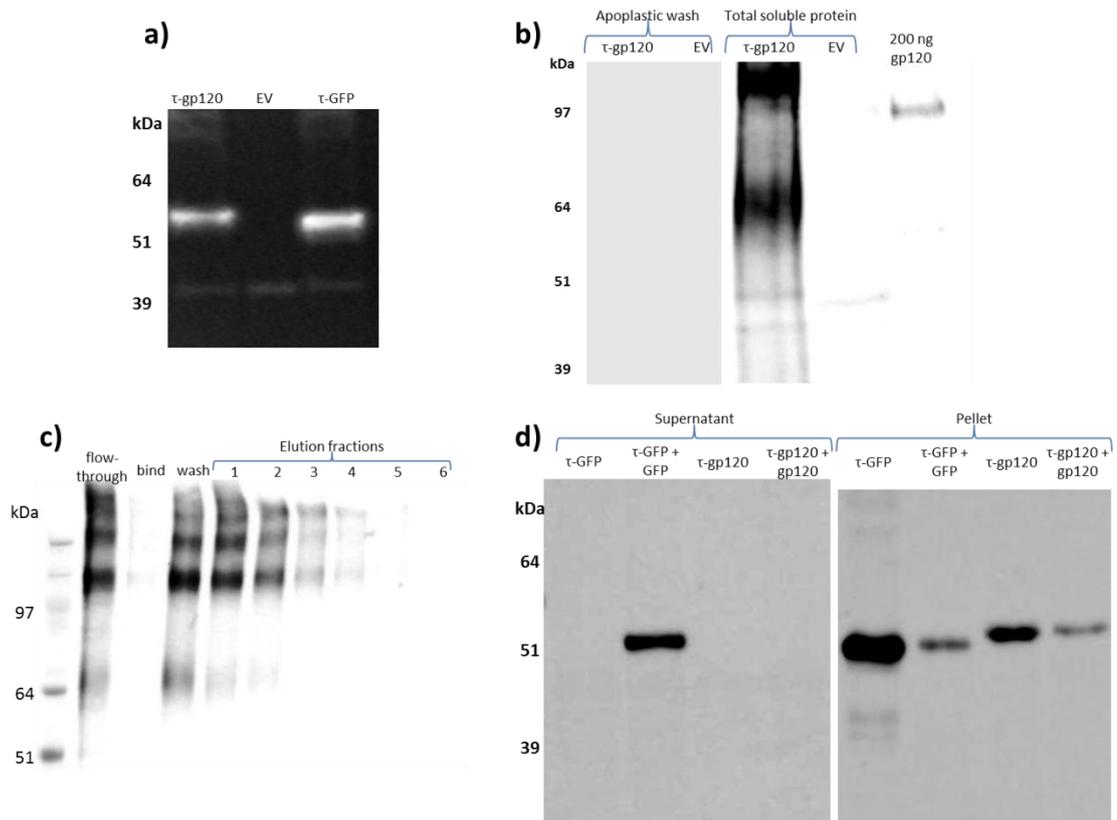
Figure 6.2 τ-gp120 and gp120 can be expressed in plants. a) Anti-HBcAg western blot of plant extract soluble fractions expressing τ-gp120, the empty vector control (EV), or τ-GFP. b) Anti-gp120 western blot of apoplastic wash (left) and total soluble protein (right) from leaves expressing τ-gp120 or the empty vector control (EV). The 200 ng of gp120 used as an internal control is commercially obtained gp120 produced in insect cells, which explains the size difference due to glycosylation. c) Anti-His western blot of the different fractions of his-tag affinity purification of plant-produced gp120. d) Western blot of supernatant (left) and resuspended pellet (right) fractions of plant extract expressing τ-GFP or τ-gp120 alone or co-expressed with its cognate antigen. This is the same blot as shown in figure 5.3.

To investigate whether τ-gp120 was able to form CLPs, a large-scale extract of leaf material was partially purified on a sucrose cushion and the interface fraction (between the 25% and 70% sucrose layers) was analysed by TEM. This revealed that τ-gp120 does indeed form particles, and that these particles can be produced and extracted from plant material in much the same way as τ-GFP (Fig. 6.3). As with τ-GFP CLPs (see chapter 5), the surface spikes of HBcAg are not clearly defined, and a minority of these particles adopt T=3 conformation (arrows). Moreover, it was found that these particles are stable in PBS for at least 6 months at 4°C (Fig. 6.3).
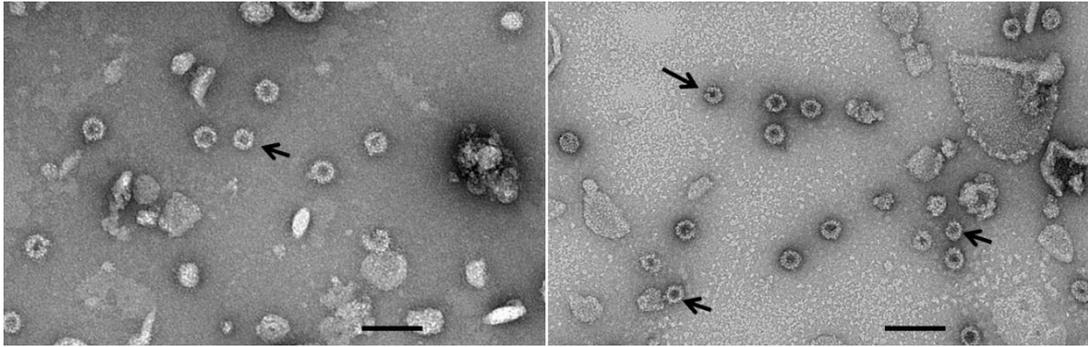
Figure 6.3 plant-produced τ-gp120 forms stable CLPs. Left: τ-gp120 CLPs produced alone in plants. Right: τ-gp120 CLPs after storage for 6 months at 4°C. Arrows indicate T=3 particles, scale bars are 100 nm.

The observation that co-expression of τ-gp120 with gp120 does not improve solubility or yield of τ-gp120 is not conclusive evidence of lack of binding. To determine whether τ-gp120 can bind gp120 *in vitro*, further tests were carried out. The first test was a sucrose gradient, in which 130 μg of concentrated tandibody (either τ-gp120 or τ-GFP as a control) was mixed with 8 μg of commercially obtained recombinant gp120 (CFAR, NIBSC). This mixture was then run on a continuous sucrose gradient, and the fractions were analysed by duplicate western blots probed with either anti-gp120 or anti-HBcAg antibodies (Fig. 6.4). Because the recombinant gp120 is produced in a baculovirus expression system, it runs at about 90 kDa instead of the native 120. For both co-infiltrations, signal corresponding to gp120 was seen to be present in low- to mid-concentration sucrose fractions, with signal corresponding to the tandibodies being present in mid-concentration sucrose fractions. The results were largely inconclusive because gp120 was found to migrate into the sucrose and partially co-localise with both tandibodies. However, the co-localisation was not complete, as would have been expected if τ-gp120 bound strongly to gp120. So this result suggests that if there is interaction between τ-gp120 and gp120, it is probably very weak.
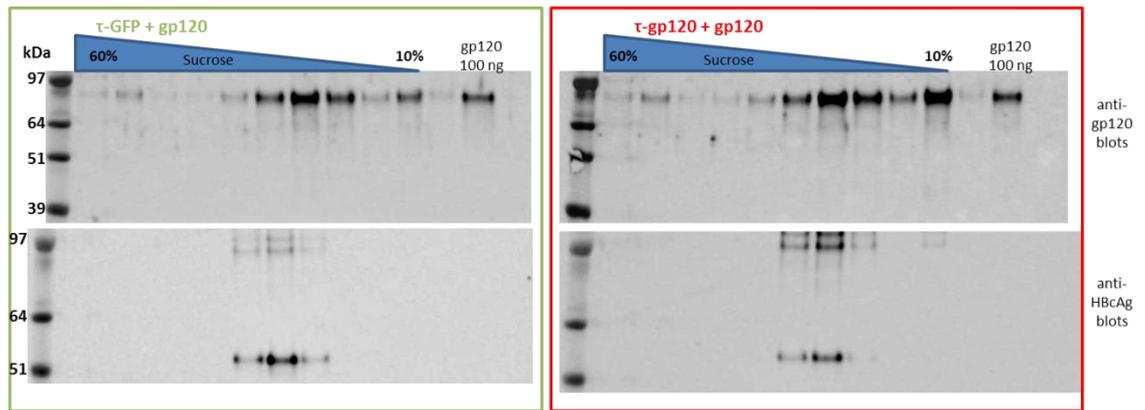
Figure 6.4 Sucrose gradient pull-down of gp120 by tandibodies. Commercially-obtained gp120 was mixed with partially purified plant-produced τ-GFP (left) and τ-gp120 (right) before being loaded onto sucrose cushions. After ultracentrifugation, the gradients were fractionated and the fractions were analysed by duplicate western blots probed with anti-gp120 (top) or anti-HBcAg (bottom) antibodies. The gp120 is heavy enough to penetrate the sucrose, and it col-localises with both tandibodies, so no conclusions can be drawn about gp120 binding to τ-gp120.

The next method used to investigate binding of τ-gp120 to gp120 was an ELISA. The protocol used was very similar to that used to determine binding of τ-GFP to GFP (see figure 5.8), except that the positive control was commercially obtained gp120 coating the wells directly, as opposed to a sandwich ELISA. The wells were coated with 200 ng of sucrose cushion – purified tandibody or commercially obtained anti-gp120 polyclonal antibody; and the amount of gp120 added per well ranged from 1.56- 12.5 ng. The goal was to have an excess of tandibody/antibody compared to gp120. Average signals for each treatment was subtracted from the background signal as determined by control wells where no gp120 was added. Neither τ-gp120 nor the τ-GFP control showed signal above background, whereas the positive control showed signal which decreased in proportion to the decreasing gp120 concentration, although this signal was quite low in general (Fig. 6.5). The result of this experiment indicated that τ-gp120 does not interact with gp120.

ELISA detection of gp120
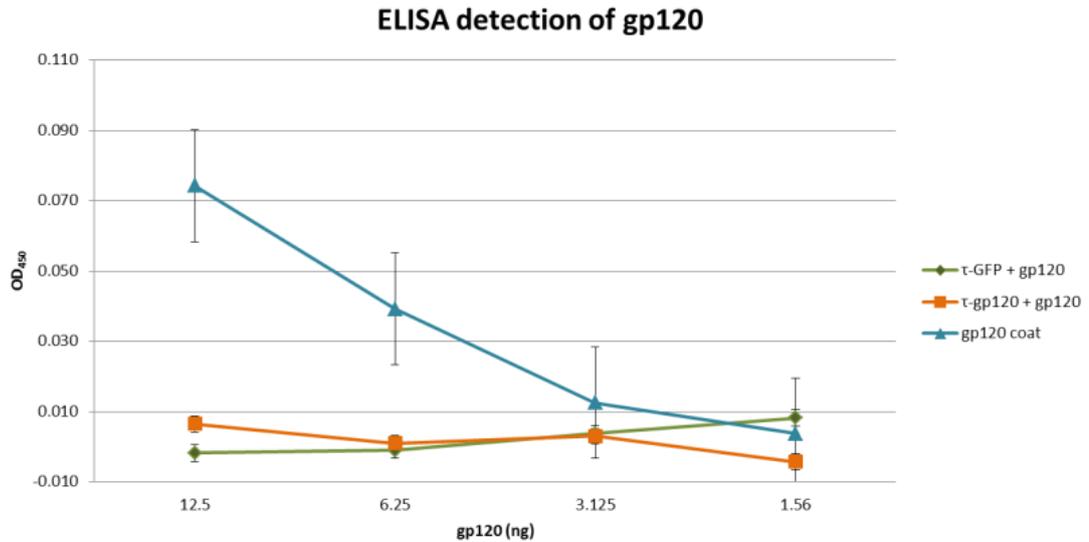
Figure 6.5 τ-gp120 CLPs do not bind to gp120 in an ELISA. Partially-purified τ-GFP and τ-gp120 CLPs were used to coat the wells of a 96-well plate in order to act like anti-gp120 antibodies in a sandwich ELISA. Pure gp120 (obtained commercially) was added at a range of concentrations to the wells and the gp120 was detected with a polyclonal anti-gp120 IgG conjugated to HRP. As a positive control, gp120 was used to coat the wells directly. Absorbance measurements indicated the amount of gp120 present in each well. Treatments were carried out in quadruplicates and average readings were subtracted from background to give net signal. While wells coated with gp120 gave signal in proportion to the amount of gp120 used, those coated with tandibodies did not. Error bars represent standard error.

This negative result suggested that the anti-gp120 nanobody on the τ-gp120 is not functional. There are two reasons why this might be the case: either the plant expression system is not conducive to proper folding of this nanobody, or the fusion of this nanobody to the tandem HBcAg affects its ability to bind its cognate antigen in some way. This could be through improper folding, or because it the presence of the CLP does not allow the nanobody to dock properly with its binding site on gp120, which is predicted to be the inside of the cavity under the bridging sheets of gp120 (Strokappe *et al.*, 2012). To investigate this, the anti-gp120 A12 nanobody sequence was cloned into pEAQ-*HT* using the AgeI and SmaI restriction sites in order to produce a 6XHis-tagged version of the A12 nanobody. The resulting pEAQ-*HT*-A12-His vector was expressed in plants. A small scale-extraction followed by His-tag affinity chromatography was carried out, and the fractions from the chromatography were analysed on a western blot probed with an anti-His antibody (Fig. 6.6). This revealed bands of the expected size (15 kDa)

for the nanobody in the total soluble protein extract, as well as in the two elution fractions. Very little signal was seen in the flow-through or the wash fractions, indicating that most of the A12-His protein binds to the nickel column, indicating that the His-tag is exposed. A faint band at about 30 kDa was seen in the first elution fraction, indicating possible dimerisation of the A12-His protein.
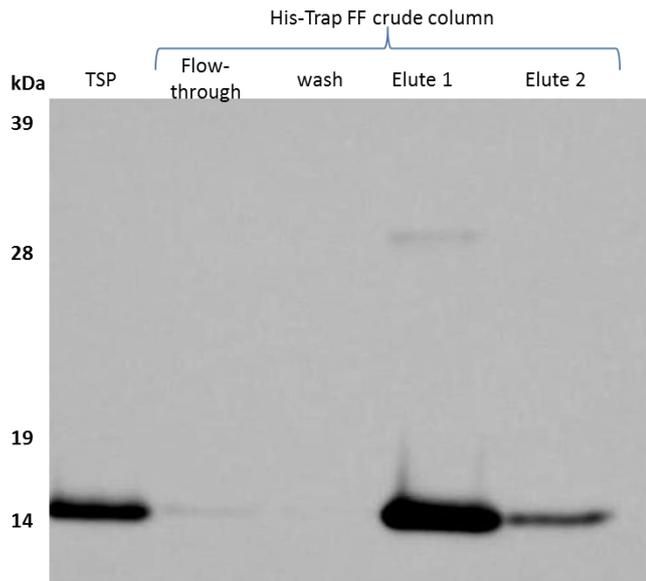
Figure 6.6 The His-tagged A12 nanobody is expressed in plants. The A12 nanobody was produced in plants using the pEAQ-*HT* vector with a C-terminal His-tag. The total soluble protein (TSP) containing recombinant A12-His was used for nickel column affinity purification. The different fractions were analysed by anti-His western blot, which revealed that the A12-His protein can be purified in this manner, indicating that the His-tag is exposed.

As an initial test to check binding of plant-produced A12-His to gp120, cobalt-coated magnetic beads (Dynabeads, Life Technologies) were used in a pull-down assay. The His-tagged A12 nanobody was fixed onto the beads, and these beads were used to pull down 10 µg of commercially obtained gp120 (NIBSC). As a control, unbound beads were mixed with the same amount (10 µg) of gp120, to make sure that any apparent binding of gp120 was to the nanobody, and not directly to the bead (Fig. 6.7). The different fractions were analysed by duplicate anti-gp120 and anti-His western blots. This revealed that gp120 (90 kDa) could be

pulled down by the beads coated with the nanobody, but not by the uncoated beads. This indicates that plant-produced A12-His does indeed bind gp120.
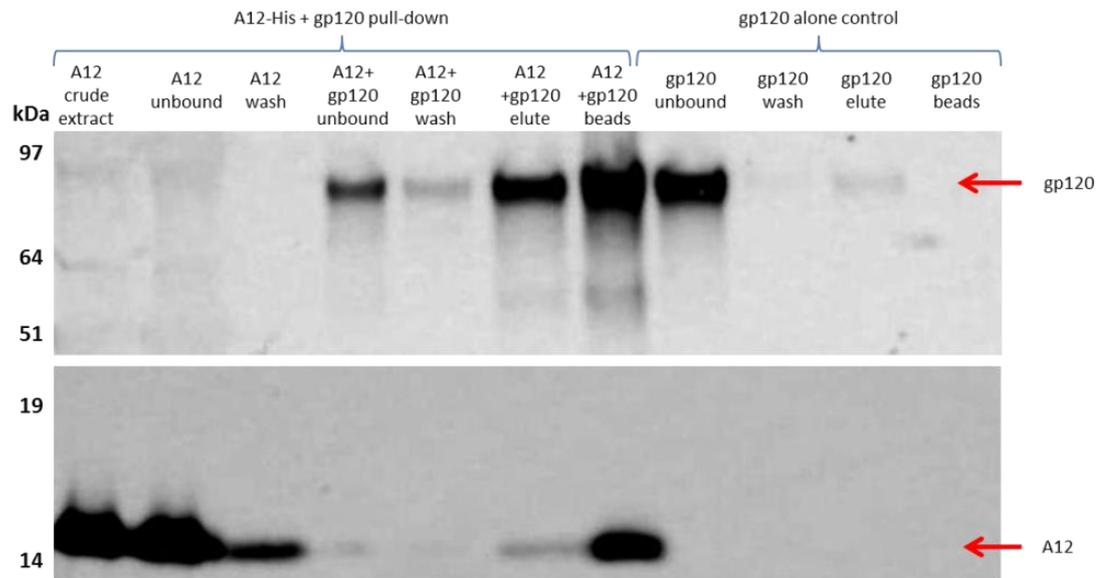
Figure 6.7 The plant-produced A12-His nanobody binds gp120 in a Dynabead pull-down assay. Cobalt-coated magnetic beads (Dynabeads) were coated with A12-His and used to pull down gp120. The gp120 co-localised with the A12-His nanobody in the elution fraction (elute), and on the beads after elution (beads). As a negative control, unbound beads were used, and these did not bind significantly to gp120 (unbound fraction). Fractions were analysed by duplicate western blots probed with anti-gp120 (top) and anti-His (bottom) antibodies.

To investigate this binding in more detail, an ELISA was set up to compare binding of A12-His and τ-gp120 to gp120, using a standard sandwich ELISA with commercially obtained polyclonal antibodies as a control. This experiment differs from that shown in figure 6.4 in three ways. Firstly, the positive control in this experiment was a sandwich ELISA in which gp120 was added to wells already coated with commercially obtained anti-gp120 polyclonal antibodies (as was done in figure 5.8). Secondly, much more protein was used to coat the wells of the plate: 500 ng per well of τ-gp120, A12-His, or anti-gp120 polyclonal antibody instead of 200 ng. Thirdly, much more gp120 was added to the wells after blocking: 50- 400 ng instead of 1.56- 12.5 ng. The results revealed net signal that decreases concomitantly with the decrease in gp120

concentration (Fig. 6.8). All net signals are relatively low, and the signal from wells coated with τ-gp120 is lower than the signal from wells coated with A12-His, which is itself lower than the signal corresponding to the commercially obtained anti-gp120 antibody. This is consistent with some low-level binding of both A12-His and τ-gp120, although A12-His seemed to bind more strongly than τ-gp120 at higher concentrations of gp120. It should be noted however that the amounts of protein used to detect this low-level binding are logs higher than what one might normally expect to use in an ELISA (see figure 5.8). This made obtaining the appropriate amount of material difficult given the low yields typically obtained for τ-gp120.
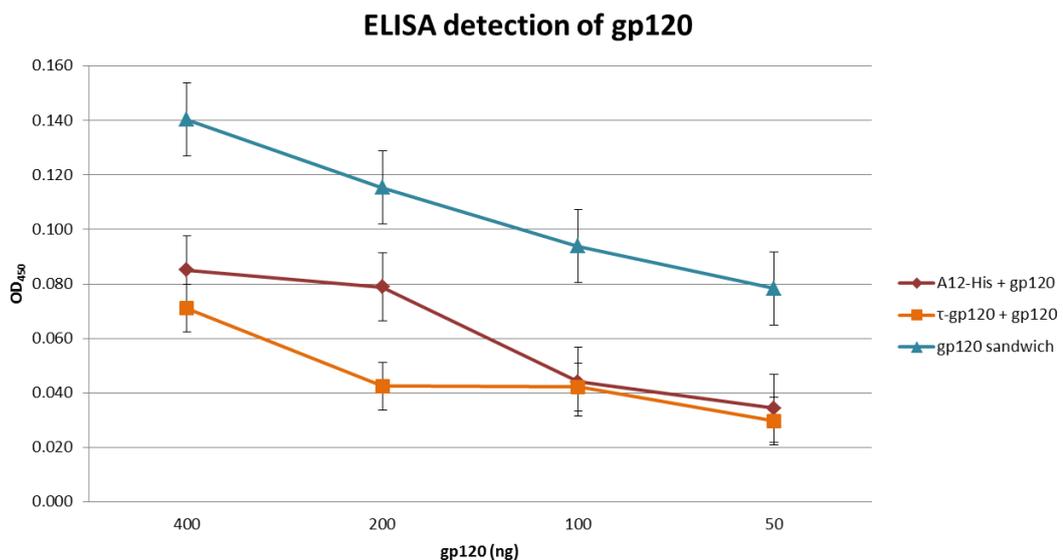
Figure 6.8 gp120 binds to A12-His more strongly than to τ-gp120 CLPs, but not as strongly as to a commercially available anti-gp120 antibody. Partially-purified τ-gp120 CLPs (orange) and A12-His nanobodies (red) were used to coat the wells of a 96-well plate in order to act like anti-gp120 antibodies in a sandwich ELISA. As a positive control, a polyclonal anti-gp120 IgG was used. Pure gp120 (obtained commercially) was added at a range of concentrations to the wells and the gp120 was detected with a polyclonal anti-gp120 IgG conjugated to HRP different to that used to coat the positive control wells. Absorbance measurements indicated the amount of gp120 present in each well. Treatments were carried out in quadruplicates and average readings were subtracted from background to give net signal. All net signal was low, despite using large quantities of protein. A12-His showed stronger binding to gp120 than τ-gp120 did, but this was weaker than gp120 binding to the anti-gp120 IgG. Error bars represent standard error.

IV - Expression and characterisation of τ-p24 and p24

A small-scale extraction was carried out on plant leaves separately expressing either pEAQ-*HT*-τ-p24 or pEAQ-*HT*-p24. The soluble protein fractions were then analysed on western blots using anti-HBcAg or anti-p24 antibodies (Fig. 6.9a). This analysis revealed that both proteins were expressed: a band corresponding to τ-p24 (55.7 kDa) was seen alongside a non-specific band that was also present in the empty vector control – infiltrated leaf tissue. A band corresponding to p24 (24 kDa) was seen in plant tissue expressing this protein, along with another band (about 48 kDa), which was considered to be a sign of dimerisation. Because co-expression of τ-GFP and GFP had resulted in increased yield and solubility of τ-GFP, τ-p24 was co-expressed with p24 alongside τ-GFP and GFP, and τ-gp120 and gp120. Small-sale extractions followed by western blot analysis of the supernatant (soluble) and resuspended pellet (insoluble) fractions were carried out (Fig. 6.9b - this is in fact the same western blot as shown in figures 5.3 and 6.2d). This revealed a band corresponding to τ-p24 (55.7 kDa) present in the insoluble fraction but not visible in the soluble fraction, similar to the situation found with τ-gp120. Moreover, the co-expression of τ-p24 with p24 (last lane on both blots) did not visibly increase the solubility of τ-p24. Furthermore, it is clear that expression levels of τ-p24 are not as high as τ-GFP. Further experiments revealed that τ-p24 yield is similar to that of τ-gp120: 2-5 μg/g FWT.
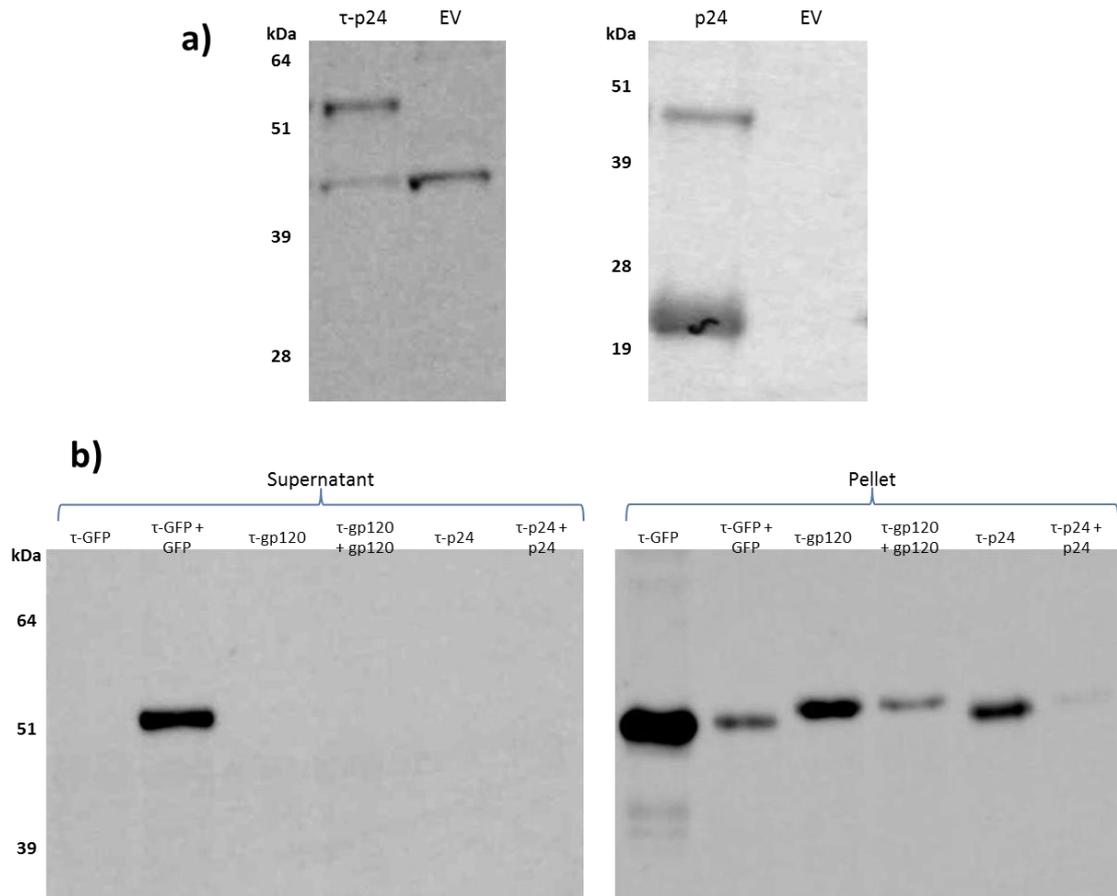
Figure 6.9 τ-p24 and p24 can be expressed in plants a) Anti-HBcAg western blot of plant extract soluble fractions expressing τ-p24, p24, or the empty vector control (EV). b) Western blots of supernatant (left) and resuspended pellet (right) fractions of plant extract expressing τ-GFP, τ-gp120, or τ-p24 alone or co-expressed with its cognate antigen. This is the same blot as shown in figures 5.3 and 6.2d.

To demonstrate that τ-p24 forms CLPs, τ-p24 with and without co-expressed p24 was partially purified on a sucrose cushion with layers of 25% and 70% sucrose. The interface between these two regions was collected and inspected by electron microscopy (Fig. 6.10). This revealed that τ-p24 does indeed form particles in plants, and that these particles can be produced, partially purified, and detected by using the same protocol as τ-GFP and τ-gp120. Moreover, τ-p24 CLPs are indistinguishable from τ-GFP or τ-gp120 CLPs under standard TEM conditions, including a similar preference for T=4 conformation over T=3 (arrows). There may be some evidence of a difference between τ-p24 CLPs produced alone or co-expressed with

p24: it seems that the CLPs produced from co-expression have a more pronounced knobbly morphology with less clearly defined spikes and apparent material projecting from the surface of the CLPs, as compared to τ-p24 CLPs produced alone (see figures 5.4 and 6.3). This could be an indication of p24 binding to the surface of τ-p24 CLPs, but it is by no means conclusive.
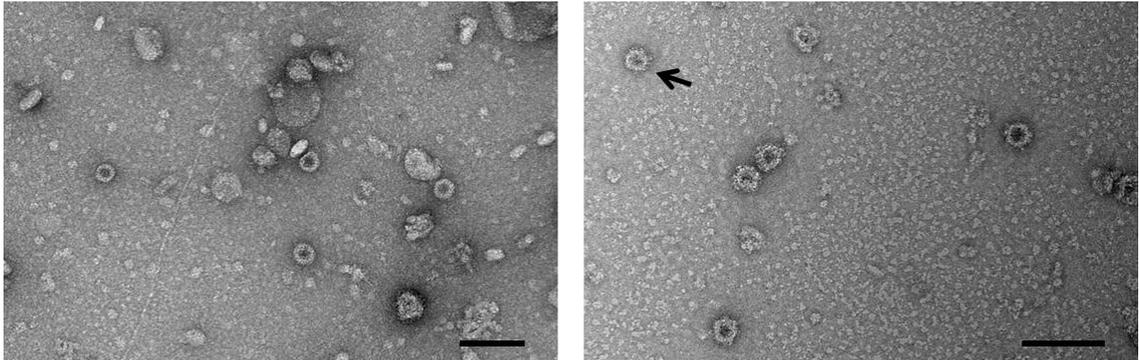
Figure 6.10 Plant-produced τ-p24 forms particles in plants when expressed alone or with p24. Left: τ-p24 CLPs produced alone in plants. Right: τ-p24 CLPs obtained from co-expression with p24. Arrow indicates T=3 particles, scale bars are 100 nm.

Once particle formation had been established, the functionality of the anti-p24 nanobody moiety of τ-p24 was studied. A sucrose cushion experiment similar to that described in chapter 5 was carried out. Briefly, plant leaf extracts from co-infiltrations of τ-p24 (or τ-GFP as a control) and p24 were run on sucrose cushions, and after ultracentrifugation, four fractions were collected: the supernatant (s), the top of the 25% sucrose layer (T), the interface between the 25% and 70% layers (M), and the bottom of the 70% sucrose layer (B). These fractions were analysed by duplicate western blots with either an anti-HBcAg antibody or an anti-p24 antibody used for detection (Fig. 6.11). The results revealed that p24 (24 kDa monomer band or 48 kDa dimer band) on its own or in the presence of τ-GFP particles (a negative control in this case) separates into two distinct species: a "light" species which remains in the supernatant or at the top of the 25% sucrose layer, and a "heavy" species which

migrates to the bottom of the gradient. Very little p24 was detected in the interface fraction. This suggests that the p24 forms two discrete species, as opposed to a continuum of forms of increasing size. The p24 capsid protein is known to be capable of oligomerisation under the right conditions (Momany *et al.*, 1996; Berthet-Colominas *et al.*, 1999), and the presence of 48 kDa dimer bands in all fractions indicates that the sizes are due to more than dimerisation. In any case, when τ-p24 is present in the plant extract, the amount of the light species of p24 is substantially reduced, and the p24 co-localises with τ-p24, mostly at the bottom of the cushion. This strongly suggests that τ-p24 interacts with at least the light form of p24, presumably by binding to it and causing it to sediment at the bottom of the sucrose cushion.
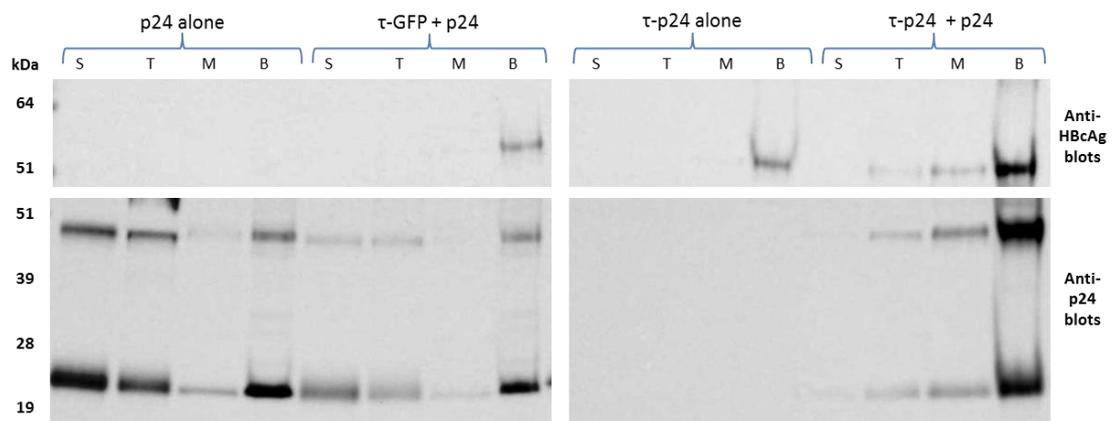
Figure 6.11 Plant-produced τ-p24 affects p24 in a sucrose gradient pull-down. Leaves producing p24 alone, τ-GFP and p24, τ-p24 alone, or τ-p24 and p24 were prepared. Extracts were subjected to sucrose cushions, and duplicate anti-HBcAg (top) and anti-p24 (bottom) western blots of the different fractions were prepared. When p24 is produced alone or with τ-GFP, there are two distinct species of p24 (24 kDa), which localise in the supernatant (S), and at the bottom (B) of the cushion, respectively. When p24 is produced with τ-p24, the light species in the supernatant almost completely disappears, suggesting that τ-p24 CLPs have caused it to sediment at the bottom of the cushion. S: supernatant, T: top of the sucrose cushion, M: interface between the 25% and 70% sucrose fractions, B: bottom of the sucrose cushion.

To further study the binding of τ-p24 to p24, a sandwich ELISA, similar to that described in figure 5.8, was performed. The wells were coated with 125 ng of tandibody (τ-p24 or τ-GFP as a control) or commercially obtained anti-p24 polyclonal antibody, and the amount of p24

added per well ranged from 1.5- 10 ng, the goal being to have excess of tandibody/antibody compared to p24. The presence of p24 was then detected using an anti-p24 antibody conjugated to HRP (different to that used to coat the surface of the positive control wells). Average signals for each treatment was subtracted from background signal as determined by control wells where no p24 was added. The results show that there is stronger net signal present in wells coated with τ-p24 than in wells coated with τ-GFP, although this signal is lower than that obtained from wells coated with the commercially obtained anti-p24 antibody (Fig. 6.12). This suggests that τ-p24 binds to p24, although the overall net signal was relatively weak, and background was relatively strong compared to the experiment carried out with τ-GFP and GFP (see figure 5.8).

ELISA detection of p24
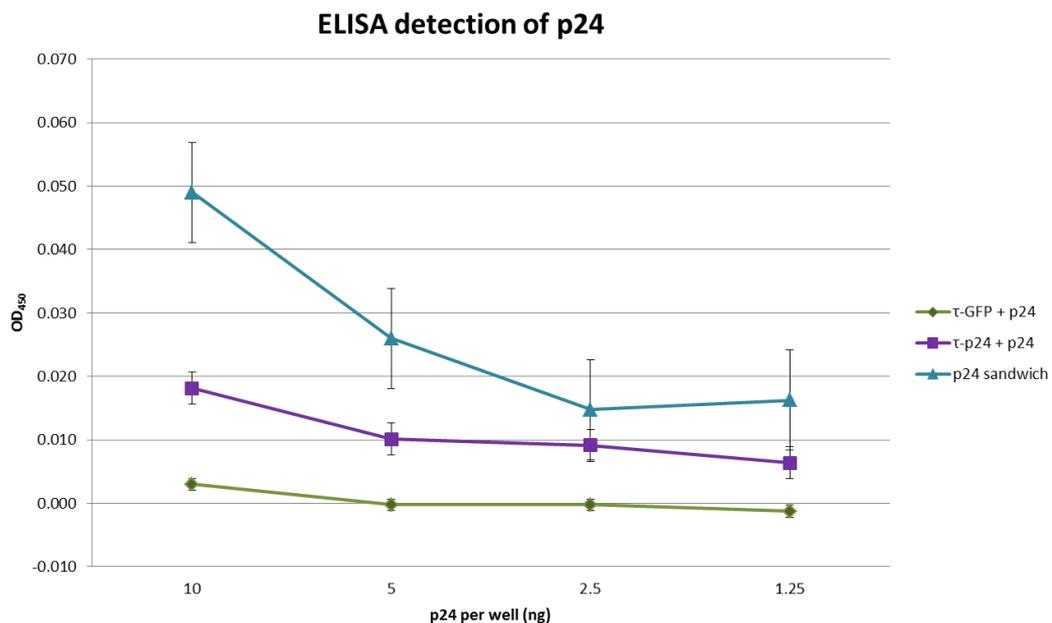
Figure 6.12 τ-p24 CLPs bind to p24 in an ELISA. Partially-purified τ-GFP and τ-p24 CLPs were used to coat the wells of a 96-well plate in order to act like anti-p24 antibodies in a sandwich ELISA. Pure p24 (commercially obtained) was added at a range of concentrations to the wells and this was detected with a polyclonal anti-p24 IgG conjugated to HRP. As a positive control, a commercially obtained polyclonal anti-p24 IgG different to that used for detection was used as part of a standard sandwich ELISA. Absorbance measurements indicated the amount of GFP present in each well. Treatments were carried out in quadruplicates and average readings were subtracted from background to give net signal. While the τ-GFP CLPs did not bind p24, τ-p24 CLPs bound did, although not as clearly as the anti-GFP IgG. Error bars represent standard error.

To further investigate binding, Surface Plasmon Resonance was used. In order to do this, a C-terminal His-tagged p24 construct was prepared in the pEAQ-*HT* vector. The pEAQ-*HT*-p24His plasmid was expressed in *N. benthamiana,* and a large-scale extraction followed by His-tag affinity chromatography was used to determine whether the protein was expressed and whether its C-terminal His-tag was exposed. The different fractions were analysed by a western blot probed with an anti-p24 antibody (Fig. 6.13). A band of the correct size (24 kDa) was seen in the soluble fraction of the crude extract, along with a 48 kDa dimer band, as was seen with non His-tagged p24. Moreover, p24His was detected in the flow-through from the affinity chromatography as well as the wash fraction, but the strongest signal was seen in the elution fraction. This indicates that p24His is expressed and the His-tag is exposed, and the protein can be purified by affinity chromatography.
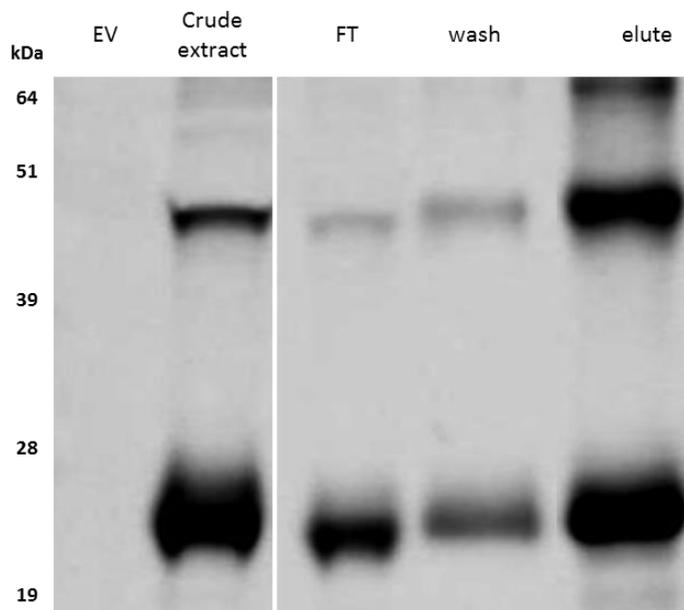
Figure 6.13 Plant-produced C-terminally His-tagged p24 is expressed in plants. The pEAQ-*HT* vector was used to produce p24 with a C-terminal His-tag. Soluble protein extract from leaves expressing this construct were subjected to affinity chromatography, and the fractions were analysed on a western blot probed with an anti-p24 antibody. While some of the p24 did not bind to the column (FT, flow-through), and some bound but was washed off during

the washing step (wash), most of the p24 bound tightly and was recovered during elution. This indicates that the His-tag is exposed.

This His-tagged p24 could be used for SPR by using a Ni-NTA Biacore chip, the same as was used for studying binding of τ-GFP to GFP (see figure 5.9). As with the SPR carried out with τ-GFP, Dr. Clare Stevenson (John Innes Centre) provided expert guidance and help with the data collection and analysis. The His-tagged p24 (at a concentration of 3 µM) was bound to the chip and then either τ-p24 or τ-GFP (negative control) at concentrations of 4 nM were flowed over the chip at 2, 5, or 10 µl/min, with a contact time on the chip of either 60, 120, or 180 seconds. The resulting sensograms were zeroed at injection of the tandibody, and these adjusted sensograms are shown in figure 6.14. These show a response that is consistently higher for τ-p24 (orange lines) than τ-GFP (green lines), regardless of flow rate or contact time. However, this response is only present during the injection of the tandibody: if tandibodies are binding p24, the binding is only transient. It should be noted that all responses appear to decrease over time, even during injection. This reflects p24-His washing off the chip spontaneously, even as the tandibodies are being flowed over the chip. A similar phenomenon was seen with GFP (see figure 5.9). In any case, the binding of the tandibody to its cognate antigen in this case contrasts markedly with τ-GFP binding to GFP (see figure 5.9). In the current case, binding is at best weak and transient, whereas τ-GFP clearly bound GFP very tightly. In fact, the only reason that τ-p24 can be said to bind to p24 from this experiment is that there is a very consistent difference between signal from τ-p24 and signal from the τ-GFP control. It is, however, possible that there were slight differences in the concentrations of the tandibody solutions, or their respective purity, and the difference in response could in fact be bulk non-specific signal. Overall, this experiment was largely inconclusive, and it suggests that if τ-p24 does bind p24, this binding is weak and transient.
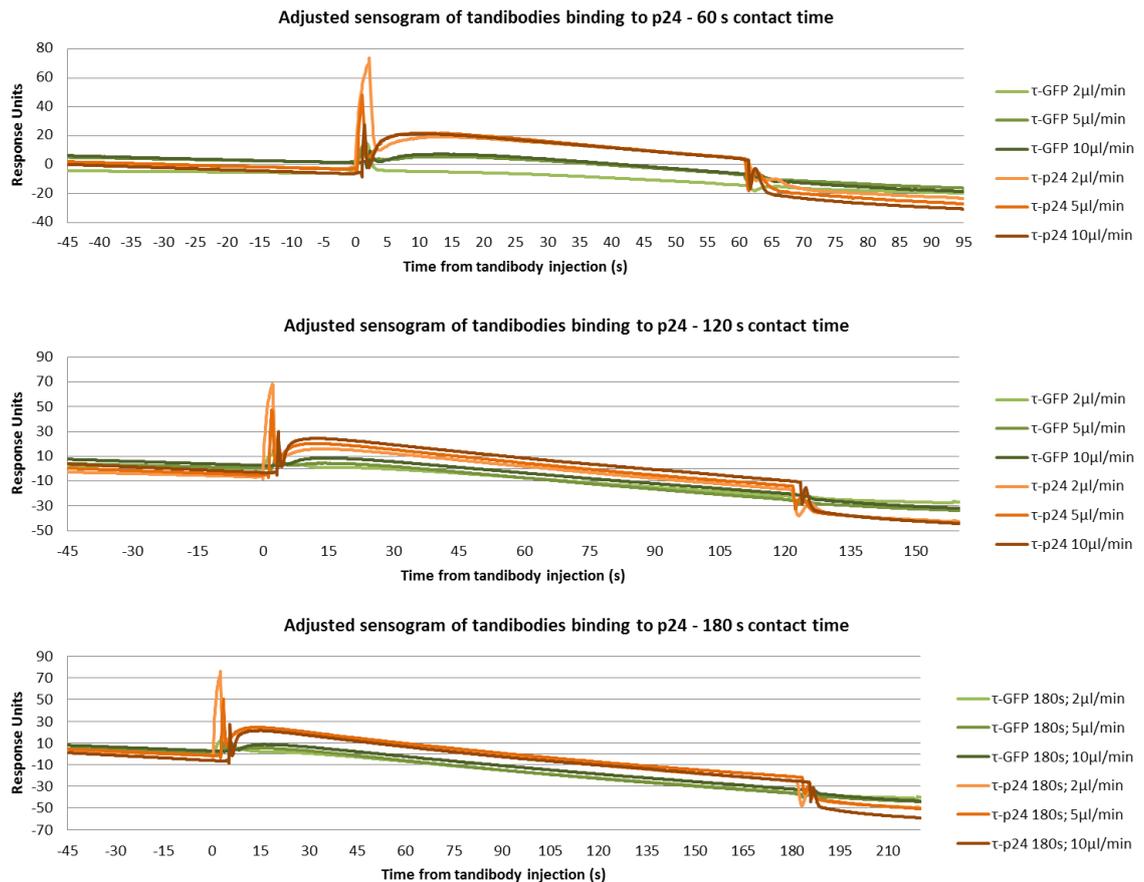
Figure 6.14 τ-p24 CLPs do not bind to p24 during surface plasmon resonance. Plant-produced His-tagged p24 was bound to the surface of a Ni-NTA SPR chip, and partially purified τ-GFP (green) or τ-p24 (orange) CLPs were flowed at varying flow rates (2, 5, and 10 µl/min) over the p24. Binding resulted in detection of Response Units. At three different contact times (corresponding to the amount of time given to the tandibodies to flow over the p24) of 60 s (top), 120 s (middle), and 180 s (bottom); τ-p24 response was very low, albeit consistently higher than τ-GFP response. Any weak and transient binding occurring during tandibody injection ended immediately after injection.

The results taken together show that τ-p24 does not bind to p24 as readily as τ-GFP binds to GFP (see chapter 5), and τ-gp120 does not bind to gp120. However, because the overall goal of the project is to develop a system for candidate vaccine production, it seemed worthwhile to make extra effort to display an antigen of medical relevance on the surface of a tandibody particle. To this end, the binding properties of τ-GFP and the medical relevance of p24 were combined: a fusion protein was created whereby solubility-enhanced GFP (sGFP, see Chapter 3) was fused to the N-terminus of p24 via a flexible 8 amino acid - long glycine rich linker

(GGSVDGGS). This fusion construct was cloned into pEAQ-*HT* via the AgeI and SmaI restriction sites to add a 6XHis tag added at the C-terminus of p24.

Co-infiltrations were set up in which the fusion protein, sGFP-p24His, was agroinfiltrated with τ-GFP, τ-gp120, or τ-p24. As controls, sGFP-p24His was also infiltrated on its own, and τ-GFP was co-infiltrated with GFP as in figure 5.6. At 6 dpi, large-scale extractions were carried out and the clarified extracts were loaded onto sucrose cushions and ultracentrifuged as described in chapter 5. The tubes were then visualised under UV light (Fig. 6.15a). This revealed that the fluorescence associated with sGFP-p24His infiltrated on its own or co-infiltrated with τ-gp120 is visible only in the supernatant (clarified extract). But when it is co-infiltrated with τ-GFP or τ-p24, the fluorescence is visible at the bottom of the sucrose cushions. This result is entirely compatible with the hypothesis that both τ-GFP and τ-p24 can bind to sGFP-p24His, but that τ-gp120 cannot. Duplicate western blots of the various fractions (bottom, middle, and top of the sucrose layers along with supernatant) were carried out with anti-HBcAg and anti-p24 antibodies (Fig. 6.15b). This showed that signal associated with sGFP-p24His (with an expected size of 52.4 ka) is present in all extracts in which it was expressed, and that it is subject to a certain amount of degradation. Moreover, the majority of the signal associated with sGFP-p24His is found in the supernatant when it is expressed on its own or co-expressed with τ-gp120. However, when sGFP-p24His is co-expressed with τ-GFP or τ-p24, the majority of the signal associated with the fusion protein is found in the bottom fraction (70 % sucrose), co-localising with the tandibody. The data from the UV fluorescence and from the western blot are concordant: the fluorescence is associated with sGFP-p24His, and this fusion protein co-migrates with τ-GFP and τ-p24, but not with τ-gp120 when the fusion protein and the tandibodies are co-expressed.
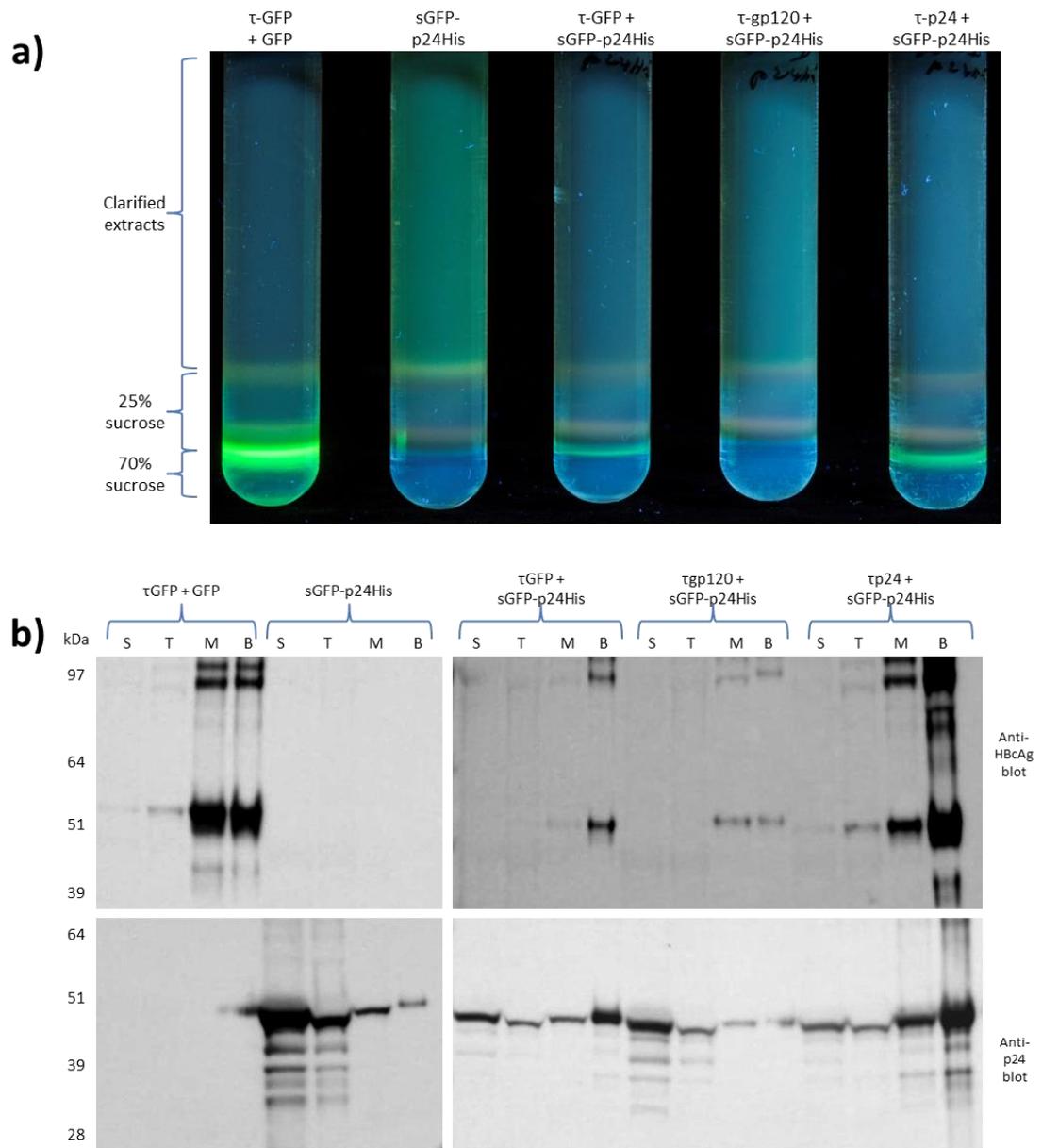
Figure 6.15 Both τ-GFP and τ-p24 bind the fusion protein sGFP-p24His. a) Leaves were co-infiltrated with different constructs as indicated above each tube, and the extracts were subjected to ultracentrifugation over sucrose cushions. These tubes were then photographed under UV light. This revealed that sGFP-p24His fluorescence is located in the supernatant (clarified extracts) when it this protein is produced alone or with τ-gp120. But when the fusion protein is co-expressed with τ-GFP or τ-p24, fluorescence sediments to the bottom of the sucrose cushion. b) The different fractions from the sucrose cushions above were analysed by duplicate western blots probed with anti-HBcAg (top) or anti-p24 (bottom) antibodies. This confirmed that sGFP-p24His co-localises with τ-GFP and τ-p24, but not with τ-gp120.

These results show that a fusion of sGFP and p24 can interact with both τ-GFP and τ-p24, which is strong evidence that it binds to both types of tandibodies. By contrast, τ-gp120, conforming to expectations, does not interact with the fusion protein. This confirms the

conclusions drawn from the pull-down experiment shown in figure 6.11, which suggested that τ-p24 binds to p24; but it conflicts with the results from SPR shown in figure 6.14, which suggested that any interaction between τ-p24 and C-terminally his-tagged p24 is, at best, weak and transient. Moreover figure 6.15b shows that co-infiltration of sGFP-p24His is associated with an increase in soluble amounts of τ-p24, to the extent that there is more τ-p24 than τ-GFP in the clarified extracts after co-infiltration. This is noteworthy because, as figure 6.9b shows, greater yields of soluble τ-GFP are typically obtained compared to soluble τ-p24, even when both are co-infiltrated with their cognate antigen. One possible explanation for this is that τ-p24 has higher affinity for sGFP-p24His than τ-GFP does. If the increase in solubility in all experiments is caused by GFP or sGFP, then it is possible that in the experiment shown in figure 6.15, τ-p24 benefits from this more than τ-GFP because τ-p24 has higher affinity for sGFP-p24His.

The bottom fractions of the co-infiltrations with τ-GFP and τ-p24 (labelled B on the western blots in figure 6.15b) were dialysed against PBS and visualised by TEM (Fig. 6.16). A high concentration of particles was seen in both samples, particularly with τ-p24, and the particles show the atypical knobbly surface morphology seen with τ-GFP particles bound to GFP (see chapter 5), with the surface spikes being far less clearly defined than those on HBcAg CLPs with no protein insert in the e1 loop. As expected, T=4 conformation dominates, with a small minority of T=3 CLPs (arrows). This is completely coherent with the results shown in figure 6.10, and with the conclusion that these particles are bound to sGFP-p24His.
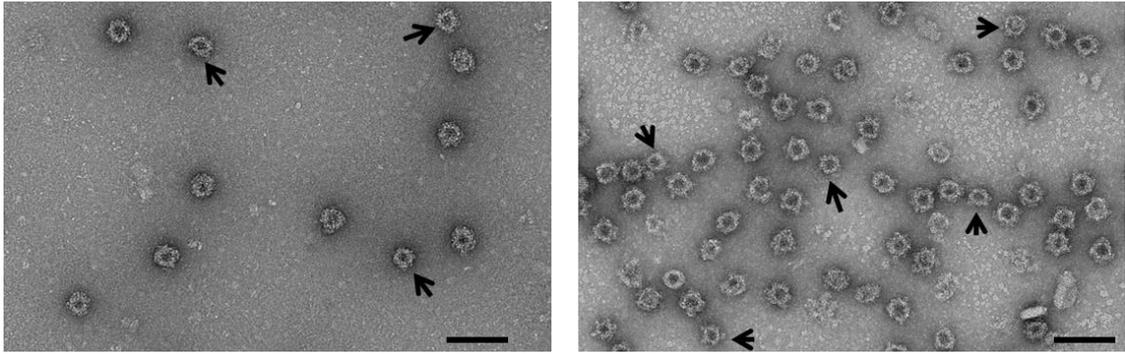
Figure 6.16 τ-GFP and τ-p24 CLPs co-expressed with sGFP-p24His exhibit "knobbly" morphology. The τ-GFP (left) and τ-p24 (right) particles obtained from the bottom fractions of the sucrose cushions shown in figure 6.15 were visualised by TEM. These are bound to the sGFP-p24His fusion protein and exhibit the knobbly morphology seen with tandibody particles, particularly when they are bound to an antigen. Arrows indicate T=3 particles, and scale bars are 100 nm.

V - Discussion

As well as the production of a further two tandibodies, the results presented in this chapter showed that HIV-1 gp120, p24, and a camelid nanobody (A12His) on its own, can be transiently expressed in plants using the pEAQ vector system. This is not the first time that that these proteins have been expressed in plant tissue. The gp140 protein, which is a synthetic construct containing gp120 along with the gp41 fusion peptide and heptad repeats, has been produced in plants (Rosenberg *et al.*, 2013). Similarly, HIV-1 p24 has been produced transiently in tobacco (Pérez-Filgueira *et al.*, 2004), and purified functional nanobodies have been obtained from plants as well (Ismaili *et al.*, 2007; Winichayakul *et al.*, 2009; Teh and Kavanagh, 2010).

The results presented in this chapter have interesting similarities and differences compared to those presented in chapter 5. The main similarity is that τ-gp120 and τ-p24 were found to form CLPs when expressed in plants, indicating that tandem HBcAg seems to be a reliable scaffold for nanobodies. However, the anti-gp120 nanobody did not appear to be fully functional when displayed on the surface of the tandem core particles. This nanobody, when expressed on its own, unconstrained by the HBcAg particle, was shown to bind gp120, although not as well or

as reliably as might have been expected given its published K_D of 0.1 nM. Indeed, it took extremely high concentrations of tandibody/nanobody and gp120 to detect any binding, which indicates that this binding was perhaps somewhat forced. But this raises the issue of comparison between the tandibody, the nanobody, and the commercially obtained polyclonal antibody used in the ELISA experiments. All have different sizes, spatial conformations, and valencies, making comparisons by mass or molar amounts somewhat difficult. In order to minimise the impact of these differences as much as possible, a conscious effort was made to coat the wells with excess of tandibody/nanobody/antibody compared to the amount of gp120 added subsequently. A similar effort was made in the sucrose gradient experiment, in which there was an excess of fused nanobody compared to gp120.

Because the commercially obtained gp120 used for these experiments came from the same source as the gp120 which was first used to identify the A12 nanobody, differences in gp120 sequence or glycosylation pattern are not likely to be a factor in the loss of function of the A12 nanobody. This suggests that this loss of function shown in these experiments is most likely due to both the plant expression system, and its fusion onto the tandem HBcAg particle. One potential factor to explain this is that the A12 produced in these experiments (both alone and as a fusion to HBcAg) has not been trafficked through the cell secretory pathway. This could affect the folding of the nanobody, since the ER and Golgi body environments are different to the cytosolic environment. That having been said, A12 was functional when expressed in *E. coli* (Strokappe *et al.*, 2012), though while bacteria do not have ER and Golgi body, the bacterial environment is again different from that of the plant cytosol.

Another potential factor is the flexibility of the nanobody when fused to the HBcAg scaffold, allowing it to bind an inner pocket of gp120. However, in this case it is expected that the nanobody fused to the surface of the HBcAg particles via long flexible linkers would have more freedom of movement than the nanobody on its own in an ELISA, since the latter would be

adsorbed onto the surface of the wells and be essentially prevented from moving. Moreover, the A12 nanobody was shown to be functional when displayed on the surface of bacteriophage as part of a phage display library (Forsman *et al.*, 2008), and when coated onto the well of a microtiter plate for ELISA (Strokappe *et al.*, 2012). The evidence therefore suggests that the lack of function of A12 here is mainly due to its expression in a plant system, with fusion to the tandem core scaffold also playing a role. These problems were not encountered with the anti-GFP Enhancer nanobody described in the previous chapter. In any case, the conclusion to be drawn from this set of experiments is that τ-gp120 can indeed form particles in plants (albeit at lower levels than τ-GFP), but that they do not interact strongly, if at all, with gp120.

The situation with τ-pgp120 contrasts with that found with τ-p24, which in several tests was found to bind p24, albeit not as clearly as τ-GFP was shown to bind GFP. However, the results relating to τ-p24 binding to p24 were somewhat contradictory. While the sucrose cushion pull-down experiments indicated binding, the ELISA painted a slightly less convincing picture, and the SPR data are underwhelming when compared to the SPR data in figure 5.9. In this sense it seems that as the sophistication of the analytical technique used increases, the apparent binding ability of τ-p24 decreases. However, there are a few practical considerations to bear in mind. In the SPR experiment (Fig. 6.14), τ-p24 was expected to bind to His-tagged p24 bound on the surface of the chip via this His-tag. However, the His-tag is on the C-terminus, and it is known that the anti-p24 nanobody binds to p24 at the C-terminus. Thus, it is possible that the physical conformation of the p24 during the SPR experiment could have had a negative impact on presentation of the antigen. The fusion protein sucrose cushion experiment shows that τ-p24 can bind to C-terminally His-tagged p24 in solution, so failure to bind during SPR would likely be due to the C-terminus of p24 facing the wrong way rather than occlusion by the His-tag. That having been said, all of the experiments done with plant-produced p24 (including the sucrose cushion pull-downs and the SPR) showed that oligomerisation occurs, and there is

literature which suggests that p24 dimerises in different conformations, including head-to-tail

N- to C- dimerisation (Berthet-Colominas *et al.*, 1999). This means that we should not treat p24

used in the SPR experiment as a homogeneous solution of monomers, and it is possible that

the p24 is being presented in a variety of conformational states in both liquid and solid phases.

Moreover, due to lack of availability of material, the concentration of tandibody used in this

SPR experiment was 4 nM, which is much lower than the 20 nM used to test for τ-GFP binding

to GFP. A low concentration of tandibody would be expected to result in low signal, but it

should still allow a difference to be seen between the test tandibody and the control, and it

should still allow the characteristics of the interaction between tandibody and antigen to be

visualised. If binding of τ-p24 to p24 was tight, the response curve would not be expected to

drop completely after the end of tandibody injection, regardless of concentration. This adds to

the overall conclusion to be drawn from the SPR experiment, which is that in that particular

case, binding of τ-p24 to p24, if there was any at all, was weak and transient.

In contrast to the SPR results, other assays present a far more optimistic picture of interaction

between τ-p24 and p24. In the first sucrose cushion experiment (Fig. 6.11), the distribution of

p24 throughout the different fractions after ultracentrifugation is clearly different when τ-p24

is present compared to when it is not. In the second sucrose cushion experiment, with the

sGFP-p24His fusion protein (Fig. 6.15b), the interaction is even clearer. The most parsimonious

explanation for this is that τ-p24 binds to p24 and causes it to sediment to the bottom of the

sucrose cushion. In the ELISA experiment (Fig. 6.12), the signal seen is as would be expected

for a tandibody that binds relatively weakly: consistent net signal which is higher than that

seen in the negative control, and which decreases with the decreasing concentration of p24.

However, the caveat with the ELISA experiment is in the strength of the net signal: the

background noise was high, leading to very low net signal after subtraction of background for

each treatment. While the end result should still be reliable, the data is not of the same quality

as the ELISA described in figure 5.8.

The apparent difference between the results of the different binding assays could reflect the differences in how the p24 is presented to the tandibody. In the sucrose cushion experiments, the two moieties are present together in the leaf tissue after co-infiltration, and in the leaf extract after harvest. This means that p24 is effectively in contact with τ-p24 in a liquid phase *in vivo* for days at 25°C pre-harvest and hours post-harvest at a range of temperatures (mostly at 4-8°C from grinding of the leaves to ultracentrifugation). In addition, there is no steric constraint on how p24 can interact with τ-p24, although it also means that all of the p24 may not be presented in the optimal conformation (whatever that may be) to τ-p24. In the ELISA, commercially-obtained, p24 is presented to the tandibodies in solution, with complete flexibility in terms of how it can bind to the nanobodies. The tandibody is then in contact with p24 for two hours at 37 °C. In the SPR experiment, the p24 had a C-terminal His-tag, was fixed onto a solid phase (the chip surface), and was not necessarily in a single conformation due to oligomerisation. Moreover the tandibody had, at most, 3 minutes of contact time at 25°C with p24. These differences in the way that τ-p24 is made to interact with p24 in these different experiments may explain the variability of the results. It is still possible, however, to conclude that τ-p24 does interact with p24, but this interaction is not as strong as the interaction between GFP and τ-GFP. Moreover, the data shown in figure 6.15 provide proof of concept for presenting a protein on the surface of a tandibody via a fusion partner which acts like a tag to link the protein of interest to a tandibody which displays a nanobody that does not recognise the protein of interest. This could be of crucial importance for displaying a protein for which there is no strongly-binding nanobodies available.

Ultimately the data presented here show that tandibody technology is generic in the sense that any nanobody is likely to allow particle formation when fused in the C-terminal loop of a tandem core construct. However, the yield of recovered tandibody particles will not always be identical from one tandibody to another: figure 6.9b clearly shows that more τ-GFP accumulates in plant leaves at 7 dpi than τ-gp120 or τ-p24, although figure 6.15b suggests that

this may depend on what proteins are bound to the tandibodies. Moreover, the capacity for

the tandibody to bind to its cognate antigen is not a given, but this may be overcome by using

a common peptide tag which could be added to antigens of interest, and against which a

functional nanobody could be used.

Chapter 7 : And now for something completely different. Producing a human enzyme in stable transgenic plants

I - Introduction

This chapter will describe the use of a derivative of the pEAQ-*HT* vector system to produce stable transgenic *N. benthamiana* which constitutively produce a human enzyme of medical relevance: human gastric lipase (hGL). Because of the difference between the work presented here and the work on HBcAg presented in previous chapters, and in the interest of clarity, this chapter will contain its own section on materials and methods.

Human gastric lipase is a human stomach enzyme which catabolises lipids in food by hydrolysing triglycerides (Ville *et al.*, 2002). There is potential for this enzyme to find uses in biotechnology, for *in vitro* digestion studies, as well as in medicine, such as for enzyme replacement therapy in patients suffering from pancreatic insufficiency. Indeed, patients suffering from this condition (which is often part of a broader condition such as cystic fibrosis or pancreatitis) have a more acidic smaller intestine, which decreases the efficiency of pancreatic lipase. Gastric lipase, with a pH optimum of 4, is still functional under these conditions, and can partially compensate for loss of pancreatic lipase activity (Abrams *et al.*, 1987; Carriere *et al.*, 2005). This enzyme has previously been transiently expressed to high levels in *N. benthamiana* using the pEAQ vector system *(Vardakou et al., 2012)*. The objective of the work presented here was to develop two stable transgenic lines of *N. benthamiana,* one of which was to express hGL-wt (unmodified hGL), while the other was to express hGL-C-His (hGL with a C-terminal 6XHis tag as an aid to purification). The ultimate goal was to produce transgenic lines that can be easily propagated and used as a steady supply of recombinant hGL for research purposes. The original plasmids that had been used for transient expression were unsuitable for the production of such stable transgenic lines because they contain the wild-

type P19 suppressor of gene silencing, which is developmentally toxic to plants (Chu *et al.*, 2000). The approach adopted was the use of the pEAQ vectors containing a modified version of the P19 (mP19) suppressor of silencing (Saxena *et al.*, 2011), which retains the ability to suppress silencing but is not developmentally toxic. Genetically stable, true-breeding transgenic lines were obtained for both constructs, and these transgenic lines were shown to produce active hGL.

II - Materials and Methods

A) Cloning

The plasmids encoding mP19 and either native or C-terminally His-tagged hGL (pEAQ-*HT*m-hGL-wt and pEAQ-*HT*m-hGL-C-His, respectively) were derived from pre-existing pEAQ-*HT*-hGL-wt and pEAQ-*HT*-hGL-C-His plasmids (Vardakou *et al.*, 2012). The modification required only a single-nucleotide change in the P19 suppressor of silencing gene: to change arginine 43 to tryptophan to create mP19. This eliminates the developmental toxicity, albeit at the cost of reduced yield of recombinant protein (Saxena *et al.*, 2011). This mutagenesis was carried out with the XL-Site-Directed Mutagenesis kit (Stratagene) for both plasmids. The introduction of the correct mutation was confirmed by sequence analysis.

B) Making transgenic plants

Plasmids pEAQ- *HT*m-hGL-wt and pEAQ-*HT*m-hGL-C-His were transformed into *A. tumefaciens* LBA4404 and tested in an agroinfiltration transient expression assay to confirm that they retained the ability to direct the expression of hGL-wt and hGL-C-His. Small-scale extractions were carried out on agroinfiltrated leaves and the soluble protein extracts were analysed by SDS-PAGE. Once the constructs were confirmed to be functional, *N. benthamiana* leaf-discs were transformed using the method described in Horsch and Klee (1986) in order to create stable transgenic lines. Briefly, *N. benthamiana* leaves were washed in 5 % (v/v) bleach

followed by 70 % (v/v) ethanol before leaf discs were cut out using a cork borer (size number 7). These were then incubated for 24 h in constant light at 25 °C on pre-callusing agar plates. These were composed of Murashige and Skoog (MS) agar (0.8% w/v agar, MS salts, 3% sucrose in 3 mM MES buffer, pH 5.7) supplemented with 1 µg/ml BAP (benzyl aminopurine); 0.1 µg/ml NAA (naphthalene acetic acid) and organic supplements (0.5 µg/ml nicotinic acid, 0.5 µg/ml pyridoxine, 0.5 µg/ml thiamine, 0.5 µg/ml glycine). The leaf discs were then dipped in a liquid culture of *A. tumefaciens* freshly grown without antibiotics but containing the appropriate plasmid. The leaf discs were then placed on fresh pre-callusing plates. After 48 h, the leaf discs were transferred to fresh pre-callusing plates which contained 500 µg/ml carbenicillin (to avoid contamination of the plates and to eliminate the agrobacteria) and 100 µg/ml kanamycin (for selection of transformed tissue). After 3-5 weeks, shoots began to appear on the growing calli. These were excised and transferred to sterile plastic rooting pots containing pre-callusing medium with only 0.6 % (w/v) agar and lacking the BAP and NAA hormones. Photographs of the different stages of transformation are shown in figure 7.1. All material on pre-callusing medium or in rooting pots was grown in a controlled-environment cabinet with 24 h light at 25 °C. Once roots had developed, the plantlets were transferred to soil and grown in the glasshouse (see chapter 2).

Figure 7.1 The different stages of transformation. *N. benthamiana* leaves (top left) were used to obtain leaf discs, which were incubated on pre-callusing medium. These developed into calli, until small shoots appeared (bottom right plate). These shoots were excised and transferred to a rooting pot (bottom left), where roots developed and the plantlet grew until it was large enough to be transferred to soil.

C) Obtaining homozygous lines from primary transformants

In order to obtain homozygous lines and to regenerate these for analysis, all transgenic plants were self-fertilised. Flowers were taped shut before they opened and only seed from taped flowers was used to grow the next generation (Fig. 7.2a). Transgenic seeds were grown in the controlled-environment cabinet on MS agar with 100 µg/ml kanamycin for selection of transgenic seeds (Fig. 7.2b). After 3-4 weeks, these were transferred to soil in the glasshouse. Homozygous lines were identified by analysing the kanamycin resistance of progeny seedlings: kanamycin is toxic to seedlings, and causes the cotyledons to bleach. The pEAQ vector used to transform the plants contained a kanamycin resistance gene in the T-DNA, so transformed plants are resistant to kanamycin. Simple Mendelian genetics therefore allows identification of homozygous T1 plants by analysing the percentage of bleached T2 seedlings obtained from self-fertilisation of a T1 plant.
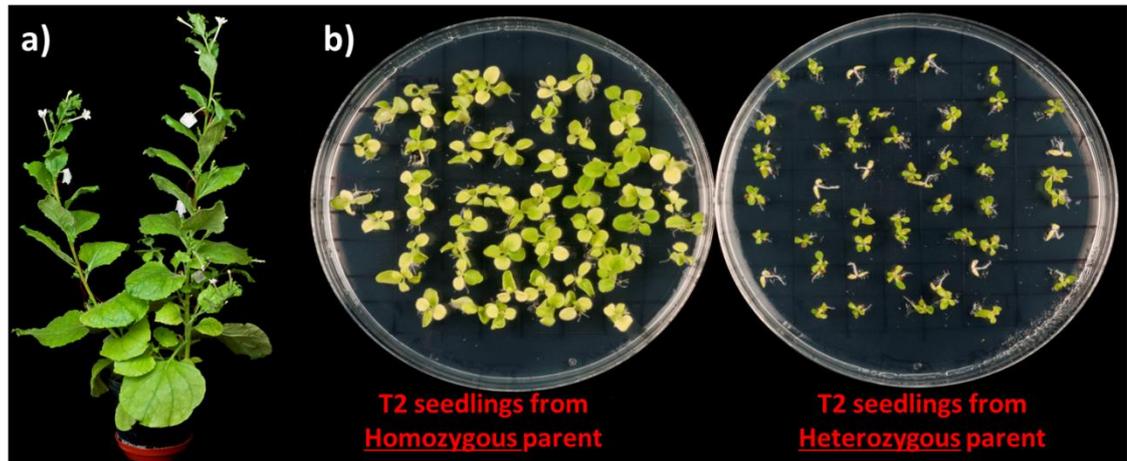
Figure 7.2 Using antibiotic selection to determine seedling zygosity. a) Developing flowers of transgenic plants are taped shut before they open to ensure self-fertilisation. b) The seeds obtained from self-fertilised flowers are sown on agar plates containing kanamycin. Transgenic seedlings carry the kanamycin resistance selection marker and develop normally, whereas non-transgenic seedlings undergo bleaching and die. T2 seedlings from a homozygous T1 parent plant will all carry the kanamycin resistance gene, so 100 % of the seedlings will survive. However, only 75 % of T2 seedlings from a heterozygous T1 parent plant will carry the resistance, so the remaining 25 % will bleach.

D) Verifying hGL expression and activity

In order to test the different lines for hGL expression, apoplastic washes were carried out (see chapter 2) using hGL extraction buffer: 90 % buffer A (50 mM NaOAc pH 4; 200 mM NaCl; 1 % (v/v) Triton X-100) and 10 % (v/v) isopropanol. The apoplastic fluid was then heat-shocked at 40°C for 10 minutes in order to denature plant proteins present in the apoplast, centrifuged to pellet insoluble material, and the supernatant was then used to analyse protein content by SDS-PAGE analysis (see chapter 2), or by a simple lipase activity assay, by measuring fluorescence released by hGL digestion of the reporter substrate methylumbeliferyl-oleate substrate (MUF-oleate, see Vardakou *et al.*, 2012). The activity of two hGL-wt lines was tested quantitatively by grinding transgenic leaves in hGL extraction buffer (described above), filtering through muslin cloth, clarifying the extract by centrifugation at 9,000 x g for 10 min, then analysing the extract on a pH-stat Titrino (718 STAT, Metrohm, UK) using the short chain triacylglycerol substrate tributyrin as described in Vardakou *et al.* (2012).

III - Results

It has previously been shown that hGL can be produced in *N. benthamiana* through transient expression using the pEAQ vector system (Vardakou *et al.*, 2012). This previous work had involved the creation of pEAQ-*HT*-hGL-wt, which does not encode a His-tag, and pEAQ-*HT*-hGL-C-His, which includes a C-terminal His-tag. Vectors for stable transgenics could therefore be easily produced by simply mutating the P19 suppressor of silencing on these extant plasmids (Saxena *et al.*, 2011). Site-directed mutagenesis was carried out on both hGL-bearing plasmids and sequencing of the mutated region of the P19 gene confirmed that they contained mP19. The ability of pEAQ-*HT*m-hGL-wt and pEAQ-*HT*m-hGL-C-His to direct the synthesis of hGL was tested alongside their parent vectors (pEAQ-*HT*-hGL-wt and pEAQ-*HT*-hGL-C-His, respectively) by transient expression. After 7 dpi, soluble fractions obtained from small-scale extractions were analysed by SDS-PAGE (Fig. 7.3). While the polypeptide chain of hGL is expected to have a size of 45.2 kDa, the protein is secreted and glycosylated, making it appear closer to 55 kDa on an SDS-PAGE gel. The extracts from leaves infiltrated with the mP19 forms of the plasmids showed expression of a protein of the expected size which was not present in the empty vector control plant extract. The amounts of hGL were lower than those seen in plant extract from leaves infiltrated with forms of the plasmids containing wild-type P19. This is consistent with the expected decrease in yield of recombinant protein which is known to occur when switching from wild-type P19 to mP19 (Saxena *et al.*, 2011).
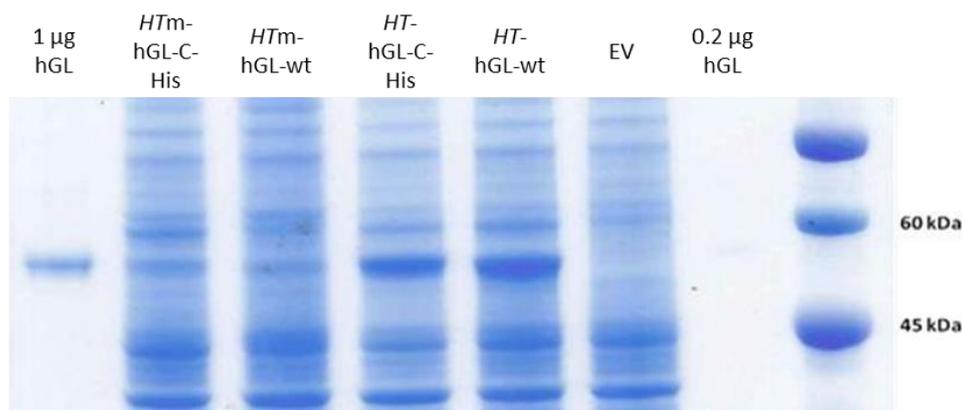
Figure 7.3 hGL can be transiently expressed in plants with both wild-type and mutant P19. Small-scale extractions were carried out on leaves that were agroinfiltrated with pEAQ vectors containing hGL-wt and hGL-C-His constructs with either mutant (*HT*m) or wild-type (*HT*) P19 suppressor of silencing. The soluble fractions were analysed by SDS-PAGE alongside purified plant-produced hGL (1 and 0.2 µg, kindly provided by Dr. Maria Vardakou) and extract from plants infiltrated with the empty-vector control. A band of the appropriate size (about 55 kDa) was seen in extracts from plants expressing hGL, but not in extracts from plants expressing the empty vector control. Clearly, yields of hGL are higher in plants infiltrated with a plasmid bearing wild-type P19 than in plants infiltrated with a plasmid bearing mutant P19.

Once the plasmids were shown to be functional, leaf discs (32 for each construct) were transformed with either pEAQ-*HT*m-hGL-wt or pEAQ-*HT*m-hGL-C-His. Because hGL is expected to be secreted and stable in the apoplast, apoplastic washes were carried out on leaves of regenerated plants in order to test for expression of hGL. These washes were analysed by SDS-PAGE (Fig. 7.4). This revealed that a protein of the expected size was visible in the apoplasts of all of the T0 lines tested (with the possible exception of hGL-wt line 9), with no protein of a similar size detected in the apoplast of a non-transgenic control plant. The results also indicate that relatively pure hGL can be obtained from apoplastic washes, as compared to soluble protein extracts shown in figure 7.3.
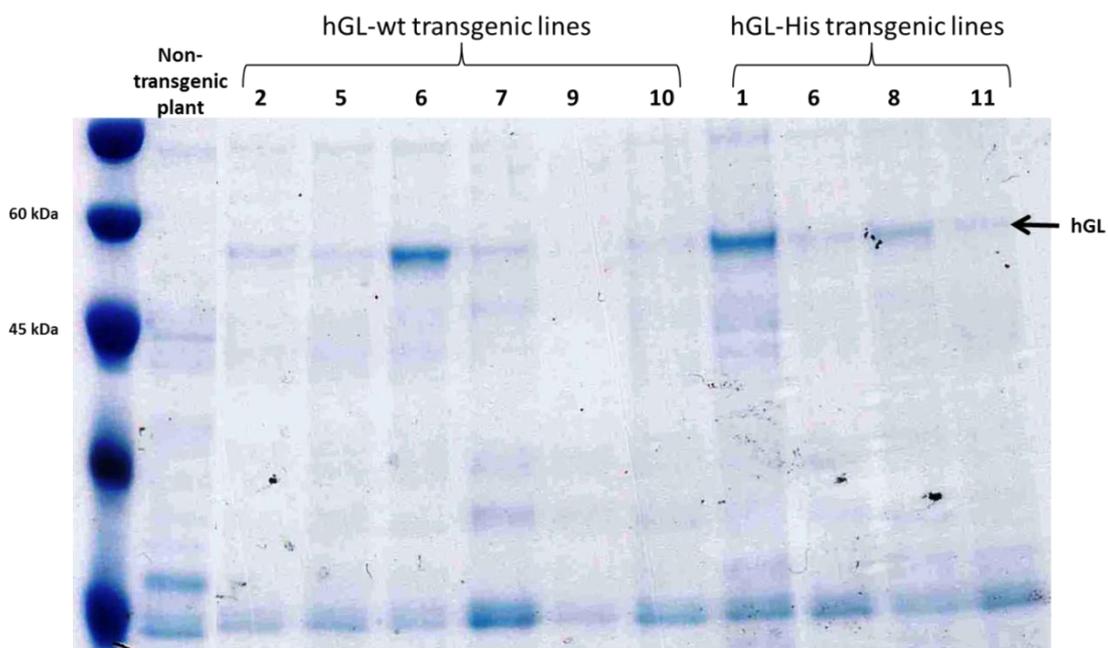
Figure 7.4 Apoplastic washes reveal hGL expression in T0 transformants. Apoplastic washes were carried out on leaves from 6 hGL-wt primary transformants and 4 hGL-C-His primary transformants, along with a non-transgenic negative control plant. These washes were analysed by SDS-PAGE, and hGL was detected in all lines except hGL-wt line 9. It should be noted that the plants (and the sampled leaves) were not all the same age when sampled.

Apoplastic washes from the hGL-wt T0 plants were then used for a non-quantitative lipase activity assay based on a synthetic fluorescent lipid analog (MUF-oleate). The positive control was an apoplastic wash from a leaf transiently expressing hGL-wt, and the negative control was an apoplastic wash from a non-transgenic leaf. This revealed that all lines tested produced active lipase, including line 9, for which the actual amount of hGL produced had been undetectable by SDS-PAGE analysis above (Fig. 7.5).
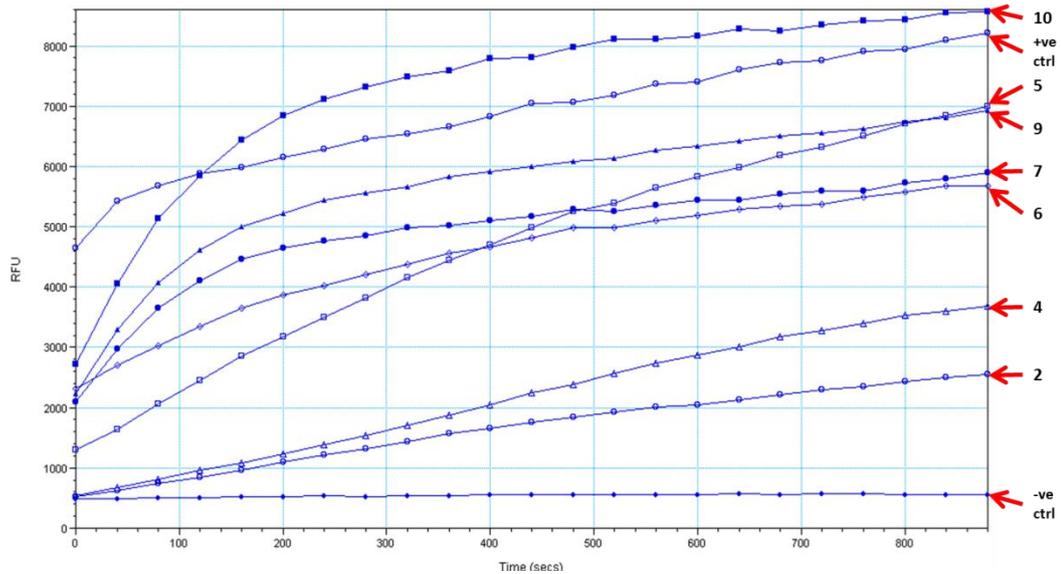
Figure 7.5 MUF-oleate digestion assay reveals that apoplastic washes from all hGL-wt T0 plants contain active lipase. Apoplastic washes from primary transformants expressing hGL-wt were used in a fluorescence-based lipase activity assay which relies on digestion of the synthetic lipid analog MUF-oleate. Because the MUF-oleate could not be added to all apoplastic washes at the same time, and because fluorescence could not be measured immediately after injection, the evolution of fluorescence through time is not comparable between lines. Numbers indicate the number of the transgenic plant, +ve ctrl is apoplastic wash from a transiently-transformed leaf, and –ve ctrl is apoplastic wash from a non-transformed leaf.

The hGL-wt and hGL-His T0 lines were self-fertilised and plants from the T1 generation were also self-fertilised so that the seeds could be used to determine which of the T1 plants were homozygous: the segregation pattern of T2 seedlings is visible on kanamycin-containing agar plates, where non-transgenic plantlets die rapidly after germination (see figure 7.2b). For hGL-wt T0, lines 2, 4, 6, and 10 were tested for homozygous T1 offspring in this way, and homozygous offspring from both plants 2 and 10 were identified. Homozygous T1 plants 2.06 and 10.03 were selected for transgenerational gene stability studies. For hGL-His, T0 lines 6, 11, and 14 were tested for homozygous T1 offspring, and homozygous offspring were identified from plants 11 and 14. Homozygous T1 plants 2.06 and 10.03 (for hGL-wt), along with 11.08 and 14.03 (for hGL-C-His) were selected for transgenerational gene stability studies. These consisted of self-fertilising the plants and analysing the offspring for presence of hGL by SDS-PAGE analysis of the apoplastic fluid. The two hGL-wt lines were propagated for a total of

six generations (seed from the T6 generation was collected and stored), and the two hGL-His

lines were propagated for a total of three generations (seed from the T3 generation was

collected and stored). At each generation, an apoplastic wash of one leaf from each plant was

tested for the presence of hGL by SDS-PAGE analysis, and all plants were determined to be

expressing the recombinant protein. Figure 7.6 shows a representative gel of these apoplastic

washes. This specific gel shows apoplastic washes from 5 individual plants from the T4

generation for each homozygous hGL-wt line 2 and 10. All plants clearly produce a protein of

the appropriate size which is not present in the apoplast of a non-transgenic control plant.
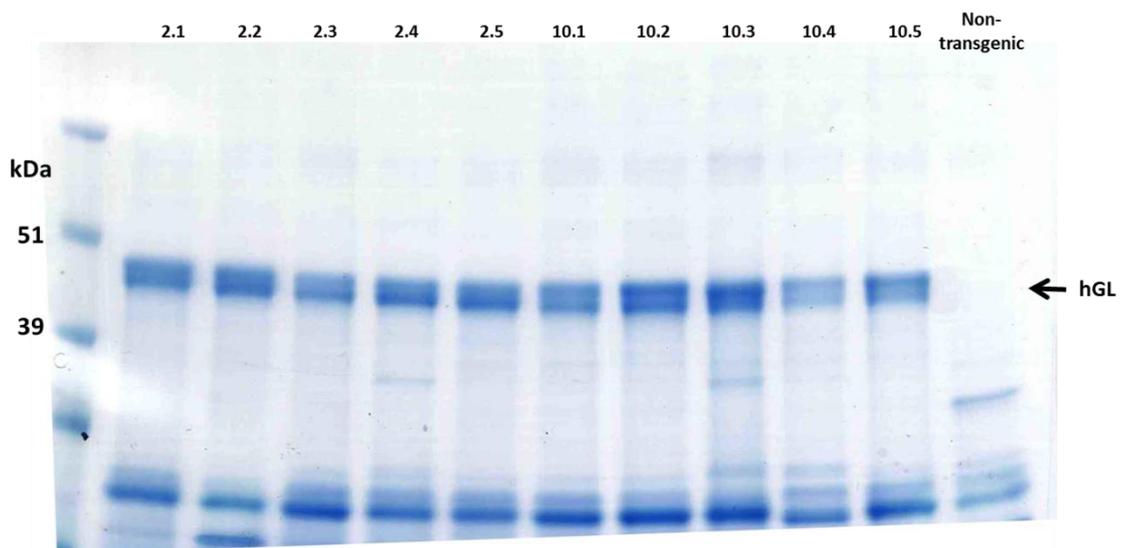
Figure 7.6 Representative analysis of homozygous line hGL expression assay. At each generation, transgenic plants were tested for hGL expression by SDS-PAGE analysis of apoplastic washes. In this specific instance, five T4 generation plants from each of homozygous hGL-wt lines 2 and 10 were tested. All plants expressed hGL.

The homozygous hGL-wt T1 plants 2.06 and 10.03 (the same as used for subsequent gene

stability studies) were used in a quantitative lipase activity assay. Leaves were blended in hGL

extraction buffer and the soluble protein extracts were tested using a pH-Stat Titrino as

described in Vardakou *et al.* (2012). This method allows the lipase activity of the extract to be

measured in enzyme units (U) per gram of fresh weight leaf tissue (FWT). The result, shown in

figure 7.7, reveals that both lines produce between 25 and 30 U/g FWT.
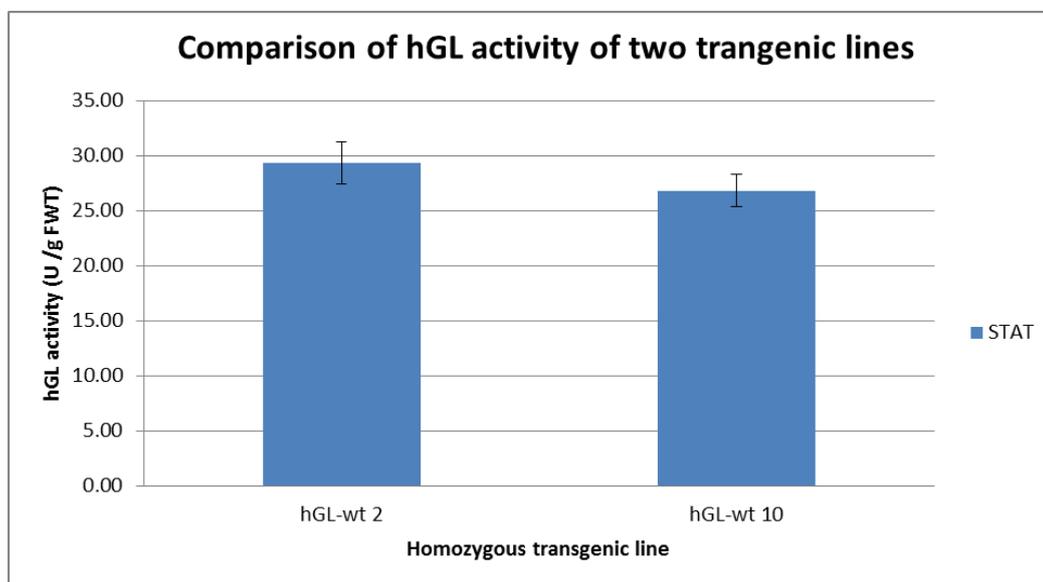
Figure 7.7 Lipase activity of soluble protein extracts from hGL-wt lines 2 and 10. Homozygous T1 plants hGL-wt lines 2 and 10 were used to quantify lipase activity in transgenic plants. Large-scale extractions were carried out and soluble protein extracts were analysed by the pH Stat Titrino method. Both lines produced between 25 and 30 units of lipase activity per gram of fresh-weight leaf tissue.

During propagation of the transgenic plants, it was noted that taped flowers often yielded

small, atrophied seed pods, or no seed pods at all. In an attempt to examine this phenomenon,

the percentage of taped flowers which yielded usable seed pods was noted, and the results are

summarised in figure 7.8. For hGL-wt lines, fertility decreased until generation T4, and then

increased slightly in generation T6 (no data was collected for generation T5). For hGL-His lines,

fertility decreased between generations T2 and T3.

Transgenerational fertility of transgenic lines
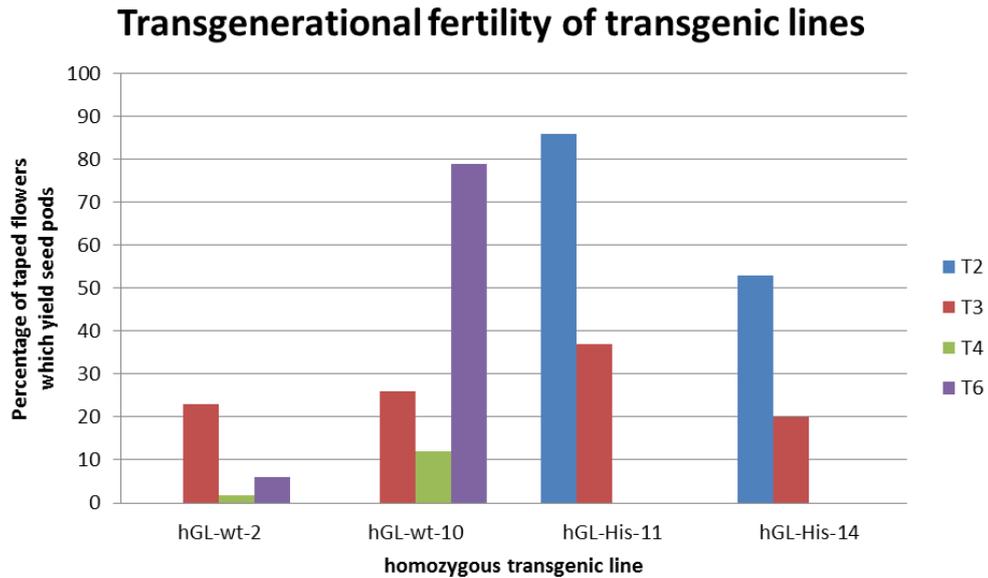
Figure 7.8 Fertility of transgenic lines varies from one generation to the next. The number of flowers taped on homozygous lines, along with the number of fertile seed pods obtained from these lines, was recorded. The percentage of taped flowers which yielded seed pods was calculated for each generation of each transgenic line. This revealed that fertility decreases in the early generations of hGL transgenics.

IV - Discussion

The results presented here demonstrate that the pEAQ-*HT* plasmid can be used to generate stable transgenic lines which constitutively produce an active enzyme. The only modification that needs to be carried out is the R43W mutation in the P19 suppressor of silencing gene; this mutation abolishes developmental toxicity, as demonstrated by Saxena *et al.* (2011). However, this study reported only the generation of GFP- producing stable transgenic lines. Attempts to create anti-HIV IgG – producing lines resulted in the creation of sterile transgenic plants (Saxena, 2012), suggesting that the ability to produce fertile lines may depend on the protein that is being expressed. The data presented here demonstrate that an active enzyme of human origin, hGL, can be produced constitutively in transgenic plants, and that the resulting plants are fertile. Preliminary enzyme activity data suggested that the lipase activity is too low for routine use of the transgenic plant extracts in *in vitro* models of digestion such as the Dynamic

Gastric Model (DGM) in use at the Institute of Food Research (IFR, Norwich, UK). That said, the hGL-His lines were never tested for lipase activity, and it could be that they produce greater lipase activity than the hGL-wt lines. Moreover, it would probably be worth testing apoplastic washes (as opposed to blended leaf extracts) from leaves of different ages, and possibly purified lipase from the different lines, before deciding that the lipase activity is insufficient in all lines. It is likely that crude extracts such as those used here contain lipids which might compete for hGL activity during the lipase activity assay. Due to time constraints, testing the activity of hGL from apoplastic washes or after purification could not be carried out.

Whatever the lipase activity of the transgenic plants, this has still been informative in terms of the production of transgenic plants using pEAQ-*HT*. Indeed, it was found that numerous fertile T0 plantlets could readily be obtained from only a few transformed leaf discs, and these T0 plants could produce homozygous offspring. These homozygous offspring were found to be fertile and their transgene was found to be stable from one generation to the next; although an interesting decrease in fertility was observed in all lines from one generation to the next, until T6 (for hGL-wt lines). This phenomenon was characterised by self-fertilised flowers which did not yield fully-developed seed pods. Often, a seed pod was observed to be forming, but it would become abortive and necrotic before reaching maturity. The seed pods that did form tended to be small and contained few seeds, although these few seeds were typically capable of germinating and producing the next generation of plants. The overall impression was that all of the self-fertilised flowers gave rise to abortive seed pods in the observed generations, but this seed pod abortion was a long process rather than a sudden event, and some pods managed to set seed before development aborted, and thus yielded some seeds. This occurrence therefore seemed to be more complex than simple low fertility: it appeared to be a sort of transgenerational seed pod degeneracy phenomenon. There does not seem to be any obvious reason why such a process might occur. It is unlikely to be a consequence of inbreeding since *N. benthamiana* is known to be self-compatible (Busot *et al.*, 2008). The cause

is therefore likely to be either environmental or epigenetic. It is interesting to note that while this effect seemed to get worse from one generation to the next, the last generation tested (T6 for hGL-wt lines) showed an improvement over the previous tested generation (T4), particularly with line hGL-wt-10. This is encouraging, since persistence in such a syndrome would ultimately lead to plants that are incapable of self-fertilising, thus preventing the inbred lines from being propagated. It would be interesting to continue propagating the plants in order to study the phenomenon further. It would also be of interest to cross the lines with one another, or with non-transgenic plants, in order to see if this effect takes place in the offspring of such hybrids.

Chapter 8 : General Conclusion

The primary aim of this thesis was to develop plant-produced HBcAg CLPs as generic carriers for antigens of interest using a variety of approaches. For the first time, the SplitCore system (Walker *et al.*, 2011) was compared to tandem core technology, and the latter was found to be a more suitable for use in a plant expression system (see chapter 3). An attempt to display an anti-phOx scFv revealed a limitation of tandem core technology: it is difficult to display proteins with more than one domain (see chapter 4). While it may be worth trying to present different scFvs on tandem cores, the data presented in chapter 4 suggest that scFv-display may well be impossible, or at least prohibitively impractical, with the tandem core system.

Tandem core technology was successfully developed for antigen presentation using two different approaches: with t-KD particles allowing the chemical conjugation of antigens of interest (see chapter 4) and through the development of tandibody particles displaying nanobodies capable of binding antigens of interest (see chapters 5 and 6). Both of these methods are potentially generic and thus could make an important contribution to the field of VLP-based vaccines.

The hypothesis on which tandibody technology was originally based is that once proof of concept had been established with one nanobody-antigen pair, it would be straightforward to switch to a different nanobody-antigen pair, with similar results. This hypothesis proved to be correct insofar as nanobodies against GFP, HIV gp120, and HIV p24 (originally from camel, llama, and alpaca, respectively) were shown to allow CLP formation when fused to the C-terminal e1 loop of a tandem core construct. The three different tandibody particles could be produced, extracted and purified in the same way. However, there was a difference in total yield between the three different tandibodies, with τ-GFP accumulating to higher levels than τ-gp120 or τ-p24. Moreover, the solubility and final yield of the tandibody particles was shown

to be affected by binding to the cognate antigen. Finally, only two of the tandibodies could be shown to bind to their cognate antigen, with the third (τ-gp120) apparently being rendered non-functional by fusion to a tandem core (and possibly to some extent by the plant expression system). Taken together, these results suggest that the tandibody system is generic in the sense of allowing the production of CLPs presenting any nanobody, but not necessarily in the sense of allowing these nanobodies to retain their ability to bind to the cognate antigen.

The problem of variation in binding ability of the presented nanobody can potentially be circumvented by presenting a nanobody specific for an antigenic tag. In this way, any protein fused to the tag could be displayed on the surface of a tandibody particle presenting the tag-specific nanobody. This was demonstrated with the fusion protein sGFP-p24His (see chapter 6), but it opens the possibility of using a more convenient small peptide tag instead of a whole protein as the fusion partner for the antigen of interest, provided a compatible tag-specific nanobody can be obtained. An attractive candidate for this approach would be the creation of a tandibody displaying the anti-EPEA nanobody (GenBank accession number 2X6M_A), a VHH of camel origin which recognises the amino acid sequence EPEA at the C-terminus of a protein (De Genst *et al.*, 2010; Djender *et al.*, 2014). Testing tandibody technology with this nanobody will undoubtedly form a major part of future work on this technology.

If a tandibody particle with this nanobody (or another which binds a different small peptide tag sequence) is found to function as well as τ-GFP in plants, then a truly generic system will have been created whereby any protein can be non-covalently attached to the surface of the same "universal" tandibody CLP so long as the antigen contains the peptide tag. While the same effect could be achieved by using a whole protein tag (as was done with sGFP-p24His), this would almost certainly be unsuitable for use in vaccines, due to regulatory distrust with regards to non-native proteins. That having been said, it is also possible that a small peptide tag such as EPEA would also be unsuitable in the eyes of regulators, as it would also, in effect,

render the antigen non-native. However, if a universal tandibody were to be deemed unsuitable for the design of human vaccines, it might still be authorised for veterinary vaccines.

Even if a universal tandibody were deemed unsuitable for vaccine design, specific nanobodies could be developed for use in the tandibody system. The work carried out in this thesis has focused on previously-characterised nanobodies with publicly available sequences, but nothing precludes future work from developing new nanobodies. In fact, nanobody phage display technology (Maass *et al.*, 2007) could be used to identify nanobodies that bind an antigen of interest with high affinity, and also function well in the tandibody system.

It may also be possible to optimise existing nanobodies for use in the tandibody system. The complementarity determining regions (CDRs) of a nanobody are the sequences which interact directly with the cognate antigen (Muyldermans, 2001; Vanlandschoot *et al.*, 2011). These three regions are usually easily identifiable in all antibody sequences due to their conserved location and hypervariable nature (Sela-Culang *et al.*, 2013). It is therefore conceivable that the CDRs of a nanobody which is not functional in the tandibody system could be transferred to the sequence of a nanobody which is. For example, the CDRs of the anti-gp120 nanobody could be used to replace the CDRs of the anti-GFP nanobody used in this thesis. In this way, it might be possible to rescue functionality of recalcitrant nanobodies. Moreover, this could potentially be an interesting tool to study the effect of each CDR: it is suspected that CDR3 is the CDR that contributes most to affinity in nanobodies (Muyldermans, 2001). Comparing a nanobody that has been modified at all three CDRs to one that has only had CDR3 modified could test this commonly-held belief. Such CDR grafting has been done before, though it is sometimes called "framework shuffling", with "framework" referring to the parts of an antibody variable region sequence that are not CDRs. Indeed, this has been done to switch CDRs between murine antibody fragments (Calcutt *et al.*, 1996), and to "humanise" mouse

antibodies by replacing human CDRs with those from a mouse antibody (Dall'Acqua *et al.*, 2005). It has also been done with nanobodies: (Vincke *et al.*, 2009) grafted CDRs from a llama nanobody onto a "humanised" camel nanobody scaffold. This "humanisation" corresponds to a series of point mutations in the framework regions of the nanobody to better mimic a consensus human sequence. It is described as a necessary step for the deployment of nanobodies as human therapeutics, for regulatory reasons. Humanisation could also be attempted in the context of tandibody technology.

The work carried out with hGL-producing stable transgenic lines demonstrates that proteins of pharmaceutical interest can be readily produced in stable transgenic lines once the constructs have been validated using the transient approach. This opens up the possibility of creating stable transgenic lines which constitutively produce tandibodies. This would be useful if a long-term, consistent supply of a particular tandibody is required, as the time-consuming agroinfiltration step would be removed once stable transgenic lines are obtained. This would be of particular relevance if a universal tandibody, as described above, were to be developed. Because stable transgenic plants yield lower quantities of recombinant protein than transiently-expressed plants, future work may focus on further optimising the pEAQ vector system for the production of stable transgenic plants – perhaps by placing the wild-type P19 suppressor of silencing under the control of an inducible promoter.

The work presented in this thesis has made significant contribution to the development of genetically engineered HBcAg as a tool for vaccine design. This is of particular significance to the field of VLP-based vaccines, which is likely to be the key contributor to the next generation of prophylactic and therapeutic vaccines. Future work will make the leap from the synthetic biology presented in this thesis to the immunology that will determine how well these new tools perform *in vivo*. Light will be thrown on the usefulness of tandem core particles and their efficacy as antigen carriers.

Bibliography

Abrams, C.K., Hamosh, M., Dutta, S.K., Hubbard, V.S., and Hamosh, P. (1987). Role of nonpancreatic lipolytic activity in exocrine pancreatic insufficiency. Gastroenterology **92,** 125-129.

Ahmad, Z.A., Yeap, S.K., Ali, A.M., Ho, W.Y., Alitheen, N.B., and Hamid, M. (2012). scFv antibody: principles and clinical application. Clinical & developmental immunology **2012,** 980250.

Arora, U., Tyagi, P., Swaminathan, S., and Khanna, N. (2012). Chimeric Hepatitis B core antigen virus-like particles displaying the envelope domain III of dengue virus type 2. Journal of nanobiotechnology **10,** 30.

Arora, U., Tyagi, P., Swaminathan, S., and Khanna, N. (2013). Virus-like particles displaying envelope domain III of dengue virus type 2 induce virus-specific antibody response in mice. Vaccine **31,** 873-878.

Bachmann, M.F., and Jennings, G.T. (2010). Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns. Nat Rev Immunol **10,** 787-796.

Bandurska, K., Brodzik, R., Spitsin, S., Kohl, T., Portocarrero, C., Smirnov, Y., Pogrebnyak, N., Sirko, A., Koprowski, H., and Golovkin, M. (2008). Plant-produced hepatitis B core protein chimera carrying anthrax protective antigen domain-4. Hybridoma **27,** 241-247.

Barton, K.A., Binns, A.N., Matzke, A.J., and Chilton, M.D. (1983). Regeneration of intact tobacco plants containing full length copies of genetically engineered T-DNA, and transmission of T-DNA to R1 progeny. Cell **32,** 1033-1043.

Bechtold, N., and Pelletier, G. (1998). In Planta AgrobacteriumMediated Transformation of Adult Arabidopsis thaliana Plants by Vacuum Infiltration. In Arabidopsis Protocols, J. Martinez-Zapater and J. Salinas, eds (Humana Press), pp. 259-266.

Bechtold, N., Ellis, J., and Pelletier, G. (1993). In planta Agrobacterium mediated gene transfer by infiltration of adult Arabidopsis thaliana plants. Comptes rendus de l'Academie des sciences **316,** 1194-1200.

Beekwilder, J., van Houwelingen, A., van Beckhoven, J., and Speksnijder, A. (2008). Stable recombinant alpaca antibodies for detection of Tulip virus X. Eur J Plant Pathol **121,** 477-485.

Berthet-Colominas, C., Monaco, S., Novelli, A., Sibai, G., Mallet, F., and Cusack, S. (1999). Head-to-tail dimers and interdomain flexibility revealed by the crystal structure of HIV-1 capsid protein (p24) complexed with a monoclonal antibody Fab. EMBO J **18,** 1124-1136.

Beterams, G., Bottcher, B., and Nassal, M. (2000). Packaging of up to 240 subunits of a 17 kDa nuclease into the interior of recombinant hepatitis B virus capsids. FEBS Lett **481,** 169-176.

Bevan, M. (1984). Binary Agrobacterium vectors for plant transformation. Nucleic Acids Res **12,** 8711-8721.

Birkett, A., Lyons, K., Schmidt, A., Boyd, D., Oliveira, G.A., Siddique, A., Nussenzweig, R., Calvo-Calle, J.M., and Nardin, E. (2002). A Modified Hepatitis B Virus Core Particle Containing Multiple Epitopes of the Plasmodium falciparum Circumsporozoite Protein Provides a Highly Immunogenic Malaria Vaccine in Preclinical Analyses in Rodent and Primate Hosts. Infection and Immunity **70,** 6860-6870.

Blokhina, E.A., Kuprianov, V.V., Stepanova, L.A., Tsybalova, L.M., Kiselev, O.I., Ravin, N.V., and Skryabin, K.G. (2013). A molecular assembly system for presentation of antigens on the surface of HBc virus-like particles. Virology **435,** 293-300.

Borisova, G.P., Berzins, I., Pushko, P.M., Pumpen, P., Gren, E.J., Tsibinogin, V.V., Loseva, V., Ose, V., Ulrich, R., Siakkou, H., and et al. (1989). Recombinant core particles of hepatitis B virus exposing foreign antigenic determinants on their surface. FEBS Lett **259,** 121-124.

Brown, A.L., Francis, M.J., Hastings, G.Z., Parry, N.R., Barnett, P.V., Rowlands, D.J., and Clarke, B.E. (1991). Foreign epitopes in immunodominant regions of hepatitis B core particles are highly immunogenic and conformationally restricted. Vaccine **9,** 595-601.

Busot, G.Y., McClure, B., Ibarra-Sanchez, C.P., Jimenez-Duran, K., Vazquez-Santana, S., and Cruz-Garcia, F. (2008). Pollination in Nicotiana alata stimulates synthesis and transfer to the stigmatic surface of NaStEP, a vacuolar Kunitz proteinase inhibitor homologue. Journal of experimental botany **59,** 3187-3201.

Cabantous, S., Terwilliger, T.C., and Waldo, G.S. (2005). Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. Nat Biotech **23,** 102-107.

Calcutt, M.J., Komissarov, A.A., Marchbank, M.T., and Deutscher, S.L. (1996). Analysis of a nucleic-acid-binding antibody fragment: Construction and characterization of heavy-chain complementarity-determining region switch variants. Gene **168,** 9-14.

Canaan, S., Dupuis, L., Rivière, M., Faessel, K., Romette, J.-L., Verger, R., and Wicker-Planquart, C. (1998). Purification and Interfacial Behavior of Recombinant Human Gastric Lipase Produced from Insect Cells in a Bioreactor. Protein Expression and Purification **14,** 23-30.

Cañizares, M.C., Liu, L., Perrin, Y., Tsakiris, E., and Lomonossoff, G.P. (2006). A bipartite system for the constitutive and inducible expression of high levels of foreign proteins in plants. Plant Biotechnology Journal **4,** 183-193.

Carriere, F., Grandval, P., Renou, C., Palomba, A., Prieri, F., Giallo, J., Henniges, F., Sander-Struckmeier, S., and Laugier, R. (2005). Quantitative study of digestive enzyme secretion and gastrointestinal lipolysis in chronic pancreatitis. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association **3,** 28-38.

Chackerian, B., Lowy, D.R., and Schiller, J.T. (1999). Induction of autoantibodies to mouse CCR5 with recombinant papillomavirus particles. Proc Natl Acad Sci U S A **96,** 2373-2378.

Chakrabarti, B.K., Kong, W.P., Wu, B.Y., Yang, Z.Y., Friborg, J., Ling, X., King, S.R., Montefiori, D.C., and Nabel, G.J. (2002). Modifications of the human immunodeficiency virus envelope glycoprotein enhance immunogenicity for genetic immunization. J Virol **76,** 5357-5368.

Chambers, M.A., Dougan, G., Newman, J., Brown, F., Crowther, J., Mould, A.P., Humphries, M.J., Francis, M.J., Clarke, B., Brown, A.L., and Rowlands, D. (1996). Chimeric hepatitis B virus core particles as probes for studying peptide-integrin interactions. J Virol **70,** 4045-4052.

Chesnut, J.D., Baytan, A.R., Russell, M., Chang, M.-P., Bernard, A., Maxwell, I.H., and Hoeffler, J.P. (1996). Selective isolation of transiently transfected cells from a mammalian cell population with vectors expressing a membrane anchored single-chain antibody. Journal of Immunological Methods **193,** 17-27.

Chilton, M.D., Drummond, M.H., Merio, D.J., Sciaky, D., Montoya, A.L., Gordon, M.P., and Nester, E.W. (1977). Stable incorporation of plasmid DNA into higher plant cells: the molecular basis of crown gall tumorigenesis. Cell **11,** 263-271.

Chu, M., Desvoyes, B., Turina, M., Noad, R., and Scholthof, H.B. (2000). Genetic Dissection of Tomato Bushy Stunt Virus p19-Protein-Mediated Host-Dependent Symptom Induction and Systemic Invasion. Virology **266,** 79-87.

Clarke, B.E., Newton, S.E., Carroll, A.R., Francis, M.J., Appleyard, G., Syred, A.D., Highfield, P.E., Rowlands, D.J., and Brown, F. (1987). Improved immunogenicity of a peptide epitope after fusion to hepatitis B core protein. Nature **330,** 381-384.

Conway, J.F., Cheng, N., Zlotnick, A., Stahl, S.J., Wingfield, P.T., Belnap, D.M., Kanngiesser, U., Noah, M., and Steven, A.C. (1998). Hepatitis B virus capsid: localization of the putative immunodominant loop (residues 78 to 83) on the capsid surface, and implications for the distinction between c and e-antigens. J Mol Biol **279,** 1111-1121.

Cooper, A., Tal, G., Lider, O., and Shaul, Y. (2005). Cytokine Induction by the Hepatitis B Virus Capsid in Macrophages Is Facilitated by Membrane Heparan Sulfate and Involves TLR2. The Journal of Immunology **175,** 3165-3176.

Crabbe, T., Weir, A.N., Walton, E.F., Brown, M.E., Sutton, C.W., Tretout, N., Bonnerjea, J., Lowe, P.A., and Yarranton, G.T. (1996). The Secretion of Active Recombinant Human Gastric Lipase bySaccharomyces cerevisiae. Protein Expression and Purification **7,** 229-236.

Crovari, P., Crovari, P.C., Petrilli, R.C., Icardi, G.C., and Bonanni, P. (1987). Immunogenicity of a yeast-derived hepatitis B vaccine (Engerix-B) in healthy young adults. Postgraduate medical journal **63 Suppl 2,** 161-164.

Crowther, R.A., Kiselev, N.A., Böttcher, B., Berriman, J.A., Borisova, G.P., Ose, V., and Pumpens, P. (1994). Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. Cell **77,** 943-950.

Cruz, S.S., Chapman, S., Roberts, A.G., Roberts, I.M., Prior, D.A., and Oparka, K.J. (1996). Assembly and movement of a plant virus carrying a green fluorescent protein overcoat. Proc Natl Acad Sci U S A **93,** 6286-6290.

D'Aoust, M.-A., Couture, M., Ors, F., Trépanier, S., Lavoie, P.-O., Dargis, M., Vézina, L.-P., and Landry, N. (2009). Recombinant Influenza Virus-like particles (VLPs) produced in transgenic plants expressing hemagglutinin, **WO 2009/076778.**

D'Aoust, M.A., Couture, M.M., Charland, N., Trepanier, S., Landry, N., Ors, F., and Vezina, L.P. (2010). The production of hemagglutinin-based virus-like particles in plants: a rapid, efficient and safe response to pandemic influenza. Plant Biotechnol J **8,** 607-619.

Dall'Acqua, W.F., Damschroder, M.M., Zhang, J., Woods, R.M., Widjaja, L., Yu, J., and Wu, H. (2005). Antibody humanization by framework shuffling. Methods **36,** 43-60.

Dalsgaard, K., Uttenthal, A., Jones, T.D., Xu, F., Merryweather, A., Hamilton, W.D., Langeveld, J.P., Boshuizen, R.S., Kamstrup, S., Lomonossoff, G.P., Porta, C., Vela, C., Casal, J.I., Meloen, R.H., and Rodgers, P.B. (1997). Plant-derived vaccine protects target animals against a viral disease. Nat Biotechnol **15,** 248-252.

De Filette, M., Min Jou, W., Birkett, A., Lyons, K., Schultz, B., Tonkyro, A., Resch, S., and Fiers, W. (2005). Universal influenza A vaccine: optimization of M2-based constructs. Virology **337,** 149-161.

de Framond, A.J., Barton, K.A., and Chilton, M.-D. (1983). Mini-Ti: A New Vector Strategy for Plant Genetic Engineering. Nat Biotech **1,** 262-269.

De Genst, E.J., Guilliams, T., Wellens, J., O'Day, E.M., Waudby, C.A., Meehan, S., Dumoulin, M., Hsu, S.T., Cremades, N., Verschueren, K.H., Pardon, E., Wyns, L., Steyaert, J., Christodoulou, J., and Dobson, C.M. (2010). Structure and properties of a complex of alpha-synuclein and a single-domain camelid antibody. J Mol Biol **402,** 326-343.

Denis, J., Acosta-Ramirez, E., Zhao, Y., Hamelin, M.E., Koukavica, I., Baz, M., Abed, Y., Savard, C., Pare, C., Lopez Macias, C., Boivin, G., and Leclerc, D. (2008). Development of a universal influenza A vaccine based on the M2e peptide fused to the papaya mosaic virus (PapMV) vaccine platform. Vaccine **26,** 3395-3403.

Desmyter, A., Decanniere, K., Muyldermans, S., and Wyns, L. (2001). Antigen Specificity and High Affinity Binding Provided by One Single Loop of a Camel Single-domain Antibody. Journal of Biological Chemistry **276,** 26285-26290.

Dhanasooraj, D., Kumar, R.A., and Mundayoor, S. (2013). Vaccine delivery system for tuberculosis based on nano-sized hepatitis B virus core protein particles. International journal of nanomedicine **8,** 835-843.

Djender, S., Beugnet, A., Schneider, A., and de Marco, A. (2014). The Biotechnological Applications of Recombinant Single-Domain Antibodies are Optimized by the C-Terminal Fusion to the EPEA Sequence (C Tag). Antibodies **3,** 182-191.

Dolk, E., van der Vaart, M., Lutje Hulsik, D., Vriend, G., de Haard, H., Spinelli, S., Cambillau, C., Frenken, L., and Verrips, T. (2005). Isolation of Llama Antibody Fragments for Prevention of Dandruff by Phage Display in Shampoo. Applied and Environmental Microbiology **71,** 442-450.

Dorokhov, Y.L., Skulachev, M.V., Ivanov, P.A., Zvereva, S.D., Tjulkina, L.G., Merits, A., Gleba, Y.Y., Hohn, T., and Atabekov, J.G. (2002). Polypurine (A)-rich sequences promote cross-kingdom conservation of internal ribosome entry. Proc Natl Acad Sci U S A **99,** 5301-5306.

Du, C., Chan, W.C., McKeithan, T.W., and Nickerson, K.W. (2005). Surface display of recombinant proteins on Bacillus thuringiensis spores. Appl Environ Microbiol **71,** 3337-3341.

Dugdale, B., Mortimer, C.L., Kato, M., James, T.A., Harding, R.M., and Dale, J.L. (2013). In plant activation: an inducible, hyperexpression platform for recombinant protein production in plants. The Plant cell **25,** 2429-2443.

Dyson, M.R., and Murray, K. (1995). Selection of peptide inhibitors of interactions involved in complex protein assemblies: association of the core and surface antigens of hepatitis B virus. Proceedings of the National Academy of Sciences **92,** 2194-2198.

Ellis, R.W., and Gerety, R.J. (1989). Plasma-derived and yeast-derived hepatitis B vaccines. American journal of infection control **17,** 181-189.

Eriksson, M., Andreasson, K., Weidmann, J., Lundberg, K., Tegerstedt, K., Dalianis, T., and Ramqvist, T. (2011). Murine polyomavirus virus-like particles carrying full-length human PSA protect BALB/c mice from outgrowth of a PSA expressing tumor. PLoS One **6,** e23828.

Feldmann, K.A., and Marks, D.M. (1987). Agrobacterium-mediated transformation of germinating seeds of Arabidopsis thaliana: A non-tissue culture approach. Mole Gen Genet **208,** 1-9.

Forsman, A., Beirnaert, E., Aasa-Chapman, M.M.I., Hoorelbeke, B., Hijazi, K., Koh, W., Tack, V., Szynol, A., Kelly, C., McKnight, Á., Verrips, T., Haard, H.d., and Weiss, R.A. (2008). Llama Antibody Fragments with Cross-Subtype Human Immunodeficiency Virus Type 1 (HIV-1)-Neutralizing Properties and High Affinity for HIV-1 gp120. Journal of Virology **82,** 12069-12081.

Fraley, R.T., Rogers, S.G., Horsch, R.B., Sanders, P.R., Flick, J.S., Adams, S.P., Bittner, M.L., Brand, L.A., Fink, C.L., Fry, J.S., Galluppi, G.R., Goldberg, S.B., Hoffmann, N.L., and Woo, S.C. (1983). Expression of bacterial genes in plant cells. Proc Natl Acad Sci U S A **80,** 4803-4807.

Francis, M.J., Hastings, G.Z., Brown, A.L., Grace, K.G., Rowlands, D.J., Brown, F., and Clarke, B.E. (1990). Immunological properties of hepatitis B core antigen fusion proteins. Proc Natl Acad Sci U S A **87,** 2545-2549.

Gallina, A., Bonelli, F., Zentilin, L., Rindi, G., Muttini, M., and Milanesi, G. (1989). A recombinant hepatitis B core antigen polypeptide with the protamine-like domain deleted self-assembles into capsid particles but fails to bind nucleic acids. Journal of Virology **63,** 4645-4652.

Garabagi, F., McLean, M.D., and Hall, J.C. (2012). Transient and stable expression of antibodies in Nicotiana species. Methods in molecular biology (Clifton, N.J.) **907,** 389-408.

Gehin, A., Gilbert, R., Stuart, D., and Rowlands, D. (2004). Hepatitis B core antigen fusion proteins, **US 2001/0223965 A1.**

Geisler, K., Hughes, R.K., Sainsbury, F., Lomonossoff, G.P., Rejzek, M., Fairhurst, S., Olsen, C.E., Motawia, M.S., Melton, R.E., Hemmings, A.M., Bak, S., and Osbourn, A. (2013). Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. Proc Natl Acad Sci U S A **110,** E3360-3367.

Gleba, Y., Klimyuk, V., and Marillonnet, S. (2007). Viral vectors for the expression of proteins in plants. Current opinion in biotechnology **18,** 134-141.

Goldbach, R.W., and Wellink, J. (1996). Polyhedral Virions and Bipartite RNA Genomes. In The Plant Viruses, B.D. Harrison and A.F. Murant, eds (New York: Plenum), pp. 35-76.

Golden, A., Austen, D.A., van Schravendijk, M.R., Sullivan, B.J., Kawasaki, E.S., and Osburne, M.S. (1998). Effect of promoters and signal sequences on the production of secreted HIV-1 gp120 protein in the baculovirus system. Protein Expr Purif **14,** 8-12.

Gomez, E., Lucero, M.S., Chimeno Zoth, S., Carballeda, J.M., Gravisaco, M.J., and Berinstein, A. (2013). Transient expression of VP2 in Nicotiana benthamiana and its use as a plant-based vaccine against infectious bursal disease virus. Vaccine **31,** 2623-2627.

Gorantala, J., Grover, S., Rahi, A., Chaudhary, P., Rajwanshi, R., Sarin, N.B., and Bhatnagar, R. (2014). Generation of protective immune response against anthrax by oral immunization with protective antigen plant-based vaccine. J Biotechnol **176,** 1-10.

Grene, E., Mezule, G., Borisova, G., Pumpens, P., Bentwich, Z., and Arnon, R. (1997). Relationship between antigenicity and immunogenicity of chimeric hepatitis B virus core particles carrying HIV type 1 epitopes. AIDS Res Hum Retroviruses **13,** 41-51.

Greve, H.D., Leemans, J., Hernalsteens, J.-P., Thia-Toong, L., Beuckeleer, M.D., Willmitzer, L., Otten, L., Montagu, M.V., and Schell, J. (1982). Regeneration of normal and fertile plants that express octopine synthase, from tobacco crown galls after deletion of tumour-controlling functions. Nature **300,** 752-755.

Grimsley, N., Hohn, B., Hohn, T., and Walden, R. (1986). "Agroinfection," an alternative route for viral infection of plants by using the Ti plasmid. Proc Natl Acad Sci U S A **83,** 3282-3286.

Grohs, B.M., Niu, Y., Veldhuis, L.J., Trabelsi, S., Garabagi, F., Hassell, J.A., McLean, M.D., and Hall, J.C. (2010). Plant-produced trastuzumab inhibits the growth of HER2 positive cancer cells. Journal of agricultural and food chemistry **58,** 10056-10063.

Groppelli, E., Belsham, G.J., and Roberts, L.O. (2007). Identification of minimal sequences of the Rhopalosiphum padi virus 5' untranslated region required for internal initiation of protein synthesis in mammalian, plant and insect translation systems. J Gen Virol **88,** 1583-1588.

Han, J.A., Kang, Y.J., Shin, C., Ra, J.S., Shin, H.H., Hong, S.Y., Do, Y., and Kang, S. (2014). Ferritin protein cage nanoparticles as versatile antigen delivery nanoplatforms for dendritic cell (DC)-based vaccine development. Nanomedicine : nanotechnology, biology, and medicine **10,** 561-569.

Haq, T.A., Mason, H.S., Clements, J.D., and Arntzen, C.J. (1995). Oral immunization with a recombinant bacterial antigen produced in transgenic plants. Science **268,** 714-716.

Harmsen, M.M., and De Haard, H.J. (2007). Properties, production, and applications of camelid single-domain antibody fragments. Applied microbiology and biotechnology **77,** 13-22.

Hayashi, N., Kipriyanov, S., Fuchs, P., Welschof, M., Dörsam, H., and Little, M. (1995). A single expression system for the display, purification and conjugation of single-chain antibodies. Gene **160,** 129-130.

Herrera-Estrella, L., Block, M.D., Messens, E., Hernalsteens, J.P., Montagu, M.V., and Schell, J. (1983). Chimeric genes as dominant selectable markers in plant cells. EMBO J **2,** 987-995.

Hiatt, A., Cafferkey, R., and Bowdish, K. (1989). Production of antibodies in transgenic plants. Nature **342,** 76-78.

Hilleman, M. (2011). Three Decades of Hepatitis Vaccinology in Historic Perspective. A Paradigm of Successful Pursuits. In History of Vaccine Development, S.A. Plotkin, ed (Springer New York), pp. 233-246.

Hilleman, M.R. (1987). Yeast recombinant hepatitis B vaccine. Infection **15,** 3-7.

Hoekema, A., Hirsch, P.R., Hooykaas, P.J.J., and Schilperoort, R.A. (1983). A binary plant vector strategy based on separation of vir- and T-region of the Agrobacterium tumefaciens Ti-plasmid. Nature **303,** 179-180.

Horsch, R.B., and Klee, H.J. (1986). Rapid assay of foreign gene expression in leaf discs transformed by Agrobacterium tumefaciens: Role of T-DNA borders in the transfer process. Proc Natl Acad Sci U S A **83,** 4428-4432.

Huang, Y., Liang, W., Wang, Y., Zhou, Z., Pan, A., Yang, X., Huang, C., Chen, J., and Zhang, D. (2005). Immunogenicity of the epitope of the foot-and-mouth disease virus fused with a hepatitis B core protein as expressed in transgenic tobacco. Viral immunology **18,** 668-677.

Huang, Z., Santi, L., LePore, K., Kilbourne, J., Arntzen, C.J., and Mason, H.S. (2006). Rapid, high-level production of hepatitis B core antigen in plant leaf and its immunogenicity in mice. Vaccine **24,** 2506-2513.

Ismaili, A., Jalali-javaran, M., Rasaee, M.J., Rahbarizadeh, F., Forouzandeh-moghadam, M., and Memari, H.R. (2007). Production and characterization of anti-(mucin MUC1) single-domain antibody in tobacco (Nicotiana tabacum cultivar Xanthi). Biotechnology and Applied Biochemistry **047,** 11-19.

Janeway, C.A., Jr. (1989). Approaching the asymptote? Evolution and revolution in immunology. Cold Spring Harbor symposia on quantitative biology **54 Pt 1,** 1-13.

Janssens, M.E., Geysen, D., Broos, K., De Goeyse, I., Robbens, J., Van Petegem, F., Timmermans, J.P., and Guisez, Y. (2010). Folding properties of the hepatitis B core as a carrier protein for vaccination research. Amino Acids **38,** 1617-1626.

Jegerlehner, A., Tissot, A., Lechner, F., Sebbel, P., Erdmann, I., Kundig, T., Bachi, T., Storni, T., Jennings, G., Pumpens, P., Renner, W.A., and Bachmann, M.F. (2002). A molecular assembly system that renders antigens of choice highly repetitive for induction of protective B cell responses. Vaccine **20,** 3104-3112.

Jobling, S.A., Jarman, C., Teh, M.-M., Holmberg, N., Blake, C., and Verhoeyen, M.E. (2003). Immunomodulation of enzyme function in plants by single-domain antibody fragments. Nat Biotech **21,** 77-80.

Kanagarajan, S., Muthusamy, S., Gliszczyńska, A., Lundgren, A., and Brodelius, P. (2012). Functional expression and characterization of sesquiterpene synthases from Artemisia annua L. using transient expression system in Nicotiana benthamiana. Plant Cell Rep **31,** 1309-1319.

Kanekiyo, M., Wei, C.J., Yassine, H.M., McTamney, P.M., Boyington, J.C., Whittle, J.R., Rao, S.S., Kong, W.P., Wang, L., and Nabel, G.J. (2013). Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. Nature **499,** 102-106.

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

Kazaks, A., Lachmann, S., Koletzki, D., Petrovskis, I., Dislers, A., Ose, V., Skrastina, D., Gelderblom, H.R., Lundkvist, A., Meisel, H., Borisova, G., Kruger, D.H., Pumpens, P., and Ulrich, R. (2002). Stop codon insertion restores the particle formation ability of hepatitis B virus core-hantavirus nucleocapsid protein fusions. Intervirology **45,** 340-349.

Kessans, S.A., Linhart, M.D., Matoba, N., and Mor, T. (2013). Biological and biochemical characterization of HIV-1 Gag/dgp41 virus-like particles expressed in Nicotiana benthamiana. Plant Biotechnol J **11,** 681-690.

Kirchhofer, A., Helma, J., Schmidthals, K., Frauer, C., Cui, S., Karcher, A., Pellis, M., Muyldermans, S., Casas-Delucchi, C.S., Cardoso, M.C., Leonhardt, H., Hopfner, K.-P., and Rothbauer, U. (2010). Modulation of protein properties in living cells using nanobodies. Nat Struct Mol Biol **17,** 133-138.

Kitidee, K., Nangola, S., Gonzalez, G., Boulanger, P., Tayapiwatana, C., and Hong, S.-S. (2010). Baculovirus display of single chain antibody (scFv) using a novel signal peptide. BMC Biotechnology **10,** 80.

Koletzki, D., Biel, S.S., Meisel, H., Nugel, E., Gelderblom, H.R., Kruger, D.H., and Ulrich, R. (1999). HBV core particles allow the insertion and surface exposure of the entire potentially protective region of Puumala hantavirus nucleocapsid protein. Biological chemistry **380,** 325-333.

Koletzki, D., Lundkvist, A., Sjolander, K.B., Gelderblom, H.R., Niedrig, M., Meisel, H., Kruger, D.H., and Ulrich, R. (2000). Puumala (PUU) hantavirus strain differences and insertion positions in the hepatitis B virus core antigen influence B-cell immunogenicity and protective potential of core-derived particles. Virology **276,** 364-375.

Kratz, P.A., Bottcher, B., and Nassal, M. (1999). Native display of complete foreign protein domains on the surface of hepatitis B virus capsids. Proc Natl Acad Sci U S A **96,** 1915-1920.

Kretzschmar, T., Aoustin, L., Zingel, O., Marangi, M., Vonach, B., Towbin, H., and Geiser, M. (1996). High-level expression in insect cells and purification of secreted monomeric single-chain Fv antibodies. Journal of Immunological Methods **195,** 93-101.

Krugman, S., and Davidson, M. (1987). Hepatitis B vaccine: prospects for duration of immunity. The Yale journal of biology and medicine **60,** 333-339.

Kushnir, N., Streatfield, S.J., and Yusibov, V. (2012). Virus-like particles as a highly efficient vaccine platform: diversity of targets and production systems and advances in clinical development. Vaccine **31,** 58-83.

Kusnadi, A.R., Nikolov, Z.L., and Howard, J.A. (1997). Production of recombinant proteins in transgenic plants: Practical considerations. Biotechnology and Bioengineering **56,** 473-484.

Lachmann, S., Meisel, H., Muselmann, C., Koletzki, D., Gelderblom, H.R., Borisova, G., Kruger, D.H., Pumpens, P., and Ulrich, R. (1999). Characterization of potential insertion sites in the core antigen of hepatitis B virus by the use of a short-sized model epitope. Intervirology **42,** 51-56.

Larsen, J., and Curtis, W. (2012). RNA viral vectors for improved Agrobacterium-mediated transient expression of heterologous proteins in Nicotiana benthamiana cell suspensions and hairy roots. BMC Biotechnology **12,** 21.

Lee, L.A., and Wang, Q. (2006). Adaptations of nanoscale viruses and other protein cages for medical applications. Nanomedicine : nanotechnology, biology, and medicine **2,** 137-149.

Lee, S.W., Markham, P.F., Coppo, M.J., Legione, A.R., Markham, J.F., Noormohammadi, A.H., Browning, G.F., Ficorilli, N., Hartley, C.A., and Devlin, J.M. (2012). Attenuated vaccines can recombine to form virulent field viruses. Science **337,** 188.

Lico, C., Chen, Q., and Santi, L. (2008). Viral vectors for production of recombinant proteins in plants. Journal of Cellular Physiology **216,** 366-377.

Liu, L., and Lomonossoff, G.P. (2002). Agroinfection as a rapid method for propagating Cowpea mosaic virus-based constructs. Journal of Virological Methods **105,** 343-348.

Liu, L., Grainger, J., Cañizares, M.C., Angell, S.M., and Lomonossoff, G.P. (2004). Cowpea mosaic virus RNA-1 acts as an amplicon whose effects can be counteracted by a RNA-2-encoded suppressor of silencing. Virology **323,** 37-48.

Love, A.J., Chapman, S.N., Matic, S., Noris, E., Lomonossoff, G., and Taliansky, M. (2012). In planta production of a candidate vaccine against bovine papillomavirus type 1. Planta **236,** 1305-1313.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. Journal of structural biology **128,** 82-97.

Ma, J.K., Hiatt, A., Hein, M., Vine, N.D., Wang, F., Stabila, P., van Dolleweerd, C., Mostov, K., and Lehner, T. (1995). Generation and assembly of secretory antibodies in plants. Science **268,** 716-719.

Maass, D.R., Sepulveda, J., Pernthaner, A., and Shoemaker, C.B. (2007). Alpaca (Lama pacos) as a convenient source of recombinant camelid heavy chain antibodies (VHHs). J Immunol Methods **324,** 13-25.

Maclean, J., Koekemoer, M., Olivier, A., Stewart, D., Hitzeroth, I., Rademacher, T., Fischer, R., Williamson, A., and Rybicki, E. (2007). Optimization of human papillomavirus type 16 (HPV-16) L1 expression in plants: comparison of the suitability of different HPV-16 L1 gene variants and different cell-compartment localization. J Gen Virol **88,** 1460 - 1469.

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

Marillonnet, S., Giritch, A., Gils, M., Kandzia, R., Klimyuk, V., and Gleba, Y. (2004). In planta engineering of viral RNA replicons: efficient assembly by recombination of DNA modules delivered by Agrobacterium. Proc Natl Acad Sci U S A **101,** 6852-6857.

Marks, J.D., Griffiths, A.D., Malmqvist, M., Clackson, T.P., Bye, J.M., and Winter, G. (1992). By-passing immunization: building high affinity human antibodies by chain shuffling. Bio/technology (Nature Publishing Company) **10,** 779-783.

Marshall, A. (2012). Existing agbiotech traits continue global march. Nat Biotechnol **30,** 207.

Martin, J., Odoom, K., Tuite, G., Dunn, G., Hopewell, N., Cooper, G., Fitzharris, C., Butler, K., Hall, W.W., and Minor, P.D. (2004). Long-term excretion of vaccine-derived poliovirus by a healthy child. J Virol **78,** 13839-13847.

Marton, L., Wullems, G.J., Molendijk, L., and Schilperoort, R.A. (1979). In vitro transformation of cultured cells from Nicotiana tabacum by Agrobacterium tumefaciens. Nature **277,** 129-131.

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

Mason, H.S., Lam, D.M., and Arntzen, C.J. (1992). Expression of hepatitis B surface antigen in transgenic plants. Proceedings of the National Academy of Sciences **89,** 11745-11749.

Matić, S., Masenga, V., Poli, A., Rinaldi, R., Milne, R.G., Vecchiati, M., and Noris, E. (2012). Comparative analysis of recombinant Human Papillomavirus 8 L1 production in plants by a variety of expression systems and purification methods. Plant Biotechnology Journal **10,** 410-421.

Mazeike, E., Gedvilaite, A., and Blohm, U. (2012). Induction of insert-specific immune response in mice by hamster polyomavirus VP1 derived virus-like particles carrying LCMV GP33 CTL epitope. Virus Res **163,** 2-10.

McAleer, W.J., Buynak, E.B., Maigetter, R.Z., Wampler, D.E., Miller, W.J., and Hilleman, M.R. (1992). Human hepatitis B vaccine from recombinant yeast. 1984. Biotechnology **24,** 500-502.

McLain, L., Porta, C., Lomonossoff, G.P., Durrani, Z., and Dimmock, N.J. (1995). Human immunodeficiency virus type 1-neutralizing antibodies raised to a glycoprotein 41 peptide expressed on the surface of a plant virus. AIDS Res Hum Retroviruses **11,** 327-334.

Mechtcheriakova, I.A., Eldarov, M.A., Nicholson, L., Shanks, M., Skryabin, K.G., and Lomonossoff, G.P. (2006). The use of viral vectors to produce hepatitis B virus core particles in plants. Journal of Virological Methods **131,** 10-15.

Meijer, H.A., and Thomas, A.A.M. (2002). Control of eukaryotic protein synthesis by upstream open reading frames in the 5'-untranslated region of an mRNA. Biochem. J. **367,** 1-11.

Meshcheriakova, I.A., El'darov, M.A., Beales, L., Skriabin, K.G., and Lomonossoff, G.P. (2008). [Production of hepatitis B virus core particles protein in plants, by using cowpea mosaic virus-based vector]. Voprosy virusologii **53,** 15-20.

Middelberg, A.P., Rivera-Hernandez, T., Wibowo, N., Lua, L.H., Fan, Y., Magor, G., Chang, C., Chuan, Y.P., Good, M.F., and Batzloff, M.R. (2011). A microbial platform for rapid and low-cost virus-like particle and capsomere vaccines. Vaccine **29,** 7154-7162.

Minor, P.D., John, A., Ferguson, M., and Icenogle, J.P. (1986). Antigenic and molecular evolution of the vaccine strain of type 3 poliovirus during the period of excretion by a primary vaccinee. J Gen Virol **67 (Pt 4),** 693-706.

Minskaia, E., and Ryan, M.D. (2013). Protein coexpression using FMDV 2A: effect of "linker" residues. BioMed research international **2013,** 291730.

Minskaia, E., Nicholson, J., and Ryan, M.D. (2013). Optimisation of the foot-and-mouth disease virus 2A co-expression system for biomedical applications. BMC Biotechnol **13,** 67.

Miyamoto, K., Shimizu, T., Lin, F., Sainsbury, F., Thuenemann, E., Lomonossoff, G., Nojiri, H., Yamane, H., and Okada, K. (2012). Identification of an E-box motif responsible for the expression of jasmonic acid-induced chitinase gene OsChia4a in rice. Journal of Plant Physiology **169,** 621-627.

Momany, C., Kovari, L.C., Prongay, A.J., Keller, W., Gitti, R.K., Lee, B.M., Gorbalenya, A.E., Tong, L., McClure, J., Ehrlich, L.S., Summers, M.F., Carter, C., and Rossmann, M.G. (1996). Crystal structure of dimeric HIV-1 capsid protein. Nature structural biology **3,** 763-770.

Monger, W., Alamillo, J.M., Sola, I., Perrin, Y., Bestagno, M., Burrone, O.R., Sabella, P., Plana-Duran, J., Enjuanes, L., Garcia, J.A., and Lomonossoff, G.P. (2006). An antibody derivative expressed from viral vectors passively immunizes pigs against transmissible gastroenteritis virus infection when supplied orally in crude plant extracts. Plant Biotechnol J **4,** 623-631.

Monie, A., Hung, C.F., Roden, R., and Wu, T.C. (2008). Cervarix: a vaccine for the prevention of HPV 16, 18-associated cervical cancer. Biologics : targets & therapy **2,** 97-105.

Montague, N.P., Thuenemann, E.C., Saxena, P., Saunders, K., Lenzi, P., and Lomonossoff, G.P. (2011). Recent advances of Cowpea mosaic virus-based particle technology. Hum Vaccin **7**.

Mortimer, E., Maclean, J.M., Mbewana, S., Buys, A., Williamson, A.L., Hitzeroth, II, and Rybicki, E.P. (2012). Setting up a platform for plant-based influenza virus vaccine production in South Africa. BMC Biotechnol **12,** 14.

Muyldermans, S. (2001). Single domain camel antibodies: current status. J Biotechnol **74,** 277-302.

Muyldermans, S., and Lauwereys, M. (1999). Unique single-domain antigen binding fragments derived from naturally occurring camel heavy-chain antibodies. Journal of molecular recognition : JMR **12,** 131-140.

Muyldermans, S., Cambillau, C., and Wyns, L. (2001). Recognition of antigens by single-domain antibody fragments: the superfluous luxury of paired domains. Trends in Biochemical Sciences **26,** 230-235.

O'Brien, G.J., Bryant, C.J., Voogd, C., Greenberg, H.B., Gardner, R.C., and Bellamy, A.R. (2000). Rotavirus VP6 expressed by PVX vectors in Nicotiana benthamiana coats PVX rods and also assembles into viruslike particles. Virology **270,** 444-453.

Ono, Y., Onda, H., Sasada, R., Igarashi, K., Sugino, Y., and Nishioka, K. (1983). The complete nucleotide sequences of the cloned hepatitis B virus DNA; subtype adr and adw. Nucleic Acids Res **11,** 1747-1757.

Pastori, C., Tudor, D., Diomede, L., Drillet, A.S., Jegerlehner, A., Rohn, T.A., Bomsel, M., and Lopalco, L. (2012). Virus like particle based strategy to elicit HIV-protective antibodies to the alpha-helic regions of gp41. Virology **431,** 1-11.

Pedelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T.C., and Waldo, G.S. (2006). Engineering and characterization of a superfolder green fluorescent protein. Nat Biotech **24,** 79-88.

Peng, K.-W., Donovan, K.A., Schneider, U., Cattaneo, R., Lust, J.A., and Russell, S.J. (2003). Oncolytic measles viruses displaying a single-chain antibody against CD38, a myeloma cell marker. Blood **101,** 2557-2562.

Pérez-Filgueira, D.M., Brayfield, B.P., Phiri, S., Borca, M.V., Wood, C., and Morris, T.J. (2004). Preserved antigenicity of HIV-1 p24 produced and purified in high yields from plants inoculated with a tobacco mosaic virus (TMV)-derived vector. Journal of Virological Methods **121,** 201-208.

Perez Aguirreburualde, M.S., Gomez, M.C., Ostachuk, A., Wolman, F., Albanesi, G., Pecora, A., Odeon, A., Ardila, F., Escribano, J.M., Dus Santos, M.J., and Wigdorovitz, A. (2013). Efficacy of a BVDV subunit vaccine produced in alfalfa transgenic plants. Vet Immunol Immunopathol **151,** 315-324.

Peyret, H., and Lomonossoff, G.P. (2013). The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. Plant Mol Biol **83,** 51-58.

Pietrzak, M., Shillito, R.D., Hohn, T., and Potrykus, I. (1986). Expression in plants of two bacterial antibiotic resistance genes after protoplast transformation with a new plant expression vector. Nucleic Acids Res **14,** 5857-5868.

Pleckaityte, M., Zvirbliene, A., Sezaite, I., and Gedvilaite, A. (2011). Production in yeast of pseudotype virus-like particles harboring functionally active antibody fragments neutralizing the cytolytic activity of vaginolysin. Microbial Cell Factories **10,** 109.

Plummer, E.M., and Manchester, M. (2010). Viral nanoparticles and virus-like particles: platforms for contemporary vaccine design. Wiley interdisciplinary reviews. Nanomedicine and nanobiotechnology.

Pogue, G.P., Lindbo, J.A., Garger, S.J., and Fitzmaurice, W.P. (2002). MAKING AN ALLY FROM AN ENEMY: Plant Virology and the New Agriculture. Annual Review of Phytopathology **40,** 45-74.

Porta, C., Spall, V.E., Loveland, J., Johnson, J.E., Barker, P.J., and Lomonossoff, G.P. (1994). Development of cowpea mosaic virus as a high-yielding system for the presentation of foreign peptides. Virology **202,** 949-955.

Pumpens, P., and Grens, E. (1999). Hepatitis B core particles as a universal display model: a structure-function basis for development. FEBS Lett **442,** 1-6.

Pumpens, P., and Grens, E. (2001). HBV core particles as a carrier for B cell/T cell epitopes. Intervirology **44,** 98-114.

Ravin, N.V., Kotlyarov, R.Y., Mardanova, E.S., Kuprianov, V.V., Migunov, A.I., Stepanova, L.A., Tsybalova, L.M., Kiselev, O.I., and Skryabin, K.G. (2012). Plant-produced recombinant influenza vaccine based on virus-like HBc particles carrying an extracellular domain of M2 protein. Biochemistry. Biokhimiia **77,** 33-40.

Riechmann, L., and Muyldermans, S. (1999). Single domain antibodies: comparison of camel VH and camelised human VH domains. Journal of Immunological Methods **231,** 25-38.

Rohll, J.B., Holness, C.L., Lomonossoff, G.P., and Maule, A.J. (1993). 3'-Terminal Nucleotide Sequences Important for the Accumulation of Cowpea Mosaic Virus M-RNA. Virology **193,** 672-679.

Rosenberg, Y., Sack, M., Montefiori, D., Forthal, D., Mao, L., Hernandez-Abanto, S., Urban, L., Landucci, G., Fischer, R., and Jiang, X. (2013). Rapid high-level production of functional HIV broadly neutralizing monoclonal antibodies in transient plant expression systems. PLoS One **8,** e58724.

Rossi, L., Hohn, B., and Tinland, B. (1996). Integration of complete transferred DNA units is dependent on the activity of virulence E2 protein of Agrobacterium tumefaciens. Proc Natl Acad Sci U S A **93,** 126-130.

Roy, P., and Noad, R. (2008). Virus-like particles as a vaccine delivery system: myths and facts. Hum Vaccin **4,** 5-12.

Ruedl, C., Schwarz, K., Jegerlehner, A., Storni, T., Manolova, V., and Bachmann, M.F. (2005). Virus-like particles as carriers for T-cell epitopes: limited inhibition of T-cell priming by carrier-specific antibodies. J Virol **79,** 717-724.

Sainsbury, F., and Lomonossoff, G.P. (2008). Extremely High-Level and Rapid Transient Protein Production in Plants without the Use of Viral Replication. Plant Physiology **148,** 1212-1218.

Sainsbury, F., Liu, L., and Lomonossoff, G. (2009a). Cowpea Mosaic Virus-Based Systems for the Expression of Antigens and Antibodies in Plants. In Recombinant Proteins From Plants, L. Faye and V. Gomord, eds (Humana Press), pp. 25-39.

Sainsbury, F., Thuenemann, E.C., and Lomonossoff, G.P. (2009b). pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. Plant Biotechnology Journal **7,** 682-693.

Sainsbury, F., Lavoie, P.-O., D'Aoust, M.-A., Vézina, L.-P., and Lomonossoff, G.P. (2008). Expression of multiple proteins using full-length and deleted versions of cowpea mosaic virus RNA-2. Plant Biotechnology Journal **6,** 82-92.

Sainsbury, F., Saxena, P., Geisler, K., Osbourn, A., and Lomonossoff, G.P. (2012a). Methods in Enzymology. (Elsevier).

Sainsbury, F., Saxena, P., Geisler, K., Osbourn, A., and Lomonossoff, G.P. (2012b). Chapter Nine - Using a Virus-Derived System to Manipulate Plant Natural Product Biosynthetic Pathways. In Methods in Enzymology, A.H. David, ed (Academic Press), pp. 185-202.

Sainsbury, F., Sack, M., Stadlmann, J., Quendler, H., Fischer, R., and Lomonossoff, G.P. (2010). Rapid Transient Production in Plants by Replicating and Non-Replicating Vectors Yields High Quality Functional Anti-HIV Antibody. PLoS ONE **5,** e13976.

Samac, D.A., Hironaka, C.M., Yallaly, P.E., and Shah, D.M. (1990). Isolation and Characterization of the Genes Encoding Basic and Acidic Chitinase in Arabidopsis thaliana. Plant Physiol **93,** 907-914.

Saxena, P. (2012). Development of RNA-free particles of Cowpea mosaic virus for applications in Nanotechnology. In John Innes Centre (Norwich, UK: University of East Anglia), pp. 167.

Saxena, P., Hsieh, Y.-C., Alvarado, V.Y., Sainsbury, F., Saunders, K., Lomonossoff, G.P., and Scholthof, H.B. (2011). Improved foreign gene expression in plants using a virus-encoded suppressor of RNA silencing modified to be developmentally harmless. Plant Biotechnology Journal **9,** 703-712.

Scheres, S.H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. Journal of structural biology **180,** 519-530.

Schmiedl, A., Breitling, F., Winter, C.H., Queitsch, I., and Dübel, S. (2000). Effects of unpaired cysteines on yield, solubility and activity of different recombinant antibody constructs expressed in E. coli. Journal of Immunological Methods **242,** 101-114.

Schob, H., Kunz, C., and Meins, F., Jr. (1997). Silencing of transgenes introduced into leaves by agroinfiltration: a simple, rapid method for investigating sequence requirements for gene silencing. Molecular & general genetics : MGG **256,** 581-585.

Schödel, F., Moriarty, A.M., Peterson, D.L., Zheng, J.A., Hughes, J.L., Will, H., Leturcq, D.J., McGee, J.S., and Milich, D.R. (1992). The position of heterologous epitopes inserted in hepatitis B virus core particles determines their immunogenicity. J Virol **66,** 106-114.

Schouten, A., Roosien, J., de Boer, J.M., Wilmink, A., Rosso, M.-N., Bosch, D., Stiekema, W.J., Gommers, F.J., Bakker, J., and Schots, A. (1997). Improving scFv antibody expression levels in the plant cytosol. FEBS letters **415,** 235-241.

Scotti, N., and Rybicki, E.P. (2013). Virus-like particles produced in plants as potential vaccines. Expert Rev Vaccines **12,** 211-224.

Sela-Culang, I., Kunik, V., and Ofran, Y. (2013). The structural basis of antibody-antigen recognition. Frontiers in immunology **4,** 302.

Shank-Retzlaff, M.L., Zhao, Q., Anderson, C., Hamm, M., High, K., Nguyen, M., Wang, F., Wang, N., Wang, B., Wang, Y., Washabaugh, M., Sitrin, R., and Shi, L. (2006). Evaluation of the thermal stability of Gardasil. Hum Vaccin **2,** 147-154.

Skamel, C., Ploss, M., Bottcher, B., Stehle, T., Wallich, R., Simon, M.M., and Nassal, M. (2006). Hepatitis B virus capsid-like particles can display the complete, dimeric outer surface protein C and stimulate production of protective antibody responses against Borrelia burgdorferi infection. J Biol Chem **281,** 17474-17481.

Skrastina, D., Petrovskis, I., Petraityte, R., Sominskaya, I., Ose, V., Lieknina, I., Bogans, J., Sasnauskas, K., and Pumpens, P. (2013). Chimeric derivatives of hepatitis B virus core particles carrying major epitopes of the rubella virus E1 glycoprotein. Clinical and vaccine immunology : CVI **20,** 1719-1728.

Smith, G.P., and Petrenko, V.A. (1997). Phage Display. Chemical Reviews **97,** 391-410.

Smolenska, L., Roberts, I.M., Learmonth, D., Porter, A.J., Harris, W.J., Wilson, T.M., and Santa Cruz, S. (1998). Production of a functional single chain antibody attached to the surface of a plant virus. FEBS Lett **441,** 379-382.

Sominskaya, I., Skrastina, D., Dislers, A., Vasiljev, D., Mihailova, M., Ose, V., Dreilina, D., and Pumpens, P. (2010). Construction and immunological evaluation of multivalent hepatitis B virus (HBV) core virus-like particles carrying HBV and HCV epitopes. Clinical and vaccine immunology : CVI **17,** 1027-1033.

Stahl, S.J., and Murray, K. (1989). Immunogenicity of peptide fusions to hepatitis B virus core antigen. Proc Natl Acad Sci U S A **86,** 6283-6287.

Strokappe, N., Szynol, A., Aasa-Chapman, M., Gorlani, A., Forsman Quigley, A., Hulsik, D.L., Chen, L., Weiss, R., de Haard, H., and Verrips, T. (2012). Llama antibody fragments recognizing various epitopes of the CD4bs neutralize a broad range of HIV-1 subtypes A, B and C. PLoS One **7,** e33298.

Sugiyama, Y., Hamamoto, H., Takemoto, S., Watanabe, Y., and Okada, Y. (1995). Systemic production of foreign peptides on the particle surface of tobacco mosaic virus. FEBS Lett **359,** 247-250.

Sun, Q.-Y., Ding, L.-W., Lomonossoff, G.P., Sun, Y.-B., Luo, M., Li, C.-Q., Jiang, L., and Xu, Z.-F. (2011). Improved expression and purification of recombinant human serum albumin from transgenic tobacco suspension culture. Journal of Biotechnology **155,** 164-172.

Teh, Y.-H., and Kavanagh, T. (2010). High-level expression of Camelid nanobodies in Nicotiana benthamiana. Transgenic Research **19,** 575-586.

Thuenemann, E.C. (2010). Virus-like particle production using Cowpea Mosaic Virus-based vectors. In John Innes Centre, Biological Chemistry (Norwich: University of East Anglia), pp. 693.

Thuenemann, E.C., Meyers, A.E., Verwey, J., Rybicki, E.P., and Lomonossoff, G.P. (2013a). A method for rapid production of heteromultimeric protein complexes in plants: assembly of protective bluetongue virus-like particles. Plant Biotechnol J **11,** 839-846.

Thuenemann, E.C., Lenzi, P., J. Love, A., Taliansky, M., Bécares, M., Zuñiga, S., Enjuanes, L., Zahmanova, G.G., Minkov, I.N., Matić, S., Noris, E., Meyers, A., Hattingh, A., Rybicki, E.P., Kiselev, O.I., Ravin, N.V., Eldarov, M.A., Skryabin, K.G., and Lomonossoff, G.P. (2013b). The use of transient expression systems for the rapid production of virus-like particles in plants. Current Pharmaceutical Design **19,** 5564-5573.

Tsuda, S., Yoshioka, K., Tanaka, T., Iwata, A., Yoshikawa, A., Watanabe, Y., and Okada, Y. (1998). Application of the Human Hepatitis B Virus Core Antigen from Transgenic Tobacco Plants for Serological Diagnosis. Vox Sanguinis **74,** 148-155.

Turpen, T.H., Reinl, S.J., Charoenvit, Y., Hoffman, S.L., Fallarme, V., and Grill, L.K. (1995). Malarial epitopes expressed on the surface of recombinant tobacco mosaic virus. Biotechnology (N Y) **13,** 53-57.

Ulrich, R., Lundkvist, A., Meisel, H., Koletzki, D., Sjolander, K.B., Gelderblom, H.R., Borisova, G., Schnitzler, P., Darai, G., and Kruger, D.H. (1998). Chimaeric HBV core particles carrying a defined segment of Puumala hantavirus nucleocapsid protein evoke protective immunity in an animal model. Vaccine **16,** 272-280.

Ulrich, R., Borisova, G.P., Gren, E., Berzin, I., Pumpen, P., Eckert, R., Ose, V., Siakkou, H., Gren, E.J., von Baehr, R., and et al. (1992). Immunogenicity of recombinant core particles of hepatitis B virus containing epitopes of human immunodeficiency virus 1 core antigen. Arch Virol **126,** 321-328.

Ulrich, R., Koletzki, D., Lachmann, S., Lundkvist, A., Zankl, A., Kazaks, A., Kurth, A., Gelderblom, H.R., Borisova, G., Meisel, H., and Kruger, D.H. (1999). New chimaeric hepatitis B virus core particles carrying hantavirus (serotype Puumala) epitopes: immunogenicity and protection against virus challenge. J Biotechnol **73,** 141-153.

Urban, J.H., Schneider, R.M., Compte, M., Finger, C., Cichutek, K., Álvarez-Vallina, L., and Buchholz, C.J. (2005). Selection of functional human antibodies from retroviral display libraries. Nucleic Acids Research **33,** e35.

Usha, R., Rohll, J.B., Spall, V.E., Shanks, M., Maule, A.J., Johnson, J.E., and Lomonossoff, G.P. (1993). Expression of an Animal Virus Antigenic Site on the Surface of a Plant Virus Particle. Virology **197,** 366-374.

Valenzuela, P., Medina, A., Rutter, W.J., Ammerer, G., and Hall, B.D. (1982). Synthesis and assembly of hepatitis B virus surface antigen particles in yeast. Nature **298,** 347-350.

van der Linden, R.H.J., Frenken, L.G.J., de Geus, B., Harmsen, M.M., Ruuls, R.C., Stok, W., de Ron, L., Wilson, S., Davis, P., and Verrips, C.T. (1999). Comparison of physical chemical properties of llama VHH antibody fragments and mouse monoclonal antibodies. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology **1431,** 37-46.

van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. Journal of structural biology **116,** 17-24.

Vanlandschoot, P., Stortelers, C., Beirnaert, E., Ibañez, L.I., Schepens, B., Depla, E., and Saelens, X. (2011). Nanobodies®: New ammunition to battle viruses. Antiviral Research **92,** 389-407.

Vaquero, C., Sack, M., Chandler, J., Drossard, J., Schuster, F., Monecke, M., Schillberg, S., and Fischer, R. (1999). Transient expression of a tumor-specific single-chain fragment and a chimeric antibody in tobacco leaves. Proc Natl Acad Sci U S A **96,** 11128-11133.

Vardakou, M., Sainsbury, F., Rigby, N., Mulholland, F., and Lomonossoff, G.P. (2012). Expression of active recombinant human gastric lipase in Nicotiana benthamiana using the CPMV-HT transient expression system. Protein Expression and Purification **81,** 69-74.

Ville, E., Carriere, F., Renou, C., and Laugier, R. (2002). Physiological study of pH stability and sensitivity to pepsin of human gastric lipase. Digestion **65,** 73-81.

Vincke, C., Loris, R., Saerens, D., Martinez-Rodriguez, S., Muyldermans, S., and Conrath, K. (2009). General Strategy to Humanize a Camelid Single-domain Antibody and Identification of a Universal Humanized Nanobody Scaffold. Journal of Biological Chemistry **284,** 3273-3284.

Walker, A., Skamel, C., and Nassal, M. (2011). SplitCore: An exceptionally versatile viral nanoparticle for native whole protein display regardless of 3D structure. Sci. Rep. **1**.

Walker, A., Skamel, C., Vorreiter, J., and Nassal, M. (2008). Internal core protein cleavage leaves the hepatitis B virus capsid intact and enhances its capacity for surface display of heterologous whole chain proteins. Journal of Biological Chemistry **283,** 33508-33515.

Wang, K., Herrera-Estrella, L., Van Montagu, M., and Zambryski, P. (1984). Right 25 bp terminus sequence of the nopaline T-DNA is essential for and determines direction of DNA transfer from agrobacterium to the plant genome. Cell **38,** 455-462.

Wang, Y.S., Ouyang, W., Liu, X.J., He, K.W., Yu, S.Q., Zhang, H.B., Fan, H.J., and Lu, C.P. (2012). Virus-like particles of hepatitis B virus core protein containing five mimotopes of infectious bursal disease virus (IBDV) protect chickens against IBDV. Vaccine **30,** 2125-2130.

Warzecha, H., Mason, H.S., Lane, C., Tryggvesson, A., Rybicki, E., Williamson, A.L., Clements, J.D., and Rose, R.C. (2003). Oral immunogenicity of human papillomavirus-like particles expressed in potato. J Virol **77,** 8702-8711.

Webster, D.E., Wang, L., Mulcair, M., Ma, C., Santi, L., Mason, H.S., Wesselingh, S.L., and Coppel, R.L. (2009). Production and characterization of an orally immunogenic Plasmodium antigen in plants using a virus-based expression system. Plant Biotechnol J **7,** 846-855.

Whitacre, D.C., Lee, B.O., and Milich, D.R. (2009). Use of hepadnavirus core proteins as vaccine platforms. Expert Rev Vaccines **8,** 1565-1573.

Willats, W.G. (2002). Phage display: practicalities and prospects. Plant Mol Biol **50,** 837-854.

Winichayakul, S., Pernthaner, A., Scott, R., Vlaming, R., and Roberts, N. (2009). Head-to-tail fusions of camelid antibodies can be expressed in planta and bind in rumen fluid. Biotechnol Appl Biochem **53,** 111-122.

Wullems, G.J., Molendijk, L., Ooms, G., and Schilperoort, R.A. (1981a). Retention of tumor markers in F1 progeny plants from in vitro induced octopine and nopaline tumor tissues. Cell **24,** 719-727.

Wullems, G.J., Molendijk, L., Ooms, G., and Schilperoort, R.A. (1981b). Differential expression of crown gall tumor markers in transformants obtained after in vitro Agrobacterium tumefaciens-induced transformation of cell wall regenerating protoplasts derived from Nicotiana tabacum. Proc Natl Acad Sci U S A **78,** 4344-4348.

Wynne, S.A., Crowther, R.A., and Leslie, A.G.W. (1999). The crystal structure of the human hepatitis B virus capsid. Mol Cell **3,** 771-780.

Yin, Y., Li, H., Wu, S., Dong, D., Zhang, J., Fu, L., Xu, J., and Chen, W. (2011). Hepatitis B virus core particles displaying Mycobacterium tuberculosis antigen ESAT-6 enhance ESAT-6-specific immune responses. Vaccine **29,** 5645-5651.

Yon, J., Rud, E., Corcoran, T., Kent, K., Rowlands, D., and Clarke, B. (1992). Stimulation of specific immune responses to simian immunodeficiency virus using chimeric hepatitis B core antigen particles. J Gen Virol **73 (Pt 10),** 2569-2575.

Yoshikawa, A., Tanaka, T., Hoshi, Y., Kato, N., Tachibana, K., Iizuka, H., Machida, A., Okamoto, H., Yamasaki, M., Miyakawa, Y., and et al. (1993). Chimeric hepatitis B virus core particles with parts or copies of the hepatitis C virus core protein. J Virol **67,** 6064-6070.

Zabel, F., Kundig, T.M., and Bachmann, M.F. (2013). Virus-induced humoral immunity: on how B cell responses are initiated. Current opinion in virology **3,** 357-362.

Zheng, J., Schödel, F., and Peterson, D.L. (1992). The structure of hepadnaviral core antigens - Identification of free thiols and determination of the disulfide bonding pattern. Journal of Biological Chemistry **267,** 9422-9429.

Appendix 1: Table of primers

The following table lists the DNA oligonucleotides used for cloning, sequencing, or colony PCR. Restriction sites are shown in red; stop or start codons in bold, italics and underlined; Kozak consensus sequence in bold and underlined; glycine-rich linker sequences in bold.

name	sequence 5'-3'	Tm (°C)	purpose
C1	AACGTTGTCAGATCGTGCTTCGGCACC	63	colony PCR
C3	CTGAAGGGACGACCTGCTAAACAGGAG	63	colony PCR
pMA-tHB5'	GTAATACGACTCACTATAGGGCGAATTGGCG	73	amplify t-EL from 5' end of pMA plasmid from GeneArt
pMA-tHB3'	GTTAATGCAGCTGGCACGACAGGTTTCCCG	80	amplify t-EL from 3' end of pMA plasmid from GeneArt
pMA-wtHB5'	GGTGGTAGTGGTGGGACCGGTACAATGGATATC	78	amplify 5' end of m-EL from t-EL & mutagenise AgeI-like site
pMA-CoreN5'	GTAATACGACTCACTATAGGGCGAATTGGC	71	amplify CoreN from 5' end of pMA plasmid from GeneArt
pMA-CoreN3'	CCCATGAGGCCAGTCTTGTGCTCCAGGTAC	79	amplify CoreN from 3' end of pMA plasmid from GeneArt
pMA-CoreC5'	GTAATACGACTCACTATAGGGCGAATTGGCGG	75	amplify CoreC from 5' end of pMA plasmid from GeneArt
pMA-coreC3'	CAGCTGGCACGACAGGTTTCCCGACTG	80	amplify CoreC from 3' end of pMA plasmid from GeneArt
5'UTRinsert	GTTTTCCCGTGGTTTTCGAACTTG	70	sequence ORF from 5'UTR of CPMV RNA-2
5'HBcAg-ce/1	GAACTGATGACCCTGGCAACCTG	71	sequence insert in ce/1 loop of HBcAg construct from 5' end
5'anti-phOx	GTTATGCACAGAACTTCCAGGGC	68	sequence anti-phOx scFv

3'anti-phOx	GTTGAACGTACCAGGAGACATAATTATTTCC	68	sequence anti-phOx scFv
3'HBcAg-ce/1	GGTGTTCACGTAATTAACAACCAGATCAC	68	sequence insert in ce/1 loop of HBcAg construct from 3' end
3'UTRinsert	GCACACCGAATAACAGTAAATTCAAACTAAAG	68	sequence ORF from 3'UTR of CPMV RNA-2
5'CoreN	GGATATCGACCCATATAAAGAATTTG	64	sequence HBcAg from 5' end
3'CoreNR	CTTCGCCAATTGCCAGTGACAAG	72	sequence in 3' direction from 3' end of CoreN
3'CoreNF	CTTGTCACTGGCAATTGGCGAAG	72	sequence in 5' direction from 3' end of CoreN
5'CoreCF	GTAGTCGACGTTACGATACTTCGCAAC	68	sequence in 5' direction from 5' end of CoreC
5'CoreCR	GTTGCGAAGTATCGTAACGTCGACTAC	68	sequence in 3' direction from 5' end of CoreC
3'CoreC	CCTAACATTGGCTTTCACG	60	sequence HBcAg from 3' end
5'solGFPR	GTGAACAGCTCCTCGCCCTTGC	73	sequence in 3' direction from 5' of sGFP1-10
5'solGFPF	GCAAGGGCGAGGAGCTGTTCAC	73	sequence in 5' direction from 5' of sGFP1-10
solGFPF	CTGGTGAACCGCATCGAGCTGAAG	75	sequence in 5' direction from middle of sGFP1-10
solGFPR	CTTCAGCTCGATGCGGTTCACCAG	75	sequence in 3' direction from middle of sGFP1-10
3'solGFPF	GATCACATGGTCCTGCACGAGTACG	72	sequence in 5' direction from 3' end of sGFP (middle of sGFP11)
3'solGFPR	CGTACTCGTGCAGGACCATGTGATC	72	sequence in 3' direction from 3' end of sGFP (middle of sGFP11)
gp120F1	CCT GTT CCT GAT CTT CAG CCT GCT GC	63	sequence gp120 from modified N-terminal secretion signal
gp120R1	CAC ACC CAG AGG CTC AAT CTT CAC CAC C	64	sequence gp120 from C-terminus
gp120F2	GGCTTGCCCTAAGGTGAGCTTCGAGC	64	sequence gp120 from middle of gene
gp120R2	GTT ACC GAT CTT ACC AAT GGT CAC GAA AGC C	63	sequence gp120 from middle of gene
VHH2F	GACCGGT**AACAATGGC**TGTTCAGCTTCAAG	76	clone gp120 from pMA plasmid adding AgeI site and **Kozak sequence**

VHH2R	GCCCGGGAGATGAAACGGTCACC	77	clone gp120 from pMA plasmid adding XmaI site
p24F	GTCTTGTCCAGAGCTCACCGGTAAC	70	clone p24 from backbone of pMA plasmid
p24R	GTATCCCGGGAAGCACTCTAGCC	70	clone p24 from pMA plasmid, replacing stop codon and XhoI site with SmaI site
GFP-p24_F	CGTACCGGTAACAATGGTGAGCAAGGGCGAGG	82	clone sGFP from pEAQ-*HT*-t-sGFP, adding **Kozak sequence**, ***start codon*** and AgeI site
GFP-p24_R	GTATCCCGGGAAGCACTCTAGCCTTATGTCCAGGACC	80	clone p24 from pMA, replacing stop codon with SmaI site
GFP-p24_2R	CACCGTCGACACTTCCTCCCTTGTACAGCTCGTCC	83	clone sGFP from pEAQ-*HT*-t-sGFP, adding SalI site and **GGS**
GFP-p24_2F	GAAGTGTCGACGGTGGTTCTATGCCTAGGACTCTTAACG	79	clone p24 from pMA adding SalI site and **GGS**
sGFP seqF	CAACTTCAACAGCCACAACGTCTATATCACCGCC	79	sequence middle of sGFP in forward direction
p24 seqR	CCACGGTATTAAGCATGGTGTTAAGATCCTGAGGGGTAGC	80	sequence middle of p24 in reverse direction

Appendix 2: Publications

During the course of this thesis, a review of the use of the pEAQ vectors was published, along with a research article introducing tandem core technology including the use of tandibodies:

Peyret, H., and Lomonossoff, G.P. (2013). The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. Plant Mol Biol, **83**(1-2): 51-8.

Peyret, H., Gehin, A., Thuenemann, E.C., Blond, D., El Turabi, A., Beales, L., Clarke, D., Gilbert, R.J.C., Fry, E.E., Stuart, D.I., Holmes, K., Stonehouse, N.J., Whelan, M., Rosenberg, W., Lomonossoff, G.P., Rowlands, D.J. (2015). Tandem fusion of hepatitis B core antigen allows assembly of virus-like particles in bacteria and plants with enhanced capacity to accommodate foreign proteins. PLOS One, *In Press*