Plant Biotechnology Journal

Plant Biotechnology Journal (2008) **6**, pp. 914–929 doi: 10.1111/j.1467-7652.2008.00365.x

Plastid transformation of high-biomass tobacco variety Maryland Mammoth for production of human immunodeficiency virus type 1 (HIV-1) p24 antigen

Matthew S. McCabe[1], Manfred Klaas[1], Nuria Gonzalez-Rabade[2], Miranda Poage[1], Jesus A. Badillo-Corona[2], Fei Zhou[3], Daniel Karcher[3], Ralph Bock[3], John C. Gray[2] and Philip J. Dix[1,*]

[1]*Department of Biology, National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland*

[2]*Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK*

[3]*Max-Planck-Institut für Molekulare Pflanzenphysiologie, Am Muhlenberg 1, D-14476 Potsdam-Golm, Germany*

Received 3 March 2008;
revised 3 June 2008;
accepted 30 June 2008.
**Correspondence* (fax +353-(0)1-708-3845; e-mail phil.dix@nuim.ie)

Summary

Chloroplast transformation of the high-biomass tobacco variety Maryland Mammoth has been assessed as a production platform for the human immunodeficiency virus type 1 (HIV-1) p24 antigen. Maryland Mammoth offers the prospect of higher yields of intact functional protein per unit floor area of contained glasshouse per unit time prior to flowering. Two different transformation constructs, pZSJH1p24 (for the insertion of a native *p24* cDNA between the *rbcL* and *accD* genes) and pZF5 (for the insertion of a chloroplast-codon-optimized *p24* gene between *trnfM* and *trnG*) were examined for the production of p24. Plants generated with construct pZSJH1p24 exhibited a normal green phenotype, but p24 protein accumulated only in the youngest leaves (up to approximately 350 μg/g fresh weight or ~2.5% total soluble protein) and was undetectable in mature leaves. In contrast, some of the plants generated with pZF5 exhibited a yellow phenotype (pZF5-yellow) with detectable p24 accumulation (up to approximately 450 μg/g fresh weight or ~4.5% total soluble protein) in all leaves, regardless of age. Total protein in pZF5-yellow leaves was reduced by ~40%. The pZF5-yellow phenotype was associated with recombination between native and introduced direct repeat sequences of the *rbcL* 3′ untransformed region in the plastid genome. Chloroplast-expressed p24 was recognized by a conformation-dependent monoclonal antibody to p24, and p24 protein could be purified from pZF5-yellow leaves using a simple procedure, involving ammonium sulphate precipitation and ion-exchange chromatography, without the use of an affinity tag. The purified p24 was shown to be full length with no modifications, such as glycosylation or phosphorylation, using N-terminal sequencing and mass spectrometry.

Keywords: chloroplast transformation, high-biomass tobacco, human immunodeficiency virus type 1 (HIV-1) p24 antigen.

Introduction

Human immunodeficiency virus type 1 (HIV-1) p24 is a non-glycosylated 231-amino-acid protein that forms the conical core of the HIV-1 virus encapsulating the genomic RNA–nucleocapsid complex. In HIV-1-infected mammalian cells, p24 is synthesized as part of a polyprotein precursor, known as Pr55Gag, which is cleaved by the viral protease into four mature major Gag proteins: p17 matrix protein (MA), core antigen or capsid protein (CA-p24), nucleocapsid protein (NC-p7) and p6 (Freed, 1998; Obregon *et al*., 2006).

HIV-1 p24 is widely used in enzyme-linked immunoassay-based diagnostic testing for HIV infection (Iweala, 2004). Antibodies to gp41 and p24 antigens are the first detectable serological markers following HIV infection (Allain *et al*., 1986). HIV-1 p24, as part of the Gag polyprotein, is included as a component of many of the composite subunit HIV-1 vaccines that are currently being tested (International AIDS Vaccine Initiative, 2006; Obregon *et al*., 2006). A translational fusion between HIV-1 p24 and a non-toxic component of the *Bacillus anthracis* lethal toxin (Lu *et al*., 2000) has also been tested recently in Phase 1 and 2 clinical trials (Girard *et al*., 2006;

International AIDS Vaccine Initiative, 2006, 2006/2007; Nkolola and Essex, 2006). In addition, immunization and challenge studies in cats have demonstrated that HIV-1 p24 confers significant cross-protection against feline immunodeficiency virus, probably via a T-cell-mediated response (Coleman *et al*., 2005).

HIV-1 p24 can be expressed and purified from recombinant expression systems, such as *Escherichia coli* (Ehrlich *et al*., 1990; Hausdorf *et al*., 1994; Narciandi *et al*., 2006) and insect cells (Mills and Jones, 1990; Gay *et al*., 1998). Reported yields of purified recombinant p24 protein from *E. coli* are 12 mg/L (Hausdorf *et al*., 1994). In insect cell cultures, recombinant p24 has been reported to accumulate to 50 mg/L.

Plants are an attractive system for the production of injectable recombinant proteins as they do not harbour viruses or prions that are pathogenic in humans. Tobacco is a non-food crop in which leaf tissue may be harvested for protein production prior to flowering and seed set (Fischer *et al*., 2004). For these reasons, tobacco is more attractive in terms of transgene containment than are plant protein production systems based on food crops, such as maize and rice, where the recombinant proteins are extracted from seed. Transgene containment is likely to be a particularly important issue for plants transformed or transfected with HIV antigen genes.

There have been reports of the expression of HIV-1 p24 protein in transgenic tobacco (Zhang *et al*., 2002; Obregon *et al*., 2006) and in tobacco transfected with engineered tobacco mosaic virus (TMV) (Zhang *et al*., 2000; Pérez-Filgueira *et al*., 2004). p24 accumulated to approximately 35 μg/g leaf fresh weight [0.35% total soluble protein (TSP)] in transgenic tobacco (Zhang *et al*., 2000) and to 80 μg/g leaf fresh weight (~0.8% TSP) using TMV transfection (Pérez-Filgueira *et al*., 2004). However, with both nuclear transformation and viral transfection, there is a risk of transgene escape.

The use of plastid transformation for the high-level expression of recombinant therapeutic proteins is now well established (Staub *et al*., 2000; Daniell *et al*., 2001; Tregoning *et al*., 2003; Birch-Machin *et al*., 2004; Molina *et al*., 2004; Watson *et al*., 2004; Koya *et al*., 2005; Daniell, 2006; Glenz *et al*., 2006; Arlen *et al*., 2007; Chebolu and Daniell, 2007). Chloroplast transformation has many advantages over other plant recombinant protein production systems, particularly for the production of non-glycosylated proteins (Maliga, 2002, 2003; Daniell, 2006). The principal advantage of this technology is that, unlike most cases of nuclear transformation, commercially viable yields of recombinant therapeutic protein, e.g. 25% TSP for tetanus toxin fragment C vaccine (Tregoning *et al*., 2003), can be obtained through stable chloroplast transformation. Furthermore, recombinant proteins expressed via plant nuclear transformation or viral transfection can potentially be glycosylated with allergenic non-mammalian glycans (Sriraman *et al*., 2004). Recombinant proteins produced by chloroplast transformation are not glycosylated. Chloroplast transformation also provides a transgene containment advantage as, unlike nuclear genomes, chloroplast genomes are transmitted with negligible frequency in the pollen of most plant species (Ruf *et al*., 2007; Svab and Maliga, 2007).

As p24 is an HIV-1 antigen, production in plants is likely to be required to be in contained glasshouses and not in the field. This limitation of space means that the optimization of the yield of intact functional protein per unit floor area per unit time prior to flowering is a major objective for the production of p24 and other proteins that have similar containment requirements. So far, chloroplast transformation for the production of recombinant proteins has mainly been carried out in the small-leaved tobacco variety Petite Havana. This is largely because of the ease of chloroplast transformation of this variety. However, Petite Havana may not be ideal for recombinant protein production as it produces relatively small amounts (100–200 g) of leaf biomass prior to flowering. It is also a day-neutral variety, so that flowering cannot be controlled by manipulation of the day length.

Maryland Mammoth is potentially an excellent candidate for the production of pharmaceutical proteins in chloroplasts. It is a large tobacco variety that can grow to over 3 m in height and produces 1–2 kg of leaf biomass per plant prior to flowering. In addition, Maryland Mammoth is a short-day flowering variety, so that flowering is inhibited if the photoperiod is more than 14 h (Garner and Allard, 1920, 1923).

In this article, we describe an assessment of two different chloroplast transformation constructs for the production of p24 in the high-biomass variety Maryland Mammoth. One of the constructs generates phenotypically normal Maryland Mammoth plants in which p24 accumulates to detectable levels only in young leaves. A second construct generates yellow Maryland Mammoth plants with delayed flowering in which there is significant accumulation of HIV-1 p24 protein in all leaves, regardless of age. The chloroplast-expressed p24 was recognized by a conformation-dependent monoclonal antibody to p24, and p24 protein, with no detectable post-translational modification, could be readily purified from leaves using a simple procedure that did not employ an affinity tag.

Results

Expression constructs

Two tobacco chloroplast transformation DNA constructs, pZSJH1p24 and pZF5 (Figure 1), were assessed for the

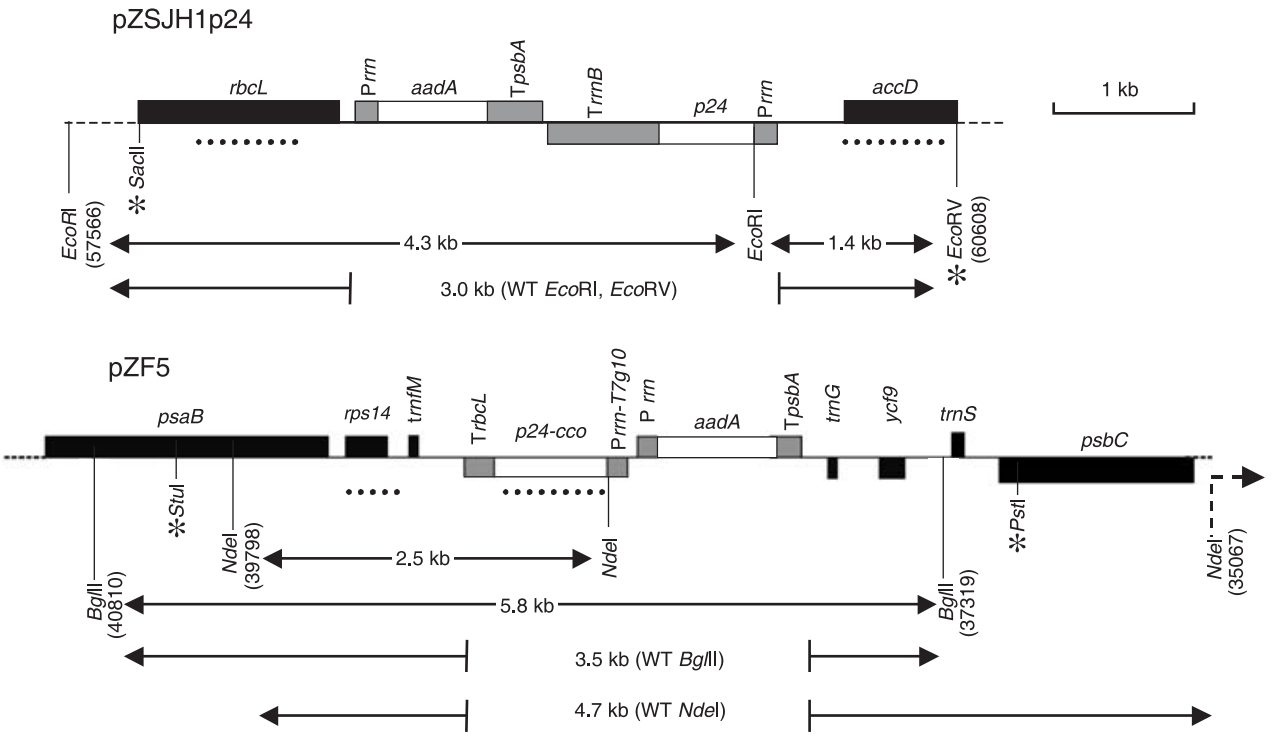

Figure 1 Chloroplast transformation constructs pZSJH1p24 and pZF5. Filled black boxes represent tobacco chloroplast genes. In pZSJH1p24, the native human immunodeficiency type 1 (HIV-1) *p24* cDNA, with the chimeric *rrn* promoter (P*rrn*) and *Escherichia coli rrn*B terminator (T*rrnB*), was inserted just downstream of the chimeric *aadA* selectable marker gene in pZS197 (Svab and Maliga, 1993). In pZF5, a synthetic chloroplast-codon-optimized HIV-1 *p24* gene (*p24-cco*), with the *rrn* promoter + T7g10 leader (*Prrn-T7g10*) and the *rbcL* terminator (T*rbcL*), was inserted upstream of *aadA* in pRB95 (Ruf *et al*., 2001). Genes above the line are transcribed from left to right; genes below the line are transcribed from right to left. The expected restriction enzyme digestion fragment sizes in transformed and non-transformed (wild-type, WT) chloroplast genomes are indicated by double ended arrows. The limit of the vector sequence is delineated by *. Regions probed in Southern blots are indicated by dotted lines.

expression of p24 protein in Maryland Mammoth. pZSJH1p24 consists of native BH10 strain HIV-1 *p24* cDNA (accession number M15654) (Mills and Jones, 1990), linked to P*rrn* with the *rbcL* 5′ untransformed region (5′UTR) and the *E. coli rrnB* 3′UTR (T*rrnB*) (Amann and Brosius, 1985), and inserted into a polylinker in pZSJH1 (a modified version of pZS197) (Birch-Machin *et al*., 2004). This adds an extra three amino acid residues (methionine-asparagine-serine, Met-Asn-Ser) to the N-terminus of the p24 protein. This vector targets genes to the intergenic spacer between the *rbcL* and *accD* genes in the large single-copy region of the chloroplast genome. The *p24* gene is in the opposite orientation to the *aadA* gene, such that these two genes are convergently transcribed. This format has been used previously for the expression of the bovine rotavirus VP6 cDNA (Birch-Machin *et al*., 2004) and green fluorescent protein (GFP) (Newell *et al*., 2003). pZF5 contains a synthetic chloroplast-codon-optimized HIV-1 *p24* gene (*p24-cco*) (accession number EU670669) linked via *Nhe*I and *Bam*HI sites to the promoter sequence P*rrn-T7g10* and *rbcL* 3′UTR, as described previously for pHK40 (Kuroda and Maliga, 2001), and inserted into the multiple cloning site of pRB95 (Ruf *et al*., 2001). The introduced genes are targeted to the intergenic spacer between the *trnfM* and *trnG* genes in the large single-copy region of the chloroplast genome. As a result of cloning, the *p24-cco* gene codes for an extra five amino acids (methionine-alanine-serine-glycine-serine, Met-Ala-Ser-Gly-Ser) at the N-terminus and an extra two amino acids (glutamic acid and phenylalanine, Glu and Phe) before the last amino acid (leucine, Leu) at the C-terminus of the p24-cco protein. The *p24-cco* gene is in the opposite orientation to the selectable marker gene, *aadA*, such that the two genes are divergently transcribed.

Chloroplast transformation

Chloroplast transformation of Maryland Mammoth was achieved with similar transformation frequencies to those obtained with Petite Havana when leaves from internodal cuttings propagated *in vitro* were bombarded. However, initial attempts to transform chloroplasts of Maryland Mammoth were unsuccessful when leaves from plants grown from seed *in vitro* were bombarded. The bombardment of

100 leaves from seed-grown plants with constructs pRB95 and pZSJH1 produced only spontaneous spectinomycin-resistant shoots and no transformants. Compared with Petite Havana, Maryland Mammoth plants grown *in vitro* from seed developed slowly with thin stems and shortened internodes. However, vigorous growth of Maryland Mammoth *in vitro* was obtained when the plants were initiated from cuttings, including one or two internodal sections, taken from seed-grown plants *in vitro*. The bombardment of leaves taken from these vigorously growing plants yielded similar transformation frequencies with pZSJH1p24 and pZF5. Bombardment of 28 Maryland Mammoth leaves (one bombardment per leaf) with pZSJH1p24 resulted in the regeneration of spectinomycin-resistant shoots from 22 independent leaf squares. When these shoots were screened by long-range and short-range polymerase chain reaction (PCR), six were identified as chloroplast transformants. For pZF5, the bombardment of 25 Maryland Mammoth leaves resulted in the regeneration of spectinomycin-resistant shoots from 21 independent leaf squares, and seven of these shoots were identified as chloroplast transformants by PCR analysis. This suggests that, in terms of transformation frequency, there is no advantage in targeting either of the two different regions of the chloroplast genome.

Phenotype

To check seed batches and compare growth, non-transformed Maryland Mammoth plants were grown from seed in multipurpose compost, without the addition of fertilizer, in 40-L pots under natural light in a glasshouse in the summer (Figure 2a). The plants grew with a single large stem to 2.1 m in height and flowered (first flower opened) at 75 days after sowing. Plants produced approximately 1.5 kg of leaf material at 75 days after sowing. Petite Havana plants grown under the same conditions flowered at 58 days after sowing and produced significantly less biomass (Figure 2a).

The transformation of Maryland Mammoth with pZSJH1p24 generated plants with the wild-type (WT) phenotype (Figure 2b). No differences in growth or phenotype were observed

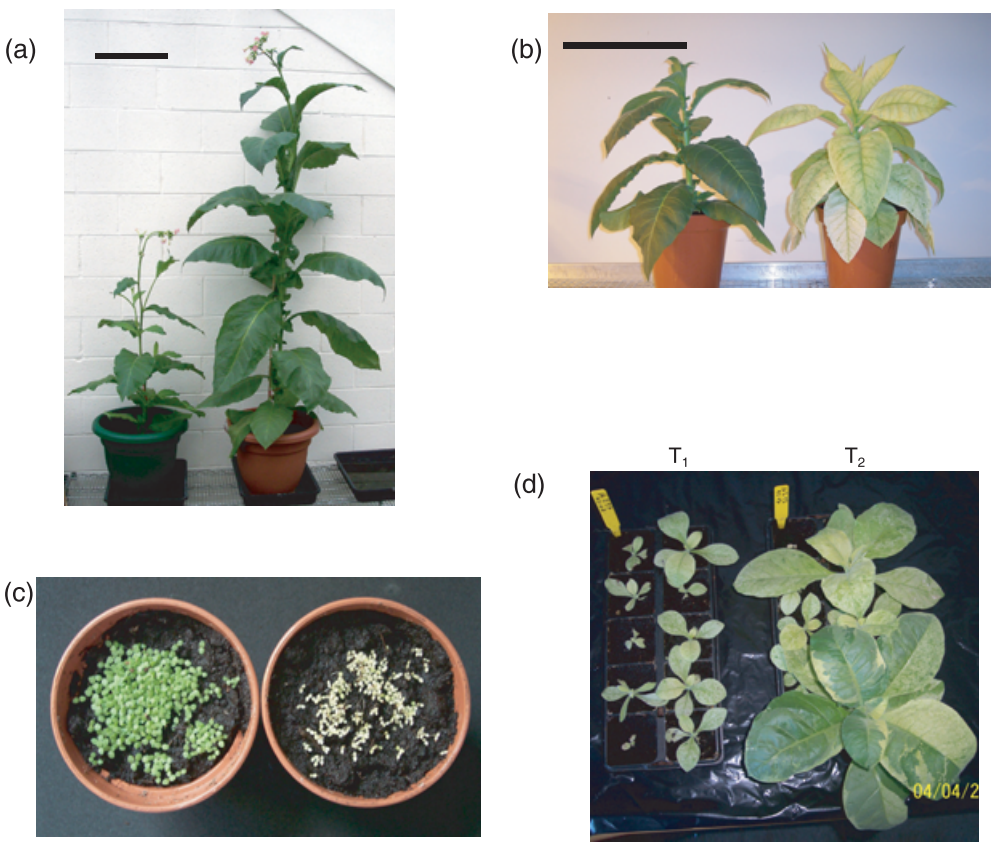

Figure 2 Growth of tobacco plants. (a) Maryland Mammoth tobacco plant (right) at start of flowering, 75 days after sowing the seed. Petite Havana plant (left) at start of flowering, 58 days after sowing the seed. Plants were grown in 40-L pots under natural daylight from July to September. Scale bar represents 37 cm. (b) pZSJH1p24 Maryland Mammoth T_1 plant 56 days after sowing the seed (left). pZF5-yellow Maryland Mammoth T_1 plant 70 days after sowing the seed (right). Scale bar represents 30 cm. (c) pZSJH1p24 (left) and pZF5-yellow T_1 (right) seedlings 7 days after sowing on compost in 9-cm-diameter pots. (d) pZF5-yellow T_1 and T_2 plants showing revertant sectors in a T_2 plant.

between pZSJH1p24 T_1 plants and WT plants grown in the glasshouse in 40-L pots in multipurpose potting compost without added fertilizer. Each plant produced approximately 1.5 kg fresh weight of leaf tissue at 60 days after sowing (prior to flowering). No statistically significant difference in leaf biomass was observed between transplastomic and WT plants. Maryland Mammoth plants transformed with pZSJH1p24 flowered 3 weeks later than pZSJH1p24-transformed Petite Havana plants when grown under defined conditions in a growth room with a 16-h/8-h light/dark cycle.

The transformation of Maryland Mammoth with pZF5 gave rise to two phenotypes: a yellow phenotype, 'pZF5-yellow', with yellow leaves and pale green (or occasionally dark green) flecks or sectors with a reduced growth rate and delayed flowering (~120 days after sowing) (Figure 2b) that occurred in around 50% of the plants, and a normal phenotype, 'pZF5-green'. pZF5-green and pZF5-yellow plants often regenerated from the same explant during the first round of regeneration. pZF5-yellow plants eventually grew to normal size in smaller pots, but were more difficult to grow in 40-L pots because of a poor rooting system that could not support the weight of the larger plants. pZF5-yellow plants produced normal flowers, seed pods and seeds after self-pollination. There was no segregation of the yellow phenotype in pZF5-yellow T_1 seedlings (Figure 2c), suggesting the insertion of the pZF5 expression cassette in the chloroplast genome and not the nuclear genome. A reversion to the green phenotype (sectoring) occurred in one in 10 T_2 plants derived from one T_1 plant (Figure 2d). No segregation of spectinomycin resistance was observed in pZF5-yellow and pZF5-green T_1 seedlings (approximately 300 seedlings germinated on 500 mg/L spectinomycin).

Analysis of transgene integration

The confirmation of the insertion of the p24 expression constructs in the chloroplast genome was obtained by Southern blot analysis on extracted total leaf DNA. The positions of the restriction sites and probes used in Southern blot analysis of the plants obtained after bombardment with pZSJH1p24 and pZF5 are shown in Figure 1. For the analysis of pZSJH1p24 plants, total DNA from T_1 plants digested with *Eco*RI and *Eco*RV, and hybridized with a probe specific for either the *rbcL* gene or the *accD* genes, gave strong bands of the expected sizes (Figure 3a). Additional fainter bands were also observed. These are often observed in Southern blots of stable chloroplast-transformed tobacco plants even after six to seven rounds of regeneration on spectinomycin in T_0 plants and three to four sexual cycles without antibiotic selection. Such bands have been attributed to promiscuous plastid DNA in the nucleus and/or the mitochondrion (Hager *et al*., 1999; Ruf *et al*., 2000).

For pZF5 plants, total DNA from pZF5-green T_1 plants and pZF5-yellow T_1 plants was digested with either *Bgl*II or *Nde*I, and Southern blots were hybridized to probes specific for either the tobacco chloroplast *rps14* gene or the *p24-cco* gene. Differences in banding patterns between green and yellow plants in these Southern blots suggested that there were rearrangements at the site of insertion (Figure 3b). In *Bgl*II digests of total genomic DNA extracted from pZF5-green and pZF5-yellow plants, the *rps14* probe hybridized to a band close to the expected size of 5.8 kb and to an additional band of approximately 3 kb. The strengths of the signals for the two bands were different for DNA from pZF5-green and pZF5-yellow plants. There was a strong signal for the expected 5.8-kb band and a weak signal for the 3-kb band in the pZF5-green plants and vice versa for the pZF5-yellow plants. In *Bgl*II-digested DNA from non-transformed (WT) plants, the *rps14* probe hybridized to a band close to the expected size of 3.5 kb. In the same digests, the *p24-cco* probe gave a strong band close to 5.8 kb in DNA from both pZF5-green and pZF5-yellow plants, and a weak band of approximately 3 kb was visible only in DNA from the pZF5-yellow plants.

For *Nde*I-digested DNA from pZF5-green and pZF5-yellow plants, the *rps14* probe hybridized with an expected band of 2.5 kb and three other bands of approximately 2, 5 and 6 kb. The expected 2.5-kb band was strongest in pZF5-green plants, whereas the 2-kb band was strongest in pZF5-yellow plants. For *Nde*I-digested DNA hybridized with the *p24-cco* probe, there was a single strong 2.5-kb band in DNA from pZF5-green plants, and a 2.5-kb band plus a faint 2-kb band in DNA from pZF5-yellow plants.

The additional 3-kb *Bgl*II band and 2-kb *Nde*I band observed with the *rps14* probe may be explained by recombination between the direct repeats of the 141-bp *rbc*L 3'UTR in the *p24-cco* gene and the native chloroplast *rbcL* gene. This would result in the excision of a 21-kb section of the chloroplast genome between *trnfM* and the 3'UTR of the *rbcL* gene, yielding a truncated 138-kb chloroplast genome containing the introduced *p24-cco* and *aadA* genes and the excised 21-kb region. The additional 3-kb *Bgl*II and 2-kb *Nde*I bands observed with the *rps14* probe indicate that this putative 21-kb region has circularized (Figure 3c). The Southern blots also suggest that the putative 21-kb circle is much more abundant in pZF5-yellow plants than in pZF5-green plants. Attempts to detect the putative 21-kb circle by Southern analysis of undigested DNA from pZF5-yellow

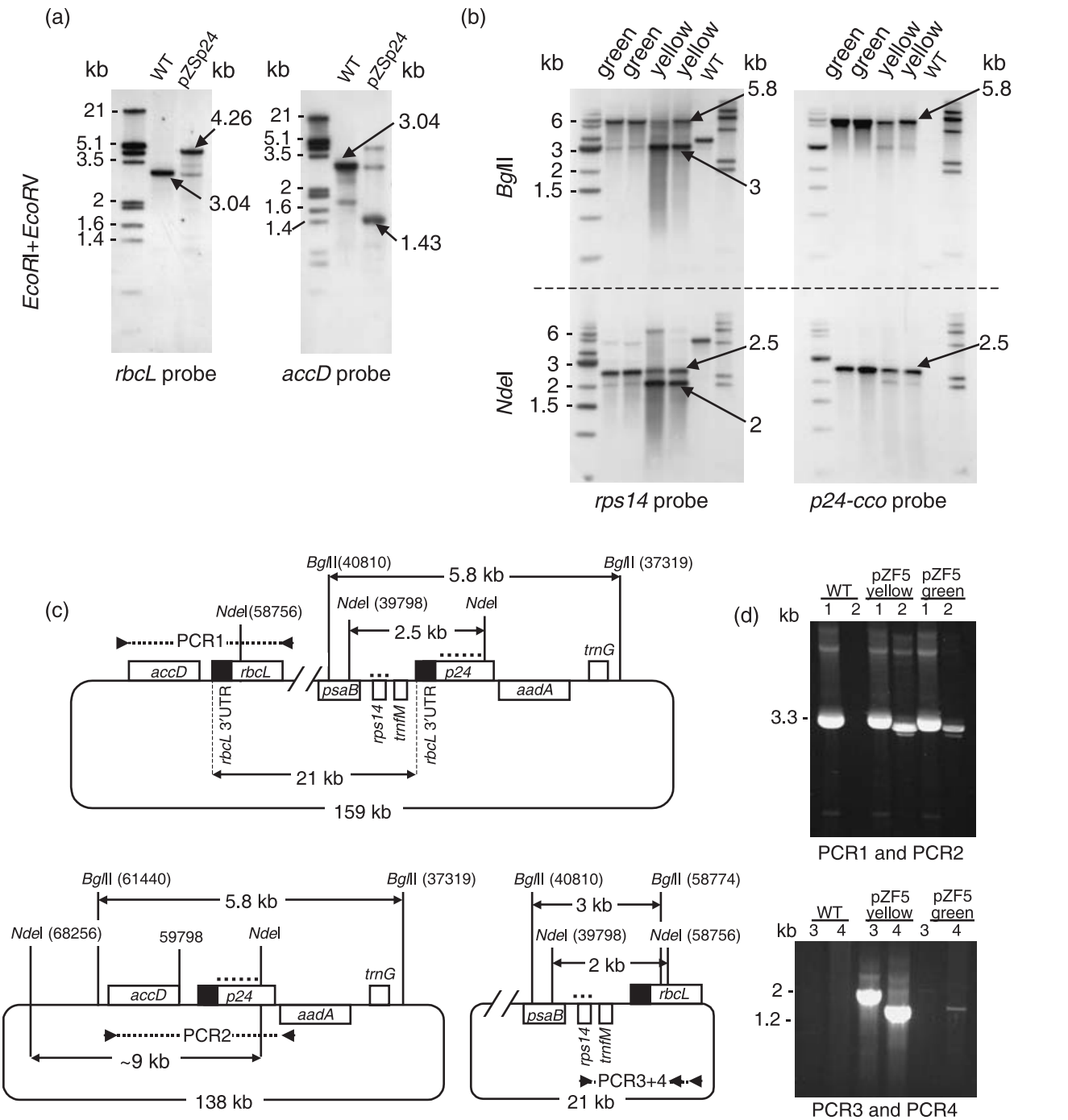

Figure 3 Southern blot and polymerase chain reaction (PCR) analysis of pZF5 and pZSJH1p24 plants. (a) For pZSJH1p24, 0.5 μg of total genomic DNA was digested with *Eco*RV and *Eco*RI and hybridized separately with PCR DIG-labelled probes specific for the tobacco chloroplast *rbcL* and *accD* genes. (b) For pZF5, 1 μg of total genomic DNA from each of two pZF5-green and two pZF5-yellow T_0 plants was digested with *Bgl*II or *Nde*I, subjected to electrophoresis on a 0.8% agarose gel, and then blotted and hybridized with PCR DIG-labelled probes specific for the tobacco chloroplast *rps14* or *p24-cco* gene. (c) Proposed recombination products resulting from recombination between introduced and native *rbcL* 3′ untransformed regions (3′UTRs) (indicated by filled boxes). The nucleotide positions of restriction sites on the tobacco chloroplast genome (GenBank accession no. Z00044) are indicated in brackets. The two diagonal lines on the 159-kb genome and 21-kb circle indicate the 17 chloroplast genes, open reading frames (ORFs) and *ycfs* between *psaB* and *rbcL* that have been omitted from the diagram for clarity. Dotted lines highlight the binding sites of *p24-cco* and *rps14* probes used on Southern blots. (d) PCR analysis of recombined transformed chloroplast genomes in pZF5-yellow and pZF5-green plants. The positions of primer binding sites for reactions PCR1–4 are shown in (c). WT, wild-type.

plants were unsuccessful, probably because it is too close to the size of undigested DNA.

With the *p24-cco* probe, the expected 5.8-kb *Bgl*II fragment in the putative 138-kb truncated chloroplast genome could not be distinguished from the 5.8-kb *Bgl*II fragment in the full-length 159-kb transformed chloroplast genome, and a predicted 9-kb *Nde*I fragment for the truncated genome was not observed (Figure 3b,c). This was either because of the low abundance of the 139-kb product, or because further recombination events occurred between the three predicted products that eliminated the predicted 9-kb band. Such recombination events may also explain the faint 6-kb band in the *rps14*-probed *Nde*I digest. The faint 3-kb and 2-kb bands in *p24-cco*-probed DNA from pZF5-yellow plants must be caused by cross-reaction of the PCR-DIG-labelled probe, as a very faint WT band was also visible with this probe. DIG labelling of the synthetic *p24-cco* gene from the pZF5 plasmid DNA template required a low annealing temperature because of the high proportion of T residues in the gene. This may have allowed non-specific labelling of chloroplast genes.

Recombination products in pZF5 plants were further analysed by diagnostic PCR. Primer combinations PCR1–4 (Figure 3c,d) were used to amplify products from total leaf DNA of pZF5-yellow and pZF5-green plants. Significant differences were observed in the amounts of PCR3 and PCR4 amplified from pZF5-green and pZF5-yellow plants (Figure 3d), confirming that the predicted 21-kb circle was more abundant in pZF5-yellow than pZF5-green plants. Sequencing of the PCR2 product, using a forward primer that bound to the 3′ end of the *p24-cco* gene, confirmed that the *accD* gene was adjacent to the *p24-cco* gene in the putative 138-kb circle. Similarly, sequencing of the PCR3 product using a forward primer specific for the 3′ end of *rbcL* and a primer specific for the start of *trnfM* confirmed that *accD* was adjacent to *p24-cco* in the putative 21-kb circle. PCR amplification products were not visible when primer combinations PCR1–4 were used to analyse DNA extracted from vector-only control Maryland Mammoth plants transformed with pRB95 (data not shown), eliminating the possibility of template switching.

Analysis of P*rrn* promoters in pZF5-green and pZF5-yellow plants

A 42-bp deletion in the P*rrn* promoter that drives the *aadA* gene has been detected in the pZF5 DNA construct in *E. coli* (Zhou *et al*., 2008). The adjacent P*rrn* promoter that drives the p24 gene is of full length. In transplastomic Maryland Mammoth plants obtained with pZF5, inversion or 'flip-flop' of the full-length and deleted P*rrn* promoters occurs. This 'flip-flop' was also observed in Petite Havana with a similar construct carrying the HIV-1 *nef* gene (Zhou *et al*., 2008). Sequencing of the P*rrn* promoters, amplified by PCR from total genomic DNA, showed that, in pZF5-yellow plants, the deleted P*rrn* promoter was driving the *aadA* gene and the full-length P*rrn* promoter was driving the *p24-cco* gene (Figure 4). By contrast, in pZF5-green plants, the deleted P*rrn* promoter was driving the *p24-cco* gene and the full-length promoter was driving the *aadA* gene (Figure 4). The analysis of transcripts by real-time PCR on cDNA showed that transcripts of *p24* were higher in pZF5-yellow than in pZF5-green plants, whereas transcripts of *aadA* were higher in pZF5-green than in pZF5-yellow plants (Figure 4). This shows that the 42-bp deletion reduces the level of transcripts from the P*rrn* promoter.

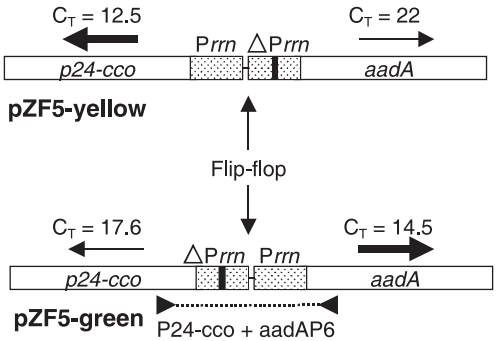

Figure 4 'Flip-flop' recombination of deleted and non-deleted P*rrn* promoters in pZF5-green and pZF5-yellow plants. Threshold cycle (C_T) values from quantitative polymerase chain reaction (PCR) of *p24-cco* and *aadA* in pZF5-yellow and pZF5-green total leaf RNA are shown. The approximate positions of primers P24-cco and aadAP6, used for the amplification and sequencing of the two P*rrn* promoters, are shown by inverted arrows at the ends of the dotted line.

Analysis of p24 accumulation and stability

In pZF5-yellow plants, the p24 protein was usually visible after sodium dodecylsulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie blue staining of 30 μg of crude soluble protein extract from young and mature leaves (data not shown). p24 was not detectable, even by Western blot, in spectinomycin-resistant T_1 pZF5-green plants that contained the p24 gene at the correct insertion site as shown by Southern analysis (Figure 3b). This lack of expression was probably a result of the deleted P*rrn* promoter driving the *p24-cco* gene in these plants. In pZSJH1p24 plants, p24 was detectable by Western blot, but was not usually visible on a Coomassie blue-stained gel.

To examine the accumulation profiles of *p24* and *p24-cco* in young and old leaves of pZSJH1p24 and pZF5-yellow Maryland Mammoth plants, respectively, leaf extracts were prepared on the same day from all leaves of a 12-leaf stage pZSJH1p24 T_1 plant and a 12-leaf stage pZF5-yellow T_1 plant grown in the same culture room at the same time, and later from plants at the 25-leaf or 29-leaf stages (Figure 5a). Western blot analysis showed that p24 accumulated in only the younger leaves of pZSJH1p24 plants, as shown previously for Petite Havana plants transformed with this construct (Zhou *et al.*, 2008). p24 protein accumulated in leaves 6–12 of plants at the 12-leaf stage, but this protein was no longer detectable in these leaves in plants at the 29-leaf stage. This suggests that p24 is not stable in older leaves of pZSJH1p24 plants. In contrast, p24 accumulated to levels detectable by Western blot in all leaves of a pZF5-yellow plant at the 12-leaf stage, and the protein was retained in older leaves of a pZF5-yellow plant at the 25-leaf stage.

The amounts of extractable p24 present in leaf tissue of pZSJH1p24 and pZF5-yellow plants at the 12-leaf stage were determined by enzyme-linked immunosorbent assay (ELISA) using a conformation-dependent antibody (Figure 5b). This confirmed the different patterns of accumulation of p24 in pZSJH1p24 and pZF5-yellow plants demonstrated by Western blotting (Figure 5a). The largest amounts of p24 were detected in leaves 6–8 of pZF5-yellow plants, and in leaf 11 of pZSJH1p24 plants (Figure 5b). p24 accumulated at up to 450 μg/g fresh weight in pZF5-yellow plants and up to 350 μg/g fresh weight in pZSJH1p24 plants. These values correspond to ~4.5% TSP in pZF5-yellow plants and ~2.5% TSP in pZSJH1p24 plants. The amount of native TSP per gram of fresh leaf weight in pZF5-yellow plants was approximately 60% of that in pZSJH1p24 and non-transformed Maryland Mammoth leaves.

To further determine the difference in stability of p24 and p24-cco, the accumulation of these proteins was monitored in a single leaf from a pZSJH1p24 and pZF5 plant, respectively, over a 19-day period (Figure 5c). This showed that, in the pZSJH1p24 leaf, there was a significant decrease in p24 to undetectable levels by day 19, whereas, in the pZF5-yellow leaf, p24-cco protein accumulation was relatively stable. This stability suggests that Maryland Mammoth pZF5-yellow plants are likely to be a good source of p24 protein for purification.

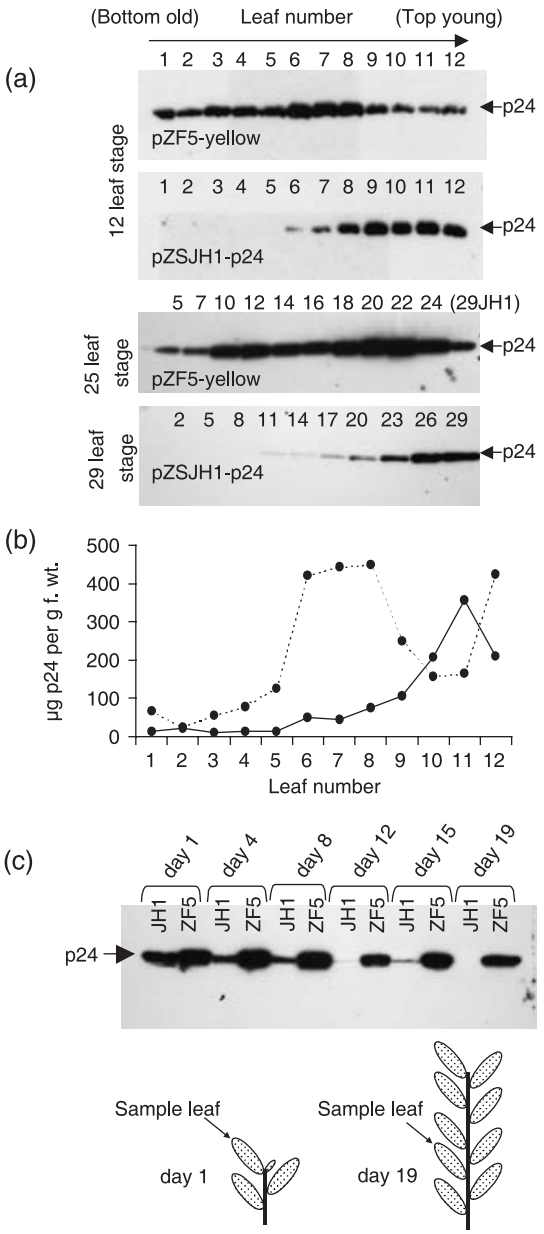

Figure 5 Analysis of p24 protein accumulation in leaves from pZF5-yellow and pZSJH1p24 plants. (a) Top two panels: Western blots of proteins extracted on the same day from leaves 1–12 (numbered from the bottom oldest leaf) of a pZF5-yellow plant at the 12-leaf stage (114 days after sowing) and a pZSJH1p24 plant at the 12-leaf stage (63 days after sowing); 20 μg of total soluble protein were loaded in each lane. Bottom two panels: Western blots of proteins extracted on the same day from pZF5-yellow and pZSJH1p24 plants at the 25- and 29-leaf stage, respectively. Each lane was loaded with 10 μL of a crude extract of 100 mg of leaf in 400 μL of extraction buffer. The signal intensity of the sample from leaf 29 of pZSJH1p24 (29 JH1) is similar on both blots. All four blots were incubated with affinity-purified polyclonal anti-p24 antibody (D7320). (b) Capture enzyme-linked immunosorbent assay (ELISA) of the same protein extracts as shown in (a) from 12-leaf stage plants, showing the amount (in micrograms) of extractable p24 per gram fresh weight of harvested leaves. (c) Analysis of stability of p24 in the same leaf over a 19-day period in a pZSJH1p24 T_1 plant (JH1) and a pZF5 T_1 (Zf5) plant. Both plants were at the three-leaf stage at day 1 and the nine-leaf stage at day 19.

Purification of p24 protein from pZF5-yellow plants

A purification procedure based on the method described by Prongay *et al*. (1990) was established to determine the properties of the chloroplast-expressed p24 protein in leaves of

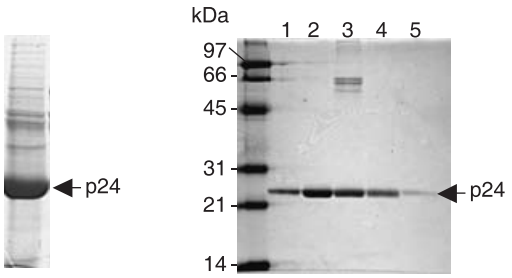

Figure 6 Purification and analysis of recombinant p24 protein from pZF5-yellow leaves. Left panel: Coomassie blue-stained gel of p24 extracted from pZF5-yellow leaves after two rounds of ammonium sulphate precipitation and resuspension in sodium citrate/citric acid buffer pH 4.5. Right panel: Coomassie blue-stained gel of p24 extracted from pZF5-yellow leaves and purified by cation-exchange chromatography. Lanes 1–5 are fractions eluted from a Mono S-Sepharose column at pH 6.2, 6.5, 6.7, 6.9 and 7.2.

pZF5-yellow plants. p24 protein at ~90% purity could be obtained simply by carrying out two rounds of ammonium sulphate precipitation on the cleared homogenate of fresh leaves of pZF5-yellow plants (Figure 6). The inclusion of SDS (0.1% w/v) or Triton-X-100 (1% v/v) in the homogenization medium did not increase the yield of p24 significantly. Extraction of the leaves with 8 M urea also had no effect on the yield of p24, indicating that recombinant HIV-1 p24 is in a soluble form in the chloroplast. HIV-1 p24 was efficiently separated from other proteins in the ammonium sulphate precipitate by cation-exchange chromatography on a spin column of Mono S-Sepharose using step-wise elution with sodium citrate/citric acid buffers of increasing pH (Figure 6). SDS-PAGE of the protein fractions and subsequent staining with Coomassie blue (Figure 6) or silver staining (data not shown) showed that elution at pH 6.5 yielded the purest p24 material.

N-terminal sequencing and electrospray ionization mass spectrometry

N-terminal sequencing of the p24 protein enriched by ammonium sulphate precipitation from pZSJH1p24 Maryland Mammoth plants and of the purified pZF5-yellow p24-cco protein showed that *N*-formylmethionine had been removed in both cases (Figure 7a). Electrospray ionization quadrupole time-of-flight mass spectrometry (ESI-QTOF-MS) analysis was used to determine the molecular mass of the purified protein. ESI-QTOF-MS analysis of the pH 6.5 fraction from the ion-exchange purification of pZF5-yellow p24-cco protein resulted in a major peak of 26 155.31 Da (Figure 7b). This was close to the theoretical mass (26 158 Da) of a 237-amino-acid p24-cco protein with an N-terminal sequence of ASGSPIVQNI. A

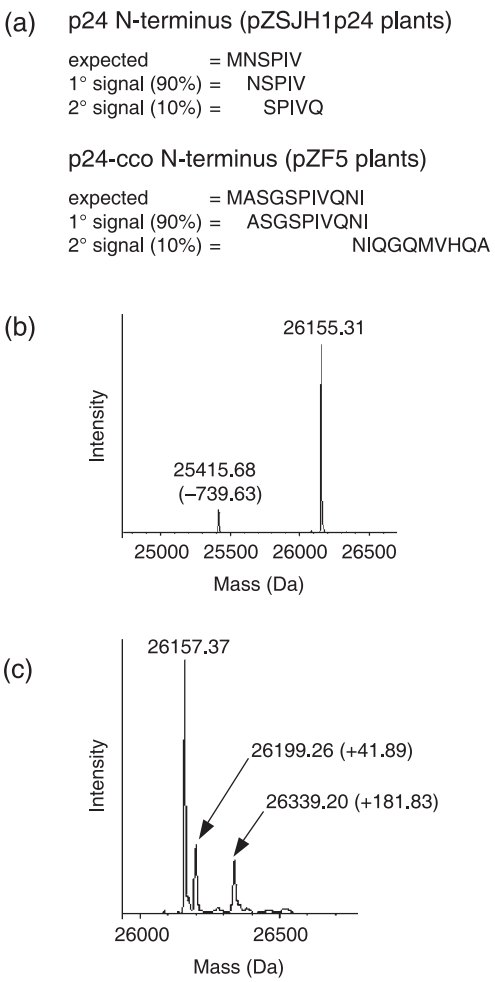

Figure 7 N-terminal sequencing and electrospray ionization quadrupole time-of-flight mass spectrometry (ESI-QTOF-MS) of purified p24. (a) N-terminal sequence of ammonium sulphate-enriched p24 from pZSJH1p24 plants and purified pZF5 chloroplast p24 pH 6.5 fraction. (b) Mass spectrum of purified pZF5 chloroplast p24 pH 6.5 fraction showing a major peak at 26 155.31 Da. The theoretical mass of a 237-amino-acid pZF5-expressed p24 protein with an N-terminus of ASGSPIVQ is 26 158 Da. The minor peak at 25 415.618 Da corresponds to the loss of eight amino acid residues (ASGSPIVQ, 739.83 Da) from the pZF5 p24 N-terminus. (c) Mass spectrum of purified pZF5 chloroplast p24 pH 6.2, 6.7, 6.9, 7.2 fractions combined showing the major peak at 26 157.37 Da. Masses indicated are labels that were added to the peaks by the software linked to the QTOF instrument. For additional peaks, masses relative to the major peak are indicated in brackets.

minor peak at 25 415.68 Da was close to the theoretical mass (25 418 Da) of a truncated p24-cco protein with an N-terminal sequence of NIQGQMVHQA. The 3-Da difference between the determined and predicted masses of the proteins is within the accuracy of the method (±1–3 Da) in this mass range. No other mass peaks were detected in the pH 6.5 fraction, suggesting that the p24 protein in this fraction was not modified post-translationally. To examine

whether modified forms of p24 occurred in other fractions, the pooled pH 6.2, 6.7, 6.9 and 7.2 fractions eluted from Mono S-Sepharose were examined by ESI-QTOF-MS (Figure 7c). In addition to the major form at 26 157.37 Da, two extra peaks representing mass increases of 41.89 Da and 181.83 Da were detected (Figure 7c). The additional peak of 181.83 Da is probably caused by covalent attachment of 4-(2-aminoethyl) benzenesulphonyl fluoride HCl (AEBSF), which is one of the components of the Halt Protease Inhibitor Cocktail (Pierce/ Thermo Fisher Scientific, Rockford, IL, USA) used during the purification of p24. AEBSF commonly leads to a mass increase of 183 Da in protein samples isolated in the presence of this inhibitor [Len Packman, Protein and Nucleic Acid Chemistry Facility (PNAC), Department of Biochemistry, University of Cambridge, Cambridge, UK, pers. commun.]. The peak corresponding to +41.89 Da could be caused by formylation plus methylation (+42 Da). Recombinant HIV-1 p24 has been reported to be formylated at the N-terminal proline (Pro) in CEM/LAV-1 T-cell lines (Fuchigami *et al*., 2002). Further study is needed to confirm this. No mass increases that could be attributed to glycosylation or phosphorylation were observed in ESI-QTOF-MS analysis of p24 purified from chloroplasts, or by matrix-assisted laser desorption ionization (MALDI)-TOF-MS analysis of tryptic digests (data not shown), despite the inclusion of phosphatase inhibitor (Halt Phosphatase Inhibitor Cocktail, Pierce) throughout the protein extraction and purification procedure.

Discussion

This study has shown that the high-level expression of a chloroplast-codon-optimized gene for HIV-1 p24 protein occurs in all leaves of mature Maryland Mammoth plants, providing gram-scale amounts of protein from a single plant. Purification of the protein and characterization by mass spectrometry showed that p24 protein expressed in tobacco chloroplasts is not modified by glycosylation or phosphorylation, and is therefore a good candidate as an injectable HIV-1 antigen.

There are surprisingly few published reports of chloroplast transformation of tobacco varieties other than the small variety Petite Havana. This is possibly the result of the low transformation frequencies that have been obtained with several other tobacco varieties. Chloroplast transformation of the TMV-resistant variety Samsun with an *aadA* gene was used for insertional inactivation of the *ycf 9* gene (Mäenpää *et al*., 2000). Chloroplast-transformed Samsun and the nicotine-free variety 22X-1 (derived from the commercial variety K327) have also been used to express a thermostable microbial cellulase (Yu *et al*., 2007). The expression of human interferon-α was compared between Petite Havana and a low-nicotine variety tobacco (Arlen *et al*., 2007). Trehalose phosphate synthase has been expressed in tobacco cultivars Xanthi and Burley (Lee *et al*., 2003). For Maryland Mammoth, the seemingly trivial step of initiating plants from nodal cuttings rather than directly from seed was necessary to generate leaves that were transformable. This step may be applicable to other tobacco varieties that are either recalcitrant to chloroplast transformation or give very low transformation frequencies.

Yields of p24 in Maryland Mammoth leaves obtained by chloroplast transformation were up to 450 μg/g fresh weight (equivalent to ~4.5% TSP). This is similar to the amounts accumulated in young leaves of Petite Havana (Zhou *et al*., 2008), but the p24-cco protein accumulates in all the leaves of Maryland Mammoth pZF5-yellow plants, giving a much higher potential yield. The amounts of p24 produced by chloroplast transformants far exceed the yields of p24 reported for nuclear transformation and viral transfection. As mentioned earlier, only 0.35% TSP (~35 μg/g fresh weight of leaf tissue) was reported in nuclear-transformed tobacco (Zhang *et al*., 2002). Pérez-Filgueira *et al*. (2004) reported up to 80 μg/g fresh weight (~0.8% TSP) of p24 in tobacco leaf tissue transfected with a TMV-derived vector.

The production and purification of p24-cco from pZF5-yellow leaves were relatively straightforward and did not require the use of an affinity tag. Purification using affinity tags requires expensive resins that cannot be employed on a large scale. After the homogenization of 100 g of leaf tissue, two ammonium sulphate precipitations and resuspension in sodium citrate buffer yielded approximately 20–30 mg of 90% pure protein. However, more than 90% of the protein was lost in subsequent ion-exchange steps as a result of protein precipitation; therefore, further optimization is required. The purified protein also retained four extra amino acids (ASGS) at the N-terminus that may or may not alter the immunological properties of the protein. These properties are currently being tested in mice. As for human somatotrophin (Staub *et al*., 2000), a ubiquitin fusion strategy could be used to achieve p24 with a non-modified N-terminus (i.e. Pro), although such a protein may be unstable in leaves.

The difference in the patterns of accumulation of p24-cco protein in pZF5-yellow plants and of p24 in pZSJH1p24 plants is intriguing and could be the result of a number of factors, although protein stability appears to be the major determinant. Differences in the 5′UTRs upstream of the *p24* and *p24-cco* genes may result in different rates of translation of the transcripts from the two genes. Both genes were under the control of P*rrn*, but the *p24-cco* coding region in pZF5 was

preceded by the T7g10 leader, whereas the *p24* coding region in pZSJH1p24 was preceded by part of the *rbcL* 5′UTR. The T7g10 leader has been used in constructs for the high-level accumulation of neomycin phosphotransferase (23% TSP) (Kuroda and Maliga, 2001) and human somatotrophin (7% TSP) (Staub *et al*., 2000). However, when systematic comparisons were carried out, the T7g10 leader did not produce higher levels of β-glucuronidase (GUS) (Herz *et al*., 2005) or p24 protein expression in plastids (Zhou *et al*., 2008). The loss of p24 protein in leaves of pZSJH1p24 plants suggests that this protein is degraded as the leaves age, whereas the p24-cco protein is stable in the leaves of pZF5-yellow plants. It was initially thought that the most likely explanation for the differences in protein stability was the difference between the N-terminal sequences of p24 and p24-cco. The *p24* gene in pZSJH1p24 codes for Met-Asn at the N-terminus, whereas the *p24-cco* gene in pZF5 codes for Met-Ala. According to the 'N-end rule', certain N-terminal amino acid residues can have a destabilizing effect on a protein, whereas others have a stabilizing effect (Baker and Varshavsky, 1991). N-terminal sequencing of p24-cco purified from pZF5-yellow plants showed that the N-terminal amino acid was Ala. The initiating *N*-formylmethionine residue was probably removed from p24-cco by the sequential action of peptide deformylase and methionine aminopeptidase (MAP). This occurs in about 55% of plastid-encoded proteins (Giglione and Meinnel, 2001). *N*-terminal Met excision by MAP occurs if the second residue in the target chain is Ala, Gly, Pro, cysteine (Cys), Ser, threonine (Thr) or valine (Val) (Giglione *et al*., 2003). These amino acids have short side-chains. Met is retained if the second amino acid has a long side-chain, as in Asn, arginine (Arg), aspartic acid (Asp), Glu, isoleucine (Ile), Leu and lysine (Lys). On this basis, the predicted N-termini would be Ala for p24-cco and Met for p24. Although both Ala and Met are stabilizing *N*-terminal amino acids for certain proteins (Mogk *et al*., 2007), *N*-terminal Met excision is crucial for stabilizing chloroplast-encoded proteins, such as the photosystem II D2 subunit (Giglione *et al*., 2003). Retention of the N-terminal Met in D2 by substituting the second amino acid (Thr2Asp or Thr2Glu), so that MAP activity was prevented, severely reduced protein stability (Giglione *et al*., 2003). However, N-terminal sequencing of p24 expressed in pSZJH1p24 plants showed that having Asn as the second amino acid in the p24 sequence does not appear to prevent N-terminal Met excision. Nevertheless, plastid-expressed rotavirus VP6 protein produced from the same pZSJH1 vector as p24, and thus coding for the same N-terminus as p24 (Met-Asn-Ser), was also found to be unstable in tobacco leaves (Birch-Machin *et al*., 2004). Extension of the N-terminus of plastid-expressed GUS with Met-Ala-Ser-Ile-Ser enhanced protein accumulation by an order of magnitude from 0.77% TSP without the N-terminal extension to 8.8% TSP with the extension (Herz *et al*., 2005). However, plastid-expressed GFP with Met-Asn-Ser at the N-terminus, by expression from a pZSJH1 vector, was stable and accumulated at detectable levels in all leaves of a flowering tobacco plant (Newell *et al*., 2003; data not shown). This indicates that the putative destabilizing effect of Met-Asn-Ser at the N-terminus does not necessarily apply to all proteins.

Recombination between the two 141-bp *rbcL* 3′UTRs in direct orientation in pZF5 plants has not been reported previously, and was not detected when the same vector format was used for high-level expression of a p24-Nef fusion protein in Petite Havana (Zhou *et al*., 2008). *rbcL* 3′UTRs have also been used previously for the expression of TetC from the *trnV/rps12/7* intergenic spacer in the inverted repeat region of the chloroplast genome, and recombination with the native *rbcL* 3′UTR was not reported (Tregoning *et al*., 2003). The 21-kb proximity of the native chloroplast *rbcL* gene to the *trnG/trnfM* intergenic spacer insertion site may be related to recombination between the introduced and native *rbcL* 3′UTRs in pZF5 plants. However, the proximity of repeated sequences on the same chloroplast genome does not necessarily increase the likelihood of their recombination. Recombination causing a large inversion between an introduced *psbA* 3′UTR in the *rbcL/accD* intergenic spacer and native *psbA* 3′UTRs in transformed chloroplast genomes has also been reported, despite their separation by nearly 60 kb (Rogalski *et al*., 2006). A similar deletion event to that described for the present work was reported by Kode *et al*. (2006). In that case, recombination between direct 649-bp repeats of the native and an introduced *atpB* gene promoter region led to the excision of a 6.1-kb section of a transformed chloroplast genome.

Plastid extrachromosomal elements resulting from chloroplast transformation have been described previously, but these were maintained by antibiotic selection (Staub and Maliga, 1994; Ahlert *et al*., 2003; Kode *et al*., 2006). The 21-kb excision product in pZF5-yellow plants is stable *in vivo*. The reduced native TSP (mostly ribulose-1,5-bisphosphate carboxylase/oxygenase, RUBISCO) and the yellow phenotype in pZF5-yellow plants must be somehow related to the abundance of the 21-kb circle and lack of full-length chloroplast genomes, as indicated by the Southern blots. However, it is only possible to speculate that the *rbcL* gene and the other genes encoded on the 21-kb circle are somehow less well expressed than on the full-length chloroplast genome. Alternatively, it is also possible that the accumulation of the p24-cco protein (up to 4.5% TSP) causes yellowing in pZF5-yellow plants, whereas the non-codon-optimized p24

expressed from pZSJH1p24 (up to 2.5% TSP) does not. In chloroplast-transformed tobacco transformed with different versions of genes (one AT rich and the other GC rich) coding for the same protein (TetC), plants transformed with *tetC-AT* were yellow and slow growing, whereas plants transformed with *tetC-GC* were normal (Tregoning *et al.*, 2003). The different phenotypes in this case were ascribed to a two-fold increase in percentage TSP of the recombinant TetC. In the present study, the pZF5-yellow phenotype is probably not caused by the level of p24-cco accumulation, as older pZF5-yellow leaves contained less than 2.5% TSP and remained yellow.

It is interesting that, in pZF5-green plants, the degree of recombination between the *rbcL* 3′UTR repeats was reduced relative to that in pZF5-yellow plants, and 'flip-flop' inversion of the deleted and full-length P*rrn* had occurred, so that p24 protein accumulation did not occur in pZF5-green plants. It would be interesting to determine whether these two recombination events are independent or connected, and what induces them. In practical terms, if it was possible to control the 'flip-flop' recombination between the deleted and non-deleted P*rrn* promoters, this could potentially be an inducible system for the expression of recombinant proteins in chloroplasts.

This work demonstrates that HIV-1 p24 protein of the correct size, without any post-translational modifications, such as glycosylation or phosphorylation, and detectable with conformation-dependent antibody, can be produced in large amounts in Maryland Mammoth tobacco plastids. Further optimization and refinement of expression are necessary to utilize fully the potential of the high biomass yield of Maryland Mammoth. In addition, improvements in the purification of p24 are needed to produce protein of high purity in high yield. However, the prospects for an improved production platform for the large-scale production of pharmaceutical proteins in plants appear to be very good.

Experimental procedures

Plant material

Nicotiana tabacum cv. Maryland Mammoth seeds were obtained from IPK Gatersleben, Germany.

p24 constructs

pZF5 and pZSJH1p24 construction has been described in Zhou *et al.* (2008). The sequence of the *p24-cco* gene used in pZF5 is available under GenBank accession number EU670669. The sequence of the *p24* gene used in pZSJH1p24 is available under GenBank accession number KO3455.1 (nucleotides 1186–1878).

Chloroplast transformation

Seeds were surface sterilized in 20% domestic bleach for 10 min, rinsed three times with sterile distilled water and then allowed to germinate in the light on half-strength Murashige and Skoog (MS) salts and vitamins, 1% sucrose and 0.7% Phytagar (Duchefa, Biochemie BV, Haarlem, the Netherlands). After germination, seedlings were transferred to an agar medium containing full-strength MS salts and B5 vitamins (Duchefa). When plants were at the four- or five-leaf stage, they were cut into several nodal sections and re-rooted on agar medium containing MS salts and B5 vitamins. Leaves (length, 3–5 cm) from plants derived from nodal cuttings were used for bombardment. Chloroplast transformation was carried out as described by Birch-Machin *et al.* (2004). Spectinomycin-resistant shoots were screened for the insertion of the *aadA* gene and the gene of interest by long-range PCR, as described below. Shoots identified as transformants were subjected to a further three rounds of regeneration on 500 mg/L spectinomycin; T_0 plants were then transferred to soil and allowed to self-pollinate and set seed. To determine homoplasmy, approximately 300–400 T_1 seeds were surface sterilized and germinated on agar medium containing half-strength MS and 1% sucrose supplemented with 500 mg/L spectinomycin dihydrochloride pentahydrate (Duchefa). The absence of segregation of spectinomycin resistance was considered as confirmation of homoplasmy.

PCR analysis

pZF5 plants were screened by long-range PCR for insertion of the *p24-cco* and *aadA* genes into the intergenic spacer between *trnfM* and *trnG* in the chloroplast genome (Figure 1) using primers psaB (5′-TTCTTCGTAGGTCATTTGTGGC-3′) and psbC (5′-CATTTATTTT-TCTCGTTGCTGGTCAT-3′), and for short-range PCR using primers psaB and aadAP6 (5′-AGCAAATCAATATCACTGTGTGGC-3′). Primers psaB and pcbC bind to sequences on the tobacco chloroplast genome external to the chloroplast sequence in pZF5, which is derived from pRB95 (Ruf *et al.*, 2001), and amplify a product corresponding to nucleotide positions 36 808 to 40 342 on the tobacco chloroplast genome (Wakasugi *et al.*, 1998). aadAP6 binds in reverse orientation just upstream of the P*rrn* promoter on the *aadA* gene. pZSJH1p24 plants were screened for insertion of the *p24* gene and the *aadA* gene into the intergenic spacer between the *rbcL* and *accD* genes on the tobacco chloroplast genome by long-range PCR using primers rbcL (5′-CCTCAACCTGGAGTTCCACCTGAAGAAGCAGGG-3′) and accD (5′-GCTGTCACTCCACCAGTTGTCGGAGATGTAAGG-3′), and by short-range PCR using primers rbcL and aadAP6. Primers rbcL and accD bind to sequences on the tobacco chloroplast genome external to the chloroplast sequence in pZSJH1p24, which is derived from pZS197 (Svab and Maliga, 1993), and amplify a product corresponding to nucleotide positions 57 724 to 61 081 on the tobacco chloroplast genome. PCR was carried out using AccuTaq (Sigma-Aldrich, St. Louis, MO, USA) according to the manufacturer's instructions. PCR conditions for all primer combinations were 39 cycles of 98 °C (30 s), 55 °C (30 s), 68 °C (1 min per kb of expected amplification product), followed by an extension step at 68 °C (10 min). Total DNA (1 μL), extracted from leaves using the GenElute™ Plant Genomic DNA Miniprep Kit (Sigma), was used as template in each 50-μL reaction.

For the analysis of possible recombination products in pZF5 plants, primer combinations were as follows: PCR1, primers accD and rbcL;

PCR2, primers accD and aadAP6; PCR3, primers rps14.2 (5′-CATTATGGCAAGGAAAAGTT-3′) and rbcL; PCR4, primers rps14.2 and rbcLF6 (5′-GGCTGAAACAGGTGAAATCAAAGG-3′). rbcLF6 binds approximately 600 bp downstream of the start of *rbcL*. PCR1–4 were carried out using AccuTaq (Sigma) according to the manufacturer's instructions, with 35 cycles of 98 °C (30 s), 55 °C (30 s), 68 °C (5 min), followed by an extension step at 68 °C (7 min). A 1 : 20 dilution of total DNA, extracted from leaves using the GenElute™ Plant Genomic DNA Miniprep Kit (Sigma), was used as template in each 50-μL reaction.

For sequencing orientation of P*rrn* promoters in the 'flip-flop' recombination, primers P24-cco (5′-CCTTCAGATAAAGCAGAAAACATAGGAATTACTTCAGGA-3′) and aadAP6 were used to amplify the promoter region. The PCR product was cloned into a PCR cloning vector (PCR2.1 TOPO, Invitrogen Corporation, Carlsbad, CA, USA), and end sequenced using M13 forward and reverse primers.

For quantitative PCR, RNA was extracted from leaves using Tri Reagent® (Applied Biosystems, Foster City, CA, USA), according to the manufacturer's instructions. First-strand cDNA synthesis was performed using SuperScript III™ (Invitrogen), according to the manufacturer's instructions, with oligo$(dT)_{20}$ (2.5 μM) and random hexamers (2.5 ng/μL). Real-time PCR was performed using Power SYBR Green PCR Master Mix (Applied Biosystems) on a Chromo 4 real-time PCR instrument (MJ Research, Inc., South San Francisco, USA). Primers 5′-CTTTATCTGAAGGAGCTACTCCTC-3′ and 5′-GTCCAGGAGCAATAGGTCCAGCAT-3′ were used to amplify the *p24-cco* transcript cDNA, and primers 5′-GAGCGCCATCTCGAACCGACGTTGC-3′ and 5′-GACTTCTACAGCGCGGAGAATCTC-3′ were used to amplify the *aadA* transcript cDNA.

Southern blot analysis

Non-radioactive Southern blot analysis was performed as described by McCabe *et al*. (1997), except that DNA was prepared according to Frey (1999). For pZF5 plants, a probe specific for the synthetic *p24-cco* gene in pZF5 was prepared by PCR DIG probe synthesis from pZF5 template using the primers 5′-CATATGGCTAGCGGATCCCTATT-3′ and 5′-tctagaggaattctactcagctttatgtc-3′. A probe specific for the tobacco chloroplast *rps14* gene was also prepared by PCR DIG probe synthesis from pZF5 template using the primers 5′-AAACAAGATGTTGCGGAGAC-3′ and 5′-CATTATGGCAAGGAAAAGTT-3′. These amplify a product corresponding to nucleotides 38 365 to 38 885 on the tobacco chloroplast genome.

For pZSJH1p24 plants, probes specific for the tobacco chloroplast *accD* and *rbcL* genes were prepared by PCR DIG probe synthesis from pZSJH1p24 template. Primers 5′-TTTATTCGGCGTTAAAGACATTCG-3′ and 5′-TCCATAACAATTTTCGCATTGAACC-3′ amplify a product corresponding to nucleotides 60 026 to 60 524 in the *accD* gene, and primers 5′-GGCTGAAACAGGTGAAATCAAAGG-3′ and 5′-TCCTAAAGTTCCTCCACCGAACTG-3′ amplify a product corresponding to nucleotides 58 281 to 58 795 in the tobacco chloroplast genome.

Protein immunoblot analysis

Proteins were extracted from leaves by homogenization in cold protein extraction buffer [50 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulphonic acid (HEPES)/KOH pH 7.5, 1 mM disodium ethylenediaminetetraacetate, 10 mM potassium acetate, 5 mM magnesium acetate, 2 mM phenylmethylsulphonylfluoride (PMSF), 30 mM 2-mercaptoethanol] using 4 mL of extraction buffer for every 1 g of leaf tissue (typically 100 mg of leaf tissue in 400 μL of buffer). Proteins were subjected to SDS-PAGE with a 12% resolving gel and a 5% stacking gel, and then transferred to Optitran BA-S85 reinforced nitrocellulose (Whatman/GE Healthcare, Maidstone, Kent, UK) using a Trans-blot SD semi-dry transfer cell (Bio-Rad Laboratories, Hercules, CA, USA). The blots were incubated at room temperature for 1 h in blocking solution (phosphate-buffered saline, 0.1% Tween 20, 2% non-fat dried milk), and then overnight at 4 °C in sheep polyclonal anti-HIV-1-p24 antibody (code no. D7320, Aalto Bioreagents, Dublin, Ireland) diluted to 1 : 500 in blocking solution. The blots were then washed in phosphate-buffered saline (PBS) and 0.1% Tween 20, and incubated in horseradish peroxidase-conjugated sheep anti-goat monoclonal antibody (code A9452, Sigma) diluted to 1 : 1000 in blocking solution. The blots were then developed by chemiluminescence and exposed to BioMax Light Film (Eastman Kodak, Rochester, NY, USA).

ELISA

Leaves (100 mg) were homogenized in 400 μL of protein extraction buffer [2.5% bovine serum albumin (BSA), 0.1% Tween 20, 2 mM PMSF, 30 mM 2-mercaptoethanol]. Maxisorp 96-well microplates (Nunc/Thermo Fisher Scientific, Rockford, IL, USA) were coated overnight at 4 °C with 50 μL of sheep polyclonal anti-HIV-1-p24 antibody (code no. D7320, Aalto Bioreagents) diluted at a concentration of 7.5 μg/mL in 100 mM sodium carbonate/bicarbonate buffer pH 9.6. The wells were washed three times with 0.1% Tween 20/PBS and then blocked with 200 μL of blocking solution (2.5% BSA/PBS without Tween) for 2 h at 37 °C (or overnight at 4 °C). Blocking solution was removed by inverting the plate, and replaced with 100 μL of leaf protein extract diluted 1 : 200 in 2.5% BSA/PBS. After incubation for 2 h at 37 °C, the wells were washed four times with 0.1% Tween 20/PBS. Conformation-dependent biotinylated anti-p24 (code no. BC1071-BIOT, Aalto Bioreagents), diluted 1 : 500 in 2.5% BSA/PBS, was then added to each well, and the plate was incubated at 37 °C for 1 h. After four washes with 0.1% Tween 20/PBS, 100 μL streptavidin–alkaline phosphatase conjugate (Hoffmann-La Roche, Basel, Switzerland), diluted to 1 : 1000 in 2.5% BSA/PBS, was added to the wells, and the plate was incubated for 30 min at 37 °C. The wells were then emptied and washed four times with 0.1% Tween 20/PBS, and 200 μL of *p*-nitrophenyl phosphate (1.0 mg/mL), prepared with SIGMA*FAST* tablets according to the manufacturer (Sigma), were added. The plate was incubated at room temperature for 10 min, and colour development was stopped by the addition of 100 μL of 1 M sodium hydroxide. Optical densities were read at 405 nm on an MWG Lambda Scan 200 (MWG, Ebersberg, Germany). Amounts of p24 in leaf extracts were determined by comparison with recombinant HIV-1 p24 purified from *E. coli* (code no. AG6054, Aalto Bioreagents) using concentration values indicated by the supplier. To calculate the amount of p24 as percentage TSP, the amount of total protein in leaf extracts was measured using the Bio-Rad protein assay, with BSA as a standard, according to the manufacturer's instructions.

Purification of p24

For protein extraction, typically 100 g of leaves were homogenized in 400 mL of cold extraction buffer (50 mM HEPES/KOH pH 7.5,

1 mM disodium ethylenediaminetetracetate, 10 mM potassium acetate, 5 mM magnesium acetate, 2 mM PMSF, 30 mM 2-mercaptoethanol). For proteins analysed by mass spectrometry, Halt Protease Inhibitor Cocktail (Pierce) and Halt Phosphatase Inhibitor Cocktail (Pierce) were added, according to the manufacturer's instructions. The homogenate was filtered through miracloth and centrifuged at 16 000 ***g*** at 4 °C for 30 min. Ammonium sulphate was then slowly added to the supernatant to a concentration of 176 g/L with continuous stirring. The solution was incubated on ice for 2 h and then centrifuged for 30 min at 16 000 ***g***. The supernatant was discarded and the pelleted protein was resuspended in 10 mL of 50 mM Tris-HCl pH 8.0, 30 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA) and 2 mM PMSF. The protein solution was passed through a DE52 cellulose column (1 cm × 7 cm), buffer exchanged by concentrating to 2 mL in a 15-mL Amicon Ultra centrifugal filter (Amicon/Millipore Coporation, Billerica, MA, USA) with a 10-kDa cut-off (Millipore), and then resuspended in 15 mL of 10 mM sodium citrate/citric acid buffer pH 4.5. Concentration and resuspension were repeated four times. The final concentration was to 2 mL. Alternatively, the pelleted protein was resuspended directly in 10 mM sodium citrate/citric acid buffer pH 4.5, and then desalted by four 7.5-fold concentrations and resuspensions in 10 mM sodium citrate/citric acid buffer pH 4.5 using a 15-mL Amicon Ultra centrifugal filter with a 10-kDa cut-off (Millipore). The solution was then passed through a 0.45-μm filter and loaded on to 500-μL capacity Mono S-Sepharose cation-exchange spin columns, according to the manufacturer's instructions (Sartorius-Vivascience, Aubagne, France). The protein was eluted from the spin column by the step-wise addition of 10 mM sodium citrate/citric acid buffers of increasing pH, according to the manufacturer's instructions.

N-terminal sequencing and mass spectrometry

Pure protein (~50 pmol) was transferred to poly(vinylidene difluoride) (PVDF) membrane by semi-dry blotting, and the membrane was stained with Coomassie blue to locate the protein. The N-terminal sequence of the blotted protein was determined by Edman degradation at the Protein and Nucleic Acid Chemistry Facility (PNAC), Department of Biochemistry, University of Cambridge, Cambridge, UK. Mass determination of chloroplast-derived p24-cco protein in solution was obtained by electrospray ionization mass spectrometry conducted on a high-mass quadrupole time-of-flight instrument adapted for a QSTAR XL platform (ESI-TOF-MS Q-STAR, ABI/MDS Sciex, South San Francisco, USA) at the Department of Chemistry, University of Cambridge, Cambridge, UK. Peptide analysis on tryptic digests of p24-cco was performed by nano-LC-MALDI-TOF-MS, also at the Department of Chemistry, University of Cambridge, Cambridge, UK, and MALDI-TOF-MS at the Mass Spectrometry Unit, Biology Department, National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland.

Acknowledgements

We are grateful to Dr Christine Newell and Sue Aspinall for help and advice. We also thank Caroline Batchelor (Mass Spectrometry Unit, Biology Department, National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland) and Dr Sean Doyle (Biotechnology, Biology Department, National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland) for help and advice with the analysis and purification of proteins, Sarah Maslen (Department of Chemistry, University of Cambridge, Cambridge, UK) for mass spectrometry analyses, and Mike Weldon and Len Packman [Protein and Nucleic Acid Chemistry Facility (PNAC), Department of Biochemistry, University of Cambridge, Cambridge, UK] for *N*-terminal sequencing and discussions of post-translational modification. This work was supported by grants from the European Commission FP6 PharmaPlanta Integrated project to PD, JCG and RB.

References

Ahlert, D., Ruf, S. and Bock, R. (2003) Plastid protein synthesis is required for plant development in tobacco. *Proc. Natl. Acad. Sci. USA*, **100**, 15 730–15 735.

Allain, J.P., Laurian, Y., Paul, D.A. and Senn, D. (1986) Serological markers in early stages of human immunodeficiency virus infection in haemophiliacs. *Lancet*, **2**, 1233–1236.

Amann, E. and Brosius, J. (1985) 'ATG vectors' for regulated high-level expression of cloned genes in *Escherichia coli*. *Gene*, **40**, 183–190.

Arlen, P.A., Falconer, R., Cherukumilli, S., Cole, A., Cole, A.M., Oishi, K.K. and Daniell, H. (2007) Field production and functional evaluation of chloroplast-derived interferon-alpha2b. *Plant Biotechnol. J.* **5**, 511–525.

Baker, R.T. and Varshavsky, A. (1991) Inhibition of the N-end rule pathway in living cells. *Proc. Natl. Acad. Sci. USA*, **88**, 1090–1094.

Birch-Machin, I., Newell, C.A., Hibberd, J.M. and Gray, J.C. (2004) Accumulation of rotavirus VP6 protein in chloroplasts of transplastomic tobacco is limited by protein stability. *Plant Biotechnol. J.* **2**, 261–270.

Chebolu, S. and Daniell, H. (2007) Stable expression of Gal/GalNAc lectin of *Entamoeba histolytica* in transgenic chloroplasts and immunogenicity in mice towards vaccine development for amoebiasis. *Plant Biotechnol. J.* **5**, 230–239.

Coleman, J.K., Pu, R., Martin, M., Sato, E. and Yamamoto, J.K. (2005) HIV-1 p24 vaccine protects cats against feline immunodeficiency virus infection. *AIDS*, **19**, 1457–1466.

Daniell, H. (2006) Production of biopharmaceuticals and vaccines in plants via the chloroplast genome. *Biotechnol. J.* **10**, 1071–1079.

Daniell, H., Lee, S.B., Panchal, T. and Wiebe, P.O. (2001) Expression of the native cholera toxin B subunit gene and assembly as functional oligomers in transgenic tobacco chloroplasts. *J. Mol. Biol*. **311**, 1001–1009.

Ehrlich, L.S., Krausslich, H.G., Wimmer, E. and Carter, C.A. (1990) Expression in *Escherichia coli* and purification of human immunodeficiency virus type 1 capsid protein (p24). *AIDS Res. Hum. Retrovir*. **10**, 1169–1175.

Fischer, R., Stoger, E., Schillberg, S., Christou, P. and Twyman, R.M. (2004) Plant-based production of biopharmaceuticals. *Curr. Opin. Plant Biol*. **7**, 152–158.

Freed, E.O. (1998) HIV-1 Gag proteins: diverse functions in the virus life cycle. *Virology*, **251**, 1–15.

Frey, J.E. (1999) Genetic flexibility of plant chloroplasts. *Nature*, **398**, 115–116.

Fuchigami, T., Misumi, S., Takamune, N., Takahashi, I., Takama, M. and Shoji, S. (2002) Acid-labile formylation of amino terminal proline of human immunodeficiency virus type 1 p24 (gag) was found by proteomics using two-dimensional gel electrophoresis and matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry. *Biochem. Biophys. Res. Commun*. **293**, 1107–1113.

Garner, W.W. and Allard, H.A. (1920) Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants. *J. Agric. Res*. **18**, 553–606.

Garner, W.W. and Allard, H.A. (1923) Further studies in photoperiodism, the response of the plant to relative length of day and night. *J. Agric. Res*. **23**, 871–920.

Gay, B., Tournier, J., Chazal, N., Carrière, C. and Boulanger, P. (1998) Morphopoietic determinants of HIV-1 Gag particles assembled in baculovirus-infected cells. *Virology*, **247**, 160–169.

Giglione, C. and Meinnel, T. (2001) Organellar peptide deformylases: universality of the N-terminal methionine cleavage mechanism. *Trends Plant Sci*. **6**, 566–571.

Giglione, C., Vallon, O. and Meinnel, T. (2003) Control of protein life-span by N-terminal methionine excision. *EMBO J*. **22**, 12–23.

Girard, M.P., Osmanov, S.K. and Kieny, M.P. (2006) A review of vaccine research and development: the human immunodeficiency virus (HIV). *Vaccine*, **24**, 4062–4081.

Glenz, K., Bouchon, B., Stehle, T., Wallich, R., Simon, M.M. and Warzecha, H. (2006) Production of a recombinant bacterial lipoprotein in higher plant chloroplasts. *Nat. Biotechnol*. **24**, 76–77.

Hager, M., Biehler, K., Illerhaus, J., Ruf, S. and Bock, R. (1999) Targeted inactivation of the smallest plastid genome-encoded open reading frame reveals a novel and essential subunit of the cytochrome b6f complex. *EMBO J*. **18**, 5834–5842.

Hausdorf, G., Gewiess, A., Wray, V. and Porstmann, T. (1994) A recombinant human immunodeficiency virus type-1 capsid protein (rp24): its expression, purification and physico-chemical characterisation. *J. Virol. Methods*, **50**, 1–9.

Herz, S., Füssl, M., Steiger, S. and Koop, H.-U. (2005) Development of novel types of plastid transformation vectors and evaluation of factors controlling expression. *Transgenic Res*. **14**, 969–982.

International AIDS Vaccine Initiative (2006) URL http://www.iavi.org.

International AIDS Vaccine Initiative (2006/2007) URL http://www.iavireport.org/specials/OngoingTrialsofPreventiveHIVVaccines.pdf, http://chi.ucsf.edu/vaccines/vaccines?page=vc-03-00#S1X [accessed on July 2008].

Iweala, O.I. (2004) HIV diagnostic tests: an overview. *Contraception*, **70**, 141–147.

Kode, V., Mudd, E.A., Iamtham, S. and Day, A. (2006) Isolation of precise plastid deletion mutants by homology-based excision: a resource for site-directed mutagenesis, multi-gene changes and high-throughput plastid transformation. *Plant J*. **46**, 901–909.

Koya, V., Moayeri, M., Leppla, S.H. and Daniell, H. (2005) Plant-based vaccine: mice immunized with chloroplast-derived anthrax protective antigen survive anthrax lethal toxin challenge. *Infect. Immun*. **73**, 8266–8274.

Kuroda, H. and Maliga, P. (2001) Complementarity of the 16S rRNA penultimate stem with sequences downstream of the AUG destabilizes the plastid mRNAs. *Nucleic Acids Res*. **29**, 970–975.

Lee, S.B., Kwon, H.B., Kwon, S.J., Park, S.C., Jeong, M.J., Han, S.E., Byun, M.O. and Daniell, H. (2003) Accumulation of trehalose within transgenic chloroplasts confers drought tolerance. *Mol. Breed*, **11**, 1–13.

Lu, Y., Friedman, R., Kushner, N., Doling, A., Thomas, L., Touzjian, N., Starnbach, M. and Lieberman, J. (2000) Genetically modified anthrax lethal toxin safely delivers whole HIV protein antigens into the cytosol to induce T cell immunity. *Proc. Natl. Acad. Sci. USA*, **97**, 8027–8032.

Mäenpää, P., Gonzalez, E.B., Chen, L., Khan, M.S., Gray, J.C. and Aro, E.M. (2000) The *ycf 9* (orf 62) gene in the plant chloroplast genome encodes a hydrophobic protein of stromal thylakoid membranes. *J. Exp. Bot*. **51**, 375–382.

Maliga, P. (2002) Engineering the plastid genome of higher plants. *Curr. Opin. Plant Biol*. **5**, 164–172.

Maliga, P. (2003) Progress towards commercialisation of plastid transformation technology. *Trends Biotechnol*. **21**, 20–28.

McCabe, M.S., Power, J.B., de Laat, A.M. and Davey, M.R. (1997) Detection of single-copy genes in DNA from transgenic plants by nonradioactive Southern blot analysis. *Mol. Biotechnol*. **7**, 79–84.

Mills, H.R. and Jones, I.M. (1990) Expression and purification of p24, the core protein of HIV, using a baculovirus–insect cell expression system. *AIDS*, **4**, 1125–1131.

Mogk, A., Schmidt, R. and Bukau, B. (2007) The N-end rule pathway for regulated proteolysis: prokaryotic and eukaryotic strategies. *Trends Cell Biol*. **17**, 165–172.

Molina, A., Hervás-Stubbs, S., Daniell, H., Mingo-Castel, A.M. and Veramendi, J. (2004) High-yield expression of a viral peptide animal vaccine in transgenic tobacco chloroplasts. *Plant Biotechnol. J*. **2**, 141–153.

Narciandi, R.E., Motlongo, J. and Figueroa, L.P. (2006) High level production of the recombinant gag24 protein from HIV-1 in *Escherichia coli*. *Afr. J. Biotechnol*. **5**, 103–107.

Newell, C.A., Birch-Machin, I., Hibberd, J. and Gray, J.C. (2003) Expression of green fluorescent protein from plastid and bacterial promoters in tobacco chloroplasts. *Transgenic Res*. **12**, 631–634.

Nkolola, J.P. and Essex, M. (2006) Progress towards an HIV-1 subtype C vaccine. *Vaccine*, **24**, 391–401.

Obregon, P., Chargelegue, D., Drake, P.M., Prada, A., Nuttall, J., Frigerio, L. and Ma, J.K. (2006) HIV-1 p24-immunoglobulin fusion molecule: a new strategy for plant-based protein production. *Plant Biotechnol. J*. **4**, 195–207.

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. *J. Virol. Methods*, **121**, 201–208.

Prongay, A.J., Smith, T.J., Rossmann, M.G., Ehrlich, L.S., Carter, C.A. and McClure, J. (1990) Preparation and crystallization of a human immunodeficiency virus p24-Fab complex. *Proc. Natl. Acad. Sci. USA*, **87**, 9980–9984.

Rogalski, M., Ruf, S. and Bock, R. (2006) Tobacco plastid ribosomal protein S18 is essential for cell survival. *Nucleic Acids Res*. **34**, 4537–4545.

Ruf, S., Biehler, K. and Bock, R. (2000) A small chloroplast-encoded protein as a novel architectural component of the light-harvesting antenna. *J. Cell Biol*. **149**, 369–377.

Ruf, S., Hermann, M., Berger, I.J., Carrer, H. and Bock, R. (2001) Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. *Nat. Biotechnol*. **9**, 870–875.

Ruf, S., Karcher, D. and Bock, R. (2007) Determining the transgene containment level provided by chloroplast transformation. *Proc. Natl. Acad. Sci. USA*, **104**, 6998–7002.

Sriraman, R., Bardor, M., Sack, M., Vaquero, C., Faye, L., Fischer, R., Finnern, R. and Lerouge, P. (2004) Recombinant anti-hCG antibodies retained in the endoplasmic reticulum of transformed plants lack core-xylose and core-alpha (1,3)-fucose residues. *Plant Biotechnol. J*. **2**, 279–287.

Staub, J.M., Garcia, B., Graves, J., Hajdukiewicz, P.T., Hunter, P., Nehra, N., Paradkar, V., Schlittler, M., Carroll, J.A., Spatola, L., Ward, D., Ye, G. and Russell, D.A. (2000) High-yield production of a human therapeutic protein in tobacco chloroplasts. *Nat. Biotechnol*. **18**, 333–338.

Staub, J.M. and Maliga, P. (1994) Extrachromosomal elements in tobacco plastids. *Proc. Natl. Acad. Sci. USA*, **91**, 7468–7472.

Svab, Z. and Maliga, P. (1993) High-frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. *Proc. Natl. Acad. Sci. USA*, **90**, 913–917.

Svab, Z. and Maliga, P. (2007) Exceptional transmission of plastids and mitochondria from the transplastomic pollen parent and its impact on transgene containment. *Proc. Natl. Acad. Sci. USA*, **104**, 7003–7008.

Tregoning, J.S., Nixon, P., Kuroda, H., Svab, Z., Clare, S., Bowe, F., Fairweather, N., Ytterberg, J., van Wijk, K.J., Dougan, G. and Maliga, P. (2003) Expression of tetanus toxin fragment C in tobacco chloroplasts. *Nucleic Acids Res*. **31**, 1174–1179.

Wakasugi, T., Sugita, M., Tsudzuki, T. and Sugiura, M. (1998) Updated gene map of tobacco chloroplast DNA. *Plant Mol. Biol. Rep*. **16**, 231–241.

Watson, J., Koya, V., Leppla, S.H. and Daniell, H. (2004) Expression of *Bacillus anthracis* protective antigen in transgenic chloroplasts of tobacco, a non-food/feed crop. *Vaccine*, **22**, 4374–4384.

Yu, L.X., Gray, B.N., Rutzke, C.J., Walker, L.P., Wilson, D.B. and Hanson, M.R. (2007) Expression of thermostable microbial cellulases in the chloroplasts of nicotine-free tobacco. *J. Biotechnol*. **131**, 362–369.

Zhang, G., Leung, C., Murdin, L., Rovinski, B. and White, A. (2000) In planta expression of HIV-1 p24 protein using an RNA plant virus-based expression vector. *Mol. Biotechnol*. **14**, 99–107.

Zhang, G.G., Rodrigues, L., Rovinski, B. and White, K.A. (2002) Production of HIV-1 p24 protein in transgenic tobacco plants. *Mol. Biotechnol*. **20**, 131–136.

Zhou, F. Badillo-Corona, J.A., Karcher, D., Gonzalez-Rabade, N., Piepenburg, K., Borchers, A.-M.I., Maloney, A.P., Kavanagh, T.A., Gray, J.C. and Bock, R. (2008) High-level expression of HIV antigens from the tobacco and tomato plastid genomes. *Plant Biotechnol. J*. doi: 10.1111/j.1467-7652.2008.00356.x.