

Expression of recombinant proteins lacking methionine as N-terminal amino acid in plastids: Human serum albumin as a case study

Alicia Fernández-San Millán, Inmaculada Farran, Andrea Molina,
Angel M. Mingo-Castel, Jon Veramendi *

Instituto de Agrobiotecnología, Universidad Pública de Navarra-CSIC, Campus Arrosadía, 31006 Pamplona, Spain

Received 28 February 2006; received in revised form 7 August 2006; accepted 15 August 2006

Abstract

Removal of the N-terminal methionine of a protein could be critical for its function and stability. Post-translational modifications of recombinant proteins expressed in heterologous systems may change amino-terminal regions. We studied the expression of mature proteins lacking methionine as the N-terminal amino acid in tobacco chloroplasts, using human serum albumin (HSA) as an example. Two approaches were explored. First, we fused the Rubisco small subunit transit peptide to HSA. This chimeric protein was correctly processed in the stroma of the chloroplast and rendered the mature HSA. The second approach took advantage of the endogenous N-terminal methionine cleavage by methionine aminopeptidase. Study of this protein processing reveals a systematic cleavage rule depending on the size of the second amino acid. Analysis of several foreign proteins expressed in tobacco chloroplasts showed a cleavage pattern in accordance to that of endogenous proteins. This knowledge should be taken into account when recombinant proteins with N-terminus relevant for its function are expressed in plastids.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Plastid transformation; Human serum albumin; Methionine removal; N-end rule; Post-translational modifications; Transit peptide

1. Introduction

The biosynthesis of the majority of proteins from all living organisms begins with either methionine or N-formylmethionine (in prokaryotes, mitochondria

and chloroplasts). The amino-terminal regions of many proteins are subjected to a variety of post-translational modifications, such as methylation, acetylation or phosphorylation of the N-terminal residue, or removal of one to several residues (e.g. the processing of leader peptides). Thus, the N-terminal methionine can be cleaved by a methionine aminopeptidase (MAP) with a very broad specificity. This specificity is determined by the size of the side chain of the

* Corresponding author. Tel.: +34 948 168033;
fax: +34 948 232191.

E-mail address: jon@unavarra.es (J. Veramendi).

penultimate residue (Sherman et al., 1985). The amino-terminal sequence may include information relevant for subcellular targeting or regulating protein life-span (Gigliione et al., 2003; Varshavsky, 1996). Moreover, removal of the N-terminal translation initiator methionine from a recombinant protein is often critical for its function. For instance, no removal of the N-terminal methionine of onconase, a cytotoxic ribonuclease with antitumor properties, expressed in *Escherichia coli*, resulted in little catalytic activity and reduced thermostability (Newton et al., 1998). Similarly, the proinflammatory cytokine RANTES was completely inactive with an additional methionine at the N-terminus (Proudfoot et al., 1996). In addition, changes in the immunological properties of the protein bearing an extra methionine have been reported for the recombinant bovine growth hormone (Erhard et al., 1994) and the bacterial staphylokinase (Schlott et al., 1994). Therefore, when these kinds of proteins are expressed in heterologous systems, it is crucial to make sure that the final amino acid sequences of recombinant and native proteins coincide. This will facilitate approval of the product for commercialisation by legal authorities. For instance, the validation process for products of recombinant DNA technology in Europe includes the partial amino acid sequence analysis to confirm the correct sequence of biopharmaceuticals (European Directorate for the Quality of Medicines, <http://www.pheur.org/site/page.628.php>).

Plastid transformation technology directs the transgene to the plastid genome of the plant cell instead to the nuclear genome. The main advantage of this approach is the high and reproducible yield: 5–20% of the recombinant protein in relation to the total soluble protein; this yield is 5–100 times higher than standard percentages obtained with nuclear transformation. Other advantages are lack of gene silencing, transgene containment due to the maternal inheritance of plastids in most of the cultivated plants and multigene engineering in a single step by using polycistrons (Daniell et al., 2002). The biotechnological applications of plastid transformation include agronomic traits (such as herbicide or insect resistance) and the expression of foreign proteins for different purposes. Tobacco chloroplasts are very attractive for expression of recombinant proteins of industrial or pharmaceutical use (Daniell et al., 2005; Maliga, 2003, 2004). For instance, the cellulytic enzyme xylanase (Leelavathi et al., 2003), and

therapeutic proteins such as the human somatotropin (Staub et al., 2000), the anthrax protective antigen (Watson et al., 2004) or the tetanus toxin (Tregoning et al., 2003) have been expressed in tobacco chloroplasts.

In order to obtain a recombinant protein with a non-methionine N-terminus, Staub et al. (2000) designed a ubiquitin–somatotropin fusion to yield somatotropin with a phenylalanine N-terminus in the chloroplast. Ubiquitin was only partially processed (efficiency 30–80%). The most abundant protein species, among the cleaved molecules, had an N-terminal proline residue, indicating the removal of one additional amino acid beyond phenylalanine. As a result, the overall efficiency was rather low. Up to now this represents the only attempt to properly express foreign proteins lacking N-terminal methionine in chloroplasts.

In the present work, we explored two alternative approaches for the production of recombinant proteins with a non-methionine N-terminus in tobacco chloroplasts. First, was the fusion of a plastid transit peptide (TP) to the protein encoded by the transgene. The TP directs a nuclear-encoded protein to the chloroplast. Upon import of the native protein, the TP is immediately cleaved by a general stromal processing peptidase (SPP) (Richter and Lamppa, 1998; VanderVere et al., 1995). SPP cleaves a broad range of precursors in a single endoproteolytic step and the pre-sequence is subsequently degraded by an ATP-dependent metallopeptidase (Richter and Lamppa, 1999). Thus, we fused the sequence of a chloroplast TP to the transgene and introduced the construct into the plastid genome. The selected transgene was human serum albumin (HSA). HSA is the intravenous protein mostly used worldwide for therapeutic purposes. Currently several companies commercialise recombinant HSA produced in yeast or mammal cells, though prices are 100–1000 times higher than HSA obtained from blood-donors. Alternative production systems, such as transgenic plants, have been explored for HSA expression (Arokiaraj et al., 2002; Farran et al., 2002). Growing plants in the field is fairly cheap and can be scaled up at low cost in relation to fermentation processes (10–50 times cheaper; Kusnadi et al., 1997).

HSA is normally synthesised by the hepatocytes in the liver and translated as a prepro-sequence, targeting the protein to the secretory pathway. The N-terminal amino acid of the mature HSA is aspartic acid. We previously expressed HSA in tobacco chloroplasts, with

an extra methionine at the N-terminal sequence of the mature protein (Fernández-San Millán et al., 2003). Up to now this has been the highest expression level obtained in plants (11% of the total protein). However, the amino-terminal region of the recombinant protein was different to that of the native protein, a difference that could potentially affect HSA functionality. It has been shown that some properties of HSA, such as association with nickel and copper, depend on the first three N-terminal amino acids (Peters, 1995). Latta et al. (1987) demonstrated that recombinant HSA, with an extra N-terminal methionine, when expressed in *E. coli*, was unable to bind nickel.

The second approach for proper expression of foreign proteins takes advantage of the endogenous N-terminal methionine cleavage mechanism of the chloroplast. The N-terminal sequences of 58 proteins of the nearly 100 encoded by the chloroplast genome have been determined to date. This represents an extensive set of data on N-terminal protein processing and reveals a systematic rule of cleavage (Giglione and Meinel, 2001). We expressed several foreign proteins in the chloroplast and analysed whether their N-terminal sequences were obtained accordingly. This study will allow for the prediction of N-methionine cleavage of recombinant proteins expressed in plastids. This could be used as a biotechnological tool for foreign protein production in chloroplasts with a N-terminal amino acid sequence identical to the native protein.

2. Materials and methods

2.1. Construction of the chloroplast expression vectors

The 171 bp of the *TP* of the Rubisco small subunit (RbcS) was amplified from tobacco genomic DNA by PCR with the following primers: TPforward (5'-GGAGGCAACCATGGCTTCCTCAGTTCTTTCC-3') and TPreverse (5'-GCATTGAACGCGTCCGCCGTT-3'). These primers included a *Nco*I restriction site at the 5' end and a *Mlu*I site, designed by silent mutagenesis, at the 3' end. The product was cloned in the pGEM-T vector (Promega) and the *Eco*RI fragment subcloned in the intermediate vector pSK+ (Stratagene). The sequence of the mature HSA was amplified

by PCR using the plasmid pTUB3HSA (Farran et al., 2002) as template and the primers DHSaforward (5'-CGGACGCGTTCAATGCGATGCACACAAGAGTGAGG-3') and HSAreverse (5'-GCGGCCGCGATGTTATAAGCCTAA-3'). The 1783 bp *Eco*RI fragment was subcloned in the intermediate vector pSK+. The HSA *Mlu*I–*Eco*RV fragment was fused to the 3' end of the *TP* sequence. The *TP*–HSA *Nco*I–*No*I fragment was fused to the 3' end of the 5' untranslated region (UTR) of the *psbA* tobacco gene (Fernández-San Millán et al., 2003). The complete fragment (promoter and 5'-UTR of *psbA*, *TP* and HSA) was obtained by *Eco*RI–*No*I digestion and inserted in the pLD final vector (kindly provided by Henry Daniell, University of Central Florida, USA) to result in the pLD-TP-Asp-HSA vector.

The same strategy was used to construct the vector pLD-TP-Ala-HSA with the HSA sequence lacking the first three nucleotides at the 5' end, but the primer AHSaforward (5'-CGGACGCGTTCAATGCGCACACAAGAGTGAGG-3') was used instead of DHSaforward.

2.2. Chloroplast transformation of tobacco

Leaves from *in vitro* grown tobacco (cv. Petit Havana) plants were bombarded using the Bio-Rad PDS-1000/He biolistic device as described previously (Daniell, 1997) to generate TP-Asp-HSA and TP-Ala-HSA transgenic plants. In addition, other plastid-transformed tobacco plants expressing the following proteins were analysed in this study: the mature HSA with an additional methionine at the N-terminus (Fernández-San Millán et al., 2003), the green fluorescent protein fused upstream of the canine parvovirus 2L21 epitope (GFP-2L21; Molina et al., 2004), the cholera toxin B subunit fused to 2L21 (CTB-2L21; Molina et al., 2004), the human cardiotrophin-1 (CT1; unpublished results), and the human interferon α 2b (IFN α 2b; Daniell et al., 2005).

2.3. Southern blot and HSA transcript quantification

Total DNA was extracted from leaves of transformed and control plants (DNeasy kit, Qiagen). Total DNA (15 μ g) was digested with *Bam*HI, electrophoresed on 0.7% agarose gels and transferred to

nylon membranes. The probe to check homoplasmy was a 0.81 kb *Bg*/II–*Bam*HI fragment and that for HSA detection was a 0.75 kb *Nco*I fragment (Fig. 2b). Hybridisation was performed using the chemiluminiscent AlkPhos direct labelling-detection system (Amersham).

Transcript quantification was performed by real time PCR. Three plants (30 days after transplanting) for each line and three replicates per plant were analysed. Untransformed plants grown in the same conditions were used as controls. Total RNA was extracted (RNeasy Plant Kit, Qiagen) and analysed by the SyBR Green protocol (Applied Biosystems) at the Scientific Park (Universidad Autónoma de Madrid). Primers for HSA detection were RTHSAforward (5'-TGCATTGCCGAAGTGGAA-3') and RTHSAreverse (5'-AAAACATGCCCAGGAAGACATC-3'). The plastidial 16S rRNA was used as constitutive control of plastid RNA loaded per reaction. Primers for 16S detection were 16Sforward (5'-CTTTTAAAGTCCGCGTCAAA-3') and 16Sreverse (5'-TCTTTCGATCTCTACGCATTTC-3').

2.4. Protein extraction, Western blot, ELISA and chlorophyll quantification

Mature leaves from tobacco plants grown in a walk-in room at 28 °C, 70% relative humidity, 16 h photoperiod and 250 $\mu\text{mol}/\text{m}^2/\text{s}$ photosynthetic photon flux were used for protein extraction. Transformed and untransformed leaf discs were ground in liquid nitrogen with a Microdismembrator (Braun, Germany). Pulverized material (100 mg) was resuspended in 700 μl of total protein extraction buffer (250 mM Tris–HCl, 4% SDS, 10% glycerol, 10% β -mercaptoethanol, pH 6.8) and boiled for 5 min. After centrifugation at 14,000 rpm, the supernatant was collected and subjected to western or ELISA analysis (Human Albumin ELISA Quantitation kit, Bethyl, Texas, USA). Protein quantification was performed by the BCA Protein Assay (Pierce) after two consecutive precipitations with acetone to avoid interferences of the extraction buffer with the assay buffer. Bovine serum albumin was used as standard.

Protein samples were separated by 10% SDS-PAGE and transferred to a nitrocellulose membrane for immunoblotting. The primary antibody (rabbit anti-HSA, Nordic Immunology, Tilburg, Holland) and the

secondary antibody (peroxidase-conjugated goat anti-rabbit, Southern Biotechnology Associates, Birmingham, USA) were used at 1:10,000 dilutions. Detection was performed using the ECL Western blotting system (Amersham).

Chlorophyll content was measured by the portable SPAD-502 (Minolta Camera Co., Japan). This is a non-destructive method based on the reflected light by the leaf that is directly proportional to the chlorophyll content (Yadava, 1986).

2.5. N-terminal sequencing of recombinant proteins expressed in tobacco chloroplasts

Since most of the HSA accumulated in the insoluble fraction (Fernández-San Millán et al., 2003), soluble proteins were eliminated by extraction with 1:7 (w/v) bicarbonate buffer (15 mM Na_2CO_3 , 35 mM NaHCO_3 , pH 9.6). Insoluble protein fractions of HSA transgenic plants were resuspended in total protein extraction buffer (see Section 2.4), separated by 10% SDS-PAGE, transferred to a PVDF membrane, stained with Coomassie Brilliant Blue without acetic acid (1 min, shaking) and destained in 50% methanol. After the membrane was dried, the band corresponding to HSA was excised and subjected to N-terminal microsequencing by Edman reaction (Sequencing Service, Universidad de Barcelona, Spain). Samples analysed by MALDI-TOF or ESI mass spectrometry (MS) were extracted in the same way, separated by SDS-PAGE and stained with Coomassie Brilliant Blue. The corresponding bands were directly excised from the polyacrylamide gel and analysed. MS was performed in the Proteomic Service (Centro Nacional de Biotecnología, Madrid).

Soluble proteins from GFP-2L21 transgenic plants were extracted with bicarbonate buffer incubated at 70 °C for 15 min, centrifuged at 14,000 rpm for 10 min and the supernatant's pH lowered to 5.5 with potassium acetate. After neutralisation with NaOH, samples were centrifuged. Supernatants were processed as described above for microsequencing by Edman reaction or MALDI-TOF MS.

Soluble proteins from CTB-2L21 transgenic plants were extracted with bicarbonate buffer, incubated at 60 °C for 15 min, centrifuged at 14,000 rpm for 10 min and the supernatant subjected to ammonium sulphate precipitation (40–80%). Precipitated proteins were

resuspended in the extraction buffer and processed as described above for MALDI-TOF MS analysis.

Samples from CT1 and IFN α 2b transgenic plants were resuspended in protein extraction buffer (20 mM sodium phosphate pH 7.4, 0.5 M NaCl, 0.1% Triton X-100), incubated 45 min and centrifuged at 14,000 rpm for 10 min. All the process was performed at 4 °C. Supernatants were filtered (0.45 μ m pore diameter) and purified by affinity chromatography (HisTrap HP columns, Amersham), taking advantage of the N-terminal His-tag. Purified proteins were processed as described above for MALDI-TOF MS analysis.

3. Results and discussion

3.1. Selection of the TP for fusion with HSA

Since more than 3000 nuclear genes encode proteins imported by the chloroplast (Martin et al., 2002), there is a high number of potential TPs processed in the stroma by the SPP that could be used in this study. Around 600 of these TPs have been sequenced and analysed; they show low homology at both the amino acid and the secondary structure levels (Theg and Scott, 1993). Considering that the first N-terminal amino acids of the mature protein seem to be important for a proper cleavage (Berry-Lowe and Schmidt, 1991), we looked in the SWISS-PROT database for a TP that, after processing in the stroma, rendered a mature protein with aspartic acid as N-terminal amino acid, coincident with the first amino acid of the mature HSA. None fulfilled this condition. Then, we valued that the cleavage site of the TP by the SPP were well characterised. We screened several candidates and

selected the TP of the small subunit of the Rubisco (RbcS). This sequence has been studied in great detail (Bruce, 2001) and sequenced in more than 20 species. The TP of RbcS is 44–59 amino acids long (57 in the case of tobacco). Functional studies of RbcS-TP have utilized several approaches: point mutations or partial deletions followed by *in vitro* translation and chloroplast import, and construction of chimeric genes (TP fused with different proteins) for processing analysis and final location of the protein of interest (Berry-Lowe and Schmidt, 1991; Dieryck et al., 1997; Knight et al., 1993; Nawrath et al., 1994). These studies demonstrated that the TP of RbcS could direct the import of proteins of diverse origin to the chloroplast and be processed by SPP. In an experiment closer to our objective, the *RbcS* gene, including the sequence of the TP, was relocated to the tobacco plastid genome (Whitney and Andrews, 2001a). It was observed that the TP was cleaved completely. We then fused the coding sequence of the RbcS-TP to the cDNA of the mature HSA (TP-Asp-HSA).

To predict the result of this strategy we used the ChloroP program (<http://www.cbs.dtu.dk/services/ChloroP>, Emanuelsson et al., 1999), to analyse the cleavage prediction of the TP in the chimeric protein TP-Asp-HSA. ChloroP uses a scoring matrix based on TP-containing proteins extracted from SWISS-PROT and reliably predicted cleavage sites for soluble proteins targeted to the stroma (Gomez et al., 2003), such as RbcS. The cleavage site score for the last amino acid of the TP (cysteine) in TP-Asp-HSA was lower than that for RbcS, indicating a lower probability of cleavage (Fig. 1). ChloroP predicted amino acid 20 of TP (alanine) to be the most probable cleavage site in TP-Asp-HSA. For this reason we analysed an alternative

		Mature protein	Cleavage site score after Cys 57	Cleavage probability after Cys 57
	<i>Rubisco small subunit transit peptide</i>			
RbcS	MASSVLSSAAVATRSNVAQANMVAPFTGLKSAASFPVSRKQNLDTISIASNGGRVQC	↓ MQVWP...	10.452	High
TP-Asp-HSA	MASSVLSSAAVATRSNVAQANMVAPFTGLKSAASFPVSRKQNLDTISIASNGGRVQC	↓ DAHKSE...	6.779	Medium
TP-Ala-HSA	MASSVLSSAAVATRSNVAQANMVAPFTGLKSAASFPVSRKQNLDTISIASNGGRVQC	↓ AHKSE...	10.900	High

Fig. 1. Cleavage prediction of the transit peptide of the Rubisco small subunit fused to HSA with an N-terminal aspartic acid (TP-Asp-HSA) or alanine (TP-Ala-HSA), compared to the native protein (RbcS). Arrows indicate the most probable cleavage site for each construct. Analysis was performed with ChloroP program.

chimeric protein, eliminating the first aspartic acid of the mature HSA, leaving alanine as N-terminal amino acid (TP-Ala-HSA). The probability of proper cleavage of TP-Ala-HSA predicted by ChloroP was even higher than that of the native protein (Fig. 1). However, elimination of the first amino acid of mature HSA would probably alter the interaction of HSA with copper and nickel, which depends on the three N-terminal amino acids (Peters, 1995).

3.2. Vector construction and chloroplast transformation

Two transformation vectors were designed. The first included the sequence coding for the TP of RbcS fused to the cDNA of the mature HSA. The second was equivalent to the previous vector but lacked the first three nucleotides at the 5' end of the cDNA of the mature HSA. Sequences were amplified by PCR, cloned and sequenced. These chimeric genes were inserted immediately downstream of the promoter and 5'-UTR of the *psbA* gene. In a previous study this promoter maximized the expression of HSA in tobacco chloroplasts (Fernández-San Millán et al., 2003). Finally, the products were introduced in the tobacco plastid transformation vector pLD resulting in the vectors pLD-TP-Asp-HSA and pLD-TP-Ala-HSA (Fig. 2). The functionality of both vectors was determined in *E. coli* by Western blot. A band with a size slightly larger than that of the commercial HSA was detected in both clones (data not shown), indicating that the TP was not processed, as expected, in *E. coli*. Whitney and Andrews (2001a) also observed that *E. coli* was unable to process the TP of RbcS.

Chloroplast transformation of tobacco was performed by leaf particle bombardment as previously described (Daniell, 1997), using spectinomycin as selective agent. Developed shoots were initially tested for specific plastome integration by PCR (data not shown). After a second round of selection, regenerated plants were analysed by Southern blot. The 0.81 kb probe, homologous to the flanking regions *trnI* and *trnA* was used to check homoplasmy. DNA from non-transformed plants produced a 7.45 kb fragment while transformed plants rendered a 10.88 kb fragment (Fig. 2a). Out of 12 transformants analysed, 11 were homoplasmic and one was an escape (lane 7, Fig. 2c). To confirm that the 10.88 kb fragment con-

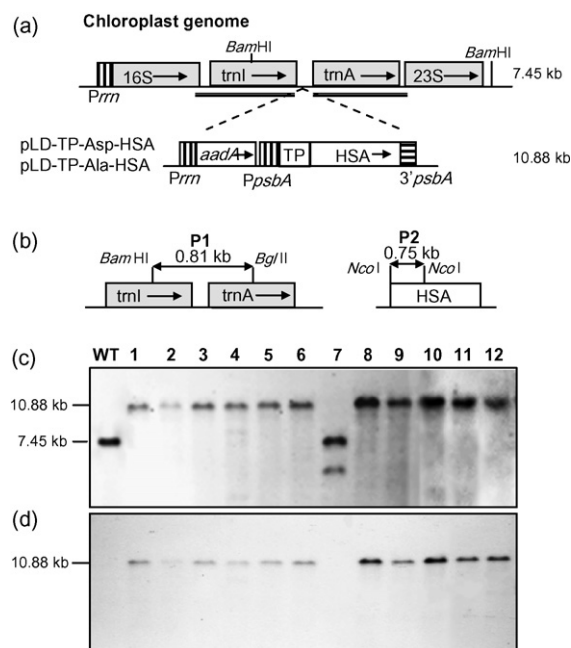


Fig. 2. Vector constructs and integration of transgenes into the chloroplast genome. (a) Regions for homologous recombination are underlined in the native chloroplast genome. The *aadA* gene was driven by the *rrn* promoter and the transgene was driven by the *psbA* promoter and terminator. The pLD-TP-Asp-HSA vector includes the DNA sequence corresponding to the mature HSA and the pLD-TP-Ala-HSA lacks the first codon (encoding aspartic acid), leaving alanine as the N-terminal amino acid. Arrows within boxes indicate the direction of transcription. Numbers to the right indicate the predicted hybridising fragments when total DNA digested with *Bam*HI was probed with P1 or P2. (b) The 0.81 kb fragment (P1) of the targeting region for homologous recombination and the 0.75 kb *HSA* fragment (P2) were used as probes for Southern blot analysis. (c and d) Southern blot analysis of independent pLD-TP-Asp-HSA (1–6) and pLD-TP-Ala-HSA (7–12) lines. Blots were probed with P1 (c) and P2 (d). WT, wild-type Petit Havana plant.

tained the *HSA* sequence, the same blot was hybridised with a 0.75 kb probe homologous to *HSA*. As expected, hybridisation was observed in the transgenic lines but neither in the wild-type plant nor in the escape line (Fig. 2d).

3.3. Expression of HSA in chloroplast transgenic plants

Protein extracts from both transgenic lines (TP-Asp-HSA and TP-Ala-HSA) were analysed by Western

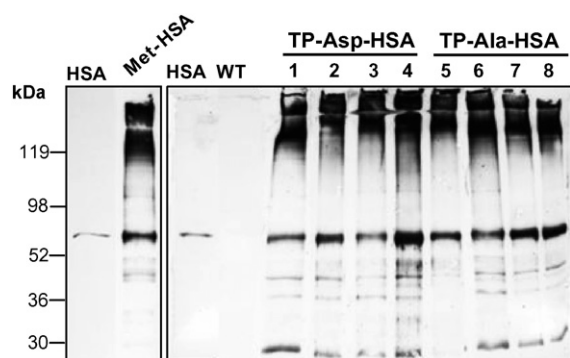


Fig. 3. Western blot analysis of HSA expression in transgenic chloroplasts. Four independent lines (1–4) of TP-Asp-HSA and another four of TP-Ala-HSA (5–8) are shown. Blots were detected using a polyclonal anti-HSA as primary antibody. Five micrograms of total protein, extracted from plants 30 days after transplanting, were loaded per well. Met-HSA, transformed plant which expressed the mature HSA with an additional N-terminal methionine; HSA, 40 ng of commercial HSA; WT, wild-type Petit Havana plant.

blot (Fig. 3). A sample from pLDapsbA-HSA transformed plants, which expressed the mature HSA with an additional methionine in the N-terminus (Met-HSA) (Fernández-San Millán et al., 2003), was included as control. The signal detected in the Western blot was specific for HSA because no band was present in the wild-type plant (Fig. 3). The size of the monomeric bands was similar in all three cases, with or without TP, regardless of the presence of aspartic acid or alanine after the transit peptide. No immunoreactive bands of 72.2 kDa, corresponding to the HSA fused to the TP, were detected. This result indicates that the TP seemed to be cleaved in both constructs. The HSA expressed in the transgenic lines had a similar electrophoretic mobility to the commercial HSA. Low molecular mass signals detected probably correspond to product degradation. As previously described (Fernández-San Millán et al., 2003), most of the HSA was aggregated due to the formation of inclusion bodies in the chloroplast.

The accumulation of HSA was measured over the life cycle of the plant to investigate whether the TP could interfere with HSA production. Recombinant protein levels increased with age reaching the maximum value 60 days after transplanting (Fig. 4a), when first fruits appeared and secondary shoots developed. This is in accordance with previous observations made by Molina et al. (2004). Transgenic plants with the HSA

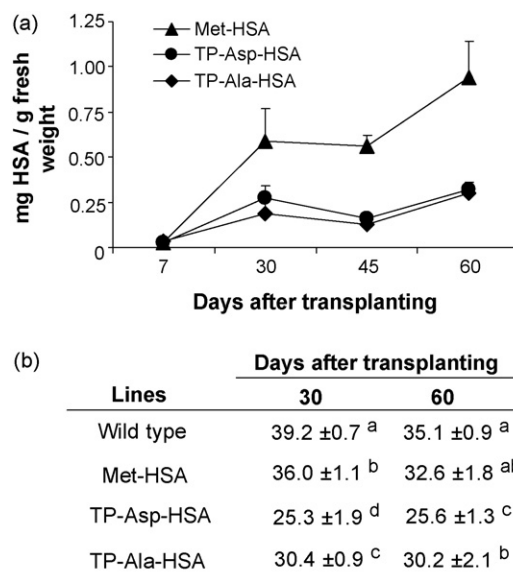


Fig. 4. (a) Recombinant protein accumulation in plants grown in the greenhouse at different developmental stages. HSA was measured by enzyme-linked immunosorbent assay. (b) Relative chlorophyll content measured 30 or 60 days after transplanting. Data are presented as the means ± S.D. of measurements on three mature leaves from five individual plants per construct ($n=15$). Analysis of variance and the Duncan test ($P \leq 0.05$) were used to test the significance of differences between means. Different letters within each column indicate statistically different values. Met-HSA, transformed plant which expressed the mature HSA with an additional N-terminal methionine.

fused to the RbcS-TP accumulated a lower amount of HSA than plants expressing Met-HSA (Fig. 4a). These differences increased with plant age. In addition, TP-Asp-HSA and TP-Ala-HSA transgenic lines displayed a slightly pale green colour in comparison with wild-type plants. Chlorophyll measurements, performed at two developmental stages, were significantly lower in the transgenic lines, especially in the construct TP-Asp-HSA (Fig. 4b). The decrease in chlorophyll may not be attributable to HSA accumulation, since the Met-HSA line does not show a greater decrease. Considering that the only difference between these three lines is the presence or absence of the TP fused to HSA, the problem should be related to the RbcS-TP. It may be hypothesised that upon expression of the fusion TP-HSA in the stroma, the SPP cleaving activity would be insufficient to cope with the excess of substrate, the recombinant protein and the diver-

sity of precursor proteins targeted to the chloroplasts. Richter and Lamppa (1998) demonstrated that an array of proteins essential for photosynthetic reactions was cleaved by SPP, such as RbcS, plastocyanin, Rbc activase and an oxygen evolving enhancer. Pleiotropic effects such as chlorophyll content decrease may then follow.

We did not expect differences at the translational level between the three lines, since each of them was under control of the 5'-UTR of the *psbA* gene. To explain the observed differences in HSA accumulation, we measured the stability of the protein in the stroma of the chloroplast. It is well known that the N-terminal sequence has a major influence in the stability of the protein. Giglione et al. (2003) demonstrated that N-terminal methionine excision was critical for the chloroplast D2 protein life-span. The three constructs expressed in tobacco chloroplasts will probably differ in the N-terminal amino acid, which are methionine, alanine or aspartic acid. However, pulse-chase labelling experiments showed that the half-life of the three proteins was similar, between 2 and 4 h (data not shown).

To check the transcriptional activity and mRNA stability, real time PCR was performed in all three lines. Even though the same promoter and terminator were used, Met-HSA transgenic line produced significantly higher amounts of HSA transcripts than the two lines with HSA fused to the *RbcS-TP* sequence (Fig. 5). This may indicate that the presence of the *RbcS-TP* presequence destabilises the HSA mRNA and may be the cause of lower levels of recombinant protein.

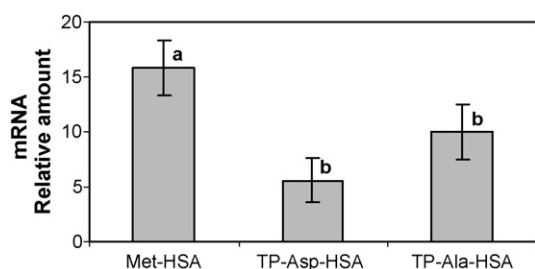


Fig. 5. Relative amount of HSA transcripts in the three different transgenic lines measured by quantitative real time PCR. Data are presented as the means $\pm$ S.D. of measurements from nine samples per construct. Different letters indicate statistically different values at $P \leq 0.05$ using the Duncan test.

3.4. The TP of TP-Asp-HSA is properly cleaved in the stroma of the chloroplast releasing the mature HSA

HSA produced in tobacco chloroplasts was mainly present as inclusion bodies (Fernández-San Millán et al., 2003), a property that we exploited for purification, since HSA is present primarily as aggregates in the insoluble fraction (Fig. 3). Protein insoluble fractions were extracted from leaves and separated by SDS-PAGE. After Coomassie blue staining, a band of 66.5 kDa was detected in samples from transgenic plants but was absent in untransformed plants (data not shown). This band was excised from the gel and analysed. Samples from TP-Asp-HSA were digested with trypsin and subjected to MALDI-TOF MS, which revealed several peptides that matched with HSA sequence. A particular peptide of 1149.5 m/z corresponding to the 10 N-terminal amino acids of the mature HSA was detected (Fig. 6a). Peptides corresponding to HSA molecules without cleavage of the

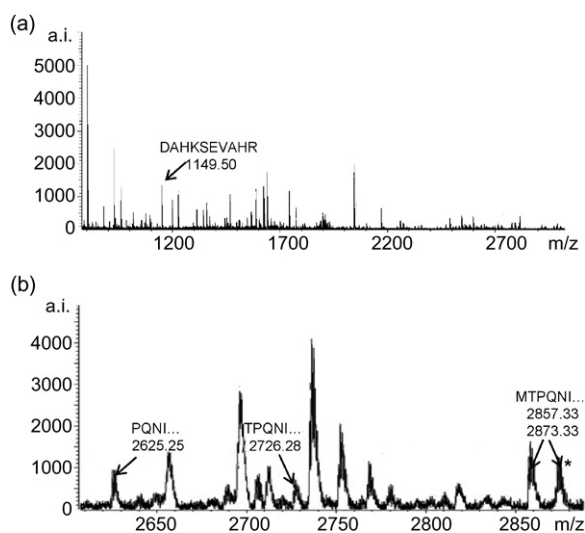


Fig. 6. Analysis of the N-terminal sequence of recombinant proteins expressed in chloroplasts by MALDI-TOF mass spectrometry. (a) Chromatogram of a sample from TP-Asp-HSA transgenic plant after tryptic digestion. The peptide corresponding to the N-terminal region of the mature HSA is marked. (b) Chromatogram of a sample from CTB-2L21 transgenic plant. The beginning of each peptide, corresponding to the N-terminal region of CTB-2L21, is shown. The asterisk represents the peptide with the N-terminal methionine sulfoxide.

TP, or even HSA with inaccurate TP cleavage were not found. We have also verified that solubilisation of TP-Asp-HSA aggregates with guanidine rendered a unique molecule of 66.5 kDa in a Western blot (data not shown). Therefore, MS and Western blot results indicate the presence of a completely processed HSA species in transgenic chloroplasts, despite the negative prediction of the ChloroP program.

Similarly, we analysed the HSA from TP-Ala-HSA samples. Despite that variety of endopeptidases used for digest (e.g. Trypsin, EndoV8, Endo GlucC and Arg-C), we were unable to detect the N-terminal region of HSA by MS. Variation in one amino acid residue could interfere with peptide ionization and further detection. Due to the variability in peptide ionization, some not detected by MALDI-TOF MS are more easily identified by ESI MS. However, this method also failed to achieve positive results. Edman degradation was also unsuccessful. The N-terminus was blocked, indicating a post-translational modification of the first amino acid. If we assume a correct cleavage of the TP, alanine would be the N-terminal amino acid. This alanine should be a good candidate for post-translational modifications, and it would explain blockade of the Edman reaction. There are examples of modifications of the alanine as N-terminal amino acid in a variety of plastid proteins. For example, the alanine of the PSII reaction centre J protein was acetylated (Gomez et al., 2002) and those of the ribosomal L2 and S14 proteins were methylated (Yamaguchi and Subramanian, 2000; Yamaguchi et al., 2000). Other authors encountered similar problems when N-terminal microsequencing of some foreign Rubisco subunits expressed in tobacco chloroplasts returned no sequence, probably due to post-translational modifications (Whitney and Andrews, 2001b; Whitney et al., 2001). Despite this problem, we presume that TP-Ala-HSA was correctly processed, rendering Ala-HSA in the stroma of the chloroplast. We based this conclusion on the following results: the ChloroP cleavage prediction for TP-Ala-HSA was higher than that for the native RbcS (Fig. 1) and the size of the protein detected by immunoblotting was similar to the mature HSA and Asp-HSA (Fig. 3).

In conclusion, we have developed a strategy to produce HSA with identical sequence to that found in humans. The fusion of the RbcS-TP to the mature HSA was properly cleaved by SPP in the stroma of

the chloroplast, producing HSA lacking methionine as N-terminal amino acid. Further studies with different proteins fused to this TP should validate this approach.

3.5. Analysis of N-terminal maturation of recombinant proteins expressed in plastids reveals compliance with the rules of native plastid proteins

Currently, the N-termini of 58 of the approximately 100 proteins encoded by the plastid genome have been determined. From this extensive data set, it has been concluded that N-terminal processing is systematic and that more than 60% of the plastid proteins lose the N-formylmethionine or even several additional residues (Giglione and Meinel, 2001). Plastidial MAP is responsible for methionine cleavage that depends on the length and radius of gyration of the side chain of the second amino acid (Sherman et al., 1985). If the second residue is Ala, Gly, Ser, Pro or Thr, the first methionine is systematically eliminated. The N-terminal methionine is retained when the side chain is large (Leu, Ile, Lys, Asp, Glu, Arg, Asn, Val). This rule allowed prediction of cleavage sites for plastid proteins with undetermined N-terminus (<http://www.isv.cnrs-gif.fr/tm/maturation/images/chloro.html>; Giglione and Meinel, 2001). In order to determine whether foreign proteins comply with the rules proposed for native plastid proteins, several recombinant proteins expressed in tobacco chloroplasts were purified and the N-terminal sequence analysed.

Table 1 shows our results as well as previous data from other authors regarding foreign proteins expressed in tobacco chloroplasts by plastid transformation. When the second residue was aspartic acid (Met-HSA, IFN α 2b), lysine (tetanus toxin) or arginine (RbcS from algae), methionine was retained. The nature of the third amino acid was irrelevant in the process (alanine in Met-HSA or histidine in IFN α 2b). N-terminal methionine was removed when the second amino acid was serine (GFP-2L21, bacterial Rbc), glycine (CT1), or alanine (somatotropin). These results are in accordance with the rules established for native proteins.

The case of CTB-2L21 protein was unexpected because methionine was mainly retained (despite threonine being the second residue), but molecules without one or two N-terminal amino acids were also detected (Table 1 and Fig. 6b). The presence of proline in the

Table 1

Amino-terminal sequences of recombinant proteins expressed in chloroplasts by plastid transformation

Protein	N-terminus as deduced from nucleotide sequence	N-terminus of the mature protein	Analysis method	Reference
Met-HSA	MDAHK	MDAHK	Edman	This study
IFN α 2b-His tag	MDHHH	MDHHH	MALDI-TOF	This study
Tetanus toxin	MKNLD	MKNLD	Edman and ESI-MS	Tregoning et al. (2003)
RbcS Gs	MRITQ	MRITQ	Edman	Whitney et al. (2001)
RbcS Pt	MRLTQ	MRLTQ	Edman	Whitney et al. (2001)
GFP-2L21	MSKGE	SKGE	Edman and MALDI-TOF	This study
CT1-His tag	MGHHH	GHHH	MALDI-TOF	This study
Bacterial Rbc	MSPQT	PQT	Edman	Whitney and Andrews (2001b)
Somatotropin	MAFPA	AFPA (>95%)	Edman and ESI-MS	Staub et al. (2000)
CTB-2L21	MTPQN	MTPQN (>50%) TPQN (<25%) PQN (<25%)	MALDI-TOF	This study
TP-Asp-HSA	tp-DAHKS	DAHKS	MALDI-TOF	This study
TP-Ala-HSA	tp-AHKSE	Blocked	Edman and MALDI-TOF	This study

HSA: human serum albumin; IFN: human interferon; RbcS: Rubisco small subunit; Gs: *Galdieria sulphuraria*; Pt: *Phaeodactylum tricornutum*; GFP-2L21: green fluorescent protein fused upstream of the canine parvovirus epitope 2L21; CT1: human cardiophorin-1; CTB-2L21: cholera toxin B subunit fused upstream of 2L21; TP: Rubisco small subunit transit peptide.

third position could influence N-terminal excision. It has been reported that a proline at position 3 partially inhibited MAP action in *E. coli* and, in the case of methionine processing, a significant protein fraction also lost the second residue (Ben-Bassat et al., 1987; Hirel et al., 1989). Furthermore, removal of the two N-terminal residues, when the third is proline, was observed in chloroplasts for the native Rbc large subunit (Houtz et al., 1989) and for Rbc from bacterial origin expressed in tobacco chloroplasts (Whitney and Andrews, 2001b).

In the case of TP-Asp-HSA, TP was cleaved by SPP and no further processing was detected.

In view of these results, we conclude that the N-terminal processing rule of native plastidial proteins maybe valid for foreign proteins expressed in transgenic chloroplasts. This knowledge should be taken into account when foreign proteins, whose N-terminus is important for its function, are produced by plastid transformation. Moreover, this behaviour could serve as a tool to generate foreign proteins without N-terminal methionine if the second amino acid allows the processing by MAP. Finally, in those cases where the second residue does not allow to get rid of methionine, like HSA, this study demonstrates that transit peptides (e.g. RbcS-TP) fused to the protein are a promising alternative.

Acknowledgments

We are grateful to Henry Daniell for the pLD vector and IFN α 2b seeds and to Alberto Paradela for his technical assistance with mass spectrometry. This work was supported by the Ministerio de Educación y Ciencia (Grant BIO2002-02851) and the Departamentos de Industria y Educación (Gobierno de Navarra).

References

- Arokiajaj, P., Ruker, F., Obermayr, E., Shamsul Bahri, A.R., Haf-sah, J., Carter, D.C., Yeang, H.Y., 2002. Expression of human serum albumin in transgenic *Hevea Brasiliensis*. Malaysian Rubber Borrad 5, 157–166.
- Ben-Bassat, A., Bauer, K., Chang, S.Y., Myambo, K., Boosman, A., Chang, S., 1987. Processing of the initiation methionine from proteins: properties of the *Escherichia coli* methionine aminopeptidase and its gene structure. J. Bacteriol. 169, 751–757.
- Berry-Lowe, S., Schmidt, B.J., 1991. Chloroplast protein transport. In: Bogorad, L., Vasil, I.K. (Eds.), The Molecular Biology of Plastids and Mitochondria. Academic Press, Orlando, pp. 257–301.
- Bruce, B.D., 2001. The paradox of plastid transit peptides: conservation of function despite divergence in primary structure. Biochim. Biophys. Acta 1541, 2–21.
- Daniell, H., 1997. Transformation and foreign gene expression in plants by microprojectile bombardment. Meth. Mol. Biol. 62, 463–489.

- Daniell, H., Chebolu, S., Kumar, S., Singleton, M., Falconer, R., 2005. Chloroplast-derived vaccine antigens and other therapeutic proteins. *Vaccine* 23, 1779–1783.
- Daniell, H., Khan, M.S., Allison, L., 2002. Milestones in chloroplast genetic engineering: an environmentally friendly era in biotechnology. *Trends Plant Sci.* 7, 84–91.
- Dieryck, W., Pagnier, J., Poyart, C., Marden, M.C., Gruber, V., Bournat, P., Baudino, S., Merot, B., 1997. Human haemoglobin from transgenic tobacco. *Nature* 386, 29–30.
- Emanuelsson, O., Nielsen, H., von Heijne, G., 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* 8, 978–984.
- Erhard, M.H., Kellner, J., Schmidhuber, S., Schams, D., Losch, U., 1994. Identification of antigenic differences of recombinant and pituitary bovine growth hormone using monoclonal antibodies. *J. Immunoassay* 15, 1–19.
- Farran, I., Sanchez-Serrano, J.J., Medina, J.F., Prieto, J., Mingo-Castel, A.M., 2002. Targeted expression of human serum albumin to potato tubers. *Transgenic Res.* 11, 337–346.
- Fernández-San Millán, A., Mingo-Castel, A.M., Miller, M., Daniell, H., 2003. A chloroplast transgenic approach to hyper-express and purify human serum albumin, a protein susceptible to proteolytic degradation. *Plant Biotech. J.* 1, 71–79.
- Gigliione, C., Meinel, T., 2001. Organellar peptide deformylases: universality of the N-terminal methionine cleavage mechanism. *Trends Plant Sci.* 6, 566–572.
- Gigliione, C., Vallon, O., Meinel, T., 2003. Control of protein lifespan by N-terminal methionine excision. *Embo J.* 22, 13–23.
- Gomez, S.M., Bil, K.Y., Aguilera, R., Nishio, J.N., Faull, K.F., Whitelegge, J.P., 2003. Transit peptide cleavage sites of integral thylakoid membrane proteins. *Mol. Cell Proteomics* 2, 1068–1085.
- Gomez, S.M., Nishio, J.N., Faull, K.F., Whitelegge, J.P., 2002. The chloroplast grana proteome defined by intact mass measurements from liquid chromatography mass spectrometry. *Mol. Cell Proteomics* 1, 46–59.
- Hirel, P.H., Schmitter, M.J., Dessen, P., Fayat, G., Blanquet, S., 1989. Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Natl. Acad. Sci. U.S.A.* 86, 8247–8251.
- Houtz, R.L., Stults, J.T., Mulligan, R.M., Tolbert, N.E., 1989. Post-translational modifications in the large subunit of ribulose biphosphate carboxylase/oxygenase. *Proc. Natl. Acad. Sci. U.S.A.* 86, 1855–1859.
- Knight, J.S., Madueno, F., Gray, J.C., 1993. Import and sorting of proteins by chloroplasts. *Biochem. Soc. Trans.* 21, 31–36.
- Kusnadi, A., Nikolov, Z., Howard, J., 1997. Production of recombinant proteins in transgenic plants: practical considerations. *Biotechnol. Bioeng.* 56, 473–484.
- Latta, M., Knap, M., Sarmientos, P., Brefort, G., Becquart, J., Guerrier, L., Jung, G., Mayaux, J.F., 1987. Synthesis and purification of mature human serum albumin from *Escherichia coli*. *BioTechnology* 5, 1309–1314.
- Leelavathi, S., Gupta, N., Maiti, S., Ghosh, A., Reddy, V.S., 2003. Overproduction of an alkali and thermostable xylanase in tobacco chloroplasts and efficient recovery of the enzyme. *Mol. Breeding* 11, 59–67.
- Maliga, P., 2003. Progress towards commercialization of plastid transformation technology. *Trends Biotechnol.* 21, 20–28.
- Maliga, P., 2004. Plastid transformation in higher plants. *Annu. Rev. Plant Biol.* 55, 289–313.
- Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., Penny, D., 2002. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. U.S.A.* 99, 12246–12251.
- Molina, A., Hervas-Stubbs, S., Daniell, H., Mingo-Castel, A.M., Veramendi, J., 2004. High-yield expression of a viral peptide animal vaccine in transgenic tobacco chloroplasts. *Plant Biotech. J.* 2, 141–153.
- Nawrath, C., Poirier, Y., Somerville, C., 1994. Targeting of the polyhydroxybutyrate biosynthetic pathway to the plastids of *A. thaliana* results in high levels of polymer accumulation. *Proc. Natl. Acad. Sci. U.S.A.* 91, 12760–12764.
- Newton, D.L., Boque, L., Wlodawer, A., Huang, C.Y., Rybak, S.M., 1998. Single amino acid substitutions at the N-terminus of a recombinant cytotoxic ribonuclease markedly influence biochemical and biological properties. *Biochemistry* 37, 5173–5183.
- Peters, T.J., 1995. All About albumin. Academic Press, San Diego.
- Proudfoot, A.E., Power, C.A., Hoogewerf, A.J., Montjovent, M.O., Borlat, F., Offord, R.E., Wells, T.N., 1996. Extension of recombinant human RANTES by the retention of the initiating methionine produces a potent antagonist. *J. Biol. Chem.* 271, 2599–2603.
- Richter, S., Lamppa, G.K., 1998. A chloroplast processing enzyme functions as the general stromal processing peptidase. *Proc. Natl. Acad. Sci. U.S.A.* 95, 7463–7468.
- Richter, S., Lamppa, G.K., 1999. Stromal processing peptidase binds transit peptides and initiates their ATP-dependent turnover in chloroplasts. *J. Cell Biol.* 147, 33–44.
- Schlott, B., Hartmann, M., Guhrs, K.H., Birch-Hirschfeld, E., Pohl, H.D., Vanderschueren, S., Van de Werf, F., Michael, A., Collen, D., Behnke, D., 1994. High yield production and purification of recombinant staphylokinase for thrombolytic therapy. *Biotechnology (NY)* 12, 185–189.
- Sherman, F., Stewart, J.W., Tsunasawa, S., 1985. Methionine or not methionine at the beginning of a protein. *Bioessays* 3, 27–31.
- Staub, J.M., García, B., Graves, J., Hajdukiewicz, P.T.J., Hunter, P., Nehra, N.S., Paradkar, V., Schlittler, M., Carroll, J.A., Spatola, L., Ward, D., Ye, G., Russell, D.A., 2000. High-level production of a human therapeutic protein in tobacco chloroplasts. *Nat. Biotechnol.* 18, 333–338.
- Theg, S.M., Scott, S.V., 1993. Protein import into chloroplasts. *Trends Cell Biol.* 3, 186–190.
- Tregoning, J.S., Nixon, P., Kuroda, H., Svab, Z., Clare, S., Bowie, F., Fairweather, N., Ytterberg, J., van Wijk, K.J., Dougan, G., Maliga, P., 2003. Expression of tetanus toxin fragment C in tobacco chloroplasts. *Nucl. Acids Res.* 31, 1174–1179.
- VanderVere, P.S., Bennett, T.M., Oblong, J.E., Lamppa, G.K., 1995. A chloroplast processing enzyme involved in precursor maturation shares a zinc-binding motif with a recently recognized family of metalloendopeptidases. *Proc. Natl. Acad. Sci. U.S.A.* 92, 7177–7181.

- Varshavsky, A., 1996. The N-end rule: functions, mysteries, uses. *Proc. Natl. Acad. Sci. U.S.A.* 93, 12142–12149.
- Watson, J., Koya, V., Leppla, S.H., Daniell, H., 2004. Expression of *Bacillus anthracis* protective antigen in transgenic chloroplasts of tobacco, a non-food/feed crop. *Vaccine* 22, 4374–4384.
- Whitney, S.M., Andrews, T.J., 2001a. The gene for the ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) small subunit relocated to the plastid genome of tobacco directs the synthesis of small subunits that assemble into Rubisco. *Plant Cell* 13, 193–205.
- Whitney, S.M., Andrews, T.J., 2001b. Plastome-encoded bacterial ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) supports photosynthesis and growth in tobacco. *Proc. Natl. Acad. Sci. U.S.A.* 98, 14738–14743.
- Whitney, S.M., Baldet, P., Hudson, G.S., Andrews, T.J., 2001. Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. *Plant J.* 26, 535–547.
- Yadava, U.L., 1986. A rapid and nondestructive method to determine chlorophyll content in intact leaves. *Hortscience* 21, 1449–1450.
- Yamaguchi, K., Subramanian, A.R., 2000. The plastid ribosomal proteins. Identification of all the proteins in the 50 S subunit of an organelle ribosome (chloroplast). *J. Biol. Chem.* 275, 28466–28482.
- Yamaguchi, K., von Knoblauch, K., Subramanian, A.R., 2000. The plastid ribosomal proteins. Identification of all the proteins in the 30 S subunit of an organelle ribosome (chloroplast). *J. Biol. Chem.* 275, 28455–28465.