Running head: Plastid proteostasis in transplastomic plants

Author for correspondence:
Michele Bellucci, Istituto di Genetica Vegetale, Consiglio Nazionale delle Ricerche (CNR), via della Madonna Alta 130, 06128 Perugia, Italy.
Tel.: +39 075 5014864; Fax: +39 075 5014869
e-mail: michele.bellucci@igv.cnr.it

Journal research area: Updates
Plastid proteostasis and heterologous protein accumulation in transplastomic plants

Francesca De Marchis¹, Andrea Pompa¹, Michele Bellucci¹*

Istituto di Genetica Vegetale, Consiglio Nazionale delle Ricerche (CNR), via della Madonna Alta 130, 06128 Perugia, Italy
This work was supported by the project PRIN 20073YHRLE of the Ministero dell’Istruzione, dell’Università e della Ricerca, Italy, and by the COST Action FA0804, Molecular Farming: Plants as a Production Platform for High Value Proteins.

1 All authors contributed equally to this article.

*Corresponding author; e-mail michele.bellucci@igv.cnr.it

Introduction

The insertion of transgenes into the plastid genome (plastome) has proved to be an effective alternative to nuclear transformation for producing stable transformants expressing heterologous proteins. Recently, methodology, characteristics and possible fields of application of plastome transformation have been extensively reviewed (Lössl and Waheed, 2011; Scotti et al., 2012). From these reports, plastid transformation turns out to be a very promising tool for the production of recombinant proteins in plants, yet some limitations must still be overcome. Although the plastome of many species has been successfully transformed, an efficient and reproducible transformation procedure easily adoptable by any plant laboratory has been established only in tobacco, and to a minor extent in two other solanaceous species (potato and tomato), as well as in few other species (Maliga and Bock, 2011). Moreover, up to now the steady-state level of a recombinant protein coded by a gene inserted into the plastome has failed to be predictable. In spite of a huge number of reports on successful foreign protein production in plastids, above all proteins with a pharmaceutical interest, in many cases the expression level of other recombinant proteins in the transplastomic plants appears to be very low or even undetectable (e.g. Birch-Machin et al., 2004; Bellucci et al., 2005; Wirth et al., 2006). We believe that this is mainly due to our limited knowledge of the mechanisms in plastids influencing the maintenance of protein homeostasis, or proteostasis. This term refers to a complex network of biological pathways capable of maintaining the dynamic equilibrium of the protein pool, thus allowing the cell to get used to environmental changes (Balch et al., 2008). This concept can also be extended to subcellular level, especially for those organelles with an endosymbiotic origin like the plastids and mitochondria (Waller, 2012). These organelles retain a whole series of mechanisms for the preservation of their protein balance, including specific proteases, transcriptional and translational control, as well as molecular chaperones and enzymes useful in protein folding. Therefore, it is important to develop basic studies in factors regulating protein synthesis, stability, folding, targeting, and accumulation in plastids, including the protein quality control which contributes to the functional integrity of...
proteins. Our intention here is not to provide a comprehensive review of the mechanisms that regulate plastid proteostasis, but to discuss some recent insights into this field which might bring beneficial applications in plastid biotechnology for transgene expression and foreign protein accumulation.

**Transcriptional control is an important step but not the main regulator of plastid gene expression**

In spite of the prokaryotic origin and the similarity to its ancestors of the gene expression machinery, the presence of introns, RNA editing and processing of mRNA polycistronic precursors, make the chloroplast a unique organelle which combines eukaryotic and prokaryotic features (Stern et al., 2010). This uniqueness can be noticed in at least two cases: i) the dual genetic origin of the plastid proteome, made up of proteins encoded either by the plastome or by the nucleus; ii) the presence of unique ribosomal proteins, named plastid-specific ribosomal proteins, together with ribosomal proteins of prokaryotic origin (Tiller et al., 2012). As an endosymbiont, chloroplast has almost totally replaced the control of gene expression at a transcriptional level switching to a predominantly post-transcriptional control (Marín-Navarro et al., 2007). However, some transcriptional regulation may also take place because plastid transcription is performed by three RNA polymerases, two monomeric, nucleus-encoded, phage-type enzymes (dubbed RPOTp and RPOTmp), and one multimeric, plastid-encoded, eubacterial-type enzyme (dubbed PEP), which need to interact with up to six nucleus-encoded transcription initiation factors in order to perform their activity (Liere and Börner, 2007; Lerbs-Mache, 2011). Hence, transcription is considered an important step for transgene expression and it can be controlled by using appropriate promoters in the plastid transformation vectors. Most of the chloroplast transformation vectors utilise the strong promoter of the plastid rRNA operon (*Prrn*), but also other promoters like the plastidial *psbA, clpP* or the bacterial *trc* are sometimes used. The technology of plastid transformation, from the engineering of plastid transgenes in *E.coli* cloning vectors to the obtainment of stable transplastomic plants, is clearly described in a recent review (Maliga and Bock, 2011), but here we intend to underline the importance of some specific motifs present in the chloroplast transformation vectors for transgene expression.

In addition to the promoters, the transgene expression cassettes commonly employed for plastid vector construction harbour other regulatory elements like 5’-untranslated regions (5’-UTRs) and 3’-untranslated regions (3’-UTRs). These regions are both involved in the regulation of RNA stability and translation efficiency of plastid transcripts (Zou et al., 2003; Pfalz et al., 2009). The
exact modality for the correct positioning of the chloroplast ribosome on the translation initiation codon of mRNAs is still beyond our understanding, but most of the plastid mRNAs possess a Shine-Dalgarno (SD) sequence in the 5’-UTR. In prokaryotes the SD sequence (GGAGG) is located between 4 and 12 nucleotides upstream from the start codon and, being complementary to the 3’ terminus of the 16S rRNA, mediates the binding of the ribosome to an mRNA. In order to find out the mechanism of SD sequence recognition, Drechsel and Bock (2011) analyzed translation initiation from mRNAs containing multiple SD sequences, demonstrating that the plastid ribosomes displayed a bias toward preferentially utilizing the 5’-most SD sequence. The fact that both the location of plastidial SD elements is much more variable than in E.coli and that for some plastid genes SD-like sequences have turned out to be less important for translation initiation, leads to the possibility that other mechanisms for translation initiation may exist (Marìn-Navarro et al., 2007; Scharff et al., 2011). Indeed, detailed analyses of the plastid 5’-UTRs have evidenced the presence of cis-elements which facilitate the interaction with nucleus-encoded RNA-binding proteins (RBPs) (Merhige et al., 2005). These associations between protein factors and the respective RNA sequences influence the translation initiation process. There is a little concern for the use of homologous regulatory elements in chloroplast transformation vectors due to the risk of potential recombination with the endogenous ones (Rogalski et al., 2006). Therefore, for example, to transform the tobacco plastome Elghabi and colleagues (2011) utilized promoter and 3’UTR sequences from Chlamydomonas reinhardtii in order to avoid unwanted recombination with the resident copies in the plastome. Conversely, in other labs the use of homologous regulatory sequences is preferred as shown in the case of the psbA 5’-UTR. It has been demonstrated that, despite the possible accumulation of foreign proteins using a heterologous psbA 5’-UTR, it occurs with a reduced efficiency in comparison to the results achieved when the homologous regulatory elements are used (Ruhlman et al., 2011). The reason for these results is probably in the preferential association to stromal RBPs with the endogenous psbA 5’-UTR, whereas the heterologous one is an ineffective competitor. On the basis of the results obtained and after a taxonomically analysis of the region 200 bases upstream of the translation start codon, the authors suggested the use of species-specific regulatory elements for significant accumulation of foreign protein in transplastomic plants. The 5’-UTR of gene 10 of the bacteriophage T7 (T7g10) has been successfully used in plastids achieving high protein expression (Kuroda and Maliga 2001), as well as intact or truncated 5’-UTR from highly expressed genes like rbcL, psbA, atbP (Herz et al., 2005; Valkov et al., 2011). The inverted repeat sequence, typically present in the 3’UTR region of many plastid genes, can potentially form a stem-loop structure, thus becoming important in the stabilization and the accumulation of the transcript coded by the upstream sequence. This concept has been recently
highlighted in transplastomic tobacco plants expressing several GFP gene constructs with the same 5’- regulatory sequence but different 3’-UTRs (Tangphatsornruang et al., 2011). Generally, 3’ regulatory elements derived from plastid or bacterial genes are used in plastid transformation vectors. The E. coli rrnB 3’ region seems to be the most stabilizing element both in tobacco and in potato transplastomic plants, although a great mRNA stability and abundance does not always correspond to a high heterologous protein expression (Valkov et al., 2011; Tangphatsornruang et al., 2011). Therefore, a list of the best promoters, 5’UTRs and 3’UTRs combinations utilised in different experiments to achieve the highest transcript accumulation and stability can surely be made (Herz et al, 2005; Maliga, 2002; Maliga and Bock, 2011; Tangphatsornruang et al., 2011; Valkov et al., 2011), but it is quite clear that the amount of transgene mRNA, even with a good stability, does not always correspond to an elevated foreign protein synthesis and accumulation inside plastids (Eberhard et al., 2002). This is because the translational control is much more prevalent than transcriptional regulation for transgene expression (see the next chapter) and this in many cases smoothes the differences eventually present in mRNA abundance. Do the sequences downstream of the translation initiation codon affect transcript stability or translation? The answer in both cases is affirmative. Experimental results obtained by fusing N-terminal segments of highly expressed proteins in plastids to the transgene coding region, indicate that it is possible to stabilize the cyanovirin-N mRNA by sequence insertions ≥60 nt between the T7g10 5’UTR and the cyanovirin-N coding region (Elghabi et al., 2011). With a similar experimental approach it is also possible to demonstrate that 14 N-terminal amino acids encoded in the rbcL plastid gene are important determinants of translation efficiency (Kuroda and Maliga, 2001). These results encourage many scientists to empirically optimize the sequence of the region located immediately downstream of the start codon (Gray et al., 2011). If processing of polycistronic transgenic transcripts into monocistronic mRNAs is needed to increase the efficiency of translation, a small intercistronic expression element can be used to separate foreign genes of the same operon (Zhou et al., 2007).

Another aspect concerning heterologous protein expression inside plastids is that, so far, almost all plastid transformation vectors have been designed to obtain constitutive expression of the transgene. However, expression of a transgene product may result in a mutant phenotype or growth retardation due to the toxicity of the metabolite expressed, or due to its interaction with photosynthesis or the plastid endomembrane system (Lössl et al., 2003; Hennig et al., 2007). A massive expression of a highly stable protein can also lead to the exhaustion of chloroplast protein synthesis capacity resulting in a delayed growth (Oey et al., 2009). To avoid deleterious effects caused by constitutive transgene expression, different strategies have been adopted such as the one
based on a nuclear-encoded, chloroplast targeted, ethanol-inducible T7 RNA polymerase that controls plastid transgene transcription from a T7 promoter system (Lössl et al., 2005), or that depends on control of transgene expression using the bacterial lac repressor (Mühlbauer et al., 2005). However, the use of these systems is strongly limited by the requested transformation of two separate cell compartments (nucleus and plastid), or by the missing of a tight regulation of gene expression in the uninduced status. Recently, the use of engineering riboswitches (Verhounig et al., 2010) has been proposed as a tool to control chloroplast transgene expression without the use of additional nuclear or plastid transgenes. The use of a synthetic theophylline riboswitch displays great regulatory properties, because the undetectable basal GFP expression in the transplastomic tobacco plants becomes rapidly detectable after the application of theophylline metabolite.

**Translational control and protein conformation**

Proteins coded by genes located in the plastome are synthesized on plastid 70S ribosomes and the expression of these genes is mainly regulated during translation of the corresponding mRNAs (Zerges, 2004) and at post-translational level as well. Indeed, after synthesis these proteins must rapidly fold into stable three-dimensional structures which often contemplate post-translational modifications to perform their biological mission, avoiding aberrant folding and aggregation with the help of molecular chaperones (Hartl et al., 2011). The combined action of molecular chaperones and proteases maintains the quality control over protein structure and function (Wickner et al., 1999). Protein quality control components of plastids are very similar to bacteria. Chloroplasts contain members of five major chaperone families (Hsp100, Hsp90, Hsp70, Hsp60 or type 1 chaperonins, small heat shock proteins or sHSPs), and several co-chaperones, for example, J-domain proteins, GrpE and Cpn10/20 (Schroda 2004; Willmund et al., 2008). Chloroplasts also contain various proteases of bacterial ancestry which are involved in numerous aspects of the biogenesis and the maintenance of these organelles. These proteases include the multi-gene families of stromal serine Clp proteases and their AAA+ chaperones, thylakoid-bound FtsH metalloproteases, Lon, Deg, and several other proteases characterized in recent years (Adam et al., 2006).

Expression of the *psbA* gene, coding for the D1 protein, one of the core proteins of the Photosystem II (PSII), is an example of the complex regulatory strategies utilized in the chloroplasts of higher plants. D1 is damaged by light-induced oxidation and it is degraded and re-synthesized during PSII repair cycle. In the plant cells a constant pool of *psbA* mRNA associated with the polysomes does exist, but the translation of D1 can be completed only when a PSII complex without D1 is available to accept the integration of the new D1 protein (Mulo et al., 2012).
This observation leads to the conclusion that translation elongation is the prevalent mechanism for regulation of the psbA gene expression. The existence of a quality control on D1 is evident during the PSII repair cycle, which requires post-translational phosphorylation of the D1 protein to regulate its degradation. Light induces phosphorylation and damage of the D1 protein, and the damaged PSII complexes migrate from the grana stacks to the stroma lamellae, where the D1 protein is dephosphorylated and degraded by the FtsH and DegP proteases (Tikkanen et al., 2011). Replacement of damaged D1 protein also requires the assistance of molecular chaperones which are integral thylakoid membrane proteins (LPA1 and PAM68), or a luminal thylakoid protein (CYP38) (Mulo et al., 2012).

**Protein sequences impacting protein stability**

Studies on transplastomic plants have confirmed that the expression of genes integrated into the plastome is mainly regulated after the translation of the mRNAs into the corresponding protein, whose stability is the key aspect which determines foreign protein accumulation (Bellucci et al., 2005; Zhou et al., 2008; Oey et al., 2009). Small proteins are empirically known to be more unstable (Ortigosa et al., 2010), but people working on plastid biotechnology are very interested in finding general determinants of protein stability in the plastids. For example, which are the signals on the protein structure that regulate the turnover of the photodamaged PSII proteins or the proteins of the oxygen-evolving complexes? Two recent reports suggest that in plastids the major stability determinants reside in the N-terminus of the protein (Apel et al., 2010; Adam et al., 2011). This may help to explain why unstable foreign proteins result to be stabilized in plastids after addition of N-terminal peptide fusions (Ye et al., 2001; Lenzi et al, 2008).

It is generally accepted that the identity of the N-terminal amino acid residue of a protein is related to its *in vivo* half-life, a principle known as the N-end rule (Varshavsky, 1996). N-terminal amino acids are classified as stabilizing or destabilizing residues, and in the eukaryotic cytosol proteins carrying a destabilizing N-residue are degraded by the ubiquitin-proteasome system. Primary destabilizing residues are recognized directly by the E3 ubiquitin ligases, conversely secondary or tertiary destabilizing residues must be enzymatically modified (Cys also requires a previous non-enzymatic oxidation) before the corresponding protein can be targeted by E3 ubiquitin ligases to the proteasome. The N-end rule pathway is evolutionary conserved being present in all organisms examined (Graciet et al., 2010), including bacteria, even if the prokaryotic version of the rule differs in the hierarchical order of the destabilizing amino acid residues (Mogk et al., 2007), and in the associated proteolytic machinery represented by the Clp protease (Román-Hernández et al., 2009). Apel and co-workers by fusing N- and C-terminal sequences from five plastid proteins to
GFP (2010) showed that the N-terminus harbours important information affecting the half life of the protein, whereas the C-terminal seems to have scarce influence on protein half life. They also tested in one of the N-fused GFP construct all twenty amino acids in the position after the initiator Met, suggesting for plastids the existence of an N-end rule-like pathway which is not fully homologous to the bacterial one. However, the observed differences in protein accumulation conferred by the five different N-terminal sequences cannot be explained only by the identity of the penultimate amino acid, so additional major determinants of plastid protein stability must reside in the N-terminal region of the protein. In an attempt to increase, in transplastomic tobacco plants, the accumulation and stability of the rotavirus VP6 protein, even Inka Borchers and colleagues (2012) indicated the vital importance of the N-terminal part for protein stability. The additional extra amino acids (up to five) at the 5’ end of the VP6 coding region increased the expression of the foreign gene in the transplastomic plants and stabilized the protein in ageing leaves. However, the N- and the C-termini are not the only ones to influence protein stability because in other systems, for example in *Escherichia coli*, the three dimensional structure of internal protein domains turns out to be a key determinant of protein stability (Zoldák et al., 2009).

*Protein modifications influencing protein stability*

The first co-translational modification applied to almost all plastid-encoded proteins is (like in bacteria and mitochondria) the N-terminal deformylation by the enzyme peptide deformylase (PDF). After the removal of the N-formyl group, if the second residue is small and uncharged (Cys or Ala, for example), the protein can then undergo N-terminal Met excision (NME), catalyzed by methionine aminopeptidase (MAP), which involves roughly 60-70% of the proteins in a given proteome (Meinnel et al., 2006 and references inside). Inhibition of NME induced by PDF inactivation indicates that NME is involved in the regulation of protein turnover in the chloroplast (Giglione et al., 2003). In particular, retention of the N-Met in D1 and D2 proteins, that usually undergo NME, can result in acceleration of their degradation (Giglione et al., 2003). D1 and D2 proteins are primary targets of photodamage so a lot of work has been done on the proteases involved in their degradation, especially for D1 (Chi et al., 2012). In *Arabidopsis* NME seems to regulate the accumulation of D1 and D2 proteins, in concerted action with chloroplast proteases (Adam et al., 2011). The authors showed that correctly NME-processed N termini of D1 and D2 are indispensable for both FtsH-mediated quality control of the proteins and repair-related degradation of D1 and D2. Moreover, inhibition of NME induces the degradation of the D1 and D2 unprocessed forms by proteases other than FtsH. Therefore, FtsH proteases should recognize the N terminus of
both neosynthesized or damaged D1 and D2 proteins, and the open question is how FtsH complexes can recognize the N-terminal tail of these proteins (Komenda et al., 2007).

According to the N-end rule in bacteria and chloroplasts, proteins with a N-terminal Met are very stable. On the contrary, the retained N-terminal Met in D1 and D2 proteins, due to inhibition of NME, may act as a signal for degradation. Therefore, considering that the stability of other chloroplast proteins remains unaffected by the retention of their N-Met, N-Met might be necessary but not sufficient to reduce the half-life of a protein, thus other protein modifications on or close to the N-Met are required to accelerate the protein degradation process (Meinnel et al., 2006). Indeed, other polypeptide modifications likely participate in regulating the stability of a protein. More than 300 different types of protein co- or post-translational modifications (PTMs) have been described (Zhao and Jensen, 2009), among which the N-terminal modifications can be grouped in: i) removal of one or more residues such as NME or propeptide cleavage, ii) N-blocking of the intact or modified proteins such as N-α-terminal-Acetylation (NTA) or N-Myristoylation, iii) ligation events extending the length of the polypeptide (Bienvenut et al., 2011). These modifications can influence protein localization, activity or stability. With regard to stability, some PTMs can produce destabilizing N-terminal residue in proteins through, for example, cleavage by endopeptidases in the cytosol or removal of transit peptides in the stroma, or NME. Indeed, removal of the N-terminal Met in the cytosol from the initiating motif MCGGAII and the consequently exposition of the penultimate Cys, seems to target the transcription factor RAP2.12 for degradation by the N-end rule pathway (Sasidharan and Mustroph, 2011). Moreover, in yeast NTA and NME are part of the N-end rule pathway, and NTA target protein for degradation (Hwang et al., 2010).

Proteomics can be a very useful tool to achieve a comprehensive understanding of plastid protein stability. A recent review has provided an update of the current status of plastid proteome research, highlighting the importance of creating a high-quality plastid proteome atlas which should include, thanks to the improvements in mass spectrometry (MS) instrumentation, information on protein abundance, protein-protein interactions and PTMs (van Wijk and Baginsky, 2011). Such an atlas can provide useful information for functional studies and system biology, and if enriched with data on the protein half life, it can also help to clarify the relationship between protein stability and PTMs in plastids. For instance, combining 2-DE profiling with MS and pulse-chase analyses, it has been shown in *Chlamydomonas reinhardtii* that the most unstable stromal proteins (half life < 1 hour) rarely possess an α-Acetylated N-terminus in contrast to the most abundant proteins (Bienvenut et al., 2011). This suggests that, conversely to what happens in the yeast cytosol, NTA could contribute to protein stabilization in the stromal proteome. In *Arabidopsis*, large-scale analysis of PTMs has recently been carried out using chloroplast preparations (Zybailov et al.,
2008) and suspension cells as well as whole seedling lysate (Bienvenut et al., 2012). These studies have identified N-terminal modifications of a large sets of nuclear- and plastome-encoded plastid proteins, obtaining information on the protein N termini, the processing site for transit peptides of imported plastid proteins, and the frequencies of many PTMs. NTA in particular, which is considered an essential co-translational modification occurring mainly in the cytosol of Eukaryotes, appears to be a widespread protein PTM also in the chloroplast, because a high proportion of nuclear-encoded chloroplast proteins undergo NTA of the mature protein after removal of the transit peptides (Bienvenut et al., 2012). The fact that for nuclear-encoded proteins the acetylated residues have a prevalence of A, V, S, and T, whereas for chloroplast-encoded proteins the co-translational acetylated N-termini are T, R or P, likely suggests the involvement of two different N-α-Acetyl transferases (Zybailov et al., 2008).

**Targeting**

To ensure proper folding and accumulation of a recombinant protein in the transformant cells, it is necessary to target this protein to an appropriate cell compartment (Fig. 1). The plant secretory pathway and the accumulation of foreign protein in the endoplasmic reticulum (ER) is often the chosen localization. This is because a lot of information on protein quality control is available on the ER (Vitale and Ceriotti, 2004), and this knowledge greatly helps to delineate strategies to retain and accumulate heterologous proteins of interest in this sub-cellular compartment, for example by means of retention signals such as the KDEL/HDEL motif (Ma et al., 1995) or fusing the protein of interest with a protein domain capable of inducing ER-protein bodies formation (de Virgilio et al., 2008). Before the end of the XX century, plastids were rarely the final compartment of localization of foreign proteins. Transgenes, inserted into the nucleus, contains at the 5’ end a sequence coding for a transit peptide, which permits the transport of the polypeptide from the cytosol, where the transgene mRNA is translated, to the stroma through the main pathway to plastids, the TIC/TOC system present in the inner and outer envelope (Jarvis, 2008). Recently, the advantages offered by chloroplast-based expression of heterologous genes, above all those related to the achievable high level of protein expression and transgene confinement, have convinced many researchers to use the plastome transformation technology for the accumulation of bio-molecules in the chloroplasts of algae and plants (Lössl and Waheed, 2011; Michelet et al., 2012). Foreign proteins expressed in the chloroplast can be accumulated in the soluble fraction, composed of stroma and thylakoid lumen, or in the membranous fraction, formed by envelope and thylakoid membranes. The localization of these proteins within the chloroplast allows insertion in different environments with a more or less
intense proteolytic activity. Moreover, each sub-compartment owns its proteome which contains a specific box of enzymes and other molecules involved in protein folding (Ferro et al. 2010; Agrawal et al., 2011). Targeting to the outer envelope membrane seems to be possible only for nucleus-encoded genes utilizing the TOC system or a “spontaneous” insertion mechanism. Which are the protein targeting pathways within chloroplasts? Endogenous chloroplast proteins, either nucleus-encoded or plastome encoded, share the same internal routes. Once in the stroma, a protein can fold and perform its biological role, or can be delivered to other compartments if it harbours specific sorting signals. The most studied sorting mechanisms inside the chloroplast concern the targeting of proteins to the thylakoids, even if other pathways have been identified to target proteins from the stroma to the inner envelope membrane and the space between the two envelope membranes (Jarvis, 2008; Skalitzky et al., 2011). At least four different targeting mechanisms, homologous to bacterial transport systems, have been discovered for the transport of proteins into and across the thylakoid membrane. Two of these insert membrane proteins: one employs a signal recognition particle (SRP)-dependent pathway and the other one (called spontaneous) appears to require none of the known targeting factors (Ouyang et al., 2011). The other two pathways translocate luminal proteins across the thylakoid membrane. One needs ATP hydrolysis and transports proteins in an unfolded conformation by a Sec-dependent machinery. The other one is the Tat pathway that accompanies folded proteins into the thylakoid lumen using a proton gradient force (Albiniak et al., 2012). The transgenic chloroplast system has been largely used to express recombinant proteins in the stroma, but only few reports have analyzed other alternative intra-plastidial localizations like the inner envelope membrane (Singh et al., 2008) or the thylakoid lumen (Tissot et al., 2008). The correct folding of a protein to reach its tertiary or quaternary structure often needs PTMs and one of the positive characteristics of chloroplasts is that many PTMs have been described in this organelles, giving the possibility to accumulate a wide repertoire of functional heterologous proteins. Some of the chloroplast PTMs have been cited here in the previous sections (phosphorylation, NME, NTA, N-terminal myristoylation and S-palmitoylation), but others have also been described such as lipidation, oxidation, glutathionylation and S-nitrosylation (Glenz et al. 2006, Armbruster et al., 2011). However, we need more detailed information on the occurrence of PTMs in different plastidial locations in order to decide where the recombinant protein of interest has to be addressed according to its specific folding requirements.

**Localization of recombinant protein containing covalent disulfide bonds in the chloroplast**

On this purpose, an interesting case of study is the formation of covalent disulfide bonds required by many proteins for proper folding and/or their catalytic activity as enzymes. Moreover, disulfide
linkage is also important for the immunogenicity of recombinant proteins to be used as vaccine antigens. Experimentally it has been demonstrated that recombinant disulfide-bond containing proteins can be successfully expressed in transgenic chloroplasts, both in the stroma and in the lumen of thylakoids, but it is still not clear which of these two compartments is more suitable to allocate this kind of proteins. A monomeric protein with intrachain disulphide bonds (Staub et al. 2000) or a multimeric protein with interchain disulphide bonds (Daniell et al., 2001) is biologically active when expressed in the stroma of plant chloroplasts. Furthermore, a soluble and functional monoclonal antibody forming inter-chain disulfide-bridged dimers can be expressed in the chloroplast of the alga *C. reinhardtii* (Mayfield et al., 2003). However, other recombinant proteins containing disulphide bonds form in the chloroplast dimers and higher-order oligomers which are often accumulated, likely in the stroma, as insoluble aggregates (Fernandez-San Millan et al., 2003; Arlen et al., 2007; Ruhlman et al., 2007; Boyhan and Daniell, 2011; Lee et al., 2011). Also the human transforming growth factor-β3 (TGF-β3), which is a homodimer linked by disulphide bonds, when expressed in the chloroplast accumulates in unfolded insoluble aggregates so it is necessary to refold the heterologous protein to show its biological activity (Gisby et al., 2011). Even if insoluble aggregates can be easily purified, denaturation and refolding of the recombinant protein significantly increase the cost of the production process, without ensuring the attainment of the native protein conformation (Yasukawa et al., 1995). In an attempt to increase the solubility of plastid-expressed human serum albumin (HSA) and thus avoiding the formation of insoluble inclusion bodies, Sanz-Barrio and colleagues have over-expressed tobacco thioredoxins together with HSA (Sanz-Barrio et al., 2011). Thioredoxins enhance solubility of recombinant proteins that cannot form disulphide bridges in microbial expression systems, but this strategy has failed to prevent the formation of HSA protein bodies within chloroplasts. One of the possible explanations for the formation of chloroplast inclusion bodies in the stroma, is that the high translation rate and stability of the recombinant proteins alters the organelle protein homeostasis causing the formation of aberrant protein-protein interactions (such as incorrect disulphide bonds) that cause accumulation in insoluble aggregates (Villar-Piquè et al., 2010; Gisby et al., 2011).

Recently, it has been proposed that the thylakoid lumen could be a more adequate environment than the stroma for the folding and activity of protein that require oxidation of cysteine residues. When the enzyme alkaline phosphatase, a homodimer with two intramolecular disulphide bonds for each subunit, necessary for stability and catalytic activity of the molecule, has been expressed both in the stroma and in the thylakoid lumen, the results indicate that sorting to the thylakoid lumen leads to larger amounts and more active enzyme (Bally et al., 2008). In addition, the cysteine residues of another recombinant protein, capable of forming inter-chain disulphide
bonds, can be oxidized if addressed to the thylakoid membranes (De Marchis et al., 2011). Conversely, the same protein when localized into the stroma are unable to form this type of covalent linkage. A similar conclusion has been drawn for a nanobody with intramolecular disulfide bonds expressed both in the stroma and in the thylakoid lumen (Lentz et al., 2012). This is not surprising as the thylakoid lumen, as well as the intermembrane space of mitochondria, is topologically equivalent to the periplasmic space of bacteria, and active thiol oxidation pathways are present in these compartments in order to promote protein folding and regulate protein function (Herrmann et al., 2009). Redox reactions are used for metabolic regulation in both the stroma and the thylakoid lumen (Dietz and Pfannschmidt, 2011), but in light conditions the stroma should be a reducing environment with proteins maintained in a reduced state by thioredoxins which counteract the oxidative activity of molecular oxygen, whereas luminal enzymes are oxidatively activated (Buchanan and Luan, 2005). Moreover, the evidence for functional oxidoreductases such as the ER-resident protein disulfide isomerases (PDIs), that catalyze the formation, reduction, and isomerization of disulfide bonds, has been reported in C. reinhardtii but not in the stroma of plant chloroplasts (Trebitsh et al., 2001). Even though several chloroplast PDIs are described in proteomic studies, up to now only for one Arabidopsis PDI (gene At3g54960) a chloroplast location has been experimentally demonstrated but neither its intraplastidial localization nor its biochemical function (Armbruster et al., 2009). Conversely, functional oxidoreductases have been identified in the Arabidopsis thylakoid membranes like LTO1, which has a thioredoxin-like domain exposed to the thylakoid lumen capable of catalyzing disulfide bond formation in the PsbO protein, a luminal subunit of PSII requiring for its stability the formation of a single intramolecular disulfide (Karamoko et al., 2011). Recently, other reports describe two Zn finger proteins showing disulfide isomerase activity located in the thylakoid membrane and involved in maintenance of PSII (Hall et al., 2010) or thylakoid biogenesis (Tanz et al., 2012). This information demonstrate together that a catalyzed thiol-disulfide process operates in the luminal side of the thylakoid membrane, with a thio-oxidizing pathway promoting formation of disulfide bonds and a thio-reducing pathway to transfer reducing equivalents (Motohashi and Hisabori, 2010).

Foreign protein localization in the chloroplast seems to be of vital importance for getting properly folded and functional polypeptides, therefore a thorough understanding of the mechanisms governing the various transport systems within chloroplast, as well as the identification of a complete chloroplast proteome (enzymes, chaperones, complex sub-units, etc.), is essential in order to obtain the best results. Chaperones have many functions in chloroplast biogenesis, for example protein trafficking inside chloroplasts is assisted by chaperone systems which perform an essential role in facilitating protein translocation (Su and Li, 2010) and in protecting membrane proteins from
unwanted aggregation (Jaru-Ampornpan et al., 2010). The targeting factor of the chloroplast SRP (cpSRP43) is a molecular chaperone that reverses the aggregation of its substrate proteins (Jaru-Ampornpan et al., 2010). This means that en route to their destinations, membrane proteins translocated by the SRP pathway are protected from misfolding. Recently it has been shown that a ER plant signal peptide is capable of carrying a fusion protein called zeolin, whose gene is inserted into the plastoma, to the thylakoid membranes where the recombinant protein accumulates as trimers. In the stroma the folding of the same protein without the signal peptide is hampered as the protein accumulates at low amounts in a monomeric form (De Marchis et al. 2011). Thus, we can speculate that whatever is the sorting mechanism that directs zeolin to the thylakoids, it has probably also a stabilizing effect on the recombinant protein (Fig. 2).

Conclusions

Nowadays different technologies can be adopted to transform a plant and plastid biotechnology is one of the most promising ways. The future perspectives of plastid genetic engineering have been clearly described by Clarke and colleagues (2011), and here we try to emphasize that a big effort has to be done in understanding the factors that regulate plastid proteostasis (Fig. 3). More exhaustive studies on chloroplasts are necessary to elucidate the relationship between protein stability and the characteristics of its N-terminal part. Moreover, the fast growing amount of data coming from proteomic studies can be utilized to create and/or to improve prediction tools for N-terminal modifications and their relationship with protein half life, as illustrated for the on line prediction tool TermiNator (http://www.isv.cnrs-gif.fr/terminator2/index.html), which predicts NME, NTA, N-terminal myristoylation and S-palmitoylation of either prokaryotic or eukaryotic proteins originating from organellar or nuclear genomes. The role of chaperones in different plastid compartments has also to be clearly elucidated in order to understand how these molecules assist protein folding in plastids and how this information can be used in plastid genetic engineering. For example, several chaperones have been co-expressed together with recombinant proteins in the chloroplast so as to improve the accumulation of the latter. The results are encouraging only for the bacterial putative chaperonin ORF 2 (De Cosa et al., 2001), but not for both the ER chaperone BiP and the chloroplast Trxs f and m (Bellucci et al., 2007; Sanz-Barrio et al., 2011). Another aspect to be taken into consideration is that the members of the plastid family are classified according to their metabolic specialization inside the cell and that more than ten different variants can be identified like chloroplasts, chromoplasts, amyloplasts, etc. Proteostasis and gene expression in non-green plastids is not well
described, but some results show that these systems are quite distinct from those of chloroplasts (Kahlau and Bock, 2008; Valkov et al., 2009; Zhang et al., 2012). A major challenge for the future will be the identification of substrates for the chloroplast proteases in order to better understand the determinants of protein stability. Indeed, it remains unknown which protease(s) degrade the most part of plastid proteins. Searching for possible substrates is the objective of Stanne and colleagues (2009) which identify potential substrates for Clp protease in higher plants.

In conclusion, a cocktail of basic and applied studies on plastid proteostasis is needed in the coming years to predictably express foreign proteins from the plastid genome.

Acknowledgments

We thank Prof. Ralph Bock for critical reading of the manuscript. We sincerely apologize to all colleagues whose work could not be cited because of space constraints.

LITERATURE CITED


without cleavable transit peptides: rare exceptions or a major constituent of the chloroplast proteome? Mol Plant 2:1325-1335


Bellucci M, De Marchis F, Nicoletti I, Arcioni S (2007) Zeolin is a recombinant storage protein with different solubility and stability properties according to its localization in the endoplasmic reticulum or in the chloroplast. J Biotechnol 131:97-105


rapid purification and folding into a biologically active molecule. Plant Biotechnol J 9:618-628


Lentz EM, Garaicoechea L, Alfano EF, Parreno V, Wig dorovitz A, Bravo-Almonacid FF (2012) Translational fusion and redirection to thylakoid lumen as strategies to improve the accumulation of a camelid antibody fragment in transplastomic tobacco. Planta DOI 10.1007/s00425-012-1642-x


Motohashi K, Hisabori T (2010) CcdA is a thylakoid membrane protein required for the transfer of reducing equivalents from stroma to thylakoid lumen in the higher plant chloroplast. Antioxid Redox Signal 13:1169-1176


**Figure 1.** Expression of the bean storage protein phaseolin and its mutated forms is very different according to the compartment of localization inside transgenic tobacco leaf cells. Proteins extracted from transformed tobacco plants were analysed by western blot analysis using anti-phaseolin antiserum. Lanes 1-2: phaseolin (arrow) expressed from nuclear transformed plants is secreted to the vacuole (VAC) where is degraded to 20-25 kDa proteolytic fragments (curly bracket). Phaseolin can be accumulated in the endoplasmic reticulum (ER) of transgenic nuclear plants, either adding to the polypeptide a KDEL retention signal (not shown), or (lanes 3-4) expressing phaseolin fused to a γ-zein N-terminal domain (the name of the fusion protein is zeolin). The accumulation of phaseolin in the ER was much higher than that in the vacuole, both as phaseolin-KDEL (not shown) and as zeolin (arrowhead), which forms large aggregates (vertical bar) due to the formation of inter-chain disulfide bonds (Mainieri et al., 2004). In the chloroplast (CHL) of transplastomic plants expressing the phaseolin fusion protein zeolin, the recombinant protein devoid of its signal peptide (Δzeo) was barely detectable in the stroma (str, lanes 5-6), conversely when targeted to the thylakoid membrane (thy, lanes 7-8) it is clearly detectable (De Marchis et al., 2011). Molecular mass in kDa (number at right) is indicated. Asterisk: contaminant peptides.

**Figure 2.** Model of the possible site of accumulation in the chloroplast of recombinant proteins containing covalent disulfide bonds. Recombinant disulfide-containing proteins (represented for simplicity by a monomeric protein with a intra-chain disulfide bridge) when expressed in transplastomic plants have been in most cases accumulated in the stroma. In this compartment only few studies reported the obtainment of soluble proper folded polypeptides, whereas the other recombinant proteins described were probably misfolded, often originating insoluble aggregates with incorrect disulphide bonds (depicted by red lines). Alternatively, the misfolded proteins are targeted for degradation by stromal proteases, resulting in a low level of protein expression. Recently, recombinant proteins requiring oxidation of Cys residues were also addressed by targeting signals to the thylakoid membrane or lumen, where they correctly folded. The thickness of the arrows is roughly proportional to the number of publications describing the expression of a recombinant protein in a chosen chloroplast sub-compartment (read text for more details). The yellow square with a red C letter inside indicates chaperone pathways which assist in protein folding and trafficking.

**Figure 3.** Mechanisms and main players involved in maintenance of plastid proteostasis which affects foreign protein accumulation.
Transcriptional and translational control, proteases, molecular chaperones, protein folding and targeting are graphically illustrated. See text for the discussion. The yellow square with a red C letter inside indicates chaperone pathways; the yellow square with a red P letter inside indicates proteases. The yellow triangle with black PR letters inside indicates the transgene promoter. 5’-untranslated regions (5’-UTRs) and 3’-untranslated regions (3’-UTRs) are indicated, as well as the AUG translation initiation codon and the sequences downstream of it which affect transcript stability and/or translation (downstream box region, DB). RBPs, RNA-binding proteins. PTMs, co- or post-translational modifications.
Fig. 1
Figure 1. Expression of the bean storage protein phaseolin and its mutated forms is very different according to the compartment of localization inside transgenic tobacco leaf cells.

Proteins extracted from transformed tobacco plants were analysed by western blot analysis using anti-phaseolin antiserum. Lanes 1-2: phaseolin (arrow) expressed from nuclear transformed plants is secreted to the vacuole (VAC) where is degraded to 20-25 kDa proteolytic fragments (curly bracket). Phaseolin can be accumulated in the endoplasmic reticulum (ER) of transgenic nuclear plants, either adding to the polypeptide a KDEL retention signal (not shown), or (lanes 3-4) expressing phaseolin fused to a γ-zein N-terminal domain (the name of the fusion protein is zeolin). The accumulation of phaseolin in the ER was much higher than that in the vacuole, both as phaseolin-KDEL (not shown) and as zeolin (arrowhead), which forms large aggregates (vertical bar) due to the formation of inter-chain disulfide bonds (Mainieri et al., 2004). In the chloroplast (CHL) of plastidic plants expressing the phaseolin fusion protein zeolin, the recombinant protein devoid of its signal peptide (Δzeo) was barely detectable in the stroma (str, lanes 5-6), conversely when targeted to the thylakoid membrane (thy, lanes 7-8) it is clearly detectable (De Marchis et al., 2011). Molecular mass in kDa (number at right) is indicated. Asterisk: contaminant peptides.
Fig. 2
Figure 2. Model of the possible site of accumulation in the chloroplast of recombinant proteins containing covalent disulfide bonds. Recombinant disulfide-containing proteins (represented for simplicity by a monomeric protein with an intra-chain disulfide bridge) when expressed in transplastomic plants have been in most cases accumulated in the stroma. In this compartment only few studies reported the obtainment of soluble properly folded polypeptides, whereas the other recombinant proteins described were probably misfolded, often originating insoluble aggregates with incorrect disulphide bonds (depicted by red lines). Alternatively, the misfolded proteins are targeted for degradation by stromal proteases, resulting in a low level of protein expression. Recently, recombinant proteins requiring oxidation of Cys residues were also addressed by targeting signals to the thylakoid membrane or lumen, where they correctly folded. The thickness of the arrows is roughly proportional to the number of publications describing the expression of a recombinant protein in a chosen chloroplast sub-compartment (read text for more details). The yellow square with a red C letter inside indicates chaperone pathways which assist in protein folding and trafficking.
Figure 3. Mechanisms and main players involved in maintenance of plastid proteostasis which affects foreign protein accumulation. Transcriptional and translational control, proteases, molecular chaperones, protein folding and targeting are graphically illustrated. See text for the discussion. The yellow square with a red C letter inside indicates chaperone pathways; the yellow square with a red P letter inside indicates proteases. The yellow triangle with black PR letters inside indicates the transgene promoter. 5’-untranslated regions (5’-UTRs) and 3’-untranslated regions (3’-UTRs) are indicated, as well as the AUG translation initiation codon and the sequences downstream of it which affect transcript stability and/or translation (downstream box region, DB). RBPs, RNA-binding proteins. PTMs, co- or post-translational modifications.