Title of article: Plastid biotechnology: food, fuel and medicine for the 21st century

Authors: Pal Maliga* and Ralph Bock

Addresses: Waksman Institute of Microbiology, Rutgers, The State University of New Jersey, 190 Frelinghuysen Road, Piscataway, New Jersey, 08854-8020, USA, and Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, D-14476 Potsdam-Golm, Germany
Footnotes

Financial source: Research in PM’s laboratory is supported by grants from the USDA National Institute of Food and Agriculture Biotechnology Risk Assessment Research Grant Program Award No. 2005-33120-16524 and 2008-03012. Research on chloroplast genetic engineering in RB’s laboratory is supported by grants from the Deutsche Forschungsgemeinschaft (DFG), the Bundesministerium für Bildung und Forschung (BMBF), the European Union (EU Framework program 7) and the Max Planck Society.

Corresponding author e-mail: maliga@waksman.rutgers.edu
ABSTRACT

Originating from endosymbiotic cyanobacteria, plastids are highly polyploid and have prokaryotic-like transcription and translation machineries. The majority of plastid functions is carried out not from plastid encoded genes but from ~3,000 nuclear genes that target proteins to plastids. The compact 125-kb to 155-kb plastid genome (ptDNA) encodes less than one hundred proteins. However, both plastid and nuclear-encoded traits can be modified by plastid genome engineering. Furthermore, novel genes and operons may be incorporated from heterologous sources to expand the repertoire of biosynthetic pathways in plants. In this review, background is provided on the plastid genetic system and how to obtain genetically uniform plants from plastid transformation. We focus on the principles of plastome engineering and address transgene design enabling biotechnological applications. We assess the state of plastome engineering in the microalga Chlamydomonas reinhardtii, the bryophytes Physcomitrella patens and Marchantia polymorpha, and flowering plants. In the flowering plant group, plastid transformation is routine only in tobacco, tomato, petunia, potato, soybean, lettuce and cabbage. Future applications of plastid genome manipulation are emphasized in the discussion of potential contributions to food security, biofuel production and plant-based biopharmaceutical or industrial enzyme manufacturing.
DNA-containing cellular compartments in plant cells are the nucleus, plastids and mitochondria. The nuclear genome, encoding ~29,000 to 32,000 genes, is the most common target for biotechnological applications. The number of genes encoded by the plastid and mitochondrial genomes is much smaller, ~120 and ~60, respectively. The number of processes that may be modified by engineering the organelar genomes is much higher than the number of genes would suggest. That is because ~10% of the nuclear gene products are targeted to plastids and ~10% to mitochondria (Leister, 2003), making these nuclear-encoded processes amenable to manipulation by plastome engineering. Traits that may be engineered by plastid genome manipulation are not restricted by the plant’s gene content because incorporating new genes or gene clusters from heterologous sources may expand the plastid’s biosynthetic repertoire beyond what is provided by nature.

Engineering of the plastid genomes (plastomes, plastid DNA or ptDNA) was first accomplished in *Chlamydomonas reinhardtii*, a unicellular green alga (Boynton et al., 1988), followed by plastid transformation in *Nicotiana tabacum*, a flowering plant species (Svab et al., 1990). Since 1988 plastid transformation has been expanded to a diverse group of species (see below). However, commercial applications are lagging behind and currently no crops are grown commercially utilizing this technology. This review focuses on the principles of plastome engineering and highlights recent developments. Most examples will be described in tobacco (*N. tabacum*), which is the model species of plastid engineering. Additional information on the biotechnological applications of plastid transformation can be found in recent reviews (Bock, 2007; Daniell et al., 2009) (Cardi et al., 2010).

**GENOME SORTING TO OBTAIN GENETICALLY STABLE TRANSPLASTOMIC PLANTS**

The target of the transforming DNA, the plastid genome, is highly polyploid. The number of plastids per cell and the number of ptDNA copies per plastid is dependent on the species and the cell type. For example, an *Arabidopsis thaliana* leaf mesophyll cell contains about 120 chloroplasts, and these harbor ~1,000-1,700 copies of the ~154-kb plastid genome (Zoschke et
al., 2007). *N. tabacum* leaf cells contain a comparable number (~100) chloroplasts harboring ~10,000 copies of the ~156-kb ptDNA (Shaver et al., 2006). The ptDNA in chloroplasts is localized to membranes in clusters of ~10 ptDNA copies, referred to as nucleoids (Fig. 1A) (Maliga, 2004). Obtaining plants with a genetically uniform population of transformed plastid genomes (transplastomes or T-ptDNA) is still the bottleneck that hinders rapid extension of the technology to new crops. Fig. 1 depicts the stages of plastid sorting that ultimately yield genetically stable transplastomic plants and are discussed below.

When the transforming DNA is introduced on the surface of microscopic particles (Fig. 1A,B), only one or a few chloroplasts in a leaf cell may be damaged by the impact, and only a few of the ~100 ptDNA copies incorporate the transforming DNA. However, the transforming DNA carries antibiotic-detoxifying genes conferring a selective advantage to plastids that carry the T-ptDNA. Because the selective agent can most conveniently be administered in the tissue culture environment, selective enrichment for the T-ptDNA is carried out in tissue culture cells. The tissue culture medium triggers cell division, yielding meristematic cells with 10 to 14 proplastids, each of which carries only one or two nucleoids. Reduction of plastid number from 100 to 10 to 14 greatly accelerates plastid sorting during cell division, during which plastids carrying the T-ptDNA are dividing at a faster rate (Fig. 1B). Plastids carrying only the wild-type ptDNA are ultimately lost by dilution during cell division. Depicted in the Figure is selection for an aurea construct that confers a pale pigment phenotype to plants (Lutz and Maliga, 2008; Tungsuchat-Huang et al., 2011).

In tobacco, selection for T-ptDNA is carried out on a shoot induction medium. Cells carrying T-ptDNA are relieved from inhibition by the toxic selective agent and will regenerate shoots on the selective medium (Fig. 2). Cells in the shoot derive from three developmental layers, each of which are the progeny of two to three slowly-dividing long-term stem cells (Fig. 1C). Genetic uniformity of a regenerated plant is ensured when each of the long-term stem cells in the three developmental layers carries the same T-ptDNA. Shoots emerging from a bombarded leaf are chimeric, consisting of transplastomic and wild-type sectors. The transplastomic sectors, often very small, can be visualized by the expression of aurea transgenes that confer a golden-yellow phenotype to leaves (Fig. 1C, bottom central). The pigment phenotype is due to posttranscriptional interference with the plastid *clpP1* gene expression by the aurea *aadA* transgene. Wild-type sectors in the shoot are present because transplastomic sectors protect wild
type cells against antibiotics in culture. Cross protection enables wild-type shoots to form on a selective medium even after two cycles of shoot regeneration. Genetically stable transplastomic plants can be obtained by collecting seeds from plants with a uniform phenotype after two cycles of plant regeneration, or from branches that carry transplastomic sectors in the second leaf layer, the source of germline cells (Lutz and Maliga, 2008; Tungsuchat-Huang et al., 2011). Because heteroplastomic cells are rare even in variegated plants and are localized at the edge of sectors, the aurea and green colors typically identify homoplastomic transgenic and wild type sectors, respectively. Plants in Fig. 1C were transformed with an aurea aadA gene that is selectable in culture and gives a visual phenotype in leaves (Tungsuchat-Huang et al., 2011). In the absence of a visual marker, DNA gel blot analysis is employed to identify plants with a uniform population of T-ptDNA molecules (Maliga, 2004).

THE TECHNOLOGY OF PLASTID TRANSFORMATION

Incorporation of foreign DNA is based on homologous recombination between the targeting region of the vector and the ptDNA (Fig. 3). The transformation vectors are E. coli plasmids that do not replicate in plastids, thus the marker gene encoded in the vector will be stably expressed only if incorporated in the plastid genome by homologous recombination. The transforming DNA is introduced on the surface of microscopic (0.4 μm – 1.0 μm) gold or tungsten particles or by polyethylene glycol treatment. Biolistic DNA delivery is used when the targets are plastids in intact tissue; polyethylene glycol treatment is used for DNA introduction into protoplasts (Dix and Kavanagh, 1995). The most commonly used selective marker gene is the aadA encoding spectinomycin resistance (Svab and Maliga, 1993). Kanamycin (Carrer et al., 1993), chloramphenicol (Li et al., 2011) and the amino acid analogues 4-methylindole (4MI) and 7-methyl-DL-tryptophan (7MT) (Barone et al., 2009) have also been successfully employed as selective agents.

The types of plastid genome manipulations include: knockout of plastid genes to probe function; replacement of plastid genes with mutant forms; and insertion of transgenes to confer novel functions. Replacement of the tobacco rbcL plastid gene (T-rbcL), encoding the Rubisco large subunit, with the sunflower large subunit (S-rbcL) is shown in Fig. 3A (Sharwood et al.,
The example shown in Fig. 3B is insertion of \textit{aadA} and six genes (~6.5 kb) of the luciferase (\textit{lux}) operon in the plastid genome in the \textit{trnI-trnA} intergenic region (Krichevsky et al., 2010). Thus far this is the highest number of genes inserted in the ptDNA. In both cases, the genes-of-interest (S-\textit{rbcL} and \textit{lux} operon) were incorporated in the plastid genome by homologous recombination \textit{via} the homologous flanking (“targeting”) sequences (dashed lines in Fig. 3). The recovery of T-ptDNA was subsequently facilitated by selection for the linked spectinomycin resistance (\textit{aadA}) marker gene.

Once a uniform population of engineered T-ptDNA has been obtained, the marker genes are no longer necessary to maintain the T-ptDNA. Excision of the marker gene enables multi-step transformation with the same marker gene. The metabolic burden from high-level expression of the marker gene, and opposition to antibiotic resistance markers in field crops are additional reasons that make the removal of marker genes desirable. Protocols for marker gene removal employ the Cre and Int phage site-specific recombinases that excise the marker genes \textit{via} flanking recombinase target sites (Fig. 4) (Lutz and Maliga, 2007). Alternatively, the marker gene may be removed by recombination \textit{via} flanking direct repeats (Fig. 5) (Kode et al., 2006).

Extension of the technology of plastid transformation to new crops has been more challenging than nuclear gene transformation. Although there are reports of partial success in many species, reproducible protocol for plastid transformation has been described only in tobacco (\textit{N. tabacum}) (Svab and Maliga, 1993), tomato (\textit{Solanum lycopersicum}) (Ruf et al., 2001), petunia (Zubko et al., 2004), potato (\textit{Solanum tuberosum}) (Valkov et al., 2011), soybean (\textit{Gycine max}) (Dufourmantel et al., 2004), lettuce (\textit{Lactuca sativa}) (Kanamoto et al., 2006) and cabbage (\textit{Brassica oleracea}) (Liu et al., 2007). Monocots as a group appear to be the most recalcitrant species.

\section*{ENGINEERING OF PLASTID TRANSGENES FOR HIGH-LEVEL PROTEIN EXPRESSION}

Expression of transgenes is facilitated by the availability of expression cassettes, into which coding regions can be inserted (Fig 6). These cassettes are available in plastid transformation vectors that also function as \textit{E. coli} cloning vectors, so that transformation-ready vectors can be
obtained in one cloning step. The 5′ regulatory regions are provided by a PL cassette, which includes a promoter (P) and translation control sequences (L, leader). The translation control sequences may be the mRNA 5′ Untranslated Region (5′ UTR), or the 5′ Translation Control Region (5′ TCR) that includes the 5′ UTR and the coding region’s N terminus. The role of the 5′ UTR is to stabilize the mRNA and to facilitate loading of mRNAs onto the prokaryotic-type 70S ribosomes. Loading of the mRNAs is facilitated by mRNA-16S rRNA interactions by a variant (GGAGG, GGAG) of the prokaryotic Shine-Dalgarno (SD) sequence found upstream of the AUG (or, much less frequently, GUG) translation initiation codon. A significant number of mRNAs, such as \textit{atpB}, has no recognizable SD sequence and 5′ UTR-binding proteins are thought to facilitate translation initiation in these reading frames. Efficient translation of some of the mRNAs is dependent on processing (Yukawa et al., 2007). The 3′ regulatory region or T cassette encodes the mRNA 3′ UTR, that typically harbors a stem-loop-type RNA secondary structure. The 3′ regulatory region is important for mRNA stability (Monde et al., 2000).

Most biotechnological applications utilize the promoter (P) upstream of the plastid rRNA (rrn) operon. Prrn is the strongest plastid promoter, but its native transcripts are not translated. To enable translation, Prrn is fused with the 5′ UTR or 5′ TCR of plastid or other prokaryotic-type genes. Expression of a protein using a PL-UTR cassette enables expression of the protein with its native (unmodified) N- and C-termini, because the expression signals are linked up with the coding region via restriction sites comprising the translation initiation and termination codons (Fig. 6B). Sometimes, using a PL-TCR fusing the protein of interest with an N-terminal peptide is the only approach that yields high-level protein expression. Examples are expression of the EPSPS enzyme from the CP4 coding region (Ye et al., 2001), or the Human Papillomavirus (HPV) L1 capsid protein (Lenzi et al., 2008) (Fig. 6B). The engineered plastid 5′ UTR (5′ TCR) is typically a truncated and mutant form of the corresponding native sequence: the length of the 5′ UTR is reduced to minimize the probability of unwanted homologous recombination (with the native UTR copy) and point mutations are introduced to remove undesirable restriction sites. For example, the highly expressed Prrn:LrbcL PL-TCR cassette (typically giving ~10% of total soluble protein; TSP) contains only 57 of the 189 nucleotide \textit{rbcL} leader sequence and carries a point mutation to eliminate an \textit{EcoRI} restriction site (Kuroda and Maliga, 2001). The importance of the 5′ UTR/5′ TCR was shown by protein accumulation varying as much as 10,000-fold depending on the choice of translation control signals (Maliga, 2002). The most efficient
translation control sequences derive from the *E. coli* phage T7 gene10 (T7g10) (Kuroda and Maliga, 2001), the *Bacillus thuringiensis* cry9Aa2 (Chakrabarti et al., 2006), and the plastid *rbcL* gene in its engineered form (Kuroda and Maliga, 2001). The highest level of protein expression in chloroplasts on record is >70% of TSP of a highly stable protein antibiotic in a PrmT7g10-UTR/TrbcL cassette (Oey et al., 2009) (see below). Translation of mRNAs, at least in some cases, is dependent on intercistronic transcript processing (i.e., cleavage of polycistronic into monocistronic mRNAs). Where processing of polycistronic transcripts is required, a short sequence encoding an Intercistronic Expression Element (IEE) may be inserted to trigger intercistronic cleavage (Zhou et al., 2007).

The choice of the insertion site in the plastome may have a profound affect on the level of protein accumulation. Inserting a transgene in the repeated region of the ptDNA doubles the number of transgene copies per genome, as compared to insertions in the unique regions. Insertion of transgenes between genes of a heavily transcribed operon will further increase the level of translatable mRNA, typically yielding higher protein levels (De Cosa et al., 2001).

**APPLICATIONS IN AGRICULTURE**

Currently, no transplastomic crops are grown commercially. However, proof of concept exists in tobacco for many potential applications. Commercial exploitation is hindered by the lack of the technology in the major field crops such as maize, wheat and rice, or only recent implementation of the technology (e.g., in soybean).

The first example demonstrating the great potential of the transplastomic technology came from the expression of the *B. thuringiensis* (Bt) cry1A(c) insecticidal protein in chloroplasts (McBride et al., 1995). The coding region of the bacterial Bt genes, when expressed in the plant nucleus, yielded extremely low protein levels. It was noted that the Bt genes are AT-rich as compared to plant nuclear genes, and thus led to the speculation that low expression was due to premature transcription termination, aberrant mRNA splicing, mRNA instability and/or inefficient codon usage. Expression of Bt proteins from a synthetic coding region with an increased GC content dramatically increased protein yield, from undetectable to 0.2% to 0.3% of total soluble cellular protein. However, when the unmodified bacterial coding region was
expressed in chloroplasts, the mRNA was stable and the cry1A(c) protein accumulated to 3-5% of the total soluble cellular protein, which was considered extraordinary at the time. In the meantime, several Bt proteins have been expressed in chloroplasts with a similar outcome: high-level accumulation of the Bt protein from the bacterial coding sequence. The protein levels obtained were >10% in tobacco (Chakrabarti et al., 2006) and cabbage (Liu et al., 2008) and, when including two open reading frames upstream of the cry2Aa2 operon, one of which encodes a putative chaperonin, the protein accumulated to up 45% of the total soluble protein (De Cosa et al., 2001).

Herbicide resistance is one of the most common commercial transgenic traits. Herbicide resistances in commercial crops are encoded in the nucleus, unavoidably allowing for occasional dissemination of the herbicide resistance by pollen. By now, resistances to the same set of herbicides have been obtained by expression of transgenes in plastids. Examples include resistance to glyphosate (Ye et al., 2001), Bialaphos/Liberty (Iamtham and Day, 2000; Lutz et al., 2001), isoxaflutole (IFT) (Dufourmantel et al., 2007) and sulfonylurea herbicides (Shimizu et al., 2008). If these plastid transgenes are incorporated in commercial crops, plastid localization provides an efficient containment tool (Ruf et al., 2007; Svab and Maliga, 2007). Expression of transgenes in the plastid genome to confer herbicide resistance is an example for modification of a nucleus-encoded trait (i.e., a nucleus-encoded metabolic pathway that is sensitive to a herbicide) by plastome engineering.

Improving the efficiency of photosynthesis by engineering of the photosynthetic machinery is the holy grail of plant breeding. While the technology is available, thus far very little progress has been made. The only exception is Rubisco, Nature’s main CO2-fixing enzyme, the large (catalytic) subunit of which is encoded in the plastome. Exploration of strategies has begun to examine the feasibility of supercharging photosynthesis by Rubisco engineering to drive a new green revolution (Whitney et al., 2011).

METABOLIC PATHWAY ENGINEERING

Metabolic pathway engineering offers great potential for improving the yield and nutritional quality of foodstuffs and animal feed. Moreover, it can provide an inexpensive and renewable
source of raw materials for industrial processes (‘green chemicals’). Attempts to engineer the
carotenoid biochemical pathway by using plastid transformation have been particularly
successful (Wurbs et al., 2007; Hasunuma et al., 2008; Apel and Bock, 2009) (Fig. 7).
Carotenoids are important antioxidants and represent an essential component of the human diet.
The carotenoid β-carotene (also referred to as provitamin A) is of special importance because it
provides the precursor for vitamin A, a lipid-soluble vitamin essential to all vertebrates. Vitamin
A deficiency is a global health problem and, therefore, increasing the provitamin A content of
crops represents an important goal of breeding and genetic engineering efforts. Taking advantage
of the availability of a plastid transformation protocol for tomato, a plant that accumulates
massive amounts of carotenoids in fruits during the ripening process, provitamin A production in
transplastomic tomatoes was attempted. Lycopene, the major storage carotenoid of ripe red
tomato fruits can be converted into β-carotene in a single-step enzymatic reaction catalyzed by
lycopene β-cyclases (Fig. 7A). Lycopene β-cyclase genes from various carotenoid-synthesizing
source organisms (bacteria, fungi and plants) were fused to plastid expression signals and tested
as transgenes in transplastomic tomato plants (Wurbs et al., 2007; Apel and Bock, 2009).
Synthesis of β-carotene was most efficient with a plant cyclase gene borrowed from daffodil
(Narcissus pseudonarcissus), with provitamin A levels reaching as much as 1 mg per g dry
weight of the fruit (while wild-type fruits have less than 200 ng provitamin A per g) (Apel and
Bock, 2009)(Fig. 7B). Interestingly, expression of the daffodil lycopene β-cyclase from the
tomato plastid genome did not only result in efficient conversion of lycopene to β-carotene, it
also led to a 50% increase in total carotenoid content, thus providing a welcome additional
improvement in the nutritional quality of the fruits. This finding suggests that the lycopene β-
cyclase reaction represents an important control point in tomato carotenoid biosynthesis, which,
at least in part, determines the total flux through the pathway in the fruit.

The possibility of producing novel carotenoids was assessed in a proof-of-concept study
in tobacco (Hasunuma et al., 2008). The ketocarotenoid astaxanthin is a high-value carotenoid
that is extensively used as a food and feed additive (e.g., in salmon farming, were it is
responsible for the orange color of the meat and accounts for up to 25% of the total costs). It also
has many applications in the pharmaceutical and cosmetic industry, mainly due to its much
higher antioxidant activity compared to most other carotenoids. Astaxanthin is not synthesized
by higher plants, but accumulates in some marine bacteria and algae. Its biosynthesis utilizes β-
carotene and/or zeaxanthin as precursors and involves two enzymes: β-carotene ketolase and β-carotene hydroxylase (Fig. 7A). Co-expression of genes for the two enzymes (taken from a marine bacterium of the genus Brevundimonas) in tobacco chloroplasts resulted in astaxanthin accumulation to up to more than 0.5% of the dry weight of the leaves, indicating that chloroplasts can accommodate significant amounts of novel carotenoid species. Similar to lycopene β-cyclase expression in tomatoes, this also resulted in a substantially increased total carotenoid content in the transplastomic tobacco plants (Hasunuma et al., 2008).

Engineering of the carotenoid pathway involved adding one additional gene. Because of their ability to express operons, chloroplasts were a natural choice for the production of polyhydroxybutiric acid (PHB) requiring expression of three bacterial genes. In the most successful case, the *Ralstonia eutropha* phbC-phbB-phbA genes were expressed from the T7gene10 promoter (PT7g10) (Lossl et al., 2005). The polycistronic mRNA was transcribed by a nuclear-encoded, plastid-targeted T7 RNA polymerase. The first gene, *phbC*, was translated from the T7g10 leader; *phbB* and *phbA* from the bacterial intergenic sequences. Production of PHB in chloroplasts indicates that there is sufficient conservation between the plastid and prokaryotic translation machineries to translate the bacterial proteins from the native mRNAs. Toxicity associated with PHB production was overcome by expressing the plastid-targeted T7 RNA polymerase from an ethanol-inducible promoter.

Construction of truly autonomously luminescent plants required expression of a fully functional bacterial luciferase pathway consisting of six genes (*luxCDABEG*) (Krichevsky et al., 2010). Again, *luxC*, the first gene of the *lux* operon was expressed from the plastid PrnLrbcL promoter/leader cassette, the rest of the *lux* operon genes from their native translation control sequences. Emission of visible light detectable by the naked eye attested to the plastid’s ability to properly interpret the prokaryotic expression signals.

Part of lipid biosynthesis takes place inside plastids, making it amenable to engineering by plastid transformation. Over-expression of the plastid *accD* gene, encoding the beta-carboxyl transferase subunit of acetyl-CoA carboxylase (ACCase), increased leaf lipid content (Madoka et al., 2002). Expression of a Δ⁹ desaturase gene in tobacco chloroplasts from either the wild potato species *Solanum commersonii* or the cyanobacterium *Anacystis nidulans*, resulted in altered fatty acid profiles and an increase in their unsaturation level both in leaves and seeds (Craig et al., 2008). Given the interest in biofuel production, increasing lipid content by plastids engineering
CHLOROPLASTS AS BIOREACTORS FOR MOLECULAR FARMING

Arguably the most alluring feature of transplastomic plants is their enormous capacity to accumulate foreign proteins (Maliga, 2004; Bock, 2007; Daniell et al., 2009). This is, at least in part, due to the high number of chloroplasts per cell, the large volume of the cell occupied by the chloroplast compartment and the high copy number of the plastid genome with hundreds or even thousands of copies being present in a single cell. The carbon dioxide-fixing enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) provides a case in point for the high protein accumulation capacity of the chloroplast. Its large subunit is encoded in the plastid genome and, in C3 plants, the Rubisco enzyme can accumulate to up to more than 20-30% of the total soluble protein in leaves (Whitney et al., 2011).

High yields of recombinant protein are particularly crucial to the production of pharmaceutical proteins in plants, an area of biotechnology commonly referred to as molecular farming. Proteinaceous biopharmaceuticals include, for example, antibodies, vaccines and various antimicrobials. Pharmaceutical proteins can either be purified from the plant or, in the case of vaccines, can be expressed in edible plants and administered orally. In both cases, high expression levels are of pivotal importance. The costs of downstream processing usually make up by far the largest fraction of the costs for plant-made pharmaceuticals and protein purification costs are, in general, negatively correlated with expression levels. High protein accumulation levels are absolutely essential for the production of edible vaccines, because stimulation of the mucosal immune system generally requires much higher doses of the antigen than conventional vaccination by injection into the bloodstream.

With the concept of oral vaccination in mind, a large number of antigens have been attempted to produce to high levels in chloroplasts; for recent reviews see (Bock and Warzecha, 2010; Cardi et al., 2010). Expression levels have been very variable ranging from nearly undetectable to as much as 40% of the total soluble protein (Tregoning et al., 2003; Zhou et al., 2008). However, so far only very few of the chloroplast-produced antigens have successfully been tested in oral vaccination experiments (Davoodi-Semiromi et al.; Tregoning et al., 2003).
Moreover, all studies conducted so far have been confined to small experimental animals (usually mice) and the concept of oral vaccination with transplastomic plants still awaits its validation in large mammals and ultimately in humans.

Extraordinarily high expression levels have recently been obtained with protein antibiotics derived from phage lytic proteins (so-called endolysins). These proteins efficiently digest the cell wall of pathogenic bacteria and, for this reason, have considerable potential as next-generation antibiotics suitable to control pathogens that have acquired resistance to most conventional antibiotics that are currently in clinical use. The chloroplast may be an ideal site for production of these protein antibiotics because (i) chloroplasts do not have a cell wall and, therefore, should be able to accommodate large amounts of these proteins and (ii) chloroplasts possess a very similar set of proteases as bacteria. The latter is important considering that phage endolysin proteins have evolved high levels of resistance to degradation by prokaryotic proteases. When expressed from the tobacco plastid genome, endolysins (targeted against pathogenic streptococci, including *Streptococcus pneumoniae*, the causal agent of pneumonia) indeed turned out to be extremely stable (Oey et al., 2009; Oey et al., 2009). Maximum protein accumulation reached unprecedented levels of up to more than 70% of the total soluble protein of the plant (Oey et al., 2009). Importantly, the plastid-produced endolysins were highly active and efficiently killed the target pathogens.

The attainable high expression levels also make transplastomic plants an attractive production platform for other high-value proteins, for example, industrial enzymes. With the growing interest in biofuels, enzymes and enzyme cocktails that are potentially suitable to digest lignocellulosic biomass into fermentable sugars have received particular attention. Several recent studies have demonstrated that the chloroplast compartment can accumulate very high levels of these enzymes (including endocellulases, exocellulases, β-glucosidases, xyloglucanases, pectate lyases and cutinases) (Yu et al., 2007; Verma et al., 2010; Kim et al., 2011), although extreme overaccumulation can cause mutant phenotypes (Petersen and Bock, 2011). For industrial enzymes expressed to high levels in plants, protein purification to homogeneity is often unnecessary and crude protein extracts can be used. Therefore, transplastomic plants expressing cell wall-degrading enzymes can potentially provide a cheap source of enzymes for the production of cellulosic ethanol. However, key issues in the conversion of lignocellulosic biomass into fuels remain to be addressed in order to make biofuels an economically viable
solution. First and foremost, the requirement for expensive (e. g., thermoacidic) pretreatments of the biomass needs to be bypassed. This will require efficient methods for lignin decomposition and subsequent removal or degradation of the released phenolic compounds.

**EXPRESSION OF RECOMBINANT PROTEINS IN CHLAMYDOMONAS CHLOROPLASTS**

A number of recent studies have explored the potential in molecular farming of the unicellular green alga *C. reinhardtii* (Coragliotti et al., 2011). Although microbial cultivation in bioreactors is considerably more expensive than growth of higher plants in soil, these extra costs are less of an issue in molecular farming compared to the food sector, especially if one considers that pharmaceutical proteins are usually high-value products and are needed in limited quantities. Moreover, production in fermenters offers significant advantages, especially by (i) providing a fully contained production facility, (ii) relying on established procedures in microbial biotechnology and (iii) potentially posing lower regulatory hurdles. Considerable progress has recently been made with expressing pharmaceutical proteins from the chloroplast genome of *Chlamydomonas*. In general, the capacity of the *Chlamydomonas* chloroplast to synthesize and/or accumulate foreign proteins appears to be much lower than that of higher plants. Unlike in flowering plants, codon optimization of transgene coding regions is essential (Franklin et al., 2002). Levels of protein expression are 2%-3% of total soluble cellular protein (Rasala et al.), although 10% and 20% protein expression levels were also obtained from the *psbA* expression signals in non-photosynthetic *psbA* knockout algae (Manuell et al., 2007; Surzycki et al., 2009). High-level recombinant protein expression in the *psbA* knockout plastids may be due to compensatory over-activation of transgene expression by *psbA*-specific trans-acting factor(s) and/or the lack of competition for these factors by the native mRNA. Incorporation of *psbA* at an ectopic location and expression from heterologous signals enabled photosynthetic competence and accumulation of the recombinant protein at somewhat reduced level (Manuell et al., 2007). Current efforts focus on the construction of *Chlamydomonas* strains in which limitations of protein expression have been removed by nuclear mutations (Neupert et al., 2009; Michelet et al., 2011).
FUTURE DIRECTIONS

Plastid engineering is currently pursued in three distinct systems: microalgae, bryophytes and flowering plants. Plastid transformation was first implemented in a microalga, *C. reinhardtii*. However, biotechnology-oriented research with *Chlamydomonas* was initiated relatively late, when recognition of the importance of codon modification enabled expression of recombinant proteins at commercially viable levels (Franklin et al., 2002). *Chlamydomonas* is likely to remain a useful expression vehicle for pharmaceutical proteins. At the same time, the hunt is on for new species that may be more suitable for biofuel production.

Plastid transformation in the bryophytes *Physcomitrella patens* (Sugiura and Sugita, 2004) and *Marchantia polymorpha* (Chiyoda et al., 2007) is relatively recent and has not yet been explored for the expression of recombinant proteins. However, by targeted nuclear gene replacements, *Physcomitrella* strains were created with non-immunogenic humanized glycan patterns and the moss is grown in a contained tissue culture system for recombinant protein production. These photo-bioreactors were proven to be superior to currently used mammalian cell lines in producing antibodies with enhanced effectiveness (Decker and Reski, 2007). Thus, a contained production system is already available. Plastid-based protein expression will further enhance the utility of the system.

A major task in flowering plants will be implementation of the technology in major crops, a requirement for agronomic applications. Implementation in new crops will depend on the availability of a genetic line that is suitable for repeated cycles of plant regeneration (to attain homoplasmy) and a suitable selectable marker gene, as discussed above. As far as industrial and pharmaceutical applications are concerned, in view of cost considerations and scalability, flowering plans will be the likely source of bulk industrial enzymes. The tobacco (*N. tabacum*) plastid expression system is already suitable to perform this role. However, it is not optimal for the production of edible vaccines due to its alkaloid content. From the available choices, the efficient lettuce system could serve as a vehicle for the expression of oral vaccines (Ruhlman et al., 2007). There is a need for robust, high-level regulated gene expression so that vegetative growth and the production phases can be separated and production of proteins induced prior to
harvest. Prototypes of such regulatory elements are metabolite-activated riboswitches (Verhounig et al., 2010) and inducible expression systems (Tungsuchat et al., 2006), which however currently still lack the robustness that is required for practical applications.

LITERATURE CITED


Coragliotti AT, Beligni MV, Franklin SE, Mayfield SP (2011) Molecular Factors Affecting the Accumulation of Recombinant Proteins in the *Chlamydomonas reinhardtii* Chloroplast. Mol Biotechnol in press


Lutz KA, Maliga P (2008) Plastid genomes in a regenerating tobacco shoot derive from a small number of copies selected through a stochastic process. Plant J 56: 975-983


kinetically equivalent source Rubiscos and can support tobacco growth. Plant Physiol 146: 83-96


FIGURE LEGENDS

**Figure 1.** Sorting of transformed ptDNA (T-ptDNA) at the organelle and cellular levels yield homoplasmic plants. A, Replication and sorting of T-ptDNA at the organelle level yields homoplasmic organelles. Sorting is facilitated by conversion of chloroplasts (CHL) to proplastids (PP), which contain only one to two nucleoids instead of ten. Wild type and transplastomic ptDNA (T-ptDNA) (blue circles and red circles, respectively) are anchored to membranes by proteins (black dots) in nucleoids (N). Sorting of ptDNA and T-ptDNA in heteroplasmic nucleoids (#1) yields nucleoids with only T-ptDNA (#1a) and wild-type ptDNA (#1b). For details see ref. (Maliga, 2004). B, Division and sorting of plastids yields genetically stable transplastomic plants. Sorting is accelerated by reduction from ~100 chloroplasts in leaf cells to ~10-14 proplastids in meristematic cells. In the cells the nucleus (Nu) is also marked. C, The plastid genotype of long-term stem cells in the three layers (L1, L2, L3) of the shoot apex determines the plastid genotype in leaves. PZ and CZ are the peripheral and central zones, respectively. On the left is a shoot apex with T-ptDNA in all three layers and a homoplasmic plant carrying only T-ptDNA encoding the aurea spectinomycin resistance \((aadA)\) gene. The variegated plant in the center has cells with wild type- and T-ptDNA in its shoot apex. The regenerated plant on the right has only wild-type ptDNA. Fig. 1C is modified from (Lutz and Maliga, 2008); the plants were described in (Tungsuchat-Huang et al., 2011).

**Figure 2.** Transplastomic clones are identified as green shoots in bombarded tobacco leaf culture on spectinomycin medium. The aurea \(aadA^{au}\) gene (Tungsuchat-Huang et al., 2011) enables greening and shoot regeneration in the culture shown here, and causes intense golden-yellow leaf pigmentation in plants (Fig. 1C).

**Figure 3.** Plastid genome manipulation is based on homologous recombination between ptDNA and the targeting regions in the vector. A, Replacement of the tobacco \(rbcL\) gene (\(T-rbcL\)) with the sunflower homologue (\(S-rbcL\)). The sunflower \(S-rbcL\) is incorporated in the tobacco ptDNA only if recombination is via the \(atpB\) and \(accD\) genes (dotted lines). Recombination adjacent to \(aadA\) (arrows) confers spectinomycin resistance, but the tobacco \(T-rbcL\) is retained. Out of six transplastomic lines three carried the \(aadA\) only; two incorporated \(S-rbcL\) and one had
recombination within rbcL (Kanevski et al., 1999). B, Insertion of the lux operon in trnI/trnA intergenic region. Note that the lux operon is transcribed from the aadA promoter and the gene cluster has only a single 3’-UTR (Krichevsky et al., 2010).

**Figure 4.** Excision of marker genes by site-specific recombinase enzymes. A, Marker genes in the plastid transformation vectors are flanked by loxP or attP/attB sequences (triangles) that are the targets for site-specific recombinases. B, The marker genes are efficiently removed when a gene encoding a plastid-targeted Cre or Int recombinase is introduced into the nucleus by transformation or pollination (Lutz and Maliga, 2007). TP1-ptDNA and TP2-ptDNA refer to the marker-containing and marker-free transplastomes.

**Figure 5.** Marker-free plastids by repeat-mediated deletion of the marker gene. In the transplastome (T-ptDNA), the aadA marker gene, expressed in the P2/T2 cassette, disrupts the hppd herbicide tolerance gene encoding 4-hydroxyphenylpyruvate dioxygenase (HPPD), an enzyme in the tocopherol biosynthetic pathway. The hppd coding region is flanked by the P1/T1 cassette, but is not expressed due to disruption by aadA. Note the 403-nt duplicated segment (darker colour) flanking the aadA. Deletion of aadA via the 403-nt repeats (arrows) reconstitutes a functional hppd gene enabling expression of herbicide resistance in seedlings (Dufourmantel et al., 2007). For a review see (Kode et al., 2006).

**Figure 6.** Transgene assembly in cassettes for protein expression. A, Shown are schematic maps of transgenes transcribed from a PL-UTR (Promoter-Leader-Untranslated Region; upper) and PL-TCR (Promoter-Leader-Translation Control Region; lower) cassette. BamHI (GGATCC), NcoI (CCATGG) and NheI (GCTAGC) restriction sites may be present in the same cassette (GGATTCCATGGCTAGC), while the NcoI and NdeI (CATATG) sites, each of which includes the translation initiation codon (ATG in bold), are incompatible. B, DNA sequence of the Prrn promoter with the atpB UTR (plasmid pJST12) (Tregoning et al., 2003)and TCR (plasmid pHK30) (Kuroda and Maliga, 2001). Note that expression in the PL-TCR cassette yields a fusion protein with 14 amino acids derived from the plastid atpB gene and two amino acids encoded in the NheI restriction site, whereas the PL-UTR transgene encodes an unmodified
protein. In bold are shown the Prrn promoter conserved -35 (TTGACG) and -10 (TATATT) promoter elements.

**Figure 7.** Metabolic engineering of the carotenoid pathway in transplastomic plants. A, Schematic overview of the carotenoid biosynthetic pathway. Lycopene, the major storage carotenoid in tomato, is indicated in red, β-carotene (provitamin A) in orange. Enzymes that have been used for plastid genome engineering are given in italics. Parts of the pathway not occurring in higher plants are shown in blue. Multiple arrows denote conversions involving multiple enzymatic steps. The reversible reactions of the xanthophyll cycle are indicated by double arrows. (B) Phenotypes of fruits from transplastomic tomato plants expressing a lycopene β-cyclase transgene from daffodil (*Narcissus pseudonarcissus*). Fruits from a wild-type plant (upper two panels), and a transplastomic line (lower two panels) were harvested at different ripening stages and photographed from the side and from the bottom. The orange color of the ripe transplastomic tomatoes comes from the efficient conversion of the red storage carotenoid lycopene into the orange provitamin A (β-carotene). The provitamin A levels reached 1 mg/g dry weight, while wild-type fruits have less than 200 ng provitamin A per g. Figure modified from ref. Apel and Bock, 2009.
A

DNA

Promoter + 5'-UTR

PL

ATG

Coding region

TAA

3'-UTR

EcoRI

SacII

BamHI

NdeI

Mcl

Nhel

XbaI

HindIII

DNA

Promoter + 5'-TCR

PL

ATG

Coding region

TAA

3'-UTR

EcoRI

SacII

BamHI

NdeI

Mcl

Nhel

XbaI

HindIII

B

Prnr promoter + atpB TCR in plasmid pHK30

SacI

1 gacgccGCTC CCCGCGCGTTC GTTCAATGAG AATGGATAAG AGGCTCGTGG
-35
-10

51 GATTGACGTG AGGCGGCCAGG GATGGCTATA TTTCTGGGAG AATTAACCGA

101 TCGACGTGCa AGCGGACATT TATTTTaaAT TCGATAATTT TTTGCAAAAAC

151 ATTTGACAT ATTTATTTAT TTTATATTAT TTTGAGATCA AATTAACCGA

201 TCTGGTTCTG GGGTTCCAC Ggctagc

S G S G V S T A S

NheI

Prnr promoter + atpB UTR in plasmid pJST12

SacI

1 gacgccGCTC CCCGCGCGTTC GTTCAATGAG AATGGATAAG AGGCTCGTGG
-35
-10

51 GATTGACGTG AGGCGGCCAGG GATGGCTATA TTTCTGGGAG AATTAACCGA

101 TCGACGTGCa AGCGGACATT TATTTTaaAT TCGATAATTT TTTGCAAAAAC

151 ATTTGACAT ATTTATTTAT TTTATATTAT TTTGAGATCA AATTAACCGA

NdeI