Progress towards commercialization of plastid transformation technology

Pal Maliga

Waksman Institute, Rutgers, The State University of New Jersey, 190 Frelinghuysen Road, Piscataway, NJ 08854-8020, USA

Tobacco chloroplasts are ready to be tested as a platform for the expression of recombinant proteins on a commercial scale. They hold the promise of reproducible yields of 5–25% of total soluble cellular protein in leaves and reliability has been achieved through refinement of an expression toolkit that includes vectors, recently developed expression cassettes and systems for marker gene removal. Implementation of plastid transformation technology in other crops, however, has met with difficulty and has delayed agronomic applications.

There is a pressing need for increasing the production capacity of recombinant proteins, including vaccines, antibodies and industrial enzymes. Protein expression in plants is advocated by many as the emerging answer [1–4]. In this review the case is made that tobacco chloroplasts are an attractive choice. There is also interest to extend plastid transformation to major agronomic crops for gene containment. Given that plastids are not transmitted by pollen in most crops, incorporation of transgenes in the plastid genome is a seductive concept. However, the technology is not yet available in the major agronomic crops. Here, recent advances in plastid transformation are reviewed and the bottlenecks associated with implementing plastid transformation in new crops are assessed. For additional information on plastid transformation and its applications see recent reviews [5–8].

The technology

The plastids of higher plants have their own genome – the plastome – which is between 120 and 180 kb in size. The plastid genetic system is highly polyploid – there can be as many as 1000 to 10 000 identical copies of the genome per cell. The challenge of plastome engineering is to uniformly alter all genome copies because genetically stable transgenic plants are obtained only if all the genome copies are identical. Plastid transformation in higher plants is based on DNA delivery by the biolistic process [9] or polyethylene glycol (PEG) treatment [10,11] and homologous recombination between the plastid-targeting sequences of the transformation vector and the targeted region of the plastid genome [9] (Fig. 1).

Biolistic delivery and PEG-induced DNA uptake were not the only approaches used in attempts to introduce transforming DNA in the plastome. The first attempt at plastid transformation used an *Agrobacterium* binary transformation vector [12]. Sophisticated nuclear targeting of the *Agrobacterium*-delivered DNA explains why attempts to transform plastids with *Agrobacterium* vectors so far have failed [13]. Another approach for the delivery of DNA into plastids is microinjection, which yields transient gene expression but not, as yet, stable plastid transformation [14]. Introduction of transforming DNA into isolated chloroplasts, followed by incorporation of the treated organelles into plants, has also been proposed [15,16] but has not yet been put into practice.

Integration of foreign DNA using homologous targeting was not the only approach tested for maintenance of transforming plastid DNA. Shuttle vectors for episomal maintenance in *Escherichia coli* and plastids have been designed as an alternative to stable integration. Vectors with a plastid 'origin of replication' (*ori*) sequence were episomally maintained in plastids if the plants were grown on a selective (spectinomycin) medium to which the shuttle plasmid marker gene confers resistance. These shuttle plasmids are not practical because they are rapidly lost in the absence of selection [17,18]. Shuttle vectors with a chloramphenicol resistance marker gene were also described [19] but have not yet been shown to transform chloroplasts.

Important discoveries that laid the foundation for higher plant plastid expression technology are listed in Table 1. Key discoveries in the unicellular alga *Chlamydomonas reinhardtii*, which influenced the development of plastid transformation in higher plants, are also listed.

Vectors

Initial transformation vectors carried a plastid 16S rRNA (*rrn16*) gene with point mutations that confer spectinomycin and streptomycin resistance [9]. The recessive *rrn16* marker genes were ~100-fold less efficient than the currently used *aadA* gene. The *aadA* gene encodes aminoglycoside 3′-adenylyltransferase, an enzyme that inactivates spectinomycin and streptomycin by adenylation [20]. Spectinomycin selection, on average, yields one transplastomic line in a bombarded leaf sample [20], although the values in this laboratory were found to vary between 0.5 and 5.0. Transplastomic clones are identified as shoots on spectinomycin medium, which suppress greening and shoot formation of wild-type tobacco cells (Fig. 2). The *neo* gene is an alternative marker for plastid transformation, which confers kanamycin resistance. The

Corresponding author: Pal Maliga (maliga@waksman.rutgers.edu).

 0167-7799/03/$ - see front matter PII: S0167-7799(02)00007-0

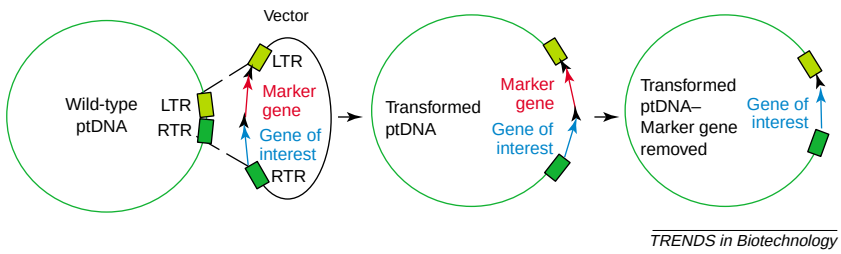

Fig. 1. Transformed plastid genome forms by two recombination events through homologous targeting sequences. Plastid genome segments included in vector are marked as left and right targeting regions (LTR,RTR), respectively. In future vectors, marker genes will be flanked by directly oriented *loxP* sites (filled triangles) for removal of marker genes by the CRE site-specific recombinase [53,54].

Table 1. Evolution of the plastid expression technology

Year	Milestone	Refs
1988	*Chlamydomonas reinhardtii*	
	1st stable plastid transformation	[64]
	DNA delivery: biolistic	
	Approach: homologous targeting	
	Selection: photosynthetic competence	
1990	*Nicotiana tabacum* (tobacco)	
	1st table plastid transformation	[9]
	DNA delivery: biolistic	
	Approach: homologous targeting	
	Selection: Spectinomycin–streptomycin resistance (*rrn16*)	
1991	*Chlamydomonas reinhardtii*	
	Selection: Spectinomycin–streptomycin resistance (*aadA*)	[65]
1993	*Nicotiana tabacum* (tobacco)	
	DNA delivery: PEG	[10,11]
	Selection: Spectinomycin–streptomycin resistance (*aadA*)	[20]
	Selection: Kanamycin resistance (*neo*)	[21]
	Protein expression: 1st high-level foreign protein; 2.5% GUS (bacterial)	[66]
1994	*Nicotiana tabacum* (tobacco)	
	Protein expression: nuclear-controlled plastid gene expression	[67]
1995	*Nicotiana tabacum* (tobacco)	
	Selection: editing-based marker gene	[68]
	Protein expression: polycistronic expression unit	[62]
	New agronomic trait: *Bacillus thuringiensis* insecticidal protein	[36]
	Marker gene elimination: co-transformation	[49]
1996	*Chlamydomonas reinhardtii*	
	Marker gene elimination: loop-out by homologous recombination and co-transformation	[50]
1998	*Arabidopsis thaliana*	
	1st stable plastid transformation	[56]
1999	*Solanum tuberosum* (potato)	
	1st stable plastid transformation	[55]
	Oryza sativa (rice)	
	1st stable plastid transformation	[59]
2000	*Nicotiana tabacum* (tobacco)	
	Protein expression: 1st human protein	[41]
	Marker gene elimination: loop-out by homologous recombination	[48]
2001	*Lycopersicon esculentum* (tomato)	
	Protein expression: 1st foreign protein expressed in fruit	[28]
	Nicotiana tabacum (tobacco)	
	Protein expression: photoautotrophic plant with heterologous RUBISCO	[32]
	Protein expression: Super-expression cassettes (>10%)	[22,23,31]
	Marker gene elimination: CRE-*lox* system	[53,54]
	New agronomic trait: glyphosate (herbicide) tolerance	[31]
	New agronomic trait: PPT (herbicide) resistance	[45]

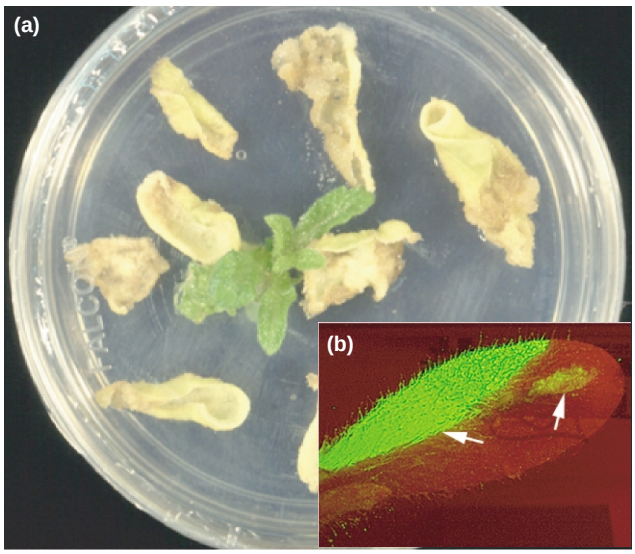

Fig. 2. Transplastomic clones are identified as green shoots on spectinomcyin medium. (a) Spectinomycin inhibits greening and shoot regeneration of wild type, but not of transplastomic, tobacco cells [9]. (b) The shoots are chimeric, visualized by the expression of green fluorescent protein in chloroplasts [59]. Genetically stable plants are obtained by shoot regeneration from the transformed sectors.

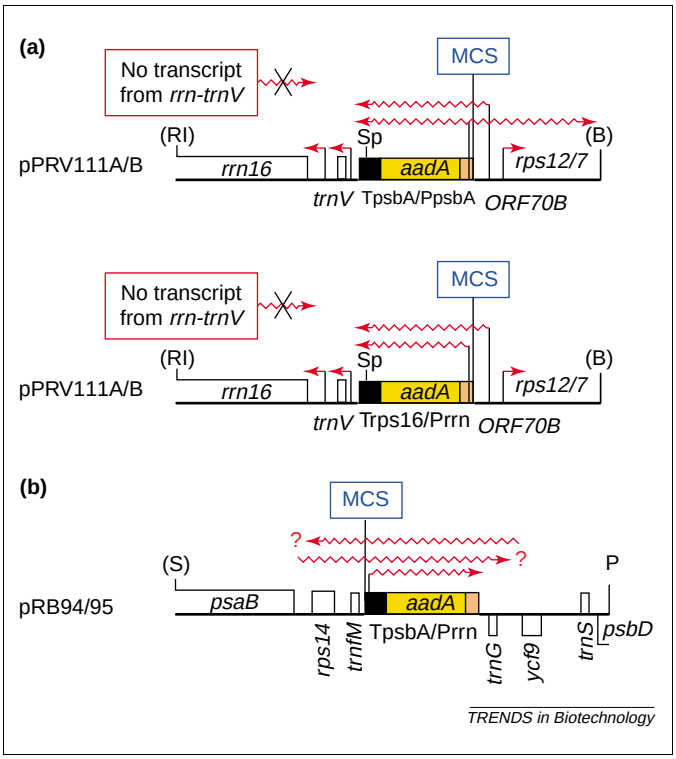

Fig. 3. Plastid transformation vectors. (a) The plastid repeat vector (pPRV) family (GenBank accession numbers U12809–U12815) [26]. Cloning convenience is ensured by alternative expression signals for *aadA* in the pPRV111 and pPRV112 series, and the pUC multiple cloning sites (MCSs) available in both orientations (A or B). There is no read-through transcription of transgenes in MCSs from the *rrn* operon. (b) The pRB94/pRB95 vectors (EMBL Accession numbers AJ312392, AJ312393) [28]. Vectors differ with respect to the orientation of the Bluescript MCS. No information is published on read-through transcription at the MCS.

first version of the *neo* gene was inefficient, yielding one transplastomic line per 25 bombarded samples [21]. The new generation of *neo* genes [22,23] is as efficient as *aadA* (author's unpublished results). A different kanamycin resistance gene (*aphA*-6) with a relatively high transformation efficiency (0.5 per bombarded leaf) was reported recently [24]. Another potentially useful marker is a plant nuclear gene encoding betaine aldehyde dehydrogenase, which confers resistance to the toxic compound betaine aldehyde (BA) [25]. BA selection is claimed to be 25-fold more efficient than spectinomycin selection [16]. Thus far there is no independent confirmation of the use of BA selection, nor is there a publication from the authors' laboratory reporting selection of transplastomic clones by resistance to BA alone. For additional information on plastid marker and reporter genes see Ref. [7].

Plastid vectors that are based on homologous targeting have left and right plastid-targeting regions, marked LTR and RTR in Fig. 1. Targeting regions have no special properties other than that they are homologous to the chosen target site and are at least 1–2 kb in size. Insertion of foreign DNA has been successful at 16 sites distributed in all regions of the plastid genome. Based on size variability among plastid genomes, incorporation of up to 50-kb of DNA (20–30 genes) is a realistic goal. Many different constructs have been used for plastid transformation [7] but there are only two plastid vector families with multiple cloning sites to expedite gene assembly from modular elements (Fig. 3). Both carry spectinomycin resistance (*aadA*) as a marker gene. The plastid repeat vector (pPRV) series targets insertions in the *trnV-rps12* intergenic region [26] (Fig. 3a). The pPRV112 vectors are suitable for expression of transcriptionally regulated passenger genes because there is no read-through transcription from the rRNA operon [26]. The divergent, weak read-through transcript derived from the *aadA* (*psbA*) promoter of the pPRV111 vectors is a useful reference transcript to study promoter activity [27]. Genes inserted in the *trnV-rps12* intergenic region are present in two copies because the insertion site is in the inverted repeat region of the tobacco plastid genome. Vectors pRB94 and pRB95 target insertions in the *trnfM-trnG* intergenic regions in the large unique region of the plastid genome (Fig. 3b) [28].

Several proteins have been expressed from promoterless constructs by transcriptional read-through by cloning blunt-ended coding region fragments (with a ribosome-binding site) into plasmid pZS197 at a unique restriction site between *aadA* and its 3′-end, TpsbA [29]. The *aadA* gene from plasmid pZS197 has been re-cloned to yield the 'universal vector' targeting insertions at the *trnA-trnI* intergenic region. Unique sites in the T-cassette of *aadA* are used for cloning as in plasmid pZS197 [29]. Plasmid pZS197 lacks a polycloning site because it was not intended to be used as a vector [20].

Engineering for high-level protein accumulation

Plastid transgenes are typically expressed in a 5′ PL- (PL, promoter and leader) and a 3′ T-cassette (T, terminator). The PL cassette includes a promoter and translation control sequences. The translation control sequences can be the mRNA 5′-untranslated region (UTR) (Fig. 4a), or the 5′-translation control

Table 2. Accumulation of recombinant proteins in chloroplasts

Protein	TSP concentration	Plasmid	Promoter[a]	5′-UTR[b]	N-terminal fusion	Coding region	3′-UTR[a]	Ref.
GUS	2.5%	pJS80	P*psbA*	*psbA*	–	*uidA*	*psbA*	[66]
NPTII	1.0%	pTNH32	Prrn	*rbcL*	*rbcL*; 5AA	*neo*	*psbA*	[21]
NPTII	2.5%	pHK31	Prrn	*atp*B	–	*neo*	*rbcL* (1)	[69]
NPTII	7.0%	pHK30	Prrn(1)	*atp*B	*atp*B; 14AA	*neo*	*rbcL* (1)	[22]
NPTII	4.0%	pHK60	Prrn	*atp*B	*atp*B; 14AA[c]	*neo*	*rbcL* (1)	[22]
NPTII	4.7%	pHK35	Prrn	*rbcL*	–	*neo*	*rbcL* (1)	[69]
NPTII	10.8%	pHK34	Prrn(2)	*rbcL*	*rbcL*; 14AA	*neo*	*rbcL* (1)	[22]
NPTII	0.31%	pHK64	Prrn	*rbcL*	*rbcL*; 14AA[c]	*neo*	*rbcL* (1)	[22]
NPTII	23.0%	pHK40	Prrn	*T7 g10*	–	*neo*	*rbcL* (1)	[23]
NPTII	16.4%	pHK38	Prrn	*T7 g10*	*T7 g10*;14AA	*neo*	*rbcL* (1)	[23]
NPTII	0.16%	pHK39	Prrn	*T7 g10*	synthetic	*neo*	*rbcL* (1)	[23]
TetC	25.0%	pJST10	Prrn	*T7 g10*	–	*tetC-AT*	*rbcL* (1)	[38]
TetC	10.0%	pJST11	Prrn	*atpB*	–	*tetC-GC*	*rbcL* (1)	[38]
TetC	10.0%	pJST12	Prrn	*T7 g10*	–	*tetC-AT*	*rbcL* (1)	[38]
EPSPS	0.001%	PMON30123	Prrn	*rbcL*	–	CP4bact	*rps16* (2)	[31]
EPSPS	0.002%	PMON30130	Prrn	*rbcL*	–	CP4synt	*rps16* (2)	[31]
EPSPS	0.2%	PMON38773	Prrn	*T7 g10*	–	CP4bact	*rps16* (2)	[31]
EPSPS	0.3%	PMON38798	Prrn	*T7 g10*	–	CP4synt	*rps16* (2)	[31]
EPSPS	>10.0%	PMON45259	Prrn	*T7 g10*	GFP; 14AA	CP4synt	*rps16* (2)	[31]
GFP	5.0%	PMON30125	Prrn(3)	*rbcL*	–	*gfp*	*rps16* (2)	[55]
GFP	5.0%	MR220	Prrn(3)	*rbcL*	–	*gfp*	*rps16* (2)	[46]
AAD-GFP	8.0%	pMSK56	Prrn(1)	*atp*B	*atp*B; 14AA	*aadA-gfp*	*psbA*	[59]
AAD-GFP	18.0%	pMSK57	Prrn(2)	*rbcL*	*rbcL*; 14AA	*aadA-gfp*	*psbA*	[59]
Ubiquitin-somato tropin	7.0%	PMON38794	Prrn	*T7 g10*	–		*rps16* (2)	[41]
PAT	>7.0%	pKO3	Prrn(1)	*atpB*	*atp*B; 14AA	*s-bar*	*rbcL* (1)	[45]
PAT	>7.0%	pKO18	Prrn(1)	*atpB*	*atp*B; 14AA	*b-bar2*	*rbcL* (1)	[45]
Cry1Ac	5.0%	pZS224	Prrn(3)	*rbcL*	–	*cry1A(c)*	*rps16* (2)	[36]
Cry2Aa2	3.0%	pZS-KM-cry2A		Bt	–	*cry2Aa2*	?	[70]
Cry2Aa2, 20-kDa; 29 kDa; CTB	45.3%	pLD-BD Cry2Aa2	2x-Prrn	Bt	–	*orf1-orf2-cry2Aa2*	*psbA*	[63]
	4.1%	pLD-LH-CTB	2x-Prrn	synthetic	–	*ctxB*	*psbA*	[71]

Extended, with permission, from Ref. [7].
[a]Arbitrary number identifies identical PL (promotor and leader) and T (terminator) cassettes.
[b]Same origin of 5′-UTR (untranslated region) does not mean identical 5′-UTR sequences.
[c]Alternative codons.

region (TCR) that includes the 5′-UTR and an N-terminal segment of the coding region (Fig. 4b). The mRNA 5′-UTR typically includes a stem-loop structure, which is required for mRNA stability, and sequences that facilitate loading of mRNAs onto ribosomes. The T-cassette encodes the mRNA 3′-UTR, which also includes a stem-loop structure. The 3′-UTR functions as an inefficient terminator of transcription and is required for mRNA stability [30]. Gene construction is facilitated by a gene assembly system that relies on having the same restriction sites at the boundaries of the PL and T cassettes and coding regions (Fig. 4).

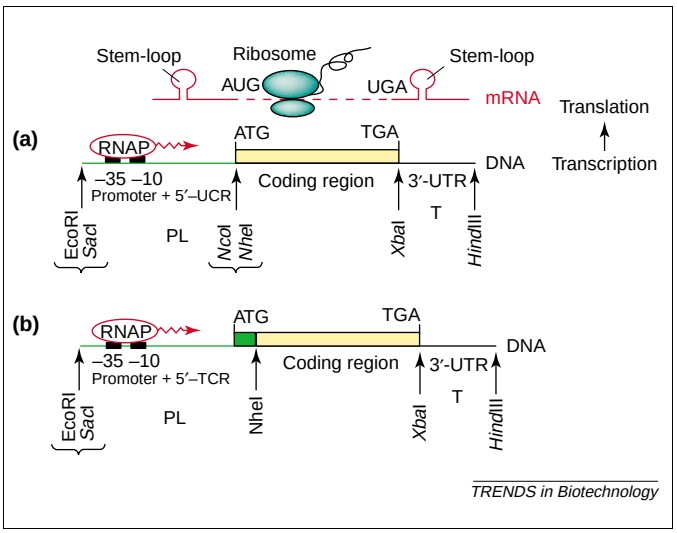

Fig. 4. Modular design of plastid promoter-leader (PL) and terminator (T) cassettes. PL cassettes encode a promoter and (a) 5′-untranslated region (5′-UTR) or (b) a 5′-translation control region (5′-TCR, 5′-UTR plus a segment of the coding region). Proteins are expressed from *Nhe*I/*Xba*I or *Nco*I/*Xba*I fragments [22,23]. Stem-loop structures formed by nucleotide pairing in the 5′-UTR and 3′-UTR and the AUG translation initiation and UGA stop codons are marked. (Reproduced, with permission, from Ref. [7]).

There are a few empiric rules that should be followed to obtain high-level protein accumulation. The first requirement is expression of the transgene from a strong promoter to ensure high levels of mRNA. Promoter choice is usually not a variable because most laboratories use the strong plastid rRNA operon (*rrn*) promoter (Prrn, Table 2). The second requirement is that the mRNA should be stabilized by a 3′-UTR. It appears that only modest improvements in protein accumulation can be obtained by choosing alternative 3′-UTRs. Thus, in most constructs the same T-cassettes are included. Targets for engineering are the 5′-UTR and coding region N-terminus because dramatic changes in protein accumulation have been obtained by choosing alternative 5′-UTRs.

An example for posttranscriptional regulation in plastids is accumulation of 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) in a 10 000-fold range from the same promoter [31] (Table 2). Levels of EPSPS were 100-fold higher when the *Agrobacterium* CP4 EPSPS was expressed from mRNA with the T7 phage gene 10 (T7 g10)

Box 1. Regulation of mRNA translation in plastids

Initiation is the rate-limiting step of plastid mRNA translation (Fig. I). Initial interaction of the small (30S) ribosomal subunit with the mRNA 5′-UTR, followed by recruitment of the large (50S) ribosomal subunit is shown. Many (40%) plastid mRNAs have ribosome binding site (RBS) complementary to the 16S rRNA [a] that can be GGAGG as in *Escherichia coli* [b], GGA [c]) or distinct, such as RBS1 (AAG) and RBS2 (UGAUGAU) [d]. The GGAGG sequence is between 12 and 7 nucleotides upstream of the AUG translation initiation codon. (a) Translation initiation from these mRNAs is dependent on interaction between the mRNA and the 16S rRNA 3′-end. Some (6%) of the plastid mRNAs have no GGAGG sequence or a variant [a]. (b) Translation initiation from these mRNAs is probably dependent on translational activating proteins (AP) that bind both the mRNA and the rRNA. In most (52%) mRNAs, the GGAGG sequence or a variant is further upstream than the prokaryotic consensus position [a]. Translation initiation from these mRNA is likely to depend on both mRNA-rRNA and mRNA-activating protein interactions. (c) The role for AP is to fold the mRNA to bring the distant RBS to the prokaryotic consensus position. The plastid S1 ribosomal protein binds to the RBS (GGAG) region of the *psbA* 5′-UTR [e], which interacts with ribosomes 33 nucleotides upstream of the AUG [b]. For reviews see Refs [f,g].

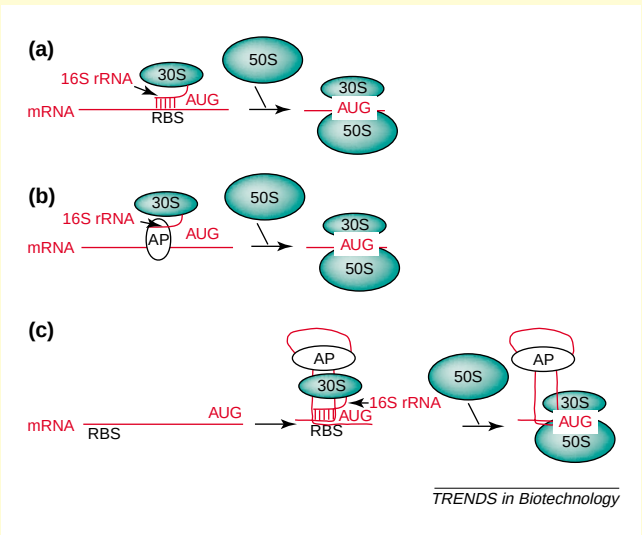

Fig. I. Models for translation initiation in plastids.

References

a Bonham-Smith, P.C. and Bourque, D.P. (1989) Translation of chloroplast-encoded mRNA: potential initiation and termination signals. *Nucleic Acids Res.* 17, 2057–2080
b Kim, J. and Mullet, J.E. (1994) Ribosome-binding sites on chloroplast rbcL and psbA mRNAs and light-induced initiation of D1 translation. *Plant Mol. Biol.* 25, 437–448
c Hirose, T. *et al.* (1998) Translation of tobacco chloroplast rps14 mRNA depends on a Shine-Dalgarno-like sequence in the 5′-untranslated region but not on internal RNA editing in the coding region. *FEBS Lett.* 430, 257–260
d Hirose, T. and Sugiura, M. (1996) *Cis*-acting elements and *trans*-acting factors for accurate translation of chloroplast psbA mRNAs: development of an *in vitro* translation system from tobacco chloroplasts. *EMBO J.* 15, 1687–1695
e Alexander, C. *et al.* (1998) Characterization of protein-binding to the spinach chloroplast psbA mRNA 5′ untranslated region. *Nucleic Acids Res.* 26, 2265–2272
f Barkan, A. and Goldschmidt-Clermont, M. (2000) Participation of nuclear genes in chloroplast gene expression. *Biochimie* 82, 559–572
g Somanchi, A. and Mayfield, S.P. (2001) Regulation of chloroplast translation. *Regulation of Photosynthesis Advances in Photynthesis and Respiration* (Vol. 11) (Aro, A.M., Anderson, B. eds), pp. 137–151, Kluwer Academic Publishers

leader instead of the *rbcL* leader. An additional >30-fold increase was obtained when the EPSPS N-terminus was translationally fused with 14 N-terminal amino acids of the green fluorescent protein (GFP) cumulatively yielding >10% of total soluble protein (TSP) in leaves. The translational fusion and choice of 5′-TCR probably affected translation rate and protein stability.

Systematic studies using *neo* as a reporter gene also confirmed the importance of posttranscriptional regulation in protein expression. The inclusion of 14 *rbcL* or *atpB* N-terminal amino acids with the cognate 5′-UTR increased NPTII levels from 4.7 to 10.8% (plasmids pHK35 and pHK34; Table 2) and from 2.5 to 7% (Plasmids pHK31 and pHK30; Table 2), respectively. Silent mutations downstream of the *rbcL* AUG codon in the PL cassette of plasmid pHK64 reduced translation efficiency 35-fold in the absence of a change in protein or mRNA stability [22]. Expression of NPTII from alternative T7 g10 5′-TCRs (Nt-pHK38 and Nt-pHK39) yielded a sevenfold difference in the mRNA level and a 100-fold difference in protein accumulation [23]. The highest level of NPTII obtained was >23% TSP [23].

The importance of the 5′-UTR in protein accumulation can be explained by translation initiation being the rate-limiting step. Translation initiation from plastid mRNAs is facilitated by mRNA–rRNA and mRNA–protein interactions (Box 1). The mRNA sequence encoding the N-terminus is not neutral because it might form a secondary structure with the 5′-UTR to facilitate or block translation initiation. Different proteins are encoded by different mRNAs and therefore the same protein might be well expressed in one cassette but poorly expressed in another (otherwise efficient) cassette. The pragmatic approach is to express the same protein in two or three cassettes with different types of translation control.

Expression of heterologous proteins

In this section discussion will be limited to expression of bacterial and human proteins in plastids (Table 2). Expression problems related to engineering Rubisco, the key enzyme of photosynthesis, and to attempts at engineering metabolic traits have been reviewed elsewhere [7,32].

When expressed in the plant nucleus, the *Bacillus thuringiensis* (Bt) insecticidal protein genes were unstable yielding many truncated polyadenylated mRNAs and only a small fraction of full-size transcript. The codon usage of the high AT insecticidal protein genes was shown to be different from the codon usage in highly expressed plant nuclear genes, which are more GC-rich (reviewed in Ref. [33]). Synthetic, more GC-rich codon-optimized Bt genes yielded stable mRNA and boosted protein accumulation from plant nuclear genes [34]. Expression of the unmodified bacterial Bt protoxin gene was also attempted in

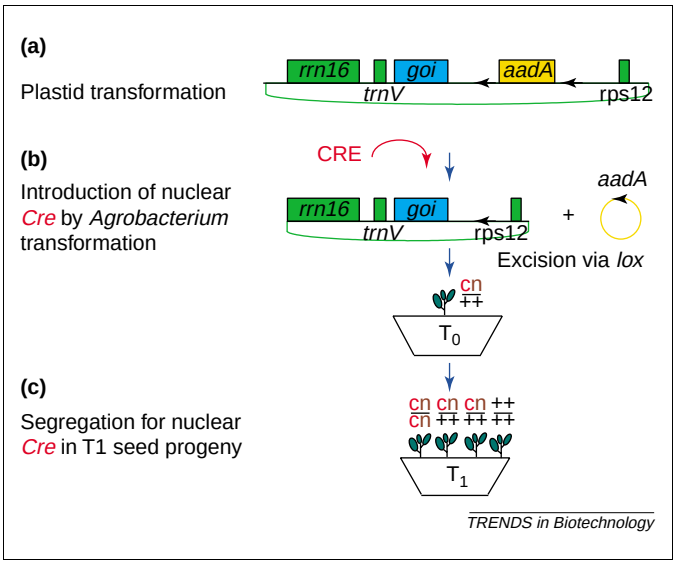

Fig. 5. Elimination of marker genes using the CRE-*lox* site specific recombination system. (a) The *aadA* marker gene in the plastid genome is flanked by two directly oriented *lox* sites (arrowheads). The gene of interest (*goi*) is shown in blue. (b) The nuclear-encoded, plastid targeted CRE is introduced by *Agrobacterium* transformation into the nucleus by selection for a linked kanamycin resistance (*neo*) gene. The *aadA* gene is simultaneously excised from all the plastid genome copies. (c) *Cre* and *Neo* are lost by segregation in the seed progeny. Abbreviations: c, *Cre*; n, *Neo*.

chloroplasts, which naturally have AT-rich genes, 72.1% in the coding region of genetic system genes and 69.1% in photosynthetic genes [35]. The mRNA expressed from the bacterial Bt coding segment in chloroplasts yielded a stable mRNA and the Bt protoxin accumulated 3–5% of TSP in leaves [36].

Expression of the *Clostridium tetani* toxin C-terminal fragment (TetC), a potential subunit vaccine, was problematic in yeast because the AT-rich *C. tetani* DNA yielded truncated mRNAs owing to the presence of several fortuitous polyadenylation sites. TetC accumulation was obtained when a codon optimized high-GC gene that lacked the polyadenylation sites was expressed in the yeast nucleus [37]. In tobacco plastids both the bacterial high-AT and synthetic high-GC mRNAs were stable. TetC from the bacterial gene accumulated to ~25% of TSP in leaves [38].

What the *B. thuringiensis* insecticidal protein gene and the *C. tetani* tetanus toxin gene have in common is relatively AT-rich DNA. Stability of AT-rich mRNAs in plastids is probably the consequence of conservation of a prokaryotic type RNA metabolism that is likely to interpret RNA signals as in bacteria. For example, a poly(A) tail stabilizes mRNAs in the cytoplasm but targets them for degradation in plastids as in *E. coli* [39,40]. Thus far, every bacterial gene expressed in plastids yielded stable mRNAs (Table 2), including the high-GC *bar* gene (68.3% GC), eliminating the need for synthetic genes. There is no information on the stability of human cDNAs in plastids because the only human protein expressed in chloroplasts, somatotropin, was expressed from a synthetic mRNA codon optimized for *E. coli* [41].

The second concern, when expressing a protein in a heterologous host, is codon optimization. Owing to the redundancy inherent in the genetic code, up to six nucleotide triplets can encode the same amino acid. Organisms with AT-rich genomes tend to favor alternative AT-rich codons reflecting tRNA abundance. The problem of expressing eukaryotic cDNAs with a large number of AGA or AGG codons, which are infrequently used in *E. coli* (1.4 or 2.1 per 1000 codons), is well understood [42,43]. In *E. coli*, incorporation of codons present up to 4.3 per 1000 codons could have a deleterious effect on translation efficiency and/or accuracy when present in clusters or in large numbers [43]. Although the tobacco plastid genome is relatively AT rich [44], there is no example for extremely low codon usage frequencies. Based on 19 cDNA sequences (4245 codons) in the National Center for Biotechnology Information database the least frequently used codons are UGC and CGC (3.5 per 1000 codons) followed by CCG, CGG and AGG (4.0, 4.2 and 4.9 per 1000 codons, respectively). Indeed, bacterial genes with a high GC content (*bar*, 68.3% GC; >7% TSP) could be expressed in plastids without codon modification [45]. Comparison of EPSPS synthase accumulation from the bacterial and synthetic CP4 genes indicates that codon modification yields only about a twofold increase in protein yield [31] (Table 2). TetC accumulation from the high-AT bacterial gene was also only ~2.5-fold higher (~25% of TSP) than the level obtained from the high-GC gene [38], whereas codon optimization of a gene encoding GFP did not yield a readily measurable gain [46]. The relatively modest (few-fold) increase obtained by codon modification in *Nicotiana tabacum*, a higher plant, contrasts the 80-fold increase obtained by codon optimization in the chloroplasts of *Chlamydomonas reinhardtii*, a unicellular alga [47].

Marker elimination system

Three factors provided the impetus to develop systems for the elimination of marker genes from the plastid genome: (1) the need to re-use the relatively few available plastid marker genes; (2) regulatory concerns to release antibiotic resistance genes in field crops; and (3) metabolic burden imposed by expression of marker genes. The first method developed for plastid marker gene removal relies on loop-out through directly repeated sequences [48]. This system is difficult to control because transformation and marker gene elimination occur simultaneously and the desired deletion derivative eventually sorts out in the seed progeny (discussed in Ref. [7]). Another approach involves co-transformation with two independently targeted plasmids and segregating out the genome copies with the marker gene while non-transformed copies are still present. The feasibility of co-transformation without physical linkage to a marker gene has been shown in tobacco plastids [49] but was used to recycle marker genes only in *Chlamydomonas* [50].

A more efficient approach benefits from the protocols developed for marker gene removal from the nucleus [51,52] and uses the P1 phage CRE-*loxP* site-specific recombination system [53,54]. According to the CRE-*loxP* scheme, the marker gene (flanked by two directly oriented *lox* sites) and the gene of interest are introduced into the plastid genome in the absence of CRE activity. When elimination of the marker gene is required, a gene encoding a plastid-targeted CRE site-specific recombinase

is introduced into the nucleus that, subsequent to its import in plastids, excises sequences between the *loxP* sites. *Cre* could be introduced by a second, *Agrobacterium*-mediated transformation [53,54] (Fig. 5) or by crossing [54]. The nuclear *Cre* is subsequently removed by segregation in the seed progeny. In tobacco, introduction of the nuclear *Cre* gene into the nucleus of transplastomic plants by *Agrobacterium* transformation extends the time needed to obtain marker-free plants by only one month. In an ideal case, it takes about six months to obtain a marker-free transplastomic tobacco plant that expresses a novel recombinant protein.

Implementation of the technology in new crops

Although the technology is >10 years old, plastid transformation has been achieved in relatively few species. Plastid transformation is a tissue-culture-dependent process because exposure of plastids (cells) to the selective agent and gradual replacement of plastid genome copies can be best accomplished in the cell culture environment. Sustained plant regeneration capability is an important characteristic because it allows selective elimination of wild-type plastid genome copies before plant regeneration. Plastid transformation has been most successful in solanaceous crops with similar tissue culture properties, including tobacco [9], potato [55] and tomato [28]. Plastid transformation in *Arabidopsis* was feasible but inefficient [56], probably owing to inefficient incorporation of the transforming DNA. The regenerated *Arabidopsis* plants were sterile – a consequence of plant regeneration from tissue culture. Plastid transformation was also described in *Brassica napus* [57] and *Lesquerella fendleri* [58], two species related to *Arabidopsis*. The transplastomic *Lesquerella* plants are fertile. In rice, transformation of plastid genomes in embryogenic cells could be obtained relatively easily. From the embryogenic cultures, plants were regenerated with chloroplasts expressing the GFP fusion protein. However, plant regeneration from cultured rice cells occurred before achieving the homoplastomic state and only a relatively small fraction of chloroplasts expressed the GFP fusion protein in the leaves [59]. Thus, it appears that plastid transformation in different taxonomic groups requires different approaches.

Conclusions and future directions

Cassettes promising high levels of protein expression and marker elimination systems are recent additions to the toolkit required for the production of recombinant proteins in tobacco chloroplasts. Tools developed in tobacco should be applicable to tomato without modification. In tomato, the fruits (chromoplasts) rather than leaves (chloroplasts) will be the targets for protein expression [28]. As to the potential of tomato fruits, it is encouraging that the marker gene product accumulates at ~0.5% TSP, which is about half the level accumulating in leaves. Using more efficient translation control signals might significantly increase protein levels in tomato fruits. Tissue-specific expression of proteins in tomato fruits without expressing them in leaves will be a subject of intense research to ensure protein accumulation in fruits without interference with the photosynthetic machinery of the plastid.

There are several plant-based protein production platforms that rely on protein expression from nuclear genes and viral vectors in leaf, seed, tuber and tissue culture cells [2–4]. Each of the platforms has strengths and weaknesses. Tobacco is emerging as the candidate for leaf-based protein production used by Large Scale Biology (formerly Biosource Technologies Inc., Vacaville, CA, USA) [60] and CropTech Corp [61]. The strength of the tobacco plastid system is speed: it takes about nine months starting with DNA to have enough seed for production on a large scale [20]. The plastid system is versatile because bacterial genes, and hopefully human cDNAs, can be directly expressed without re-synthesis and codon modification. Multi-subunit complex proteins can be expressed from polycistronic mRNAs [62,63] and human proteins can correctly fold and form disulfide bonds in plastids [41]. Expression of relatively large proteins is feasible in plastids indicated by the accumulation the 133 kDa Cry1A(c) protoxin at 3–5% TSP in leaves [36]. There is much to be learned about post-translational protein modification of heterologous proteins in plastids. It appears, however, that proteins in plastids are not glycosylated [38]. A further advantage is that transgenes incorporated in the plastid genome are contained efficiently because plastids in tobacco (as in most crops) are not transmitted by pollen.

The plastid-based expression platform in tobacco is ready to be tested. However, production platforms are only the first step towards plant-based manufacturing of therapeutics, diagnostics and technical enzymes. Commercialization of the technology now depends on the availability of production technologies, protein purification and quality control.

Acknowledgements

The National Science Foundation, the Rockefeller Foundation Rice Biotechnology Program, Monsanto Co., and a Rutgers FandA Special Project Grant supported research in author's laboratory.

References

1 Kusnadi, A. *et al.* (1997) Production of recombinant proteins in transgenic plants: practical considerations. *Biotechnol. Bioeng.* 56, 473–484
2 Ma, J.K. (2000) Genes, greens and vaccines. *Nat. Biotechnol.* 18, 1141–1142
3 Giddings, G. *et al.* (2000) Transgenic plants as factories for biopharmaceuticals. *Nat. Biotechnol.* 18, 1151–1155
4 Stoger, E. *et al.* (2002) Plantibodies: applications, advantages and bottlenecks. *Curr. Opin. Biotechnol.* 13, 161–166
5 Heifetz, P.B. and Tuttle, A.M. (2001) Protein expression in plastids. *Curr. Opin. Plant Biol.* 4, 157–161
6 Bock, R. (2001) Transgenic plastids in basic research and plant biotechnology. *J. Mol. Biol.* 312, 425–438
7 Maliga, P. (2002) Engineering the plastid genome of higher plants. *Curr. Opin. Plant Biol.* 5, 164–172
8 Staub, J.M. (2002) Expression of recombinant proteins via the plastid genome. *Handbook of Industrial Cell Culture: Mammalian, Microbial and Plant Cells* (Parekh, S.R., Vinci, V.A. eds), pp. 261–280, Humana Press
9 Svab, Z. *et al.* (1990) Stable transformation of plastids in higher plants. *Proc. Natl Acad. Sci. USA* 87, 8526–8530
10 Golds, T. *et al.* (1993) Stable plastid transformation in PEG-treated protoplasts of *Nicotiana tabacum*. *Biotechnology* 11, 95–97

11 O'Neill, C. *et al.* (1993) Chloroplast transformation in plants: polyetylene glycol (PEG) treatment of protoplasts is an alternative to biolistic delivery systems. *Plant J.* 3, 729–738
12 De Block, M. *et al.* (1985) Chloroplast transformation by *Agrobacterium tumefaciens*. *EMBO J.* 4, 1367–1372
13 Ward, D.V. *et al.* (2002) Agrobacterium VirE2 gets the VIP1 treatment in plant nuclear import. *Trends Plant Sci.* 7, 1–3
14 Knoblauch, M. *et al.* (1999) A galistan expansion femtosyringe for micrionjection of eukaryotic organelles and prokaryotes. *Nat. Biotechnol.* 17, 906–909
15 Daniell, H. and McFadden, B.A. (1987) Uptake and expression of bacterial and cyanobacterial genes by isolated cucumber etioplasts. *Proc. Natl Acad. Sci. USA* 84, 6349–6353
16 Daniell, H. *et al.* (2002) Milestones in chloroplast genetic engineering: an environmentally friendly era in biotechnology. *Trends Plant Sci.* 7, 84–91
17 Staub, J.M. and Maliga, P. (1994) Extrachromosomal elements in tobacco plastids. *Proc. Natl Acad. Sci. USA* 91, 7468–7472
18 Staub, J.M. and Maliga, P. (1995) Marker rescue from the Nicotiana tabacum plastid genome using a plastid *Escherichia coli* shuttle vector. *Mol. Gen. Genet.* 249, 37–42
19 Daniell, H. (1993) Foreign gene expression in chloroplasts of higher plants mediated by tungsten particle bombardment. *Methods Enzymol.* 217, 536–556
20 Svab, Z. and Maliga, P. (1993) High-frequency plastid transformation in tobacco by selection for a chimeric aadA gene. *Proc. Natl Acad. Sci. USA* 90, 913–917
21 Carrer, H. *et al.* (1993) Kanamycin resistance as a selectable marker for plastid transformation in tobacco. *Mol. Gen. Genet.* 241, 49–56
22 Kuroda, H. and Maliga, P. (2001) Sequences downstream of the translation initiation codon are important determinants of translation efficiency in chloroplasts. *Plant Physiol.* 125, 430–436
23 Kuroda, H. and Maliga, P. (2001) Complementarity of the 16S rRNA penultimate stem with sequences downstream of the AUG destabilizes the plastid mRNAs. *Nucleic Acids Res.* 29, 970–975
24 Hunag, F.C. *et al.* (2002) Efficient plastid transformation in tobacco using the *aphA-6* gene and kanamycin selection. *Mol. Genet. Genom.* 268, 19–27
25 Daniell, H. *et al.* (2001) Marker free transgenic plants: engineering the chloroplast genome without the use of antibiotic selection. *Curr. Genet.* 39, 109–116
26 Zoubenko, O.V. *et al.* (1994) Efficient targeting of foreign genes into the tobacco plastid genome. *Nucleic Acids Res.* 22, 3819–3824
27 Allison, L.A. and Maliga, P. (1995) Light-responsive and transcription-enhancing elements regulate the plastid *psbD* core promoter. *EMBO J.* 14, 3721–3730
28 Ruf, S. *et al.* (2001) Stable genetic transformation of tomato plastids: foreign protein expression in fruit. *Nat. Biotechnol.* 19, 870–875
29 Daniell, H. *et al.* (1998) Containment of herbicide resistance through genetic engineering of the chloroplast genome. *Nat. Biotechnol.* 16, 345–348
30 Monde, R.A. *et al.* (2000) The sequence and secondary structure of the 3′-UTR affect 3′-end maturation. RNA accumulation, and translation in tobacco chloroplasts. *Plant Mol. Biol.* 44, 529–542
31 Ye, G.N. *et al.* (2001) Plastid-expressed 5-enolpyruvylshikimate-3-phosphate synthase genes provide high level glyphosate tolerance in tobacco. *Plant J.* 25, 261–270
32 Whitney, S.M. and Andrews, T.J. (2001) Plastome-encoded bacterial ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) supports photsynthesis and growth of tobacco. *Proc. Natl Acad. Sci. USA* 98, 14738–14743
33 Murray, E.E. *et al.* (1991) Analysis of unstable RNA transcripts of insecticidal crystal protein genes of *Bacillus thuringiensis* in transgenic plants and electroporated protoplasts. *Plant Mol. Biol.* 16, 1035–1050
34 Perlak, F.J. *et al.* (1991) Modification of the coding sequence enhances plant expression of insect control protein genes. *Proc. Natl Acad. Sci. USA* 88, 3324–3328
35 Shimada, H. and Sugiura, M. (1991) Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. *Nucleic Acids Res.* 19, 983–995
36 McBride, K.E. *et al.* (1995) Amplification of a chimeric *Bacillus* gene in chloroplasts leads to an extraordinary level of an insecticidal protein in tobacco. *Biotechnology* 13, 362–365
37 Romanos, M.A. *et al.* (1991) Expression of tetanus toxin fragment C in yeast: gene synthesis is required to eliminate fortuitous polyadenylation sites in AT-rich DNA. *Nucleic Acids Res.* 19, 1461–1467
38 Maliga, P. *et al.* (2003) Tobacco chloroplasts as a platform for vaccine production. In *10th IAPTC&B Congress, Plant Biotechnology 2002 and Beyond* (Vasil, I.K. *et al.*, eds), Kluwer Academic Publishers, in press
39 Komine, Y. *et al.* (2002) Evidence for *in vivo* modulation of chloroplast RNA stability by 3′-UTR homopolymeric tails in *Chlamydomonas reinhardtii*. *Proc. Natl Acad. Sci. USA* 99, 4085–4090
40 Hayes, R. *et al.* (1999) Degrading chloroplast mRNA: the role of polyadenylation. *Trends Biochem. Sci.* 24, 199–202
41 Staub, J.M. *et al.* (2000) High-yield production of a human therapeutic protein in tobacco chloroplasts. *Nat. Biotechnol.* 18, 333–338
42 Makrides, S.C. (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* 60, 512–538
43 Kane, J.F. (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* 6, 494–500
44 Shinozaki, K. *et al.* (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J.* 5, 2043–2049
45 Lutz, K.A. *et al.* (2001) Expression of bar in the plastid genome confers herbicide resistance. *Plant Physiol.* 125, 1585–1590
46 Reed, M.L. *et al.* (2001) High-level expression of a synthetic red-shifted GFP coding region incorporated into the chloroplasts. *Plant J.* 27, 257–265
47 Franklin, S. *et al.* (2002) Development of a GFP reporter gene for *Chlamydomonas reinhardtii* chloroplast. *Plant J.* 30, 733–744
48 Iamtham, S. and Day, A. (2000) Removal of antibiotic resistance genes from transgenic tobacco plastids. *Nat. Biotechnol.* 18, 1172–1176
49 Carrer, H. and Maliga, P. (1995) Targeted insertion of foreign genes into the tobacco plastid genome without physical linkage to the selectable marker gene. *Biotechnology* 13, 791–794
50 Fischer, N. *et al.* (1996) Selectable marker recycling in the chloroplast. *Mol. Gen. Genet.* 251, 373–380
51 Ow, D.W. (2002) Recombinase-directed plant transformation for the post-genomic era. *Plant Mol. Biol.* 48, 183–200
52 Hare, P.D. and Chua, N.H. (2002) Excision of selectable marker genes from transgenic plants. *Nat. Biotechnol.* 20, 575–580
53 Hajdukiewicz, P.T.J. *et al.* (2001) Multiple pathways for Cre/*lox*-mediated recombination in plastids. *Plant J.* 27, 161–170
54 Corneille, S. *et al.* (2001) Efficient elimination of selectable marker genes from the plastid genome by the CRE-*lox* site-specific recombination system. *Plant J.* 27, 171–178
55 Sidorov, V.A. *et al.* (1999) Stable chloroplast transformation in potato: use of green fluorescent protein as a plastid marker. *Plant J.* 19, 209–216
56 Sikdar, S.R. *et al.* (1998) Plastid transformation in *Arabidopsis thaliana*. *Plant Cell Rep.* 18, 20–24
57 Hou, B.K. *et al.* (2002) Chloroplast transformation in oilseed rape. *Transgenic Res.* in press
58 Skarjinskaia, M. *et al.* (2002) Plastid transformation in *Lesquerella fendleri*, an oilseed Brassicacea. *Transgenic Res.* in press
59 Khan, M.S. and Maliga, P. (1999) Fluorescent antibiotic resistance marker to track plastid transformation in higher plants. *Nat. Biotechnol.* 17, 910–915
60 McCormick, A.A. *et al.* (1999) Rapid production of specific vaccines for lymphoma by expression of the tumor-derived single-chain Fv epitopes in tobacco plants. *Proc. Natl Acad. Sci. USA* 96, 703–708
61 Cramer, C.L. *et al.* (1999) Transgenic plants for therapeutic proteins: linking upstream and downstream strategies. *Curr. Top. Microbiol. Immunol.* 240, 95–118
62 Staub, J.M. and Maliga, P. (1995) Expression of a chimeric uidA gene indicates that polycistronic mRNAs are efficiently translated in tobacco plastids. *Plant J.* 7, 845–848
63 De Cosa, B. *et al.* (2001) Overexpression of the Bt cry2Aa2 operon in chloroplasts leads to formation of insecticidal crystals. *Nat. Biotechnol.* 19, 71–74
64 Boynton, J.E. *et al.* (1988) Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles. *Science* 240, 1534–1538

65 Goldschmidt-Clermont, M. (1991) Transgenic expression of aminoglycoside adenine transferase in the chloroplast: a selectable marker of site-directed transformation of chlamydomonas. *Nucleic Acids Res.* 19, 4083–4089
66 Staub, J.M. and Maliga, P. (1993) Accumulation of D1 polypeptide in tobacco plastids is regulated via the untranslated region of the psbA mRNA. *EMBO J.* 12, 601–606
67 McBride, K.E. *et al.* (1994) Controlled expression of plastid transgenes in plants based on a nuclear DNA-encoded and plastid-targeted T7 RNA polymerase. *Proc. Natl Acad. Sci. USA* 91, 7301–7305
68 Chaudhuri, S. *et al.* (1995) Site-specific factor involved in the editing of the psbL mRNA in tobacco plastids. *EMBO J.* 14, 2951–2957
69 Maliga, P. (2002) Chloroplasts for the production of recombinant proteins. *PS2001 Proceedings: 12th International Congress on Photosynthesis, Melbourne, Australia, 2001*, CSIRO Publishing, ISBN 0643067116; www.publish.csiro.au/ps2001
70 Kota, M. *et al.* (1999) Overexpression of the *Bacillus thuringiensis* (Bt) Cry2Aa2 protein in chloroplasts confers resistance to plants against susceptible and Bt-resistant insects. *Proc. Natl Acad. Sci. USA* 96, 1840–1845
71 Daniell, H. *et al.* (2001) Expression of the native cholera toxin B subunit gene and assembly of functional oligomers in transgenic tobacco chloroplasts. *J. Mol. Biol.* 311, 1001–1009