S40-001

Chloroplasts for the production of recombinant proteins

P Maliga[1,2], H Kuroda[1], S Corneille[1], K Lutz[1], Z Svab[1]

[1]*Waksman Institute, Rutgers University, 190 Frelinghuysen Road, Piscataway, NJ, 08854-8020, USA. maliga@waksman.rutgers.edu*

[2]*Department of Plant Science, Rutgers University, 59 Dudley Road, New Brunswick, NJ, 08901, USA*

Keywords: chloroplast, marker gene elimination, plastid transformation, protein expression, translation

Introduction

There is a significant interest in using transgenic plants as factories for biopharmaceuticals and industrial enzymes (Giddings et al., 2000; Ma, 2000; Walmsley and Arntzen, 2000). High-level protein accumulation is important to keep production costs low and for efficacy in some applications, such as oral vaccines, which involve the direct use of plant tissues. Protein expression from plastid transgenes yields consistently high protein levels making plastid transformation an attractive alternative to expression from nuclear genes. Plastid transformation involves implantation of foreign genes in the organelle's plasmid-like double-stranded circular DNA genome which may be present in up to 10,000 copies per cell. We describe here a system for protein expression in tobacco chloroplasts which includes transformation vectors, expression cassettes and a system for marker gene elimination. Other applications of plastid transformation have been covered in several recent reviews (Bock, 2001; Bogorad, 2000; Heifetz and Tuttle, 2001; van Bel et al., 2001).

Plastid transformation vectors

Plastid transformation involves insertion of foreign genes by two homologous recombination events *via* two flanking "targeting" sequences (Fig. 1). Introduction of the transforming DNA may be obtained by the biolistic process (Svab et al., 1990) or by polyethylene glycol treatment (Golds et al., 1993; Koop et al., 1996; O'Neill et al., 1993). Incorporation of the transgenes into the plastid genome is only the first step in obtaining a genetically stable plant as it takes several cell generations to dilute out all wild-type plastid genome copies.

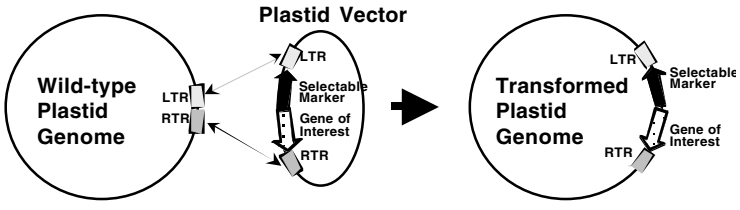

Fig. 1. Homologous recombination between the plastid genome and plastid transformation vector (arrows; left) and the transformed plastid genome. LTR and RTR are the left and right targeting regions.

Transformed plastid genome copies are amplified by selection for spectinomycin resistance (*aadA,* encoding aminoglycoside 3"-adenyltransferase) (Svab and Maliga, 1993) or, less efficiently, by selection for kanamycin resistance (*neo* or *kan*, encoding neomycin phosphotransferase or NPTII) (Carrer et al., 1993).

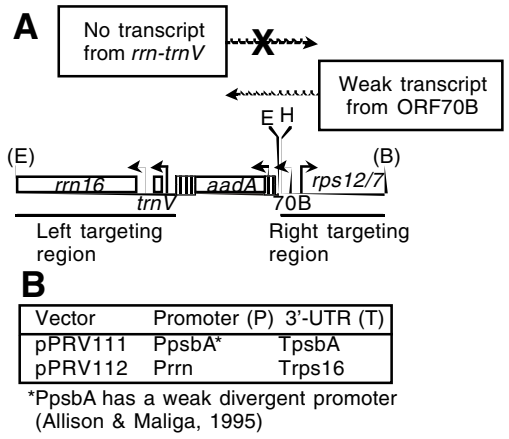

Fig. 2. Plastid transformation vectors. (A) Map of vectors pPRV111 and pPRV112 (Zoubenko et al., 1994). Marked are the spectinomycin resistance (*aadA*) gene, the *Eco*RI((E) and *Hin*dIII (H) sites of the pUC119 polycloning site, plastid genes *rrn16*, *trnV* and transcription initiation sites (horizontal arrows). (B) Expression signals of the *aadA* genes in the vectors.

Transgenes have been inserted at numerous sites in the plastid genome. For protein expression we have developed a family of plastid vectors targeting insertions in the *trnV-rps12/7* intergenic region in the repeated region of the plastid genome (pPRV=plastid repeat vector) (Zoubenko et al., 1994). To facilitate cloning, unique restriction sites have been inserted adjacent to the selective marker. During transformation the targeted region of the plastid genome is replaced by the targeting region of the vector and the vector backbone (pUC119) is lost. The *aadA* gene in the pPRV111 and pPRV112 vectors is expressed from different signals to allow the choice of different promoters and terminators for the marker gene (*aadA)* and the linked gene of interest (Fig. 2). The pPRV vectors on average yield one transformation event per bombarded leaf sample. There is no readthrough transcription at the insertion site from the rRNA operon, but there is low-level transcription from the ORF70B promoter adjacent to the *rps12/7* operon (Fig. 2)(Zoubenko et al., 1994). If expression of a transgene by readthrough transcription is undesirable, it should be oriented divergently relative to the rRNA operon.

Cassettes for high-level protein expression

Plastid gene expression is regulated both at the transcriptional and posttranscriptional levels. Protein levels in chloroplasts depend on mRNA abundance, which is determined by promoter strength and mRNA stability. However, high mRNA levels are no assurance of high-level protein accumulation as post-transcriptional processes ultimately determine obtainable protein levels. The modules that we have designed for gene assembly are as follows: (i) Promoter and the 5' untranslated region (5'-UTR) or the promoter and the 5' translation control region (5'-TCR, including 5'-UTR and coding region N-terminus); (ii) the coding region, and (iii) the 3'-untranslated region (3'-UTR). Suitable restriction sites flanking the modules facilitate gene assembly in a pUC119 plasmid polycloning site (Fig. 3).

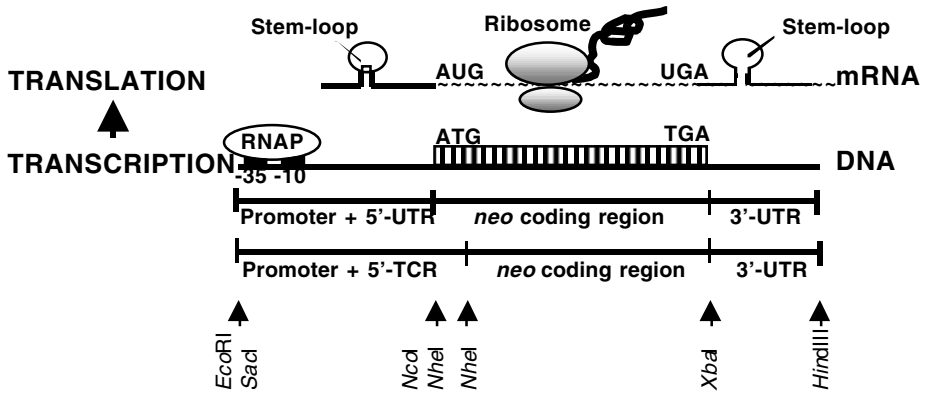

Fig. 3. DNA and RNA signals for plastid gene expression.

A large number of plastid promoters have been identified that are recognized by the *E. coli* like, plastid-encoded RNA polymerase (PEP) or the phage type, nuclear-encoded plastid RNA polymerase (NEP) (Liere and Maliga, 2001). For protein expression, we have developed derivatives of one of the strongest PEP promoters, which drives the expression of the plastid rRNA operon (*rrn*). The products of the *rrn* operon are ribosomal RNA subunits and not protein. To ensure translation of Prrn-derived transcripts, the promoter was fused with sequences which are part of the 5'-UTR of highly expressed plastid and phage genes. The 5'-UTR typically forms a stabilizing stem-loop structure and is the target of RNA binding proteins which regulate translation and mRNA turnover (Barkan and Goldschmidt-Clermont, 2000) (Alexander et al., 1998). The early Prrn derivatives contain 34 bp of the *rrn* 5' transcribed region and 18 nucleotides of the *rbcL* 5'UTR including the GGGAGGG ribosome binding site (pZS197; Svab and Maliga, 1993). Expression of transgenes from this promoter results in accumulation of 1% of the total soluble cellular protein (TSP) as NPTII (Carrer et al., 1993) and 3-5% Bt protoxin (McBride et al., 1995).

In the new generation of Prrn derivatives the promoter is directly linked up with the 5'-UTR of plastid genes at the transcription start sites (Kuroda and Maliga, 2001a; Kuroda and Maliga, 2001b). In one set of constructs Prrn was fused with sequences derived from plastid genes: either the processed form of the *rbcL* 5'-UTR (59 nt of the 182 nt 5'-UTR) or the processed form of the *atpB* 5'-UTR (90 nt). For both Prrn derivatives two constructs were prepared: one that has only the processed 5'-UTR and one with the 5'-TCR. The 5'-TCR constructs include the processed 5'-UTR and 42 nucleotides downstream of the AUG. Comparison of plants transformed with the 5'-UTR constructs and plants transformed with their 5'-TCR counterparts revealed consistently improved NPTII accumulation if sequences downstream of the AUG were included (Table 1). Silent mutations downstream of the AUG significantly reduced protein accumulation from the *rbcL* (35-fold reduction), but not from the *atpB* construct (Kuroda and Maliga, 2001b). Thus, introduction of silent mutations downstream of the AUG may be explored to enhance protein accumulation from poorly-expressed mRNAs.

A further source of translation control signals was the T7 phage gene 10 (T7g10), which was shown to promote high-level protein accumulation in bacteria (Studier et al., 1990).

Table 1. NPTII accumulation in tobacco leaves from Prrn promoter derivatives with plastid and phage 5'-UTR and TCR sequences.

Vector	5'-UTR or 5'-TCR	% TSP in leaves	% mRNA
pHK30	atpB/TCR	7.02±0.82	70.05±12.33
pHK31	atpB/UTR	2.52±0.79	100
pHK34	rbcL/TCR	10.83+3.84	48.91+22.65
pHK35	rbcL/UTR	4.68+1.84	21.41+7.88
pHK38	T7g10/TCR	16.39+3.42	47.59+19.06
pHK40	T7g10/UTR	23.00+5.40	90.27+31.83

Again, two constructs were prepared one with the T7g10 5'-UTR and one with the 5'-TCR. Interestingly, this time the 5'-UTR construct was more efficient yielding 23% NPTII as compared to 16% NPTII obtained from the 5'-TCR construct (Table 1). In *E. coli*, enhancing the complementarity between sequences downstream of the AUG and the 16S rRNA penultimate stem resulted in increased protein accumulation without a significant affect on mRNA stability. Unlike in *E. coli*, increasing the complementarity downstream of the T7g10 AUG with the plastid 16S rRNA penultimate stem dramatically (100-fold) reduced NPTII accumulation (0.16%) and destabilized the mRNA, pinpointing differences between the prokaryotic and plastid expression systems (Kuroda and Maliga, 2001a).

Transcription of the *rrn* operon is light regulated (Shiina et al., 1998). However, the Prrn-derived transcript of early constructs accumulates to comparable levels in the light and in the dark suggesting that the mRNA contains a stability element that compensates for the reduced rates of transcription in the dark (Allison and Maliga, 1995). Proteins expressed from the Prrn derivatives were shown to accumulate in all plastid types, including leaf chloroplasts, non-green plastids in roots and the chromoplasts in flower petals, reflecting the need for rRNA expression in the plastids of all cell types (Khan and Maliga, 1999).

The role of the 3'-UTR is to protect the plastid mRNAs from degradation. The 3'-UTR contains a stem-loop structure that is the binding site for proteins regulating mRNA processing and turn-over (Hayes et al., 1999). We routinely use the *psbA*, *rbcL* and *rps16* 3'-UTRs in our constructs.

The CRE-lox system for efficient elimination of the selectable marker gene

The protein levels obtained from the selectable marker genes are in the range of 1% to 10% of TSP, a significant metabolic burden on the plants. Furthermore, it may be difficult to obtain regulatory approval to release transplastomic crops carrying the spectinomycin resistance (*aadA*) gene. Fortunately, it is not necessary to have antibiotic resistance genes in transplastomic crops. One approach is to search for an alternative marker system. A useful marker not involving antibiotics could be selection for resistance to betaine aldehyde based on plastid expression of the plant nuclear betaine aldehyde dehydrogenase (BADH) gene, as long as it is widely applicable (Daniell et al., 2001).

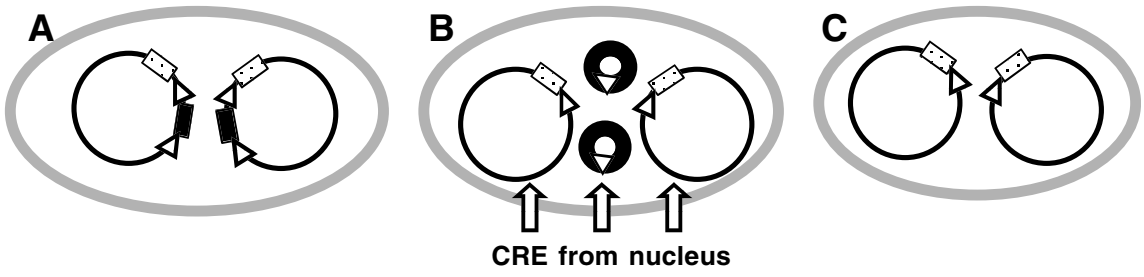

Fig. 4. CRE-lox site-specific recombinase for elimination of marker genes in plastids. (A) Chloroplast with transformed plastid genome. Marker gene (black box) flanked by *lox* sites (triangles) and the gene of interest (open box). (B) The *Cre* gene is introduced into the nucleus. The plastid-targeted CRE is imported into plastids and simultaneously excises marker genes from all genome copies. Note circularized marker gene and modified plastid genome with one *lox* site. (C) Marker-free plastid genomes with one *lox* site and the gene of interest. The marker gene is rapidly lost in somatic cells. The nuclear *Cre* gene segregates out in the seed progeny.

A better alternative is to remove the selective marker altogether once it has fulfilled its useful purpose. The scheme developed by Iamtham and Day (Iamtham and Day, 2000) exploits formation of multiple recombination products in the same plastid. Fortuitous deletion derivatives in which the marker gene is removed are obtained by homologous recombination via short (in the specific case 174 bp) direct repeats and are identified in the somatic progeny of the initially transformed cell by phenotype or by brute-force screening. The problem with such an approach is that transformation and marker gene elimination are occurring simultaneously. Transformation and marker gene elimination are separate processes in the CRE-*lox* plastid marker elimination system developed independently by Hajdukiewicz et al.(Hajdukiewicz et al., 2001) and Corneille et al.(Corneille et al., 2001) (Fig. 4). According to the CRE-*lox* 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 *lox* sites. *Cre* may be introduced by a second, *Agrobacterium*-mediated transformation or by crossing. The nuclear *Cre* is subsequently removed by segregation in the seed progeny. In tobacco, introduction of the nuclear *Cre* gene by *Agrobacterium* transformation extends the time needed to obtain marker-free transplastomic plants only by about a month. Thus, it takes a total of four to five months to obtain a marker-free transplastomic tobacco plant expressing a novel recombinant protein.

Acknowledgements

This research was supported by NSF Grants MCB 96-308763, MCB 99-05043 and Rutgers F&A Special Project Grant to PM.

References

Alexander, C., Faber, N. and Klaff, P. (1998) *Nucleic Acids Res.*, **26**, 2265-2272.
Allison, L.A. and Maliga, P. (1995) *EMBO J.*, **14**, 3721-3730.
Barkan, A. and Goldschmidt-Clermont, M. (2000) *Biochimie*, **82**, 559-572.
Bock, R. (2001) *J. Mol. Biol.*, in press.
Bogorad, L. (2000) *Trends Biotech.*, **18**, 257-263.
Carrer, H., Hockenberry, T.N., Svab, Z. and Maliga, P. (1993) *Mol. Gen. Genet.*, **241**, 49-56.

Corneille, S., Lutz, K., Svab, Z. and Maliga, P. (2001) *Plant J.*, **72**, 171-178.
Daniell, H., Muthukumar, B. and Lee, S.B. (2001) *Curr. Genet.*, **39**, 109-116.
Giddings, G., Allison, G., Brooks, D. and Carter, A. (2000) *Nat. Biotechnol.*, **18**, 1151-1155.
Golds, T., Maliga, P. and Koop, H.U. (1993) *Biotechnology*, **11**, 95-97.
Hajdukiewicz, P.T.J., Gilbertson, L. and Staub, J.M. (2001) *Plant J.*, **27**, 161-170.
Hayes, R., Kudla, J. and Gruissem, W. (1999) *Trends Biochem. Sci.*, **24**, 199-202.
Heifetz, P.B. and Tuttle, A.M. (2001) *Curr. Opin. Plant Biol.*, **4**, 157-161.
Iamtham, S. and Day, A. (2000) *Nat. Biotechnol.*, **18**, 1172-1176.
Khan, M.S. and Maliga, P. (1999) *Nat. Biotechnol.*, **17**, 910-915.
Koop, H.U., Steinmüller, K., Wagner, H., Rössler, C., Eibl, C. and Sacher, L. (1996) *Planta*, **199**, 193-201.
Kuroda, H. and Maliga, P. (2001a) *Nucleic Acids Res.*, **29**, 970-975.
Kuroda, H. and Maliga, P. (2001b) *Plant Physiol.*, **125**, 430-436.
Liere, K. and Maliga, P. (2001) Plastid RNA polymerases in higher plants. In Anderson, B. and Aro, E.M. (eds.), *Regulation of Photosynthesis.* Kluwer Academic Publishers, Dordrecht, pp. 29-49.
Ma, J.K. (2000) *Nat. Biotechnol.*, **18**, 1141-1142.
McBride, K.E., Svab, Z., Schaaf, D.J., Hogan, P.S., Stalker, D.M. and Maliga, P. (1995) *Biotechnology*, **13**, 362-365.
O'Neill, C., Horvath, G.V., Horvath, E., Dix, P.J. and Medgyesy, P. (1993) *Plant J.*, **3**, 729-738.
Shiina, T., Allison, L. and Maliga, P. (1998) *Plant Cell*, **10**, 1713-1722.
Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) *Meth. Enzymol.*, **185**, 60-89.
Svab, Z., Hajdukiewicz, P. and Maliga, P. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 8526-8530.
Svab, Z. and Maliga, P. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 913-917.
van Bel, A.J., Hibberd, J., Prufer, D. and Knoblauch, M. (2001) *Curr. Opin. Biotechnol.*, **12**, 144-149.
Walmsley, A.M. and Arntzen, C.J. (2000) *Curr. Opin. Biotechnol.*, **11**, 126-129.
Zoubenko, O.V., Allison, L.A., Svab, Z. and Maliga, P. (1994) *Nucleic Acids Res.*, **22**, 3819-3824.