12

Self-Processing Polyproteins: A Strategy for Co-expression of Multiple Proteins in Plants

GARRY A. LUKE[1*], PABLO DE FELIPE[1], VANESSA M. COWTON[1], LORRAINE E. HUGHES[1], CLAIRE HALPIN[2] AND MARTIN D. RYAN[1]

[1]*Centre for Biomolecular Sciences, School of Biology, Biomolecular Sciences Building, University of St. Andrews, North Haugh, St. Andrews, KY16 9ST, UK and* [2]*Plant Research Unit, School of Life Sciences, University of Dundee at SCRI, Invergowrie, Dundee, DD2 5DA, UK*

Introduction

The past few years have witnessed tremendous progress in the introduction of foreign genes into plant cells. Plant transformation has become a valuable tool for conferring resistance to a variety of pests, microbial diseases, and weed control agents, and for improving nutritional content and post-harvest characteristics (Mentag *et al.*, 2003; Qaim and Zilberman, 2003; Cao *et al.*, 2005; Paine *et al.*, 2005). Gene transfer into plants also has the potential to facilitate the production of new and improved raw materials for a wide range of industries: for instance, to optimize or engineer oil-producing plants to produce industrial lubricants, plastics, biodegradable detergents, and bio-fuels (McLaren, 2005). In medical science, transgenic plants are being developed for the production of high value molecules, such as therapeutic proteins and vaccines. This subject area has been reviewed several times in recent years (Hiatt *et al.*, 1989; Ma *et al.*, 2003; Mor *et al.*, 2004). Not surprisingly, the majority of experiments to date have focused primarily on genetic transformation with single target genes of scientific or applied interest. In most cases, trait genes are transferred with a selectable/visual reporter because the efficiency of plant transformation is less than optimal in most crops. These genes either

*To whom correspondence may be addressed (gal@st-andrews.ac.uk)

Abbreviations: ACK, acetate kinase; CaMV, cauliflower mosaic virus; CAT, chloramphenicol acetyltransferase; CHYSEL, cis-acting hydrolase element; ER, endoplasmic reticulum; eIF4G, eukaryotic initiation factor 4G; FMDV, foot-and-mouth disease virus; GUS, β-glucuronidase; HRV, human rhinovirus; IRES, internal ribosomal entry site; ORF, open reading frame; SMV, soybean mosaic virus; SRP, signal recognition particle; TEV, tobacco etch potyvirus; TMV, tobacco mosaic virus; TVMV, tobacco vein mottling virus; TP, transit peptide.

Biotechnology and Genetic Engineering Reviews – Vol. 23, December 2006
0264-8725/07/23/239-252 $20.00 + $0.00 © Lavoisier/Intercept, 14 rue de Provigny, F-94236 Cachan cedex, France

confer resistance to an antibiotic or herbicide, or provide a rapid screen for transformed tissue. However, many agronomic traits, as well as complex metabolic or regulatory pathways, usually involve the action of several genes working in concert, and so full exploitation of the potential for plant engineering will doubtless necessitate the manipulation of multiple transgenes (Halpin, 2005).

Traditional strategies for co-expressing multiple transgenes *in planta* include: (1) crossing plants that contain the genes of interest; (2) re-transformation or (3) co-transformation with multiple plasmids, or (4) with single plasmids on which several transgenes are linked (Halpin *et al.*, 2001). These methods may not result in coordinate expression, especially when the transgenes are unlinked and integrated within different loci of the genome. Biotechnological exploitation requires these genes be introduced into the target species, such that they are co-expressed at the appropriate level and time, and crucially, in the appropriate tissues and subcellular localization. Delivery of tailored gene products to targeted subcellular organelles might provide more favourable environments for certain biochemical reactions, and for accumulating large amounts of some gene and protein products.

Recent developments have provided exciting alternatives to the problems of expression of multiple foreign genes in plants, based on strategies exploited by several animal picornaviruses and plant potyviruses to process their polyproteins. In this review, we provide a brief introduction to the different approaches used to introduce multiple genes into plant genomes. Then, we describe in detail the strategy of coordinated protein expression *via* self-processing polyprotein systems, and we consider the potentials and problems of this novel strategy.

Expression of multiple unlinked transgenes in plants

Strategies used to introduce multiple foreign genes into higher plants have proceeded classically by performing single transformation experiments, then characterizing the progeny, and subsequently crossing suitable lines. Sequential sexual crossing has allowed the accumulation of four independent transgenes in a single plant, enabling it to produce functional antibodies (Ma *et al.*, 1995). Crossing has some drawbacks: the procedure is relatively time-consuming, certainly if more than two transgenes need to be combined, and the different transgenes in the resulting progeny are inserted at different loci, which complicates the subsequent breeding process. A second approach is to introduce the different transgenes *via* sequential single-gene transformation steps. Re-transformation has enabled simultaneous antisense downregulation of three synthase genes in potato, resulting in the production of a novel starch that can withstand up to five freeze–thaw cycles (Jobling *et al.*, 2002). Integrating multiple transgenes by repetitive insertion of single transgenes has as its main drawbacks the time and effort required for the recovery of the transgenic plants and the need for a different selectable marker for each transformation step.

Co-transformation and polycistronic vectors as a means for multi-transgene stacking

Achieving coordinate, high level and stable expression of multiple transgenes in

plants is currently difficult. Expression levels are notoriously variable and influenced by factors that act independently on transgenes at different genetic loci. Instability of expression due to loss, re-arrangement, or silencing of transgenes may occur, and is exacerbated by increasing numbers of transgenic loci and repeated homologous sequences. Even linking two or more genes with a T-DNA, such as the engineering of 'Golden Rice', does not necessarily result in coordinate expression. Golden Rice is the name coined to describe the genetically modified rice that produces carotenoids in the endosperm of the grain, giving rise to a characteristic yellow colour (Ye *et al.*, 2000). In engineering the provitamin A (β-carotene) biosynthetic pathway into rice endosperm, *Agrobacterium*-mediated co-transformation using two plasmid constructs expressing three proteins was required. Plasmid pB19hpc encoded phytoene synthase (*psy*) originating from daffodil (*Narcissus pseudonarcissus*) and a bacterial phytoene desaturase (*crtI*) originating from *Erwinia uredovora* (with a functional transit peptide) – placed under the control of the endosperm-specific glutelin and CaMV (cauliflower mosaic virus) 35S promoters, respectively. The second construct, plasmid pZLcyH, encoded (daffodil) lycopene β-cyclase (*lcy*), plus a functional transit peptide, under the control of the CaMV 35S promoter (Ye *et al.*, 2000). The long-term functioning of this pathway is, therefore, dependent upon the genetic stability of multiple loci with repeated sequences. A second generation Golden Rice (Golden Rice 2), expressing the efficacious maize *psy/crtI* transgene combination, promises a greater impact on vitamin A deficiency and related health issues (Paine *et al.*, 2005). Transfer of multiple genes *via Agrobacterium*-mediated transformation, although possible, is technically demanding, and becomes more and more problematic as the number of plasmids increase when using co-transformation, or the size of the T-DNA increases when using polycistronic vectors.

Co-expression of multiple proteins from a proteolytically processed polyprotein

PRINCIPLE OF THE POLYPROTEIN APPROACH

Linking proteins in the form of polyproteins is a replication strategy adopted by positive-stranded RNA viruses. Virus-encoded proteinases co- and post-translationally process the polyprotein into mature viral proteins. The great merit of this expression strategy is that multiple proteins may be linked together in a single open reading frame (ORF), expressed under the control of a single promoter. The polyprotein then undergoes processing to produce the mature products. Harnessing this 'single-transgene' approach for viral vector-mediated delivery of multiple, individual protein functions is an interesting alternative and has the advantages of relative simplicity in terms of plant transformation technology and the number of transformants that need to be screened is much lower. A possible disadvantage of this strategy is that these proteinases may also play a role in various aspects of virus–host interactions, from genome expression and amplification to virus dissemination through the infected host.

APPLICATIONS OF ARTIFICIAL POLYPROTEINS USING PROTEINASES

Potyviral NIa proteinase

In an effort to accumulate multiple proteins in transgenic plants, Marcos and Beachy (1994, 1997) designed an expression cassette based on the 49 kDa nuclear inclusion (NIa) proteinase from tobacco etch potyvirus (TEV). The NIa protein is one of three proteinases in TEV that are responsible for self-processing the viral polyprotein, and is active both *in cis* (intramolecular proteolysis) and *in trans* (intermolecular proteolysis). NIa recognizes a specific heptapeptide sequence [Glu-x-Tyr-x-Gln/Gly (or Ser)], where x is any amino acid, and is accountable for several processing events of the large viral peptide (Parks and Dougherty, 1991; Marcos and Beachy, 1994; Dasgupta *et al.*, 1998). In this case, the linker cassette included the NIa coding sequence flanked by cleavage recognition sites, cleaving to create both its N- and C-termini – cleavage in *cis* (see *Figure 12.1, panel a*). The linker peptide connected the tobacco mosaic virus (TMV) and soybean mosaic virus (SMV) coat proteins. While proper processing of the multifunctional polypeptide was demonstrated *in planta*, the coat proteins accumulated only to low levels, irrespective of their location in the cassette, and, accordingly, the levels of coat protein-mediated resistance to TMV were low. A similar proteolytic strategy using the TEV NIa proteinase was successfully adopted by Liang and colleagues to simultaneously express two pathogen resistance-enhancing proteins, a *Trichoderma harzinum* endochitinase (Ech42) and a wheat oxalate oxidase (OxO), in *Arabidopsis* (Liang *et al.*, 2005). Although secretion of Ech42 and OxO were not detected in the transformed plants (despite the fact that they each contained a native signal sequence for secretion), the polyprotein was properly cleaved after translation, and both products exhibited functional enzymatic activity *in vivo*. The homologous proteinase of tobacco vein mottling virus (TVMV), a closely related potyvirus, cleaves at similar, yet distinct, recognition sites (Parks and Dougherty, 1991). A similar expression cassette based on this proteinase was used to transgenically express the first two enzymes involved in the mannityl opine biosynthetic pathway (von Bodman *et al.*, 1995). The presence of the opine intermediates was demonstrated, although the level of accumulation of the enzymes was not determined in this study. To be of agronomic value, levels of the foreign protein must accumulate to significant levels in the transformed plant. Therefore, it may be desirable to target protein products into a suitable subcellular compartment. Dasgupta and colleagues constructed and introduced in tobacco a series of genes encoding polyproteins containing the TVMV NIa proteinase, along with two other reporter genes, those encoding the *Escherichia coli* acetate kinase (ACK) and chloramphenicol acetyl-transferase (CAT) enzymes (Dasgupta *et al.*, 1998). Another component of the polyprotein precursor was chloroplast targeting information (a transit peptide (TP) from a pea rbcS gene). Results showed that a polyprotein construct driven by a single promoter can be processed efficiently into its components by NIa proteinase in transgenic plants, and that subcellular localization information can be incorporated into the polyprotein so as to target the individual components of the polyprotein. However, because cleavage by NIa occurs post-translationally, it is possible for chloroplast localization to outstrip processing, resulting in the mislocalization of incompletely processed polypeptides

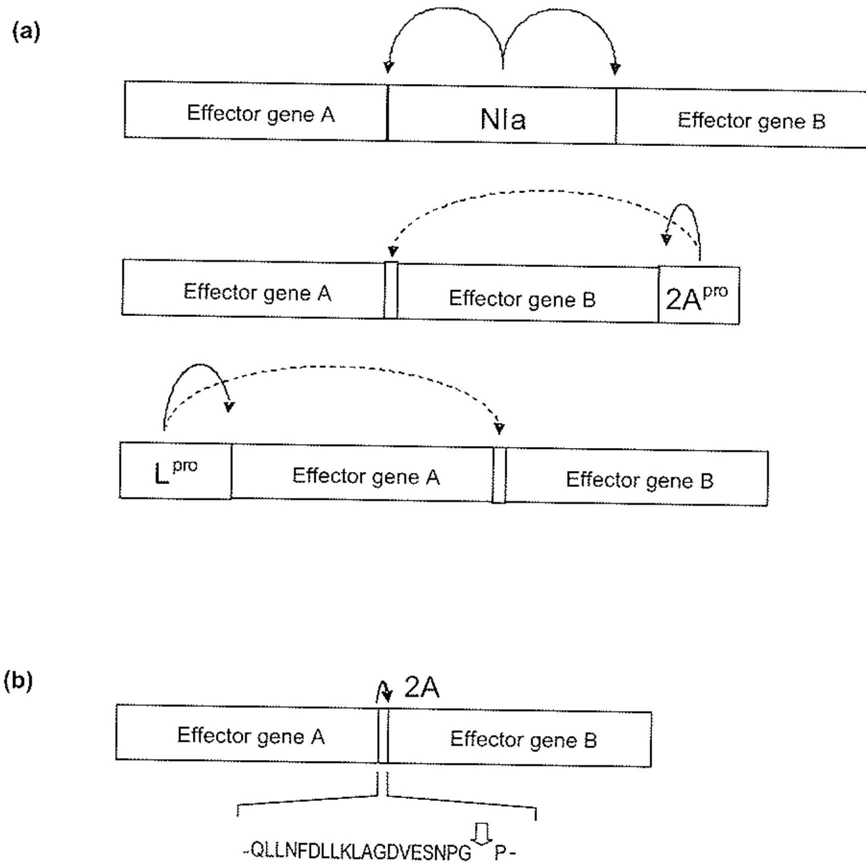

Figure 12.1. Polyprotein expression systems. (a) Proteolytic processing of dual-reporter polyproteins by NIa, 2A^pro^, and L^pro^. Linker sites (shaded) between each effector gene (2A^pro^/L^pro^) contain a cleavage site that is processed by the proteinase either in an intra- (*in cis* – solid line), or an intermolecular manner (*in trans* – dashed line). (b) CHYSEL processing by the FMDV 2A. The amino acid sequence of 2A is shown below with the open arrow indicating the site of cleavage.

to the choloroplast. Nevertheless, these reports suggest the potential of potyvirus-encoded NIa proteinases as tools for attaining coordinate expression of multiple proteins in targeted plant cells.

FMDV L and HRV 2A proteinases

We have used the foot-and-mouth disease virus (FMDV) L and human rhinovirus (HRV) 2A proteinases in the construction of other self-processing polyproteins (*Figure 12.1, panel a*). The leader protein (L^{pro}) of FMDV is a cysteine proteinase with a papain-like fold that releases itself from the polyprotein by cleaving at its own C-terminus (reviewed in Ryan and Flint, 1997). In contrast, HRV $2A^{pro}$, a zinc-containing cysteine proteinase with a chymotrypsin-like fold, cleaves at its N-terminus to separate the structural and replicative protein precursors (Toyoda *et al.*, 1986; Petersen *et al.*, 1999). Like many viral proteins, both L^{pro} and $2A^{pro}$ are

multifunctional enzymes. In addition to processing the viral polyprotein, both proteinases cleave the mammalian host cell protein eukaryotic initiation factor 4G (eIF4G), albeit at different but proximal sites (Kuechler *et al.*, 2002; Gradi *et al.*, 2004). eIF4G plays an important role during protein synthesis initiation by coordinating the protein complex recognizing the capped mRNA with the rest of the translation machinery. The rationale for exploring the utility of these proteinases in plants is that while they have evolved to cleave specific mammalian host-cell proteins, the plant homologues of these proteins (eIF4G and eIFiso4G) show sequence variation, notably in the corresponding region of their mammalian counterparts, which are cleaved by the virus-encoded proteinases. Indeed, we have analysed polyproteins containing L^{pro} and $2A^{pro}$ in *E. coli*, *in vitro* translation systems (rabbit reticulocyte lysates, wheat germ extracts) and a limited number of transgenic tobacco plants.

We found that in all these expression systems, the $2A^{pro}$- and L^{pro}-based polyproteins cleaved at their own termini (L^{pro} – C-terminus; $2A^{pro}$ – N-terminus) in a rapid, intramolecular manner, as anticipated. They also mediated cleavage at the proteinase cleavage linker sites in a slower, intermolecular manner, as hoped (Luke *et al.*, manuscript in preparation). These proteinases proved, therefore, to be non-toxic to plants, and functioned as 'restriction proteinases'. A major drawback was encountered, however, in that constructs encoding L^{pro} proved to be toxic to *Agrobacterium tumefaciens*. This precludes, therefore, this particular method of plant transformation for constructs comprising this proteinase (Cowton, 2000).

Co-expression of multiple proteins by manipulation of translation

OPERON-LIKE mRNAS: PLASTID TRANSFORMATION

In addition to expression as nuclear transcription units, foreign genes can also be expressed as plastid genes. Many chloroplast genes are grouped in bacterial-type operons (groups of genes with a single promoter). Compared to conventional nuclear transformation, engineering foreign genes through chloroplast genomes offers several potential advantages, including high transgene expression levels (Staub *et al.*, 2000), absence of epigenetic effects (gene silencing and position effects), easy transgene stacking in operons (Matzke and Matzke, 1998; Bock, 2001), and increased biosafety because, in most cases, plastid DNA is not transmitted in pollen (Daniell *et al.*, 2002). Gene escape through pollen, the toxicity of transgenic pollen to non-target insects, and the possibility of insects developing resistance to insecticidal proteins, due to low levels of expression, are serious environmental concerns. The introduction of genes *via* chloroplast genetic engineering was advanced as a potential solution to these problems. The landmark study of de Cosa and colleagues demonstrates the ability of transgenic chloroplasts to express bacterial operons, opening the door for multigene engineering *via* the chloroplast genome (de Cosa *et al.*, 2001). The *Bacillus thuringiensis cry2Aa2* operon encoding the Cry2Aa2 insecticidal protein and two other proteins (including a putative chaperonin that facilitates folding of Cry2Aa2) was introduced into the chloroplast genome. Expression of the operon resulted in the accumulation of 46.1% of total soluble protein, even in senescing bleached old leaves, arguing

against the possibility of gene silencing in this system. Due to the simultaneous expression of the gene, coding for the chaperonin, operon-derived Cry2Aa2 formed proteolytically stable, cuboidal crystals. Insects that are normally difficult to control were killed after consuming transgenic leaves.

Although plastid transformation is a promising option, it is a difficult technology. Several therapeutic proteins and agronomic traits have been highly expressed *via* the tobacco chloroplast genome, but extending this concept to important crops has been a major challenge. Thus far, only a limited number of plants have been transformed successfully by plastome manipulation (Sikdar *et al.*, 1998; Sidorov *et al.*, 1999; de Cosa *et al.*, 2001; Skarjinskaia *et al.*, 2003). On top of that is the low success rate of gene insertion into the chloroplast genome and the need for extensive selection procedures (van Bel *et al.*, 2001). Making this a routine method will depend on progress in simplifying the technology and extending the crop range.

INTERNAL INITIATION OF TRANSLATION AND IRESs

Translation of most eukaryotic mRNAs is cap-dependent, and proceeds with the recruitment of small ribosomal subunits to the 5' end of the transcript and 'scanning' to the first suitable initiation codon. However, a few animal mRNAs have been shown to contain an internal ribosomal entry site (IRES). These sequences mediate cap-independent ribosomal attachment to an internal position in the mRNA. Detecting the activity of putative IRESs usually involves the construction of dicistronic mRNAs, in which the IRES element is an intergenic region between two reporter genes, and assaying the second reporter gene's activity (Pelletier and Sonenberg, 1988). So far, this translation assay has been used to identify IRES elements in viral genomes and in a broad range of cellular mRNAs from insects to mammals, although no such sequences have been identified in endogenous plant genes (Jackson and Kaminski, 1995; Martinez-Salas *et al.*, 2004). In plants, it follows that those viral mRNAs that naturally lack a 5' cap structure must use a translation initiation mechanism that bypasses the normal cap recognition step. Indeed, the 5' leaders of several plant viral mRNAs, including members of the *Potyviridae*, *Comoviridae*, *Tobamoviridae*, and *Luteoviridae*, are responsible for conferring cap-independent translation (Carrington and Freed, 1990; Thomas *et al.*, 1991; Gallie *et al.*, 1995; Ivanov *et al.*, 1997; Niepel and Gallie, 1999; Skulachev *et al.*, 1999). IRESs have been used successfully in gene therapy research in animal systems, and IRESs from different viruses have been tested and shown to function in plant systems (Urwin *et al.*, 2000, 2002; Dorokhov *et al.*, 2002; Jaag *et al.*, 2003). Since genes are under the control of the same promoter, and integrated into the same place within the genome, transgenes expressed in this way are coordinately regulated. A major limitation of this strategy is that proteins are expressed at different levels due to the relative inefficiency of internal initiation. Expression of a gene downstream of an IRES is characteristically only ~10% of the upstream gene (reviewed in de Felipe, 2002).

TRANSLATIONAL 'SKIPPING' USING FMDV 2A POLYPROTEINS

Unlike the enteroviruses and rhinoviruses, where 2A sequences are recognizable as those of a proteolytic enzyme, the aphthoviruses and cardioviruses do not have a 2A

proteinase, but instead rely on an oligopeptide with a 'self-cleaving' activity. The 18 amino acid oligopeptide 2A region of FMDV plays a major role in polyprotein processing, but is not a proteinase (Ryan and Drew, 1994). The sole function of the 2A protein is to direct a single 'cleavage' at its own C-terminus, and function only in *cis* (*Figure 12.1, panel b*). The core sequences adjacent to the cleavage site are strongly conserved and always contain the essential [Asp-x-Glu-x-Asn-Pro-Gly-Pro] motif. The cleavage occurs between the last two residues of this sequence (Gly/Pro). Site-directed mutagensis of this conserved motif uniformly show either reduced or no cleavage activity (Donnelly *et al.*, 2001b). '2A-like' sequences are found not only in other picornaviruses (cardio-, teschoviruses), but in human type C rotaviruses, reoviruses (insect cypoviruses), and in a wide range of insect +ve strand RNA viruses. Furthermore, 2A-like sequences are found in repeated sequences in *Trypanosoma cruzi* (Chagas disease) and a different type of repeat sequence in *T. brucei* (African trypanosomiasis – sleeping sickness) (Donnelly *et al.*, 2001b). Recently, we have shown that 'cleavage' is by a novel translational effect (ribosome 'skipping'), rather than proteolysis – a full description of the proposed CHYSEL (**c**is-acting **hy**drolase **el**ement) model is presented in Ryan *et al.* (1999) and Donnelly *et al.* (2001a). Cleavage gives rise to two alternative outcomes: (1) translation terminates at the C-terminus of 2A in the absence of a stop codon; or (2) while the synthesis of the glycyl-prolyl peptide bond (-NPG P-) is 'skipped', translation of the remainder of the downstream protein continues. In this manner, multiple, discrete ('cleaved'), translation products can be synthesized from a single open reading frame.

Subcellular targeting of proteins from a 2A polyprotein

Since 2A separation is co-translational, one major aspect affecting its biotechnological utility is the effect of including co-translational signal sequences at different sites within the self-processing polyprotein. Protein targeting occurs either co-translationally (targeting to endoplasmic reticulum [ER], Golgi apparatus, secretory vesicles, plasma membrane, vacuoles in plants, and lysosomes in animals) or post-translationally (targeting to nucleus, mitochondria, plastids, chloroplast, etc.), and is orchestrated by specific targeting signals encoded within the polypeptide sequence of proteins destined for these locations. By the inclusion of a range of co-translational signal sequences into artificial reporter polyprotein systems, we have shown that:

- co-translational signal sequences located downstream of 2A were recognized by the signal recognition particle (SRP) and the protein translocated into the cellular secretory pathway. This supports our hypothesis that 2A-mediated cleavage occurs within the ribosome; such signal sequences only function as nascent N-terminal features (de Felipe and Ryan, 2004). A post-translational or delayed co-translational cleavage would not allow targeting proteins to the ER using the co-translational SRP route, as the signal peptide needs to be recognized very quickly while still exposed in the vicinity of the exit tunnel of the ribosome. This feature of 2A enables us to co-express a cytosolic protein followed by a protein co-translationally targeted to the secretion pathway or the plasma membrane;
- a single signal sequence located at the N-terminus of a 2A-based polyprotein

system leads to the 'slipstream' translocation of proteins downstream of 2A into the lumen of the ER, even though they do not possess any signal sequence. This is the case for mammalian, but not for plant or yeast translocon systems, where proteins downstream of 2A and lacking any signal sequences remain in the cytoplasm (de Felipe *et al.*, 2003; de Felipe and Ryan, 2004; El Amrani *et al.*, 2004);

- post-translational signal sequences, either upstream or downstream of 2A, function perfectly. All reports indicate the correct post-translational targeting to the nucleus, chloroplast, membranes, or cytosolic tubules formed by the movement proteins of certain plant viruses (de Felipe and Ryan, 2004; El Amrani *et al.*, 2004).

Utility of 2A in plant biotechnology

Due to its small size, FMDV 2A (and related sequences or derivatives) presents an ideal candidate for engineering into plant expression vectors. To demonstrate the ability of FMDV 2A to function independently of other FMDV sequences, 2A activity was studied in artificial self-processing polyproteins comprising the reporter proteins, CAT and β-glucuronidase (GUS), flanking the foot-and-mouth disease virus (FMDV) 2A sequence in a single open reading frame. Expression of [CAT2AGUS] in wheatgerm lysate and transgenic tobacco resulted in consistent detection of freed CAT and/or GUS proteins, indicating that FMDV 2A protein functioned properly in plant cells. Moreover, the 19 amino acid 2A-spanning sequence, which remains attached to CAT after cleavage from GUS, did not compromise CAT activity (Halpin *et al.*, 1999). The length and/or sequence of the 2A or '2A-like' used is important in determining the efficacy of dissociation. This offers researchers bespoke solutions to a wide range of co-expression problems. Minimum primary cleavage maps to a 13 amino acid element, which includes most of the 2A region and terminates with the essential octomer (Donnelly *et al.*, 2001b). Longer versions of 2A with a few extra sequences derived from the capsid protein immediately upstream of 2A process more completely. For example, a truncated 16 amino acid version of 2A was used to demonstrate the potential for producing large quantities of useful peptides (such as vaccines) as components of decorated plant virus particles fusing the peptide *via* 2A to viral coat proteins (CP). Assembly of 'decorated' virions required the presence of free coat protein subunits, in addition to the fusion protein subunits. Cleavage of this short version of 2A was inefficient, and both uncleaved and cleaved products of the peptide-2A–CP fusion accumulated, allowing the assembly of viral particles 'decorated' with peptide (Santa-Cruz *et al.*, 1996, more examples cited in de Felipe *et al.*, 2006).

In plant biotechnology, 2A has been used in several examples of metabolome engineering and the introduction of novel product traits. Increasing the resistance of crops to osmotic stresses was one of the first objectives of plant metabolic engineering (LeRudulier *et al.*, 1984), and remains a major goal today. Of special interest is the metabolism of trehalose, a compound considered not only as a source of energy, but mainly as an osmoprotectant. The engineering of trehalose accumulation in plants has been undertaken not only to improve stress resistance, but also to produce trehalose at low cost for use as a stabilizing agent for pharmaceuticals and other

products. Trehalose is biosynthesized by the trehalose-6-phosphate synthase/phosphatase (TPS1/TPS2) enzyme complex in a two-step pathway. Both TPS1 and TPS2 genes of yeast were introduced simultaneously into potato plants as a TPS2-2A-TPS1 polyprotein in an attempt to generate abiotic stress-tolerant plants. Although coordinate expression of both proteins was not shown, the resulting transgenic potato plants showed very strong drought tolerance (Kwon *et al.*, 2004).

Using metabolic engineering, Ralley and colleagues modified the carotenoid pathway in tobacco and tomato to produce astaxanthin, a red pigment of considerable economic value (Ralley *et al.*, 2004). This was achieved by expressing the two enzymes necessary for astaxanthin synthesis, 4,4'-β-oxygenase (crtW) and 3,3'-β-hydroxylase (crtZ) (from *Paracoccus* sp.), simultaneously as a crtW-2A-crtZ polyprotein. The pea Rubisco small subunit transit peptide sequence was used as an N-terminal signal for targeting the two enzymes into chromoplasts. Subsequent cleavage of the polyprotein, targeting of the two enzymes to the plastid, and enzyme activity was shown for both gene products.

Two classes of alcohol soluble seed storage proteins found in the endosperm of maize contain unusually large amounts of the sulfur containing amino acids, cysteine, and methionine. These two proteins, the β and δ zeins, were introduced into plants with a view to improving the sulfur content and nutritional value of plants (Randall *et al.*, 2004). The β and δ zeins were fused with a synthetic 2A peptide, and the β-zein-2A-δ-zein construct introduced into tobacco. Effective processing of the polyprotein was observed, resulting in the production, in transgenic plants, of the individual β and δ zein proteins.

General considerations when using 2A polyproteins

When using the 2A system, it should be borne in mind that: (1) the 2A oligopeptide remains as a C-terminal extension of the upstream fusion partner; and (2) attachment of a proline residue at the N-terminus of the downstream protein could have adverse effects on the enzymatic activity and/or biological function of certain proteins. The structure of the C-terminus may play a critical role in the localization or function of a protein. The need to target proteins to different subcellular locations within plant cells by C-terminal localization signals may be compromised if they contain a 2A-extension. Recently, strategies have been devised that allow removal of the 2A linker. By combining at the N-terminal of the FMDV 2A sequence, the fourth linker peptide (LP4) of the polyprotein precursor originating from *I. balsamina*, Francois and colleagues succeeded in the amino- and carboxy-terminal cleavage of the linker peptide by endogenous proteinases acting on the LP4, and by self-cleavage mediated by the FMDV 2A sequence, respectively (Francois *et al.*, 2004). Fang and colleagues were able to engineer an antibody expression cassette in which monoclonal antibody heavy and light chain sequences are linked by a combination of a furin cleavage site next to the 2A sequence (Fang *et al.*, 2005). Removal of 2A-derived amino acid residues by furin cleavage generates an antibody that more closely resembles the native protein, thereby eliminating possible adverse effects. The presence of an N-terminal proline residue does not normally interfere with function – it does, however, confer high protein stability (Varshavsky, 1992). To date, all the different proteins fused to the 2A moiety have shown activity. More-

over, proteins that require authentic termini, or are N-/C-terminally modified, can be introduced as the first or final polyprotein domain, respectively. However, the 2A 'tag' may remain – a number of studies have demonstrated the value of using antibodies directed against 2A to detect the gene cloned upstream, in particular when several 'gene-2A' cassettes are linked in tandem.

Closing remarks

The importance of combining two or more desirable genes or transgenes in one plant is receiving increased attention in review articles on the future of plant biotechnology. Although conventional methods exist for combining transgenes in a single plant, such methods are, for the most part, relatively crude. They generally suffer the drawbacks of being both time-consuming and work-intensive, and cannot ensure coordinate expression of different transgenes – this has to be extensively screened and selected for. There is a real and urgent need to improve the efficiency of conventional methods for stacking genes or to develop new methods for achieving the same ends. Until this happens, the full potential of plant genetic manipulation, either as a basic research tool or as a commercial enterprise, cannot be realized. We, and others, have developed the utility of the FMDV 2A oligopeptide sequence to produce artificial, self-processing polyprotein systems. With advances in our understanding of how this self-processing system operates, the possibility of manipulating the timing, type, and site of delivery of diverse protein functions is therefore becoming more and more feasible. Though the system described herein will continue to make important contributions in applications of plant transgenesis, a number of different virus-encoded proteinases, some of which have yet to be exploited fully, will likely complement the current armoury. A pick 'n' mix approach using both types of polyprotein mechanism (translational 'skipping' and proteolysis) can be envisaged readily. Indeed, a system could be developed whereby the proteinase would be constitutively expressed in the target tissue of a plant and the 'substrate' polyprotein (alone) needs to be transfected. Finally, despite the diverse and widespread beneficial applications of biotechnology products, resistance from the public to genetically modified plants may well seriously inhibit the introduction of potentially beneficial crops (Falk *et al.*, 2002). It is, therefore, clear that stringent controls and the provision of accessible information are requisites to assure and encourage public acceptance (Sharma *et al.*, 2002).

Acknowledgements

We gratefully acknowledge the Biotechnology and Biological Sciences Research Council for supporting this work.

References

BOCK, R. (2001). Transgenic chloroplasts in basic research and plant biotechnology. *Journal of Molecular Biology* **312**, 425–438.

CAO, M.-X., HUANG, J.-Q., WEI, Z.-M., YAO, Q.-H., WAN, C.-Z. AND LU, J.-A. (2005). *Agrobacterium*-mediated multiple gene transformation in rice using a single vector. *Journal of Integrative Plant Biology* **47**, 233–242.

CARRINGTON, J.C. AND FREED, D.D. (1990). Cap-independent enhancement of translation by a plant potyvirus 5' nontranslated region. *Journal of Virology* **64**, 1590–1597.

COWTON, V.M. (2000). Coordinated expression *via* artificial self-processing polyprotein systems. PhD thesis, University of St. Andrews.

DANIELL, H., KHAN, M.S. AND ALLISON, L. (2002). Milestones in chloroplast genetic engineering: an environmentally friendly era in biotechnology. *Trends in Plant Science* **7**, 84–91.

DASGUPTA, S., COLLINS, G.B. AND HUNT, A.G. (1998). Coordinated expression of multiple enzymes in different subcellular compartments in plants. *The Plant Journal* **16**, 107–116.

DE COSA, B., MOAR, W., LEE, S.-B., MILLER, M. AND DANIELL, H. (2001). Over-expression of the *Btcry*2Aa2 operon in chloroplasts leads to formation of insecticidal crystals. *Nature Biotechnology* **19**, 71–74.

DE FELIPE, P. (2002). Polycistronic viral vectors. *Current Gene Therapy* **2**, 355–378.

DE FELIPE, P. AND RYAN, M.D. (2004). Targeting of proteins derived from self-processing polyproteins containing multiple signal sequences. *Traffic* **5**, 616–626.

DE FELIPE, P., HUGHES, L.E., RYAN, M.D. AND BROWN, J.D. (2003). Co-translational, intraribosomal cleavage of foot-and-mouth disease virus 2A peptide. *Journal of Biological Chemistry* **278**, 11441–11448.

DE FELIPE, P., LUKE, G.A., HUGHES, L.E., GANI, D., HALPIN, C. AND RYAN, M.D. (2006). *E unum pluribus*: multiple proteins from a self-processing polyprotein. *Trends in Biotechnology* **24**, 68–75.

DONNELLY, M., LUKE, G.A., MEHROTRA, A. *ET AL.* (2001a). Analysis of aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect : a putative ribosomal 'skip'. *Journal of General Virology* **82**, 1013–1025.

DONNELLY, M., HUGHES, L.E., LUKE, G.A. *ET AL.* (2001b). The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. *Journal of General Virology* **82**, 1027–1041.

DOROKHOV, Y.L., SKULACHEV, M.V., IVANOV, P.A. *ET AL.* (2002). Polypurine (A)-rich sequences promote cross-kingdom conservation of internal ribosome entry. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 5301–5306.

EL AMRANI, A., BARAKATE, A., ASKARI, B.M. *ET AL.* (2004). Coordinate expression and independent subcellular targeting of multiple proteins from a single transgene. *Plant Physiology* **135**, 16–24.

FALK, M.C., CHASSY, B.M., HARLANDER, S.K., HOBAN, T.J., MCGLOUGHLIN, M.N. AND AKHLAGHI, A.R. (2002). Food biotechnology: benefits and concerns. *Journal of Nutrition* **132**, 1384–1390.

FANG, J., QIAN, J.-J., YI, S., HARDING, T.C., TU, G.H., VAN ROEY, M. AND JOOSS, K. (2005). Stable antibody expression at therapeutic levels using the 2A peptide. *Nature Biotechnology* **23**, 584–590.

FRANCOIS, I.E.J.A., HEMELRIJCK, W.V., AERTS, A.M. *ET AL.* (2004). Processing in *Arabidopsis thaliana* of a heterologous polyprotein resulting in differential targeting of the individual plant defensins. *Plant Science* **166**, 113–121.

GALLIE, D.R., TANGUAY, R. AND LEATHERS, V. (1995). The tobacco etch viral 5' leader and poly(A) tail are functionally synergistic regulators of translation. *Gene* **165**, 233–238.

GRADI, A., FOEGER, N., STRONG, R. *ET AL.* (2004). Cleavage of eukaryotic translation initiation factor 4GII within foot-and-mouth disease virus-infected cells: identification of the L-protease cleavage site *in vitro*. *Journal of Virology* **78**, 3271–3278.

HALPIN, C. (2005). Gene stacking in transgenic plants – the challenge for 21st century plant biotechnology. *Plant Biotechnology Journal* **3**, 141–155.

HALPIN, C., COOKE, S.E., BARAKATE, A., EL AMRANI, A. AND RYAN, M.D. (1999). Self-processing 2A-polyproteins – a system for coordinate expression of multiple proteins in transgenic plants. *The Plant Journal* **17**, 453–459.

HALPIN, C., BARAKATE, A., ASKARI, B.M., ABBOTT, J.C. AND RYAN, M.D. (2001). Enabling technologies for manipulating multiple genes on complex pathways. *Plant Molecular Biology* **47**, 295–310.

HIATT, A., CAFFERKEY, R. AND BOWDISH, K. (1989). Production of antibodies in transgenic plants. *Nature* **342**, 76–78.

IVANOV, P.A., KARPOVA, O.V., SKULACHEV, M.V. *ET AL.* (1997). A tobamovirus genome that contains an internal ribosome entry site function *in vitro*. *Virology* **232**, 32–43.

JAAG, H.M., KAWCHUK, L., ROHDE, W., FISCHER, R., EMANS, N. AND PRUFER, D. (2003). An unusual internal ribosome entry site of inverted symmetry directs expression of a potato leafroll polerovirus replication-associated protein. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 8939–8944.

JACKSON, R.J. AND KAMINSKI, A. (1995). Internal initiation of translation in eukaryotes: the picornavirus paradigm and beyond. *RNA* **1**, 985–1000.

JOBLING, S.A., WESTCOTT, R.J., TAYAL, A., JEFFCOAT, R. AND SCHWALL, G.P. (2002). Production of freeze–thaw-stable potato starch by antisense inhibition of three starch synthase genes. *Nature Biotechnology* **20**, 295–299.

KUECHLER, E., SEIPELT, J., LIEBIG, H.-D. AND SOMMERGRUBER, W. (2002). Picornavirus proteinase-mediated shutoff of host cell translation: direct cleavage of a cellular initiation factor. In: *Molecular biology of picornaviruses*. Eds. B.L. Semler and E. Wimmer, pp 301–311. Washington, USA: ASM Press.

KWON, S.-J., HWANG, E.-W. AND KWON, H.-B. (2004). Genetic engineering of drought-resistant potato plants by co-introduction of genes encoding trehalose-6-phosphate synthase and trehalose-6-phosphate of *Zygosaccharomyces rouxii*. *Korean Journal of Genetics* **26**, 199–206.

LERUDULIER, D., STROM, A.R., DANDEKAR, A.M., SMITH, L.T. AND VALENTINE, R.C. (1984). Molecular biology of osmoregulation. *Science* **224**, 1064–1068.

LIANG, H., GAO, H., MAYNARD, C.A. AND POWELL, W.A. (2005). Expression of a self-processing, pathogen resistance-enhancing gene construct in *Arabidopsis*. *Biotechnology Letters* **27**, 435–442.

MA, J.K.-C., HIATT, A., HEIN, M.B. *ET AL.* (1995). Generation and assembly of secretory antibodies in plants. *Science* **268**, 716–719.

MA, J.K.-C., DRAKE, P.M. AND CHRISTOU, P. (2003). The production of recombinant pharmaceutical proteins in plants. *Nature Reviews Genetics* **4**, 794–805.

MARCOS, J.F. AND BEACHY, R.N. (1994). *In vitro* characterization of a cassette to accumulate multiple proteins through synthesis of a self-processing polypeptide. *Plant Molecular Biology* **24**, 495–503.

MARCOS, J.F. AND BEACHY, R.N. (1997). Transgenic accumulation of two plant virus coat proteins on a single self-processing polypeptide. *Journal of General Virology* **78**, 1771–1778.

MARTINEZ-SALAS, E., FERNANDEZ-MIRAGALL, O., REIGADAS, R., PACHECO, A. AND SERRANO, P. (2004). Internal ribosome entry site elements in eukaryotic genomes. *Current Genomics* **5**, 259–277.

MATZKE, A.J.M. AND MATZKE, M.A. (1998). Position effects and epigenetic silencing of plant transgenes. *Current Opinion in Plant Biology* **1**, 142–148.

MCLAREN, J.S. (2005). Crop biotechnology provides an opportunity to develop a sustainable future. *Trends in Biotechnology* **23**, 339–342.

MENTAG, R., LUCKEVICH, M., MORENCY, M.J. AND SEGUIN, A. (2003). Bacterial disease resistance of transgenic hybrid poplar expressing the synthetic antimicrobial peptide D4E1. *Tree Physiology* **23**, 405–411.

MOR, T.S., MASON, H.S., KIRK, D.D., AMTZEN, C.J. AND CARDINEAU, G.A. (2004). Plants as production and delivery vehicle for orally delivered subunit vaccines. In: *New generation vaccines* 3rd Edition. Eds. M.M. Levine, R. Rapuoli, M.A. Liu and M.F. Good, pp 305–312. New York–Basel: Marcel Dekker, Inc.

NIEPEL, M. AND GALLIE, D.R. (1999). Identification and characterization of the functional elements within the tobacco etch virus 5' leader required for cap-independent translation. *Journal of Virology* **73**, 9080–9088.

PAINE, J.A., SHIPTON, C.A., CHAGGAR, S. *ET AL.* (2005). Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nature Biotechnology* **23**, 482–487.

PARKS, T.D. AND DOUGHERTY, W.G. (1991). Substrate recognition by the Nla proteinase of two potyviruses involves multiple domains: characterization using genetically engineered hybrid proteinase mutations. *Virology* **182**, 17–27.

PELLETIER, J. AND SONENBERG, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* **334**, 320–325.

PETERSEN, J.F.W., CHERNEY, M.M., LIEBIG, H.-D., SKERN, T., KUECHLER, E. AND JAMES, M.N.G. (1999). The structure of the 2A proteinase from a common cold virus: a proteinase responsible for the shut-off of host-cell protein synthesis. *EMBO Journal* **18**, 5463–5475.

QAIM, M. AND ZILBERMAN, D. (2003). Yield effects of genetically modified crops in developing countries. *Science* **299**, 900–902.

RALLEY, L., ENFISSI, E.M.A., MISAWA, N., SCHUCH, W., BRAMLEY, P.M. AND FRASER, P.D. (2004). Metabolic engineering of ketocarotenoid formation in higher plants. *The Plant Journal* **39**, 477–486.

RANDALL, J., SUTTON, D., GHOSHROY, S., BAGGA, S. AND KEMP, J.D. (2004). Coordinate expression of β– and δ–zeins in transgenic tobacco. *Plant Science* **167**, 367–372.

RYAN, M.D. AND DREW, J. (1994). Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *EMBO Journal* **134**, 928–933.

RYAN, M.D. AND FLINT, M. (1997). Virus-encoded proteinases of the picornavirus super-group. *Journal of General Virology* **78**, 699–723.

RYAN, M.D., DONNELLY, M.L.L., LEWIS, A., MEHROTRA, A.P., WILKIE, J. AND GANI, D. (1999). A model for non-stoichiometric, co-translational protein scission in eukaryotic ribosomes. *Bioorganic Chemistry* **27**, 55–79.

SANTA-CRUZ, S., CHAPMAN, S., ROBERTS, A.G. AND ROBERTS, I.M. (1996). Assembly and movement of a plant virus carrying a green fluorescent overcoat. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 6286–6290.

SHARMA, H.C., CROUCH, J.H., SHARMA, K.K., SEETHARAMA, N. AND NASH, C.T. (2002). Applications of biotechnology for crop improvement: prospects and constraints. *Plant Science* **163**, 381–395.

SIDOROV, V.A., KASTEN, D., PANG, S.Z., HAJDUKIEWICZ, P.T.J., STAUB, J.M. AND NEHRA, N.S. (1999). Technical advance: stable chloroplast transformation in potato: use of green fluorescent protein as a plastid marker. *The Plant Journal* **19**, 209–216.

SIKDAR, S.R., SERINO, G., CHAUDHURI, S. AND MALIGA, P. (1998). Plastid transformation in *Arabidopsis thaliana*. *Plant Cell Reports* **18**, 20–24.

SKARJINSKAIA, M., SVAB, Z. AND MALIGA, P. (2003). Plastid transformation in *Lesquerella fendleri*, an oilseed Brassicacea. *Transgenic Research* **12**, 115–122.

SKULACHEV, M.V., IVANOV, P.A., KARPOVA, O.V. *ET AL.* (1999). Internal initiation of translation directed by the 5′ untranslated region of the tobamovirus subgenomic RNA I_2. *Virology* **263**, 139–154.

STAUB, J.M., GARCIA, B., GRAVES, J. *ET AL.* (2000). High-yield production of a human therapeutic protein in tobacco chloroplasts. *Nature Biotechnology* **18**, 333–338.

THOMAS, A.A.M., TER HAAR, E., WELLINK, J. AND VOORMA, H.O. (1991). Cowpea mosaic virus middle component RNA contains a sequence that allows internal binding of ribosomes and that requires eukaryotic initiation factor 4F for optimal translation. *Journal of Virology* **65**, 2953–2959.

TOYODA, H., NICKLIN, M.J.H., MURRAY, M.G. *ET AL.* (1986). A second virus-encoded proteinase involved in proteolytic processing of poliovirus polyprotein. *Cell* **45**, 761–770.

URWIN, P.E., YI, L., MARTIN, H., ATKINSON, H.J. AND GILMARTIN, P.M. (2000). Functional characterization of the EMCV IRES in plants. *The Plant Journal* **24**, 583–589.

URWIN, P.E., ZUBKO, E.I. AND ATKINSON, H.J. (2002). The biotechnological application and limitation of IRES to deliver multiple defence genes to plant pathogens. *Physiological and Molecular Plant Pathology* **P61**, 103–108.

VAN BEL, A.J.E., HIBBERD, J., PRUFER, D. AND KNOBLAUCH, M. (2001). Novel approach in plastid transformation. *Current Opinion in Biotechnology* **12**, 144–149.

VARSHAVSKY, A. (1992). The N-end rule. *Cell* **69**, 725–735.

VON BODMAN, S.B., DOMIER, L.L. AND FARRAND, S.K. (1995). Expression of multiple eukaryotic genes from a single promoter in *Nicotiana*. *Bio-Technology* **13**, 587–591.

YE, X., AL-BABILI, S., KLOTI, A. *ET AL.* (2000). Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* **287**, 303–305.