Final manuscript of a book chapter published:

The plant endoplasmic reticulum: methods and protocols **[ISBN: 9781071637098] / edited by Verena Kriechbaumer (Springer, 2024).**

Gene stacking and stoichiometric expression of ER-targeted constructs using "2A" self-cleaving peptides

Tatiana Spatola Rossi[1], Mark Fricker[2] and Verena Kriechbaumer[1]

[1] Endomembrane Structure and Function Research Group, Department of Biological and Medical Sciences, Oxford Brookes University, OX3 0BP Oxford, UK

[2] Department of Biology, University of Oxford, Oxford, OX1 3RB, UK

Corresponding author:

Dr Verena Kriechbaumer

Endomembrane Structure and Function Research Group

Department of Biological and Medical Sciences

Oxford Brookes University

OX3 0BP Oxford, UK

vkriechbaumer@brookes.ac.uk

Running head: ER protein expression with self-cleaving peptides

Gene stacking and stoichiometric expression of ER-targeted constructs using "2A" self-cleaving peptides

Tatiana Spatola Rossi, Mark Fricker and Verena Kriechbaumer

[1] Endomembrane Structure and Function Research Group, Department of Biological and Medical Sciences, Oxford Brookes University, OX3 0BP Oxford, UK

[2] Department of Biology, University of Oxford, Oxford, OX1 3RB, UK

Abstract

Simultaneous stoichiometric expression of multiple genes plays a major part in modern research and biotechnology. Traditional methods for incorporating multiple transgenes (or "gene stacking") have drawbacks such as long time frames, uneven gene expression, gene silencing and segregation derived from the use of multiple promoters. 2A self-cleaving peptides have emerged over the last two decades as a functional gene stacking method, and have been used in plants for the co-expression of multiple genes under a single promoter. Here we describe design features of multicistronic polyproteins using 2A peptides for co-expression in plant cells and targeting to the endoplasmic reticulum (ER). We designed up to quad-cistronic vectors that could target proteins in tandem to the ER. We also exemplify the incorporation of self-excising intein domains within 2A polypeptides, to remove residue additions. These features could aid in the design of stoichiometric protein co-expression strategies in plants in combination with targeting to different subcellular compartments.

Key words Self-cleaving peptides, gene stacking, subcellular targeting, recombinant protein, endoplasmic reticulum

1. Methods for gene stacking

Multigene engineering is essential in modern research for the production of novel traits, the modification of metabolic pathways, and the expression of protein complexes, amongst others [1,2]. Historically, gene stacking in plants was achieved by crossing transgenic lines harbouring individual transgenes. However, this method is very time consuming and labour intensive, and can lead to the segregation of transgenes in subsequent generations due to integration at different genomic loci. Co-transformation with separate plasmids is an alternative strategy that can reduce the time requirement, although this method presents the drawbacks of requiring multiple selection markers, and a lower probability of co-transformation as the gene number increases (Figure 1A). Alternatively, vectors containing multiple expression cassettes have been developed (Figure 1B), such as Modular Cloning (MoClo) [3], MultiSite Gateway Technology [4], or "2in1" vectors [5], amongst others, which allow the incorporation of two or more genes in a single transformation event. These methods have been used successfully for genetic engineering. However, they sometimes result in uncoordinated gene expression due to the use of separate promoters, or gene silencing. Therefore, methods based on polyproteins have been developed, which allow co-expression under a single promoter and co-translational cleavage to give separate proteins. These include the use of internal ribosome entry sites (IRES) and 2A self-cleaving peptides (Figure 1C, D).

IRES sequences are present in the 5'UTR region of some genes (mainly in viruses) and mediate a cap independent initiation of translation [6]. IRES vary in length and neither their amino acid sequence nor their secondary structure are conserved between organisms, so their activity has to be determined empirically [7]. Synthetic polypeptides containing IRES sequences have been used and shown to mediate the translation of two genes [6,8]. However, these sequences have the drawbacks of being fairly long (ranging from 130 bp to 1kb, but typically around 500 bp or more), and of having a relatively low translational efficiency. For example, the downstream gene has been reported to translate at levels of around 10% of the upstream gene [1,9].

2A self-cleaving peptides are short (18-22 amino acid) viral sequences that produce translation of two or more proteins with co-translational cleavage via a “ribosomal skipping” mechanism [10]. A conserved sequence in their C-terminus is thought to induce hydrolysis of a peptidyl(2A)-tRNAGly ester bond, releasing the nascent polypeptide whilst continuing with downstream translation [10]. The most commonly used 2A peptide derives from the foot-and-mouth disease virus (FMDV), termed F2A. Self-cleaving peptides from other viruses have also been identified, such as E2A (from equine rhinitis virus), P2A (from porcine teschovirus -1) and T2A (*Thosea asigna* virus) [11]. All 2A peptides contain the conserved C-terminal sequence –GDVExNPGP, with cleavage occurring between the terminal glycine and proline residues [11]. This process leaves the final proline residue attached to the N-terminus of the downstream protein and the remainder of the self-cleaving peptide fused to the C-terminus of the upstream protein.

2A peptides have a high cleavage efficiency and have been used successfully for polyprotein construction in a variety of systems including mammalian, insect, plant, fungal and yeast cells, for biomedical and biotechnological purposes [12]. It has been widely reported that this system is more effective and efficient for co-expression of genes than the use of IRES, due to their shorter length, higher translational levels of downstream proteins, functionality across different cell types, and ability to concatenate more than two transgenes [1,9,12–14].

2. Use of 2A self-cleaving peptides in plants

Gene stacking in plants mediated by self-cleaving peptides was tested more than two decades ago, by co-expressing the chloramphenicol acetyltransferase (CAT) and β-glucuronidase (GUS) reporters in wheat [15]. Since then others have made use of the system in plants to express bisoynthetic enzymes, produce protein complexes and to study protein trafficking through target-reporter systems.

As an example of introducing novel biosynthetic pathways, Ha and co-workers expressed a bicistronic vector containing the carotenoid biosynthetic genes *Psy* and *CrtI* used in Golden Rice [16,17] linked by F2A in rice endosperm [13]. They obtained β-carotene production as 44% of total carotenoid content, showing that the 2A system is viable for the co-expression of biosynthetic enzymes [13]. Other studies used 2A sequences to co-express a larger number of genes. For example, Liao *et al*., used 2As for metabolic engineering of the production of health-promoting natural products, by co-expressing 6 biosynthetic enzymes in tobacco and Arabidopsis [18]. Another project aimed at biofortification in maize expressed 11 transgenes in the anthocyanin biosynthesis pathway using 2As, to produce “Purple Embryo Maize” [19]. A recent study used 2A to further the development of gene editing in plants. They created Cas9-P2A-GFP to monitor the levels of Cas9 in Arabidopsis and create Cas9-free CRISPR plants [20].

2A peptides are attractive for the production of protein complexes, where a 1:1 stoichiometry between protein subunits is often needed. In this regard, F2A has been used to produce recombinant Ebola monoclonal antibodies in tobacco [21]. Similar expression levels of the heavy and light chains were reported, and a greater yield of assembled antibodies compared to expressing the heavy and light chains from separate transcription units. The authors hypothesized that the coordinated expression could be the cause of the higher yields.

It would be expected for proteins in a polypeptide to be produced at a near 1:1 stoichiometry when present in the same transcriptional unit. However, this is not always the case, as protein quantities can be affected by other processes such as proteasomal degradation, ribosome drop-off along the length of the transcript, the gene order, the order and type of self-cleaving peptides, and the nature of protein additions adjacent to self-cleaving peptides, amongst others [11,22]. Nevertheless, even if the expression ratio is not precisely 1:1 the stoichiometry tends to be fixed and stable, conferring a much higher coordination of gene expression compared to other co-expression strategies.

Due to this characteristic, 2As have been used in the development of ratiometric reporter systems. For example, Samalova and co-workers studied membrane trafficking in tobacco by linking a trafficked marker to a reporter marker (ie YFP-2A-secGFP), in the presence or absence of dominant mutants of effector proteins [23]. The 2A system allowed the expression of the reference fluorophore (YFP) at a constant ratio to the trafficked marker (secGFP). This was useful to normalize fluorescence of the trafficked marker to the reference, and in this way account for cell-to-cell variability in expression levels. The 2A system has been used to further study membrane trafficking in plants by linking reporter proteins to small GTPase proteins [24]. Similarly, it has been used for ratiometric analysis to investigate auxin dynamics in Arabidopsis protoplasts [25], and to study the degradation of karrikins in *Nicotiana benthamiana* [26]. Khosla and co-workers have developed a system of Gateway-compatible vectors termed pRATIO, which include a Gateway cassette for cloning a gene of interest upstream of a target fluorophore and a reference fluorophore separated by F2A [26]. These may facilitate the generation of constructs for ratiometric studies using the 2A system.

2.1. 2A self-cleaving peptides and subcellular targeting

During the design of polypeptides, the compatibility between self-cleaving peptides and different subcellular targeting signals must be considered when optimizing a gene stacking strategy. El Amrani and co-workers expressed an ER targeted protein with a cytosolic protein in tobacco (erGFP-F2A-ble), and found that the dual localization was compatible with the 2A system [27]. They also tested the co-expression of a cytosolic protein, with a protein anchored to the ER membrane in the downstream position (GUS-F2A-F5H). With this they showed that the 2A system allowed targeting to the ER when the ER targeted protein was found either in the anterior or posterior position of the polypeptide [27]. Additionally, they tested the localization of a chloroplast-targeted protein placed both in the anterior or posterior position in combination with a cytosolic protein, and found correct subcellular localization of both proteins [27].

Samalova and co-workers used F2A to study membrane trafficking, and suggested a model in which when the upstream protein localizes to the cytosol and the downstream protein has a signal peptide directing it to the ER membrane, the signal peptide is sufficient to target and cleave, without the requirement of including a self-cleaving peptide [23]. However, if both proteins translocate through the ER membrane they found that the self-cleaving peptide was necessary in order to avoid degradation. They suggested that the self-cleaving peptide orients the second signal peptide correctly to allow translocation through the ER to continue [23]. In addition, they found that ER targeted RFP was missorted to the vacuole in the presence of 2A (spRFP-F2A). This should be considered during polypeptide design. It is not known whether other 2A sequences (see Table 1) produce missorting when fused C-terminally to an RFP moiety. A later study, which also used 2As to study membrane trafficking, found that bicistronic 2A polypeptides could target to the Golgi and to the ER [24].

3. Design considerations of 2A polypeptides

3.1. Combining multiple 2A sequences

Several 2A sequences of different viral origin have been identified (Table 1) [28]. The effectiveness of self-cleaving peptides for protein co-expression lies in their ability to efficiently produce cleavage between their C-terminal glycine and proline residues, and to continue translation. In most cases the cleavage efficiency is below 100%, yielding some proportion of the uncleaved product, which may be subject to degradation or comprise a dysfunctional fusion protein. Moreover, after cleavage, ribosome drop-off may occur, resulting in lower translation of the downstream protein. In spite of these, 2As are regarded as a highly efficient system for co-expression, with high translation rates of the downstream protein, compared with other systems such as IRES [9].

The first identified and most commonly used peptide for synthetic biology purposes is F2A [29]. T2A and P2A have later emerged as sequences with higher cleavage and translation

efficiencies in bi-cistronic constructs [14,30]. They have also frequently been used in combination to generate tri-cistronic vectors [31]. Liu and co-workers performed a systematic comparison varying the gene position and the type of 2A peptide in bi-, tri and quad-cistronic constructs using mammalian cell lines [11]. They found that in bi-cistronic constructs the highest efficiency of expression was obtained using T2A, but in all cases the protein in the second position was found to have substantial lower expression. Therefore, the design of polypeptides could take into account the generally stronger expression of the first position. In the case of tri-cistronic constructs, they also found strongest expression of the first position and the best results obtained with the use of T2A followed by P2A [11]. Finally, for quad-cistronic constructs, they obtained the best expression of all proteins using T2A, followed by P2A and finaly E2A (the same combination shown in Figure 2), and they found a negative effect on repeating a same 2A sequence within a polypeptide [11]. Therefore, the repetition of the same self-cleaving peptide multiple times is not advised as it may hinder expression levels and result challenging during synthesis or cloning.

Cleavage and expression efficiency of self-cleaving peptides has also been reported to be potentially affected by the identity of the adjacent sequences, and therefore may result case-specific [9,11]. In particular, the addition of a glycine-serine-glycine (GSG) spacer following the upstream gene (Table 1) has been shown to increase the cleavage efficiency of 2A sequences [32]. In view of the context-dependent results of 2As, testing the polypeptide with the genes of interest and fluorescent proteins (as shown in Figure 2), may give an idea of the efficacy of the polypeptide strategy for the specific genes of interest. Moreover, western blots can be performed to determine the cleavage efficiency of the polypeptide (by observing or not a band corresponding to the uncleaved product). Western blots can be carried out using antibodies against the proteins of interest, alternatively anti-2A antibodies are available which bind to the C-terminal conserved region of 2A sequences.

Finally, the polypeptide design should also take into account a small number of added residues which become incorporated to the upstream protein (expanded in Section 3.2).

Table 1. Commonly used self-cleaving peptide sequences for biotechnology.

Different self-cleaving peptide sequences [28] showing the conserved C-terminal region (marked in red), and the cleavage site (marked with an asterisk). The GSG spacer [32] may be included to increase cleavage efficiency.

Name	Derived from	Sequence
F2A	Foot-and-mouth disease virus	(GSG) VKQTLNFDLLKLAGDVESNPG*P
P2A	Porcine teschovirus-1	(GSG) ATNFSLLKQAGDVEENPG*P
T2A	Thosea asigna virus	(GSG) EGRGSLLTCGDVEENPG*P
E2A	Equine rhinitis A virus	(GSG) QCTNYALLKLAGDVESNPG*P

We explored the use of different 2A sequences to recombinantly co-express three integral membrane proteins from the bacterial particulate methane monooxygenase (pMMO) enzyme in tobacco plants (proteins PmoC, PmoA, and PmoB – in the order they appear in the bacterial operon). PmoC and PmoA localized to the ER when recombinantly expressed in tobacco, due to their targeting and transmembrane domains. However, PmoB was incorrectly localised, therefore its bacterial signal peptide was replaced by a plant signal peptide from pumpkin (*Cucurbita* sp) 2S albumin [33] to target it to the ER, yielding csPmoB$_{40\text{-}431}$. We produced three polypeptides containing the pMMO subunits separated by self-cleaving peptides and found that all three subunits localized to the ER (Figure 2). In one of the polypeptides (RFP-CABClover), a C-terminal Clover fluorophore acted as a cytosolic reporter to confirm expression of the complete polypeptide (Figure 2). This showed that tandem targeting to the secretory pathway is compatible with the 2A system, and the cytosolic expression of Clover showed translation throughout the quad-cistronic polypeptide as well as correct cleavage of

3.2. Evaluation and removal of protein additions

The mode of function of 2A self-cleaving sequences is such that a short number of residues (all residues in the sequence except the C-terminal proline) become incorporated to the C-terminus of the upstream protein, whilst the proline residue becomes added to the N-terminus of the downstream protein. Therefore, the polypeptide design must contemplate whether these residue additions may impact upon protein structure or hinder function in any way, such as by interfering with binding to co-factors, disrupting protein-protein interactions or altering subcellular targeting.

As an example, crystallography data shows that the pMMO enzyme forms a heterotrimer, which is repeated three times to form a nonamer [34]. The sequence of the RFP-CABClover polypeptide (Figure 2), implies that each of the three subunits will have 20-22 added residues to its C-terminus, as they are each upstream of a self-cleaving peptide. Therefore, analysis of the quaternary structure of pMMO was carried out to evaluate whether these additions could potentially interfere with the complex assembly, and consequently hinder enzymatic function.

Protein modelling was performed to predict the structure of the subunits with the added residues. One approach used SWISS-MODEL. However, this approach only produced a structure for the regions of the sequence which aligned sufficiently with the template (PDB ID 3RFR, [35]), and the protein additions were not modelled. Other bioinformatic tools used tested included I-TASSER and Phyre2, which combine template based modelling with ab initio folding, for regions without a template. These software packages produced full models of the subunits with the additions, however, the confidence score for the additional regions was extremely low. Unsurprisingly, ab initio modelling is less reliable for protein structure prediction in comparison to template based models [36]. This is especially true for highly flexible regions which can adopt multiple conformations, such as is the case for the self-cleaving peptides, which are short and have a high content of glycine, serine and alanine residues. Therefore, modelling approaches were not considered appropriate to provide a confident result for the folding of the subunits with the protein additions. Instead, the quaternary structure of pMMO

was examined using PyMOL to evaluate whether the additions would be located at sterically restricted regions which could interfere with the native interactions.

The C-termini of PmoC and PmoB face the outside of the nonamer structure, thus, additions to these protein termini would likely not compromise interactions with the other subunits (Figure 3A). On the other hand, the C-terminus of PmoA faces the core of the nonamer and is in close proximity to PmoC and PmoB from the adjacent trimer (Figure 3A).This suggested that protein additions on this position might interfere with the ability of PmoA to interact properly in the complex. To remove the added residues, the polypeptide sequence was modified to include an **intein domain** upstream of the P2A peptide (Figure 3B). The intein domain possesses hyper-N-terminal autocleavage capacity and if fused upstream of a 2A peptide constitutes a self-excising fragment which leaves no added residues to the proteins [37].

In the case of the pMMO proteins, the N-terminus of PmoB is also a region where additional residues could be detrimental, as the N-terminal histidine is part of a copper-binding site within the protein structure. The signal peptide from *Cucurbita sp.* (Cs) used to target PmoB to the ER leaves three added residues to its N-terminus (Figure 3C). Therefore, a different signal peptide was evaluated, derived from tobacco pathogenesis-related protein 1a (PR1a), to avoid residue additions in this critical region. The signal peptide cleavage site was assessed using **SignalP 6.0** [38]. The sequence was codon optimized for tobacco and synthesized before cloning into a plant expression vector.

4. Protocol for the expression of polypeptide vectors containing fluorescent proteins and ratiometric analysis in plant cells

1. Design polypeptide taking into account the combination of self-cleaving peptides (T2A, P2A, and E2A are recommended for quad-cistronic vectors), and peptide additions to the C-terminus of proteins upstream of a 2A sequence. Add targeting signals to the individual

proteins if needed (SignalP 6.0 [38] and MuLocDeep [39] can be used to predict subcellular localization).

2. Codon-optimize the sequence and synthesize.

3. For cloning using Gateway technology [40], amplify the synthesis product with primers containing Gateway overhangs (forward primer: 5'-GGGGACAAGTTTGTACAAAAAAGCAGGC-TNN-(template-specific sequence)-3′, reverse primer: 5′-GGGGACCACTTTGTACAAGAAAGCTGG- GTN-(template-specific sequence)-3) (*see* **Note 1**).

4. Purify the amplicon by running the PCR product on a 0.7% agarose gel, cutting the band with a clean scalpel and performing gel extraction (i.e using the NEB Monarch® DNA Gel Extraction Kit) (*see* **Note 2**).

5. Perform a BP recombination reaction to incorporate the amplicon to a pDONR221 vector following instructions on the Gateway technology manual [40] (*see* **Note 3**).

6. Stop BP reaction by adding 0.5 µl of Proteinase K. Transform NEB 5-alpha competent *E. coli* cells adding kanamycin (50 mg/L) as selection agent.

7. Check the presence of the insert using colony PCR.

8. Grow a positive colony in LB medium and extract plasmid DNA using a Miniprep Kit. It is recommended to send the sample for sequencing at this step for verification.

9. Perform an LR recombination reaction, similarly to step 5, but using 1 µl of a destination vector (pDEST) vector of choice (i.e containing an N-terminal fluorescent protein fusion, C-terminal fluorescent protein fusion, or no fluorophore) (*see* **Note 4**), 1 µl of pENTRY vector, 1 µl LR clonase and 2 µl of dH2O. This reaction yields an expression vector.

10. Stop the LR reaction and transform *E. coli* as in step 6, using the appropriate antibiotic as a selection agent (this should be different to kanamycin).

11. Repeat steps 7 and 8 (without sequencing).

12. Transform competent *Agrobacterium* cells (strain GV3101) with the expression vector (*see* **Note 5** for generation of competent *Agrobacterium* cells). For transformation, thaw an aliquot of 50 µl cells on ice and add approximately 250 ng of DNA. Incubate cells on ice for 5 minutes, subsequently at -80°C for 3 minutes, and then incubate at a 37°C water bath for 4 minutes. Add 1 ml of LB media without antibiotics and incubate at 28°C whilst shaking for 2 hours. Spread on plates containing 25 mg/L rifampicin and the appropriate plasmid selection agent. Incubate plates at 28°C for three days.

13. Grow a single *Agrobacterium* colony in 3 ml LB medium containing 25 mg/L rifampicin and the selection agent overnight at 28°C with shaking at 180 rpm.

14. Perform transient transformation of tobacco via agroinfiltration as described in [41]. A summarized protocol with basic steps can also be found in [42].

15. Image cells using a confocal laser scanning microscope (CLSM). Cut a small leaf piece (approximately 6 mm^2) a scalpel and place on a glass slide with the abaxial side facing up. Pipette water on top of the specimen and cover with a coverslip. If two fluorophores are present in the construct a ratiometric analysis based on fluorescence can be carried out.

16. For two channel imaging using a Zeiss LSM880, set up a two track mode to reduce spectral bleed-through with line switching and set excitation and detection wavelengths as specified for each fluorophore (i.e. for GFP excitation at 488 nm and detection at 495-550 nm, and for RFP excitation at 561 nm with detection at 570-615 nm) (*see* **Note 6**).

17. Acquire the two-track images adjusting the laser power and gain to capture the full dynamic range (to carry out ratiometric analysis it is important to avoid saturated pixels in either channel, and to maintain the same imaging parameters in different sessions) (*see* **Notes 7-10**).

18. Carry out image analysis using ImageJ. Split the two channel images and perform background subtraction, followed by thresholding of the areas of interest (optional) (*see* **Note 11**).

19. Measure the fluorescence intensity (on an 8-bit scale) for each channel. The ratio of GFP/RFP fluorescence can be calculated and compared between samples.

5. Notes

1. Make sure that the additional nucleotides in the forward primer (NN) do not create a stop codon. Use Q5 High-Fidelity DNA polymerase for this step.

2. For large DNA fragments (>8 kb) it is recommended to add additional water to the dissolving buffer (see NEB Gel Extraction Kit instructions).

3. Volumes can be used at 1/2 of those specified in the manual i.e 1 µl pDONR vector (150 ng/ µl), 1 µl PCR product (150 ng/ µl) 1 µL BP clonase (1 µL), 2 µl dH2O. Reactions are incubated at 25 °C for 1 hour or overnight, and yield a pENTRY vector.

4. Karimi and co-workers created a set of commonly used Gateway-compatible destination vectors for *Agrobacterium*-mediated plant transformation [43]

5. Competent *Agrobacterium tumefaciens* cells (strain GV3101) are made starting with a 5ml pre-culture of cells in LB media incubated overnight at 28°C and 180 rpm, with 25 µg mL−1 rifampicin. This is used to inoculate a larger volume (200 ml) of LB in a sterile beaker, which is incubated overnight in the same conditions, and grown to an OD600 of 0.5 to 1. The culture is split into 50 ml falcon tubes, chilled on ice for 5 minutes, and centrifuged at 3500 rpm for 30 minutes at 4°C. Pellets are resuspended in 15 ml of ice cold CaCl2 solution (20mM, 15% glycerol), then spun at 5000 rpm for 10 minutes at 4°C, and again resuspended in 15 ml of ice cold CaCl2 solution. The competent cells are aliquoted, snap frozen in liquid nitrogen and stored at -80°C.

6. Typically a Zeiss PlanApo 63x/1.46 NA oil immersion objective or a PlanApo 100x/1.46 NA oil immersion objective on the LSM880 are used. Airyscan mode can be used to increase resolution. Typically 512 × 512 images are collected in 8-bit with 2-line averaging at an (x,y) pixel spacing of 20–80 nm.

7. Imaging at two days after infiltration is usually sufficient to observe fluorescence in the ER.

8. Protein expression levels may vary. Large constructs or functional proteins may induce silencing and be expressed at lower levels compared to marker proteins (such as GFP-HDEL).

9. If expression levels are very low consider co-expressing with a silencing suppressor such as P19 in *Nicotiana benthamiana* [44].

10. Replicas can be taken by infiltrating different plants (biological replicas) and acquiring various images within a same leaf piece (technical replicas) (for example 5-20, depending on the observed level of variability within the sample, if there is a lot of cell-to-cell variability a greater number of technical replicas should be acquired, or ROIs containing several cells can be sampled, to obtain a more accurate representation of the sample).

11. Different thresholding algorithms are available in ImageJ which can be tried before deciding which one segments the areas of interest better.

Acknowledgements

This work was supported by a BBSRC Doctoral Training Program PhD studentship to TSR as well as a Leverhulme Trust fund (RGP-2015-1927) and a BBSRC Engineering Biology Breakthrough Award (BB/W011166/1) to VK.

References

1. Halpin C (2005) Gene stacking in transgenic plants – the challenge for 21st century plant biotechnology. Plant Biotechnol J 3:141–55.

2. Ceccon CC, Caverzan A, Margis R, Salvadori JR, Grando MF (2020) Gene stacking as a strategy to confer characteristics of agronomic importance in plants by genetic engineering. Ciência Rural 50.

3. Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S (2011) A modular cloning system for standardized assembly of multigene constructs. PLoS One 6:e16765.

4. Cheo DL, Titus SA, Byrd DRN, Hartley JL, Temple GF, Brasch MA (2004) Concerted assembly and cloning of multiple DNA segments using in vitro site-specific recombination: functional analysis of multi-segment expression clones. Genome Res 14:2111.

5. Hecker A, Wallmeroth N, Peter S, Blatt MR, Harter K, Grefen C (2015) Binary 2in1 vectors improve in planta (co)localization and dynamic protein interaction studies. Plant Physiol 168:776–87.

6. Urwin P, Yi L, Martin H, Atkinson H, Gilmartin PM (2000) Functional characterization of the EMCV IRES in plants. Plant J 24:583–9.

7. Ngoi S, Chien A, Lee C (2004) Exploiting internal ribosome entry sites in gene therapy vector design. Curr Gene Ther 4:15–31.

8. Gouiaa S, Khoudi H, Leidi EO, Pardo JM, Masmoudi K (2012) Expression of wheat Na +/H + antiporter TNHXS1 and H +- pyrophosphatase TVP1 genes in tobacco from a bicistronic transcriptional unit improves salt tolerance. Plant Mol Biol 79:137–55.

9. Minskaia E, Ryan MD (2013) Protein coexpression using FMDV 2A: effect of “linker” residues. Biomed Res Int 1–12.

10. Donnelly MLL, Luke G, Mehrotra A, Li X, Hughes LE, Gani D, et al. (2001) Analysis of the aphthovirus 2A/2B polyprotein “cleavage” mechanism indicates not a proteolytic reaction, but a novel translational effect: A putative ribosomal “skip.” J Gen Virol 82:1013–25.

11. Liu Z, Chen O, Wall JBJ, Zheng M, Zhou Y, Wang L, et al. (2017) Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. Sci Rep 7:1–9.

12. Luke GA, Roulston C, Tilsner J, Ryan MD (2015) Growing uses of 2A in plant biotechnology. Biotechnology InTech.

13. Ha SH, Liang YS, Jung H, Ahn MJ, Suh SC, Kweon SJ, et al. (2010) Application of two bicistronic systems involving 2A and IRES sequences to the biosynthesis of carotenoids in rice endosperm. Plant Biotechnol J 8:928–38.

14. Kim JH, Lee SR, Li LH, Park HJ, Park JH, Lee KY, et al. (2011) High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. PLoS One 6.

15. Halpin C, Cooke SE, Barakate A, Amrani A El, Ryan MD (1999) Self-processing 2A-polyproteins - a system for co-ordinate expression of multiple proteins in transgenic plants. Plant J 17:453–9.

16. Ye X, Al-Babili S, Klöti A, Zhang J, Lucca P, Beyer P, et al. (2000) Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science 287:303–5.

17. Paine JA, Shipton CA, Chaggar S, Howells RM, Kennedy MJ, Vernon G, et al. (2005) Improving the nutritional value of Golden Rice through increased pro-vitamin A content. Nat Biotechnol 23:482–7.

18. Liao J, Liu T, Xie L, Mo C, Huang X, Cui S, et al. (2022) Plant metabolic engineering by

multigene stacking: synthesis of diverse mogrosides. Int J Mol Sci 23:10422.

19. Liu X, Yang W, Mu B, Li S, Li Y, Zhou X, et al. (2018) Engineering of 'Purple Embryo Maize' with a multigene expression system derived from a bidirectional promoter and self-cleaving 2A peptides. Plant Biotechnol J 6:1107–9.

20. Wang J, Chen H (2020) A novel CRISPR/Cas9 system for efficiently generating Cas9-free multiplex mutants in Arabidopsis. aBIOTECH 1:6–14.

21. Lin Y, Hung CY, Bhattacharya C, Nichols S, Rahimuddin H, Kittur FS, et al. (2018) An effective way of producing fully assembled antibody in transgenic tobacco plants by linking heavy and light chains via a self-cleaving 2A peptide. Front Plant Sci 9:1379.

22. Reinhardt A, Kagawa H, Woltjen K (2020) N-terminal amino acids determine KLF4 protein stability in 2A peptide-linked polycistronic reprogramming constructs. Stem Cell Reports 14:520–7.

23. Samalova M, Fricker M, Moore I (2006) Ratiometric fluorescence-imaging assays of plant membrane traffic using polyproteins. Traffic 7:1701–23.

24. Burén S, Ortega-Villasante C, Ötvös K, Samuelsson G, Bakó L, Villarejo A (2012) Use of the Foot-and-Mouth Disease Virus 2A peptide co-expression system to study intracellular protein trafficking in Arabidopsis. Caplan S, editor. PLoS One 7:e51973.

25. Wend S, Dal Bosco C, Kämpf MM, Ren F, Palme K, Weber W, et al (2013) A quantitative ratiometric sensor for time-resolved analysis of auxin dynamics. Sci Reports 31 3:1–7.

26. Khosla A, Rodriguez-Furlan C, Kapoor S, Van Norman JM, Nelson DC (2020) A series of dual-reporter vectors for ratiometric analysis of protein abundance in plants. Plant Direct 4:e00231.

27. El Amrani A, Barakate A, Askari BM, Li X, Roberts AG, Ryan MD, et al. (2004) Coordinate expression and independent subcellular targeting of multiple proteins from

a single transgene. Plant Physiol 135:16–24.

28. Szymczak AL, Vignali DAA (2005) Development of 2A peptide-based strategies in the design of multicistronic vectors. Expert Opin Biol Ther 5:627–38.

29. Ryan MD, King AMQ, Thomas GP (1991) Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. J Gen Virol 72:2727–32.

30. Szymczak AL, Workman CJ, Wang Y, Vignali KM, Dilioglou S, Vanin EF, et al. (2004) Correction of multi-gene deficiency in vivo using a single “self-cleaving” 2A peptide–based retroviral vector. Nat Biotechnol 22:589–94.

31. Wang L, Liu Z, Yin C, Asfour H, Chen O, Li Y, et al. (2015) Stoichiometry of Gata4, Mef2c, and Tbx5 influences the efficiency and quality of induced cardiac myocyte reprogramming. Circ Res 116:237–44.

32. Wang Y, Wang F, Wang R, Zhao P, Xia Q (2015) 2A self-cleaving peptide-based multi-gene expression system in the silkworm *Bombyx mori*. Sci Rep 5:16273.

33. Hara-Hishimura I, Takeuchi Y, Inoue K, Nishimura M (1993) Vesicle transport and processing of the precursor to 2S albumin in pumpkin. Plant Journal 4:793–800.

34. Lieberman RL, Rosenzweig AC (2005) Crystal structure of a membrane-bound metalloenzyme that catalyses the biological oxidation of methane. Nature 43:177–82.

35. Smith SM, Rawat S, Telser J, Hoffman BM, Stemmler TL, Rosenzweig AC (2011) Crystal structure and characterization of particulate methane monooxygenase from *Methylocystis* species strain M. Biochemistry 50:10231–40.

36. Lee J, Freddolino PL, Zhang Y (2017) Ab initio protein structure prediction. From Protein Structure to Function with Bioinformatics: Second Edition p. 3–35.

37. Zhang B, Rapolu M, Kumar S, Gupta M, Liang Z, Han Z, et al. (2017) Coordinated

protein co-expression in plants by harnessing the synergy between an intein and a viral 2A peptide. Plant Biotechnol J 15:718–28.

38. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. (2022) SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol 40:1023–5.

39. Jiang Y, Wang D, Yao Y, Eubel H, Künzler P, Møller IM, et al. (2021) MULocDeep: A deep-learning framework for protein subcellular and suborganellar localization prediction with residue-level interpretation. Comput Struct Biotechnol J 19:4825–39.

40. Gateway ® Technology (2003) A universal technology to clone DNA sequences for functional analysis and expression in multiple systems Vol. 70, Invitrogen.

41. Sparkes IA, Runions J, Kearns A, Hawes C (2006) Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. Nat Protoc 1:2019–25.

42. Spatola Rossi T, Pain C, Botchway SW, Kriechbaumer V (2022) FRET-FLIM to determine protein interactions and membrane topology of enzyme complexes. Curr Protoc 2:e598.

43. Karimi M, Inzé D, Depicker A (2002) GATEWAY™ vectors for Agrobacterium-mediated plant transformation. Trends Plant Sci 7:193–5.

44. Voinnet O, Rivas S, Mestre P, Baulcombe D (2003) An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. Plant J 33:949–56.

Figure legends

Figure 1. Methods for multigene engineering in plants through *Agrobacterium*-mediated transformation.

Schematic diagram showing different co-expression techniques: (A) co-expression from separate plasmids, (B) co-expression within a same T-DNA region, (C) polypeptide mediated by IRES with two different translation sites, (D) polypeptide mediated by self-cleaving peptide with co-translational cleavage. LB= left border, RB= right border, P= promoter, T= terminator.

Figure 2. Example of subcellular targeting to the ER using self-cleaving peptides.

(A i) Diagram of polypeptide containing all three subunits of the bacterial pMMO complex and Clover, with RFP fused to the N-terminal protein. The expected localization of each protein is specified to the right, and in bold are the localizations that can be verified by microscopy. (A ii) Polypeptide containing all three subunits with fusion of GFP to the ER targeted pmoB ($csPmoB_{40-431}$). (A iii) Polypeptide containing the first two subunits with fusion of GFP to PmoA. (B i) Confocal images of RFP-CABClover following transient transformation of tobacco, showing localization in the ER of PmoC (magenta), and cytosolic localization of Clover (green), indicating correct cleavage of E2A. (B ii) Confocal images of RFP-CAB-GFP showing localization in the ER of both PmoC (magenta) and $csPmoB_{40-431}$ (green). (B iii) Confocal images of RFP-CA-GFP showing localization in the ER of both PmoC (magenta) and PmoA (green). Overall, this shows that the three subunits can be targeted to the ER in tandem, and Clover can act as a cytosolic marker for expression that is not attached to any subunits. Scale bar = 2 μm.

Figure 3. Removal of added 2A sequence residues.

(A) pMMO nonamer structure (PDB ID 3RFR), with one heterotrimer colored with PmoC in dark red, PmoA in yellow and PmoB in green, the other two trimers are colored in slate. Arrows indicate the position of the C-terminus of each subunit. (B) Modified pMMO polypeptide which contains a self-excising intein domain [37]. Black dotted arrows indicate the cleavage sites of the self-excising domain and the blue dotted arrow indicates the cleavage site of the PR1 signal peptide. (C) Amino acid sequences of the signal peptide from pumpkin 2S albumin (Cs)

and from tobacco pathogenesis-related protein 1a (PR1) for targeting to the secretory pathway, showing the cleavage sites with an asterisk.