Coordinated protein co-expression in plants by harnessing the synergy between an intein and a viral 2A peptide

Bei Zhang1, Madhusudhan Rapolu1, Sandeep Kumar2, Manju Gupta2, Zhibin Liang1, Zhenlin Han1, Philip Williams3 and Wei Wen Su1,*

1Department of Molecular Biosciences and Bioengineering, University of Hawaii at Manoa, Honolulu, HI, USA
2Dow AgroSciences LLC, Indianapolis, IN, USA
3Department of Chemistry, University of Hawaii at Manoa, Honolulu, HI, USA

Received 26 August 2016; revised 16 November 2016; accepted 19 November 2016.
*Correspondence (Tel 1 808 956 3531; fax 808 956 3542; email wsu@hawaii.edu)

Keywords: FMDV 2A, gene and trait stacking, intein, molecular farming, protein expression, production of protein complex.

Introduction

Coordinated expression of multiple proteins in plants is essential for unravelling fundamental cellular mechanisms as well as development of next-generation crops with improved traits. By co-expressing multiple proteins, crops have been successfully modified to acquire enhanced abiotic tolerance, improved pathogenic resistance, as well as enriched nutritional contents (Arvinth et al., 2010; Chen et al., 1998; Sun et al., 2012). Currently, only a limited number of approaches are available for co-expression of multiple proteins in plants. Cotransformation of multiple monosticronic expression cassettes and crossing of different transgenic events containing single expression cassettes are the most prevalent approaches to create genetically modified crops with multiple stacked transgenes (James, 2010). However, these approaches require laborious screening, breeding and introgression processes. Moreover, coordinated expression of the resulting proteins often necessitates extensive tuning of the promoter and regulatory elements. Other notable techniques for multigene co-expression or gene stacking include those that are based on polycistronic or polyprotein vectors. Gene stacking in plants using polycistronic transgenes mainly operates on internal translational initiation mediated by internal ribosome entry sites (IRES) (François et al., 2002b; Ha et al., 2010). However, IRES-mediated translational initiation is less efficient compared with that of the 5′-cap-mediated initiation and results in uneven protein co-expression (François et al., 2002b; Ha et al., 2010; Mizuguchi et al., 2000).

In co-expression of multiple proteins from a polyprotein transgene, the constituent POIs are released from the polyprotein precursor during or after protein translation. Some of the polyprotein expression systems have exploited endogenous plant protease activity to liberate multiple POIs connected by protease substrate sequences (François et al., 2002a; Urwin et al., 1998; Walker and Vierstra, 2007; Zhang et al., 2011). However, processing of the polyproteins in this case can only occur in a particular cellular compartment within hosts where the specific proteases are located. Alternatively, incorporating an exogenous protease within the polyprotein system is used to overcome this problem. Some notable examples include polyprotein vectors that employ Nla protease recognition sequence together with the Nla proteinase from the tobacco etch virus (Marcos and Beachy, 1994, 1997) and a similar vector that employs a Nla proteinase from the tobacco vein mottling virus (Dasgupta et al., 1998). While proper polyprotein processing was achieved using these vectors, the protease expression levels were generally low. With the protease-based polyprotein approaches, formation of a properly folded polyprotein precedes the proteolytic processing to release the individual protein units. This could be especially problematic as the number of proteins to be expressed increases...
which results in very large polyproteins. To this end, a strategy that involves cotranslational protein cleavage would be more desirable, to avoid potential misfolding of the large polyproteins.

Polyprotein expression based on the unique ribosome skipping mechanism of the FMDV 2A (F2A) peptide (Donnelly et al., 2001b) operates cotranslationally. The F2A peptide has been used to direct multiprotein co-expression in a wide range of eukaryotic hosts (de Felipe, 2004; de Felipe et al., 2006). However, the 'remnant' 2A residues appended to the carboxyl terminus of the processed proteins could hinder protein activity and/or cellular targeting (François et al., 2004; Randall et al., 2004; Samalova et al., 2006). Removal of the extraneous 2A residues using host endogenous proteases has been attempted in plant (François et al., 2002a) and mammalian systems (Fang et al., 2005), yet the requirement of specific endogenous proteases and inability to completely avoid appending remnant protease substrate linker residues to the cleaved POIs have significantly limited its general usefulness. To resolve these problems, we have exploited in vivo self-excision of the 2A sequence extension via intein-mediated N-terminal autocleavage, by fusing an engineered mini-intein with the 2A sequence through a linker to create the 'IntF2A' self-excising domain.

Inteins mediate protein splicing in which a portion of the protein excises itself while ligating flanking protein sequences. The protein splicing element is the ‘intein’, while the protein sequences flanking the intein sequence are termed ‘exteins’. By mutating the essential C-terminal asparagine to alanine (N159A), inteins can be modified to boost their autocatalytic N-terminal autocleavage activity (i.e. cleave off protein flanking the intein’s N-terminus), with essentially diminshed splicing activity (Amitai et al., 2009; Xu and Perler, 1996). The N-terminal autocleavage efficiency can also be modulated by amino acid residues in the flanking extein regions.

Unlike other existing polyprotein vector technologies, the IntF2A-based approach enables cotranslational ‘cleavage’ via 2A’s translational recoding activity, followed by very efficient and rapid post-translational autocatalytic cleavage via intein at its N-terminal junction, and it does not require the presence of any host-specific proteases or cofactors. As such, this approach can potentially be applicable across a broad range of hosts. Also, the IntF2A-mediated in vivo polyprotein autoprocessing is not affected by the subcellular location of the protein. The present work provides detailed characterization of the IntF2A-based polyprotein expression system in plants for coordinating co-expression of multiple functional proteins, differential cellular targeting of processed proteins and production of complex protein products (by demonstrating synthesis of a functional IgG antibody).

Results
Processing of the IntF2A-based polyprotein in plants

IntF2A-based polyprotein cassettes (summarized in Figure 1) were assembled by connecting an upstream POI (POI1) and a downstream POI (POI2) with the intervening IntF2A autoprocessing domain that enables self-excision at both terminal junctions (Figure S1). To maximize the 2A activity, a 58aa F2A sequence that includes 39 aa from the C-terminal portion of the

Figure 1 IntF2A-based polyprotein cassettes investigated in this study. The intein domains are boxed and marked in orange. Intein terminal residues are underlined. Linker sequences are shown in blue wherein N-terminal autocleavage accelerating flanking extein residues are shown in underline bold letters. F2A: 2A peptide from foot-and-mouth disease virus; SP1: Arabidopsis thaliana basic chitinase signal peptide; SP2: rice a-amylase signal peptide. GFP172: Green fluorescent protein with an internal hexa-histidine Tag between amino acid residues 172 and 173. RFPStrep: monomeric cherry fluorescent protein with a C-terminal Strep Tag. mKO1FLAG: monomeric Kusabira-Orange 1 fluorescent protein with a C-terminal FLAG Tag. Anti-His Lc: light chain of anti-His Tag antibody; Anti-His Hc: heavy chain of anti-His Tag antibody.
1D capsid protein preceding the 2A was used (Donnelly et al., 2001a). After cotranslational F2A-mediated release of PO1, PO1 can be liberated by the N-terminal cleavage activity of the Ssp DnaE mini-intein with an N159A mutation. Processing of the IntF2A-based polyprotein in plants was initially characterized using Western blot analysis of the total protein extract from tobacco NT1 cells expressing the ND-1 polyprotein cassette (Figure 1). As shown in Figure 2a, essentially complete release of both POIs, that is GFP172 and RFPStrep, was observed when the samples were probed with anti-GFP or Strep Tag antibodies. The processed proteins migrated to the same position as purified protein standards (~28 kDa). The lower immunoreactive band on the Strep Tag Western blot resulted from hydrolysis of the acylimine bond at the RFP chromophore under the denaturing condition imposed by the sample heating step (Campbell et al., 2002).

Similar to undifferentiated tobacco NT1 cells, when the ND-1 polyprotein was expressed in Nicotina tabacum cv. Xanthi plants, efficient release of both upstream GFP172 and downstream RFPStrep was detected in leaf, stem and root extracts (Figure 2b). Aside from tobacco, efficient processing of the ND-1 polyprotein was observed in Nicotiana benthamiana and Romaine lettuce (Lactuca sativa L. var. longifolia) based on transient expression via agroinfiltration (Figure 2c,d). These results support the general utility of the IntF2A polyprotein system in a wide range of plant species for efficient coordinated production of multiple proteins. When examined using fluorescence microscopy, tobacco NT1 cells expressing ND-1 displayed bright fluorescence (Figure 7d). Characteristic GFP and RFP spectra, distinctive from the background autofluorescence of untransformed wild-type control, were also observed in the protein extracts of transgenic tobacco cells (Figure 3). Together, these results confirmed that constituent proteins are functional upon release from the IntF2A-based polyprotein precursor.

As modification of cellular pathways in plants often requires manipulation of multiple enzymes simultaneously, capability of the IntF2A system for co-expressing more than two POIs was investigated. Here co-expression of three proteins, that is GFP172, monomeric Kusabira-Orange 1 fluorescent protein with a FLAG Tag at the C-terminus (mK01FLAG), and RFPStrep, using the IntF2A polyprotein system, was examined in tobacco NT1 cells (ND-2; Figure 1). Essentially, complete release of all three proteins was confirmed by Western blots probed with anti-GFP, FLAG Tag and Strep Tag antibodies, respectively (Figure 4a–c). Similar to the observation in the two-protein co-expression construct, ND-1, all processed fluorescent reporters from ND-2 were verified to be functional based on fluorescence spectroscopy (Figure 4d–f).

In the IntF2A-based polyprotein system, constituent POIs were translated from a single transcript; thus, it is possible that stoichiometric protein co-expression can be achieved by the system. Figure 5 shows that processed GFP and RFP from ND-1 accumulated to similar levels inside the cells. Similarly, all three proteins expressed from ND-2 accumulated to a similar level (Figure 5). In addition, co-expression using the IntF2A polyprotein system does not compromise expression levels when compared with those using single-protein vectors, as shown in Figure 5 (in single-protein vectors, GFP172 or RFPStrep alone was expressed from the (ocs)/mas promoter).

---

**Figure 2** Characteristics of co-expressing two fluorescent proteins from the IntF2A-based polyprotein cassette ND-1 in plants. Efficient in vivo autocleavage and release of the fluorescence reporters in (a) tobacco NT1 cells, (b) different organs of Nicotina tabacum plants, (c) leaf tissue of Nicotiana benthamiana and (d) leaf tissue of Romaine lettuce, shown using Western blots probed with anti-GFP (left panel) and anti-Strep Tag (right panel) antibodies for detecting released upstream and downstream proteins of interest, respectively. Hereinafter, ‘M’ & ‘Wt’ denote molecular marker and nontransformed wild-type control, respectively.
Determination of processing sites within the IntF2A-based polyprotein

To investigate whether the observed cellular processing of polyprotein precursors indeed resulted from autoprocessing activity of the IntF2A fusion domain, N-terminal cleavage activity of the IntF2A domain was blocked by introducing a C1A mutation in the intein sequence (N(-)-1 in Figure 1). In addition, the intein sequence was flanked by amino acid residues Leu-Glu-Tyr and Gly-Gly-Ser-Arg at the N- and C-terminal junctions to further diminish its autocleavage activity. As shown in Figure 6a,b, release of PO1 was impaired by these mutations, whereas efficient release of the PO2 was still preserved. This result confirmed that the N-terminal cleavage activity of the Ssp DnaE intein domain was solely responsible for release of the PO1.

The exact processing sites within the IntF2A polyprotein were determined by N-terminal sequencing and ESI-TOFMS of the processed POIs purified from the extract of transgenic NT1 cells expressing ND-1. N-terminal amino acid sequencing indicated that the released RFPStrep (POI2) has a proline residue directly upstream of its native N-terminal sequence (VSKGEE) (Figure 6c). This result is consistent with the known mechanism of F2A-mediated processing by which peptide bond formation between the last two amino acid residues, glycine and proline, is disrupted during translation. Regarding the PO1, the molecular mass measured using ESI-TOFMS (28 590 Da) matches that of an N-terminal acetylated GFP172 plus the C-terminal N-extein linker (LEGGSKFAND) (calculated mass 28 593 Da) (Figure 6c). Note that after the initiator methionine was removed from GFP172 by methionine aminopeptidase, serine at the N-terminus, being the most common substrate for N-terminal acetyltransferase, is likely to be acetylated. The notion of post-translational modification at the N-terminus of PO1 is in agreement with our N-terminal sequencing result which indicated that the N-terminus was blocked. This result is also consistent with our previous findings of a dual-intein polyprotein expression system when used in plants (Zhang et al., 2015).

Collectively, our results support that the observed cellular processing of the IntF2A-based polyprotein was indeed mediated by the specific actions of intein and F2A sequences.

Subcellular targeting of proteins processed from the IntF2A-based polyprotein

To examine whether proteins released from the IntF2A-based polyprotein precursor can be independently targeted to different

---

**Figure 3** Processed proteins from ND-1 in tobacco NT1 cells retain proper fluorescence property based on fluorescence spectroscopy measurement (a,b). Characteristic fluorescence spectra were detected in ND-1 extracts but not in the untransformed wild-type controls (WT).

**Figure 4** Co-expression of three POIs using the IntF2A-based polyprotein vector system. Efficient in vivo processing of three POIs was confirmed by Western blots of transgenic NT1 cell extracts expressing ND-2 using anti-GFP (a), anti-mKO1 (b) and anti-Strep Tag (c) antibodies, respectively. Processed POIs in ND-2 show characteristic fluorescence spectra (d–f), which are distinctive from the background autofluorescence in untransformed wild-type controls (WT). POIs, proteins of interest.
cellular compartments, endoplasmic reticulum (ER) targeting signal was integrated upstream of either only POI1 (ND-3) or both POI1 and POI2 (ND-4) in the IntF2A polyprotein cassette. The Arabidopsis thaliana basic chitinase signal peptide (SP1) and the rice α-amylase signal peptide (SP2) were used to direct ER targeting of POI1 (GFP172) and POI2 (RFPStrep), respectively, as depicted in Figure 1. For both ND-3 and ND-4 cassettes, GFP172 was effectively released from the polyprotein precursor (Figure 7a, intracellular lanes), indicating active intein autocleavage in the ER. Moreover, GFP172 was detected in the spent media of both ND-3 and ND-4 suspension cultures, but not in the ND-1 culture media (Figure 7a, extracellular lanes). Effective release of RFPStrep with both ND-3 and ND-4 was also observed (Figure 7b, intracellular lanes) and, as anticipated, extracellular RFP was detected only in the ND-4 culture media (Figure 7b, extracellular lanes). The presence of a proline residue at the N-terminus of the SP2, as a result of F2A action, did not impair the ability of signal peptide to target the RFPStrep to ER, and inclusion of SP2 was necessary to direct the RFP for secretion since no extracellular RFP was found with ND-3. It was also noted that the C-terminal Strep Tag was removed from the RFPStrep in the ND-4 expressing cells (Figure 7c). Differential subcellular localization of the processed POIs was further examined using confocal laser scanning microscopy (Figure 7d). For the cytosolic ND-1 cassette, both GFP172 and RFPStrep localized to the cytosol as well as the nucleus. For small cytosolic proteins such as GFP and RFP, it is well known that they may translocate to the nucleus on their own via nonspecific diffusion across the nuclear pores (Seibel et al., 2007). However, when a signal peptide was appended to the GFP as in ND-3 and ND-4, the processed GFP172 was found localized to the ER network as well as the cell envelope, but not in the nucleus. Subcellular distribution of RFP fluorescence within the ND-3 cells is similar to that of ND-1. Because the translational recoding event mediated by F2A occurs cotranslationally, the presence of SP1 alone (without SP2) in ND-3 does not allow the protein downstream of F2A to enter the ER. For ND-4, while we detected extracellular RFP in the spent media (Figure 7b), RFP fluorescence was also found in the vacuole (Figure 7d).

As many nonstructured C-terminal peptide sequences have been reported to direct vacuole targeting in plants (Xiang et al., 2013), it may explain the observed vacuole-targeting property of the C-terminal F2A extension. To this end, another notable observation from this study is that some of the RFPStrep released from the ND-4 polyprotein accumulated inside the vacuole (Figure 7d, ND-4) although we did also detect extracellular RFP (Figure 7b). In an earlier study of plant membrane traffic using polyproteins (Samalova et al., 2006), secretion of a monomeric GFP (a DsRed variant (Campbell et al., 2002) very similar to the mCherry GFP used in the present study) into the apoplastic space of N. benthamiana leaf was noted when an N-terminal secretory signal peptide was incorporated. Therefore, the C-terminal Strep Tag appended to the GFP in ND-4 may be recognized as a putative vacuole sorting determinant. The Strep Tag sequence on some of the processed RFPStrep might have been degraded along the secretory pathway and avoided recognition by the vacuole sorting.

Figure 5 Comparing expression levels of POIs (GFP, mKO1 and RFP) from IntF2A-based polyprotein vectors vs. single-protein vectors in transgenic tobacco N1 cells. To estimate the POI expression levels, fluorescence intensities of individual POIs in the transgenic cell extracts were corrected for background autofluorescence, by subtracting fluorescence of wild-type extract prepared with the same total soluble protein concentration, followed by conversion to protein concentrations using a calibration curve. Data represent the mean of three transgenic callus replicates ± SD. The high P-values from one-way ANOVA indicated that the constituent POI expression levels in the same construct are not significantly different for either ND-1 or ND-2. TSP, total soluble proteins; POIs, proteins of interest.

Figure 6 Intein and F2A are responsible for autocatalytic release of the upstream and downstream POIs, respectively. (a) GFP Western blot of ND-1 and N(-)-1 cell extracts confirms that the intein domain is solely responsible for release of the upstream POI, GFP172. (b) RFP Western blot shows that inactive N-terminal intein cleavage does not impair C-terminal cleavage mediated by F2A in N(-)-1. (c) In vivo cleavage of the ND-1 polyprotein precursor occurred at the expected sites (indicated by the arrows), as confirmed using N-terminal amino acid sequencing and ESI-TOFMS analysis of released POIs. POIs, proteins of interest.

© 2016 The Authors. Plant Biotechnology Journal published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd., 15, 718-728.
receptor and, as a result, was secreted into the media. To this end, we detected a RFP band on Western blot of extracellular ND-4 samples using an anti-mCherry antibody (Figure 7b, lane 5), but not with an anti-Strep Tag antibody (Figure 7c, lane 5). Evidence from our prior study indicated very active peptidase activity in the secretory system of the tobacco cells that leads to digestion of nonstructured terminal peptide linker extensions (Zhang et al., 2011). For the RFP subpopulations with the intact Strep Tag, once translocated into the plant vacuole, the tag sequence might also be removed. When probed using an anti-mCherry antibody, the band on the Western blot corresponding to RFP in the intracellular extract for ND-4 is smaller than that of ND-1 and ND-3 (Figure 7b, lanes 3–5). Furthermore, the same RFP product is not visible when probed with an anti-Strep Tag antibody (Figure 7c, lane 5).

Processing of the IntF2A-based polyprotein without N-extein linker extension

The IntF2A-based polyprotein system described above resulted in release of POI1 with residual amino acids from the N-extein linker. While proper function of the processed POI1, that is GFP172, was preserved despite addition of this non-native sequence at its C-terminus, other proteins might not tolerate this extension. A previous study (Amitai et al., 2009) reported that intein N-terminal autocleavage can be accelerated by modifying amino acid residues proximal to its N- or C-terminal junctions (cf. Figure 1). Hence, we modified the ND-1 polyprotein cassette by removing the N-extein linker extension and substituting the C+2 and C+3 residues with N-terminal cleavage enhancers, serine and cysteine, respectively (Amitai et al., 2009). Efficient processing of this modified IntF2A-based polyprotein (SC-1; Figure 1) was demonstrated in transgenic tobacco NT1 cells using Western blots probed with both anti-GFP and Strep Tag antibodies (Figure 8a, b). The results suggested that elimination of the N-extein linker did not attenuate processing efficiency of the IntF2A domain. N-terminal amino acid sequencing together with ESI-TOFMS analysis (spectra not shown) on processed proteins revealed that processing of the SC-1 polyprotein occurred at the expected sites, and the resulting processed upstream GFP172 preserved its native C-terminus (Figure 8c).

Production of a functional IgG antibody using the IntF2A-based polyprotein system

To evaluate the applicability of the optimized IntF2A-based polyprotein system for expression of useful multimeric protein complexes, a chimeric anti-His Tag antibody was chosen as a
To construct this chimeric IgG antibody, we combined the variable domains of a murine anti-His Tag ScFv antibody (Kaufmann et al., 2002) with the constant domains of human IgG1 (Dodev et al., 2014). The antibody expression cassette (Figure 1, SC-2) was assembled by flanking the optimized IntF2A autoprocessing domain (without the N-extein linker extension) with the chimeric kappa light chain (as POI1) and gamma heavy chain (as POI2) of the anti-His Tag antibody. The assembled SC-2 was transiently expressed in N. benthamiana leaves via agroinfiltration. The antibody product was analysed 4 days postinfiltration. Processing of the anti-His Tag antibody expressed in N. benthamiana was characterized using Western blot probed with anti-human kappa light chain antibody (a) or anti-human IgG1 antibody (b). (c) Biological activity of the plant-made anti-His Tag antibody was measured based on sandwich ELISA assay against antigen GFP-His. Error bars indicate standard deviation of the mean from triplicate samples.
residual amino acids, which is superior to previously reported approaches (Fang et al., 2005). The native N-terminus of the processed downstream heavy chain fragment should also be preserved as the proline residue introduced by the action of F2A is removed upon signal peptide cleavage.

**Discussion**

In this study, we developed an IntF2A fusion protein domain with hyperactive autoprocessing activity to direct coordinated co-expression of multiple proteins in plants from a single open reading frame encoding a polyprotein precursor. This system has important advantages over the polyprotein expression system that is based on F2A alone. While F2A sequences from 18 to 58 amino acid residues have been used, it is well known that the shorter versions of F2A give lower translational recording efficiencies. On the other hand, when using longer versions of F2A, the F2A extension that remains on the POI’s C-terminus could be highly problematic (Donnelly et al., 2001a; Minskaia et al., 2013). In the IntF2A system reported here, the long 58aa version of F2A is utilized to maximize the efficiency in releasing its downstream protein, but it is removed swiftly by the action of the intein domain to eliminate potential negative effects imposed by the long F2A overhang. We have succeeded in co-expressing as many as three POIs from a single IntF2A-based polyprotein. Efficient cellular processing of the IntF2A-based polyproteins was demonstrated in cultured tobacco NT1 cells, as well as in different organs of transgenic N. tabacum plants, as well as in lettuce and N. benthamiana. POIs released from the polyprotein precursors displayed proper function and accumulated at similar levels. N-terminal amino acid sequencing together with ESI-TOFMS analysis revealed that processing of the IntF2A-based polyprotein is consistent with the known protein processing mechanism mediated by intein and the F2A peptide. By optimizing the C-extein linker residues within the IntF2A domain, released POIs can preserve its native C-terminus. Furthermore, we demonstrated important practical applications of the IntF2A technology beyond just proof of concepts with reporters, by successfully producing a correctly assembled and biologically functional IgG antibody molecule. Mass production of pharmacologically relevant proteins such as antibodies in plants through molecular farming, especially with transient expression systems, has drawn increasing interests due to its simplicity and low-cost nature compared with the conventional microbial and mammalian cell culture production methods (Arntzen, 2015; Ma et al., 2003).

Processing of the IntF2A-based polyprotein is initiated during protein translation, in which peptide bond formation between the last two amino acids of the F2A peptide, that is glycine and proline, is disrupted by the unique translational recoding activity of the F2A sequence (Donnelly et al., 2001b). It has been hypothesized that the translational recoding event requires specific interaction of the ribosome exit tunnel with the nascent 2A peptide to constrain the conformational space of the peptideyl (2A)-tRNA\(^{\text{Phe}}\) ester bond in the ribosome P-site to ‘jam’ further elongation (Minskaia et al., 2013; Roulston et al., 2016; Yan et al., 2010). Since the ribosome exit tunnel can accommodate 30–40 amino acids, the activity of F2A sequences with shorter than 30 residues may be influenced by the C-terminal sequence of the protein upstream (Minskaia and Ryan, 2013). Indeed, it is known that improved translational recoding efficiency can be achieved by employing longer versions of F2A which incorporate additional residues derived from the native viral sequence upstream of the core F2A sequence (Donnelly et al., 2001a; Minskaia et al., 2013). The long C-terminal F2A extension could, however, negatively impact protein conformation, protein trafficking, as well as post-translational modifications and protein activities that require native carboxyl terminus (Minskaia et al., 2013). Using the inF2A domain, this adverse effect is overcome by virtue of the N-terminal autoleavage activity of the intein mutant within the IntF2A fusion domain. While N-terminal intein cleavage is considered a post-translational process, previous kinetic studies revealed that N-terminal intein cleavage rates determined in vitro are much faster than the average protein synthesis rates in plant cells (Amitai et al., 2009; Li et al., 2012; Martin et al., 2001; Saleh et al., 2011; Trewavas, 1972). By incorporating N-terminal cleavage accelerating residues in the flanking extein regions, for example Asn-Asp at intein N-terminal junction or Cys-Ser-Cys bordering the intein C-terminus, cleavage rate was found to increase by fourfold compared to that with the native flanking extein residues (Amitai et al., 2009). Therefore, it is reasonable to assert that release of POI1 from the polyprotein precursor occurs immediately after the intein is folded to allow rapid removal of the IntF2A domain from the C-terminus of POI1 to avoid potential adverse effects inside the cell.

As an exemplified case where the presence of F2A at the C-terminus of a cytosolic protein may negatively impact the protein, we observed that when GFP\(_{\text{172}}\) and RFP\(_{\text{Strep}}\) were separated by the 58aa F2A sequence but without the intein domain (i.e. forming a GFP\(_{\text{172}}\)-F2A-RFP\(_{\text{Strep}}\) polyprotein termed ‘2A-1’), the amount of GFP\(_{\text{172}}\)-F2A detected was much lower than that of RFP\(_{\text{Strep}}\) (Figure S2a). In addition to fluorescence-based measurement, the observed low GFP/RFP ratio was confirmed by scanning densitometry analysis of Western blots (data not shown). Interestingly, the GFP\(_{\text{172}}\)-F2A fragment appeared smaller than the GFP\(_{\text{172}}\) standard (Figure S2b). These observations suggested that although the C-terminal F2A extension does not hamper the fluorescence function of GFP\(_{\text{172}}\), it might have triggered destabilization of GFP\(_{\text{172}}\), resulting in less protein accumulated and with a lower molecular mass. This problem can be circumvented with the IntF2A approach in which the intein domain enables rapid removal of the C-terminal F2A extension to achieve more balanced production of the constituent POIs (Figure S2a, ND-1 and SC-1).

Correct differential subcellular targeting of the POIs derived from IntF2A polyproteins to cytosol and the secretory system was demonstrated in this study, suggesting the IntF2A approach does not interfere with protein trafficking in plant cells. In the event the C-terminal F2A sequence remained attached to an ER-targeted protein, it might cause erroneous protein sorting (François et al., 2004; Samalova et al., 2006). To this end, we investigated tobacco NT-1 cells expressing the N(\(-\))-2 polyprotein (Figure 1) that consists of a secretory GFP\(_{\text{172}}\) and a cytosolic RFP\(_{\text{Strep}}\), separated by a mutated IntF2A with abolished N-terminal autoleavage activity, that is Int(N)-F2A (containing a C1A mutation in the intein). As expected, we detected GFP\(_{\text{172}}\)-Int(N)-F2A in the cell extract indicating RFP\(_{\text{Strep}}\) was released from the N(\(-\))-2 polyprotein (since F2A was active) but GFP\(_{\text{172}}\) was not (Figure S3a). In the cell extract, we also detected protein species smaller than that of the GFP\(_{\text{172}}\) standard but cross-reacted with the GFP antibody on the Western blot (Figure S3a). However, in the spent culture media, no protein products were detected on anti-GFP Western blot for the N(\(-\))-2 cells (Figure S3b). It is plausible that the lack of detectable extracellular GFP\(_{\text{172}}\)-Int(N-) F2A implied that protein mistargeting might have occurred along
the secretory pathway. This notion is substantiated by the observation of strong GFP fluorescence localized to the vacuole of the tobacco NT1 cells expressing N(-)2 under confocal fluorescence microscopy examination (Figure S3c). The vacuole targeted GFP172-IntN-F2A was partially degraded, probably due to hydrolysis by proteases that reside in the secretory pathway or the vacuole, as multiple protein bands were detected by Western blot analysis (Figure S3a). The undesired vacuole sorting mediated by the F2A sequence has been reported in earlier studies that the secretory protein anterior to the F2A peptide is prone to accumulation in the vacuole (François et al., 2004; Samalova et al., 2006). Conversely, IntF2A with an active intein (e.g. in ND-3 or ND-4) directs efficient protein secretion as shown in Figures 7a and S3b. This again highlights the importance of rapid removal of the C-terminal F2A extension to avoid potential adverse effects, for both ER-targeted and cytosolic proteins. The successful expression and secretion of GFP173 from ND-3 and ND-4 also indicated that the intein autocleavage activity is insensitive to the cellular oxidative environment, since the ER lumen is more oxidative than that of the cytoplasm, and this finding is in agreement with our previous observation (Zhang et al., 2015).

In comparison with the dual-intein-based polyprotein expression system recently reported by our group (Zhang et al., 2015), in which a pair of self-excising mini-intein variants (having N- and C-terminal autocleavage activity, respectively) fused in tandem, the IntF2A approach has a number of important advantages. These include (i) smaller molecular size of the IntF2A domain, (ii) cotranslational cleavage mediated by the F2A peptide, (iii) adding only a single proline residue to the POI trailing the autoprocessing domain (in case of a secretory POI, the proline residue will be removed along with the signal peptide by the signal peptidase and hence the POI can preserve its authentic N-terminus) and (iv) higher efficiency in releasing the downstream POI and hence more balanced co-expression among the POIs in plants. The utility of the IntF2A approach can also be further extended, for instance, by synergistic integration with the bidirectional promoter systems (Kumar et al., 2015) to further increase the number of transgenes that can be co-expressed in plants. In summary, the IntF2A-based polyprotein system enables highly efficient coordinated co-expression of multiple proteins from a single transgene in plant cells and whole plants. Its many unique advantages as described throughout this report make the IntF2A a highly powerful molecular tool for plant sciences and biotechnology.

**Experimental procedures**

**Design and assembly of genetic constructs**

For the ND series of cassettes, asparagine and aspartate as flanking N-extein amino acids have been reported to accelerate N-terminal intein cleavage (Amitai et al., 2009). The C-extein linker located between the intein and the F2A sequence consists of three native C-extein residues (Cys-Phe-Asn) followed by Gly-Ser-Gly-Ser-Arg. For the SC cassettes, the N-extein linker was completely eliminated to preserve the C-terminus of POI1. To enhance the N-terminal cleavage efficiency, the C-extein linker was modified by substituting the native C+2 and C+3 residues phenylalanine and asparagine with serine and cysteine, respectively (Amitai et al., 2009). Fluorescent proteins GFP172 (a GFP variant with a hexa-histidine sequence inserted between amino acids 172 and 173) (Paramban et al., 2004) and RFPStrep (mCherry with a C-terminal Strep Tag) were used as reporters in most polyprotein cassettes (as POI1 and POI2, respectively) for facile detection of protein processing.

Detailed experimental procedures for vector construction, nucleic acid sequence of the IntF2A domain, along with a list of primers used in this study (Appendix S1; Table S1), can be found in the supporting information.

**Plant transformation and protein extraction**

The pe1775 vectors containing IntF2A-based polyprotein sequences were transformed into Agrobacterium tumefaciens C58C1 via electroporation. Stable transformation of tobacco NT1 cells and N. tabacum plants was performed using Agrobacterium cocultivation approach as described in previous publications (Fishier and Guiltnan, 1995; Mayo et al., 2006). Hygromycin was used for transformation selection. Fluorescence and Western blot analysis were used to screen the highest expression lines which were selected for subsequent characterization. Vacuum-assisted agroinfiltration for transient expression in leaf tissues of N. benthamiana and Romaine lettuce was performed as described previously (Zhang et al., 2011).

Total soluble intracellular proteins were extracted in boric acid extraction buffer following a procedure described previously (Peckham et al., 2006). To collect secreted proteins, apoplastic or spent media of transgenic NT1 calli or suspension cultures, respectively, were filtered through a 10-μm nylon mesh filter. The crude filtrates were clarified by filtering through a Whatman #1 filter paper. Total protein concentration was determined using Bradford protein assay (Bradford, 1976) (Bio-Rad, Hercules, CA) and target proteins were tracked by fluorescence measurement using a Hitachi F-2500 fluorescence spectrophotometer (Hitachi High Technologies America, Pleasanton, CA).

**SDS-PAGE and Western blot**

Protein extracts were mixed with SDS-PAGE sample buffer and denatured at 95°C for 5 min under reducing or nonreducing conditions. After brief centrifugation, denatured protein samples were subjected to 12% SDS-PAGE and blotted onto PVDF membranes. Rabbit anti-GFP (Invitrogen, Grand Island, NY) was used to detect GFP172, while RFPStrep was detected by rabbit anti-Strep Tag antibody (Genscript, Piscataway, NJ) or rabbit anti-RFP antibody (Biovision, Milpitas, CA). Rabbit anti-human α light chain and rabbit anti-human IgG antibodies (Abcam, Cambridge, MA) were used to detect light and heavy chains of anti-His Tag antibody, respectively. Alkaline phosphatase-conjugated goat anti-rabbit antibody (Southern Biotech, Birmingham, AL) was used as the secondary antibody. Immunoreactive bands were visualized by NBT/BCIP coupled chromogenic reaction. The GFP172 and RFPStrep protein standards were expressed in *Escherichia coli* BL21 (DE3) and purified using affinity column chromatography as described previously (Zhang et al., 2015).

**Fluorescence confocal microscopy**

Subcellular sorting of processed proteins was analysed using an Olympus Fluoview FV-1000 confocal laser scanning microscope system mounted on an Olympus IX-81 inverted microscope (Nikon, Tokyo, Japan) and performed at the Biological Electron Microscopy Facility at the University of Hawaii. Cell images were observed with a UPLSAPO 20x lens [numerical apertures (NA), 0.70; Nikon, Tokyo, Japan], along with 3X digital zooming. Cells were excited with laser beams at 488 and 543 nm, respectively, for detection of GFP and RFP (mCherry) fluorescence. Filters
BA505-525 and BA560F were used for collecting fluorescence emission of GFP and RFP, respectively.

Protein purification for N-terminal amino acid sequencing and ESI-TOFMS analysis

Processed RFP$_\text{Strep}$ was purified from ND-1 and SC-1 cell extracts using Streptactin chromatography (Qiagen, Hilden, Germany) according to the procedure published previously (Zhang et al., 2015). Unbound flow-through from the Streptactin column was subject to hydrophobic interaction chromatography, followed by immobilized metal affinity chromatography (GE Healthcare, Marlborough, MA) for purification of processed GFP$_{172}$ as described previously (Peckham et al., 2006; Zhang et al., 2015). Purified GFP$_{172}$ and RFP$_\text{Strep}$ were further processed using SDS-PAGE and blotted onto PVDF membrane. Target protein bands were excised for N-terminal amino acid sequencing using Edman degradation approach (performed by the Protein Facility at Iowa State University). ESI-TOFMS analysis of purified processed GFP$_{172}$ protein was carried out as described previously (Zhang et al., 2015).

ELISA assay for chimeric antibody against hexa-histidine tagged proteins

Biological activity of plant expressed anti-His Tag antibody was assayed using sandwich ELISA. Protein extract of agroinfiltrated N. benthamiana leaves was coated in triplicate onto a 96-well plate. After blocking with 1% BSA (in PBS with 0.05% Tween 20), different concentrations (0, 0.625, 1.25, 2.5, 5 and 10 μg/mL) of C-terminal hexa-histidine tagged GFP (GFP$_\text{His}$) were added. Captured GFP$_{\text{His}}$ was detected using rabbit anti-GFP antibody (Cell Sciences, Canton, MA), followed by alkaline phosphatase-conjugated goat anti-rabbit antibody. ELISA signal was generated by incubating with chromogenic substrate p-nitrophospho- phate and measured by absorbance at 405 nm. Data presented in Figure 9 were normalized by subtracting the background generated from the wild-type leaf extract hybridizing with corresponding antigen concentrations.

Acknowledgements

This work was supported in part by the USDA, Hawaii Community Foundation, and Dow AgroSciences, LLC. We thank Professor Ho Leung Ng of the Chemistry Department at University of Hawaii for helpful discussions.

Conflict of interest

B.Z. and W.W.S. are the inventors of a US patent for the autoprocessing domain technology.

References

Amitai, G., Callahan, B.P., Stanger, M.J., Belfort, G. and Belfort, M. (2009) Modulation of intein activity by its neighboring extein substrates. Proc. Natl Acad. Sci. USA, 106, 11005–11010.

Amitzen, C. (2015) Plant-made pharmaceuticals: from ‘Edible Vaccines’ to Ebola therapeutics. Plant Biotechnol. J. 13, 1013–1016.

Arvind, S., Arun, S., Selvakumaran, R.K., Srikanth, J., Mukunthan, N., Ananda Kumar, P., Premachandran, M.N. et al. (2010) Genetic transformation and pyramiding of aprotinin-expressing sugarcane with cry1Ab for shoot borer (Chilo suppressalis) resistance. Plant Cell Rep. 29, 383–395.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248–254.

Campbell, R.E., Tour, O., Palmer, A.E., Steinbach, P.A., Baird, G.S., Zacharias, D.A. and Tsien, R.Y. (2002) A monomeric red fluorescent protein. Proc. Natl Acad. Sci. USA, 99, 7877–7882.

Chen, L., Marmeye, P., Taylor, N.J., Briard, J.-P., Espinoza, C., D’Cruz, P., Huet, H. et al. (1998) Expression and inheritance of multiple transgenes in rice plants. Nat. Biotechnol. 16, 1060–1064.

Dasgupta, S., Collins, G.B. and Hunt, A.G. (1998) Co-ordinated expression of multiple enzymes in different subcellular compartments in plants. Plant J. 16, 107–116.

Dodev, T.S., Karagiannis, P., Gilbert, A.E., Josephs, D.H., Bowen, H., James, L.K., Pax, H.J. et al. (2014) A tool kit for rapid cloning and expression of recombinant antibodies. Sci. Rep. 4, 5985.

Donnelly, M.L.L., Hughes, L.E., Luke, G., Mendoza, H., ten Dam, E., Gani, D. and Ryan, M.D. (2001a) The ‘cleavage’ activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring ‘2A-like’ sequences. J. Gen. Virol. 82, 1027–1041.

Donnelly, M.L.L., Luke, G., Mehrotra, A., Li, X., Hughes, L.E., Gani, D. and Ryan, M.D. (2000b) 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, 1013–1025.

Fang, J., Qian, J.-J., Yang, H., Harding, T.C., Tu, G.H., VanRoye, M. and Jooss, K. (2005) Stable antibody expression at therapeutic levels using the 2A peptide. Nat. Biotechnol. 23, 584–590.

de Felipe, P. (2004) Skipping the co-expression problem: the new 2A “CHYSEL” technology. Genet. Vaccines Ther. 2, 13.

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 Biotechnol. 24, 68–75.

Fisher, D. and Guiltnan, M. (1995) Rapid, efficient production of homologous transgenic tobacco plants with Agrobacterium tumefaciens: a seed-to-seed protocol. Plant Mol. Biol. Rep. 13, 278–289.

François, I.E.J.A., Bolle, M.F.C.D., Dwyer, G., Goderis, L.W.M., Wouters, P.F.J., Verhaert, P.D., Proost, P. et al. (2002a) Transgenic expression in Arabidopsis of a polyprotein construct leading to production of two different antimicrobial proteins. Plant Physiol. 128, 1346–1358.

François, I.E.J.A., Broekaert, W.F. and Carmeau, B.P.A. (2002b) Different approaches for multi-transgene-stacking in plants. Plant Sci. 163, 281–295.

François, I.E.J.A., Van Hemelrijck, W., Aerts, A.M., Wouters, P.F.J., Proost, P., Broekaert, W.F. and Carmeau, B.P.A. (2004) Processing in Arabidopsis thaliana of a heterologous polyprotein resulting in differential targeting of the individual plant defensins. Plant Sci. 166, 113–121.

Ha, S.-H., Liang, Y.S., Jung, H., Ahn, M.-J., Suh, S.-C., Kwone, S.-I., Kim, D.-H. 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, 928–938.

James, C. (2010) A global overview of biotech (GM) crops: adoption, impact and future prospects. GM Crops 1, 1–8.

Kaufmann, M., Lindner, P., Honegger, A., Blank, K., Tschopp, M., Capitani, G., Plank, H. et al. (2002) Crystal structure of the anti-His Tag antibody 3D9 single-chain fragment complexed to its antigen. J. Mol. Biol. 318, 135–147.

Kumar, S., AlAbed, D., Whitemeck, I.T., Chen, W., Berinett, S., Aubrey, A., Wang, X. et al. (2015) A combinatorial bidirectional and bicistronic approach for coordinated multi-gene expression in corn. Plant Mol. Biol. 87, 341–353.

Li, L., Nelson, C.J., Solheim, C., Whelan, J. and Millar, A.H. (2012) Determining degradation and synthesis rates of Arabidopsis proteins using the kinetics of progressive (15)N labeling of two-dimensional gel-separated protein spots. Mol. Cell Proteomics, 11, M111.010025.

Ma, J.K.C., Drake, P.M.W. and Christou, P. (2003) The production of recombinant pharmaceutical proteins in plants. Nat. Rev. Genet. 4, 794–805.

Marcos, J.F. and Beachy, R.N. (1994) In vitro characterization of a caspase to accommodate multiple proteins through synthesis of a self-processing polypeptide. Plant Mol. Biol. 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. J. Gen. Virol. 78, 1771–1778.

Martin, D.D., Xu, M.-Q. and Evans, T.C. (2001) Characterization of a naturally occurring trans-splicing intein from synechocystis sp. PCC6803. Biochemistry, 40, 1393–1402.
Mayo, K.J., Gonzales, B.J. and Mason, H.S. (2006) Genetic transformation of tobacco NT1 cells with Agrobacterium tumefaciens. Nat. Protoc. 1, 1105–1111.

Minskaia, E. and Ryan, M.D. (2013) Protein coexpression using FMDV 2A: effect of “Linker” residues. Biomed Res. Int. 2013, 12.

Minskaia, E., Nicholson, J. and Ryan, M. (2013) Optimisation of the foot-and-mouth disease virus 2A co-expression system for biomedical applications. BMC Biotechnol. 13, 1–11.

Mizuguchi, H., Xu, Z., Ishii-Watabe, A., Uchida, E. and Hayakawa, T. (2000) IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector. Mol. Ther. 1, 376–382.

Paramban, R.I., Bugos, R.C. and Su, W.W. (2004) Engineering green fluorescent protein as a dual functional tag. Biotechnol. Bioeng. 86, 687–697.

Peckham, G.D., Bugos, R.C. and Su, W.W. (2006) Purification of GFP fusion proteins from transgenic plant cell cultures. Protein Expr. Purif. 49, 183–189.

Randall, J., Sutton, D., Ghoshroy, S., Bagga, S. and Kemp, J.D. (2004) Co-ordinate expression of β- and δ-zeins in transgenic tobacco. Plant Sci. 167, 367–372.

Roulston, C., Luke, G.A., de Felipe, P., Ruan, L., Cope, J., Nicholson, J., Sukhodub, A. et al. (2016) ‘2A-Like’ signal sequences mediating translational recoding: a novel form of dual protein targeting. Traffic, 17, 923–939.

Saleh, L., Southworth, M.W., Considine, N., O’Neill, C., Benner, J., Bollinger, J.M. and Perler, F.B. (2011) Branched intermediate formation is the slowest step in the protein splicing reaction of the Ala1 KlbA Intein from Methanococcus jannaschii. Biochemistry, 50, 10576–10589.

Samalova, M., Fricker, M. and Moore, I. (2006) Ratiometric fluorescence-imaging assays of plant membrane traffic using polyproteins. Traffic, 7, 1701–1723.

Seibel, N.M., Eljouni, J., Naslaskowski, M.M. and Hampe, W. (2007) Nuclear localization of enhanced green fluorescent protein homomultimers. Anal. Biochem. 368, 95–99.

Sun, H., Lang, Z., Zhu, L. and Huang, D. (2012) Acquiring transgenic tobacco plants with insect resistance and glyphosate tolerance by fusion gene transformation. Plant Cell Rep. 31, 1877–1887.

Trewavas, A. (1972) Determination of the rates of protein synthesis and degradation in Lemna minor. Plant Physiol. 49, 40–46.

Unwin, P.E., McPherson, M.J. and Atkinson, H.J. (1998) Enhanced transgenic plant resistance to nematodes by dual proteinase inhibitor constructs. Planta, 204, 472–479.

Walker, J.M. and Vierstra, R.D. (2007) A ubiquitin-based vector for the co-ordinated synthesis of multiple proteins in plants. Plant Biotechnol. J. 5, 413–421.

Xiang, L., Etxeberria, E. and Van den Ende, W. (2013) Vacuolar protein sorting mechanisms in plants. FEBS J. 280, 979–993.

Xu, M.Q. and Perler, F.B. (1996) The mechanism of protein splicing and its modulation by mutation. EMBO J. 15, 5146–5153.

Yan, F., Doronina, V.A., Sharma, P. and Brown, J.D. (2010) Orchestrating ribosomal activity from inside: effects of the nascent chain on the peptidyltransferase centre. Biochem. Soc. Trans. 38, 1576–1580.

Zhang, B., Rapolu, M., Huang, L. and Su, W.W. (2011) Coordinate expression of multiple proteins in plant cells by exploiting endogenous kex2p-like protease activity. Plant Biotechnol. J. 9, 970–981.

Zhang, B., Rapolu, M., Liang, Z., Han, Z., Williams, P.G. and Su, W.W. (2015) A dual-intein autoprocessing domain that directs synchronized protein co-expression in both prokaryotes and eukaryotes. Sci. Rep. 5, 8541.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Figure S1 Cellular processing of the IntF2A based polyprotein.

Figure S2 Presence of F2A at the C-terminus of a cytosolic protein may negatively impact the protein.

Figure S3 C-Terminal F2A extension causes mis-targeting of POI to the vacuole when an ER targeting signal is included at the N-terminus of the POI.

Table S1 List of primers used in this study.

Appendix S1 Supplementary experimental procedures and nucleic acid sequence of the IntF2A domain.