SynBio

MDPI

Review

Self-Excising Proteins: Dual-Intein, Intein-2A, and Intein-Ubiquitin for Coordinated Multi-Gene Expression in Synthetic Biology

Kylah Lau and Wei-Wen Su *

Department of Molecular Biosciences and Bioengineering, University of Hawaii at Manoa, Honolulu, HI 96822, USA

* Correspondence: wsu@hawaii.edu

Abstract

The production of multiple proteins using a single open reading frame (sORF)/polyprotein system is a powerful strategy for coordinated multi-protein expression in eukaryotes. The most widely used approach relies on 2A peptides, but conventional 2A systems suffer from several limitations, and their viral origin makes them less than ideal for commercial crop biotechnology applications. Self-excising protein (SEP) modules are a promising alternative that enables coordinated production of multiple proteins from a single ORF encoding a polyprotein precursor. SEPs provide distinct advantages over conventional systems and effectively address many of the limitations inherent to the 2A approach. An SEP module is a fusion protein composed of an N-terminal excising domain (NED) and a C-terminal excising domain (CED) joined by a peptide linker. Using this architecture, a panel of SEP modules has been developed by pairing an engineered intein (serving as the NED) with various CEDs, including a second engineered intein, a 2A-like peptide, and ubiquitin. These modules release multiple proteins from the polyprotein precursor with nearly stoichiometric expression and clean cleavage. Coordinated coexpression using SEPs has been successfully demonstrated in several eukaryotic systems, including yeast, mammalian cells, and plants. This review offers a comprehensive analysis of the SEP technology while underscoring its major applications.

Keywords: gene stacking; metabolic engineering; plant traits; protein engineering; protein expression; trait engineering

1. Introduction

Coordinated expression of multiple recombinant proteins is a core capability across basic and applied life sciences. Many cellular processes run on multi-protein complexes: ribosomes, RNA polymerases, antibodies, viral capsids, ion channels, the proteasome, and so on. Isolated subunits typically have little intrinsic activity on their own, and many will not even fold correctly outside the complex. They only behave properly when their components are produced in roughly the right amounts inside the same cell and have a chance to assemble. That is why studying one of these cellular machines, deploying one as a recombinant system in synthetic biology, or producing one as a therapeutic or vaccine, almost always means expressing its subunits together rather than one at a time. Furthermore, achieving balanced, simultaneous expression of multiple subunits is critical for correct protein folding and complex assembly [1,2]. In antibody production, too

Check for updates

Academic Editor: Jiong Hong

Received: 1 July 2026
Revised: 29 July 2026
Accepted: 30 July 2026
Published: 31 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

much heavy chain relative to light chain gives aggregates instead of functional antibody. Viral capsids assembled from imbalanced subunits come out empty or fail to form at all. Deficiency of enzymes in a metabolic pathway stalls flux and lets intermediates accumulate, sometimes to toxic levels. Beyond ratios, the proteins must be in the same cell (complexes cannot assemble from separate cultures), expressed in compatible time windows, and not subject to silencing, position effects, or competition for folding capacity. Self-cleaving polyprotein architectures solve this coordination challenge, thereby enabling the downstream steps that depend on balanced multi-protein expression [2,3].

Currently three strategies are most commonly practiced for coexpression of multiple proteins in eukaryotic cells: (1) multiple monocistronic expression cassettes on the same or separate vectors; (2) internal ribosomal entry site (IRES) based polycistronic vectors; and (3) 2A peptide or protease substrate sequence based single open reading frame (sORF) polyprotein vectors [3,4]. With respect to the use of multiple monocistronic cassettes for coexpression, it is difficult to coordinate the expression level of each transgene, especially with multiple transgenes on separate vectors. When linked transgene cassettes are used on a single vector, since each gene cassette harbors its own set of regulatory elements, the size of the vector could become quite large especially if the transgene number increases, and could substantially attenuate the genetic transformation or transfection efficiencies. Furthermore, for linked transgenes expressed using the same type of promoter, it does not necessarily result in the same level of expression for each transgene [2].

Polycistronic strategies were developed in part to address these limitations, but introduce their own. Viral IRES-mediated translational initiation is far less efficient than 5′-cap–mediated initiation, leading to highly uneven coexpression in a cell-type–specific manner [5,6]. The 2A viral peptide avoids a separate initiation event by mediating co-translational release from a single ORF, but its efficacy varies with the upstream sequence or protein structure in a manner that remains poorly understood [3,7], and the residual 2A sequence left on the carboxyl terminus of the upstream protein can perturb its folding, targeting, or function [8]. Proteolytic strategies to remove a peptide overhang have been described, for example via endogenous kex2p-like protease activity, but they impose their own compartmental and efficiency constraints [9]. More recently, several nucleic-acid linker–based, IRES-like polycistronic systems have been reported, but they likewise suffer from imbalanced coexpression [10,11].

Collectively, these approaches share a recurring set of deficiencies: (1) inefficiency in precisely coordinating expression among multiple target proteins; (2) a requirement for viral sequences; (3) cumbersome implementation requiring extensive optimization; and (4) an inability to preserve the native terminal residues of the co-expressed proteins. Here we review an emerging alternative, mediated by chimeric self-excising proteins (SEPs), that mitigates several of these deficiencies. Once the SEP design landscape is in view, we proceed in Section 7 to a systematic comparison against IRES- and 2A-based strategies. SEPs are modular fusion proteins composed of an N-terminal excising domain (NED) and a C-terminal excising domain (CED) joined by a peptide linker. Using this architecture, a panel of SEP modules has been developed by pairing an engineered intein (serving as the NED) with various CEDs, including a second engineered intein [1], a 2A-like peptide [8,12], and ubiquitin (Ub) [13]. SEP modules substitute for conventional 2A and IRES elements within polycistronic expression cassettes. Following expression, the SEP module is excised, enabling co- or post-translational release of the adjacent proteins with near-stoichiometric expression and a near-traceless, module-dependent junction. This review provides an in-depth examination of the SEP system, including its design, performance characteristics, and associated benefits and limitations, and highlights applications of SEP technology in

gene/trait stacking, metabolic engineering, biomanufacturing of complex protein products, and synthetic biology.

2. SEP Design

SEP technology represents a newly developed approach that supports the coordinated, simultaneous expression of multiple proteins from a sORF in eukaryotes. Its main advantages against current methods include efficient cleavage, predictable stoichiometry, and minimal amino acid scars. Furthermore, some of the SEP systems rely exclusively on non-viral sequences. SEPs are modular fusion proteins composed of N- and C-terminal excising domains (NED and CED, respectively) linked by a peptide linker. In theory, the NED can be any protein that enacts a cleaving event at its N-terminus, and similarly, the CED can be any protein that enacts a cleaving event at its C-terminus [14]. Cleavage events are autocatalytic, driven by the excision-domains themselves, or proceed through ubiquitous eukaryotic mechanisms such as cleavage by common endogenous host proteases. Because no auxiliary enzymes or regulatory proteins are required, these are effectively self-excising proteins. Regarding the linker, it does more than bridge the NED and CED; it also accommodates additional design features that can be optimized to enhance overall SEP performance, for example improving folding by minimizing interference between the two EDs and reducing steric hindrance, among other features discussed in later sections. An effective SEP module directs highly efficient release of its flanking proteins of interest (POIs), as illustrated conceptually in Figure 1.

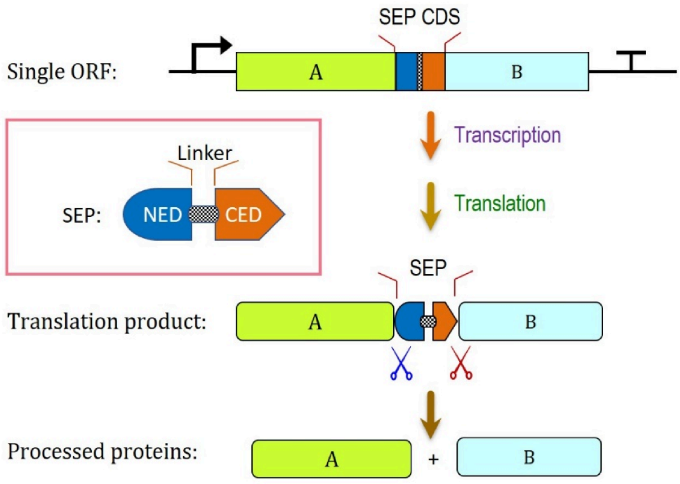


Figure 1. Working principle of SEP in coordinated multi-gene expression. A & B: target proteins coexpressed; CDS: coding sequence.

Fusing several proteins into a single SEP precursor does not, in principle, impair their folding. Because translation and folding both proceed from the N- to the C-terminus, each protein begins to fold co-translationally as it emerges from the ribosome [15,16]; an upstream protein therefore folds largely before the downstream SEP is synthesized, and cleavage releases the completed chain rather than gating the folding of proteins that have already emerged. The SEP modules themselves fold autonomously: ubiquitin reaches its native state on the millisecond timescale [17], and mini-inteins adopt the conserved Hedgehog/Intein (HINT) β-strand fold largely independently of the flanking sequence [18], while the flexible linkers noted above further decouple each SEP from its neighbors. Proteins that depend on one another to fold are released as separate polypeptides that associate as they would if expressed from separate genes, with the added advantage that a single SEP transcript supplies them at fixed stoichiometry.

In this review, we discuss a panel of SEPs that employ engineered mini-inteins with exceptional N-terminal autocatalytic cleavage activity as the NED, paired with diverse CEDs, including engineered mini-inteins with C-terminal cleavage activity, viral 2A peptides, non-viral 2A-like peptides, and Ub variants. Whereas inteins are autocatalytic: splicing and cleavage proceed through an intrinsic chemical mechanism that requires no host factors, so both N- and C-terminal cleaving variants are functional in prokaryotes and eukaryotes alike [1]. Ubiquitin and 2A are host-dependent by contrast. Ub-based processing relies on endogenous deubiquitinases (DUBs), and 2A relies on the eukaryotic ribosome; both therefore act autonomously as a CED only in eukaryotic hosts. Further details can be found in Sections 3–6.

3. Intein as a NED

Common to all SEP systems reviewed here is the use of an engineered mini-intein as the NED. Although inteins are better known for protein-purification and protein-ligation applications, their use as NEDs for protein expression is less familiar. Natively, an intein excises itself from a precursor protein, joining its two flanking sequences, the exteins, into a single chain [19]. For SEP, the intein is instead engineered to release the protein fused to its N-terminus by autocatalytic N-terminal cleavage.

Inteins occur in cis- or trans-splicing forms (Figure 2). In cis-splicing, a single polypeptide is arranged in the order N-extein, intein, C-extein; the intein folds as one domain and excises itself intramolecularly. In trans-splicing, a split intein is divided into an N-terminal fragment (IntN) and a C-terminal fragment (IntC) carried on separate polypeptides, with the N-extein fused to IntN and IntC fused to the C-extein. The two halves have strong mutual affinity, spontaneously associating to reconstitute a functional intein fold; only then can splicing proceed, joining exteins from two originally separate chains in an intermolecular reaction [19,20].

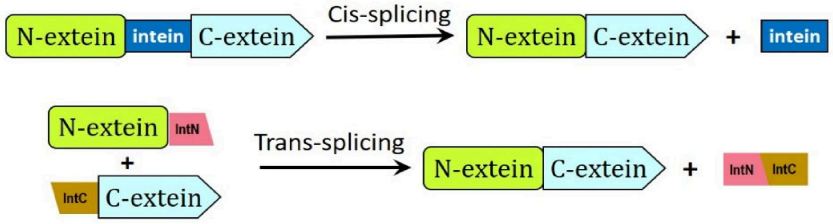


Figure 2. Cis- vs. trans-splicing via intein.

The two halves of a naturally split intein can also be joined into a contiguous mini-intein, as with the naturally split Npu DnaE mini-intein from *Nostoc punctiforme* [21]. Other contiguous inteins carry an embedded homing endonuclease (HEN) domain that can be deleted to create a mini-intein [22] and subsequently engineered into a split pair, as demonstrated for the *Mycobacterium tuberculosis* RecA intein [23]. Still others, such as the Mxe GyrA intein from the *Mycobacterium xenopi* DNA gyrase A subunit, natively lack a HEN domain [24].

The N-terminal cleavage used in SEP (Figure 3a) is a redirection of the intein's native splicing reaction (Figure 3b), which joins the two exteins with a native peptide bond through a series of acyl shifts, a transesterification, and cyclization of the intein's C-terminal asparagine. Substituting that asparagine with alanine prevents the cyclization that would complete splicing, so the acyl-shift intermediate is instead resolved by hydrolysis, freeing the N-extein instead of ligating the exteins [20]. This cleavage-competent, splicing-blocked design is central to all current SEP systems.

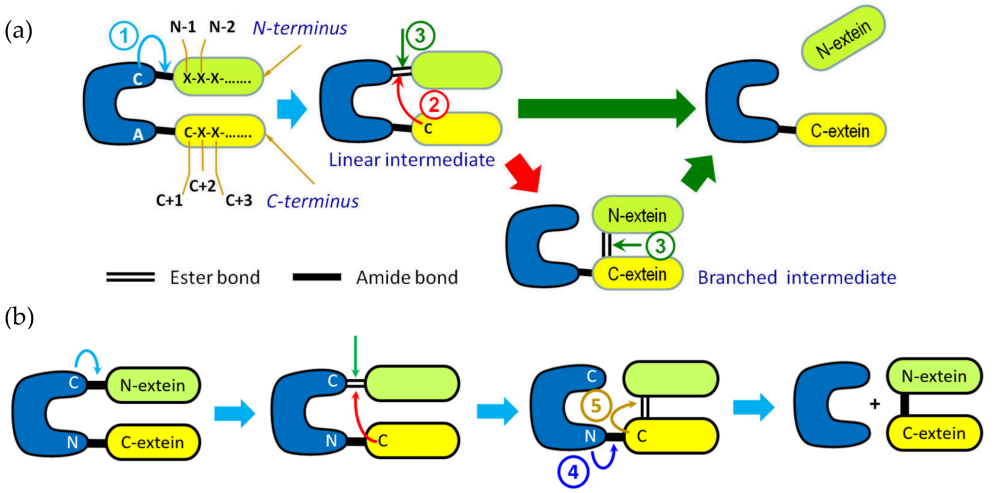


Figure 3. (**a**) N-terminal autocleavage process of an intein variant with a C-terminal Asn → Ala mutation. ① The nucleophilic thiol of the intein's N-terminal cysteine attacks the amide bond joining the N-extein to the intein, driving an N → S acyl shift that yields the linear intermediate. ② The N-extein is then relayed to the nucleophilic side chain of the C + 1 cysteine, forming a branched intermediate. ③ Hydrolysis of the unstable ester bond in the linear or branched intermediate, accelerated by mutating the flanking N- and/or C-extein residues, releases the cleaved N-extein. (**b**) Native intein splicing process. Following the same steps ① and ②, in step ④, the intein's C-terminal Asn cyclizes to a succinimide, cleaving the intein–C-extein amide bond and releasing the intein; ⑤ a spontaneous S → N acyl shift resolves the branched thioester into a native peptide bond, ligating the two exteins. A Cys at the +1 position is shown for illustration; Ser or Thr can serve the same role. Adapted from Amitai et al. [20].

The residues driving these steps sit in a conserved series of sequence blocks, A through H [19]. The block A first residue is the nucleophile (Cys, Ser, or Thr), and its identity sets the acyl shift: cysteine gives a labile thioester, whereas serine or threonine gives a more stable, less reactive ester. Cys1 is therefore favored for in vivo autocleavage, its thioester being both more readily hydrolyzed and formed faster, since thiolate is a stronger nucleophile than alkoxide at physiological pH [19]. Catalysis further depends on the block B histidine and threonine (the TxxH motif) and a block F aspartate, which polarize and strain the scissile bond and stabilize the transition state; the block B histidine is the principal catalytic residue, and its mutation cripples cleavage [19].

The exteins immediately flanking the intein also modulate activity. The N-1 residue (the last N-extein residue) sits in the junction pocket and strongly influences the shift rate: bulky hydrophobic residues slow it, and proline, which lacks a backbone amide proton and is geometrically incompatible, essentially abolishes it [20]. We have shown, however, that C-terminal Asn → Ala Ssp DnaE and Npu DnaE variants tolerate a broad range of N-1 residues during N-terminal autocleavage, although proline still suppresses activity [1,8,13].

4. Dual-Intein SEP: Using an Engineered Intein as a CED

To serve as an effective CED, efficient peptide-cleavage events need to occur at the juncture of CED and its downstream flanking protein. One of the initial CEDs investigated by the Su lab is an engineered C-terminal cleaving intein which enables autocatalytic peptide cleavage. Zhang et al. [1] reported a SEP design composed of two tandemly fused self-excising engineered inteins to direct synchronized protein coexpression in both prokaryotes and eukaryotes. This dual-intein system utilized *Ssp* DnaE and Ssp DnaB mini-inteins as NED and CED, respectively, where mutations were introduced to favor discrete cleavage over splicing. The first mutation converted the C-terminal asparagine in *Ssp* DnaE to an alanine (N159A), hence preventing asparagine cyclization and causing

this intein to favor N-terminal cleavage over splicing. The second mutation changed the N-terminal cysteine of *Ssp* DnaB to an alanine (C1A). In an intein carrying a Cys1 → Ala substitution, the N–S acyl shift and transesterification steps are blocked, uncoupling C-terminal cleavage from the splicing pathway. Cleavage then proceeds through a single step: protonation of the scissile amide nitrogen between the terminal Asn and the first C-extein residue renders the carbonyl carbon susceptible to nucleophilic attack by the Asn side chain, which cyclizes to a succinimide and releases the downstream protein with a free N-terminus. In *Ssp* DnaB this step is initiated by the penultimate His153, and its rate is pH- and temperature-dependent. Although the +1 residue has no direct catalytic role, its side chain modulates the cyclization barrier electronically: the naturally occurring nucleophiles Cys, Ser and Thr support the fastest cleavage, whereas hydrophobic substitutions such as Ala, Met, Val and Leu attenuate it by up to an order of magnitude. Notably, the +1 identity governs rate rather than extent, with cleavage reaching completion in all cases examined [25]. The combined use of these engineered inteins as a single cassette produced a polyprotein expression vector that enables independent, autocatalytic cleavage events for each flanking protein. The working principle of the dual-intein SEP for multi-protein expression is depicted in Figure 4.

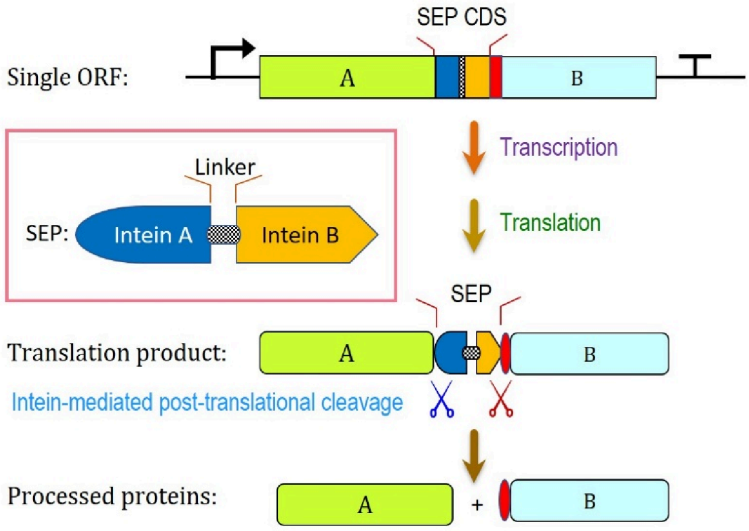


Figure 4. Mechanistic overview of dual-intein SEP mediated protein-coexpression. Note that the C + 1 residue of intein B (normally Cys, Ser, or Thr) becomes covalently appended to the N-terminus of POI-B after cleavage from the polyprotein precursor (indicated by the red oval). A & B: target proteins coexpressed.

The use of C1A mutation to block splicing and promote C-terminal cleaving, while very commonly practiced, is not universal across all inteins. For instance, the DnaE intein from *Nostoc punctiforme* was successfully converted into a C-terminal-cleaving intein (without N-terminal cleavage) by introducing a single Asp118-to-Gly mutation without mutating the Cys1 to Ala [26]. It should be noted, however, that the D118G mutation alone may not work as well in vivo if the proteins are targeted to an oxidized cell compartment, such as the endoplasmic reticulum (ER), because the potential formation of the Cys1/Cys + 1 disulfide trap may stay permanent to prevent C-terminal autocleavage [26]. This strategy may still be applicable for in vivo applications to express cytosolic proteins since cytosol is strongly reducing.

In the work of Zhang et al. [1], GFP and RFP (mCherry), along with other protein pairs including GFP/chloramphenicol acetyltransferase (CAT) and thioredoxin (Trx)/RFP, were positioned on either side of the dual-intein cassette and used as cytosolic reporter proteins to evaluate the viability of the dual-intein SEP system. The dual-intein system was demonstrated to efficiently cleave all proteins tested, in bacterial (*E. coli*), mammalian (HEK293T), and plant (*Nicotiana tabacum*) cells while preserving their activities. Inactive intein mutants with blocked N- or C-terminal auto-cleavage activity were also tested, and these controls confirmed that the observed cellular processing of the precursor polyprotein was specifically attributable to the autocatalytic cleavage activity of the dual-intein fusion domain. Furthermore, by analyzing the released RFP and GFP, purified from the transgenic tobacco NT1 cells, the exact cleavage sites within the GFP-SEP-RFP polyprotein were confirmed according to N-terminal protein sequencing and Electrospray Ionization Time-of-Flight Mass Spectrometry (ESI-TOF MS), respectively. The RFP product released from the polyprotein carries an N-terminal Ser, which corresponds to the +1 position of the DnaB intein, while the observed molecular mass of the released GFP product measured using ESI-TOF MS matches that of an N-terminal acetylated GFP [1].

Zhang et al. [1] also investigated the effect of different redox environments and temperatures on the cleavage activity of the dual-intein system. It was shown that there was little to no effect on cleavage with the different redox potentials in the cytosol of *E. coli*, HEK293T, and tobacco NT1 cells. When testing the effect of expression temperatures in *E. coli*, Zhang et al. [1] found that while N-terminal cleavage was unaffected by a 25 °C incubation compared to the usual 37 °C, C-terminal cleavage was reduced, while RFP cleavage efficiency was restored after re-incubating the cells at 37 °C. In the plant system, which is cultured at 25 °C, both N- and C-terminal cleavages functioned properly.

With the dual-intein SEP, expression levels of GFP and RFP post-cleavage are shown to be close to stoichiometric across different host organisms, including *E. coli*, HEK293T, and tobacco NT1 cells. One feature of the dual-intein system that warrants mention is the intrinsic addition of extein scars on the C-terminal POI. This occurs because C-terminal intein cleavage is affected by the C + 1 extein residue (serine, cysteine, or threonine), with a secondary contribution from C + 2. As these residues must remain integral to the dual-intein module, they are consequently retained on the downstream POI following release of the precursor. Serine at C + 1 in the construct of Zhang et al. [1] gave near-complete release of the downstream RFP in planta at 25 °C and in *E. coli* at 37 °C. Although cysteine may cleave marginally faster, it would leave an N-terminal cysteine on the released protein that is subject to N-end-rule degradation in plants via the cysteine-oxidation branch; serine therefore represents the preferred C + 1, combining efficient cleavage with a stable product N-terminus [27].

In principle, configuring a single intein for both N- and C-terminal autocleavage requires mutually incompatible mutations: N-terminal cleavage requires an active N-terminal nucleophile (Cys1) with the C-terminal Asn inactivated, whereas C-terminal cleavage requires an active C-terminal Asn with the N-terminal nucleophile inactivated [20]. Nonetheless, a single intein bearing both the C1A and N159A mutations has been reported to direct coexpression of the heavy (HC) and light (LC) chains of an IgG1 antibody from a single open reading frame in mammalian cells [28,29]. In this case, HC and LC each carried their native signal peptides. *Pyrococcus horikoshii* Pol I intein and *P. abysii* Lon intein, respectively, were used in these studies. While some assembled IgG proteins were detected in the cell culture media, considerable amounts of unprocessed proteins were also noted. Although a single intein yields a more compact sORF construct, the difficulty of optimizing both N- and C-terminal autocleavage within one intein favors the dual-intein system.

5. Intein-2A SEP: Using 2A-Peptide as CED

As an alternative to the C-terminal cleaving intein, a 2A peptide has been explored as a CED to form a SEP system with an N-terminal cleaving intein [8]. An N159A *Ssp* DnaE mini-intein was fused to the foot-and-mouth disease virus 2A (F2A) peptide for coordinated expression of multiple proteins [8]. As noted in the introduction, 2A peptides with ribosome-skipping activity are commonly used in sORF expression of multiple proteins. However, the 'remnant' 2A residues appended to the carboxyl terminus of the processed proteins could hinder protein activity and/or cellular targeting [8]. Attempts to remove residual 2A residues using endogenous host proteases in plants [30] and mammals [31] have seen limited success, owing to the need for specific proteases and the persistence of remnant linker residues on the cleaved POIs. To resolve these problems, Zhang et al. [8] leveraged intein-mediated N-terminal autocleavage to excise the 2A extension in vivo by fusing an engineered mini-intein to the 2A sequence via a linker, creating the intein-2A SEP (Figure 5).

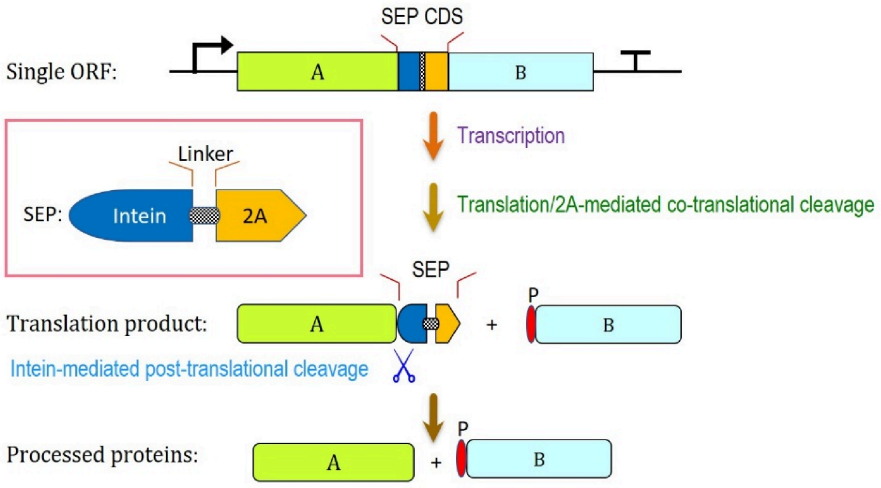


Figure 5. Mechanistic overview of multi-protein coexpression mediated by an intein-2A SEP. Intein provides precise and clean cleavage of POI-A, while 2A leaves a proline scar on the N-terminus of POI-B.

This intein-2A construct enables post-translational cleavage of the N-extein via the engineered intein, and co-translational cleavage of the C-extein via F2A. Notably, for certain inteins such as the Npu DnaE mini-intein, N-terminal cleavage proceeds extremely rapidly, approaching a cotranslational event [21]. In the work of Zhang et al. [8], both bicistronic and tricistronic expression were demonstrated using the intein-2A SEP in tobacco NT1 cells, tobacco plants, *Nicotiana benthamiana* leaves, and *Lactuca sativa* leaves. To test bicistronic expressions, GFP and RFP were used as N- and C-exteins. To assess tricistronic expression, GFP, mKO1 (an orange fluorescent protein), and RFP were arranged in series with an intein–F2A sequence between GFP and mKO1 and a second between mKO1 and RFP. In additional intein-2A polyprotein constructs, plant ER-targeting signal peptides were fused to GFP and RFP, respectively, and also to both reporters; together, the results demonstrated that released POIs undergo accurate and differential subcellular targeting in transgenic tobacco cells. The utility of the intein-2A SEP was further demonstrated in the expression of a fully assembled and functional chimeric anti-His tag antibody in *Nicotiana benthamiana*, achieved through co-expressing signal-peptide-bearing antibody light and heavy chains [8].

In 2022, the intein-2A system was further developed to incorporate non-viral 2A-like sequences, in place of conventional viral 2A peptides [12]. Two non-viral 2A-like peptides derived from purple sea urchin (*Strongylcentrotus purpuratus*) [12] and California sea slug (*Aplysia californica*) [32], respectively, were evaluated for intein-2A based multi-protein

expression [12]. An engineered (N159A) N-terminal *Ssp* DnaE intein was used as the NED. It was found that the two non-viral 2A peptides tested gave similar performance as their viral counterparts, in plant and mammalian cells. The non-viral intein-2A system was successfully used to express the heavy chain and light chain subunits of a recombinant anti-HER2 antibody. Given current consumer sentiment, food and medical products that avoid viral components are generally preferred. Accordingly, the use of non-viral 2A-like peptides provides a more favorable route for deploying intein-2A technology in the manufacture of food and medical products via genetically modified crops and animals.

6. Intein-Ub SEP: Ubiquitin and Its Variants as CED

Ub is a small (76-residue), highly conserved protein found in all eukaryotic cells [13]. It is most commonly known for its role in proteasomal degradation, where the carboxyl group of its C-terminal glycine forms an isopeptide bond with a lysine residue on a substrate protein in a process called ubiquitination, catalyzed by an E1–E2–E3 enzyme cascade. Additional Ub molecules can be conjugated through one of the seven lysine residues (or the N-terminal methionine) of a previously attached Ub, building a polyubiquitin chain; Lys48-linked chains are the canonical signal that targets the substrate to the proteasome, whereas other linkages (e.g., Lys63) serve largely non-degradative signaling roles. The reverse reaction, deubiquitination, is carried out by DUBs, which hydrolyze the isopeptide bond immediately C-terminal to the diglycine motif, releasing intact Ub. By recycling Ub and reversing or editing conjugation, these processes maintain the free Ub pool and regulate substrate protein levels within the cell [13].

Because DUBs are present in all eukaryotes, releasing the protein downstream of the CED requires no exogenous protease, making Ub a promising CED candidate for SEP-based multi-protein coexpression. A single Ub between two POIs, however, releases only the C-terminal POI cleanly; the N-terminal POI is left with Ub fused to its C-terminus. To free both proteins, Walker and Vierstra [33] inserted the last 6–14 C-terminal residues of Ub between the N-terminal POI and a full-length Ub, so that DUBs cleave at both Ub C-termini, excising the full-length Ub as a free monomer and releasing the C-terminal POI with a native N-terminus. This design does not, however, fully remove the Ub-derived C-terminal peptide which remains appended to the N-terminal POI. The influence of this residual fragment on the behavior of the appended protein merits additional examination.

To facilitate clean liberation of the POIs without altering their terminal residues, ubiquitin was appended to the C-terminus of an N-cleavage intein, yielding an intein-Ub SEP capable of directing multi-protein expression from a single ORF, as shown in Figure 6 [13]. Whereas 2A and Ub based sORF expression systems result in residual amino acid sequences being left on either or both POIs, which can affect folding pathways, stability, and function, the intein-Ub SEP system can successfully release the flanking POIs while preserving their native sequences, as confirmed using ESI-TOF MS and N-terminal protein sequencing [13]. By yielding discrete proteins that retain native, scar-free termini, the intein-Ub SEP system provides a distinct benefit for high-fidelity, multi-protein coexpression in eukaryotic systems. When a FLAG tag (DYKDDDDK) was introduced into the intein-Ub linker, anti-FLAG Western blotting of *N. benthamiana* leaf extracts expressing GFP–intein–FLAG–Ub–RFP did not reveal the intein–FLAG–Ub intermediate, suggesting its degradation in planta, whereas probing with anti-GFP and anti-RFP antibodies confirmed efficient release and accumulation of the discrete GFP and RFP products. The intein-Ub fragment ends in a diglycine (–GG) motif, a canonical C-terminal degron in animal systems [34], which may contribute to proteasomal degradation of the intein–FLAG–Ub species; however, this possibility remains unvalidated in plants. An alternative fate is that the ubiquitin moiety is released into the cellular ubiquitin pool and subsequently

degraded, although this pathway likewise requires experimental confirmation in plant systems. Importantly, GFP and RFP accumulate to approximately equimolar levels [13]. This represents an ideal scenario, i.e., robust POI expression coupled with self-destruction of the SEP once the POIs are released.

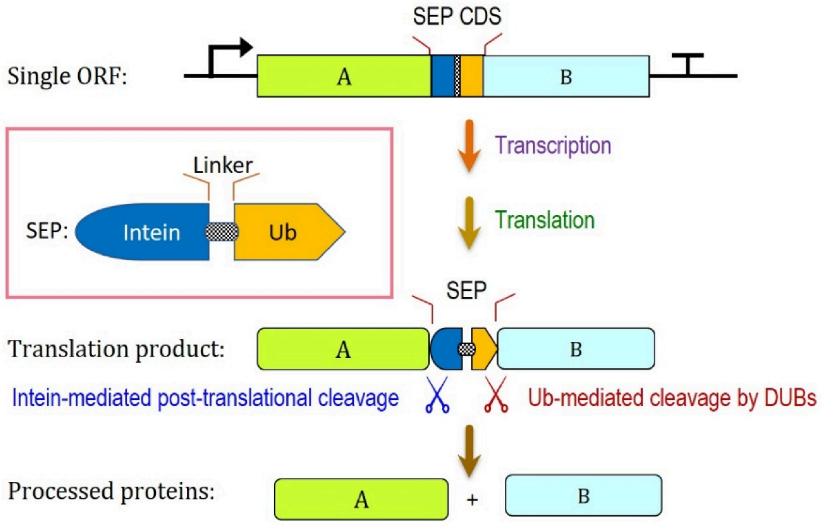


Figure 6. Mechanistic overview of intein-Ub SEP mediated protein coexpression. Note that the POIs preserve their native termini after cleavage from the polyprotein precursor. A & B: target proteins coexpressed.

7. Evaluative Comparison of SEP with Established Multi-Protein Expression Systems

Viral 2A and IRES elements are two established strategies for protein coexpression in eukaryotes. Building upon the brief overview provided in Section 1, a more detailed comparison between SEP and these current states of the art is presented in this section.

7.1. Internal Ribosome Entry Site Elements (IRES)

In typical eukaryotic translation, the ribosome binds near the 5′ cap of an mRNA strand and scans toward a start codon. IRES elements provide an alternative mechanism for translation initiation. They are structured RNA sequences found in viruses that allow ribosomes to begin translation of an ORF at an internal position within an mRNA strand rather than at the usual 5′ cap. This is commonly known as cap-independent translation. IRES elements were first discovered in two separate picornaviruses, i.e., encephalomyocarditis virus (EMCV) and poliovirus, but have since been identified in various other viral families such as flaviviruses, dicistroviruses, and retroviruses [35].

IRES elements are considered the first-generation method for eukaryotic polycistronic expression, and they were widely used until 2A peptide systems took over. By placing an IRES element between two genes of interest, the first gene is translated via normal cap-dependent reactions, while subsequent genes will be translated via IRES-mediated cap-independent reactions [6].

IRES-mediated expression of multiple genes has been applied to several different eukaryotic systems for biomanufacturing, gene therapy, and synthetic biology. For example, IRES elements have been used to coordinate expression of heavy and light chains for HER2 monoclonal antibody (mAb) production in CHO cells [36]; and they have also been used to bicistronically express an Na^+/H^+ antiporter with a H^+-pyrophosphatase in order to improve salt tolerance in tobacco plants [37]. These studies demonstrate the feasibility of IRES to achieve multi-gene coexpression from a polycistronic sequence in eukaryotes.

However, the IRES-based method is limited by the lack of stoichiometric control over the relative expression levels of each target protein. In IRES-based multicistronic cassettes, each subsequent protein is translated independently of the others through separate ribosome recruiting events, where only the first gene undergoes efficient cap-dependent translation. In comparison, cap-independent translation has been found to be up to 80% less efficient than cap-dependent translation [38]. As a result, the upstream protein in a bicistronic vector will be expressed at a much higher rate than its downstream counterparts, and this disparity in efficiency only becomes more apparent as more cistrons are incorporated.

IRES elements are also limited by their sequence variety, large size, and viral origins. All IRES elements initiate some form of internal translation, but while most viral IRES elements are dependent on their secondary RNA structures in order to carry out internal translation initiation, there is no single structural motif shared across all IRES-containing viruses. Even in picornaviruses, there are five different classifications of IRES [39]. These sequence differences contribute to substantial variability in translational efficiency and host compatibility, making the efficiency of IRES highly context dependent. Also, it should be noted that animal viral IRES sequences may not work in plants, and vice versa [40].

Viral IRES elements are relatively large, with functional elements typically spanning ~200 to 600 nt depending on class. Although this is a modest contribution to overall T-DNA size for plant transformation, the more consequential limitations of IRES in polycistronic constructs are the strongly attenuated translation of the downstream cistron and the added complexity of vector assembly [41]. Furthermore, because of their viral origins, IRES are less likely to be accepted publicly, leading to troubles with commercial application. While small, non-viral IRES elements have been found, they are often less reliable and robust than the conventional large, viral IRES elements [35]. In fact, these limitations majorly influenced the development of alternative polycistronic expression systems like 2A peptides and the SEP system.

7.2. 2A and 2A-like Peptides

Currently, the most widely adopted strategy for eukaryotic sORF-based multi-protein expression involves the use of viral 2A peptides because of their compact size and improved stoichiometric balance in comparison to IRES elements [3]. While IRES elements are on the order of 500 nt in length, 2A peptides commonly reported contain just 18–30 amino-acid residues (though sequences as long as 58 residues have also been used [8]). It was first characterized in foot-and-mouth disease virus (F2A) and later found across picornaviral and 2A-like elements, including P2A (porcine teschovirus-1), T2A (*Thosea asigna* virus), and E2A (equine rhinitis A virus) [42]. Unlike IRESes, animal virus-derived 2A is functional in plants [2]. Although routinely described as self-cleaving, 2A peptides are not proteases and no peptide bond is hydrolyzed in the conventional sense. Instead, they direct a co-translational recoding event known as ribosomal skipping or stop-carry-on translation. As the ribosome translates the conserved C-terminal motif D-(V/I)-E-X-N-P-G-P, the nascent 2A sequence within the exit tunnel impairs formation of the final glycyl-prolyl peptide bond; the ester linkage of the upstream peptidyl-tRNA is resolved and the upstream polypeptide is released, while the ribosome continues elongation from the downstream proline without dissociating [3]. The outcome is two separate polypeptides generated from a single open reading frame through one cap-dependent initiation event. Because cleavage occurs at the glycyl-prolyl junction, the upstream protein retains the 2A sequence (minus its terminal proline) as a C-terminal extension, while the downstream protein begins with an N-terminal proline.

Use of 2A and 2A-like sequences has been reported in a multitude of eukaryotic organisms including yeast, plants, and mammalian cells, for protein production, trait stacking, and metabolic engineering [2,31]. Despite being widely adopted, these systems still have significant limitations stemming from their residual peptide scars, viral origins, and variable ribosome-skipping efficiency. Due to the nature of ribosome skipping, the 2A scar left on the upstream POI and the Pro scar left on the downstream POI can potentially compromise folding, stability, subcellular targeting, and even function of the translated proteins [8]. Furthermore, besides ribosome skipping, readthrough (yielding a fusion polyprotein) or drop-off (yielding only the upstream protein with the 2A extension) may occur, which has been found to lead to an approximate 70% decrease in expression between the first and second cistrons [42]. Regarding subcellular targeting, the presence of an N-terminal signal peptide in a 2A-based expression cassette causes both the upstream and downstream proteins to be targeted to the ER due to what is known as a "slipstream" in the translocon pore in mammalian cells [43]. In another study, the 2A extension was shown to interfere with the ER-targeting signal and resulted in mistargeting of the protein to vacuole in plant (tobacco) cells, while incorporation of intein to remove the 2A extension resolved this problem [8].

7.3. Comparison of SEPs with IRES and 2A

A side-by-side comparison of the main features of SEP systems with 2A and IRES is presented in Table 1. SEP technology offers advantages in most respects, its principal trade-off being molecular size. At ~160–300 aa (~480–900 bp, depending on the type used), a SEP module is considerably larger than a 2A peptide (~18–58 aa) and adds roughly 18–34 kDa of translated polypeptide that draws on the cell's ribosomes, tRNAs, and energy without contributing to the final products. This imposes a modest metabolic cost that can slightly reduce per-cell yield and add to the synthesis and cloning burden of assembling the cassette, a burden that may grow as more cargo proteins are linked in a single construct. In practice, however, these costs remain marginal and are outweighed by the more complete and efficient separation that SEP provides, with yields of various products, including recombinant IgG antibodies, staying within the expected range [1,8,12,13].

Because the inteins used in SEP are cleavage-competent mutants rather than splicing domains, their activity depends on pH, temperature, and the residues flanking the junction, so each new junction, host, or compartment requires validation. Compartment is nonetheless less restrictive than intein biochemistry might suggest, as N-terminal cleavage proceeds efficiently in the ER lumen in planta [8]. One caveat applies to both 2A and SEP systems: equimolar synthesis fixes the ratio at which partners are made, not the ratio at which they accumulate.

Intein-Ub combines several desirable attributes: traceless cargo release, clean cleavage, non-viral origin, and modest size. It does, however, depend on host DUBs for its ubiquitin moiety to function as a CED. Because DUBs are cytosolic and nuclear, targeting the precursor to the ER with a signal peptide translocates the ubiquitin moiety into the lumen, away from them. Intein-Ub is nonetheless well suited to coexpressing cytosolic, nuclear, and plastid-targeted proteins, all of which are translated in the cytosol, so this constraint does not affect the majority of important applications, including commercial crop trait stacking, as discussed in Section 8.

Among the SEP variants, intein-2A has the smallest footprint, followed by intein-Ub and then dual-intein. It also improves on 2A used alone: although the ~20-residue 2A peptide is the most common form, skipping efficiency can be raised with a longer variant [44,45], which, fused directly to a protein, would leave a long and disordered 2A overhang. Within intein-2A this liability disappears, because the entire module, 2A

included, is cleaved from the upstream protein, leaving no overhang, and the intein-to-2A linker can be tuned to enhance skipping further. The inherent deficiencies of 2A are thus attenuated in this context.

Table 1. Comparison of IRES, 2A, and SEP systems.

Feature \ Module	IRES	2A	SEP		
			Intein-Intein	Intein-2A	Intein-Ub
Non-viral origin	No	No	Yes	Yes (with non-viral 2A)	Yes
Preserved native termini	Yes	2A scar on proximal POI proline scar on distal POI	C + 1 scar on distal POI	Proline scar left on distal POI	Yes
Stoichiometric production of released cargo POIs	frequently departing from equimolar	sub-equimolar/ more variable ▽	~equimolar ♣	~equimolar ♣	~equimolar ♣
Completeness of cargo POI release	yields individual POIs	>50% ❤	>80% ♢	>80% ❤	>90%
Applicability in cytosol & ER	Yes	Yes	Yes	Yes	cytosol §
Applicability in prokaryotes or eukaryotes	eukaryotes	eukaryotes	prokaryotes & eukaryotes	eukaryotes	eukaryotes
Module Size	~500 nt [3]	~18–58 aa [8]	Intein: 137–159 aa [1]	Intein: 137–159 aa; 2A: 18–58 aa [8,12]	Intein: 137–159 aa; Ub: 76 aa [13]
Main references	[38]	[3]	[1]	[8,12]	[13]

▽: affected by the 2A peptide type and length, and its upstream protein C-terminal sequence [3]; the 2A overhang may cause misfolding or mistargeting [8]. ♣: equimolar synthesis fixes the ratio at which cargo POIs are made, not the ratio that accumulates, which also depends on individual POI stability. ❤: depending on the 2A peptide used in releasing the downstream POI. ♢: depending on the C-cleaving intein used in releasing the downstream POI. §: Currently only tested for cytosolic proteins.

Two questions become more pressing as additional genes are linked in a single sORF: whether the size of the cargo proteins caps their number, and whether the lengthening construct introduces liabilities at the nucleic-acid level. Gene number is not limited by product size, because the independence of successive modules is built into their design rather than borrowed from the proteins they separate. Each module is self-contained: the NED intein takes its C-extein residues from an internal linker rather than from the adjacent target protein, so junction chemistry is fixed regardless of the identity, size, or number of flanking proteins, and each module adds minimal sequence to its released product (none for intein-Ub, a single Cys, Ser, or Thr for dual-intein, and a single Pro for intein-F2A). Three fluorescent proteins have accordingly been expressed in plants from one ORF using two intein-F2A modules, with no minimum target-protein size [8]. The only parameter inherited from flanking sequence is the C-terminal residue of the upstream protein, which occupies the intein's -1 position and modulates the rate of the N $\rightarrow$ X acyl shift, a sequence consideration rather than a limit on size or gene number. No intrinsic molecular-weight ceiling is known for the precursor; the operative constraints (transcript length, vector capacity, and the folding and solubility of the transiently accumulating precursor) lie outside the SEP chemistry itself.

At the nucleic-acid level, joining coding sequences head-to-tail alters the exon/intron structure of the assembled transcript, since the novel junctions can generate cryptic splice sites, a risk that scales with construct length. Because SEP and 2A are protein-coding, such a signal can be removed by a synonymous substitution that leaves the product unchanged, an option unavailable to an IRES, whose function depends on RNA secondary structure. A related consideration is intrinsic to the sORF architecture itself: because all genes share a

single reading frame, imprecise splicing frameshifts every downstream product, so junction design merits attention as more genes are linked. This is particularly pertinent in plants, where introns are frequently included by design for intron-mediated enhancement.

8. Applications

Polycistronic and single-ORF multiprotein expression technologies have become indispensable tools in synthetic biology, plant gene/trait stacking, metabolic engineering, and complex bioproduct (such as antibodies) manufacturing, enabling the coordinated expression of multiple genes from a single mRNA in eukaryotes. Some notable applications involving the use of 2A or SEP are summarized in Table 2. Many multi-protein expression applications currently supported by IRES and 2A systems could benefit from SEP technology, especially when non-viral sequences, stoichiometric expression, and clean, near-traceless cargo protein release are required.

The information in Table 1 serves as a useful guideline for selecting the appropriate SEP for multi-protein expression. Chief among the factors to consider is the subcellular localization of the target POIs (refer to Table 2). For the coexpression of cytosolic, nuclear, and plastid-targeted proteins translated in the cytosol, the intein-Ub SEP serves as an efficient platform. For POIs destined for the secretory pathway, dual-intein or intein-2A SEPs are preferred. A case in point is the expression of immunoglobulin antibodies, whose successful assembly requires not only coordinated, stoichiometric expression of the heavy and light chains, but also their proper trafficking through the secretory pathway. Following translation, N-terminal signal peptides direct the heavy and light chains into the ER lumen, where they undergo folding, disulfide bond formation, and final assembly. Zhang et al. [8] demonstrated successful intein-mediated N-terminal cleavage within this oxidative environment using an ER-targeted GFP-Ssp DnaE intein-F2A-RFP fusion construct, detecting correctly cleaved and secreted GFP in transgenic tobacco cells. Mechanistically, intein N-terminal autocleavage requires an N $\rightarrow$ S (or N $\rightarrow$ O) acyl shift that depends on the free thiol group of a cysteine (or the hydroxyl of a Ser/Thr) at position 1 of the intein. Although free thiols are highly prone to oxidation or disulfide bond formation in the ER lumen, which typically stalls the reaction, Zhang et al. [8] proved that the intein's structural pocket shields the active site well enough to outcompete this oxidative environment. Regarding C-terminal autocleavage in a dual-intein module, the C1A mutation of the CED intein completely eliminates the N-terminal cysteine, thereby isolating the C-terminal cleavage reaction. Because this reaction is driven by asparagine (Asn) cyclization, it is entirely independent of redox state and is firmly expected to function effectively in the oxidative ER [58]. Beyond immunoglobulins, this SEP selection rule applies broadly to the wide variety of metabolic enzymes targeted to the ER.

A related question for secretory-pathway targets is whether N-linked glycosylation is correctly installed on proteins released from a SEP precursor. Interference is unlikely, because glycosylation is co-translational: oligosaccharyltransferase modifies Asn-X-Ser/Thr sequons on the nascent chain as it enters the ER lumen, independently of when a module is cleaved. Modifications occurring after release likewise follow the same pathways as in conventional expression, since each product is by then a free protein indistinguishable from one made monocistronically. SEP-based expression would therefore not be expected to alter the modification of the released proteins, although this remains to be confirmed directly. The point matters wherever the products are enzymes or otherwise bioactive, since correct glycosylation can be essential to their activity, stability, or localization.

Table 2. Notable published applications associated with multi-protein expression.

Application	Example (Brief Description)	Strategy	Target Protein Localization	Host	Ref.
Plant trait/gene stacking	β-glucosidase, xylanase, & exoglucanase each linked via 2A to endoglucanase Cel5A, co-expressed for improved biomass hydrolysis & bioethanol production	2A (FMDV F2A)	Plastid (chloroplast; transit peptide on each enzyme)	Tobacco (*Nicotiana tabacum*)	[46]
	Golden Rice 2: provitamin A biofortification via maize phytoene synthase + *Erwinia* carotene desaturase	Multiple monocistronic cassettes (separate promoters)	Plastid (endosperm)	Rice (*Oryza sativa*)	[47]
	Gene stacking and stoichiometric expression of ER-targeted constructs using SEP	Intein-P2A SEP	ER	Tobacco	[48]
Metabolic engineering	Nine-gene carotenoid + violacein pathways coexpressed via T2A/bidirectional promoter	2A (T2A)	Cytosol	*Pichia pastoris*	[49]
	Triterpene friedelin via synthase co-expressed with truncated HMG1 linked by ERBV-1 2A	2A (ERBV-1)	Cytosol	Yeast (*Saccharomyces cerevisiae*)	[50]
	Engineering the hydrolyzable-tannin pathway in *N. benthamiana* by co-expressing gallic-acid biosynthetic genes	Intein-Ub SEP	Cytosol	*Nicotiana benthamiana*	[51]
Synthetic biology	Single-transcript dCas9 multi-effector regulator for activation, repression, methylation, and demethylation	2A (T2A/P2A)	Nucleus (each carries SV40 NLS)	Mammalian (HEK293)	[52]
	Anti-inflammation sense-and-respond device; 2A-linked module secretes anti-TNF-α IgG + IL-10/IL-1RA	2A	Secretory pathway → secreted	Mammalian (HEK293-derived)	[53]
	Co-expressing a Ser2P-mintbody and H2B-mRuby to image RNA Pol II Ser2P dynamics in planta	IntF2A SEP	Cytosol/nucleus	*Nicotiana Benthamiana/ Arabidopsis thaliana*	[54]
	Meiosis-specific co-expression in *Arabidopsis*	IntF2A SEP (co-express 3 proteins)	Differential post-cleavage targeting (nucleus/cytosol)	*Arabidopsis thaliana*	[55]
	Detecting membrane protein–protein interaction in planta based on dual-intein-coupled tripartite split-GFP association	Dual-intein SEP	Cytosol/nucleus	*Arabidopsis thaliana*	[56]
Antibody production	IntF2A SEP releases flanking proteins in cis from one polyprotein; demonstrated with an anti-His-tag IgG (heavy + light chain) plus fluorescent reporters	IntF2A SEP	Differential: cytosol/secretory pathway	Tobacco cells and whole plants	[8]
	Full-length IgG from a single ORF via FMDV 2A linking heavy and light chains; AAV-delivered, >1 mg/mL *in vivo*	2A (FMDV F2A)	Secretory pathway → secreted	Mammalian (HEK293; mouse)	[31]
	IgG3 heavy/light chain co-expression in *N. benthamiana*	IntF2A SEP	Secretory pathway → secreted	*Nicotiana Benthamiana*	[57]

Abbreviations: AAV, adeno-associated virus; ERBV, equine rhinitis B virus; FMDV, foot-and-mouth disease virus; H2B, H2B core histone protein; HMG1, HMG-CoA reductase I; IL-10, interleukin-10; IL-1RA, interleukin-1 receptor antagonist; IntF2A, *Ssp* DnaE mini-intein fused to FMDV 2A; Ser2P, serine-2-phosphorylated Pol II; TNF-α, tumor necrosis factor alpha.

9. Conclusions

Established multiprotein expression methods such as IRES elements and 2A peptides have contributed substantially to the fields of recombinant protein complex biomanufacturing, synthetic genetic circuits, and trait and metabolic engineering, yet each carries notable limitations. Less widely known are SEPs, an emerging class of single-ORF multiprotein expression systems that address many of these constraints through autocatalytic or endogenous processing mechanisms. SEPs have drawbacks of their own, chief among them a larger molecular footprint, since each module adds coding sequence and translational burden, but each of the current systems (dual-intein, intein-2A, and intein-Ub) offers distinct advantages that make it a competitive alternative to IRES and 2A. The appropriate SEP can be selected by matching its processing mechanism to the needs of the target protein. Proteins synthesized in the cytosol may benefit from the intein-Ub SEP, whereas proteins translated at the ER may be better served by the dual-intein or intein-2A SEP. Matching a SEP module's processing characteristics to its intended application thus provides a route to highly efficient protein production while preserving native protein structure and function.

While independent adoption of SEP technologies for metabolic engineering, synthetic biology, and antibody production has begun to emerge [48,51,54–57,59], broader uptake may require more extensive demonstration of commercially important trait genes in gene/trait stacking in crop plants, along with further research to overcome the key limitations imposed by the molecular size of the SEP module.

Author Contributions: K.L. performed the literature review, assessed the published findings, and contributed to manuscript development. W.-W.S. conceived and designed the study, integrated and critically evaluated the literature, contributed to manuscript writing, and secured all funding for the project. W.-W.S. is the primary corresponding author. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the University of Hawaii Provost Strategic Investment Initiative Award, and by the NIFA hatch project HAW05040-H and multistate project HAW05041-R to W.W.-S. at University of Hawaii.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors acknowledge Zhenlin Han and Bei Zhang of the University of Hawai'i for their significant contributions to the SEP work carried out in the Su lab and presented in this review.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhang, B.; Rapolu, M.; Liang, Z.; Han, Z.; Williams, P.G.; Su, W.W. A Dual-Intein Autoprocessing Domain that Directs Synchronized Protein Co-Expression in Both Prokaryotes and Eukaryotes. *Sci. Rep.* **2015**, *5*, 8541. [CrossRef]
2. Halpin, C. Gene stacking in transgenic plants—The challenge for 21st century plant biotechnology. *Plant Biotechnol. J.* **2005**, *3*, 141–155. [CrossRef] [PubMed]
3. de Felipe, P.; Luke, G.A.; Hughes, L.E.; Gani, D.; Halpin, C.; Ryan, M.D. E unum pluribus: Multiple proteins from a self-processing polyprotein. *Trends Biotechnol.* **2006**, *24*, 68–75. [CrossRef] [PubMed]
4. Wang, X.; Marchisio, M.A. Synthetic polycistronic sequences in eukaryotes. *Synth. Syst. Biotechnol.* **2021**, *6*, 254–261. [CrossRef] [PubMed]
5. El Amrani, A.; Barakate, A.; Askari, B.M.; Li, X.; Roberts, A.G.; Ryan, M.D.; Halpin, C. Coordinate Expression and Independent Subcellular Targeting of Multiple Proteins from a Single Transgene. *Plant Physiol.* **2004**, *135*, 16–24. [CrossRef] [PubMed]
6. Renaud-Gabardos, E.; Hantelys, F.; Morfoisse, F.; Chaufour, X.; Garmy-Susini, B.; Prats, A.C. Internal ribosome entry site-based vectors for combined gene therapy. *World J. Exp. Med.* **2015**, *5*, 11–20. [CrossRef] [PubMed]
7. Sharma, P.; Yan, F.; Doronina, V.A.; Escuin-Ordinas, H.; Ryan, M.D.; Brown, J.D. 2A peptides provide distinct solutions to driving stop-carry on translational recoding. *Nucleic Acids Res.* **2012**, *40*, 3143–3151. [CrossRef] [PubMed]

8. Zhang, B.; Rapolu, M.; Kumar, S.; Gupta, M.; Liang, Z.; Han, Z.; Williams, P.; Su, W.W. Coordinated protein co-expression in plants by harnessing the synergy between an intein and a viral 2A peptide. *Plant Biotechnol. J.* **2017**, *15*, 718–728. [CrossRef] [PubMed]
9. Zhang, B.; Rapolu, M.; Huang, L.; Su, W.W. Coordinate expression of multiple proteins in plant cells by exploiting endogenous kex2p-like protease activity. *Plant Biotechnol. J.* **2011**, *9*, 970–981. [CrossRef] [PubMed]
10. Yue, Q.; Meng, J.; Qiu, Y.; Yin, M.; Zhang, L.; Zhou, W.; An, Z.; Liu, Z.; Yuan, Q.; Sun, W.; et al. A polycistronic system for multiplexed and precalibrated expression of multigene pathways in fungi. *Nat. Commun.* **2023**, *14*, 4267. [CrossRef] [PubMed]
11. Ma, X.; Yue, Q.; Miao, L.; Li, S.; Tian, J.; Si, W.; Zhang, L.; Yang, W.; Zhou, X.; Zhang, J.; et al. A novel nucleic acid linker for multi-gene expression enhances plant and animal synthetic biology. *Plant J. Cell Mol. Biol.* **2024**, *118*, 1864–1871. [CrossRef] [PubMed]
12. Su, W.W.; Zhang, B.; Han, Z.; Kumar, S.; Gupta, M. Non-viral 2A-like sequences for protein coexpression. *J. Biotechnol.* **2022**, *358*, 1–8. [CrossRef] [PubMed]
13. Zhang, B.; Han, Z.; Kumar, S.; Gupta, M.; Su, W.W. Intein-ubiquitin chimeric domain for coordinated protein coexpression. *J. Biotechnol.* **2019**, *304*, 38–43. [CrossRef] [PubMed]
14. Su, W.W.; Zhang, B. Auto-Processing Domains for Polypeptide Expression. U.S. Patent 8,945,876, 3 February 2015.
15. Kramer, G.; Boehringer, D.; Ban, N.; Bukau, B. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat. Struct. Mol. Biol.* **2009**, *16*, 589–597. [CrossRef] [PubMed]
16. Netzer, W.J.; Hartl, F.U. Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* **1997**, *388*, 343–349. [CrossRef] [PubMed]
17. Jackson, S.E. Ubiquitin: A small protein folding paradigm. *Org. Biomol. Chem.* **2006**, *4*, 1845–1853. [CrossRef] [PubMed]
18. Ding, Y.; Xu, M.Q.; Ghosh, I.; Chen, X.; Ferrandon, S.; Lesage, G.; Rao, Z. Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing. *J. Biol. Chem.* **2003**, *278*, 39133–39142. [CrossRef] [PubMed]
19. Volkmann, G.; Mootz, H.D. Recent progress in intein research: From mechanism to directed evolution and applications. *Cell Mol. Life Sci.* **2013**, *70*, 1185–1206. [CrossRef] [PubMed]
20. Amitai, G.; Callahan, B.P.; Stanger, M.J.; Belfort, G.; Belfort, M. Modulation of intein activity by its neighboring extein substrates. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11005–11010. [CrossRef] [PubMed]
21. Iwai, H.; Züger, S.; Jin, J.; Tam, P.-H. Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme. *FEBS Lett.* **2006**, *580*, 1853–1858. [CrossRef] [PubMed]
22. Derbyshire, V.; Wood, D.W.; Wu, W.; Dansereau, J.T.; Dalgaard, J.Z.; Belfort, M. Genetic definition of a protein-splicing domain: Functional mini-inteins support structure predictions and a model for intein evolution. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 11466–11471. [CrossRef] [PubMed]
23. Mills, K.V.; Lew, B.M.; Jiang, S.-q.; Paulus, H. Protein splicing in *trans* by purified N- and C-terminal fragments of the *Mycobacterium tuberculosis* RecA intein. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 3543–3548. [CrossRef] [PubMed]
24. Telenti, A.; Southworth, M.; Alcaide, F.; Daugelat, S.; Jacobs, W.R.; Perler, F.B. The Mycobacterium xenopi GyrA protein splicing element: Characterization of a minimal intein. *J. Bacteriol.* **1997**, *179*, 6378–6382. [CrossRef] [PubMed]
25. Shemella, P.T.; Topilina, N.I.; Soga, I.; Pereira, B.; Belfort, G.; Belfort, M.; Nayak, S.K. Electronic structure of neighboring extein residue modulates intein C-terminal cleavage activity. *Biophys. J.* **2011**, *100*, 2217–2225. [CrossRef] [PubMed]
26. Ramirez, M.; Valdes, N.; Guan, D.; Chen, Z. Engineering split intein DnaE from Nostoc punctiforme for rapid protein purification. *Protein Eng. Des. Sel.* **2013**, *26*, 215–223. [CrossRef] [PubMed]
27. Varshavsky, A. The N-end rule pathway and regulation by proteolysis. *Protein Sci.* **2011**, *20*, 1298–1345. [CrossRef] [PubMed]
28. Kunes, Y.Z.; Gion, W.R.; Fung, E.; Salfeld, J.G.; Zhu, R.-R.; Sakorafas, P.; Carson, G.R. Expression of antibodies using single-open reading frame vector design and polyprotein processing from mammalian cells. *Biotechnol. Prog.* **2009**, *25*, 735–744. [CrossRef] [PubMed]
29. Gion, W.R.; Davis-Taber, R.A.; Regier, D.A.; Fung, E.; Medina, L.; Santora, L.C.; Bose, S.; Ivanov, A.V.; Perilli-Palmer, B.A.; Chumsae, C.M.; et al. Expression of antibodies using single open reading frame (sORF) vector design: Demonstration of manufacturing feasibility. *mAbs* **2013**, *5*, 595–607. [CrossRef] [PubMed]
30. François, I.E.J.A.; Bolle, M.F.C.D.; Dwyer, G.; Goderis, I.J.W.M.; Woutors, P.F.J.; Verhaert, P.D.; Proost, P.; Schaaper, W.M.M.; Cammue, B.P.A.; Broekaert, W.F. Transgenic Expression in Arabidopsis of a Polyprotein Construct Leading to Production of Two Different Antimicrobial Proteins. *Plant Physiol.* **2002**, *128*, 1346–1358. [CrossRef] [PubMed]
31. Fang, J.; Qian, J.-J.; Yi, S.; Harding, T.C.; Tu, G.H.; VanRoey, M.; Jooss, K. Stable antibody expression at therapeutic levels using the 2A peptide. *Nat. Biotechnol.* **2005**, *23*, 584–590. [CrossRef] [PubMed]
32. Odon, V.; Luke, G.A.; Roulston, C.; de Felipe, P.; Ruan, L.; Escuin-Ordinas, H.; Brown, J.D.; Ryan, M.D.; Sukhodub, A. APE-Type Non-LTR Retrotransposons of Multicellular Organisms Encode Virus-Like 2A Oligopeptide Sequences, Which Mediate Translational Recoding during Protein Synthesis. *Mol. Biol. Evol.* **2013**, *30*, 1955–1965. [CrossRef] [PubMed]

33. Walker, J.M.; Vierstra, R.D. A ubiquitin-based vector for the co-ordinated synthesis of multiple proteins in plants. *Plant Biotechnol. J.* **2007**, *5*, 413–421. [CrossRef] [PubMed]
34. Koren, I.; Timms, R.T.; Kula, T.; Xu, Q.; Li, M.Z.; Elledge, S.J. The Eukaryotic Proteome Is Shaped by E3 Ubiquitin Ligases Targeting C-Terminal Degrons. *Cell* **2018**, *173*, 1622–1635.e1614. [CrossRef] [PubMed]
35. Borman, A.M.; Le Mercier, P.; Girard, M.; Kean, K.M. Comparison of Picornaviral IRES-Driven Internal Initiation of Translation in Cultured Cells of Different Origins. *Nucleic Acids Res.* **1997**, *25*, 925–932. [CrossRef] [PubMed]
36. Ho, S.C.; Bardor, M.; Feng, H.; Mariati; Tong, Y.W.; Song, Z.; Yap, M.G.; Yang, Y. IRES-mediated Tricistronic vectors for enhancing generation of high monoclonal antibody expressing CHO cell lines. *J. Biotechnol.* **2012**, *157*, 130–139. [CrossRef] [PubMed]
37. Gouiaa, S.; Khoudi, H.; Leidi, E.O.; Pardo, J.M.; Masmoudi, K. 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.* **2012**, *79*, 137–155. [CrossRef] [PubMed]
38. Mizuguchi, H.; Xu, Z.; Ishii-Watabe, A.; Uchida, E.; Hayakawa, T. IRES-Dependent Second Gene Expression Is Significantly Lower Than Cap-Dependent First Gene Expression in a Bicistronic Vector. *Mol. Ther.* **2000**, *1*, 376–382. [CrossRef] [PubMed]
39. Martinez-Salas, E.; Francisco-Velilla, R.; Fernandez-Chamorro, J.; Embarek, A.M. Insights into Structural and Mechanistic Features of Viral IRES Elements. *Front. Microbiol.* **2018**, *8*, 2629. [CrossRef] [PubMed]
40. Jung, H.; Kim, J.-K.; Ha, S.-H. Use of animal viral internal ribosome entry site sequence makes multiple truncated transcripts without mediating polycistronic expression in rice. *J. Korean Soc. Appl. Biol. Chem.* **2011**, *54*, 678–684. [CrossRef]
41. Ha, S.-H.; Liang, Y.S.; Jung, H.; Ahn, M.-J.; Suh, S.-C.; Kweon, S.-J.; Kim, D.-H.; Kim, Y.-M.; Kim, J.-K. Application of two bicistronic systems involving 2A and IRES sequences to the biosynthesis of carotenoids in rice endosperm. *Plant Biotechnol. J.* **2010**, *8*, 928–938. [CrossRef] [PubMed]
42. Liu, Z.; Chen, O.; Wall, J.B.J.; Zheng, M.; Zhou, Y.; Wang, L.; Ruth Vaseghi, H.; Qian, L.; Liu, J. Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Sci. Rep.* **2017**, *7*, 2193. [CrossRef] [PubMed]
43. de Felipe, P.; Luke, G.A.; Brown, J.D.; Ryan, M.D. Inhibition of 2A-mediated 'cleavage' of certain artificial polyproteins bearing N-terminal signal sequences. *Biotechnol. J.* **2010**, *5*, 213–223. [CrossRef] [PubMed]
44. Donnelly, M.L.L.; Hughes, L.E.; Luke, G.; Mendoza, H.; ten Dam, E.; Gani, D.; Ryan, M.D. The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. *J. Gen. Virol.* **2001**, *82*, 1027–1041. [CrossRef] [PubMed]
45. Minskaia, E.; Nicholson, J.; Ryan, M. Optimisation of the foot-and-mouth disease virus 2A co-expression system for biomedical applications. *BMC Biotechnol.* **2013**, *13*, 67. [CrossRef] [PubMed]
46. Lee, D.S.; Lee, K.H.; Jung, S.; Jo, E.J.; Han, K.H.; Bae, H.J. Synergistic effects of 2A-mediated polyproteins on the production of lignocellulose degradation enzymes in tobacco plants. *J. Exp. Bot.* **2012**, *63*, 4797–4810. [CrossRef] [PubMed]
47. Paine, J.A.; Shipton, C.A.; Chaggar, S.; Howells, R.M.; Kennedy, M.J.; Vernon, G.; Wright, S.Y.; Hinchliffe, E.; Adams, J.L.; Silverstone, A.L.; et al. Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nat. Biotechnol.* **2005**, *23*, 482–487. [CrossRef] [PubMed]
48. Spatola Rossi, T.; Fricker, M.; Kriechbaumer, V. Gene Stacking and Stoichiometric Expression of ER-Targeted Constructs Using "2A" Self-Cleaving Peptides. In *The Plant Endoplasmic Reticulum: Methods and Protocols*; Springer: New York, NY, USA, 2024; pp. 337–351. [CrossRef] [PubMed]
49. Geier, M.; Fauland, P.; Vogl, T.; Glieder, A. Compact multi-enzyme pathways in P. pastoris. *Chem. Commun.* **2015**, *51*, 1643–1646. [CrossRef] [PubMed]
50. Souza-Moreira, T.M.; Navarrete, C.; Chen, X.; Zanelli, C.F.; Valentini, S.R.; Furlan, M.; Nielsen, J.; Krivoruchko, A. Screening of 2A peptides for polycistronic gene expression in yeast. *FEMS Yeast Res.* **2018**, *18*, foy036. [CrossRef] [PubMed]
51. Oda-Yamamizo, C.; Mitsuda, N.; Milkowski, C.; Ito, H.; Ezura, K.; Tahara, K. Heterologous gene expression system for the production of hydrolyzable tannin intermediates in herbaceous model plants. *J. Plant Res.* **2023**, *136*, 891–905. Erratum in *J. Plant Res.* **2023**, *136*, 947. [CrossRef] [PubMed]
52. Jiang, C.; Geng, L.; Wang, J.; Liang, Y.; Guo, X.; Liu, C.; Zhao, Y.; Jin, J.; Liu, Z.; Mu, Y. Multiplexed Gene Engineering Based on dCas9 and gRNA-tRNA Array Encoded on Single Transcript. *Int. J. Mol. Sci.* **2023**, *24*, 8535. [CrossRef] [PubMed]
53. Smole, A.; Lainšček, D.; Bezeljak, U.; Horvat, S.; Jerala, R. A Synthetic Mammalian Therapeutic Gene Circuit for Sensing and Suppressing Inflammation. *Mol. Ther. J. Am. Soc. Gene Ther.* **2017**, *25*, 102–119. [CrossRef] [PubMed]
54. Shibuta, M.K.; Sakamoto, T.; Yamaoka, T.; Yoshikawa, M.; Kasamatsu, S.; Yagi, N.; Fujimoto, S.; Suzuki, T.; Uchino, S.; Sato, Y.; et al. A live imaging system to analyze spatiotemporal dynamics of RNA polymerase II modification in Arabidopsis thaliana. *Commun. Biol.* **2021**, *4*, 580. [CrossRef] [PubMed]
55. Piskorz, E.W.; Steckenborn, S.; Cuacos, M.; Heckmann, S. Meiosis-specific protein co-expression in Arabidopsis thaliana: A protein delivery tool in meiotic cells. *Plant J.* **2026**, *125*, e70811. [CrossRef] [PubMed]
56. Liu, T.-Y.; Chou, W.-C.; Chen, W.-Y.; Chu, C.-Y.; Dai, C.-Y.; Wu, P.-Y. Detection of membrane protein–protein interaction in planta based on dual-intein-coupled tripartite split-GFP association. *Plant J.* **2018**, *94*, 426–438. [CrossRef] [PubMed]

57. Opdensteinen, P.; Meyer, S.; Buyel, J.F. *Nicotiana* spp. for the Expression and Purification of Functional IgG3 Antibodies Directed Against the Staphylococcus aureus Alpha Toxin. *Front. Chem. Eng.* **2021**, *3*, 737010. [CrossRef]
58. Daniele, J.R.; Chu, T.; Kunes, S. A novel proteolytic event controls Hedgehog intracellular sorting and distribution to receptive fields. *Biol. Open* **2017**, *6*, 540–550. [CrossRef] [PubMed]
59. Wu, Y.-M.; Ma, Y.-J.; Wang, M.; Zhou, H.; Gan, Z.-M.; Zeng, R.-F.; Ye, L.-X.; Zhou, J.-J.; Zhang, J.-Z.; Hu, C.-G. Mobility of FLOWERING LOCUS T protein as a systemic signal in trifoliate orange and its low accumulation in grafted juvenile scions. *Hortic. Res.* **2022**, *9*, uhac056. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.