Title: **The Protein Co-expression Problem in Biotechnology and Biomedicine: Virus 2A and 2A-like Sequences Provide a Solution.**

Authors: **Garry A. Luke & Martin D. Ryan**

Address: **Biomedical Sciences Research Complex,**
North Haugh, St Andrews KY16 9ST
Fife, Scotland, U.K.

e-mail **martin.ryan@st-andrews.ac.uk**
gal@st-andrews.ac.uk

Tel: **01334 463403**

Summary. Synthetic biology enables us to create genes virtually at will. Ensuring that multiple genes are efficiently co-expressed within the same cell – to assemble multimeric complexes, to transfer biochemical pathways, to transfer 'traits', is more problematic. Viruses such as picornaviruses accomplish exactly this task: they generate multiple, different, proteins from a single open reading frame. The study of how foot-and-mouth disease virus (FMDV) controls it's protein biogenesis lead to the discovery of a short oligopeptide sequence, '2A', that is able to mediate a co-translational 'cleavage' between proteins. 2A and '2A-like' sequences (from other viruses and cellular sequences) can be used to concatenate multiple gene sequences into a single gene, ensuring their co-expression within the same cell. These sequences are now being used in the treatment of cancer, in the production of pluripotent stem cells, to create transgenic plants and animals amongst a host of other biotechnological and biomedical applications.

Keywords. 2A oligopeptide, co-expression, gene therapy, induced pluripotent stem cells, transgenic plants, transgenic animals.

Protein Co-Expression: What's The Problem? The advances in polymerase chain reaction (PCR) technology alongside those in synthetic biology have transformed our capabilities in the design and creation of new gene structural / regulatory sequences. This in itself, however, is not enough. In the early days of human gene therapy or the production of transgenic animals or plants, only single transgenes could be used successfully - a prime example being the first human gene therapy: four year old Ashanti DeSilva, where a lesion in a single gene (adenosine deaminase - causing severe combined immune deficiency) was treated with a functional copy of the gene. However, for gene therapy of certain human genetic disorders, the production of high-value therapeutic proteins, or, the introduction of 'traits' into animals or plants *via* transgenesis, multiple different genes must be co-expressed *within the same cell.*

Co-expression of multiple proteins in bacterial cells can be solved in a relatively straightforward manner. The initiation of translation in bacteria occurs at short ribosome binding (Shine-Dalgarno) sequences proximal to the initiating AUG codon. Such sequences are internal to the messenger RNA (mRNA): this method of initiation is fundamentally different in eukaryotes where translation initiation factors must bind to a cap structure at the 5' end of an mRNA. In eukaryotes, once the stop codon has

been reached at the end of the open reading frame (ORF), translation terminates and must normally re-initiate *via* a 5′ mRNA cap structure. In bacteria, genes requiring co-ordinated co-expression can simply be concatenated using short intergenic regions encoding ribosome binding sites – an operon. Here, multiple genes (cistrons) are transcribed into a polycistronic mRNA where each gene is translated independently by internal initiation of translation at the multiple (intergenic) ribosome binding sites. Due to the completely different method of initiation of translation, this strategy of co-expression cannot be used in eukaryotes.

Whilst transfection of cells with a mixture of different plasmid DNAs (each encoding a different transgene) may produce effective co-expression taken across the *population* of cells, it will only produce a small proportion of *individual* cells co-expressing all of the products – albeit with a range of comparative levels of expression of each transgene. Naturally, the transcription of these individual, dispersed, genes must be controlled in a co-ordinated manner. Long-term stable expression requires integration of the different plasmid DNAs into the genome: without sophisticated 'targeting' strategies, the integration of individual transgenes occurs at sites dispersed – at random - throughout the genome. This leads to genetic instability in the longer-term, and the consequential loss of the desired phenotype due to the loss of expression of one, or more, of the transgenes.

Why is the ability to co-express multiple proteins within the same cell so important? An excellent example is the production of antibodies: both the heavy and light chains need to co-expressed within the same cell so that the different chains can be assembled into a functional complex as they are exported from the cell. Expressing the heavy chain in one cell and the light chain in another cell would not produce a functional product. This principle applies across the board: production of therapeutic proteins (protein complexes), gene therapy and, for example, the introduction of biochemical pathways into transgenic organisms (*e.g*. 'Golden Rice').

In attempts to improve the efficiency of co-ordinated co-expression, various strategies have been employed: placing multiple genes on the same plasmid DNA, using multiple divergent or tandem promoters – but none have proved to be a significant advance. One solution to this problem is to concatenate sequences encoding proteins *via* linkers (which comprise host-cell proteinase cleavage sites) into a single ORF. This strategy has the disadvantage of problems which may arise from tissue/species specificity of proteinase expression, and, that 'processing' of the fusion protein is post-, and not co-, translational. The latter aspect precludes strategies involving the co-expression of cellular proteins which are either secreted from the cell, located within the lumen/membranes of cytoplasmic vesicular structures (excluding mitochondria), or, are plasma membrane proteins – some 39% of all mammalian proteins. Viruses, such as FMDV, encode proteinase domains within their polyproteins which 'process' precursors into mature products. A common problem here is that although these virus-encoded proteinases are specific for the virus polyprotein, they are cytotoxic in that they have evolved to also cleave certain key host-cell proteins to manipulate host-cell metabolism or modify macromolecular structures to promote virus replication – not a desirable property for a co-expression system. Certain virus-encoded proteins have been used to great effect in cleaving expressed, purified, fusion proteins *in vitro* (the plant virus Tobacco Etch Virus proteinase, for example), but the cytotoxic 'off-target' cleavage of host-cell proteins *in vivo* remains a problem. Re-iterative rounds of transgenesis combined with selection has been used successfully in, for example, the production of plant transgenes - but this strategy simply cannot be used in many biomedical applications where the window to treat the disease in question may be limited.

Protein Co-Expression: What's The Solution? Fortunately for biotechnologists, it turns out that the co-expression problem has already been the subject of highly intensive experimentation – indeed, billions

upon billions of experiments: very elegant solutions to the co-expression problem have been developed - by viruses. The genome structure of picornaviruses, such as FMDV, have the same overall architecture as cellular mRNAs: a 5’ non-coding region (5’NCR), a single ORF (in picornaviruses encoding a ‘polyprotein’), a 3’ non-coding region and a poly-A tail (Figure 1). Picornaviruses co-express capsid proteins, proteins involved in altering host-cell macromolecular structures and metabolism, plus proteins to replicate the virus RNA genome – all from a single ORF. In most cases this is accomplished by the virus encoding its own proteinases (domains of the polyprotein) which serve to cleave the polyprotein substrate at specific sites to produce the individual ‘processing’ products. Parenthetically, these proteinases also serve to degrade specific host-cell proteins to promote virus replication (reviewed in [1]). However, relatively recently an alternative method of controlling protein biogenesis was discovered – ribosome ‘skipping’ – which has lead to a revolution in protein co-expression technology.

The N-terminal protein of the FMDV polyprotein is a proteinase (L^{pro}), which cleaves at its own C-terminus. L^{pro} also cleaves the cellular translation initiation factor eIF4G – ‘shutting off’ host-cell cap-dependent mRNA translation. The FMDV genome is, however, translated from an internal ribosome entry sequence (IRES; Figure 1, Panel A) in a cap-independent mechanism. In this manner the virus sequesters the resources of the cell for its own replication. Indeed, picornavirus IRESes provided an early method of creating a bicistronic mRNA such that 2 proteins could be co-expressed from a single mRNA (Figure 1, Panel B; discussed in [1]). The first ORF is translated in a cap-dependent manner, the second ORF in a cap-independent manner - driven by the IRES. The draw-back of this system is that the second ORF is translated only to some 10% of the first ORF: this ‘polarity’ effect being exacerbated when one uses multiple IRESes to express more than 2 genes.

A co-translational ‘cleavage’ of the FMDV polyprotein occurs at the C-terminus of 2A (only 18aa long; Figure 1, Panel A). Analyses of recombinant FMDV polyproteins indicated the FMDV 2A oligopeptide appeared to mediate this cleavage without the involvement of other FMDV proteins. This hypothesis was subsequently confirmed by inserting the FMDV 2A sequence (together with the N-terminal proline of protein 2B – collectively referred-to as ‘2A’) into artificial polyprotein systems. These systems comprised two ‘reporter’ proteins flanking FMDV 2A: a single ORF was created by removing the stop codon of the protein upstream of 2A (Figure 1, Panel C). Analyses of these types of construct in cell-free translation systems and transfected cells showed;

- the FMDV 2A oligopeptide sequence mediated a highly efficient ‘cleavage’ (>90%)
- 2A ‘cleaved’ at it’s own C-terminus – just like in the FMDV polyprotein
- 2A-mediated ‘cleavage’ was co-, and not post-, translational

Importantly, these artificial polyprotein systems, designed for analyses of the mechanism of the 2A-mediated ‘cleavage’, provided the first demonstration that 2A could be used to co-express multiple proteins [2-4].

The 2A Co-Expression System. Using this approach of artificial self-processing polyprotein systems in plant, yeast, insect and mammalian cells, the indication was that 2A could work in all eukaryotic expression systems and (unlike IRESes) worked highly efficiently in all cell types. ‘2A-like’ sequences were identified from other viruses and cellular genomes and a number proved to match, or exceed, the ‘cleavage’ efficiency of the FMDV 2A sequence [5,6]. Our analyses of the mechanism showed that 2A was neither a substrate for a host-cell proteinase, nor a proteolytic element itself, but mediated a ribosomal ‘skipping’ event in which the synthesis of a specific peptide bond was ‘skipped’: translation terminated at the C-terminus of 2A, but could re-initiate at the N-terminal proline of the downstream protein. The

'cleavage' products were, in fact, generated as discrete translation products. Our model of 2A-mediated, co-translational, 'cleavage' proposes the nascent 2A oligopeptide interacts with the ribosome exit tunnel. This interaction leads to a modification and restriction of the conformational space of the peptidyl-tRNA within the peptidyl-transferase centre of the ribosome - 'jamming' the process of polypeptide elongation. This jam is overcome by release (termination) factors eRF1 and eRF3 releasing the nascent protein, then the ribosome may re-commence elongation of the downstream protein: the individual products do not arise from a substrate undergoing proteolysis, but are actually synthesised as discrete translation products (reviewed in [2]).

This mechanism imparts no 'polarity' on the system – more than 2 proteins can be concatenated without affecting the 'cleavage' at other 2A linker sequences. The 2A system does however, have drawbacks: (i) the system does not work in prokaryotic cells, (ii) the 2A (or 2A-like) oligopeptide sequence remains as a C-terminal extension (some 18-25aa) of the upstream protein and (iii) the downstream protein must have an N-terminal proline residue. Although an N-terminal proline confers a long half-life upon a protein, it does preclude many N-terminal post-translational modifications that may be essential for activity. If this is the case, such a protein should be placed first in the chain of concatenated sequences.

For the reasons outlined above, the 2A system is not ideal - but it is the best, and has proven to be an extremely useful tool in biotechnology and biomedicine. For the first time multiple proteins could be co-expressed – in equal stoichiometry – from a single promoter. Our translational model of 2A-mediated 'cleavage' predicted that 2A modified the translational apparatus – 2A worked within the ribosome. To test this model we incorporated co-translational signal sequences immediately downstream of 2A. If the model was correct, these signal sequences would be recognised as nascent N-terminal features by signal recognition particle and target the second protein to the exocytic pathway [7]. This indeed proved to be the case, adding more support for our model - but also imparting a new dimension to the utility of the system: one could not only co-express multiple proteins, but potentially target individual components to different sub-cellular sites. It should be noted, however, that the 2A-mediated cleavage may be inhibited in the case of some proteins targeted to the exocytic pathway [8].

Although individual cells express these self-processing polyproteins at different levels, the key point is that within a cell, each component of the polyprotein is expressed at the same levels. This is shown in figure 2 (panel A), showing images of cells transfected with a plasmid encoding green fluorescent protein, linked *via* Thosea asigna virus 2A (TaV2A), to cherry fluorescent protein (cherryFP). The [GFP-TaV2A-CherryFP] ORF is translated into [GFP-TaV2A] and CherryFP – no uncleaved [GFP-TaV2A-CherryFP] is detected (data not shown). Fluorescence image analyses (Figure 2, panel A) show that these proteins are expressed to the same level within any given cell.

2A was first characterised in FMDV, although it was apparent that other picornaviruses encoded 'ribosome skipping' 2As, and that this method of controlling protein biogenesis was also used by a wide range of other RNA viruses: indeed, some of these viruses encode multiple 2A-like sequences [5]. A number of these virus '2A-like' sequences have been used in biomedicine and biotechnology (Table 1), but a wider range of other, highly efficient, 2As could be used (Table 2). Some researchers seeking to use 2A have expressed concern with regards the public acceptance of products comprising sequences derived from viruses, but we have identified other efficient 2A-like sequences from cellular genes which could be used to replace virus sequences [6, unpublished observations] (Table3).

Biotechnological Applications. Since the early observations on the properties of the 2A oligopeptide sequence were published in virology journals, many early uses involved the creation of recombinant virus genomes (*e.g*. influenza virus, poliovirus, plant potex- and comoviruses). For cellular systems, early

1 applications were relatively simple: (i) to link the sequences encoding the two different components of a
2 heterodimer (*e.g*. high-value therapeutic proteins such as interleukin 12, monoclonal antibodies), (ii) to
3 link two different reporter proteins to provide proof-of-principle for co-expression in various target cell-
4 types/organisms, or, (iii) to monitor the expression of a (trans)gene by linking it to a marker protein(s)
5 (*e.g*. GFP, luciferase). Using the latter approach, the site of expression of a therapeutic transgene within
6 an organism could be studied by histology or whole-body imaging.
7 Once the utility of the system became more apparent, more ambitious constructs were
8 assembled – notably the co-expression of all of the components of an active T-cell receptor complex (6
9 different proteins – see below). Such a dramatic demonstration of the utility of 2A helped raise the
10 'profile' and the potential of the system. Over time data accumulated from a wider range of proteins and
11 eukaryotic cell-types showing that, indeed, the 2A co-expression system worked in all eukaryotic cell-
12 types tested: yeast, fungal, plant, insect and mammalian. An impression of the utility of the this co-
13 expression system can be gained by viewing the range and huge number of publications citing the use of
14 2A (http://www.st-andrews. ac.uk/ryanlab/page10.htm).
15
16 **Transgenic Organisms.** Breeders classically 'stack' genes by a program of re-iterative crossing between
17 parents each with a desired trait, then identifying offspring expressing both traits. In some cases,
18 however, the generation time of the target species places a major constraint on this process rendering it
19 impractical. Gene stacking by genetic engineering is a term used in the plant sciences, but here we will
20 use the term, in the sense of a general procedure, for the production of both plant and animal genetically-
21 modified organisms (GMOs). In many publications researchers have chosen to provide a proof-of-principle
22 by co-expression of multiple fluorescent proteins, since these are encoded by relatively small genes, the
23 translation products are readily detected by microscopy and the efficiency of co-expression
24 demonstrated.
25 ***(i) Transgenic Plants.*** The early analyses of the mechanism of 2A-mediated 'cleavage' showed 2A was
26 equally active in animal- (rabbit reticulocyte lysates) and plant-based (wheat-germ extracts) cell-free
27 translation systems. It was also shown that (i) 2A could be used to co-express proteins in transgenic
28 tobacco cells and (ii) 2A could be used to manipulate plant virus genomes. Indeed, research was
29 conducted to use (non-GMO) plants as production platforms - programmed with recombinant plant
30 viruses encoding 2A to co-express high-value proteins, either as free proteins, or, on the surface of plant
31 virus particles. Since then, 2A has been used to create a wide range of transgenic plants; drought-resistant
32 crops, crops with improved nutritional values, the production of a new generation of golden rice, to
33 engineer crops to produce 'nutraceuticals', to engineer plant metabolic pathways and, more recently, to
34 stack the genes for glyphosate-resistance with BT-toxins (the World's two most common transgenes) in
35 the form of a self-processing polyprotein [9-17].
36 ***(ii) Transgenic Animals.*** 2A has been used successfully in the production of a wide range of transgenic
37 animals; mice [18-27], fish [28, 29], rats [30], pigs [31, 32], birds [33], amphibians [34], insects [35] and
38 sheep [36]. Again, many publications describing the generation of GMOs to provide a proof-of-principle
39 (using reporter genes), to create transgenic animals as research tools, for developmental studies, for
40 histological studies, as models of human diseases, to produce animal models of human disease, develop
41 therapies etc. – but, to date, there are no reports of the use of 2A in the introduction of traits designed to
42 increase nutritional values, enhance productivity/disease resistance or to add value to the animal product
43 – the technology remains largely as a research tool.
44

Biomedical Applications. Politics, particularly in Europe, has substantially delayed the use of GMOs in agriculture. In the field of biomedicine, however, the story could not be more different. Here, 2A has been used not only as an effective research tool to monitor the expression of therapeutic transgenes by linking them, *via* 2A, to marker proteins [37], but also as an integral part of effective therapies in the clinic. Literally hundreds of papers have been published citing the use of 2A in a broad range of biomedical applications. It is beyond the scope of this review to do justice to all of the work, but the work outlined in the sections below describes some of the truly astonishing recent advances in molecular medicine: astonishing not only in the scientific vision that underpins these advances, but also the rapidity in which these developments have reached patients.

(i) 'Transferable' Immune Responses (TIRs). Herd (or community) immunity is present when the vaccination of a portion of a population is sufficient to provide an effective measure of protection for susceptible individuals: chains of infection are likely to be broken and the basic reproduction number (R_o) of the infectious agent falls below 1. Over the past few years, however, we have witnessed the development of technologies which will bring about another form of immunity arising not from the immune repertoire of the individual in question, but from the repertoire of the 'herd': 'transferable immune responses'. Here, a susceptible individual could benefit from the immune response of another: not indirectly, by breaking a chain of transmission, but in a direct manner by the transfer of genetic information – conferring an immune trait from another individual in the population.

(a) Adoptive Cell Transfer (Engineered T-cells). In the 1980s it was shown that cancers regressed following treatment with autologous tumour-infiltrating lymphocytes. Tumour-associated antigens may be recognised by T-cell receptors (TCRs), composed of alpha and beta chains. Once reactive TCRs were identified and purified, it was possible to clone the genes encoding these chains, chains that were critical in the T-cells recognising and killing the cancerous cells. This lead to a new form of treating cancer: adoptive cell transfer (ACT). In essence, peripheral blood lymphocytes (PBLs) are isolated from the patient, transduced with genes encoding the desired TCR which targets the cancer cells (*ex vivo* gene therapy), and the (autologous) genetically-modified cells introduced back into the patient. The cytotoxic T-cells attack and destroy the cancer cells. Initially, genes encoding the alpha and beta chains were introduced as individual genes, or, linked by an IRES to improve co-expression. Again, the use of 2A to co-express these chains proved to be a substantial advance in their co-expression and assembly into a functional complex [18].

This strategy showed that the T-cell repertoire of an individual could be modified or expanded by harnessing – *via* gene transfer – the result of a successful immune response mounted by another individual. Indeed, this form of ACT using 2A to co-express TCR alpha and beta chains has been used to treat a range of cancers: metastatic melanoma, synovial cell sarcomas, colorectal cancer and renal cell carcinomas, but the list is expanding rapidly [38-45]. Naturally there are problems associated with this type of therapy, such as (i) targeting of normal tissues (autoreactivity) expressing the cognate antigen and the formation of 'chimeric' TCRs in which an exogenous (gene transferred) receptor chain associates with an endogenous chain, (ii) loss of potency during T-cell manipulation before transduction, and (iii) the relatively shorter life of peripheral blood T-cells: problems which will be overcome. ACT represents a very exciting development in the field of cancer therapy.

(b) Engineered B-cells. 'Passive' immunity against pathogens, or toxins, may be achieved simply by the transfer of antibodies (animal or human in origin) to the patient. Like TCRs, the production of functional antibodies requires co-expression (heavy and light chains) within the same cell. The use of 2A to express the two different antibody chains was carefully optimised by Cell Genesys Inc. [46, 47]. The

heavy chain (plus it's signal sequence) was linked to the light chain (plus it's signal sequence) *via* 2A within a recombinant adeno-associated virus vector (rAAV). Therapeutic levels of antibodies were secreted from rAAV transduced mouse cells. Incorporation of a furin proteinase cleavage site between the heavy chain and 2A resulted in the C-terminal extension of 2A being 'trimmed' away by the cellular furin – a proteinase located mainly in the Golgi apparatus.

This approach of heavy/light chain co-expression was transformed into a new therapeutic strategy by the production of recombinant B-cells. However, were mature B-cells to be transduced with such a [heavy chain-2A-light chain] type of construct, the same problem of mixing of the endogenous and exogenous antibody chains would be encountered, as outlined above for the T-cell receptor alpha and beta chains. The solution was to transduce (naive) human hematopoietic stem/progenitor cells (HSPCs), then to drive differentiation (*in vitro*) of the transduced HSPCs into mature B-cells: the most potent antibody-producing cells [48, 49]. This approach was adopted to produce anti-HIV antibodies, but the principle of an immune response, transferable from one individual to another, was demonstrated. Subsequently this approach was extended to T-cells by the transduction of HSPCs with disease-specific TCRs, leading to the generation of long-lasting and functional cytotoxic T-cells, solving the chain-mixing and T-cell longevity problems [50].

***(ii) Pluripotent Stem Cells*.** One major area of regenerative medicine is the use of stem cells to regenerate damaged tissues. The main sources of autologous stem cells in adults are the bone marrow, adipose tissue and blood. The problems associated with isolation of stem cells from such tissues could be overcome by *producing* stem cells from differentiated tissue, rather than their direct isolation from the body. Such a technology would also circumvent the ethical issues surrounding the isolation and use of embryonic stem cells. Combinations of genes known to be particularly important in embryonic stem cells were transduced into (differentiated) mouse fibroblasts. In this manner, four genes were identified (Oct-3/4, SOX2, c-Myc, and Klf4) which, when co-expressed in the same cell, lead to the production of induced pluripotent stem cells (iPSCs): the differentiation into a fibroblast had been reversed. Initially, these individual genes were co-expressed using multiple lentivirus vectors, but the laboratory of Shinya Yamanaka at Kyoto University chose to link these genes, *via* 2As, to create a single, self-processing, polyprotein [51]. This strategy ensured the co-expression of the multiple proteins within the same cell - vital for iPSC production. The use of 2A for co-expression has been adopted by many laboratories to produce iPSCs [52-74]. The technologies developed to produce iPSCs has advanced with astonishing rapidity – a reflection of the huge potential in the field of regenerative medicine. Patient-specific iPSCs can now be produced *ex vivo* for the administration of cells to treat disease, but one can conceive of transduction of cells *in vivo* – a gene therapy approach to tissue regeneration.

Future Perspective. Basic research into how FMDV generates multiple proteins from a single open reading frame led to the discovery of how the 2A oligopeptide mediates a co-translational 'cleavage': a discovery that has facilitated a bewildering array of biotechnological and biomedical applications.

In some ways, however, the true potential of this system remains to be exploited: to date only a few publications cite the use of 2A in engineering or the transfer of biochemical pathways. Here one thinks of 'dual use' crops (e.g. eat one part, ferment another for biofuels), improving the nutritional properties of crops, improving abiotic and abiotic/biotic stress resistance (drought/salinity/pests/viruses), creation of crops able to fix nitrogen, and the genetic modification of algae/fungi/yeasts for biofuel production. In animal biotechnology one naturally thinks of disease resistance (multivalent vaccines, anti-microbial peptides, modification of the innate immune system etc.), the production of high-value or

therapeutic proteins (*e.g*. the ill-fated PPL Therapeutics) and productivity (*e.g*. AquAdvantage salmon). In the arena of human health, 2A conferring the ability to perform more complex transgenesis has opened the door to new strategies of immuno-therapy: not only of cancer, but potentially of much wider significance: methamphetamine abuse for example [75], but also monoclonal antibodies directed against other small molecules such as the tobacco-specific nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone [76].

A Sensor of Translational Stress? A further potential future use of 2A arises from our model of the mechanism of 2A-mediated cleavage. The model invokes a key step in the re-initiation of the translation of sequences downstream of 2A, mediated by eukaryotic elongation factor 2 (eEF2) [1-4, 77]. The activity of eEF2 is regulated by eEF2 kinase. The activity of eEF2 kinase is also regulated by phosphorylation, but in this case by a series of signalling pathways related to different forms of cellular stress (hypoxia, temperature, amino-acid starvation, alcohol etc.), such that when cells become stressed both the initiation and the elongation phases of translation are down-regulated.

The hypothesis we are currently testing is that as cells become stressed (such as FMDV infection), eEF2 becomes increasingly phosphorylated, leading to a progressive reduction in the rate of polypeptide elongation. Re-initiation of translation of the downstream sequences would be very sensitive to such a reduction: termination at the 2A site would be progressively *increased* and re-initiation progressively *decreased*. The net effect being an increase in the molar ratio of upstream : downstream products as a response to increasing cellular stress. In the case of FMDV infection, this would result in the synthesis of a higher ratio of capsid proteins : replication proteins as the infectious cycle progresses. At the latter stages of infection, what remains of the cell's resources (amino-acyl tRNAs) would be progressively targeted to the synthesis of capsid, and not replication, proteins – increasing the yield of virus particles. If this indeed proves to be the case, 2A could be used not just to co-express proteins – but to be used as a 'sensor' of translational stress. Since 2A works in all eukaryotic systems tested to date, such a sensor could be used for both plant and animal biotechnologies.

Dual Protein Targeting. Recently the complete genome sequence of the purple sea urchin (*Strongylocentrotus purpuratus*) was published [78]. Our bioinformatic analyses of the genome showed 2A-like sequences were detected in two major types of gene: non-LTR retrotransposons and 'CATERPILLER' proteins of the innate immune system. We were perplexed, however, in that these latter genes 2A-like sequences occurred at the N-terminus of the ORF. What could be the purpose of a self-cleaving sequence at the N-terminus? We have recently shown that these 2A-like sequences are active in mediating 'cleavage' (to various degrees) – but that they may also function as signal sequences, targeting the (downstream, 'uncleaved') protein to the exocytic pathway. In the case of one 2A-like sequence (STR6-2A; Table 3), the wild-type sequence shows high 'cleavage' activity: if the signal sequence 'cleaves' itself from the downstream protein then the protein localises to the cytoplasm. The 2A-like signal sequence 'cleaves' itself away from the downstream protein *within the ribosome* such that the protein emerges from the ribosome without a signal at its N-terminus. If the 2A-like signal sequence does not cleave, however, the signal remains attached, is recognised by signal recognition particle, and the entire protein is targeted to the exocytic pathway. To demonstrate this effect, we fused such a 2A-like signal sequence to CherryFP. The wild-type (signal) sequence (STR6-2A^{wt}; Table 3) cleaved itself from CherryFP (data not shown), such that CherryFP – lacking any signal sequence - was localised to the cytoplasm (Figure 2, Panel B). A site-directed mutant form (STR6-2A^{mut}; Table 3) is 'cleavage' inactive (data not shown): in this case the 2A-like signal sequence remains attached to the downstream protein and targets the entire protein to the exocytic pathway (Figure 2, Panel B). This represents, therefore, a novel form of dual protein targeting.

We are currently both characterising natural sequence variants and developing mutants to provide a range of sequences which could be used to determine the proportion of the expressed protein which partitions between localisation in the cytoplasm and secretion from the cell. A single (trans)gene translation product could, therefore, be localised in both the cytoplasm *and* be secreted from the cell. This entirely new property of this class of 2A-like sequences could be of utility in biotechnological and biomedical applications.

The biology associated with this oligopeptide sequence has proven to be fascinating and the diverse uses to which to 2A has been put is amazing: to paraphrase - '*never in the field of biotechnology was so much owed by so many to so few (amino acids)*'.

1 **References:**
2
3 *** of interest**
4 **** of considerable interest**
5

6 1. Martínez-Salas E, Ryan MD. Translation and Protein Processing. In: *Picornaviruses: Molecular Biology,*
7 *Evolution and Pathogenesis*. Ehrenfeld, E., Domingo, E. & Roos, R.P. eds. ASM Press, Washington DC,
8 USA, 141- 161 (2010).
9 2. Brown JD, Ryan MD. Ribosome “Skipping”: “Stop-Carry On” or “StopGo” Translation. In: *Recoding:*
10 *Expansion of Decoding Rules Enriches Gene Expression*. Atkins J.F. and Gesteland, R.F eds. Springer,
11 101-122 (2010).
12 3. Doronina VA, Wu C, de Felipe P, Sachs MS, Ryan MD, Brown JD. Site-specific release of nascent chains
13 from ribosomes at a sense codon. *Mol. Cell. Biol*. 28(13), 4227-4239 (2008).
14
15 *** More complete elucidation of the 2A ‘cleavage’ mechanism, showing the involvement of**
16 **translation release (termination) factors.**
17
18 4. Sharma P, Yan F, Doronina V, Escuin-Ordinas H, Ryan MD, Brown JD. 2A peptides provide distinct
19 solutions to driving stop-carry on translational recoding. *Nuc. Acids Res*. 40(7), 1-9 (2011).
20 5. Luke GA, de Felipe P, Lukashev A, Kallioinen SE, Bruno EA, Ryan MD. The occurrence, function and
21 evolutionary origins of ‘2A-like’ sequences in virus genomes. *J. Gen. Virol.* 89(4), 1036-1042 (2008).
22 6. Odon V, Luke GA, Roulston C *et al*. APE-type non-LTR retrotransposons of multicellular organisms
23 encode virus-like 2A oligopeptide sequences, which mediate translational recoding during protein
24 synthesis. *Mol. Biol Evol.* 30(8), 1955-1965 (2013).
25 7. de Felipe P, Luke GA, Hughes LE, Gani D, Halpin C, Ryan MD. *E unum pluribus*: multiple proteins from
26 a self-processing polyprotein. *Trends in Biotechnology* 24(2), 68-75 (2006).
27 8. de Felipe P, Luke GA, Brown JD, Ryan MD. Inhibition of 2A-mediated 'cleavage' of certain artificial
28 polyproteins bearing N-terminal signal sequences. *Biotechnol. J.* 5(2), 213-223 (2010).
29
30 *** Identification of problems associated with 2A-mediated co-expression of some proteins targeted**
31 **to the exocytic pathway.**
32
33 9. Halpin C, Cooke SE, Barakate A, El Amrani A, Ryan MD. Self-processing polyproteins - a system for co-
34 ordinate expression of multiple proteins in transgenic plants. *Plant J.* 17(4), 453-459 (1999).
35 10. François IE De Bolle MF, Dwyer G *et al*. Transgenic expression in Arabidopsis of a polyprotein
36 construct leading to production of two different antimicrobial proteins. *Plant Physiol.* 128(4), 1346-
37 1358 (2002).
38 11. Kwon SJ, Hwang EW, Kwon HB. Genetic engineering of drought resistant potato plants by co-
39 introduction of genes encoding trehalose-6-phosphate synthase and trehalose-6-phosphate
40 phosphatase of *Zygosaccharomyces rouxii*. *Korean J. Genet.* 26(2), 199-206 (2004).
41 12. Randall J, Sutton D, Ghoshroy S, Bagga S, Kemp JD. Co-ordinate expression of beta- and delta-zeins in
42 transgenic tobacco. *Plant Sci*. 167, 367-372 (2004).
43 13. Yasuda H, Tada Y, Hayashi Y, Jomori T, Takaiwa F. Expression of the small peptide GLP-1 in transgenic
44 plants. *Transgenic Res*. 14(5), 677-684 (2005).

14. Obro J, Borkhardt B, Harholt J, Skjot M, Willats WGT, Ulvskov P. Simultaneous *in vivo* truncation of pectic side chains. *Transgenic Res*. 18(6), 961-969 (2009).
15. Ha SH, Liang YS, Jung H *et al*. Application of two bicistronic systems involving 2A and IRES sequences to the biosynthesis of carotenoids in rice endosperm. *Plant Biotech. J.* 8(8), pp. 928–938 (2010).
16. Petrie JR, Shrestha P, Belide S *et al*. Transgenic production of arachidonic acid in oilseeds. *Transgenic Res.* 21(1), 139-147 (2012).
17. Sun H, Lang Z, Zhu L, Huang D. Acquiring transgenic tobacco plants with insect resistance and glyphosate tolerance by fusion gene transformation. *Plant Cell Rep.* 31(10), 1877-1887 (2012).
18. Szymczak AL, Workman CJ, Wang Y *et al*. Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nature Biotech*. 22(5), 589-94 (2004).

**** Description of the use of 2A to generate transgenic mice expressing all of the components of a functional T-cell receptor complex.**

19. Alli R, Nguyen P, Geiger TL. Retrogenic modeling of experimental allergic encephalomyelitis associates T cell frequency but not TCR functional affinity with pathogenicity. *J. Immunol*. 181(1), 136-145 (2008).
20. Chaparro RJ, Burton AR, Serreze DV, Vignali DA, DiLorenzo TP. Rapid identification of MHC class I-restricted antigens relevant to autoimmune diabetes using retrogenic T cells. *J. Immunol. Methods* 335(1-2), 106-115 (2008).
21. O'Connell-Rodwell CE, Mackanos MA, Simanovskii D *et al*. In vivo analysis of heat-shock-protein-70 induction following pulsed laser irradiation in a transgenic reporter mouse. *J. Biomed. Optics* 13(3), Article # 030501 (2008).
22. Trichas G, Begbie J, Srinivas S. Use of the viral 2A peptide for bicistronic expression in transgenic mice. *BMC Biology* 6:40, Article #40 (2008).
23. Bettini ML, Bettini M, Vignali DA. T-cell receptor retrogenic mice: a rapid, flexible alternative to T-cell receptor transgenic mice. *Immunology* 136(3), 265-272 (2012).
24. Fang R, Peng YQ, Zheng M, Meng QY. Muscle-specific expression of delta-12 and omega-3 fatty acid desaturases and human catalase using "self-cleaving" 2A peptides in transgenic mice. *Prog. Biochem. Biophys*. 39, 175-180 (2012).
25. Rawlins EL, Perl AK. The a"MAZE"ing world of lung-specific transgenic mice. *Am. J. Respir. Cell. Mol. Biol*. 46(3), 269-282 (2012).
26. Tittel AP, Heuser C, Ohliger C *et al*. Functionally relevant neutrophilia in CD11c diphtheria toxin receptor transgenic mice. *Nat. Methods* 9, 385-390 (2012).
27. Peter M, Bathellier B, Fontinha B, Pliota P, Haubensak W, Rumpel S. Transgenic mouse models enabling photolabeling of individual neurons in vivo. *PLoS One* 8:e62132 (2013).
28. Provost E, Rhee J, Leach SD. Viral 2A peptides allow expression of multiple proteins from a single ORF in transgenic zebrafish embryos. *Genesis* 45(10), 625-629 (2007).
29. Dempsey WP, Fraser SE, Pantazis P. PhOTO zebrafish: a transgenic resource for in vivo lineage tracing during development and regeneration. *PLoS One* 7:e32888 (2012).
30. Herold MJ, van den Brandt J, Seibler J, Reichardt HM. Inducible and reversible gene silencing by stable integration of an shRNA-encoding lentivirus in transgenic rats. *Proc. Natl. Acad. Sci. USA* 105(47), 18507–18512 (2008).
31. Yang D, Wang CE, Zhao B *et al*. Expression of Huntington's disease protein results in apoptotic neurons in the brains of cloned transgenic pigs. *Hum Mol Genet*. 19(20), 3983-3994 (2010).
32. Deng W, Yang D, Zhao B *et al*. Use of the 2A peptide for generation of multi-transgenic pigs through a single round of nuclear transfer. *PLoS One* 6, e19986 (2011).

33. Bower DV, Sato Y, Lansford R. Dynamic lineage analysis of embryonic morphogenesis using transgenic quail and 4D multispectral imaging. *Genesis* 49(7), 619-463 (2011).
34. Lin G, Chen Y, Slack JM. Transgenic analysis of signaling pathways required for Xenopus tadpole spinal cord and muscle regeneration. *Anat. Rec.* (Hoboken) 295(10), 1532-1540 (2012).
35. Diao F, White BH. A novel approach for directing transgene expression in drosophila: T2A-Gal4 in-frame fusion. *Genetics* 190(3), 1139-1144 (2012).
36. Tian Y, Li W, Wang L *et al*. Expression of 2A peptide mediated tri-fluorescent protein genes were regulated by epigenetics in transgenic sheep. *Biochem. Biophys. Res. Commun*. 434(3), 681-687 (2013).
37. Osborn MJ, Panoskaltsis-Mortari A, McElmurry RT *et al*. A picornaviral '2A-like' sequence based tricistronic vector allowing for high level therapeutic gene expression coupled to a dual reporter system. *Mol. Ther*. 12, 569-574 (2005).
38. Chhabra, A., Yang, L., Wang, P., *et al*. CD4+ CD25- T Cells transduced to express MHC class I-restricted epitope-specific TCR synthesize Th1 cytokines and exhibit MHC class I-restricted cytolytic effector function in a human melanoma model. *J. Immunol.* 181(2), 1063-1070 (2008).
39. Johnson LA, Morgan RA, Dudley ME *et al*. Gene therapy with human and mouse T-cell receptors mediates cancer regression and targets normal tissues expressing cognate antigen**.** *Blood* 114(3), 535-546 (2009).

**** Gene therapy clinical trial using patient T-cells genetically modified to recognize antigens associated with metastatic melanoma.**

40. Parkhurst MR, Joo J, Riley JP *et al*. Characterization of genetically modified T-cell receptors that recognize the CEA:691-699 peptide in the context of HLA-A2.1 on human colorectal cancer cells. *Clin. Cancer Res.* 15(1), 169-180 (2009).
41. Govers C, Sebestyén Z, Coccoris M, Willemsen RA, Debets R. T cell receptor gene therapy: strategies for optimizing transgenic TCR pairing. *Trends Mol. Med*. 16(2), 77-87 (2010).
42. Hudecek M, Schmitt TM, Baskar S *et al*. The B-cell tumor-associated antigen ROR1 can be targeted with T cells modified to express a ROR1-specific chimeric antigen receptor. *Blood* 116(22), 4532-4541 (2010).
43. Leisegang M, Wilde S, Spranger S *et al*. MHC-restricted fratricide of human lymphocytes expressing survivin-specific transgenic T cell receptors. *J. Clin. Invest*. 120(11), 3869-3877 (2010).
44. Robbins PF, Morgan RA, Feldman SA *et al*. Tumor regression in patients with metastatic synovial cell sarcoma and melanoma using genetically engineered lymphocytes reactive with NY-ESO-1. *J. Clin. Oncol*. 29(7), 917-924 (2011).
45. Wang QJ, Hanada K, Feldman SA, Zhao Y, Inozume T, Yang JC. Development of a genetically-modified novel T-cell receptor for adoptive cell transfer against renal cell carcinoma. *J. Immunol. Methods* 366(1-2), 43-51 (2011).
46. Fang J, Qian J-J, Yi S *et al*. Stable antibody expression at therapeutic levels using the 2A peptide. *Nature Biotech.* 23(5), 584-590 (2005).

*** Use of 2A in the production of a monoclonal antibody in which the C-terminal extension of 2A was removed by furin during transport through the Golgi apparatus.**

1 47. Fang JM, Yi SL, Simmons A *et al*. An antibody delivery system for regulated expression of therapeutic
2 levels of monoclonal antibodies in vivo. *Mol. Therapy* 15(6), 1153-1159 (2007).
3 48. Luo XM, Maarschalk E, O'Connell RM, Wang P, Yang L, Baltimore D. Engineering human
4 hematopoietic stem/progenitor cells to produce a broadly neutralizing anti-HIV antibody after in
5 vitro maturation to human B lymphocytes. *Blood* 113(7), 1422-1431 (2009).
6
7 **** Use of 2A to co-express a therapeutic antibody in hematopoietic progenitor cells and
8 subsequent maturation *in vitro* into a B-cell secreting the antibody.**
9
10 49. Yu KK, Aguilar K, Tsai J *et al*. Use of mutated self-cleaving 2A peptides as a molecular rheostat to
11 direct simultaneous formation of membrane and secreted anti-HIV immunoglobulins. *PLoS One*
12 7(11):e50438 (2012).
13 50. Vatakis DN, Koya RC, Nixon CC., *et al*. Antitumor activity from antigen-specific CD8 T cells generated
14 in vivo from genetically engineered human hematopoietic stem cells. *Proc. Natl. Acad. Sci. U.S.A.*
15 108(51):E1408-16 (2011).
16 51. Okita K, Nakagawa M, Hyenjong H, Ichisaka T, Yamanaka S. Generation of mouse induced pluripotent
17 stem cells without viral vectors. *Science* 322(5903), 949-953 (2008).
18
19 **** Use of 2A to co-express multiple transcription factors in the production of induced pluripotent
20 stem cells.**
21
22 52. Carey BW, Markoulaki S, Hanna J *et al*. Reprogramming of murine and human somatic cells using a
23 single polycistronic vector. *Proc. Natl. Acad. Sci. USA* 106(1), 157-162 (2009).
24 53. Chang CW, Lai YS, Pawlik KM *et al*. Polycistronic lentiviral vector for "hit and run'' reprogramming of
25 adult skin fibroblasts to induced pluripotent stem cells. *Stem Cells* 27(5), 1042-1049 (2009).
26 54. Gonzalez F, Monasterio MB, Tiscornia G *et al*. Generation of mouse-induced pluripotent stem cells by
27 transient expression of a single nonviral polycistronic vector. *Proc. Natl. Acad. Sci. USA* 106(22), 8918-
28 8922 (2009).
29 55. Kuzmenkin A, Liang H, Xu G *et al*. Functional characterization of cardiomyocytes derived from murine
30 induced pluripotent stem cells in vitro. *FASEB J*. 23(12), 4168-4180 (2009).
31 56. O'Malley J, Woltjen K, Kaji K. New strategies to generate induced pluripotent stem cells. *Curr. Opin.*
32 *Biotechnol*. 20(5), 516-521 (2009).
33 57. Woltjen K, Michael IP, Mohseni P *et al*. PiggyBac transposition reprograms fibroblasts to induced
34 pluripotent stem cells. *Nature* 458(7239), 766-770 (2009).
35 58. Yusa K, Rad R, Takeda J, Bradley A. Generation of transgene-free induced pluripotent mouse stem
36 cells by the piggyBac transposon. *Nat Methods*. 6(5), 363-369 (2009).
37 59. Jincho Y, Araki R, Hoki Y *et al*. Generation of genome integration-free induced pluripotent stem cells
38 from fibroblasts of C57BL/6 mice without c-Myc transduction. *J. Biol. Chem.* 285(34), 26384-26389
39 (2010).
40 60. Li W, Wang D, Qin J *et al*. Generation of functional hepatocytes from mouse induced pluripotent
41 stem cells. *J. Cell Physiol*. 222(3), 492-501 (2010).
42 61. Carey BW, Markoulaki S, Hanna J *et al*. Reprogramming of murine and human somatic cells using a
43 single polycistronic vector. *PNAS* 106(1), 157-162 (2009).
44 62. Kaji K, Norrby K, Paca A, Mileikovsky M, Mohseni P, Woltjen K. Virus-free induction of pluripotency
45 and subsequent excision of reprogramming factors. *Nature* 458(7239), 771-775 (2009).

63. Shao L, Feng W, Sun, Y *et al*. Generation of iPS cells using defined factors linked via the self-cleaving 2A sequences in a single open reading frame. *Cell Research* 19(3), 296-306 (2009).
64. Sommer CA, Stadtfeld M, Murphy GJ, Hochedlinger K, Kotton DN, Mostoslavsky G. Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem Cells* 27(3), 543-549.
65. Woltjen K, Michael IP, Mohseni P *et al*. piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* 458(7239), 766-770 (2009).
66. Okita K, Hong H, Takahashi K, Yamanaka S. Generation of mouse-induced pluripotent stem cells with plasmid vectors. *Nat. Protoc.* 5(3), 418-428 (2010).
67. Somers A, Jean JC, Sommer CA *et al*. Generation of transgene-free lung disease-specific human induced pluripotent stem cells using a single excisable lentiviral stem cell cassette. *Stem Cells* 28(10), 1728-1740 (2010).
68. Ma J, Guo L, Fiene SJ *et al*. High purity human-induced pluripotent stem cell-derived cardiomyocytes: electrophysiological properties of action potentials and ionic currents. *Am. J. Physiol. Heart Circ. Physiol.* 301(5), H2006-2017 (2011).
69. Montserrat N, Garreta E, González F *et al*. Simple generation of human induced pluripotent stem cells using poly-B-amino esters as the non-viral gene delivery system. *J. Biol. Chem.* 286(14), 12417-12428 (2011).
70. Papapetrou EP, Sadelain M. Generation of transgene-free human induced pluripotent stem cells with an excisable single polycistronic vector. *Nat. Protoc*. 6, 1251-1273 (2011).
71. Fan Y, Luo Y, Chen X, Li Q, Sun X. Generation of human β-thalassemia induced pluripotent stem cells from amniotic fluid cells using a single excisable lentiviral stem cell cassette. *J Reprod Dev*. 58(4), 404-409 (2012).
72. Ramos-Mejía V, Montes R, Bueno C *et al*. Residual expression of the reprogramming factors prevents differentiation of iPSC generated from human fibroblasts and cord blood CD34+ progenitors. *PLoS One*. 7(4):e35824 (2012).
73. Sommer AG, Rozelle SS, Sullivan S *et al*. Generation of human induced pluripotent stem cells from peripheral blood using the STEMCCA lentiviral vector. *J Vis Exp*. Oct 31;(68). doi:pii: 4327. 10.3791/4327 (2012).
74. Yao Y, Nashun B, Zhou T *et al*. Generation of CD34+ cells from CCR5-disrupted human embryonic and induced pluripotent stem cells. *Hum. Gene Ther.* 23(2), 238-242 (2012).
75. Chen YH, Chen CH. The development of antibody-based immunotherapy for methamphetamine abuse: immunization, and virus-mediated gene transfer approaches. *Curr. Gene Ther*. 13(1), 39-50 (2013).
76. Wanczyk H, Barker T, Rood D *et al*. Cloning and Characterization of a hybridoma secreting a 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK)-specific monoclonal antibody and recombinant F(ab). *Toxins (Basel).* 5(3), 568-589 (2013).
77. Donnelly MLL, Luke GA, Mehrotra A *et al*. 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(5), 1013-1025 (2001).

*** Description of a model of translational, rather than proteolytic, mechanism of 2A-mediated 'cleavage'.**

78. Sodergren E, Weinstock GM, Davidson EH *et al*. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* 314(5801), 941-952 (2006).

Legends to Figures.

Figure 1. The genome structure of FMDV and picornavirus sequences used in co-expression. The overall architecture of the picornavirus genome resembles that of a cellular mRNA. The virus RNA cap structure comprises an oligopeptide (VPg), rather than the ^{7me}G of mRNAs. The FMDV 5’ non-coding region (NCR) comprises an internal ribosome entry sequence (IRES) preceding the single, long, open reading frame (ORF; boxed area). The polyprotein comprises the L proteinase (L^{pro}), the capsid proteins domain, 2A and two domains which together comprise the replication proteins. The sequence of 2A is shown together with the site of ‘cleavage’ (arrow) and the N-terminal proline of protein 2B, immediately downstream of 2A. The short 3’NCR bears a poly-A tail (Panel A). The first bicistronic mRNAs utilised the cap-independent mode of translation conferred by the IRES. The first ORF is translated in the canonical manner for the 5’ ^{7me}G-cap structure, whilst translation of the second ORF is cap-independent driven by the IRES – although only to some 10% of the first ORF (Panel B). Gene sequences 1 (stop codon removed) and 2 are concatenated into a single (trans)gene *via* a 2A linker. The translation products are synthesised in an equimolar ratio, although; (i) protein 1 upstream of 2A bears a C-terminal extension of 2A and (ii) protein 2 bears an N-terminal proline residue (Panel C).

Figure 2. Use of 2A in protein coexpression and dual protein targeting. Sequences encoding GFP and CherryFP were linked *via* TaV2A into a single ORF. HeLa cells transfected with this construct produced the ‘cleavage’ products [GFP-TaV2A] and CherryFP. Image analyses show individual cells expresses each product to the same level, although different cells have different expression levels (A). The wild-type STR6-2A sequence and a point-mutated (‘cleavage’ inactive) form (STR6-2A^{mut}) were fused to the N-terminus of CherryFP. The wild-type sequence is highly active in mediating ‘cleavage’ and, therefore, CherryFP emerges from the ribosome lacking any signal sequence: it is localised throughout the cytoplasm and diffuses into the nucleus. The mutant, ‘cleavage’ inactive, form remains fused to the N-terminus of CherryFP, is recognised by signal recognition particle (SRP) and targets CherryFP to the exocytic pathway: transiting through the ER and the characteristic perinuclear crescent shape of the Golgi apparatus (B).

1

Executive summary

Features of the 2A co-expression system

- Multiple coding sequences are concatenated into a single gene (single open reading frame) *via* 2A linker sequences
- Each component is translated as a discrete product
- No theoretical upper limit as to the number of genes which may be co-expressed
- Co- and post-translational protein targeting (signal) sequences may be incorporated within the polyprotein: proteins may be co-expressed *and* targeted to different sub-cellular sites
- Complex traits / components of multimeric complexes can be co-expressed from a single transgene

Drawbacks of the 2A co-expression system

- 2A remains as a C-terminal extension of the upstream protein
- Protein downstream of 2A bears an N-terminal proline residue (may preclude certain post-translational modifications)

Advantages over IRES co-expression

- 2A co-expression system works in all tissue-types / eukaryotic organisms
- Smaller size (~50-100bp)
- No 'polarity' effect: each translation product synthesised in equimolar quantities

Future perspective

- Substantial expansion of the use of 2A-mediated co-expression in the field of transferable immune responses (TIRs): to combat cancer, inherited genetic disorders and infectious agents
- Widespread use of 2A in the production of patient-specific induced pluripotent stem cells
- Genetic modification of fungi and yeasts for biotechnology
- Genetic modification of domesticated animals to improve health, increase yields, introduce complex traits
- 'Stacking' transgenes in the genetic modification of crop species: use of 2A in the production of 'dual-use' / increased nutritive value / disease resistance / abiotic stress resistance / pesticide tolerance
- Potential novel use as a sensor of cellular stress
- Use of N-terminal signal sequence 2As for dual protein targeting of an expressed protein: localisation within the cytoplasm *and* entry into the exocytic pathway

2

1 **Acknowledgements.** The authors gratefully acknowledge the support of the UK Biotechnology and 2 Biological Sciences Research Council (BBSRC), the Wellcome Trust and the UK Medical Research Council 3 (MRC).

4

5

6

7 **Table 1. Examples of 2A / 2A-like sequences used in biomedicine and biotechnology.**

Virus	Abbreviation	2A / 2A-like sequence	References
Picornaviridae			
Foot-and-mouth disease virus	FMDV	-PVKQLLNFDLLKLAG**D**V**E**S**NPG P**-	9,11,12,15,18,47,51,62.
Equine rhinitis A virus	ERAV	-QCTNYALLKLAG**D**V**E**S**NPG P**-	18,61,62.
Porcine teschovirus-1	PTV1	-ATNFSLLKQAG**D**V**E**E**NPG P**-	28,37,44,61.
Tetraviridae			
Thosea asigna virus	TaV	-EGRGSLLTCG**D**V**E**S**NPG P**-	18,22,37,39,61,62.

8 The –DxExNPG P- motif conserved amongst 2A/2A-like sequences is shown in bold.

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

1

2

3

Table 2. Active 2A sequences in viruses.

Virus	Abbreviation	2A Sequence
Positive-stranded RNA viruses		
Picornaviruses (primarily mammals)		
Theiler's murine encephalomyelitis virus	TMEV	-FREFFKAVRGYHADYYKQRLIH**D**V**E**M**NPG P**-
Encephalomyocarditis virus	EMCV	-VFGLYRIFNAHYAGYFADLLIH**D**I**E**T**NPG P**-
Saffold virus	SAF-V	-FTDFFKAVRDYHASYYKQRLQH**D**V**E**T**NPG P**-
Equine rhinitis B virus	ERBV-1	-EATLSTILSEGATNFSLLKLAG**D**V**E**L**NPG P**-
Ljungan virus	LV	-YFNIMHSDEMDFAGGKFLNQCG**D**V**E**T**NPG P**-
Iflaviruses (insects)		
Infectious flacherie virus	IFV	-PSIGNVARTLTRAEIEDELIRAGI**E**S**NPG P**-
Ectropis oblique picorna-like virus	EoPV-2A_1	-PSIGNVARTLTRAEIEDELIRAGI**E**S**NPG P**-
	EoPV-2A_2	-TRGGLQRQNIIGGGQRDLTQDG**D**I**E**S**NPG P**-
Perina nuda picorna-like virus	PnPV-2A_1	-GQRTTEQIVTAQGWVPDLTVDG**D**V**E**S**NPG P**-
	PnPV-2A_2	-TRGGLRRQNIIGGGQKDLTQDG**D**I**E**S**NPG P**-
Tetraviruses (insects)		
Euprosterna elaeasa virus	EeV	-RRLPESAQLPQGAGRGSLVTCG**D**V**E**E**NPG P**-
Providence virus	PRV-2A_1	-LEMKESNSGYVVGGRGSLLTCG**D**V**E**S**NPG P**-
	PRV-2A_2	-NSDDEEPEYPRGDPIEDLTDDG**D**I**E**K**NPG P**-
	PRV-2A_3	-TIMGNIMTLAGSGGRGSLLTAG**D**V**E**K**NPG P**-
Dicistroviruses (insects)		
Cricket paralysis virus	CrPV	-LVSSNDECRAFLRKRTQLLMSG**D**V**E**S**NPG P**-
Acute bee paralysis virus	ABPV	-TGFLNKLYHCGSWTDILLLLSG**D**V**E**T**NPG P**-
Double-stranded RNA viruses		
Rotaviruses (mammalian)		
Bovine rotavirus C	BoRV-C	-GIGNPLIVANSKFQIDRILISG**D**I**E**L**NPG P**-
Human rotavirus C	HuRV-C	-GAGYPLIVANSKFQIDKILISG**D**I**E**L**NPG P**-
New Adult diarrhoea virus	ADRV-N	-FFDSVWVYHLANSSWVRDLTRECI**E**S**NPG P**-
Cypoviruses (insect)		
Bombyx mori cypovirus 1	BmCPV-1	-RTAFDFQQDVFRSNYDLLKLCG**D**I**E**S**NPG P**-
Operophtera brumata cypovirus-18	OpbuCPV-18	-IHANDYQMAVFKSNYDLLKLCG**D**V**E**S**NPG P**-
Totiviruses (crustaceans)		
Infectious myonecrosis virus	IMNV-2A_1	-WDPTYIEISDCMLPPPDLTSCG**D**V**E**S**NPG P**-
	IMNV-2A_2	-RDVRYIEKPEDKEEHTDILLSG**D**V**E**S**NPG P**-

The –DxExNPG P- motif conserved amongst 2A/2A-like sequences is shown in bold.

1

2 **Table 3. Active 2A cellular sequences.**

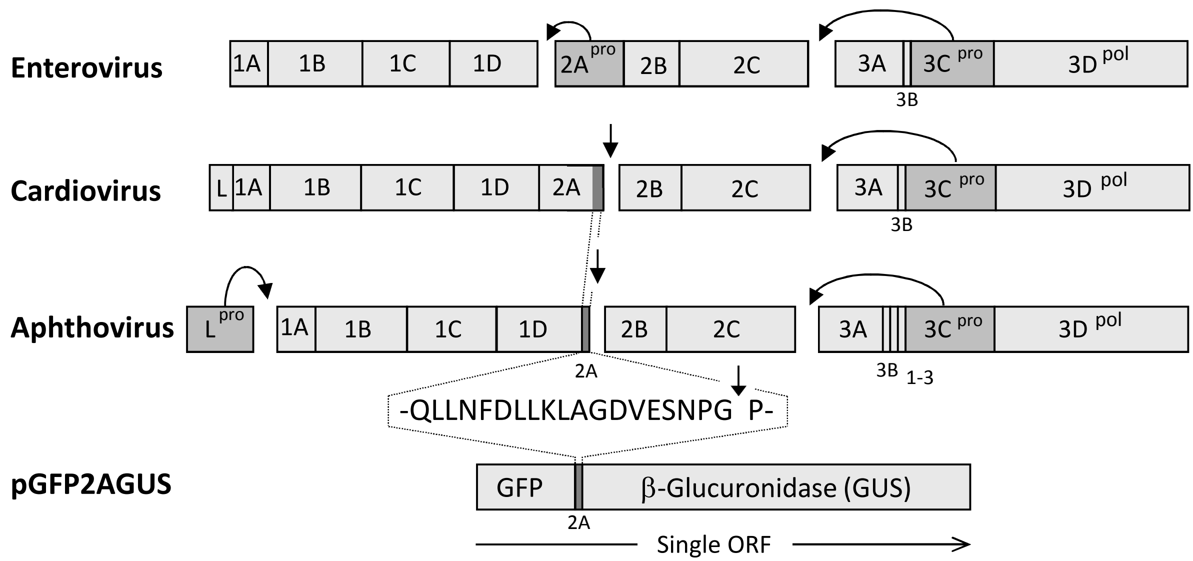
Cellular 2A	Name	2A sequence
Non-LTR retrotransposons [6]		
Trypanosoma *spp*		
T.brucei	*Ingi*	-RSLGTCKRAISSIIRTKMLVSG**D**V**E**E**NPG P**-
T.cruzi	*L1Tc*	-QRYTYRLRAVCDAQRQKLLLSG**D**I**E**Q**NPG P**-
Strongylocentrotus purpuratus (purple sea urchin)		
	STR-32_SP	-NSSCVLNIRSTSHLAILLLLSGQV**E**P**NPG P**-
	STR-51_SP	-SRPILYYSNTTASFQLSTLLSG**D**I**E**P**NPG P**-
	STR-61_SP	-GARIRYYNNSSATFQTILMTCG**D**VDP**NPG P**-
	STR-69_SP	-CRRIAYYSNSDCTFRLELLKSG**D**IQS**NPG P**-
	STR-197_SP	-KHPILYYTNGESSFQIELLSCG**D**INP**NPG P**-
Crassostrea gigas (Pacific oyster)		
	CR1-1_CGi	-SRHIVVYNFYLQFFMFLLLLCG**D**I**E**V**NPG P**-
Lottia gigantean (Owl limpet)		
	CR1-1_LG	-TLLNDTFSSILYYCFILIIRSG**D**I**E**L**NPG P**-
Aplysia californica (California sea slug)		
	ingi-1_AC	-PGFFLGGQHNPAWLARLLILAG**D**V**E**Q**NPG P**-
CATERPILLER proteins (unpublished)		
S.purpuratus		
	STR6-2A^{wt}	MDGFCLLYLLLILLMRSG**D**V**E**T**NPG P**-
	STR6-2A^{mut}	MDGFCLLYLLLILLMRSG**D**V**E**T**N**A**G P**-

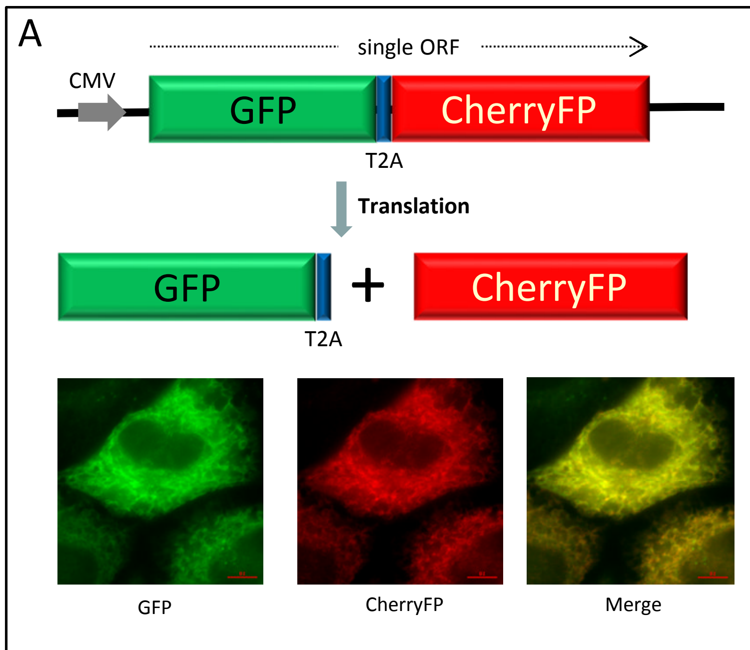
3 STR6-2A^{mut}: site of mutation (Pro → Ala) to create a cleavage inactive mutant - underlined

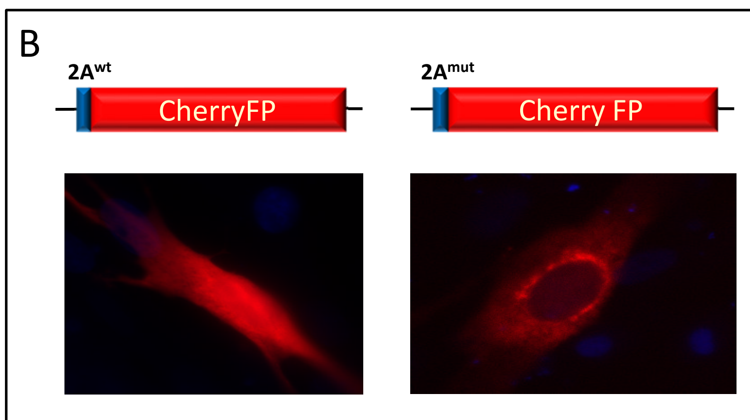
4

5

Figure 1




A
single ORF
CMV
GFP
CherryFP
T2A
Translation
GFP
+
CherryFP
T2A
GFP
CherryFP
Merge


B
2A^{wt}
CherryFP
2A^{mut}
Cherry FP