

Commentary

Open Access

## Skipping the co-expression problem: the new 2A "CHYSEL" technology

Pablo de Felipe\*

Address: Centre for Biomolecular Sciences, School of Biology, Biomolecular Sciences Building, University of St. Andrews, North Haugh, St. Andrews KY16 9ST, Scotland, UK

Email: Pablo de Felipe\* - pdf@st-andrews.ac.uk

\* Corresponding author

Published: 13 September 2004

Received: 30 June 2004

*Genetic Vaccines and Therapy* 2004, **2**:13 doi:10.1186/1479-0556-2-13

Accepted: 13 September 2004

This article is available from: <http://www.gvt-journal.com/content/2/1/13>

© 2004 de Felipe; licensee BioMed Central Ltd.

This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

The rapid progress in the field of genomics is increasing our knowledge of multi-gene diseases. However, any realistic hope of gene therapy treatment for those diseases needs first to address the problem of co-ordinately co-expressing several transgenes. Currently, the use of internal ribosomal entry sites (IRESs) is the strategy chosen by many researchers to ensure co-expression. The large sizes of the IRESs (~0.5 kb), and the difficulties of ensuring a well-balanced co-expression, have prompted several researchers to imitate a co-expression strategy used by many viruses: to express several proteins as a polyprotein. A small peptide of 18 amino acids (2A) from the foot-and-mouth disease virus (FMDV) is being used to avoid the need of proteinases to process the polyprotein. FMDV 2A is introduced as a linker between two proteins to allow autonomous intra-ribosomal self-processing of polyproteins. Recent reports have shown that this sequence is compatible with different sub-cellular targeting signals and can be used to co-express up to four proteins from a single retroviral vector. This short peptide provides a tool to allow the co-expression of multiple proteins from a single vector, a useful technology for those working with heteromultimeric proteins, biochemical pathways or combined/synergistic phenomena.

### Introduction

For the last 20 years, the gene therapy field has centred many of its efforts on finding ways to deliver a therapeutic gene to certain target cells in order to produce a therapeutic result. It was soon clear that it was necessary to deliver at least two genes, because a reporter/marker gene was needed in order to track the expression of the therapeutic gene (normally not easy to detect). There has been a large increase in vector development during these years, with the appearance of many new viral and non-viral vectors. However, since the late 1980s, few improvements have been made 'inside' those vectors. The linkage of the two genes of interest (therapeutic and reporter) has remained the same. The different strategies known for co-expression

were reported during the 1980s -splicing, multiple promoters, fusions, reinitiation and IRESs-, at the same time that the first gene therapy experiments were being performed (for a review [1]). During the 1990s, nearly all those strategies were abandoned in favour of the IRESs. In bicistronic mRNAs bearing an IRES sequence, the first cistron is translated by scanning ribosomes that enter via the 5' end. The cloning of an IRES sequence downstream of the first cistron, allows the internal entry of ribosomes that translate the second cistron. As each cistron is translated from a different translational initiation event, both translations are uncoupled, and the proteins are not obtained in an equimolecular proportion ("imbalance") leading to a large excess of the first protein.

The drive to co-express more than two genes, opening the door to therapies for multi-gene deficiencies, was halted by the inability of vector technology to guarantee a reliable co-expression. Nevertheless, IRESs were the first strategy that met with some success, and several polycistronic vectors able to co-express up to 4 genes were developed during the 1990's [2]. However, two main problems blocked the successful use of large and complex polycistronic vectors: the large size and imbalance of most IRESs which makes it very difficult to predict the level of expression of the downstream cistron [3].

This commentary discusses several recent publications that use self-processing polyproteins as a novel strategy for co-ordinated co-expression of several genes.

### Discussion

Although gene therapy has employed the viruses as vectors, the co-expression strategies previously described have not taken advantage of the dominant ways in which viruses achieve co-expression in cells. It is the polyprotein strategy that many viruses use to co-express most of their proteins, or even all of them (as in picornaviruses). Not surprisingly, this strategy is indeed used by cells, although not very often, in particular for the co-ordinated secretion of different proteins and peptides. Recently, several groups have been trying to utilize this co-expression strategy. One of the possibilities is to introduce the target site for a cellular proteinase between two cistrons cloned in frame forming a single open reading frame (ORF; [4]). In this way the polyprotein is synthesized as a fusion protein that post-translationally is proteolytically cleaved to yield the discrete proteins of interest. Unfortunately, this strategy has several practical difficulties: (1) the polyprotein to be cleaved must reside, or at least pass through, the same compartment as the proteinase, (2) the cleavage is not always independent of the context, (3) the cleavage may be incomplete and unpredictable, (4) efficient cleavage will only be produced in cells actively expressing the proteinase, and (5) the post-translational cleavage is not compatible with all possible sub-cellular targetings. In many ways, a co-translational strategy such as reinitiation, which would be independent of cellular or viral factors, would be desirable. In reinitiation, ribosomes first translate an upstream cistron, although highly inefficiently, then resume translation of the second, downstream, cistron. Reinitiation was indeed tried in the 1980s, but the large imbalance makes it unsuitable for co-expression of even two genes (reviewed in [1]).

#### **The foot-and-mouth disease virus (FMDV) 2A sequence as a co-expression tool**

Picornaviruses, the same family of viruses to first provide the IRESs, encode all their proteins in a long single ORF that is cleaved post-translationally by viral proteinases.

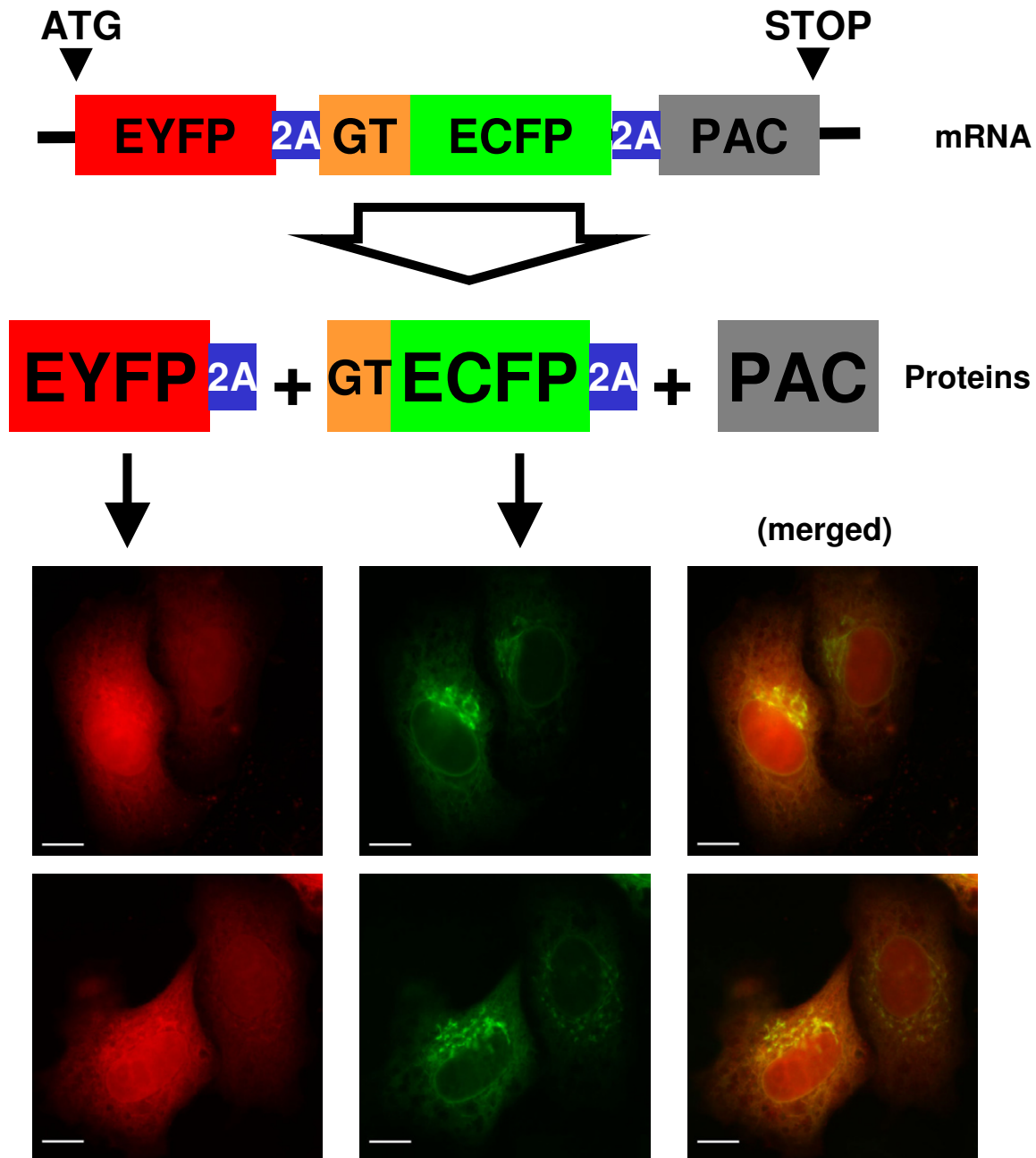
However, it was described in the 1980's that at one position, the polyprotein of some picornaviruses (such as FMDV) underwent a rapid co-translational self-processing. It was soon realised that the key was a small 18aa peptide (2A) that directed its own separation from the growing polyprotein. During the last decade, this mechanism has been studied in detail, resulting in a simple model: the small 2A peptide, during its translation, interacts with the exit tunnel of the ribosome to induce the "skipping" of the last peptide bond at the C-terminus of 2A. The crucial point is that the ribosome is able to continue translating the downstream gene, after releasing the first protein fused in its C-terminus to 2A (reviewed in [5]). This type of sequence has been termed CHYSEL (*cis*-acting *h*ydrolase *e*lement). From a biotechnological standpoint, all that is needed is to clone the coding sequence of 2A, followed by the codon for the first amino acid of the next FMDV protein (2B), in frame between the two genes one wishes to co-express. The synthesis of the peptide bond between the last amino acid (Gly) of 2A and the first (Pro) of 2B is skipped, producing an upstream protein with a C-terminal tail of 18aa (2A) and a downstream protein with a Pro at the N-terminus. The extra sequences have minimal effect on the activity of most proteins and none on their stability. In fact, the 2A peptide has been used as an efficient tag for immunoprecipitation and Western blotting, although commercial antibodies are not yet available. Interestingly, additional CHYSEL sequences have been found in viruses other than FMDV (for a review of these "2A-like" sequences, see [5]).

#### **Broad applicability of 2A**

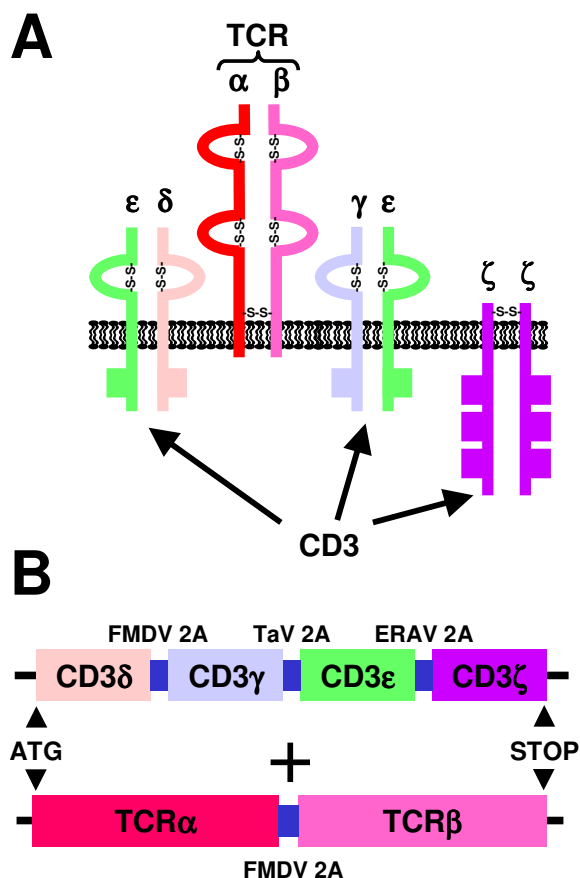
The initial publications using this strategy have shown that 2A skipping can be used in the typical viral vectors used for gene therapy (retrovirus and adeno-associated virus) to reliably co-express many reporter proteins (neomycin phosphotransferase, NEO; puromycin N-acetyl transferase, PAC; green fluorescent protein, GFP, etc) and therapeutic proteins (*Herpes simplex* virus-1 thymidine kinase, HSV1TK; interleukin-12, IL-12; viral antigens, etc.) in transient transduced or stable cells lines and in animals. A full list of publications using 2A is available on the web [6]. Several publications in the past few months have shown the potential of this new co-expression strategy [7-9].

#### **Co-ordinating the co-expression of all your genes**

Up to four genes have been successfully co-expressed from plasmids and retrovirus using several copies of the FMDV 2A or other 2A-like sequences (to avoid direct repeats in retroviruses) [8,9]. Not only was co-expression effective, its co-ordination was also apparent [7,9] (Fig. 1), and the imbalance in the level of the proteins expressed was low (determined to 1.2 [8]). These properties allowed

**Figure 1**

**Co-ordinated co-expression to different compartments in HeLa cells.** A single ORF was designed with the fluorescent genes *eyfp* and *ecfp* plus the puromycin resistant gene *pac* [9]. These genes were cloned flanking FMDV 2A sequences. An internal signal-anchor from the human  $\beta$ -1,4 galactosyltransferase (GT) was fused to the 5' end of the *ecfp* for Golgi targeting. During its translation, the self-processing of this polyprotein produced EYFP-2A that diffused to the cytoplasm and nucleus (due to its small size), while GT-EYFP-2A was co-translationally targeted to the Golgi apparatus by the GT signal (some protein also stays in the endoplasmic reticulum, due to the continuous cycling between these compartments). Two fields are shown, in both cases the cell on the left shows a high level of expression of both proteins that were expressed at lower levels in the cell on the right, illustrating the co-ordination obtained with the 2A co-expression strategy. PAC was able to confer resistance to puromycin. Images were taken 48 hours post-transfection. Bar represents 10  $\mu$ m.



**Figure 2**  
**Self-processing polyproteins to reconstruct the TCR:CD3 complex.** (A) Schematic diagram of the TCR:CD3 complex spanning the cytoplasmic membrane. The T-cell receptor (TCR) is formed by two subunits and the other four proteins assemble in three dimers to form the CD3 complex. The square boxes in the cytoplasmic sequences of the CD3 subunits represent the immunoreceptor tyrosine-based activation motifs (ITAMs). (B) To express the TCR:CD3 complex in cells, two retroviral vectors were designed to carry the two ORFs drawn here [8]. In the retrovirus encoding the four CD3 subunits, three different 2A sequences were used to avoid deletions due to direct repetitions.

polycistronic vectors bearing *pac* in the last position to easily generate stable cell lines co-expressing two upstream genes [7,9].

#### Putting your proteins where they should be

The CHYSEL strategy of co-expression is also compatible with the most disparate sub-cellular localisations [7-10].

Proteins processed by 2A from polyproteins were targeted to the cytosol, nucleus, mitochondria, endoplasmic reticulum, Golgi apparatus, plasma membrane (both, by transmembrane proteins and by cytosolic attachment due to myristoylation) and the extra-cellular compartment. Post-translationally targeted cytosolic proteins as well as co-translationally secreted and transmembrane proteins type I, II and III, have been successfully co-translated. Only one combination of co-translational signals was not correctly targeted [9].

#### Designing complex polyproteins for multi-gene deficiency

The results reported in reference [8] should be particularly interesting for researchers in the gene therapy field. They provide a good example of the potential of the 2A co-expression strategy, introducing up to four genes in a single vector. Furthermore, they show the utility of this strategy to reconstruct a very delicate heteromultimeric protein complex on the cell surface (T-cell receptor:CD3 complex, TCR:CD3; Fig. 2A). It is known that all six subunits are necessary for the efficient formation of the TCR:CD3 complex and just two retroviral vectors were sufficient to reconstruct it in transfected 293T or infected 3T3 cells: one encoding both subunits of the T-cell receptor and the other the four subunits of the CD3 complex (Fig. 2B).

Lethally irradiated  $CD3\epsilon^{AP/AP} \times CD3\zeta^{-/-}$  mice (lacking all four CD3 subunits) were transplanted with bone marrow from wt C57BL/6 mice or  $CD3\epsilon^{AP/AP} \times CD3\zeta^{-/-}$  mice transduced with a retrovirus encoding the four CD3 subunits, and in both cases TCR surface expression was detected and the T cells proliferated normally after immune stimulation. Bone marrow from  $CD3\epsilon^{AP/AP} \times CD3\zeta^{-/-}$  mice without CD3 transduction did not restore T-cell development. T cells were also reconstituted in sub-lethally irradiated  $RAG-1^{-/-}$  mice (lacking mature T and B lymphocytes) in which bone marrow from  $CD3\epsilon^{AP/AP}$  mice (lacking CDE and with a severe inhibition of  $CD3\gamma$  and  $CD3\delta$ ), transduced with a retrovirus encoding these three subunits (via two 2A sequences), was used for a transplant into the  $RAG-1^{-/-}$  mice. The same experiment using three vectors encoding the CD3 subunits separately was unsuccessful.

#### Conclusions

The development of FMDV 2A as a cloning tool is an example of how dangerous pathogenic viruses can be harnessed by biotechnology for human benefit. Their molecular "tricks" (as IRES or CHYSEL sequences) are gradually becoming part of the biotechnologists' toolbox. The development of the polycistronic vectors here discussed is a big step forward, a decade and a half after the launching of the very first gene therapy trial with the aim of introducing in blood cells just a single therapeutic gene, adenosine deaminase (ADA), and the NEO marker [11]. These results represent a considerable advance in the correction

of diseases that involve heteromultimeric proteins, several enzymes involved in a biochemical pathway or various proteins for combined/synergistic effects. 2A is not a magic tool that is going to solve all our problems, but it will help to pave the way for gene therapy.

### Competing interests

None declared.

### Acknowledgements

I would like to thank Drs. M. D. Ryan, M. C. Thomas and M. C. López for critical reading of the manuscript, and Drs. G. Luke and L. E. Hughes for helpful discussions on the topics of this paper. The author is supported by the Biotechnology and Biological Sciences Research Council (BCB).

### References

1. de Felipe P: **Polycistronic viral vectors.** *Curr Gene Ther* 2002, **2**:355-378.
2. Fussenegger M: **The impact of mammalian gene regulation concepts on functional genomic research, metabolic engineering, and advanced gene therapies.** *Biotechnol Prog* 2001, **17**:1-51.
3. 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.
4. Gäken J, Jiang J, Daniel K, van Berkel E, Hughes C, Kuiper M, Darling D, Tavassoli M, Galea-Lauri J, Ford K, Kemeny M, Russell S, Farzaneh F: **Fusagene vectors: a novel strategy for the expression of multiple genes from a single cistron.** *Gene Ther* 2000, **7**:1979-1985.
5. Ryan MD, Luke G, Hughes LE, Cowton VM, ten Dam E, Li X, Donnelly MLL, Mehrotra A, Gani D: **The aphto- and cardiovirus "primary" 2A/2B polyprotein "cleavage".** In: *Molecular Biology of Picornaviruses* Edited by: Semler BL, Wimmer E. Washington, ASM Press; 2002:213-223.
6. **Dr. Martin Ryan's laboratory web page** [<http://www.st-andrews.ac.uk/ryanlab/Index.htm>]
7. Lorens JB, Pearsall DM, Swift SE, Peelle B, Armstrong R, Demo SD, Ferrick DA, Hitoshi Y, Payan DG, Anderson D: **Stable, stoichiometric delivery of diverse protein functions.** *J Biochem Biophys Methods* 2004, **58**:101-110.
8. Szymczak AL, Workman CJ, Wang Y, Vignali KM, Dilioglou S, Vanin EF, Vignali DAA: **Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector.** *Nat Biotechnol* 2004, **22**:589-594.
9. de Felipe P, Ryan MD: **Targeting of Proteins Derived from Self-Processing Polyproteins Containing Multiple Signal Sequences.** *Traffic* 2004, **5**:616-626.
10. El Amrani A, Barakate A, Askari BM, Li X, Roberts AG, Ryan MD, Halpin C: **Coordinate expression and independent subcellular targeting of multiple proteins from a single transgene.** *Plant Physiology* 2004, **135**:16-24.
11. Muul LM, Tuschong LM, Soenen SL, Jagadeesh GJ, RamseY WJ, Long Z, Carter CS, Garabedian EK, Alleyne M, Brown M, Bernstein W, Schurman SH, Fleisher TA, Leitman SF, Dunbar CE, Blaese RM, Candotti F: **Persistence and expression of the adenosine deaminase gene for 12 years and immune reaction to gene transfer components: long-term results of the first clinical gene therapy trial.** *Blood* 2003, **101**:2563-2569.

Publish with **BioMed Central** and every scientist can read your work free of charge

*"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."*

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

