CHLOROPLAST RNA TRANSPORT SYSTEM OPTIMISATION FOR HETEROLOGOUS PROTEIN EXPRESSION

Master´s degree final project

MSc Molecular and Cellular Plant Biotechnology

Valencia, 2017

Joan Márquez Molins

Scientific Director

José Antonio Navarro Bohigues

Academic Tutor

María Purificación Lisón Párraga

Abstract

Chloroplast transgenesis is an emerging approach for the production of heterologous proteins. However, it is limited by the necessity of genomic sequence knowledge and the lengthy procedures required for obtaining a homo-plastid configuration, among others. That constrains could be avoided with an expression system based on specific RNA traffic to the chloroplast. In the present study, an expression cassette based on a viroid-derived sequence as targeting signal, was attempted to optimize by increasing protein expression levels and reducing biosafety concerns. Influence of several 5' untranslated leader sequences in protein expression levels was analysed. Additionally, the viroid-derived sequence was reduced in two shorter versions and a sequence with an analogous secondary structure was also designed. These variants did not present protein accumulation in the chloroplast despite that correspondent RNAs were presumably detected in the organelle. According to our results, the RNA targeting signal encodes a putative chloroplast transport peptide which is translated from a non-canonical start. Finally, 5' untranslated chloroplastic translation enhancers were not suitable for transient expression because of the early expression in *Agrobacterium tumefaciens*.

Keywords

Chloroplast, ELVd, RNA traffic, heterologous protein expression

INDEX

INTRODUCTION

1. INTRODUCTION

1.1 Heterologous protein expression: the importance of molecular farming.

One of the most important fields in biotechnology is the production of desired proteins that can be used in therapy or have an application in industrial processes. Metabolic engineering also uses the expression of heterologous proteins but for expanding metabolic pathways or increasing the amount of certain valuable metabolites (Raab et al., 2005). Many systems for protein expression have been described in prokaryotes as in eukaryotes, e.g., yeast, chinese hamster ovary (CHO) cells, baculovirus and insect cells among others. All of them have advantages and disadvantages which determine their field of useful application.

The use of plants for this purpose has been exploited long time ago in what is denominated molecular farming, and more exactly molecular pharming, in the case of therapeutic proteins. Many diverse and valuable recombinant proteins have been produced in plants, including subunit vaccines, antibodies and antibody fragments, hormones, blood products, cytokines, and enzymes (Khan et al., 2012). The advantages of employing plants as bioreactors for recombinant protein generation are numerous, including low cost of production, easier scale-up, cost-effective storage, and absence of animal pathogens and bacterial toxins in protein preparations. To this list, it can be added that their use also minimizes ethical concerns (Tak et al., 2016). However, several constraints hinder the widespread use of plants. These restrictions can explain why commercialization of plant-produced recombinant proteins has progressed slowly ever since plants were demonstrated as feasible protein factories two decades ago (Park and Wi, 2016). Obstacles to commercial-scale production include difficulties in obtaining high quality, quantity and homogeneity of the final product, the challenge of processing plant-derived pharmaceutical macromolecules under good manufacturing practice conditions, concerns about biosafety by regulatory agencies and the high cost associated with their regulatory approval.

Molecular farming disposes of several strategies for heterologous protein expression in plants (Xu et al., 2012). These strategies can be classified in two main groups, according to whether they involve transient or stable plant transformation.

1.1.1 Stable transformation

Stable transformation approaches implicate the production of true-breeding lines of genetically transformed plants. Plant transformation involves the integration of heterologous genes under the control of specific promoters into the host plant genome, and subsequent recovery of fertile

plants from the transformed cell (Jones and Sparks, 2009). Integration of heterologous genes can occur by physical methods or be mediated by *Agrobacterium tumefaciens.* Stable transformation is much more time-consuming than transient expression, but it has a higher reproducibility and a potential large-scale production, as the integrated sequence is heritable (Peyret and Lomonossoff, 2015).

Transgenic plants were first developed by exogenous DNA incorporation mediated by *A. tumefaciens* (Chilton et al., 1977). The genetic components carried by *A. tumefaciens* required for plant tumour inducing are: the Ti (tumour inducing) plasmid that harbours the transfer DNA (T-DNA) and the virulence *vir* region; and three chromosomal virulence loci (*chvA chvB* and *pscA*) (Ziemienowicz, 2014). The T-DNA is transferred to the plant genome and consist of opine and phytohormone genes flanked by a 25-bp border sequence in a directly repeated orientation. Binary vectors derived from the Ti plasmid carry the *vir* region and the T-DNA that, instead of bacterial genes, contains the desired transgene as well as a plant selectable marker. In most cases, markers are genes that confer resistance against antibiotics or herbicides. This method is especially effective among dicotyledonous plants, and can be classified according to whether the transformation target is a tissue culture or a zygotic/gametic cell. The last strategy is available only for some model species and it has the advantage to avoid *in vitro* cell culture as transgenic plants can be identified at seed germination.

There is a variety of non-biological transformation methods such as electroporation, polyethylene glycol (PEG) or calcium treatment, silicon carbide whiskers, microinjection, laser-mediated transfer or particle bombardment, which rely entirely on physical principles for introducing naked nucleic acids into the cell (Darbani *et al.*, 2008). Particle bombardment or biolistic uses heavy metal particles coated with DNA, shot at high speed (Sanford et al., 1993). The transformation efficiency might be lower than with *A. tumefaciens* and integration patterns are more complex, with multi copy insertions that could cause gene silencing (Darbani *et al.*, 2008). Despite those disadvantages, biolistic is still a valuable transformation method because of its effectivity in transforming both nuclear and organellar genomes.

1.1.2 Transient expression

Transient expression is based on the use of modified pathogens to infect plant tissues in a short period of time (Sack et al., 2015). Pathogen infection enables the expression of foreign sequences and consequently, recombinant protein production in what is essentially a batch process (Pogue et al., 2002). For this purpose, *A. tumefaciens*, plant viruses and hybrid vectors with components of both systems, named “deconstructed” vectors, have been commonly used (Peyret and Lomonossoff, 2015). When fresh leaves are used, transient expression systems can yield in the

range of grams of recombinant protein within a few weeks. Advantages of transient expression approaches are not only the lower upfront investment cost, but also the fact that they are particularly convenient for products that must be manufactured rapidly in response to an emergency (Vamvaka et al., 2014).

A. tumefaciens transient plant transformation is a widely used technique that relies on T-DNA transfer without further selection (Krenek et al., 2015). Indirect evidence indicates that transient expression predominantly occurs from T-DNA copies that are not integrated into the host genome (Lacroix and Citovsky, 2013). Expression usually peaks 2-4 days post infection and declines thereon (Lacroix and Citovsky, 2013). The most common methods of *A. tumefaciens* entry to the plant are the vacuum infiltration, which involves the use of negative pressure to flood the intercellular space of a plant leaf with the bacteria and the syringe infiltration in which a needle-less syringe is used. *A. tumefaciens* containment is the major problem of this approach.

Viral based systems are engineered by the insertion of a foreign sequence into a viral genome, either by fusion, replacement or addition (Mortimer et al., 2015). *Potato virus* X (Ruiz *et al.*, 1998) *Tobacco mosaic virus* (Avesani et al., 2007) and geminiviruses (Turnage et al., 2002) are some of the plant viruses most commonly used as viral expression vectors. High expression of foreign sequences is provided because viruses multiply within infected cells, and therefore the foreign sequence is amplified. However, mechanical inoculation is generally plant-species and tissue-limited and not every cell in the inoculated leaf is infected. Even though this problem can be circumvented by delivering full-length virus vectors via agroinfection, there is still an important issue to be solved. Very often viral encapsidation and stability require genomes no longer than a given length, a phenomenon that seriously constrains the size of the protein that can be expressed.

A new generation of expression vectors combines the use of partially deconstructed RNA viral genomes to maximize protein accumulation and agrobacterium-mediated gene transfer as a means of spreading DNA copies of the above viral RNAs (Peyret and Lomonossoff, 2015). This strategy involves the use of one or more “deconstructed” replicons, where viral genes not strictly necessary to the production of the recombinant protein have been removed (e.g. MP and CP), enabling larger inserts. On occasion, the host plant can be genetically modified to provide missing necessary functions of the “deconstructed” viruses. Inducible promoters strictly regulated (Aoyama and Chua, 1997; Caddick et al., 1998) have been used to control the initiation and maintenance of vector replication (Mortimer et al., 2015). Further development of this strategy requires the use of transgenic plants in inducible virus platforms such as magnICON®, developed by Icon Genetics, (Werner et al., 2011) and In Plant Activation (INPACT) (Dugdale et al., 2013).

Those strategies combine advantages of both, transient and stable transformation, resulting in a tightly regulated transgene expression and amplification.

1.1.3 Subcellular protein targeting

Regardless of the expression system used, some important issues need to be addressed concerning the subcellular compartment where the protein is located. Cytoplasm is not usually a convenient environment for protein accumulation, due to the presence of reduction-oxidation reactions which difficult protein folding, high protease levels, proteasome degradation and the lack of post-translational modifications (Habibi et al., 2017). Therefore, directing recombinant proteins to subcellular organelles serves to avoid the liability of cytoplasm targeting. Proteins can be targeted to the endoplasmic reticulum (ER) (Aebi, 2013), subcellular bodies (Alvarez et al., 2010), apoplast (Ramírez et al., 2000) and plastids. In addition, they can even be secreted from plant roots directly to a hydroponic medium (Horvath et al., 2000; Doran, 2006). There are several types of plastids, but the chloroplast has emerged as the most promising for biotechnology applications (Michelet et al., 2011).

1.2 Chloroplasts as bio-factories

Chloroplast is an organelle with prokaryotic evolutionary origin, responsible for many metabolic processes, including photosynthesis and biosynthesis of diverse essential primary and secondary metabolites (Nakayama and Archibald, 2012). Chloroplast transgenesis, also known as transplantomics, is an emerging system for protein production (Bock, 2014). This technology provides a prokaryotic-like environment that enables the expression of recombinant proteins at high levels, usually higher than in the cytosol. This is most probably because of the elevated number of chloroplastic genomes per cell and the absence of gene-silencing phenomena in these organelles. It has been reported the production of foreign proteins in transgenic chloroplasts to levels exceeding 10 % of the total soluble protein of the plant (Daniell et al., 2009; Bock, 2007). However, in other cases, expression levels have been unexpectedly low, and it is not clear why a given transgene works or not in the chloroplasts. Moreover, transplantomics is an ecologically safe technology as the risk of transgene dispersion through the pollen is dramatically reduced by mother inheritance predominance of subcellular organelles.

Some metabolic or genetic disorders, including Alzheimer's, diabetes, hypertension, haemophilia, and retinopathy have been treated with plastid-made biopharmaceuticals (Daniell *et al.*, 2016). Additionally, nearly 50 subunit vaccine candidates have been expressed in chloroplast (Chan and Daniell, 2015; Lössl and Waheed, 2011; Waheed et al., 2015). Industrial-scale production of

therapeutic transplastomic plants has even been developed for producing coagulation factor IX (Su et al., 2015).

Unfortunately, chloroplast transformation is based on homologous recombination and that cause two major restrictions i) knowledge of the genomic sequence is required and ii) specific genome constructions are needed. Moreover, there are challenging technical factors such as a limited number of selection markers and the lengthy procedures required to recover fully segregated plants (Ahmad et al., 2016). These restrictions significantly reduce the number of species susceptible to the application of this technology and highlight the necessity of developing novel universal strategies that simplify the expression of foreign genes in the chloroplast (Day and Goldschmidt-Clermont, 2011). The discovery of a novel signalling mechanism of RNA traffic to the chloroplast has opened a new possibility (Gómez and Pallás, 2010b). Developing a system that allows protein translation in the chloroplast based on RNA traffic without the need of obtaining transgenic plastids is a very interesting matter of research.

1.3 RNA traffic to the chloroplast: can a pathogenic RNA be converted into a biotechnological tool?

RNAs are targeted to specific organelles mainly by active transport (Kloc et al., 2002). RNA cis elements or zip codes recognised by specific RNA-binding activities and other structural or sequence motifs must be responsible for transport specificity. According to current knowledge, accessory proteins such as adaptors for cytoskeletal movement and, in some instances, components for translation machinery may have a significant role on RNA transport (Michaud et al., 2010). The detailed mechanism of RNA transport is far from being known, as there are many dark-aspects of the RNA-sorting to specific cell-destinations to be elucidated.

1.3.1 RNA traffic to chloroplasts and viroids.

The vast majority of the chloroplast proteins are encoded in the nucleus, synthesised as cytosolic preproteins, and imported into the chloroplast thanks to N-terminal, cleavable transit peptides (cTPs). To cross the outer and inner envelope membranes (OEM, IEM) of the chloroplasts, these cTPs are recognized by the chloroplast transport machinery composed of two proteins complexes: TOC and TIC (Translocon of the Outer and Inner membranes of the Chloroplasts, respectively). In contrast, only a few proteins are encoded in the plastid genome, and therefore translated in this organelle (e.g. the large subunit of RuBisCO). However, a third group of proteins lacks a canonical cTP. It is not clear whether they are transported as proteins by an alternative pathway or their corresponding mRNAs are trafficked to the chloroplast and then translated by chloroplast

ribosomes. RNA import to that organelle may have a biological significance, but more experiments are required for knowing how important it is. In this sense, viroids have recently emerged as valuable tools in not only studying RNA long distance transport but also RNA inter-organelle trafficking.

Viroids are small single-stranded RNA (246-401 nucleotides) plant pathogens that have been considered the smallest autonomous infectious nucleic acids known so far (Flores et al., 2005; Flores et al., 2015). Their replication and movement rely completely on host factors as their small RNA genome apparently does not code for any protein. Viroidal RNAs have a circular structure and high degree of self-complementary that promotes their compact folding which presumably protects them from degradation by RNA silencing mechanisms (Elena et al., 2009). More than 30 species of viroids have been discovered since Diener *et al* identified *Potato spindle tuber Viroid* (PSTV) in 1971 (Diener, 1971; Diener, 2003) and their phenotypic effects range from severe symptoms to latent infections (Flores et al., 2005). Current taxonomy is formed by two families: *Pospiviroidae* and *Avsunviroidae* (Di Serio et al., 2014). Members of the family *Pospiviroidae* replicate in the nucleus through an asymmetric rolling-circle mechanism and fold into a rod-like structure that contains a central conserved region (CCR). In contrast, the family *Avsunviroidae* is formed by viroids lacking a CCR and that replicate in the chloroplast through a symmetric pathway of the rolling-circle mechanism. Members of this family are additionally characterised by having autocatalytic activity due to the presence of special RNA structures known as hammerhead ribozymes, which are responsible of RNA cleavage. Viroids in the family *Avsunviroidae* show an extremely narrow host range restricted to the species in which they were discovered or to some related species. The study of viroid's pathogenic cycle, particularly how these non-coding RNAs move to different subcellular locations, can help understanding RNA specific movement in plants.

1.3.2 Non coding RNA (ncRNA) signalling to the chloroplast

Viroids belonging to the family *Avsunviroidae* are the only functional RNAs known to traffic selectively into chloroplasts so far, but, as is often the case, it is possible that a common cell process has been taken over by the pathogen. In fact, a novel signalling mechanism of RNA traffic to the chloroplast was proposed by Nicolaï *et al* (Nicolaï et al., 2007). The authors showed that the mRNA coding for the eukaryotic translation factor 4E (eIF4E), an essential regulator of translation, entered the chloroplast and was located in the stroma. Furthermore, eIF4E RNA triggered the entry of a fused heterologous GFP mRNA into the chloroplast compartment. Neither eIF4E nor GFP translation was observed inside the chloroplasts and therefore, the authors hypothesized that chloroplast sequestration of the eIF4E mRNA was a way to regulate cytosolic translation activity.

More recently Gómez and Pallás also showed that a ncRNA derived from *Eggplant latent viroid* (ELVd), a viroid of the family *Avsunviroidae*, acting as a 5' untranslated region (UTR) mediates the functional import of the green fluorescent protein (GFP) messenger RNA (mRNA) into the chloroplast (Gómez and Pallás, 2010b). Transient expression in *N. benthaminana* leaves infiltrated with *A. tumefaciens* carrying this construction, and subsequent RT-PCR analysis of the RNA isolated from chloroplasts revealed that the ELVd-derived RNA also delivered heterologous RNAs to the chloroplast. In contrast to that observed with the eIF4E mRNA, laser scanning confocal microscopy (LSCM) analysis revealed that unmodified GFP exhibits nucleus-cytoplasmic distribution, but when the viroidal sequence was fused, GFP fluorescence was observed in the chloroplast stroma (Figure1). Therefore, ELVd-derived ncRNA could also act as an untranslated signal capable of mediating chloroplast transport and stable expression of foreign proteins in the chloroplast.

These findings provide a conceptual framework for the development of a new and improved protein expression system. The use of this ELVd-derived sequence could be used as biotechnological tool for mRNA delivery to the chloroplast, and subsequent protein translation in that organelle, avoiding the necessity of transforming cell chloroplasts. Obtaining a homo-plastid transformation is a hard procedure, and it is limited to a restricted number of plant species. These disadvantages highlight the importance of a transient expression system suitable for testing chloroplast translation of candidate proteins.

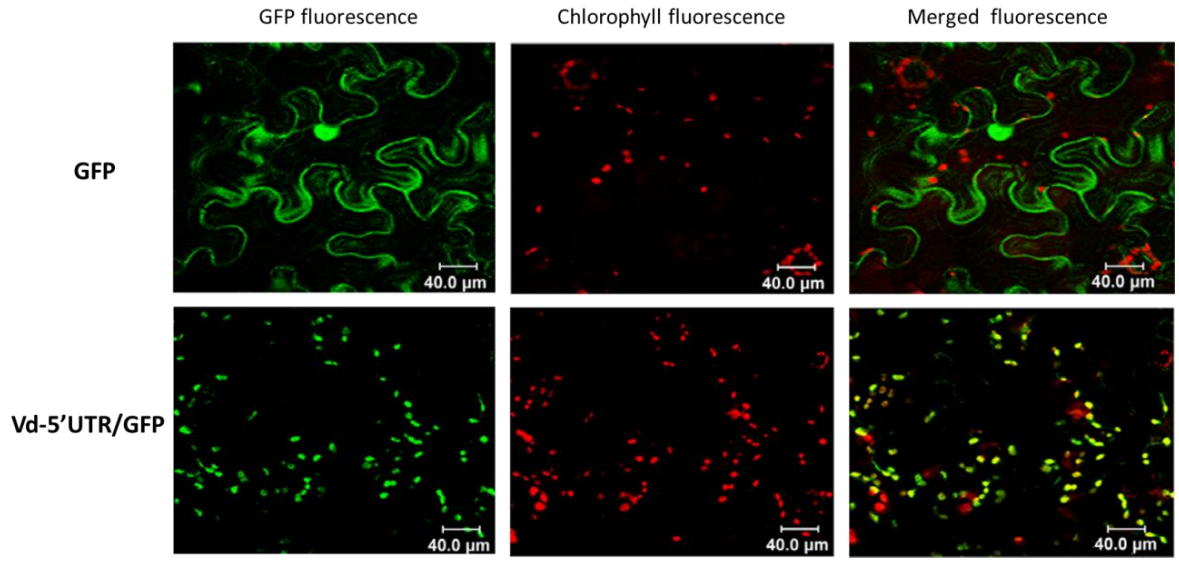


Figure 1. The GFP arising from vd-5'UTR/GFP transcripts accumulate specifically in chloroplasts. Confocal microscope observation of the *N. benthamiana* leaves expressing unmodified GFP (upper panels) and vd-5'UTR-GFP (inferior panels). Figure modified from Gómez and Pallás, 2010b.

1.3.3 Eggplant latent viroid and the potential structural motifs involved in chloroplast targeting.

Eggplant latent viroid (ELVd) has been the last member of the family *Avsunviroidae* to be formally characterised, and has been classified in its own genus, *Elaviroid.* ELVd has been described as a friendly experimental system because of its particular properties (Daròs, 2016). ELVd consists of a single-stranded, circular, non-coding RNA of 332-335 nucleotides that is folded in a branched quasi-rod-like minimum free-energy conformation (Daròs, 2016). Its secondary structure has been elucidated, resulting in similar bifurcated conformation for both polarity strands (Figure 2). These structures have been supported by *in silico*, *in vivo* and *in vitro* data (Giguère et al., 2014). ELVd has only been shown to infect eggplant (*Solanum melongena* L.) being apparently symptomless. This viroid is replicated in the chloroplast presumably by a nuclear encode polymerase (NEP) and, in a recent study, its transcription initiation sites have been mapped in its *in vivo* conformation (López-Carrasco et al., 2016). As the other members of its family, this viroid is characterised by an extremely high mutation rate and by having hammerhead ribozymes in both polarities.

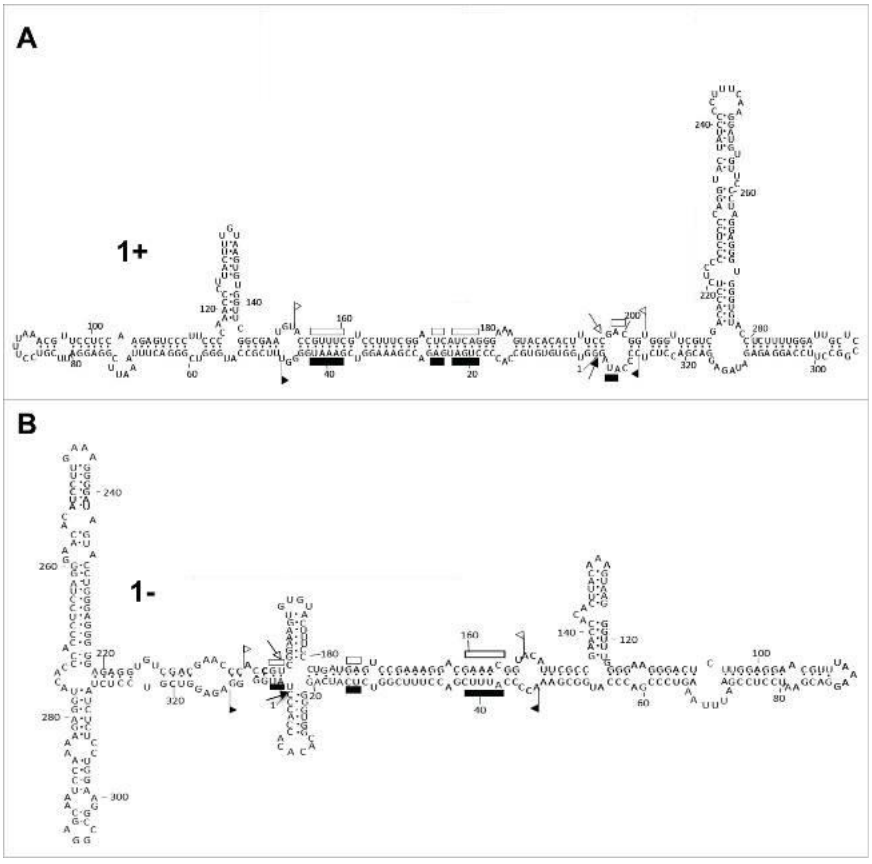


Figure 2. Conformations of minimum free energy predicted for both strands of ELVd. A) Structure generated for ELVd (+) RNA by the *Mfold* and *RNA structure* software for RNA folding. B) Structure generated for ELVd (−) RNA by the 3 software. Sequences forming the hammerhead structures are delimited by flags, motifs conserved in natural hammerhead structures are denoted by bars, and self-cleavage sites are marked by arrows. Solid and open symbols refer to plus and minus polarities, respectively. The same numbers are used for both polarities. Figure modified from López-Carrasco *et al.*, 2016.

The ELVd-derived sequence that enable the GFP mRNA traffic into chloroplasts consisted of a fragment of the minus strand (arbitrarily defined as the less abundant in infected tissues) of ELVd-AJ536613 (position 54–267, note that the numbering of the minus strand goes backwards), followed by a fragment of the plus strand of ELVd-AJ536613 (position 54–261, including two mutations). This ELVd-derived sequence did not contain either an AUG *in frame* with that of GFP, or sense viroid ribozymes that can mediate mRNA cleavage.

To have a more detailed picture and to delimit the sequences and/or structures involved in the process, the ELVd-derived RNA was dissected into three arbitrary regions (Gómez and Pallás, 2010a). Results showed that the 110-nucleotide-long central fragment was sufficient for directing RNA to the chloroplast, resulting in dual expression of GFP in that organelle and also in the nucleus-cytoplasm. Additionally, although the right and left fragments of the ELVd-derived sequence were unable to mediate mRNA import to the chloroplast, their fusion to the central fragment increased the efficiency of chloroplast mRNA traffic, decreasing GFP fluorescence in the nucleus-cytoplasm.

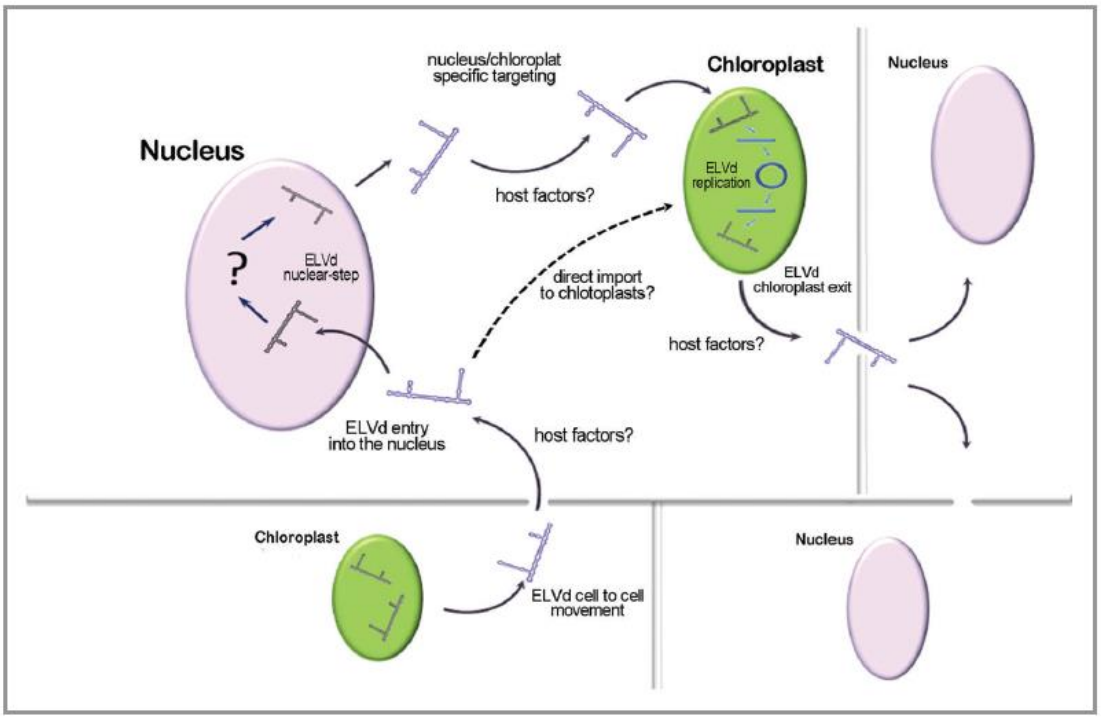


Figure 3. Proposed model to explain the specific compartmentalization of members of the family *Avsunviroidae* into chloroplasts. In this transport pathway (likely mediated by a RNA domain localized in the left terminal region), after the viroid invades the cell cytoplasm, it is imported into the nucleus, by means of an unknown host-dependent mechanism. The viroid then uses this organelle as a sub-cellular shuttle for delivery into the chloroplast, where replication takes place. The possibility that the Avsunviroidae can directly traffic from the cytoplasm to the chloroplast cannot be excluded. Figure reproduced from Gómez and Pallás, 2012a.

ELVd intracellular movement is more complex than previously thought, as it also includes a previous step of nuclear import. According to that, it has been proposed a model (Figure 3) in which ELVd would move first from the cytoplasm to the nucleus, and therefore it would be delivered to the chloroplast (Gómez and Pallás, 2012b). This model explains viroidal RNA accumulation in the chloroplast and also the detection of a small fraction in the nucleus (Gómez and Pallás, 2012a).

OBJECTIVES

2. OBJECTIVES

The main objective of this work is to optimise the ELVd-derived chloroplast transport and expression system to increase its productivity, and therefore, developing a biotechnological tool, ecologically safe, which could be used for testing chloroplastic translation of heterologous proteins without having to transform chloroplasts. Particular objectives can be summarised as:

1. Identification and characterisation of the minimum structure of the viroidal ncRNA involved in a chloroplastic efficient transport.
2. Enhancing heterologous protein expression in the chloroplast by modifying 5' untranslated regions.
3. Design of an ncRNA sequence not-related to ELVd but with analogous secondary structure that retains chloroplast transport functionality.

MATERIAL AND METHODS

3. MATERIAL AND METHODS

3.1 Plant material and agro-infiltration.

Nicotiana benthamina leaves of 3-4 weeks old plants were agroinfiltrated with a culture of *A. tumefaciens* strain C58C1, previously transformed with the desired construct. The overnight grown bacterial culture was diluted in infiltration buffer (MES 0.1 M, $MgCl_2$ 1 M) up to an optical density at 600 nm (OD600) of 0.2 and injected on the abaxial side of the leaves using a 1 ml needle-less syringe. Agroinfiltrated plants were grown in a phytotron and analysed two days post-agroinfiltration. The phytotron growing conditions were as follow: photoperiod of 16 hours under visible light (wavelength between 400-700 nm) with an irradiance of 65-85 $\mu mol \cdot m^{-2}s^{1}$ and 8 hours of darkness; temperature cycles of 25 °C (light) and 18 °C (darkness) and relative humidity of 60-65 % (light) and 95-100 % (darkness).

3.2 Basic cloning methods

Agarose gel electrophoresis

DNA fragments were analysed using agarose gels of percentages between 1-2 % depending on their size. Different percentages of agarose were prepared by melting solid agarose in TAE buffer 1X (40 mM Tris, 20 mM sodium acetate, 1 mM EDTA, pH 8). For DNA staining, the melted agarose was mixed with ethidium bromide (EtBr) to a final concentration of 0.1 μg/ml before being poured into a gel casting tray. Samples were mixed with loading buffer (10 mM Tris-HCl, pH 7.6, 60 % glycerol, 60 mM EDTA, pH 8, 0.03 % bromophenol blue and 0.03 % xylene cyanol). Electrophoresis was run at 100 V for about 20-40 minutes depending on the gel. Gels were visualized using a transilluminator Vilber Lourmat EXC-20-M. DNA fragment size was estimated by comparing with a 1 kb Plus DNA ladder (Invitrogen). Images were acquired with G:BOX EF gel documentation system (Syngene).

DNA fragments purification from agarose gel

DNA fragments were extracted from agarose gel with the *GeneJET Gel Extraction Kit* (Thermo Fisher Scientific) according to the protocol supplied by the manufacturer.

Isolation of plasmid DNA (Minipreparation of plasmid DNA)

Plasmids were isolated by using the NucleoSpin® Plasmid kit of Macherey-nagel according to the protocol supplied by the manufacturer.

Plasmid detection by colony polymerase chain reaction (PCR)

Colony polymerase chain reaction (PCR) allows identifying bacterial colonies that have a certain plasmid by performing a PCR with specific primers and using an isolated colony as source of DNA template. Individual colonies obtained from the transformation of a plasmid were picked up using an automatic pipette and a fine tip. Colonies were pipetted up and down into each reaction tube containing the following reaction mix:

GoTaq Flexi buffer (Promega)	1X
$MgCl_2$	2 mM
dNTPs	0.2 mM
Forward and reverse primers	0.2 µM
GoTaq DNA polymerase (Promega)	0.025 U/µl
Nuclease free water	Up to 5 µl

Reaction tubes were transferred to a thermocycler and the PCR was set up as follows: an initial denaturation step of 5 min at 95 °C, 30 cycles consisting of 30 seconds at 94°C, 30 seconds at 55°C and 1 minute per kb at 68 °C and a final step of 5 min at 68 °C. During the PCR, the tips were dipped into 5 ml growth media with appropriate antibiotics. After agarose gel electrophoresis, positives were cultured overnight. Alternatively, the tips were streaked onto another agar plate containing the appropriate antibiotics and grown overnight.

3.3 Cloning strategy

In order to clone the different modifications of the ELVd-derived sequence into a binary plasmid suitable for transient expression, a three-step cloning strategy was followed. The first step was the obtaining of the desired ELVd-derived variants. That was achieved either by PCR, using specific primers (S1), or directly by ordering their synthesis to IDT as DNA fragments (gBlock® Gene Fragments). PCR was performed with proofreading PrimeSTAR™ HS DNA Polymerase (Takara). The general PCR reaction mixture was 5X PCR Buffer, 2.5 mM dNTP mixture, 0.2 µM (final conc.) each primer, 500 ng template, 2.5U PrimeSTAR™ and sterilized water up to 50 µl. PCR conditions were 30 cycles of 10 sec at 98°C, 5 sec at 55°C and 1 min/kb at 72°C. Each DNA fragment was ligated for 1 h at room temperature (RT) to the commercial plasmid pJET1.2 (Thermo Fisher Scientific) using T4 DNA ligase (Promega). Ligation mixture was transformed into *Escherichia coli* DH5α electro-competent cells. Electroporation conditions were 1650 mV for 3-6 ms. After the pulse, competent cells were resuspended in 500 µl of LB medium without any antibiotic, and let shaking for 45-60 minutes at 37 °C to facilitate cell recovery and antibiotic resistance gene expression. After that, the

bacterial culture was plated on a petri dish containing solid LB medium with the corresponding antibiotic (in this case ampicillin) and let grown overnight at 37 °C.

Individual colonies were analysed by colony PCR to select some that have the plasmid with the desired insertion. Positive bacteria were grown as mentioned before to extract plasmid DNA. In order to verify that the sequence of the inserted fragment was correct, recombinant pJET1.2 plasmids were sequenced by the IBMCP DNA sequencing facility.

In the second step, recombinant pJET1.2 was digested with BamHI and KpnI restriction enzymes to release the ELVd-derived fragment. Digestion mixture was loaded in an agarose gel and run as mentioned before. The band corresponding to the expected size fragment was shown by the electrophoresis size separation and excised to be purified. Afterwards, the fragment was ligated into a plasmid that contains an YFP expression cassette (pIVEX 2.3), upstream the yellow fluorescent protein (YFP) open reading frame (ORF). Thus, the ELVd-derived and the YFP sequences were arranged in an expression cassette, flanked by two HindIII restriction sites, and under the control of the 35S duplicated promoter (35Sx2) of the *Cauliflower mosaic virus*, the 5' leader sequence of the *Alfalfa mosaic virus* (AMV) RNA 4 and the potato protease inhibitor II terminator (PoPit). pIVEX 2.3 with the expression cassette was previously digested with BamHI and KpnI and loaded in an agarose electrophoresis, the band corresponding to the linear plasmid was excised from the gel and purified. Ligation product was transformed to *E. coli* cells as described before, and cells were grown in LB medium with kanamycin. Recombinant pIVEX 2.3 was obtained and verified in an analogous manner as in the previous step.

Finally, in a third step, the recombinant pIVEX 2.3 was digested with HindIII to release the expression cassette that, after agarose gel purification, was ligated to a binary plasmid named pMOG800 (Knoester et al 1988). Ligation product was transformed into *E. coli* cells and plasmid DNA with the correct insertion was obtained and confirmed as previously described. 25 ng of this plasmid were used to transform *A. tumefaciens* C58C1 electro-competent cells (resistant to rifampicin) in the same conditions as the *E. coli* electroporation was done. Those cells were grown in solid LB medium with kanamycin and rifampicin for 2 days at 28 °C. Some individual colonies were selected to be grown in LB medium for transient expression or to be stored as a glycerinate.

The introduction of the 5'-leader sequence of *Tobacco mosaic virus* (TMV-Ω) and the synthetic 5' untranslated region (UTR) synJ, in addition to the elimination of 5' leader sequence of the *Alfalfa mosaic virus* (AMV) RNA 4, were obtained by a slightly different process because the first pJET.2 cloning step was avoided. This is because the desired clones with the YFP expression cassette were obtained directly by inverse PCR using the corresponding recombinant pIVEX 2.3 plasmid as

template (S2). Inverse PCR primers are designed to extend away from each other rather than toward each other as in regular PCR. Reactions were performed using PrimeSTAR™ HS DNA Polymerase as described before. After that, DpnI digestion was performed for 2 hours in order to eliminate the PCR template vector (DpnI only cleaves Dam methylated DNA). PCR amplified DNA was phosphorylated with T4 DNA polynucleotide kinase (20 U) (Promega) in the presence of ATP (5 µM) for 30 minutes at 37 °C. Self-ligation of the resulting PCR product was performed overnight at 16 °C.

3.4 RNA secondary structure predictions and synthetic sequence design

All RNA structures in this work were generated by *RNAfold*, software based on minimal free energy predictions, using default parameters. The designing process of a synthetic RNA sequence with the same secondary structure as the ELVd-derived sequence consisted in manually changing the complementary ribonucleotides of the stem structures, but maintaining all ribonucleotides in the loops. After every pair of ribonucleotides was changed, the predicted structure was obtained with the program in order to verify that it was identical to the original. Therefore, the technique consisted in an iterative process of changing nucleotides and corroborating that any change does not alter the global secondary structure. Resulting synthetic sequence was denominated CTVL (Chloroplast Transporter Viroid Like).

3.5 Confocal microscopy and fluorescence intensity quantification

Subcellular localisation of the fluorescence and imaging was conducted with an inverted Zeiss LSM 780 confocal microscope and ZEN software (Carl Zeiss). Images were acquired using an objective plan-apochromat 40x/1.4 Oil DIC M27. 0.5 cm diameter leaf dishes were cut and mounted in water. The chlorophyll and the YFP were excited with the 488 nm and 514 nm lines of an argon ion multilaser, respectively. The YFP emitted fluorescence was collected between 530 and 570 nm and chlorophyll autofluorescence was collected from 700 nm and beyond. We ensured that there was no cross-talk between the YFP and chlorophyll signals by using sequential instead of simultaneous scanning settings.

Fluorescence intensity was quantified using Fiji software. Chloroplasts in the same plane were automatically selected by creating a duplicated of the original image and considering the same threshold to highlight the structures with a minimum area of 7 arbitrary units. This resulted in a binary image with the structures in black and the background in white. After that, measurements were redirected to the original image and the particles were analysed by comparing intensity with the original image. Results were manually checked to discard intensities that correspond to

structures in a different plane. For intensity comparison, a triplicated sample was considered. Each sample consists of 10 images acquired in the same conditions as the other samples. Once analysed the images, mean and standard deviation of the intensity quantification was calculated. In addition, sample mean and standard deviation was calculated using 8 values, discarding the two images with the minimum and maximum intensity. These values were statistically compared using GraphPad Prism 6.01 to test their significance. T-test-unpaired with Welch correction, which not assume equal standard deviation of the two populations, was used. Approximately 150-200 chloroplasts were analysed per sample.

3.6 RNA extraction

RNA extraction was performed with Ribozol™ RNA Extraction Reagent ™ (Amresco). *N. benthamina* leaves were grinded in liquid nitrogen with a ceramic mortar and pestle, until the tissue was turned into a fine powder. An amount of 50 mg was collected per sample in a 1.5 ml tube and 1 ml of Ribozol was added to each tube and mixed thoroughly. After that, 200 µl of chloroform was added, vigorously mixed with the Ribozol extract and afterwards centrifuged for 15 minutes at 13 000g at 4 °C. Supernatant was collected in new tubes and 1 volume of isopropanol was added to each one, samples were incubated for 15 minutes at room temperature and afterwards centrifuged for 15 minutes at 13 000g at 4 °C. Supernatant was discarded and 1 ml of 70 % ethanol was added to wash the resulting pellets, and centrifuged for 5 minutes at 13 000g at 4 °C. Pellets were resuspended with water (90 µl), and after that, 10 µl of sodium acetate 3 M and 2.5 V of absolute ethanol were added. Samples were incubated for 30 minutes at -80 °C and centrifuged 15 minutes at 13 000g at 4 °C. Supernatant was discarded and the resulting pellet was resuspended with water. RNA was quantified in the NanoDrop™2000 microvolume spectrophotometer (Termo Fisher Scientific) for further analysis.

3.7 Northern Blot analysis

Northern Blot was performed as previously described (Pallás et al., 1998), using an YFP probe to detect the messenger RNA of interest and the ribosomal RNA 25S to normalise band intensity according to the quantity of RNA transferred to the membrane.

RNA electrophoresis

First, electrophoresis gel casting tray and comb were sterilised with hydrogen peroxide (H_2O_2) for 15 minutes to inactivate RNases. Agarose was dissolved in water by heating and afterwards MOPS buffer (20 mM 3-(N-morpholino)-propanesulfonic acid, 5 mM sodium acetate, 1 mM EDTA) and formaldehyde were added, resulting in a final percentage of 1.5 % agarose, 1 % MOPS and 2 %

formaldehyde. Samples were mixed with 1 V of freshly prepared 2X RNA loading buffer (to prepare 1 ml mix, 670 µl formamide, 200 µl 37 % formaldehyde, 130 µl 10X MOPS, 0.1 % bromophenol blue, 0.1 % xylene cyanol and 10 µg/ml ethidium bromide) and heated for 10 minutes at 65 °C to denaturalise the RNA. Samples were loaded in the gel and run for 1 hour at 80 V in 1X MOPS as running buffer. The integrity of total RNAs was verified by visualising EtBr-stained ribosomal RNAs in a UV transilluminator as described before.

RNA transference by capillary blotting

RNA was transferred by capillarity from the agarose gel to a nylon membrane with positive charge (Roche). RNA gel was submerged in buffer 10X SSC (0.75 M NaCl, 0.075 M sodium citrate, pH 7) for 15 minutes. Capillary transfer was set up using 10X SSC. A glass plate was placed over a plastic tray with 400-500 ml of 10X SSC. Two pieces of wet 3MM filter paper were placed over the glass plate with their two ends dipping into the 10X SSC buffer reservoirs as a bridge. Then, the agarose gel was placed upside down on top of the 3MM papers and overlaid with the nylon membrane. Two pieces of 3MM filter paper and a stack of regular filter paper (10 mm thick) were placed over the membrane. Finally, a glass plate and a 500 ml bottle of water were positioned on top of the stack to serve as a weight. RNA transference onto the nylon membrane was performed overnight.

Hybridisation and development

Nucleic acids transferred to the membrane were covalently fixed by using ultraviolet light (700x100 J/cm^2). After that, the membrane was introduced in an hybridisation tube and blocked with 10 ml of hybridisation solution (50 % formamide, 5X SSC, 0.1 % N-Lauroyl sarcosine, 0.02 % SDS and 2X blocking agent) for two hours at 68 °C in a rotary oven. Afterwards, digoxigenin(DIG)-labelled probe, previously denatured for 10 minutes at 80 °C, was added to the hybridisation tube and incubated overnight at 68 °C in the same oven. Subsequently, two washing steps of 5 minutes were performed at room temperature with 2X SSC and 0.1 % SDS, and two more washing steps of 15 minutes at 68 °C with 0.1% SSC and 0.1% SDS. Next steps for DIG detection were performed at room temperature in the hybridisation tube. First step was performed with washing buffer one (0.1 M maleic acid, 0.15 M NaCl, pH 7.5) with 0.3 % Tween 20 for 10 minutes. Second step consists of membrane blocking with buffer one plus 1X blocking agent (Roche) for 30 minutes. After that, it was incubated with anti-dioxigenin-AP Fab fragments 1:10000 (Roche) diluted in buffer one and 1X blocking agent for 30 minutes. Further steps were two washes with buffer one with 0.3 % Tween 20 for 15 minutes and finally one with buffer three (0.1M Tris-HCl, pH 9.5, 0.1 M NaCl). The membrane was transferred to a plastic bag and the chemiluminescent substrate CSPD diluted 1:100 in buffer three was added contacting with the membrane and kept in dark for 5 minutes.

After that, the substrate was eliminated and the chemiluminescent image was acquired with Fujifilm LAS 3000 Imager (Fuji). Band intensity was quantified using Image Gauge V4.0 (Fuji) and statistically compared using T student test.

Probe stripping

Previously to the second hybridisation with the ribosomal RNA 25S probe, the YFP probe was removed from the membrane. For stripping, the membrane was incubated twice for 60 minutes at 80 °C in a solution containing 50 % deionised formamide, 5 % SDS and 50 mM Tris-HCl, pH 7.5. Finally, it was rinsed with 2X SSC for 5 minutes and hybridised again as described above.

3.8 Chloroplast isolation

Agroinfiltrated leaves were homogenised at 4 °C with 10 ml of extraction buffer (300 mM sorbitol, 1 mM $MgCl_2$, 50 mM HEPES-KOH, pH 7.8, 2 mM EDTA, 0.04 % β-mercaptoethanol and 0.1 % polyvinilpyrrolidone). The homogenate was filtered through rayon-polyester Miracloth (Millipore) and centrifuged 10 minutes at 1500g at 4 °C. The supernatant was discarded and the resulting pellet was resuspended in 2 ml of the isolation buffer (300 mM sorbitol, 1 mM $MgCl_2$, 50 mM HEPES-KOH, pH 7.8, and 2 mM EDTA). 1 ml of the resuspended extract was loaded in a discontinuous Percoll® (Sigma) gradient (5 ml 10 %-2.5 ml 80 %), and was centrifuged at 8000g at 4 °C for 20 minutes in a swinging rotor. Intact chloroplasts were in a green band between 10 % and 80 % Percoll® layers. Chloroplasts were removed from the gradient with a Pasteur pipette and washed with one volume of isolation buffer, and afterwards centrifuged 5 minutes at 1000g. The supernatant was discarded and the resulting pellet, which contains the chloroplasts, was stored at -20 °C. After that, RNA extraction was proceeded as described above. Protein extraction was also performed to analyse the degree of purity and/or contamination of the isolated chloroplasts.

3.9 Protein extraction

N. benthamina leaves were homogenised with Laemmli buffer (62.5 mM Tris-HCl, pH 6.8, 15 % glycerol, 2 % SDS, 350 mM DTT) for protein extraction. Samples were boiled for 15 minutes and centrifuged at 13000g for 15 minutes. The resulting pellet was discarded and the supernatant was stored at -20 °C for further analysis.

3.10 Western blot analysis

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was prepared using the MiniProtean 3 system (BioRad). The resolving gel was 12 % acrylamide (40 % acrylamide mix, 1 M Tris pH 6.8, 10 % SDS, 0.01 % ammonium persulfate, 0.001 % TEMED) and the staking gel was 5 % acrylamide (40 % acrylamide mix, 1 M Tris, pH 6.8, 10 % SDS, 0.01 % ammonium persulfate, 0.001 % TEMED). After gel polymerisation, electrophoresis buffer tank was filled with protein electrophoresis buffer (0.025M Tris-HCl, pH 8.3, 0.2 M glycine and 1 % SDS). Samples were boiled at 95 °C for 5 minutes in order to be sure that proteins are completely denaturalised. Then, samples and also a pre-stained protein molecular weight ladder (Fermentas) were loaded and run for 2 hours at 100 V.

When protein staining was required, a solution of Coomassie brilliant blue (0.05 % Coomassie brilliant blue R-250, 10 % acetic acid and 50 % methanol) was used to incubate the gel for 10 minutes in a shaker. After that, it was incubated with a distaining solution (10 % acetic acid and 50 % methanol) for 15 min in a shaker. Distaining solution was renewed several times until the stained protein bands could be visualised against a light blue background.

Protein blotting by wet transference

Gels with separated proteins were transferred to polyvinylidene fluoride (PVDF) membranes (GE Healthcare Life Sciences). Previously to the transference, the membranes were submerged in 100 % methanol for 5 minutes. After methanol membrane activation, the transference was done using the wet transfer method and the Mini Trans-Blot Cell system (BioRad). The transfer sandwich (foam pad-3MM filter paper-gel-PVDF membrane-3MM filter paper-foam pad) was assembled under cold transfer Towbin buffer without methanol (25 mM Tris-HCl, pH 8.3, 192 mM glycine) into a holder cassette. The sandwich was placed vertically in the core assembly module between the wire electrodes and, finally, accommodated into a buffer tank filled with cold transfer buffer. Transference conditions were one hour at 100 V.

Inmunodetection

After the transference was finished, the membrane was washed with TBST solution (1X TBS: 20 mM Tris-Cl, pH 7.5 and 150 mM NaCl with 0.1 % Tween 20) for 5 minutes and incubated with blocking solution (TBST solution and 5 % skimmed milk powder) for at least an hour and a half at 4 °C in soft agitation. Afterwards, the membrane was washed with TBST solution for 5 minutes and incubated overnight with hybridisation solution (2.5 % skimmed milk powder and TBST solution) that contained the corresponding primary antibody. Western Blot analysis was done for detecting

the large subunit of RuBisCO (chloroplast marker), the UPD-glucose pyrophosphorylase (cytoplasm marker) and YFP in infiltrated leaf tissue. All three proteins were detected using primary antibodies obtained in rabbit at 1:10,000, 1:3,000 and 1:10,000 dilutions, respectively. Two washes of 10 minutes each were performed with TBST to eliminate the primary antibody that was not specifically bound. Anti-rabbit secondary antibody conjugated with peroxidase was employed at 1:10,000. Once the membrane was incubated with the secondary antibody for at least one hour, three washes with TBST solution were carried out. Finally, the membrane was washed for 10 minutes with 1X TBS.

Photographic film development

Chemiluminescent substrate (Pierce™ ECL Western Blotting Substrate) was prepared according to the manufacturer instructions (Thermo Fisher Scientific) and applied onto the membrane for 5 minutes in darkness. For signal detection, the membrane was exposed to an autoradiography photographic film (Amersham Hyperfilm™ ECL) and developed with a M35X-Omat automatic processor (Kodak).

3.11 RT-PCR

Prior to RT-PCR analysis, RNA samples, which had been previously quantified, were treated with DNAase I for 30 minutes at 37 °C, to eliminate possible DNA contamination. Then, 50 mM EDTA was added to the RNA samples, and next, they were heated for 10 minutes at 65 °C to inactivate the enzyme. Superscript™ III One-Step RT-PCR with Platinum® Taq was used for one-step RT-PCR (Thermo Fisher Scientific). Reactions were carried out in a 5 µl final volume containing 2X Reaction Mix, 200 ng of RNA, 0.5 µl of each primer (10 µM), 0.4 µl of SuperScript™ III RT/ Platinum ® Taq enzyme mix, 0.1 µl RiboLock RNase inhibitor (40 U/µL) and nuclease-free water up to 10 µl. RT-PCR conditions were an initial step of 30 min at 55 °C followed by 23, 25 or 27 PCR cycles of denaturation (94 °C), annealing (60 °C) and extension (72 °C). The duration of all steps in the cycles were 30 s

RESULTS

4. RESULTS

The ELVd-derived sequence was previously cloned upstream of a yellow fluorescent protein (YFP) reporter gene, and downstream of a strong constitutive transcription promoter. This later consisted of a duplicated *Cauliflower mosaic virus* (CaMV) 35S promoter (S3), which is significantly stronger than the natural 35S promoter, with the untranslated leader sequence of *Alfalfa mosaic virus* (AMV) RNA 4 as a translational enhancer (Kay et al., 1987) (Figure 4). Additionally, the terminator sequence of the *Solanum tuberosum* proteinase inhibitor II gene (PoPit) was used as transcription terminator (Genovés et al., 2006). This expression cassette was inserted into a binary plasmid (pMOG800) and the resulting construct was denominated p[^{AMV}ELVd-derived]YFP.

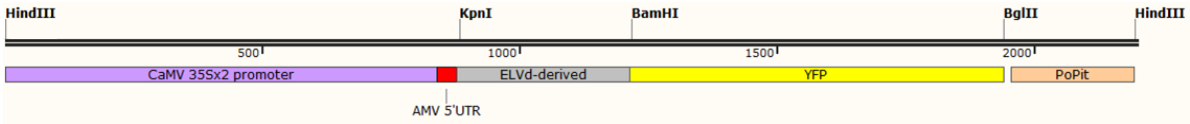


Figure 4. Schematic representation of the expression cassette of the p[^{AMV}ELVd-derived]YFP clone. Restriction enzyme sites relevant for the cloning process are in bold.

The Gómez and Pallás initial construct consisted of the same ELVd-derived sequence and, instead of YFP, GFP as reporter (Gómez and Pallás, 2010b). This change was made to avoid chlorophyll and GFP crosstalk or bleed-through that could lead to false GFP localisation in chloroplast when imaging in laser scanning confocal microscopy (LSCM). Crosstalk can often occur when two or more fluorescent markers are simultaneously excited using a single excitation laser line and, due to the spectral properties of the fluorescence, their emission ranges overlap. GFP and chlorophyll, specially chlorophyll b, are both usually excited in LCSM using the 488 nm line of the argon ion laser. As GFP emission range overlap with chlorophyll emission range at their tails and particularly under conditions of high laser intensity, chlorophyll auto-fluorescence can be registered and misinterpreted as GFP fluorescence. Despite the YFP emission peak (527 nm) is closer to the chlorophyll emission peak (around 650 nm for chlorophyll b) than this of the GFP (507 nm), YFP excitation is achieved by setting the argon ion laser to 514 nm line that is far apart from the chlorophyll excitation peak. Therefore, sequential acquisition using both laser lines (488 nm for chlorophyll and 514 nm for YFP) serves to specifically detect and clearly separate YFP fluorescence from chlorophyll autofluorescence.

In addition to this GFP-to-YFP change, several variants of the initial YFP expression cassette were obtained for system optimisation. The different constructs were named according to the modifications made on the initial YFP expression cassette. In order to increase protein expression levels in the chloroplast, 5' UTR sequences were changed as will be further detailed. Additionally, some deletions and a non-related-ELVd sequence having a secondary structure analogue to the

ELVd-derived sequence were also cloned to study their functionality in delivering YFP RNA to chloroplasts.

4.1 Analysis of different 5' UTR leader sequence effect on protein expression.

As a first approach to optimise the ELVd-derived-YFP expression cassette, we decided to test different 5' UTR leader sequences to increase protein expression levels in chloroplasts. The expression of a gene is the outcome of the interplay of multiple processes and both transcriptional and posttranscriptional parameters can be regulated for controlling expression. At transcriptional level, improvement of this cassette was unlikely as duplicated 35S is already among the strongest constitutive promoters described so far. Consequently, to realise the full potential of the duplicated 35S promoter strength we focused on posttranscriptional parameters, and more specifically in testing different 5' UTR leader sequences, because they have a significant role in determining translation rate (Fan et al., 2012). As indicated before, the 5' UTR leader sequence used in p[^{AMV}ELVd-derived]YFP was the 5' UTR of the AMV RNA 4 (Datla et al., 1993). According to the literature, this 45 nt long sequence is one of the strongest translation enhancers described in plants so far (Kanoria and Burma, 2012), but we decided to test other leader sequences described as such to optimise the YFP expression cassette.

Table 1. Sequences of the different 5' UTR analysed in the present section

Name	Length (nt)	Sequence (5' to 3')
synJ	28	ACACGCTGGAATTCTAGTATACTAAACC
Ω	77	ACCTCGAGTATTTTTACAACAATTACCAACAACAACAAACAACAAACAACA TTACAATTACTATTTACAATTACACC
AMV	45	ACCTCGAGTTTTTATTTTTAATTTTCTTTCAAATACTTCCATCCC

First, the AMV RNA 4 leader sequence was changed to another leader sequence of viral origin, the 5' UTR of the *Tobacco mosaic virus* (TMV) RNA, denominated Ω (Gallie et al., 1988). The resulting construct was called p[^{TMV}ELVd-derived]YFP. Additionally, a 5' UTR of synthetic origin, denominated synJ (Kanoria and Burma, 2012) was also cloned replacing the AMV 5' UTR, and resulting in p[^{synJ}ELVd-derived]YFP construct (Figure 5A and Table 1)

LCSM imaging of leaf tissue agro-infiltrated with bacterial cultures either harbouring p[^{AMV}ELVd-derived]YFP, p[^{TMV}ELVd-derived]YFP or p[^{synJ}ELV-derived]YFP construct revealed that YFP was specifically accumulated in chloroplasts in all cases (Figure 5B). Apparently, chloroplast fluorescence intensity of images from the tissue expressing p[^{synJ}ELV-derived]YFP was much lower

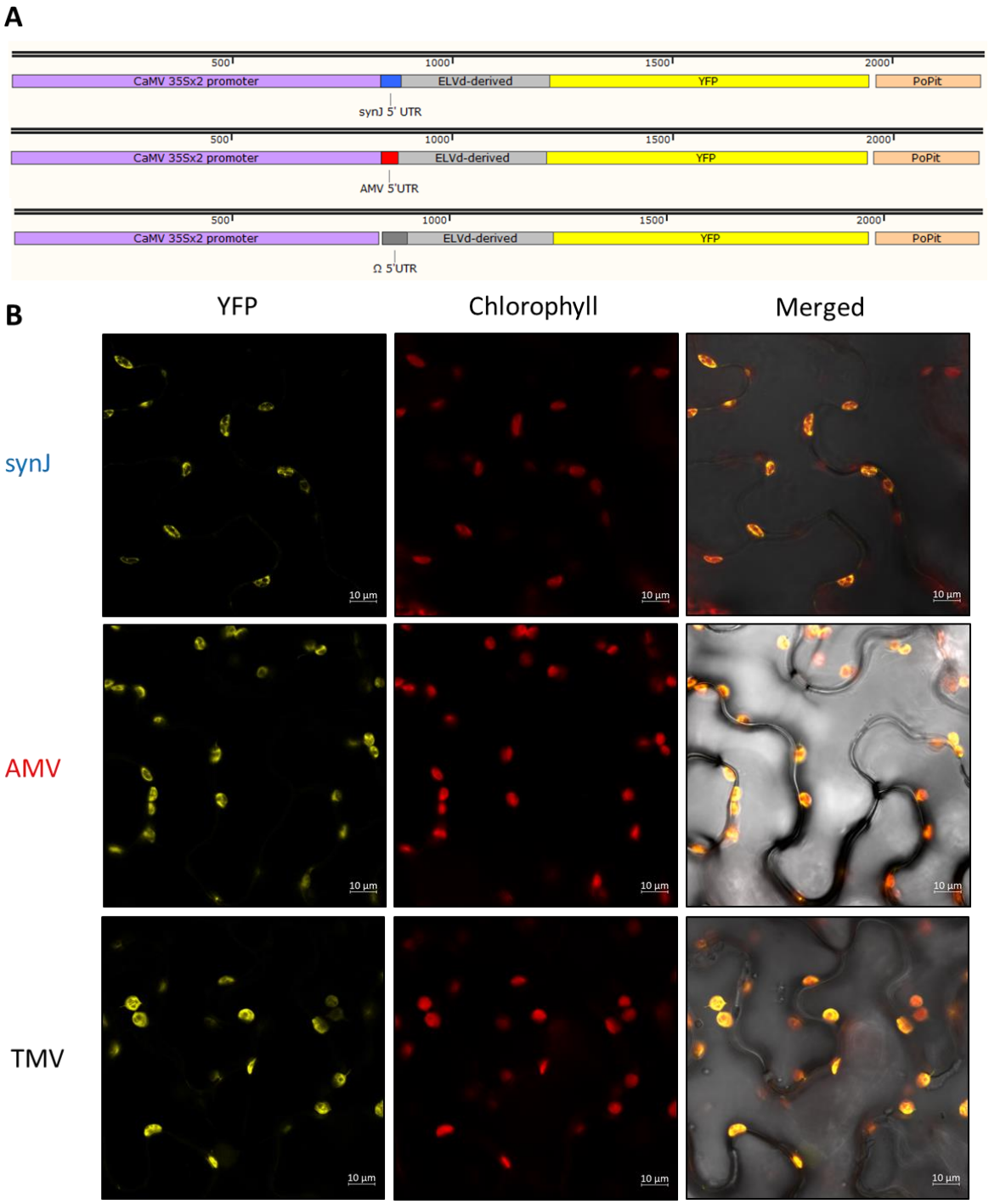


Figure 5. A) Schematic representation of the YFP expression cassette of different 5' UTR constructs. p[^{synJ}ELV-derived]YFP (up), p[^{AMV}ELVd-derived]YFP (middle) and p[^{TMV}ELVd-derived]YFP (down). **B)** Laser scanning confocal microscopy images of agroinfiltrated *N. bethamiana* leaf tissue expressing different 5' UTR-ELVd-derived constructs. p[^{synJ}ELV-derived]YFP (upper panels), p[^{AMV}ELVd-derived]YFP (middle panels) and p[^{TMV}ELVd-derived]YFP (bottom panels). Chlorophyll fluorescence is used as a control of chloroplast location. For qualitative purpose, these images were acquired by setting the microscope at different gain values; therefore the fluorescence intensity levels cannot be compared among them.

than the intensity observed in leaves expressing p[^{AMV}ELVd-derived]YFP, whereas the fluorescence of chloroplasts from the tissue expressing p[^{TMV}ELVd-derived]YFP was slightly more intense than in the case of p[^{AMV}ELVd-derived]YFP (Figure 6A). To test if the apparent intensity differences were

statistically significant, relative fluorescence from randomly selected chloroplasts was quantified as described in Materials and Methods to apply a T-Student significance test for the mean (Figure 6B and 6C). Differences were statistically significant in both cases (S3).

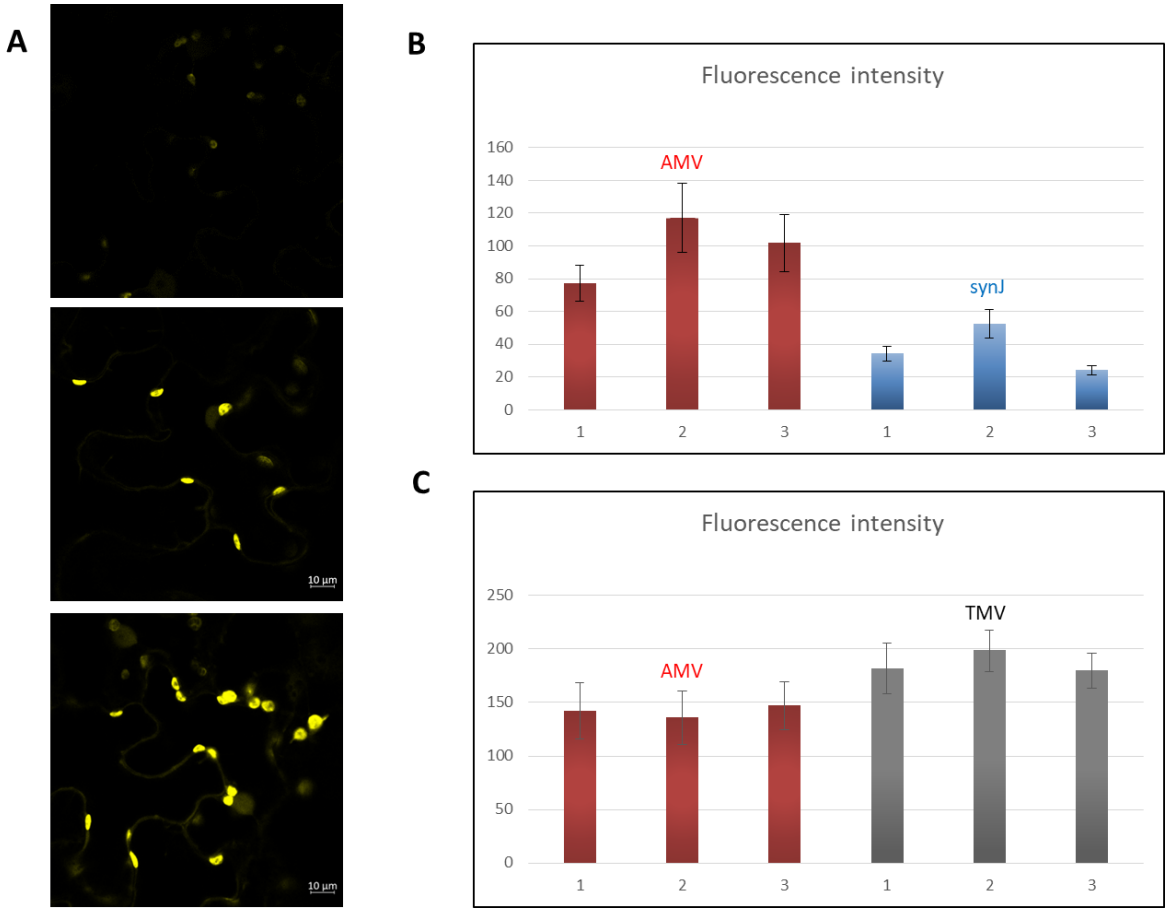


Figure 6. Quantitative and qualitative analyses of fluorescence intensity in chloroplasts. **A)** Confocal microscopy images, acquired setting up the microscope at identical gain and laser intensity conditions, of agroinfiltrated *N. bethamiana* leaf tissue expressing p[^{synJ}ELVd-derived]YFP (upper panel), p[^{AMV}ELVd-derived]YFP (middle panel) and p[^{TMV}ELVd-derived]YFP (bottom panel). Each bar represents the mean of the fluorescence intensity in chloroplasts from 8 randomly selected microscope fields of view from a leaf disk. Mean and standard deviation are showed. Approximately 150-200 chloroplasts were analysed per sample. **B)** Bar graph showing the relative YFP fluorescence intensity comparison between *N. benthamiana* agroinfiltrated leaf tissue expressing p[^{AMV}ELVd-derived]YFP construct and p[^{synJ}ELVd-derived]YFP construct **C)** Bar graph showing the relative fluorescence intensity comparison between *N. benthamiana* agroinfiltrated leaf tissue expressing p[^{AMV}ELVd-derived]YFP construct and p[^{TMV}ELVd-derived]YFP construct.

A Northern blot analysis was carried out in order to analyse whether the increase in the protein levels could have any relation with a change in the amount of transcripts or was strictly due to a translation enhancement. Equal amount of total RNAs extracted from agroinfiltrated leaves expressing each of the three construct were electrophoresed in denaturing conditions. As loading control, total RNAs were stained with ethidium bromide and 25S ribosomal RNA (rRNA) was specifically detected by Northern blot (Figure 7A and 7B, respectively). Next, the same membrane was stripped to remove the attached 25S rRNA riboprobe and hybridised for a second time using a riboprobe against the YPP mRNA. YFP mRNA signal intensity was quantified and normalised

respect to 25S rRNA signals (Figure 7C and 7D). Results revealed that p[^{synJ}ELVd-derived]YFP expressing samples showed the lowest transcript amount, which correlates with the observed fluorescence levels (Figure 7A and 7B). p[^{TMV}ELVd]YFP seems to produce YFP mRNA levels superior to p[^{AMV}ELVd-derived]YFP clone, therefore it is also in agreement with fluorescence quantification.

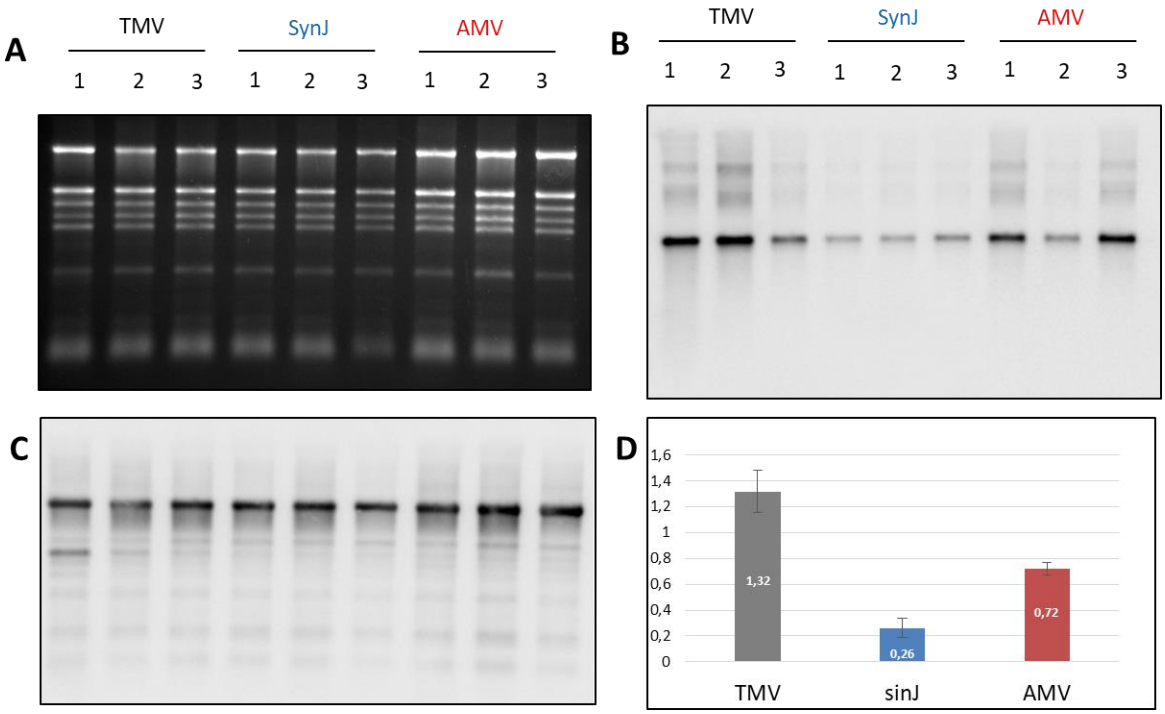


Figure 7. Northern blot analysis of total RNA extracted from agroinfiltrated *N. bethamiana* leaf tissue expressing [ELVd-derived]YFP RNA under the control of the indicated 5' leader sequence. **A)** RNA electrophoresis in denaturing conditions stained with ethidium bromide previous to membrane transference. **B)** The image corresponds to the chemiluminescent detection of the 25S ribosomal RNA used as loading control of RNA quantity. Loaded RNA was supposed to be adjusted to 500 ng per sample. **C)** The image corresponds to the chemiluminescent detection of the YFP mRNA. Three samples derived each one from a distinct leaf were analysed per construct (1, 2 and 3). **D)** Bar graph showing the comparison among the means of the normalised YFP hybridisation signal intensities in arbitrary units.

4.2 Identification of the minimal structural motif of the ELVd-derived sequence involved in chloroplast RNA targeting

Determination of the ELVd-derived minimal sequence fully efficient for leading RNA traffic to the chloroplast would be convenient for further application. Thus, two shorter versions of the ELVd-derived RNA were constructed by removing sequences at the 5' and 3' ends but keeping its central core folding structure, because it was assumed that this RNA secondary structure was responsible for RNA trafficking to the chloroplast (see Introduction). The shortest version consisted of a 181-nt length sequence (ELVd-derived$_{181}$) that conserved the unpaired nucleotides of the central loop, three stem-loops that formed a trident-like structure and a small stem-loop opposite to that trident-like structure (Figure 8A, RNA structure on the left). The longest version consisted of a 247-

nt length sequence (ELVd-derived$_{247}$) that had an additional long stem-loop bifurcated in two small additional branches (Figure 8A, RNA structure on the middle). The only differences between ELVd-derived$_{247}$ and full-length ELVd-derived sequences are three stem-loops that correspond to the extremes of the sequence (Figure 8A, RNA structure on the right). ELVd-derived$_{181}$ and ELVd-derived$_{247}$ sequences were included into a YFP expression cassette similar to that of p[^{AMV}ELVd-derived]YFP and cloned into a binary vector to generate p[^{AMV}ELVd-derived$_{247}$]YFP and p[^{AMV}ELVd-derived$_{181}$]YFP (Figure 8B).

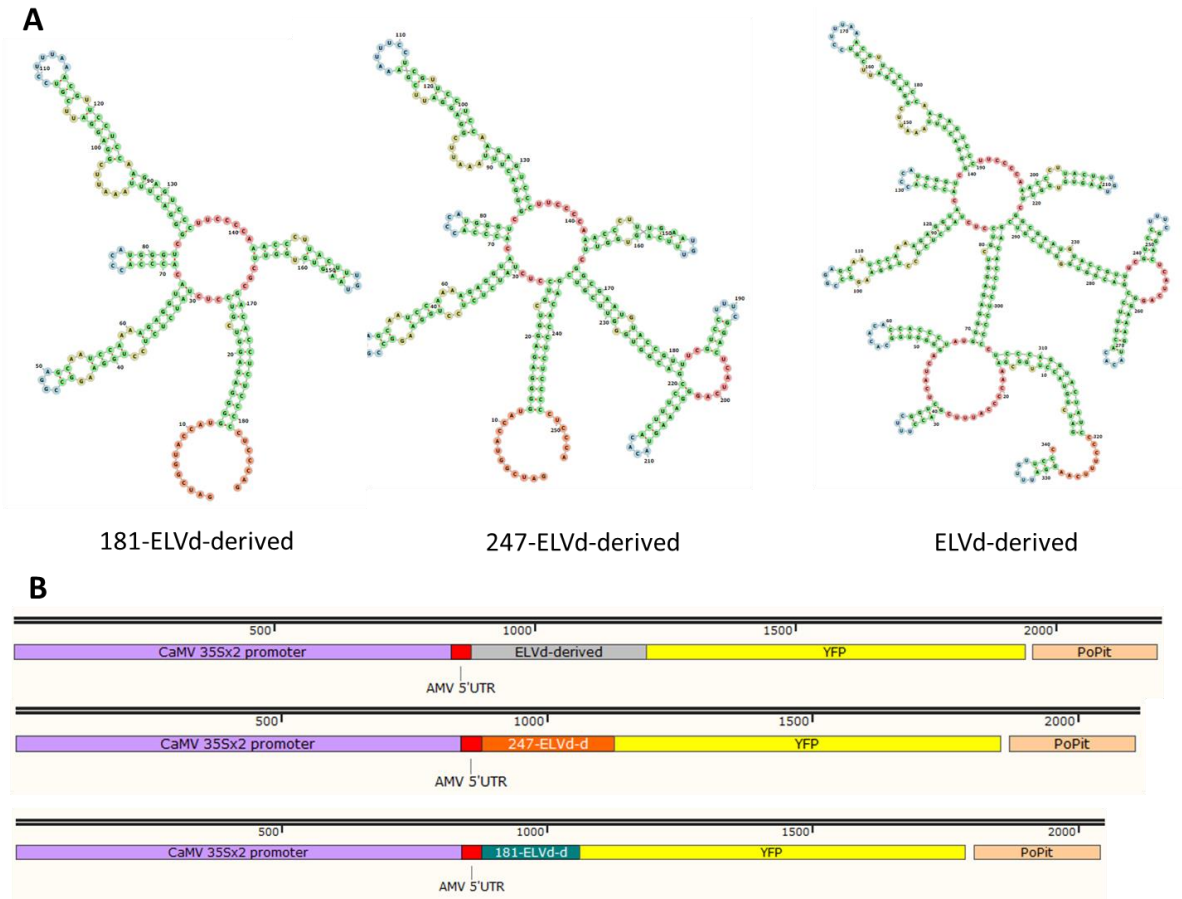


Figure 8. A) RNA secondary structures predicted with the *RNAfold* web server of the ELVd-derived$_{181}$ sequence (left), ELVd-derived$_{247}$ sequence (middle) and ELVd-derived sequence (right). The represented structures also include seven non-related nucleotides adjacent to the viroidal sequence. **B)** Schematic representation of the YFP expression cassette of p[^{AMV}ELVd-derived]YFP (up), p[^{AMV}ELVd-derived$_{247}$]YFP (middle) and p[^{AMV}ELVd-derived$_{181}$]YFP (down) plasmids.

LSCM images revealed that YFP expressed from p[^{AMV}ELVd-derived$_{181}$]YFP clone showed high expression levels but was localised in the nucleus and cytoplasm, as expected from an untagged YFP translated in the cytoplasm (Figure 9, upper panels). YFP expressed from p[^{AMV}ELVd-derived$_{247}$]YFP had low expression levels and showed a dual subcellular localisation pattern, being mainly found in the nuclei and cytoplasm but very occasionally in the chloroplasts (Figure 9, middle panels). From these observations, it seems that ELVd-derived$_{181}$ RNA was unable to be targeted to the chloroplasts whereas ELVd-derived$_{247}$ RNA apparently underwent a dramatic reduction of the chloroplast RNA targeting efficiency.

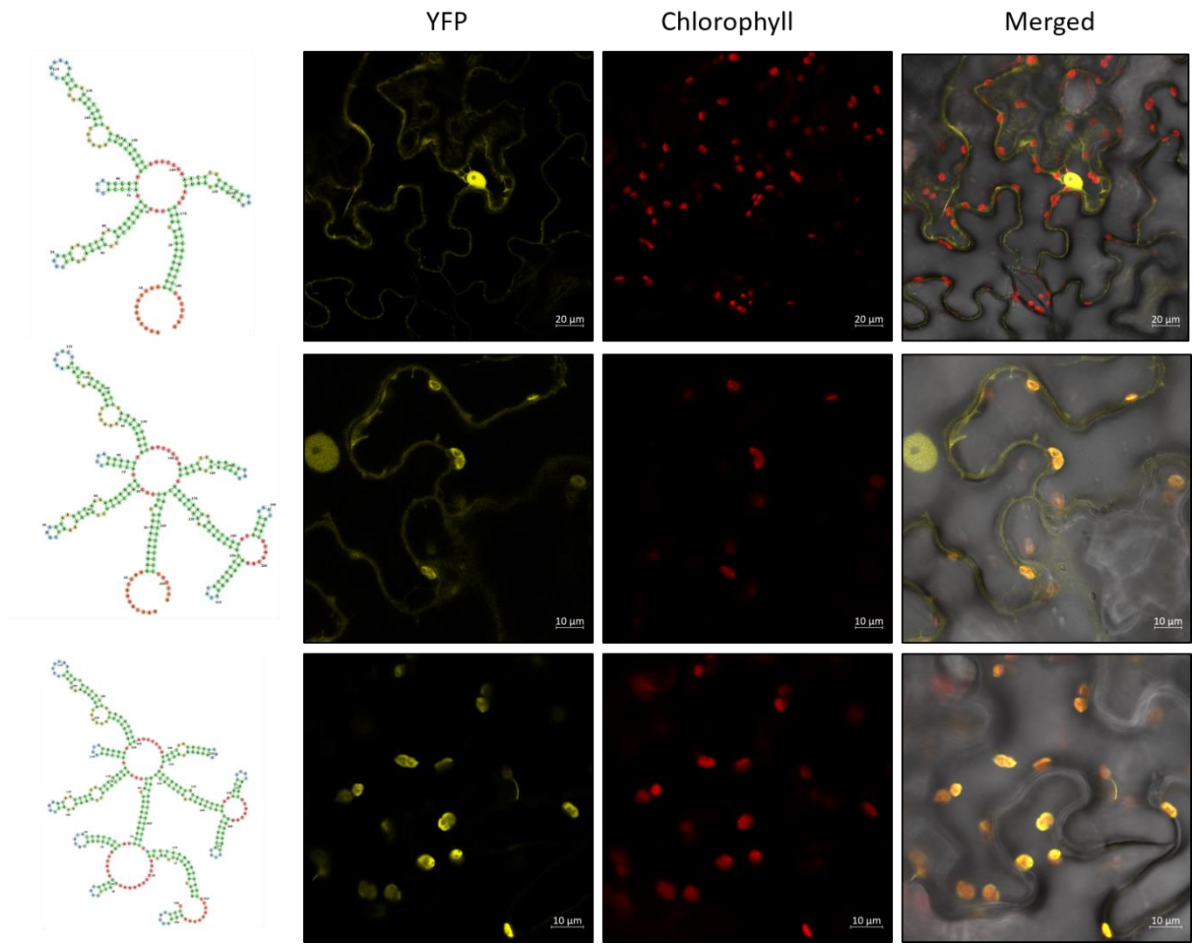


Figure 9. Laser scanning confocal microscopy images of agroinfiltrated *N. benthamiana* leaf tissue expressing YFP from p[^{AMV}ELVd-derived$_{181}$]YFP (upper panels), p[^{AMV}ELVd-derived$_{247}$]YFP (middle panels) and p[^{AMV}ELVd-derived]YFP (bottom panels). Chlorophyll fluorescence was used as chloroplast marker. These images were acquired by setting the microscope at different gain values and laser intensities; therefore the intensity levels cannot be compared.

4.3 Designing of a non-viroid-related RNA showing an ELVd-derived analogous secondary structure and its functional analysis in chloroplast RNA transport.

For biosecurity concerns, we were searching for a sequence not recognised as derived from a viroid and that consequently could not be related to a plant pathogen. Considering the full ELVd-derived RNA as a starting point, a different non viroid-related RNA that fold in an ELVd-derived analogous secondary structure was designed using the *RNAfold* web server. This sequence with the same predicted secondary structure was denominated CTVL (chloroplast transporter viroid-like) (Figure 10A). CTVL conserved 63 % of the original ELVd-derived sequence, but it was arranged in such a way that it was different enough to be not recognised in a BLAST® (Basic Local Alignment Search Tool) search as ELVd; therefore CTVL can be considered as a truly synthetic sequence. As mentioned in Materials and Methods, the CTVL design consisted in a hard trial and error procedure of switching the sequence in base-paired regions of the stems from one side to the

other and keeping all ribonucleotides in the loops. The final CTVL and the ELVd-derived predicted structures are compared in Figure 10A. Except for the stem connecting the 5' and 3' CTVL ends, which is three base pairs shorter than in the ELVd-derived RNA structure (see the bottom right side of the represented structures), both secondary structures were undistinguishable. CTVL sequence was included into a YFP expression cassette similar to that of p[^{AMV}ELVd-derived]YFP and cloned into a binary vector to generate p[^{AMV}CTVL]YFP (Figure 10B).

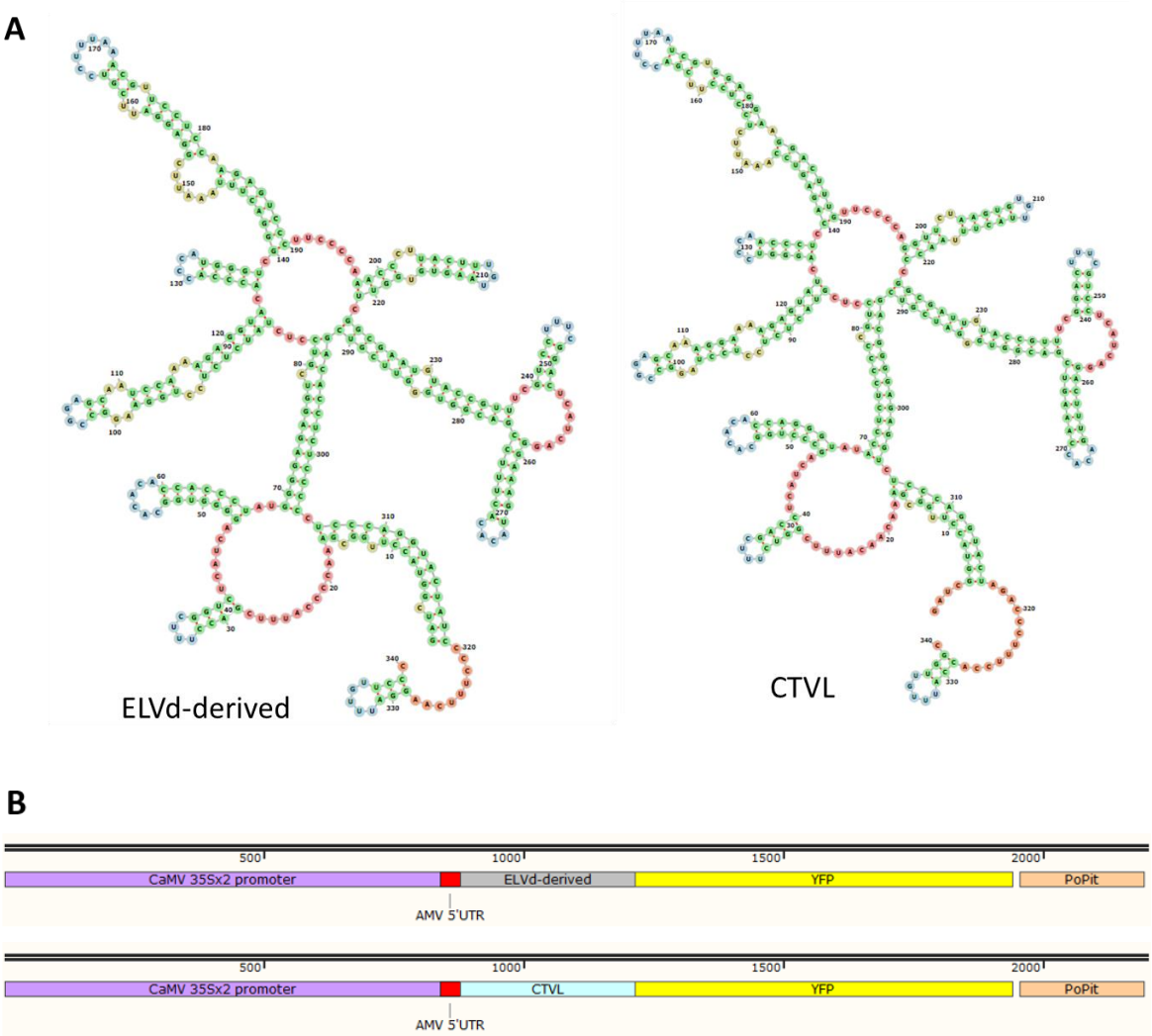


Figure 10. A) RNA secondary structures predicted with the *RNAfold* web server of the ELVd-derived sequence (left) and the CTVL sequence (right). **B)** Schematic representation of the YFP expression cassette of p[^{AMV}ELVd-derived]YFP (up) and p[^{AMV}CTVL]YFP (down).

LSCM images revealed that the subcellular distribution pattern of the YFP resulting from p[^{AMV}CTVL]YFP expression was nucleus-cytoplasmic (Figure 11); therefore, that imitating structure did not seem to keep the chloroplast transport functionality of the ELVd-derived sequence.

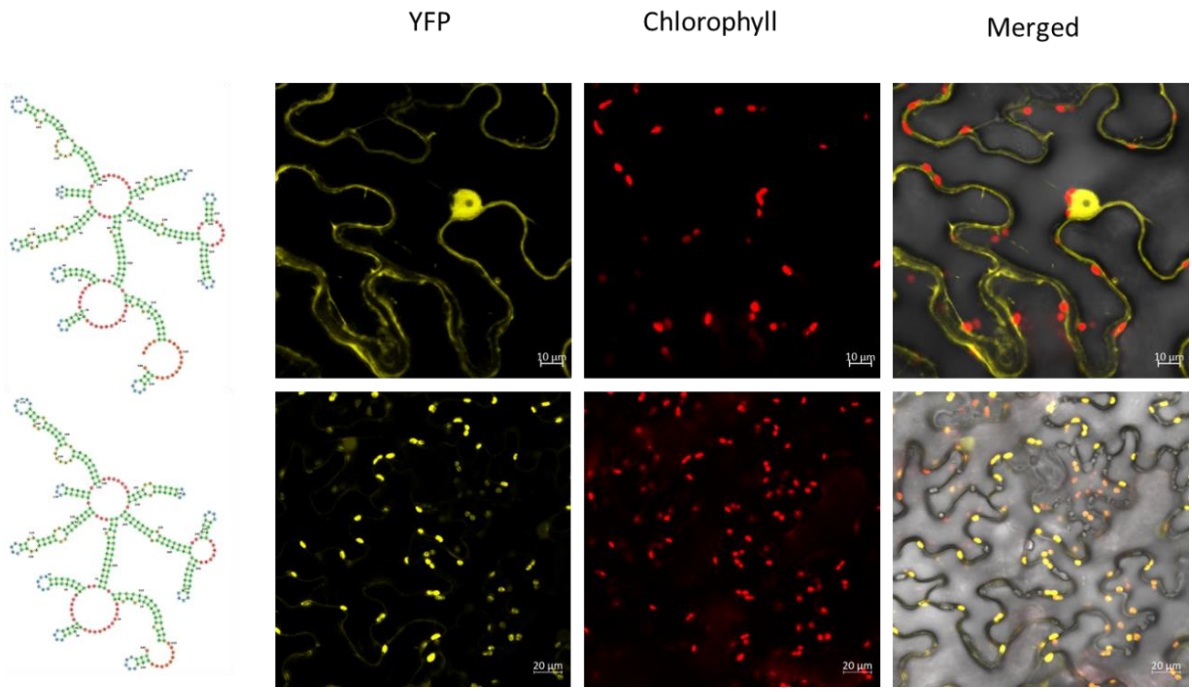


Figure 11. Laser scanning confocal microscopy images of *N. benthamiana* leaf tissue agroinfiltrated with a bacteria culture harbouring p[^{AMV}CTVL]YFP (upper panels) and p[^{AMV}ELVd-derived]YFP (bottom panels). Chlorophyll fluorescence was used as chloroplast marker. These images were acquired by setting the microscope at different gain values and laser intensities; therefore the intensity levels cannot be compared.

4.4 YFP mRNA detection in isolated chloroplasts.

In order to corroborate LSCM results, total RNA from purified chloroplasts was isolated to analyse the presence/absence of the YFP mRNA by RT-PCR. Chloroplasts were isolated through a Percoll® gradient as described in Material and Methods (Figure 12A). Agroinfiltrated *N. benthamiana* leaves expressing YFP from p[^{AMV}ELVd-derived]YFP, p[^{AMV}ELVd-derived$_{181}$]YFP, p[^{AMV}ELVd-derived$_{247}$]YFP, p[^{AMV}YFP] and p[^{AMV}CTVL]YFP were used as chloroplasts source. YFP expression cassette of p[^{AMV}YFP] contains neither original nor modified ELVd-derived sequences. Therefore, it was used as a control of an YFP mRNA accumulated and translated in the cytoplasm but not delivered to chloroplasts. Chloroplast integrity was verified by optical fluorescence microscopy (Figure 12A). Western blot analysis for detecting a cytoplasmic protein (UPD-glucose pyrophosphorylase) (Figure 12B) and a chloroplastic protein as a control (RuBisCO) (Figure 12C) were performed to assess chloroplast purity. In that analysis, the cytoplasmic protein was not detected in chloroplast samples, suggesting that there were no cytoplasmic contaminants, whereas the chloroplast-encoded protein was detected in both total and chloroplast samples. Additionally, an electrophoresis of RNA isolated from chloroplast samples was performed to analyse the RNA integrity and to check that there were no cytoplasmic RNAs present. As expected, typical rRNAs of eukaryotic size (18S and 25S rRNA) were not detected, whereas chloroplast rRNAs, which are in agreement with rRNAs of prokaryote size (16S and 23S rRNA), were present (Figure 12D).

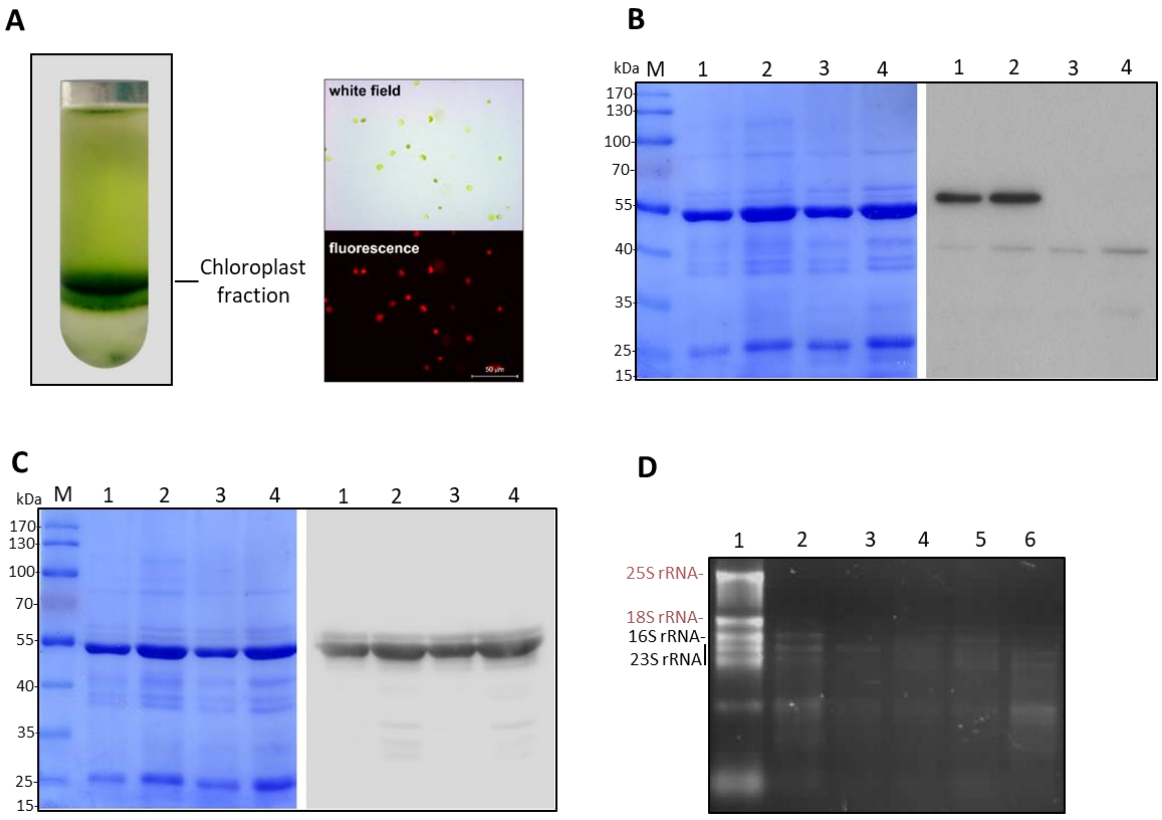


Figure 12. A) Image of a representative Percoll® gradient after centrifugation in a swinging rotor (left). Intact chloroplasts are in a green layer between 10 % and 80 % of Percoll® gradient. **B and C)** Western blot analysis of total proteins extracted from agroinfiltrated *N. benthamiana* leaf tissue. The left panels correspond to proteins stained with Coomassie blue and the right panels to chemiluminescent specific detection. Total protein sample (lanes 1 and 2), chloroplast protein sample (lanes 3 and 4). Primary antibody employed was against large subunit of RuBisCO a chloroplastic encoded protein (**B**), and UDP-glucose pyrophosphorylase, a cytoplasmic protein (**C**). **D)** 1.5 % agarose gel electrophoresis of total RNAs extracted from agroinfiltrated leaf tissue (total sample) and from chloroplasts of that tissue (chloroplastic samples). Total sample: p[^{AMV}ELVd-derived]YFP (1). Cloroplastic samples: p[^{AMV}ELVd-derived]YFP (2), p[^{AMV}ELVd-derived$_{181}$]YFP (3), p[^{AMV}ELVd-derived$_{247}$]YFP (4), p[^{AMV}YFP] (5) and p[^{AMV}CTVL]YFP (6).

RNA extracted from isolated chloroplasts was analysed by semiquantitative RT-PCR to detect YFP RNAs ([^{AMV}ELVd-derived]YFP RNA, [^{AMV}ELVd-derived$_{181}$]YFP RNA, [^{AMV}ELVd-derived$_{247}$]YFP RNA [^{AMV}CTVL]YFP RNA and ^{AMV}YFP RNA). Detection of the rbcL mRNA encoding the RuBisCO large subunit was used as RT-PCR control (Figure 13, upper panel). As seen in Figure 13 (bottom panel), ^{AMV}YFP RNA was detected in chloroplast RNAs from tissue expressing YFP by p[^{AMV}YFP], although in significantly smaller levels than in the other samples. Detection of a non-chloroplast delivered YFP transcript in chloroplast RNAs could have been caused because RT-PCR is a very sensitive technique. However, the number of cycles was low (23, 25 and 27) and the control without any template indicated that reaction components were uncontaminated (data not shown). Consequently, it seems that ^{AMV}YFP RNA detection was caused by some kind of chloroplast sample contamination, most likely nuclear. Chloroplast isolation was repeated several times but in all cases ^{AMV}YFP transcript was weakly detected in this control. In this scenario, we decided to consider the ^{AMV}YFP transcript levels as a background above which YFP RNA detection in

chloroplast could actually be considered positive. [^{AMV}ELVd-derived]YFP transcript was the most abundant in chloroplasts (Figure 13, bottom panel, 25 cycles, lane 5), but [^{AMV}ELVd-derived$_{247}$]YFP transcript, could be also detected in chloroplasts at higher levels than ^{AMV}YFP transcript (Figure 13, bottom panel, lane 3). Interestingly, [^{AMV}ELVd-derived$_{181}$]YFP RNA and, although to a minor extent, [^{AMV}CTVL]YFP RNA were detected in chloroplasts even though nucleus-cytoplasmic YFP distribution was detected in both cases (Figure 13, bottom panel, lanes 2 and 4). According to these results, a relation between YFP accumulation in the chloroplast and the correspondent transcript presence exists, but it seems that transcript presence in chloroplast does not guarantee the corresponding protein accumulation in this organelle.

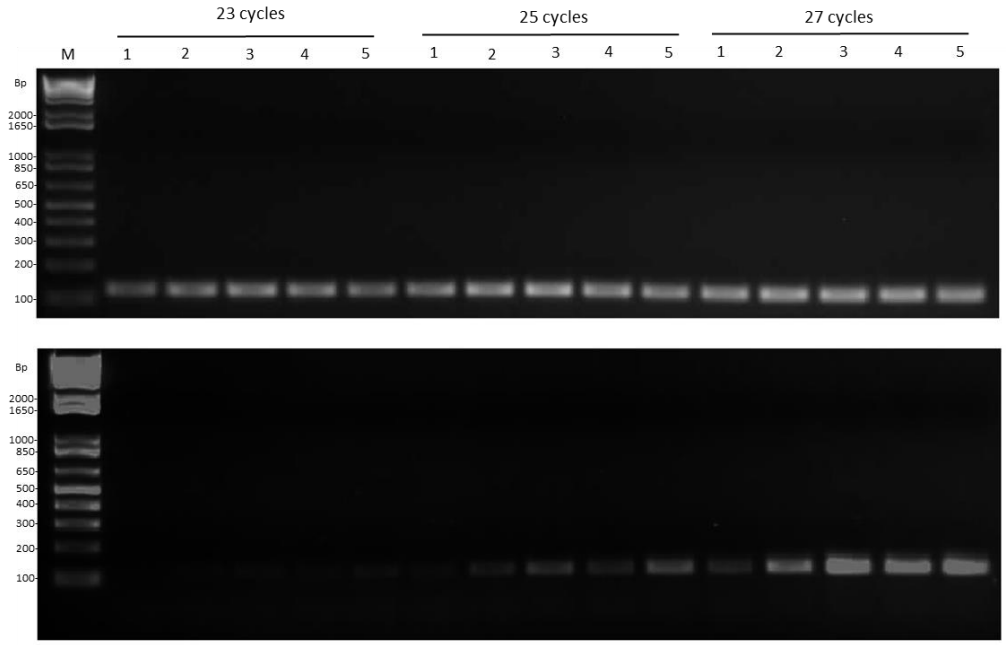


Figure 13. rbcL and YFP RNA detection in chloroplast samples by semiquantitative RT-PCR. 2 % agarose gel electrophoresis analysis of the RT-PCR amplified products obtained using specific primers for the chloroplast rbcL mRNA (upper electrophoresis) and for the different YFP RNAs (lower electrophoresis). RNA was extracted from chloroplasts isolated from agroinfiltrated leaf tissue expressing p[^{AMV}YFP] (1), p[^{AMV}CTVL]YFP (2), p[^{AMV}ELVd-derived$_{247}$]YFP (3), p[^{AMV}ELVd-derived$_{181}$]YFP (4) and p[^{AMV}ELVd-derived]YFP (5).

4.5 Western blot analysis reveal the presence of a tagged YFP always related with YFP fluorescence in chloroplasts.

A routine Western blot analysis, performed with an antibody against the N terminal region of the YFP, revealed the presence of two bands when the ELVd-derived sequence was fused upstream of the YFP ORF in p[^{AMV}ELVd-derived]YFP. One band migrated slower than the free YFP, and the other band migrated according to the size of the free YFP (Figure 14, lanes 1). In contrast, clones expressing YFP with nucleus-cytoplasmic distribution (p[^{AMV}YFP], p[^{AMV}ELVd-derived$_{181}$]YFP and p[^{AMV}CTVL]YFP) showed a single band of 27 kDa (free YFP) (Figure 14, lanes 2, 3 and 5). In agreement with its dual subcellular location in chloroplasts and nucleus-cytoplasm, Western blot

of proteins produced by p[^{AMV}ELVd-derived$_{247}$]YFP expression also resulted in two YFP-related bands. In this case, the additional upper band was less clear and the majority of the protein had the free YFP size (Figure 14, lanes 4). Thus, we observed a relation between the YFP accumulation in the chloroplast and the presence and intensity of the upper band.

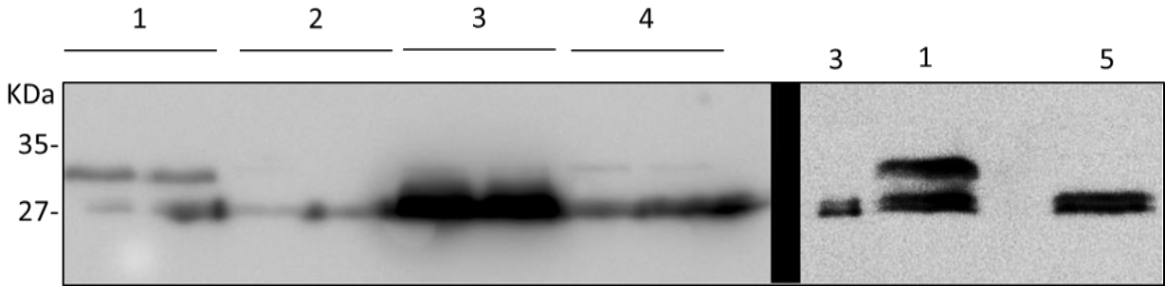


Figure 14. Western Blot analysis of proteins extracted from agroinfiltrated *N. benthamiana* leaf tissue expressing YFP from p[^{AMV}ELVd-derived]YFP (1), p[^{AMV}ELVd-derived$_{181}$]YFP (2), p[^{AMV}YFP] (3), p[^{AMV}ELVd-derived$_{247}$]YFP (4) and p[^{AMV}CTVL]YFP (5). YFP was detected with an antibody against its N terminal region. The two membranes (left and right) are separated by a black bar to show that they correspond to different experiments.

Looking through the [ELVd-derived]YFP RNA sequence, we realised that a peptide, covering the whole ELVd-derived sequence, could be translated *in frame* with the YFP ORF. However, this hypothetical ELVd-derived peptide did not have any AUG codon from which translation could start.

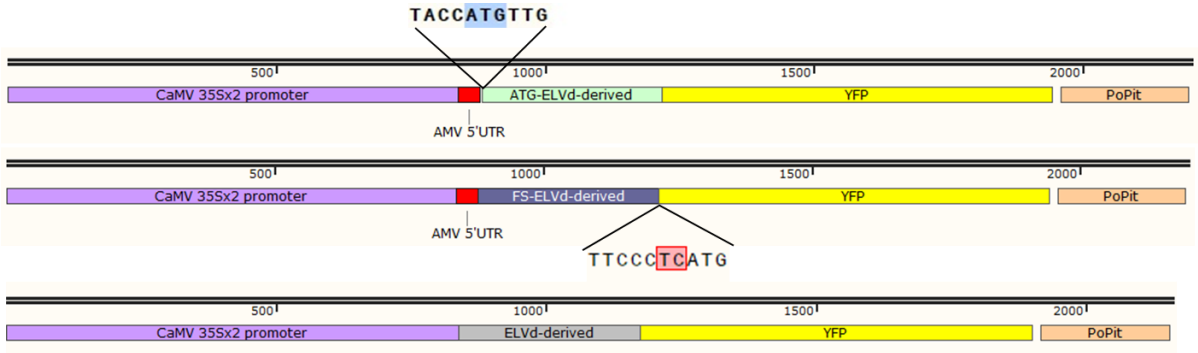


Figure 15. Schematic representation of the YFP expression cassette of p[$^{AMV}_{ATG}$ELVd-derived]YFP (up) that is identical to p[^{AMV}ELVd-derived]YFP except for an ATG insertion, p[^{AMV}ELVd-derived]$_{FS}$YFP (middle) that has a dinucleotide insertion which cause a frameshift in the ELVd-derived-YFP reading frame and p[ELVd-derived]YFP (down) that differs from p[^{AMV}ELVd-derived]YFP construct because it has not the AMV 5' UTR.

In spite of this, and in order to provide more insights about whether or not ELVd-derived RNA was really being translated, we decided to design novel clones. p[$^{AMV}_{ATG}$ELVd-derived]YFP (Figure 15, upper scheme) had an additional ATG, downstream of the AMV 5' UTR and upstream of the ELVd-derived sequence, which was *in frame* with the YFP ORF, and would force the translation of full ELVd-derived-YFP sequence. Additionally, the modification in p[^{AMV}ELVd-derived]$_{FS}$YFP (Figure 15, middle scheme) consisted of a dinucleotide insertion between the ELVd-derived sequence and the ATG starting codon of the YFP ORF that, consequently, would cause a frameshift in the putative ELVd-derived-YFP reading frame. The last modification was a deletion of the AMV 5' UTR to

analyse the influence of the leader sequence in ELVd-derived sequence putative translation (p[ELVd-derived]YFP).

Transient expression of p[$^{AMV}_{ATG}$ELVd-derived]YFP resulted in chloroplast location of the YFP fluorescence (S4). Fluorescence was considerably more intense than that observed with the original construct without the ATG insertion (Figure 16A). Fluorescence image quantification showed that chloroplast fluorescence intensity in p[$^{AMV}_{ATG}$ELVd-derived]YFP images were approximately seven fold more intense than the observed with p[^{AMV}ELVd-derived]YFP (Figure 16B). A Western blot analysis confirmed the presence of the two above described different YFP-related sizes in both clone expressions (Figure 16C). However, the major band in p[$^{AMV}_{ATG}$ELVd-derived]YFP expression was clearly that with higher molecular weight which translation is most likely facilitated by the additional initiation codon. Once again, chloroplast fluorescence intensity and Western blot YFP-related upper band detection and intensity were connected.

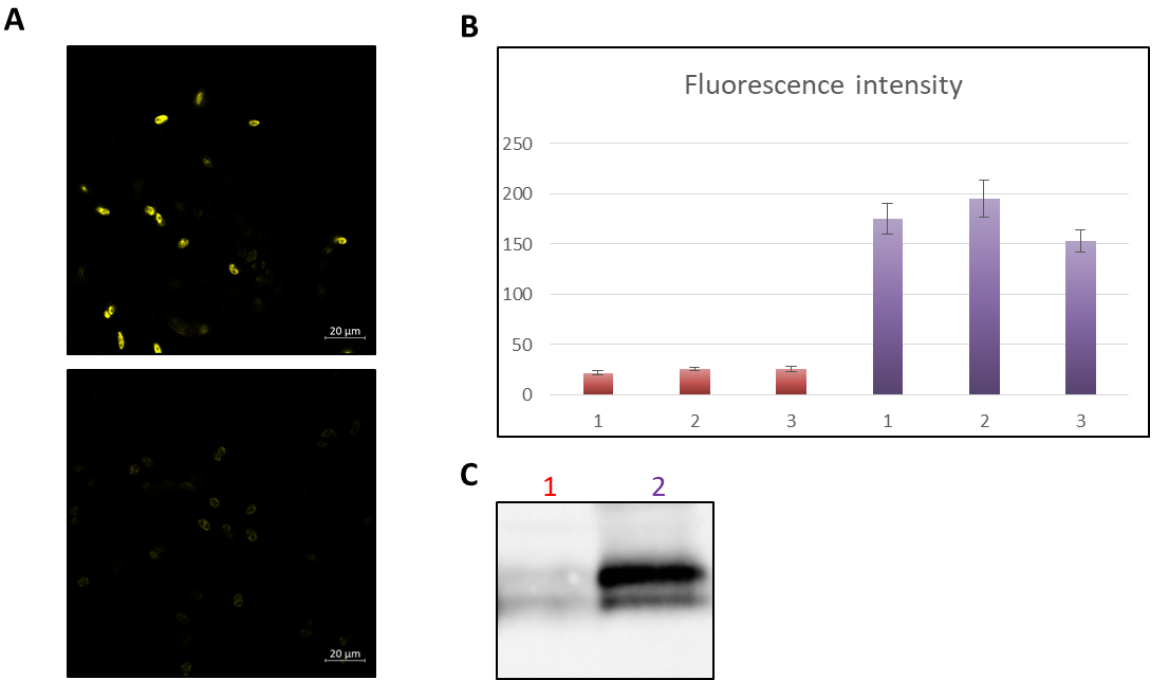


Figure 16. A) Laser scanning confocal microscopy images, acquired by setting the microscope at identical gain and laser intensity condition, of *N. bethamiana* leaf tissue agroinfiltrated with bacteria that carry p[$^{AMV}_{ATG}$ELVd-derived]YFP (upper panel) and p[^{AMV}ELVd-derived]YFP (bottom panel). **B)** Relative fluorescence intensity comparison between *N. benthamiana* leafs agro-infiltrated with bacteria that carry p[^{AMV}ELVd-derived]YFP construct (left bars) and p[$^{AMV}_{ATG}$ELVd-derived]YFP construct **C)** Western blot analysis of extracted proteins from leaf tissue agroinfiltrated with bacteria that carry p[^{AMV}ELVd-derived]YFP (1) and p[$^{AMV}_{ATG}$ELVd-derived]YFP (2). YFP was detected with an antibody against its N terminal region.

Additionally, the YFP of both p[^{AMV}ELVd-derived]$_{FS}$YFP and p[ELVd-derived]YFP transient expressions had nucleus-cytoplasm distribution (Figure 17). This indirect evidence strongly suggests that ELVd-derived peptide is being translated and that AMV 5' UTR is essential to facilitate the initiation at non-canonical codons. From these results, we hypothesised that, by

random, ELVd-derived peptide fused to the YFP could be acting as a chloroplast transit peptide (cTP) responsible for YFP accumulation in the chloroplast as will be further discussed.

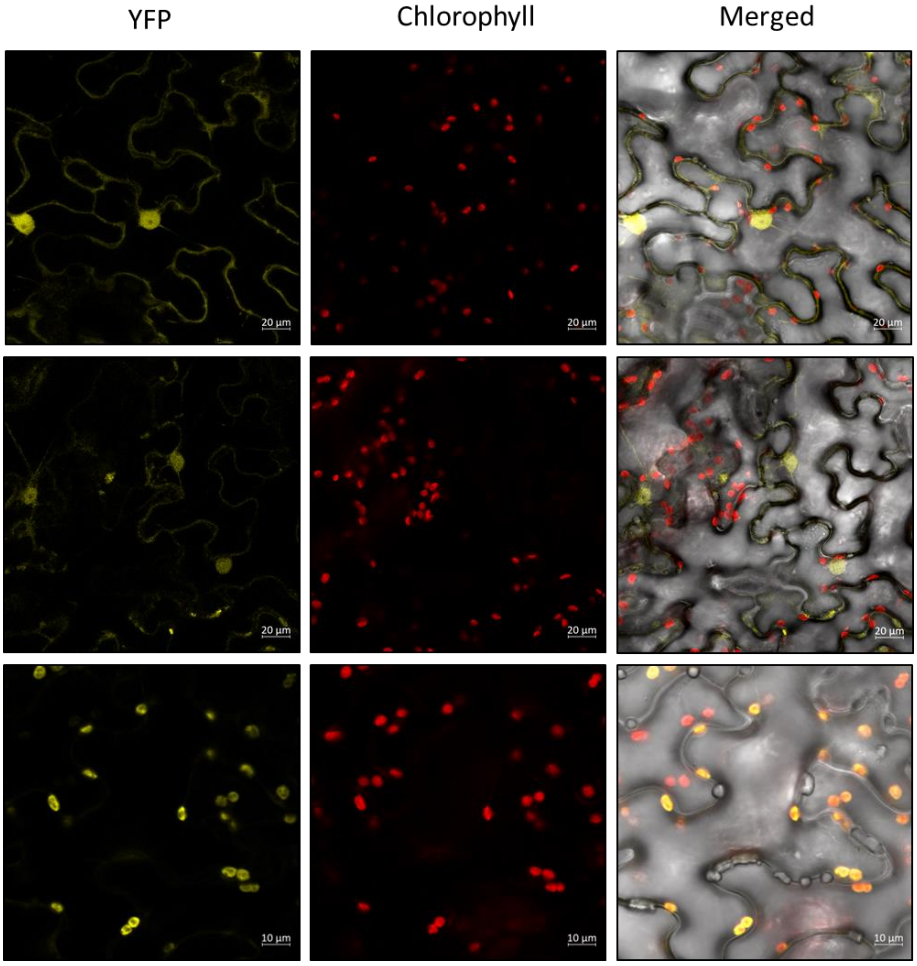


Figure 17. Laser scanning confocal microscopy images of *N. benthamiana* leaf tissue agroinfiltrated with bacteria that carry p[ELVd-derived]YFP (upper panels), p[^{AMV}ELVd-derived]$_{FS}$YFP (middle panels) and p[^{AMV}ELVd-derived]YFP (bottom panels). Chlorophyll fluorescence was used as chloroplast marker.

4.6 Analysis of two 5' UTR chloroplastic translation enhancers effect in the translation of [ELVd-derived]YFP RNA

According to above findings, it was extremely unlikely that the YFP was being translated in the chloroplast from [ELVd-derived]YFP RNA. Independently of this, YFP RNA was trafficked to these organelles delivered by ELVd-derived sequence although it fails in promoting translation. Additional sequences that favour ribosome recruitment could be necessary for YFP translation. Biosynthesis of chloroplastic proteins occurs in bacterial-type 70S ribosome and largely depends on the structure and sequence of the 5' UTR (Bock, 2014). Bacterial-type ribosome-binding sites, also referred to as Shine-Dalgarno (SD) sequences, are present in most of the plastid mRNAs and are related with the highest translation rates (Hirose and Sugiura, 2004). Ribosome recruitment is

favoured by SD sequence because it exhibits complementarity to the 3' end of the 16S ribosomal RNA.

Considering above particularities of chloroplast translation, two specific enhancers were selected. The 5' UTR of RuBisCO large subunit gene (rbcL) was chosen because it codifies for the most expressed protein in the chloroplast (Patel and Berry, 2008), and consequently it seems a good candidate for improving the expression of our desired protein in that organelle. However, 5' UTR of viral origin are even stronger than rbcL 5' UTR, and according to the bibliography, the strongest is the 5' UTR of T7 phage gene 10 (T7-10) (Oey et al., 2009). T7-10 5' UTR contains a perfect SD sequence and also drives extraordinarily high rates of mRNA translation in *Escherichia coli* (Kuroda, 2001). These sequences (Table 2) were cloned downstream the ELVd-derived sequence and upstream YFP, resulting in p[^{AMV}ELVd-derived]$_{T7\text{-}10}$YFP and p[^{AMV}ELVd-derived]$_{rbcL}$YFP constructs (Figure 18A).

Table 2 Sequences of the different 5' UTR analysed in the present section

Name	Length (nt)	Sequence (5' to 3')
rbcL 5' UTR	59	TGTCGAGTAGACCTTGTTGTTGTGAGAATTCTTAATTCATGAGTTGTAGGGAGGGATTTATG
T7-10 5'UTR	63	GGGAGACCACAACGGTTTCCCACTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATG

LSCM images revealed that YFP expression levels with p[^{AMV}ELVd-derived]$_{T7\text{-}10}$YFP were higher than with p[^{AMV}ELVd-derived]$_{rbcL}$YFP (Figure 18 A). The YFP expressed from the RNA carrying the rbcL 5' UTR was not accumulated in the chloroplast and resembled a nucleus-cytoplasmic pattern (Figure 18B, upper panels). In the case of the p[^{AMV}ELVd-derived]$_{T7\text{-}10}$YFP construct, the YFP expression was neither chloroplastic nor nucleus-cytoplasmic, instead, it was localised in small bodies randomly distributed throughout the cytoplasm (Figure 18B, middle panels). Similar bodies were occasionally observed in the case of with p[^{AMV}ELVd-derived]$_{rbcL}$YFP (data not shown).

Therefore, the addition of those 5' UTR apparently prevents RNA chloroplast transport, and in the case of T7-10 5' UTR caused an abnormal protein distribution throughout the cell. To understand what was happening, p[^{AMV}ELVd-derived]$_{T7\text{-}10}$YFP clone was co-expressed in presence of cell stress markers which may explain protein aggregation in that kind of spots. The autophagy-related protein 8 (ATG8) fused upstream of the cherry fluorescent protein (ChFP) was used as indicator of lithic vacuoles, and the eukaryotic translation initiation factor 4E (eIF4E) fused downstream of ChFP as granule stress marker. Neither of cell markers co-localized with YFP spots (Figure 19), therefore that YFP pattern might not be the consequence of an intracellular defensive mechanism against stress caused by high protein levels.

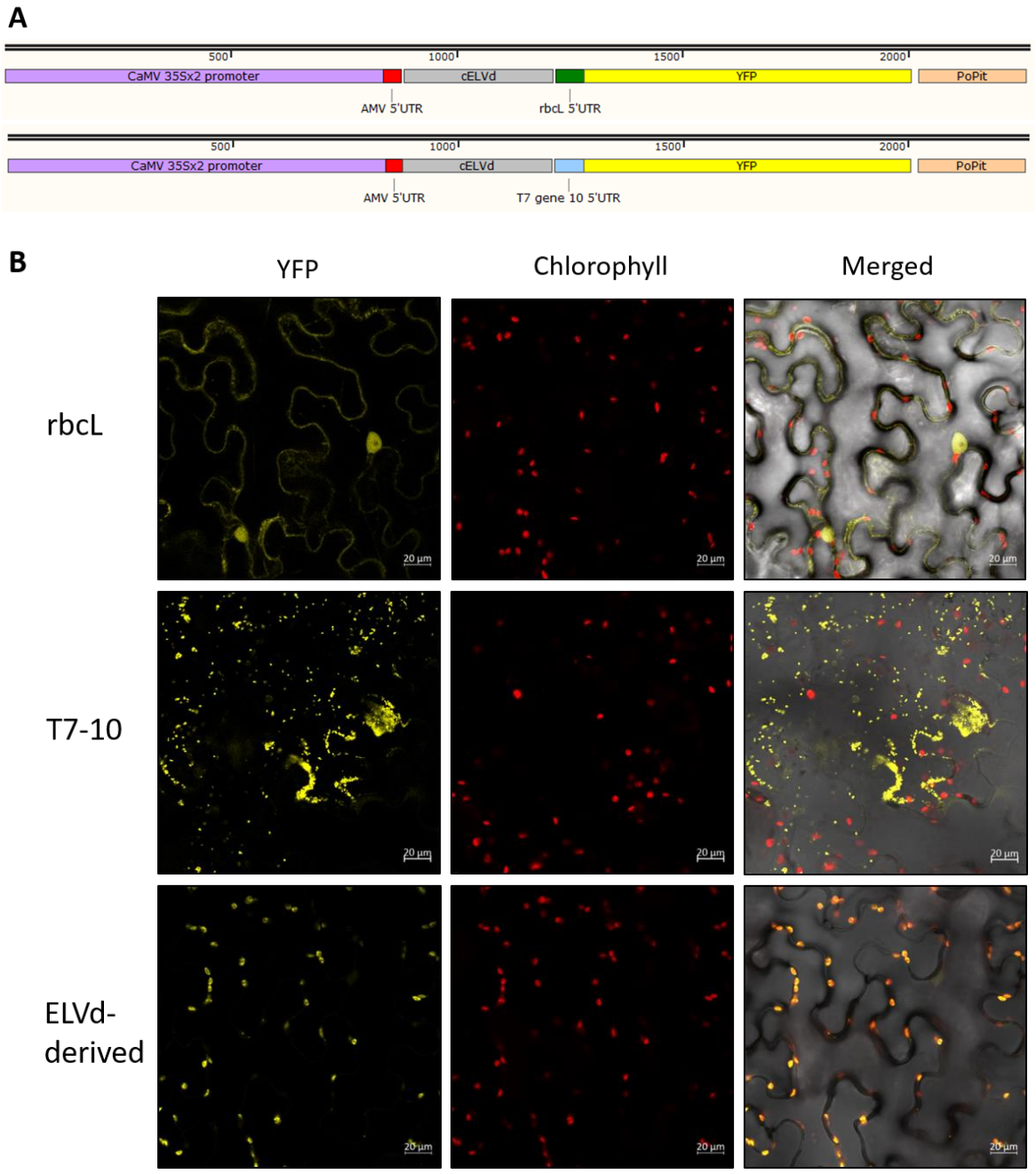


Figure 18. A) Schematic representation of the YFP expression cassette of p[^{AMV}ELVd-derived]$_{rbcL}$YFP construct (up) and p[^{AMV}ELVd-derived]$_{T7\text{-}10}$YFP construct (down). **B)** Laser scanning confocal microscopy of infiltrated tissue with bacteria carrying the p[^{AMV}ELVd-derived]$_{rbcL}$YFP construct (up) and p[^{AMV}ELVd-derived]$_{T7\text{-}10}$YFP construct (down). Chlorophyll fluorescence was used as chloroplast marker.

Finally, *A. tumefaciens* cells carrying those constructs were visualized by LSCM (Figure 20). No YFP fluorescence was detected in bacteria transformed with p[^{AMV}ELVd-derived]YFP. In contrast, bacterial cells transformed with any of the construct that carry a prokaryotic 5' UTR translation enhancer showed YFP fluorescence, more intense in the case of T7-10 5' UTR. This observation demonstrated that prokaryotic 5' UTR translation enhancers caused early expression of the YFP in bacteria. Therefore, fast accumulation of YFP could be related with the avoidance of T-DNA transfer and its proper expression.

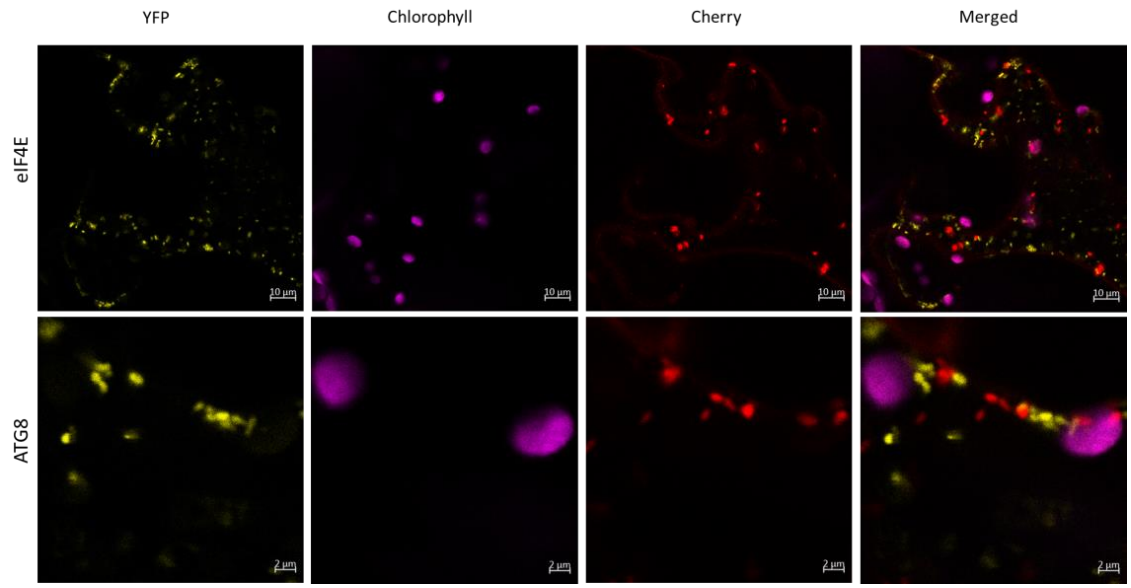


Figure 19. Laser scanning confocal microscopy images of co-agroinfiltrated *N. benthamiana* leaf tissue with bacteria carrying p[^{AMV}ELVd-derived]$_{T7\text{-}10}$YFP and a cell marker fused to the ChFP: autophagy-related protein 8 (ATG8) and eukaryotic translation initiation factor 4E (eIF4E). Chlorophyll fluorescence was used as chloroplast marker.

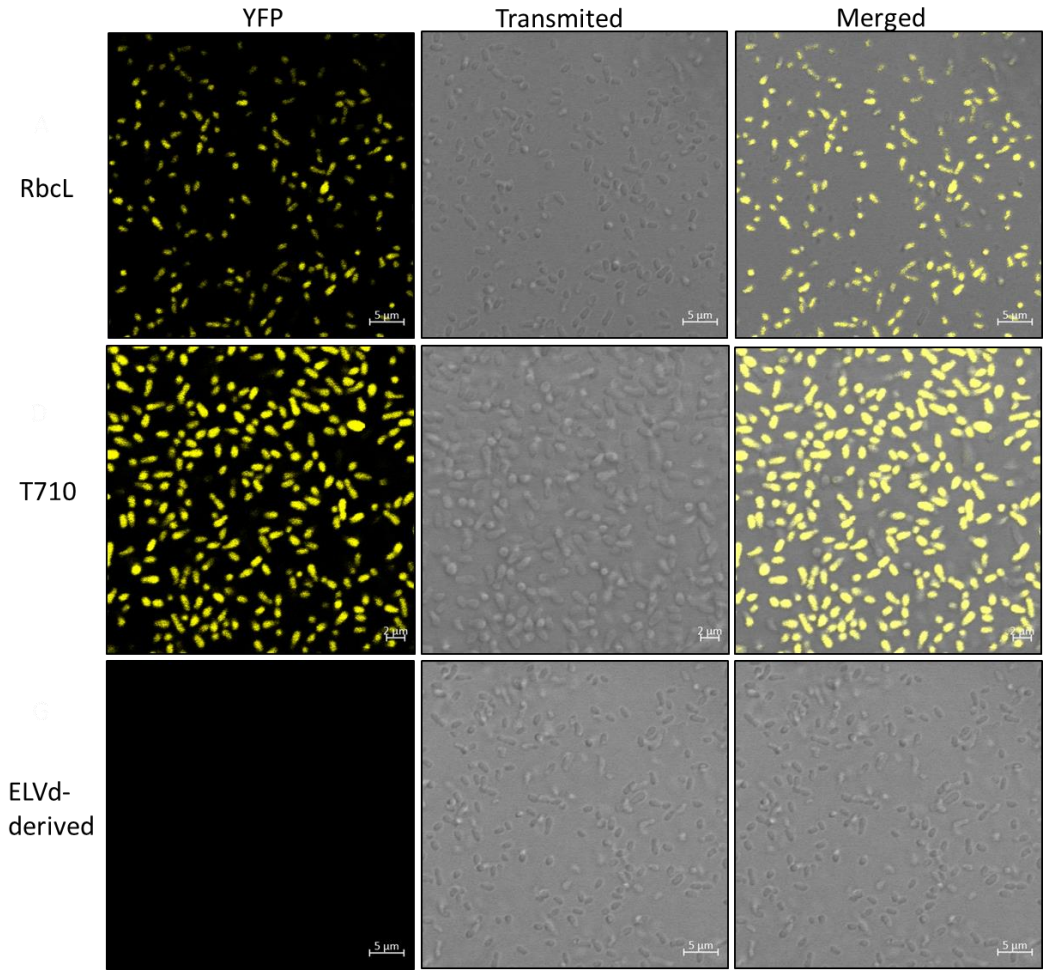


Figure 20. Laser scanning confocal microscopy images of *A. tumefaciens* cells transformed with p[^{AMV}ELVd-derived]$_{rbcL}$YFP (upper panels), p[^{AMV}ELVd-derived]$_{T7\text{-}10}$YFP (middle panels) and p[^{AMV}ELVd-derived]YFP (bottom panels). Laser intensity in p[^{AMV}ELVd-derived]$_{rbcL}$YFP images was set to twice the p[^{AMV}ELVd-derived]$_{T7\text{-}10}$YFP images.

DISCUSSION

5. DISCUSSION

5.1 Ω-TMV leader caused the highest increase in YFP levels.

Gómez and Pallás previously demonstrated, by LCSM and RT-PCR, that a ELVd modified sequence fused as a 5´ UTR in front of the GFP gene allowed, not only the import of the GFP RNA, but also a high degree of GFP expression in chloroplasts. This observation opened up new possibilities for engineering a protein expression system in chloroplasts based on specific RNA traffic. In the present study, we tried to improve this system by increasing protein expression levels and reducing biosafety concerns about the potential presence of plant pathogen sequences. We first carried out a screening for the effect of different 5' UTR leader sequences in protein expression. 5' UTR are known to influence gene expression, particularly translation rate. Hence, we compared the two most widely used 5' UTR for enhancing expression in dicotyledonous plants: Ω-TMV and AMV leaders. A third leader sequence of synthetic origin (synJ), which was described as equivalent or stronger than the formers, was also evaluated (Kanoria and Burma, 2012).

The Ω-TMV leader (77 nt) sequence is highly organised and it has been shown that can interact with several proteins (Gallie and Walbot, 1992). These interactions are related to primary sequence as Ω-TMV leader is predicted to have little secondary structure. Two repeated motifs (three copies of an eight base direct repeat and a poly(CAA) region) comprise 72 % of Ω-TMV sequence and are functionally redundant for enhancing protein translation. More specifically, it has been demonstrated that the HSP101 binds to this poly(CAA) region and mediates the binding of translation initiation factors eIF3 and eIF4G (Gallie and Walbot, 1992). The AMV leader sequence is shorter (45 nt) than Ω-TMV but it is also reported to provide a more efficient access to the translational apparatus (Datla et al., 1993). According to Kanoria and Burma, synJ (28 nt) enhances gene expression strictly at posttranscriptional level by providing a suitable translation initiation context (AAACC immediately upstream of AUG) as it was not observed any protein binding to this sequence (Kanoria and Burma, 2012).

Our results showed that Ω-TMV leader was the best 5' UTR for enhancing YFP levels followed by the AMV leader, and both were significantly more efficient than synJ. Kanoria and Burma measured the efficiency of synJ 5' UTR in tobaco transgenic lines by quantifying β-glucuronidase (GUS) activity. Their results of GUS activity in leaf tissue were not in agreement with our findings. In their study, they observed that AMV leader enhanced the maximum GUS activity and that obtained with synJ was sligthly superior to Ω-TMV leader. Those discrepant results may be influenced by the fact that their studies were performed in transgenic lines and ours were

mediated by transient expression. However, that argument seems clearly insufficient to explain the poor expression obtained with synJ in comparison with the viral leader sequences. In our construct, the synJ sequence was not probably situated immediately upstream of the starting codon as it was cloned between the 35S promoter and the ELVd-derived sequence. This could be a reason for low synJ-driven YFP expression considering that ELVd-derived peptide translation start has not been determined yet, as will be discussed later. In any case, the most plausible explanation should be related to the differences observed at transcriptional level since Northern blot results showed a positive relation between the amount of transcripts and the YFP fluorescence levels.

According to the literature, 5' UTR influence is at translational level, and therefore that differences could be related to the experimental variability of the transient expression technique. Although the main contribution of the viral leader sequences to boost gene expression is at translational level, they may also provide a better transcription context than a short synthetic sequence as synJ. In conclusion, Ω-TMV and AMV leaders seem to favour the translation start in canonical and, as we will discuss later, in non-canonical starting codons, being the Ω-TMV leader more efficient than the AMV leader. SynJ particularly low efficiency could be related with its translation enhancement mechanism and shorter length. Additionally, our results suggest that molecular mechanism that governs 5' UTR translation enhancement could be more complicated than expected.

5.2 RNA traffic into the chloroplast is not related to specific fluorescent protein accumulation in that organelle: chloroplast transit peptide hypothesis.

Plant infection capacity of the ELVd-derived RNA in our system was most likely abolished since viroidal region including the plus polarity hammerhead ribozyme, which is essential for replication, was arranged as a non-self-cleaving reverse-complement sequence. In spite of this and to maximally reduce biosafety concerns, we decided to eliminate, as much as possible, the viroidal sequences from the expression cassette. From previous published results, it was evident that central region of the ELVd-derived RNA was essential for RNA chloroplast targeting but regions at both sides of this region favourably modulated the transport (Gómez and Pallás, 2010b). Moreover, ELVd-derived RNA was predicted to fold into a significant secondary structure consisting of two loops with several radial stem-loops of different length. One of these loops corresponded to the ELVd-derived central region and was thought to be involved in RNA chloroplast transport. For that reason, we engineered two shorter versions of the ELVd-derived sequence by deleting sequences from both 5' and 3' ends but always keeping the central loop

structure. As an alternative to reduce pathogen-related sequences, ELVd-derived was turned into CTVL by completely preserving ELVd-derived conformation. In contrast, the ribonucleotide sequence was changed in such a way that viroid could not be recognised.

According to semi-quantitative RT-PCR analysis performed with RNAs extracted from chloroplast, all three RNAs were imported into this organelle. The chloroplast transport of the RNA carrying the two short ELVd-derived sequences, [^{AMV}ELVd-derived$_{247}$]YFP RNA and [^{AMV}ELVd-derived$_{181}$]YFP RNA, appeared to be as efficient as the entire sequence [^{AMV}ELVd-derived]YFP RNA. In contrast, accumulation of [^{AMV}CTVL]YFP RNA in chloroplasts was significantly lower than the others. Assuming that *RNAfold* prediction was accurate and considering that ELVd-derived and CTVL RNA predicted structures were almost identical, it can be proposed that primary structure or ribonucleotide sequence may also be important for chloroplast traffic. One consideration is that these predictions are based on minimal free energy calculations, but *in vivo* conformations may differ from *in silico* model. Even though, [^{AMV}CTVL]YFP RNA was detected in isolated chloroplast therefore we could not exclude that this RNA is truly trafficked to this organelle.

In spite of the fact that [^{AMV}ELVd-derived$_{247}$]YFP RNA, [^{AMV}ELVd-derived$_{181}$]YFP RNA and [^{AMV}CTVL]YFP RNA were presumably detected in the chloroplast, this observation did not match with their YFP fluorescence pattern (nucleus-cytoplasmic for all of them with occasional chloroplast localisation only in the case of [^{AMV}ELVd-derived$_{247}$]YFP). That discrepancy was an intriguing question, especially because [^{AMV}ELVd-derived$_{247}$]YFP RNA and [^{AMV}ELVd-derived$_{181}$]YFP RNA, both contain the 110 nt region previously identified as sufficient to mediate the corresponding mRNA traffic and fluorescent protein accumulation in chloroplasts (Gómez and Pallás, 2010b). The presence of two different YFP-related bands in Western blot analysis provided a plausible explanation. Besides the band corresponding to the free YFP (27 KDa), a bigger band that could be the result of a fusion between the YFP and an ELVd-derived peptide was detected. These two bands were not observed in the Western blot carried out by Gómez and Pallás (Figure 1, Gómez and Pallás 2010b). This discrepancy with our findings cannot be explained because their construct carried GFP instead of YFP as both proteins have approximately the same size. Probably, a single band was detected in their analysis because they used a SDS-PAGE that had a lower resolution than that used in the present work (10 % vs 12 % acrylamide, respectively).

Translation of a peptide from the ELVd-derived RNA was unexpected because it does not contain any AUG *in frame* with the YFP ORF. Nevertheless, translation could have started in a non-canonical codon, a phenomenon already described in plant cells (Gordon *et al.*, 1992), and even with biological significance (Simpson *et al.*, 2010). Translation initiation efficiency at non-AUG

triplets is significantly lower than at AUG, but triplets with one base difference from AUG are considerably more efficient than the rest (2-30 % of AUG activity). Studies on transient expression in protoplasts (Gordon *et al.*, 1992) and transgenic plants (Depeiges *et al.*, 2006) offered distinct percentages of AUG activities, but agreed that CUG is the most active followed by ACG and GUG, and the least active is AUC. In this sense, a GUG codon *in frame* with the YFP ORF was early found approximately 40 nucleotides downstream of the AMV 5' UTR end (S5). Significant increase in the chloroplastic YFP levels caused by AUG addition was in agreement with a putative non-AUG translation start in [^{AMV}ELVd-derived] RNA because the appearance of the YFP-related upper band in Western blot analysis was extremely evident. Moreover, translational frameshifting in [^{AMV}ELVd-derived]YFP RNA resulted in very low levels of nucleus-cytoplasmic YFP. This observation strongly suggested that ELVd-derived sequence was actually being translated from [^{AMV}ELVd-derived]YFP RNA and, most likely, it was involved in chloroplast localisation of the YFP. Finally and according to our results with [ELVd-derived]YFP RNA, translation start in a non-canonical codon appeared to be extremely dependent on the presence of a translation enhancer, like AMV 5'UTR, just in front of the ELVd-derived sequence.

A Shine-Dalgarno like sequence was observed in the middle region of ELVd-derived sequence (GGAGGATTCG). It was assumed that this sequence in combination with secondary or/and tertiary structure of the central region could be playing a role in protein translation in chloroplasts. In the light of above findings, we proposed that ELVd-derived sequence is being translated and that raises the question of whether translation occurs in the chloroplast or in the cytoplasm acting, just by chance, the ELVd-derived peptide as a targeting signal of protein import to the organelle (chloroplast transit peptide, ChTP). In this scenario, our results suggested that ELVd-derived peptide tag was translated in the cytoplasm because its translation impairment, either by frameshift introduction or by removing translation enhancer, caused nucleus-cytoplasmic YFP distribution. These changes with translational effect should alter neither mRNA secondary structure nor RNA chloroplast transport.

Chloroplast transit peptides are usually located at the N-terminal part of the protein. They have from 13 to more than 100 amino acids in length and generally, exhibit high abundance of hydroxylated (Ser and Thr), hydrophobic (Ala, Val, Ile, Leu, Met, Trp, Phe and Tyr) and positively charged (Arg and Lys) amino acids but few acidic residues (Asp and Glu) (Zhang and Glaser, 2002). That information is in high agreement with the putative length (90 aa) and composition of the ELVd-derived peptide (S6). Canonical cTPs are cleaved in the traffic to the chloroplast stroma mediated by TOC-TIC translocon machinery (Inaba and Schnell, 2008). However, several proteins are targeted to the chloroplast through internal, non-cleavable targeting signal. One possibility is

that they use the same machinery as canonical cTP but lost the motif for stromal processing of the peptide. According to our current findings, it seems that ELVd-derived peptide is cleaved and therefore, the increase of the YFP size is due to a part of the ELVd-derived peptide that remains after the processing. That assumption comes from the fact that the estimated molecular weight of the ELVd-derived-YFP protein, starting at GUG codon, was 37.7 kDa. However, the upper YFP-related band in Western blot analysis migrated faster than the 35 kDa marker.

There is still one thing that remains to be discussed about [^{AMV}ELVd-derived$_{247}$]YFP and [^{AMV}ELVd-derived$_{181}$]YFP RNAs. As already mentioned, the expression of both RNAs resulted in a nucleus-cytoplasmic distribution of the YFP with very occasionally chloroplast localisation in the former. However, both RNAs had a 5' end deletion that results in the elimination of the GUG non-canonical start codon present in [^{AMV}ELVd-derived]YFP RNA (S5). In spite of this, a weaker AUC start codon, 50 ribonucleotides downstream of the AMV 5' UTR end, could promote the translation of two shorter versions of the putative cTP from both RNAs. It has been proposed that a cTP is an assembly of motifs for interacting with selected translocon components (Li and Teng, 2013). It is possible that [ELVd-derived$_{247}$] peptide still have some of these motifs but not [ELVd-derived$_{181}$] peptide, which has an additional deletion in its 3' end. Extremely low levels of expression from the AUC codon compared with canonical AUG of the YFP ORF would explain the very occasional localisation of the fluorescence in chloroplast when this construct was expressed.

5.3 5' UTR chloroplastic translation enhancers were not suitable for *Agrobacterium*-mediated transient expression in plants.

The results above described clearly indicate that YFP fluorescence observation was not a good marker for studying either RNA traffic or protein expression in chloroplast. In prokaryotes, translation is supposed to be assisted by mRNA-rRNA interactions between the Shine-Dalgarno sequence, which is located upstream of the translation initiation codon, and the anti-Shine-Dalgarno sequence of the 16S ribosomal RNA. Over 90 % of chloroplast genes in land plants have an upstream sequence similar to the bacterial SD sequence (usually GGAGG) that is capable of binding to a complementary sequence close to the 3´end of the chloroplast 16S rRNA. To circumvent the inconvenient arisen from our last findings, we added well-known chloroplast specific translation enhancers, immediately upstream of the YFP ORF, as a way to achieve YFP translation in this organelle.

However, plant expression of YFP from binary plasmids carrying either the T7 gene 10 leader sequence or the Rubisco Large Subunit (rbcL) 5' leader region including its SD-like sequence

(GGGAGGG) was impaired by early YFP expression in *A. tumefaciens* cells. CaMV 35S promoter was considered to be plant specific and not active in other organisms such as bacteria, fungi or human cells. This assumption was demonstrated to be erroneous as it has been established that CaMV 35S promoter is not only active in plants but also in *E.coli*, in soil bacteria *Agrobacterium rhizogenes*, yeast, in extracts of human cancer cell lines and, finally, in *A. tumefaciens* (Yu *et al.*, 1990; Franzetti *et al.*, 1992; Pobjecky *et al.*, 1990)

Early translation in bacteria could happen because the T7 gene 10 leader is a ribosome-binding site from a bacteriophage whereas the rbcL leader region was directly acquired from a chloroplast gene and both have strong homologies to prokaryote enhancers. Consequently, it is not surprising that they were active in a prokaryotic environment. In fact, Ahmad *et al* also observed that these sequences caused a significant translation enhancement when studying the influence of several 5' UTR on *A. tumefaciens* translation (Ahmad *et al.*, 2014). We hypothesised that high expression of the binary plasmid in bacteria could hinder T-DNA transference to the plant cell by interfering with the bacterial translation rate or function of *vir* proteins.

Those prokaryote-like translation enhancers seem not suitable to be used in bacterial-mediated transient expression, but plant stable transformation could be a solution to this problem. Developing transgenic plant with this type of constructs could serve to test RNA translation in the chloroplast. Nuclear transformation is time-consuming, but it is a more standardized procedure and requires less time than having a homo-plastid configuration. Considering the many advantages of chloroplast translation for achieving high levels of heterologous protein, ELVd-derived sequence could still be useful, although further research is needed.

CONCLUDING REMARKS

6. CONCLUDING REMARKS

In conclusion, we studied the influence of 5' UTR leader sequences on agro-bacterium mediated transient expression of YFP. We found that Ω-TMV leader leads the highest protein accumulation. By using deletions and a secondary structure analogous version of the ELVd-derived RNA we showed that secondary structure but also ribonucleotide sequence of the central region were relevant in chloroplast RNA transport. In this study, it was also demonstrated that YFP accumulation in the chloroplast was not related to the presence of the corresponding RNA in this organelle. Hence, our results suggested that YFP was being translated from a non-canonical start codon located in the upstream ELVd-derived sequence. That phenomenon resulted in an YFP N terminal fusion of an ELVd-derived peptide that showed the typical features of a chloroplast transit peptide. Therefore, YFP was not being translated in the chloroplast; instead, the ELVd-derived-YFP fusion protein was trafficked to this organelle. Finally, an attempt to actually achieve chloroplast translation driven by the ELVd-derived RNA was performed by adding specific translation enhancers downstream of the ELVd-derived sequence and immediately upstream of the YFP ORF. However, translation in chloroplast was not achieved presumably because of the early expression of the YFP in *A. tumefaciens* cells.

REFERENCES

7. REFERENCES

Aebi, M. (2013) N-linked protein glycosylation in the ER. *Biochim. Biophys. Acta - Mol. Cell Res.*, **1833**, 2430–2437.

Ahmad, N., Michoux, F., Lössl, A.G. and Nixon, P.J. (2016) Challenges and perspectives in commercializing plastid transformation technology. *J. Exp. Bot.*, **67**, 5945–5960.

Ahmad, T., Venkataraman, S., Hefferon, K. and Abouhaidar, M.G. (2014) Viral and chloroplastic signals essential for initiation and efficiency of translation in Agrobacterium tumefaciens. *Biochem. Biophys. Res. Commun.*, **452**, 14–20.

Alvarez, M.L., Topal, E., Martin, F. and Cardineau, G.A. (2010) Higher accumulation of F1-V fusion recombinant protein in plants after induction of protein body formation. *Plant Mol. Biol.*, **72**, 75–89.

Aoyama, T. and Chua, N.H. (1997) A glucocorticoid-mediated transcriptional induction system in transgenic plants. *Plant J.*, **11**, 605–612.

Avesani, L., Marconi, G., Morandini, F., Albertini, E., Bruschetta, M., Bortesi, L., Pezzotti, M. and Porceddu, A. (2007) Stability of Potato virus X expression vectors is related to insert size: Implications for replication models and risk assessment. *Transgenic Res.*, **16**, 587–597.

Bock, R. (2014) Engineering chloroplasts for high-level foreign protein expression. *Methods Mol. Biol.*, **1132**, 93–106.

Bock, R. (2007) Plastid biotechnology: prospects for herbicide and insect resistance, metabolic engineering and molecular farming. *Curr. Opin. Biotechnol.*, **18**, 100–106.

Caddick, M.X., Greenland, A.J., Jepson, L., et al. (1998) An ethanol inducible gene switch for plants used to manipulate carbon metabolism. *Nat. Biotechnol.*, **16**, 177–180.

Chan, H.T. and Daniell, H. (2015) Plant-made oral vaccines against human infectious diseases-Are we there yet? *Plant Biotechnol. J.*, **13**, 1056–1070.

Chilton, M.-D., Drummond, M.H., Merlo, D.J., Sciaky, D., Montoya, A.L., Gordon, M.P. and Nester, E.W. (1977) Stable incorporation of plasmid DNA into higher plant cells: the molecular basis of crown gall tumorigenesis. *Cell*, **11**, 263–271.

Daniell, H., Chan, H.-T. and Pasoreck, E.K. (2016) Vaccination via Chloroplast Genetics: Affordable Protein Drugs for the Prevention and Treatment of Inherited or Infectious Human Diseases. *Annu. Rev. Genet.*, **50**, 595–618.

Daniell, H., Singh, N.D., Mason, H. and Streatfield, S.J. (2009) Plant-made vaccine antigens and biopharmaceuticals. *Trends Plant Sci.*, **14**, 669–679.

Darbani, B., Farajma, S., Toorchi, M., Zkerbastanabad, S., Noeparvar, S. and Stewart, N. (2008) DNA delivery methods to produce transgenic plants. *Biotechnology*, **7**, 385–402.

Daròs, J.A. (2016) Eggplant latent viroid: a friendly experimental system in the family Avsunviroidae. *Mol. Plant Pathol.*, **17**, 1170–1177.

Datla, R.S.S., Bekkaoui, F., Hammerlindl, J.K., Pilate, G., Dunstan, D.I. and Crosby, W.L. (1993) Improved high-level constitutive foreign gene expression in plants using an AMV RNA4 untranslated leader sequence. *Plant Sci.*, **94**, 139–149.

Day, A. and Goldschmidt-Clermont, M. (2011) The chloroplast transformation toolbox: Selectable markers and marker removal. *Plant Biotechnol. J.*, **9**, 540–553.

Depeiges, A., Degroote, F., Espagnol, M.C. and Picard, G. (2006) Translation initiation by non-AUG codons in Arabidopsis thaliana transgenic plants. *Plant Cell Rep.*, **25**, 55–61.

Diener, T.O. (2003) Discovering viroids: a personal perspective. *Nat. Rev. Microbiol.*, **1**, 75–80.

Diener, T.O. (1971) Potato spindle tuber "virus". IV. A replicating, low molecular weight RNA. *Virology*, **45**, 411–428.

Doran, P.M. (2006) Foreign protein degradation and instability in plants and plant tissue cultures. *Trends Biotechnol.*, **24**, 426–432.

Dugdale, B., Mortimer, C.L., Kato, M., James, T.A., Harding, R.M. and Dale, J.L. (2013) In Plant Activation: An Inducible, Hyperexpression Platform for Recombinant Protein Production in Plants. *Plant Cell*, **25**, 2429–2443.

Elena, S.F., Gómez, G. and Daròs, J.A. (2009) Evolutionary constraints to viroid evolution. *Viruses*, **1**, 241–254.

Fan, Q., Treder, K. and Miller, W.A. (2012) Untranslated regions of diverse plant viral RNAs vary greatly in translation enhancement efficiency. *BMC Biotechnol.*, **12**, 22.

Flores, R., Hernández, C., Alba, A.E.M. de, Daròs, J.-A. and Serio, F. Di (2005) Viroids and Viroid-Host Interactions. *Annu. Rev. Phytopathol.*, **43**, 117–139.

Flores, R., Minoia, S., Carbonell, A., Gisel, A., Delgado, S., López-Carrasco, A., Navarro, B. and Serio, F. Di (2015) Viroids, the simplest RNA replicons: How they manipulate their hosts for being propagated and how their hosts react for containing the infection. *Virus Res.*, **209**, 136–145.

Franzetti, B., Carol, P. and Mache, R. (1992) Characterization and RNA-binding properties of a chloroplast S1-like ribosomal protein. *J. Biol. Chem.*, **267**, 19075–19081.

Gallie, D.R., Sleat, D.E., Watts, J.W., Turner, P.C. and Wilson, T. michael A. (1988) Mutational analysis of the tobacco mosaic virus 5′-leader for altered ability to enhance translation. *Nucleic Acids Res.*, **16**, 883–893.

Gallie, D.R. and Walbot, V. (1992) Identification of the motifs within the tobacco mosaic virus 5′-leader responsible for enhancing translation. *Nucleic Acids Res.*, **20**, 4631–4638.

Genovés, A., Navarro, J.A. and Pallás, V. (2006) Functional analysis of the five melon necrotic spot virus genome-encoded proteins. *J. Gen. Virol.*, **87**, 2371–2380.

Giguère, T., Adkar-Purushothama, C.R., Bolduc, F. and Perreault, J.P. (2014) Elucidation of the structures of all members of the Avsunviroidae family. *Mol. Plant Pathol.*, **15**, 767–779.

Gómez, G. and Pallás, V. (2012a) A pathogenic non coding RNA that replicates and accumulates in chloroplasts traffics to this organelle through a nuclear-dependent step. *Plant Signal. Behav.*, **7**, 881–883.

Gómez, G. and Pallás, V. (2010a) Can the import of mRNA into chloroplasts be mediated by a secondary structure of a small non-coding RNA? *Plant Signal. Behav.*, **5**, 1517–9.

Gómez, G. and Pallás, V. (2010b) Noncoding RNA mediated traffic of foreign mRNA into chloroplasts reveals a novel signaling mechanism in plants. *PLoS One*, **5**, 1517–9.

Gómez, G. and Pallás, V. (2012b) Studies on subcellular compartmentalization of plant pathogenic noncoding RNAs give new insights into the intracellular RNA-traffic mechanisms. *Plant Physiol.*, **159**, 558–564.

Gordon, K., Fütterer, J. and Hohr, T. (1992) Efficient initiation of translation at non-AUG triplets in plant cells. *Plant J.*, **2**, 809–813.

Habibi, P., Prado, G.S., Pelegrini, P.B., Hefferon, K.L., Soccol, C.R. and Grossi-de-Sa, M.F. (2017) Optimization of inside and outside factors to improve recombinant protein yield in plant. *Plant Cell. Tissue Organ Cult.*, **130**, 449–467.

Hirose, T. and Sugiura, M. (2004) Functional Shine-Dalgarno-Like Sequences for Translational Initiation of Chloroplast mRNAs. *Plant Cell Physiol.*, **45**, 114–117.

Horvath, H., Huang, J., Wong, O., Kohl, E., Okita, T., Kannangara, C.G. and Wettstein, D. von (2000) The production of recombinant proteins in transgenic barley grains. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 1914–1919.

Inaba, T. and Schnell, D.J. (2008) Protein trafficking to plastids: one theme, many variations. *Biochem. J.*, **413**, 15–28.

Jones, H.D. and Sparks, C.A. (2009) Stable transformation of plants. *Methods Mol. Biol.*, **513**, 111–130.

Kanoria, S. and Burma, P.K. (2012) A 28 nt long synthetic 5′UTR (synJ) as an enhancer of transgene expression in dicotyledonous plants. *BMC Biotechnol.*, **12**, 85.

Kay, R., Chan, A., Daly, M. and McPherson, J. (1987) Duplication of CaMV 35S Promoter Sequences Creates a Strong Enhancer for Plant Genes. *Science (80-.).*, **236**, 1299–1302.

Khan, I., Twyman, R.M., Arcalis, E. and Stoger, E. (2012) Using storage organelles for the accumulation and encapsulation of recombinant proteins. *Biotechnol. J.*, **7**, 1099–1108.

Kloc, M., Zearfoss, N.R. and Etkin, L.D. (2002) Mechanisms of subcellular mRNA localization. *Cell*, **108**, 533–544.

Krenek, P., Samajova, O., Luptovciak, I., Doskocilova, A., Komis, G. and Samaj, J. (2015) Transient plant transformation mediated by Agrobacterium tumefaciens: Principles, methods and applications. *Biotechnol. Adv.*, **33**, 1024–1042.

Kuroda, H. (2001) Complementarity of the 16S rRNA penultimate stem with sequences downstream of the AUG destabilizes the plastid mRNAs. *Nucleic Acids Res.*, **29**, 970–975.

Lacroix, B. and Citovsky, V. (2013) The roles of bacterial and host plant factors in Agrobacterium-mediated genetic transformation. *Int. J. Dev. Biol.*, **57**, 467–481.

Li, H. min and Teng, Y.S. (2013) Transit peptide design and plastid import regulation. *Trends Plant Sci.*, **18**, 360–366.

López-Carrasco, A., Gago-Zachert, S., Mileti, G., Minoia, S., Flores, R. and Delgado, S. (2016) The transcription initiation sites of eggplant latent viroid strands map within distinct motifs in their in vivo RNA conformations. *RNA Biol.*, **13**, 83–97.

Lössl, A.G. and Waheed, M.T. (2011) Chloroplast-derived vaccines against human diseases: Achievements, challenges and scopes. *Plant Biotechnol. J.*, **9**, 527–539.

Michaud, M., Maréchal-Drouard, L. and Duchêne, A.M. (2010) RNA trafficking in plant cells: Targeting of cytosolic mRNAs to the mitochondrial surface. *Plant Mol. Biol.*, **73**, 697–704.

Michelet, L., Lefebvre-Legendre, L., Burr, S.E., Rochaix, J.D. and Goldschmidt-Clermont, M. (2011) Enhanced chloroplast transgene expression in a nuclear mutant of Chlamydomonas. *Plant Biotechnol. J.*, **9**, 565–574.

Mortimer, C.L., Dugdale, B. and Dale, J.L. (2015) Updates in inducible transgene expression using viral vectors: From transient to stable expression. *Curr. Opin. Biotechnol.*, **32**, 85–92.

Nakayama, T. and Archibald, J.M. (2012) Evolving a photosynthetic organelle. *BMC Biol.*, **10**, 35.

Nicolaï, M., Duprat, A., Sormani, R., Rodriguez, C., Roncato, M.A., Rolland, N. and Robaglia, C. (2007) Higher plant chloroplasts import the mRNA coding for the eucaryotic translation initiation factor 4E. *FEBS Lett.*, **581**, 3921–3926.

Oey, M., Lohse, M., Kreikemeyer, B. and Bock, R. (2009) Exhaustion of the chloroplast protein synthesis capacity by massive expression of a highly stable protein antibiotic. *Plant J.*, **57**, 436–445.

Pallás, V., Más, P. and Sánchez-Navarro, J. a (1998) Detection of plant RNA viruses by nonisotopic dot-blot hybridization. *Methods Mol. Biol.*, **81**, 461–468.

Park, K.Y. and Wi, S.J. (2016) Potential of plants to produce recombinant protein products. *J. Plant Biol.*, **59**, 559–568.

Patel, M. and Berry, J.O. (2008) Rubisco gene expression in C4 plants. In *Journal of Experimental Botany*. pp. 1625–1634.

Peyret, H. and Lomonossoff, G.P. (2015) When plant virology met Agrobacterium: The rise of the deconstructed clones. *Plant Biotechnol. J.*, **13**, 1121–1135.

Pobjecky, N., Rosenberg, G.H., Dinter-Gottlieb, G. and Käufer, N.F. (1990) Expression of the β-glucuronidase gene under the control of the CaMV 35S promoter in Schizosaccharomyces pombe. *MGG Mol. Gen. Genet.*, **220**, 314–316.

Pogue, G.P., Lindbo, J.A., Garger, S.J. and Fitzmaurice, W.P. (2002) Making an ally from an enemy: plant virology and the new agriculture. *Annu. Rev. Phytopathol.*, **40**, 45–74.

Raab, R.M., Tyo, K. and Stephanopoulos, G. (2005) Metabolic engineering. *Adv. Biochem. Eng. Biotechnol.*, **100**, 1–17.

Ramírez, N., Lorenzo, D., Palenzuela, D., Herrera, L., Ayala, M., Fuentes, A., Pérez, M., Gavilondo, J. and Oramas, P. (2000) Single-chain antibody fragments specific to the hepatitis B surface antigen, produced in recombinant tobacco cell cultures. *Biotechnol. Lett.*, **22**, 1233–1236.

Ruiz, M.T., Voinnet, O. and Baulcombe, D.C. (1998) Initiation and Maintenance of Virus-Induced Gene Silencing. *Plant Cell*, **10**, 937–946.

Sack, M., Hofbauer, A., Fischer, R. and Stoger, E. (2015) The increasing value of plant-made proteins. *Curr. Opin. Biotechnol.*, **32**, 163–170.

Sanford, J.C., Smith, F.D. and Russell, J.A. (1993) Optimizing The Biolistic Process for Different Biological Applications. *Methods Enzymol.*, **217**, 483–509.

Serio, F. Di, Flores, R., Verhoeven, J.T.J., Li, S.F., Pallás, V., Randles, J.W., Sano, T., Vidalakis, G. and Owens, R.A. (2014) Current status of viroid taxonomy. *Arch. Virol.*, **159**, 3467–3478.

Simpson, G.G., Laurie, R.E., Dijkwel, P.P., Quesada, V., Stockwell, P.A., Dean, C. and Macknight, R.C. (2010) Noncanonical Translation Initiation of the Arabidopsis Flowering Time and Alternative Polyadenylation Regulator FCA. *PLANT CELL ONLINE*, **22**, 3764–3777.

Su, J., Zhu, L., Sherman, A., et al. (2015) Low cost industrial production of coagulation factor IX bioencapsulated in lettuce cells for oral tolerance induction in hemophilia B. *Biomaterials*, **70**, 84–93.

Tak, H., Negi, S., Ganapathi, T.R. and Bapat, V.A. (2016) Molecular farming: Prospects and limitation. In *Banana: Genomics and Transgenic Approaches for Genetic Improvement*. pp. 261–275.

Turnage, M.A., Muangsan, N., Peele, C.G. and Robertson, D. (2002) Geminivirus-based vectors for gene silencing in Arabidopsis. *Plant J.*, **30**, 107–114.

Vamvaka, E., Twyman, R.M., Christou, P. and Capell, T. (2014) Can plant biotechnology help break the HIV-malaria link? *Biotechnol. Adv.*, **32**, 575–582.

Waheed, M.T., Ismail, H., Gottschamel, J., Mirza, B. and Lössl, A.G. (2015) Plastids: The Green Frontiers for Vaccine Production. *Front. Plant Sci.*, **6**, 1005.

Werner, S., Breus, O., Symonenko, Y., Marillonnet, S. and Gleba, Y. (2011) High-level recombinant protein expression in transgenic plants by using a double-inducible viral vector. *Proc. Natl. Acad. Sci.*, **108**, 14061–14066.

Xu, J., Dolan, M.C., Medrano, G., Cramer, C.L. and Weathers, P.J. (2012) Green factory: Plants as bioproduction platforms for recombinant proteins. *Biotechnol. Adv.*, **30**, 1171–1184.

Yu, X.B., Burke, C., Zhang, J., Marchitelli, L.J., Davis, E.A. and Ackerman, S. (1990) Transcription factor IIA of wheat and human interacts similarly with the adenovirus-2 major late promoter. *Biochem. Biophys. Res. Commun.*, **168**, 498–505.

Zhang, X.P. and Glaser, E. (2002) Interaction of plant mitochondrial and chloroplast signal peptides with the Hsp70 molecular chaperone. *Trends Plant Sci.*, **7**, 14–21.

Ziemienowicz, A. (2014) Agrobacterium-mediated plant transformation: Factors, applications and recent advances. *Biocatal. Agric. Biotechnol.*, **3**, 95–102.

SUPPLEMENTARY MATERIAL

7. SUPPLEMENTARY MATERIAL

S1. Primers for obtaining ELVd-derived cassette modifications

Fw-p[^{AMV}ELVd-derived$_{247}$]YFP	TCGGTACCATGGGGAGAGGTCGTCCTCTATCT
Rv-p[^{AMV}ELVd-derived$_{247}$]YFP	ACGGATCCCATGGGAGGGGGAGAGGTGTCGA
Fw- p[$^{AMV}_{ATG}$ELVd-derived]YFP	GATCGGTACCATGTTGGCGAAAC
Rv - p[$^{AMV}_{ATG}$ELVd-derived]YFP	TCACGGATCCGGGAACAAATCC
Fw-p[^{AMV}ELVd-derived]$_{FS}$YFP	GATCGGTACCTTGGCGAAACCC
Rv-p[^{AMV}ELVd-derived]$_{FS}$YFP	TCACGGATCCCATGAGGGAACAAATC

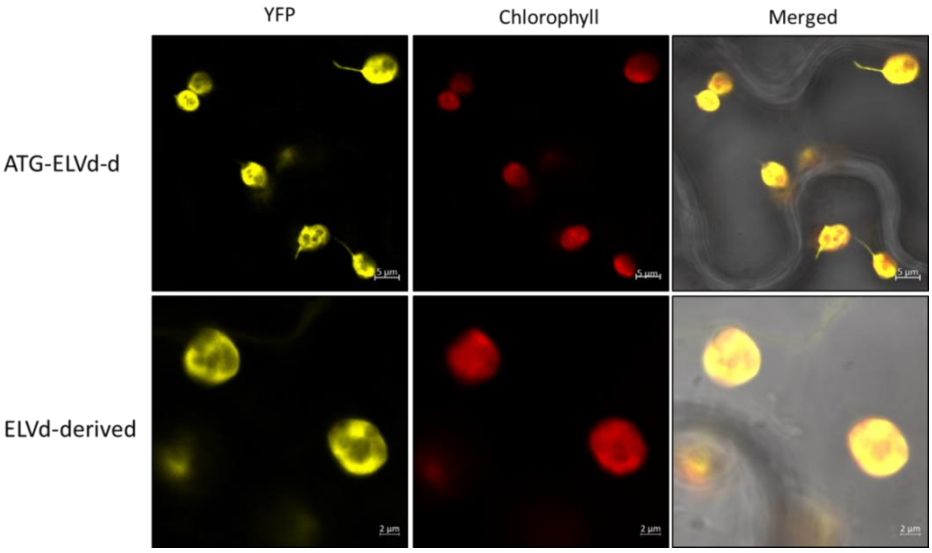
S2. Primers for inverse PCR

Fw-p[^{TMV}ELVd-derived]YFP	AACAAACAACATTACAATTACTATTTACAATTACACCGGTACCTTGGCGAAACCCC
Rv-p[^{TMV}ELVd-derived]YFP	GTTTGTTGTTGTTGGTAATTGTTGTAAAAATACCCTCTCCAAATGAAATGAACTTCC
Fw-p[^{synJ}ELVd-derived]YFP	ACACGCTGGAATTCTAGTATACTAAACCGGTACCTTGGCGAAACCCC
Rv-p[^{synJ}ELVd-derived]YFP	CCTCTCCAAATGAAATGAACTTCC
Fw-p[ELVd-derived]YFP	GGTACCTTGGCGAAACCCC
Rv-p[ELVd-derived]YFP	CCTCTCCAAATGAAATGAACTTCC

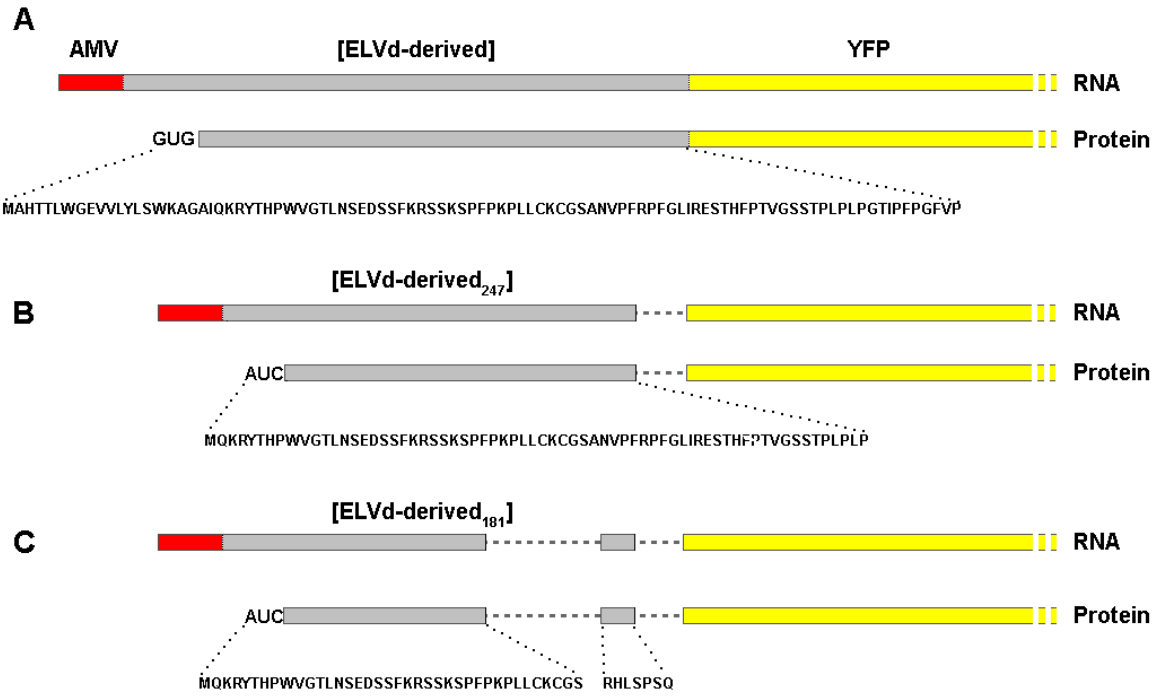
S3. p values of t-test-unpaired with Welch correction

	SynJ 1	**SynJ 2**	**SynJ 3**
AMV 1	< 0.0001	< 0.0001	< 0.0001
AMV 2	< 0.0001	< 0.0001	< 0.0001
AMV 3	< 0.0001	0.0007	< 0.0001

	TMV 1	**TMV 2**	**TMV 3**
AMV 1	0.0093	0.0012	0.0120
AMV 2	< 0.0001	< 0.0001	< 0.0001
AMV 3	0.0094	0.0009	0.0127



S4. Laser scanning confocal microscopy images of *N. benthamiana* leaf tissue agroinfiltrated with bacteria that carry p[$^{AMV}_{ATG}$ELVd-derived]YFP (upper panels) and p[^{AMV}ELVd-derived]YFP (bottom panels). Chlorophyll fluorescence was used as chloroplast marker. These Images were acquired by setting the microscope at different gain values and laser intensities; therefore, the intensity levels cannot be compared.



S5. Schematic representation of [ELVd-derived]YFP (A), [ELVd-derived$_{247}$]YFP (B) and [ELVd-derived$_{181}$]YFP (C) RNAs and the corresponding predicted proteins in case of translation initiation at the indicated non-canonical start codons. Amino acid sequences of the putative ELVd-derived peptides are shown.

S6.

Amino acid composition of the ELVd-derived peptide

Aminoacid	Number	Percentage
Ser (S)	14	12.5
Pro (P)	14	12.5
Leu (L)	11	9.8
Gly (G)	10	8.9
Thr (T)	9	8.0
Phe (F)	8	7.1
Val (V)	8	7.1
Lys (K)	7	6.2
Ala (A)	5	4.5
Arg (R)	4	3.6
His (H)	4	3.6
Glu (E)	3	2.7
Trp (W)	3	2.7
Ile (I)	3	2.7
Tyr (Y)	2	1.8
Asn (N)	2	1.8
Asp (D)	2	1.8
Cys (C)	2	1.8
Gln (Q)	1	0.9
Met (M)	0	0