Plant viral vectors: an strategy for transient expression of recombinant proteins in plants

Vetores virais de plantas: uma estratégia para a expressão transiente de proteínas recombinantes em plantas

Rafaela S. Formoso¹, Isabel L. Vighi¹, Vanessa Galli¹, Luciano S. Pinto¹*

ABSTRACT

Plants have been studied as bioreactors for over 30 years. Their advantages over other expression methods arouse interest in this heterologous protein expression system. Over the years, several methods of plant transformation were developed, and among them, the transient expression system - that occurs temporarily and without integration into the host genome. The use of viral vectors for this type of expression stands out, and the viral genome can be “full-length” or “deconstructed”. Some genome modifications in the “deconstructed” form help searching for efficient protein expression, and several types of viruses are studied to develop these vectors. This review reports different types of viral vectors, especially those particularly suitable for the development of effective vectors of fast and high yield recombinant protein production, based on Tobamoviruses, Potexviruses, Geminiviruses and Comoviruses.

Keywords: Agrobacterium; Agroinfiltration; Viral vectors; Recombinant proteins.

RESUMO

As plantas têm sido estudadas como bioreatores há mais de 30 anos. Suas vantagens sobre outros métodos de expressão despertam o interesse para serem utilizadas como sistema de expressão de proteínas heterólogas. Ao longo dos anos, diversos métodos de transformação de plantas foram desenvolvidos, entre eles, o sistema de expressão transiente - que ocorre temporariamente e sem integração no genoma do hospedeiro. Destaca-se a utilização de vetores virais para este tipo de expressão, podendo o genoma viral utilizado ser “completo” ou “desconstruído”. Algumas modificações do genoma na forma “desconstruída” ajudam na busca por uma expressão proteica eficiente, e vários tipos de vírus são estudados para desenvolver esses vetores. Esta revisão relata diferentes tipos de vetores virais, especialmente aqueles particularmente adequados para o desenvolvimento de vetores eficazes de produção de proteína recombinante rápida e de alto rendimento, baseados em Tobamovírus, Potexvírus, Geminivírus e Comovírus.


¹ Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas.
*E-mail: ls_pinto@hotmail.com
INTRODUCTION

The idea of using plants to express proteins emerged during the 1980s (FRANKEN; TEUSCHEL; HAIN, 1997), and several studies using transgenic plants for the production of antibodies (HIATT; CAFFFERKEY; BOWDISH, 1997), hormones (BARTA et al., 1986), and vaccines (THANAVALA et al., 1995) were developed. Despite the existence of well-defined systems for recombinant proteins expression - such as Escherichia coli (BAESHEN et al., 2015), insect cells (KOLLEWE, 2013), mammalian cells (ZHU, 2012) and yeasts (NIELSEN, 2014) - plants present many advantages: low production cost, high level of scalability, ability to synthesize complex proteins, low risk of contamination of human pathogen and high level of expression in less time (Figure 1).

Figure 1 - General comparison of expression hosts for heterologous protein production.

Note: content is sourced from LOH et al. (2017). The glycosylation pattern is compared to that of human counterpart. This figure was created using BioRender.

To produce transgenic plants a stable transformation is required, which involves the integration of T-DNA with the host cell genome. This type of transformation offers some advantages, such as relatively higher stability of the transferred gene and an easy process to set up and use (HWANG; YU; LAI, 2017). As an example, during the 2014 Ebola crisis, the production of ZMapp (a cocktail of three monoclonal antibodies) in transgenic tobacco plants represented a major advance on drug production in transgenic plants (TAKEYAMA; KIYONO; YUKI, 2015). However, the stable transformation technique offers disadvantages: it is a time-consuming process, which requires a lot of work before analyzing the expressed proteins; gene silencing can occur, causing damage.
to the host's genome; and there is a risk of hybridization with non-transgenic plants (Lomonossoff; Daoust, 2016).

The transient expression appears as an alternative for protein production in plants. In this case, heterologous proteins are produced without stable integration of the transgene into the plant's genome and its transference (Liu et al., 2018). One of the main advances was the use of transient expression systems based on Agrobacterium tumefaciens as a vector, which supplies the transgene through syringes directly in leaf cells by infiltration, a technique called agroinfiltration (Fahad et al., 2014).

The most widely used method to express proteins transiently in plants is the use of viral vectors. When introduced into host plant cells, viruses are designed to contain a gene of interest replicate, allowing the expression of high levels of heterologous proteins in infected plants (Lindbo, 2007; Hefferon, 2014; Ibrahim; Odon; Kormelink, 2019). This method offers many advantages over the transgenic system, including its compatibility with various plant species, reduced protein production time, and high profitability (Hefferon, 2019). As the plants used for protein production are not transgenic, it is possible to avoid controversies regarding genetically modified crops (Takeyama; Kiyono; Yuki, 2015). Over time, several techniques were developed for using vectors based on plant virus genomes, which can be divided into "full-length" strategy and "deconstructed" strategy.

"FULL VIRAL VECTOR” STRATEGY

The so-called complete virus strategy, or first-generation viral vectors, are essentially functional viruses that express the sequence that encodes a protein of interest besides all their genes. The protein of interest is expressed from a strong (duplicate) viral promoter, such as the subgenomic promoter of the coating protein, or as a fusion with the viral coating protein (Gleba; Klimyuk; Marillonnet, 2007).

This type of viral vector maintains the mechanisms for the replication cycle, which includes: viral replication, host infection, translation, assembly of mature virions, cell-to-cell movement, plant-wide movement, reprogramming of the host's biosynthetic mechanism and suppression of the silencing gene (Salazar-González; Bañuelos-Hernández; Rosales-Mendoza, 2015). For this, some viral components involved such as the capsid protein (CP) that improves the formation of replication complexes (Asurmendi et al., 2004) and long-distance movement (Hilf;
DAWSON, 1993) and the protein of movement (MP), which is an RNA-binding protein involved in virus cell-to-cell spread (CHEN et al., 2000).

Many plant viruses are used to express proteins of interest. The first virus explored, TMV (tobacco mosaic virus), was established as a viral vector and used for the production of several recombinant proteins (GLEBA; KLIMYUK; MARILLONNET, 2007). The popular vectors are derived from the potato X virus (PVX) and cowpea mosaic virus (CPMV) (DANIELL et al., 2009; IBRAHIN; ODON; KORMELINK, 2019). The expression levels reached under this full virus strategy are considered convenient, with approximately 10% of the total soluble protein, similar to the amount of coating protein produced by a wild-type virus (BAMOGO et al., 2019). However, these vectors present some limitations: the size of the protein is limited due to vector stability and the incapability to produce of hetero-oligomeric proteins (SALAZAR-GONZÁLEZ; BAÑUELOS-HERNÁNDEZ; ROSALES-MENDOZA, 2015), and the biosafety issues (BAMOGO et al., 2019). So, scientists then implied the “deconstructed viruses” strategy as an alternative to solve these problems.

DECONSTRUCTED VIRAL VECTOR' STRATEGY

“Deconstructed vectors” are composed only by the gene of interest and the basic viral replication components (YUSIBOV et al., 2013) - other functions, such as the delivery of DNA for example, are performed by non-viral elements. This strategy aims to replace a native virus ORF not necessary for its replication with the gene of interest, which allows it to be larger than the "complete virus technology" (BAMOGO et al., 2019). In this way, the "deconstructed vectors" present both advantages and disadvantages concerning the previous strategy (Table 1).

Table 1 - Comparison of advantages and disadvantages of first and second-generation vectors.

<table>
<thead>
<tr>
<th>Type of vector</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-generation Vectors</td>
<td>- Successful expression of heterologous proteins fused to proteins of CP.</td>
<td>- Restriction on protein size; - Biosafety;</td>
</tr>
<tr>
<td>Second-generation Vectors</td>
<td>- Can produce larger proteins; - More security; - Can afford the presence of reporter genes;</td>
<td>- Low capacity of systemic movement;</td>
</tr>
</tbody>
</table>
Two elements of the first-generation viral vectors that limit its functionality are the ability of the vector to move systemically, which is provided by the CP, and the low level of expression of the protein of interest, probably because part of cellular metabolism is dedicated to the synthesis of a large amount of CP. One solution was to remove the CP and replace the systemic movement capacity with the use of Agrobacterium to deliver the viral vector to the whole plant (GLEBA; KLIMYUK; MARILLONNET, 2007). In RNA viruses, PVX and CMV (cucumber mosaic virus), the absence of CP can limit cell-to-cell and systemic movement of the virus, which can result in the production of low quantities of the target protein (AZHAKANANDAM et al., 2007). The advantages of viral vectors missing the CP are the lower risk of escaping to the environment and maximization of protein production efficiency. This is due to their inability to systemically infect plants and of being transmitted by insects or other types of natural vectors, as well as the low metabolic energy directed to the expression of CP, abundant in the wild-type virus infected-cells (LINDBO, 2007).

Vector expression has been improved by genetic engineering techniques that modified transcription-processing locations, codon usage patterns, insertion of introns in the coding sequences (IBRAHIN et al., 2019), and adding sequences for reporter gene expression to track movement and expression. The target protein is usually expressed as a monomer without any component of the plant virus, being fused to a viral protein or as an immunogenic chimeric protein (SALAZAR-GONZÁLEZ et al., 2015).

Several virus-based vectors were engineered for transient expression, for example, PVX, cowpea mosaic virus (CPMV), Bean yellow dwarf virus (BeYDV) and tobacco mosaic virus (TMV). Some studies demonstrate that viral vectors for protein expression systems can already be considered consolidated, attractive, and competitive processes.

**VECTORS BASED ON TOBAMOVIRUSES**

Tobamoviruses are rigid rod-shaped single-stranded positive-sense RNA (+ssRNA) viruses. The first virus to be designed as a deconstructed vector was the TMV, due to its relatively simple genome organization and high copy number of the CP (2130 copies) (GLEBA; MARILLONNET; KLIMYUK, 2004). TMV expression system consists of two modules, one containing the replicase and the other containing an exogenous genes integration site (HEFFERON, 2017).
One of the most well-known and effective techniques for producing recombinant proteins using TMV-based vectors was developed by Icon Genetics (Figure 2A). The technique is known as Magnifection and transfect plants with recombinant virus vector modules, combining agro-infiltration with the delivery of a deconstructed vector incapable of spreading to other plants, and allows the co-expression of several necessary proteins to assembly hetero-oligomeric proteins. This produces high levels of expression, the entire process can be concentrated in a greenhouse, generating up to 500kg of protein in 1 ha per year – 1000 times greater when compared to the strategy of “full-length” virus (GLEBA; KLIMYUK; MARILLONNET, 2005). The system is capable of producing up to 5 mg/g of GFP in fresh tissue in plants. Previous studies showed that the co-expression of a viral silencing suppressor, such as Tomato bushy stunt virus P19 or potato virus A HcPro, improves the production of recombinant proteins (KOMAROVA et al., 2010).

The MagnICON vector is used to develop personalized drugs against non-Hodgkin’s lymphoma (NHL) (MCCORMICK et al., 2008). The constructs encoding tumor-derived immunoglobulin were used to transform A. tumefaciens, which then were used to transfect the leaves of N. benthamiana. The expressed proteins were purified and the derived vaccine had success in the first clinical trials, proving to be safe and economically viable (TUSE, 2006). For example, in the production of IgG, to overcome the limitation in the size of the insert, the vector MagnICON was used in combination with a non-competitive vector CP deficient, based on PVX, to efficiently produce the heavy and light IgG chains. The individual IgG chains were fused in the endoplasmic reticulum reaching a yield of 0.5 mg/g of fresh tissue weight in N. benthamiana (GIRITCH et al., 2006). On another occasion, Zahin et al. (2016), expressed the HPV16 L1 protein in leaves of N. benthamiana. The proteins were extracted from agro-infiltrated leaves and efficiently purified and a yield of 250 mg of L1 yield was obtained per kg of plant biomass. The results indicated that the transient expression in tobacco plants is a promising way to produce a cheap and effective vaccine against HPV16.

Another “deconstructed” viral vector that uses the Agrobacterium system is TRBO (TMV RNA-based overexpression), based on the tobacco mosaic virus (TMV) viral vector without the capsid protein and adding the exogenous gene ORF closer to the 3’ end of TMV RNA, which increases the expression rate several times (Figure 2B) (LINDBO, 2007b). Recently, Carlsson et al. (2020) have successfully produced human myoglobin in the leaves of Nicotiana benthamiana by transient expression using a viral
vector TRBO delivered by Agrobacterium tumefaciens, obtaining a yield of purified protein of approximately 210 mg/kg fresh weight leaf tissue.

The Fraunhofer USA company expressed a recombinant monomeric hemagglutinin protein (HAC1) in N. benthamiana leaves using a TMV-based vector and presented a potential vaccine against the pandemic H1N1 flu virus (SHOJI et al., 2013). Several other studies described protein production using TMV-based vectors for vaccines against Zika Virus (YANG et al., 2017), Chikungunya (HURTADO et al., 2019), and human papillomavirus (HPV) (ZAHIN et al., 2016).

Some authors reported some factors that influence TMV expression, for example, Matsuda et al. (2017) observed that 20º C is the ideal temperature to obtain the highest yields after inoculation and Patil et al. (2015) reported that the light intensity influences the movement of the vector inside the plant. TMV was certainly the most explored Tobamovirus for the production of viral vectors, but other researchers, such as Liu and Kearney (2010) developed a vector-based on another tobamovirus - the Sun Hemp Mosaic Virus (SHMV) (Figure 2C) - that infects legumes and incorporates some of the attributes of the Icon and TRBO strategies.

**VECTORS BASED ON POTEXVIRUSES**

Potexviruses are flexible rod-shaped positive-sense RNA viruses and the Potato virus X (PVX) is the frequent vector-based system to express in N. benthamiana (CHAPMAN; KAVANAGH; BAULCOMBE, 1992). The capsid protein gene and those responsible for the PVX movement are the targets in the construction of deconstructed PVX vectors. PVX has been used to express complete proteins, fusion proteins, and epitopes (MARDANOVA et al., 2017).

Wang et al. (2014) developed a PVX-based vector called pCaPVX440 (Figure 2D), in which two proteins can be co-expressed with two different subgenomic promoters, but the expression levels were low. In another version of the vector, pCaPVX760 (Figure 2E), in which CP was excluded, a higher yield of recombinant protein was obtained, but there is no information on the co-expression of two proteins with this vector.

In PVX, the genetic instability is mainly associated with replication and not encapsidation, as is generally the case with icosahedral viruses that have limited space within the viral capsid (PEYRET; LOMONOSOFF, 2015). To solve this problem,
Dickmeis et al. (2014) developed a PVX-based vector that contains a subgenomic promoter of the bamboo mosaic virus (BaMV) (Figure 2F). This vector is genetically stable without limiting the movement or encapsidation of the modified RNA. This new expression vector showed to be particularly useful for long-term protein production and larger inserts.

The deconstructed vector FECT, based on the monopartite potexvirus Foxtail mosaic virus (FoMV) was designed to express proteins in plants, but this vector was not able to spread systemically and was not efficient in the species tested (LIU; KEARNEY, 2010). To overcome these limitations, Bouton et al. (2018), designed an overexpression vector-based on FoMV to infect cereals. The vector (Figure 2G) was superior to the existing monocot expression vectors and the authors reported it was able to carry the encoding sequence of the GUSPlus gene of 1,800 nucleotides - while the other vectors derived from BSMV do not tolerate insertions larger than 450 to 500 (LEE; HAMMOND-KOSACK; KANYUKA, 2012). This vector allows the expression of heterologous proteins in their native forms, including those with N-terminal signal sequences for plant cell secretion.

VECTORS BASED ON COMOVIRUSES

Comoviruses are a genus of bipartite positive-strand RNA viruses. Most work in vector development is based on the Cowpea mosaic virus (CPMV) (SAINSBURY; CANIZARES; LOMONOSSOFF, 2010). It infects several legume species, N. benthamiana, and grows rapidly in its natural host, cowpea (Vigna unguiculata).

As for deconstructed vectors, a transient expression vector CPMV-HT, which is based on deletion in CPMV RNA-2, allows rapid production, with a high level of protein expression without viral replication (SAINSBURY; LOMONOSSOFF, 2008).

The pEAQ vector also allows the expression of exogenous proteins without the need for viral replication (Figure 2H) (SAINSBURY; THUENEMANN; LOMONOSSOFF, 2009). Thus, it increases the translation of proteins without the progression of the viral cycle, which could interfere on the stability of the integration of exogenous genes (MESHCHERIAKOVA; SAXENA; LOMONOSSOFF, 2014).

The pEAQ-HT vector, besides CPMV-based UTRs, includes an expression cassette for the P19 suppressor of the tomato stun virus (TBSV) gene in the T-DNA next
to the expression cassette to increase mRNA stability. The pEAQ-HT vector, in addition to CPMV-based UTRs, includes an expression cassette for the P19 suppressor of the tomato stunt virus (TBSV) gene in the T-DNA next to the HT expression cassette to increase mRNA. Together, these characteristics make pEAQ-HT a popular choice for the transient expression of recombinant proteins in *N. benthamiana*. This vector was widely used due to its ease of use, promoting a high yield of expressed recombinant protein, and direct the expression of GFP in yields above 1.5 g per kg of fresh tissue (SAINSBURY; THUENEMANN; LOMONOSSOFF, 2009). Ruiz et al. (2018), used the pEAQ-HT to express a recombinant Foot-and-mouth disease (FMDV) like particles (VLPs) via transient expression mediated by *Agrobacterium* in *N. benthamiana*. The groups of mice that received purified VLPs and the crude leaf extract produced a specific humoral response with similar antibody titers.

Among the vaccines in clinical trials, those produced in plants show great potential. One of them is the influenza vaccine developed by Medicago Inc., which uses a vector-based on CPMV to mediate the transient expression of hemagglutinin (HA) in *N. benthamiana* (MARDANOVA et al., 2016). This flu hemagglutinin is incorporated into a lipid envelope originating from plant cell membranes. The rapid scalability of this production platform could be advantageous in combating global influenza pandemics since this vaccine takes only three weeks to be produced after the identification of the circulating strain (D’AOUST et al., 2010; LINDSAY et al., 2018). This represents a reduction of four to eight-fold in the time required to manufacture flu vaccines using the traditional egg-based method (SAINSBURY, 2020).

**VECTORS BASED ON GEMINIVIRUSES**

Unlike the viruses described above, geminiviruses are double spherical viruses with single-stranded circular DNA genomes. This has some advantages over RNA viruses, as their genomes are easier to manipulate with modern molecular biology techniques, such as RT-PCR, and the circular DNA genome can handle much larger inserts than linear RNA genomes (PEYRET; LOMONOSSOFF, 2015).

A geminivirus-based expression system derived from the Bean yellow dwarf virus (BeYDV) allows for fast and high-yield protein production in tobacco plants (CHEN et al., 2011) (Figure 21). In this system, a vector carrying BeYDV replication elements is
delivered by agro-infiltration. BeYDV Rep and RepA proteins greatly amplify the transgene copy number in the plant's nucleus, leading to improved transcription and a subsequent increase in the accumulation of the target protein in just 3–5 days (HUANG et al., 2010).

Figure 2 - Simplified diagrams of deconstructed viral vectors. A- The magnICON system is composed of three modules. The recombinase module allows the recombination of the 5′ and 3′ modules to form a complete replicon. ACT2: Arabidopsis actin 2 promoter, RdRp: tobamovirus RNA-dependent RNA polymerase, MP: tobamovirus movement protein, SGP: TMV subgenomic promoter directing transcription of the gene of interest (GOI), R: recombination site, nos: nos terminator, PhiC31: streptomyces phage 31 integrase. B- Description of TRBO vector, black box represents CaMV polyA signal sequence/terminator and gray box a ribozyme (LINDBO, 2007a) C- Tobamovirus SHEC vector was the deleted-vector version of SHMV and was created by elimination of the CP gene, grey arrow: CaMV 35S promoter, dark grey box: ribozyme, black box: CaMV 3 terminator sequence, black arrows: subgenomic promoter (LIU; KEARNEY, 2010). D- Schematic representation of the overexpression vector pCaPVX440. Black arrows indicate the PVX CP subgenomic promoters, green arrow indicates the position of TMV CP subgenomic promoter (WANG et al., 2014). E- Schematic representation of pCaPVX760 and three promoters (arrows) (WANG et al., 2014). F- Schematic representation of modified PVX vector (DICKMEIS; FISCHER; COMMANDEUR, 2014). G- Schematic diagram of the constructed FoMV expression vectors. The gene of interest (GOI) in these vectors is inserted by restriction enzyme or Gateway cloning; 35S, Cauliflower mosaic virus (CaMV) 35S promoter; nos, nopaline synthase terminator; start codon of ORF5A (BOUTON et al., 2018). H- Structure of pEAQ vector. Pink boxes, CPMV-HT UTRs; Blue box, polylinker; P19, gene of silencing suppressor P19 from TBSV; NPTII, eukaryotic kanamycin resistance gene (SAINSBURY et al., 2009). I - Geminiviral BeYDV vector for recombinant proteins expression. The left (LB, pink triangle) and right border (RB, pink triangle) delineate the T-DNA construct that will be transferred into plant cells by Agrobacterium; LIR; GOI, gene of interest; SIR (CHEN et al., 2011). This figure was created using BioRender.
CONCLUSION

Plants demonstrate some advantages over other protein expression systems, although not yet sufficient to displace the main production platforms used in the organic manufacturing industry. On the other side, plants gained their space in the production of recombinant proteins. The combination of Agrobacterium with viral vectors was essential to obtain significant levels of expression. Therefore, further studies should focus on the optimization of existing vectors, trading, and promotion of these techniques so that many new products can be produced, tested, and used, turning the protein production faster, cheaper, and more productive than ever.

ACKNOWLEDGEMENTS

This study was financed in part by the CAPES and CNPq.
REFERÊNCIAS


Recebido em: 15/01/2022
Aprovado em: 16/02/2022
Publicado em: 18/02/2022