

Possible consequences of the overlap between the CaMV 35S promoter regions in plant transformation vectors used and the viral gene VI in transgenic plants

Nancy Podevin^{1,*} and Patrick du Jardin²

¹The European Food Safety Authority (EFSA); Parma, Italy; ²Gembloux Agro-Bio Tech; Plant Biology Unit; University of Liège; Gembloux, Belgium

Keywords: CaMV 35S promoter, gene VI, GM plant, plant transformation vector, transcription, risk assessment

Multiple variants of the Cauliflower mosaic virus 35S promoter (P35S) are used to drive the expression of transgenes in genetically modified plants, for both research purposes and commercial applications. The genetic organization of the densely packed genome of this virus results in sequence overlap between P35S and viral gene VI, encoding the multifunctional P6 protein. The present paper investigates whether introduction of P35S variants by genetic transformation is likely to result in the expression of functional domains of the P6 protein and in potential impacts in transgenic plants. A bioinformatic analysis was performed to assess the safety for human and animal health of putative translation products of gene VI overlapping P35S. No relevant similarity was identified between the putative peptides and known allergens and toxins, using different databases. From a literature study it became clear that long variants of the P35S do contain an open reading frame, when expressed, might result in unintended phenotypic changes. A flowchart is proposed to evaluate possible unintended effects in plant transformants, based on the DNA sequence actually introduced and on the plant phenotype, taking into account the known effects of ectopically expressed P6 domains in model plants.

Introduction

Cauliflower mosaic virus (CaMV) is a pararetrovirus (Caulimoviridae) that infects members of the Brassicaceae family. CaMV was one of first plant DNA viruses to be studied, and its double-stranded circular DNA genome [~8 kb] has been completely sequenced.¹ The genome encodes seven genes and has a large (~700 bp) and a small (~150 bp) intergenic region that contain regulatory sequences and single-stranded interruptions.

The coding sequences are either separated or overlap by several nucleotides, except for gene VI, which lies between the two intergenic regions. CaMV DNA is transcribed from two promoters in the intergenic regions into two major capped and polyadenylated transcripts, the 19S and 35S RNAs.

The regulatory elements of CaMV have been used since the 1980s to express novel genes in plants;² specifically, the 35S promoter (P35S) and terminator are widely used in research and plant biotechnology.^{3,4} The P35S is a strong constitutive promoter, generating high levels of gene expression in dicotyledonous plants. Of the 86 single transgenic plant events that have been authorised in the United States, 54 contain one or more copies of the CaMV P35S.⁵

Odell et al.⁶ demonstrated that a P35S that contains 350 bp (-343 to +8, with +1 as the transcriptional start site) is sufficient

to obtain constitutive expression, which is due to different domains^{7–10} (Fig. 1).

The CaMV genome in the region of the P35S region contains multiple overlapping domains (Fig. 1), with colinearity between regulatory regions and protein-encoding sequences.¹⁰ The 3' end of P35S overlaps with CaMV polyadenylation regions. The 5' end of P35S overlaps with the 3' end of the coding sequence of gene VI.

The product of gene VI is a multifunctional protein (P6, 62 kDa) that harbours nuclear targeting and export signals¹⁵ and ssRNA-, dsRNA- and protein-binding domains. Considerable effort has been devoted to determine the various functions of P6 (Fig. 1).^{12,16–18}

Bioinformatic tools are increasingly being used in the evaluation of transgenic crops. Guidelines, proposed by WHO/FAO¹⁹ and EFSA,²⁰ include the use of bioinformatics screening to assess the risk of potential allergenicity and toxicity. With this aim, the EFSA GMO Panel has updated its guidance for the risk assessment of GM plants and proposed to identify all new ORFs due to the transformation event.²¹ New ORFs are defined as strings of codons uninterrupted by the presence of a stop codon at the insert genomic DNA junction and within the insert.^{20,21} The putative translation products of these ORFs are then screened for similarities with known toxins and allergens.

*Correspondence to: Nancy Podevin; Email: nancy.podevin@gmail.com
Submitted: 04/22/12; Revised: 06/17/12; Accepted: 07/08/12
<http://dx.doi.org/10.4161/gmcr.21406>

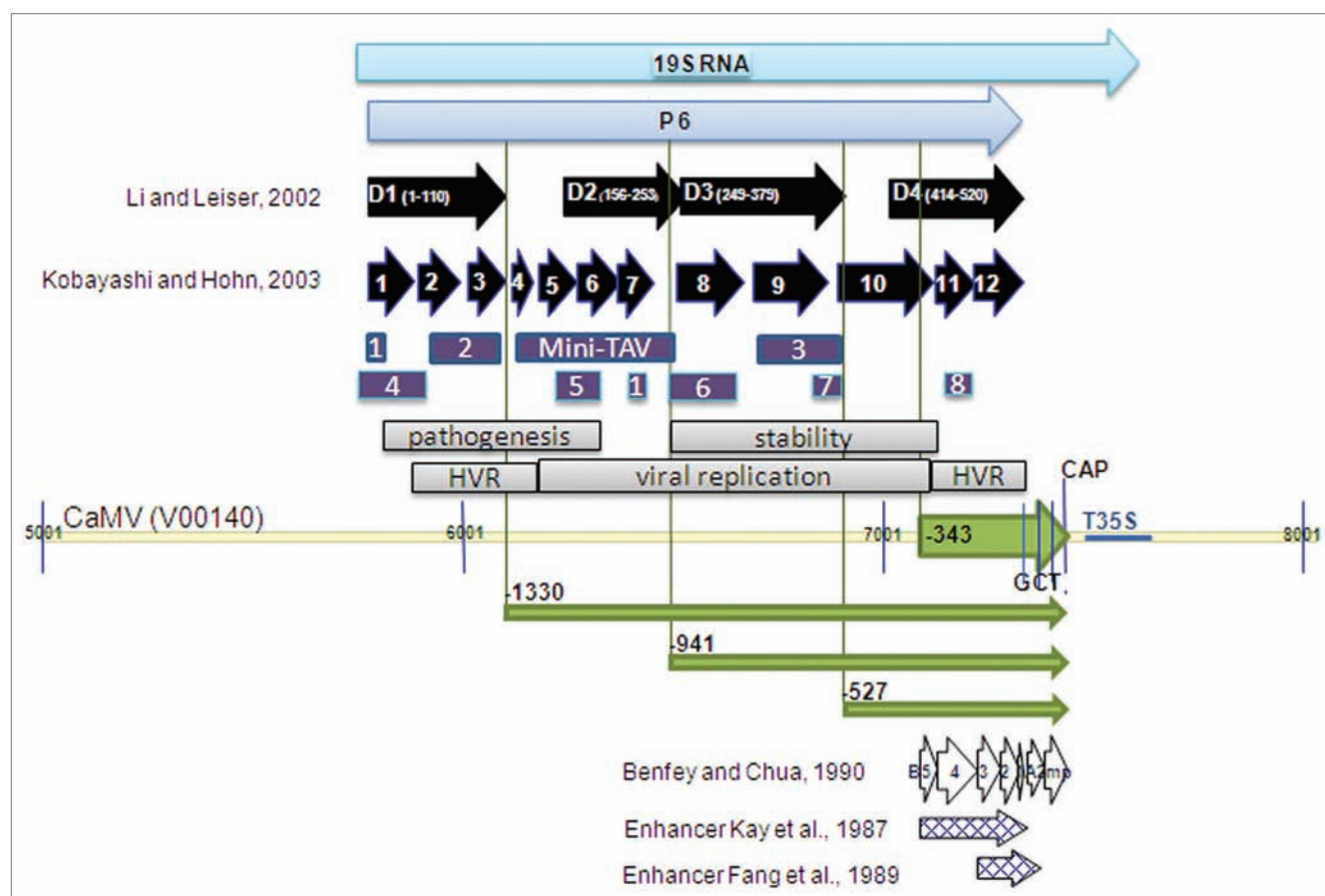


Figure 1. Schematic of overlapping region between the 35S promoter and gene VI, encoding the P6 protein. The 19S RNA (light blue arrow on top) contains only gene VI which codes for the multifunctional protein P6 (blue arrow) which is divided into four domains (D1–D4) according to Li and Leiser.¹¹ Black boxes 1 to 12 indicate the deletion mutants described by Kobayashi and Hohn.¹² Functional domains are indicated in purple: (1) nuclear localization signal, (2) virulence and avirulence Vi/Av, (3) domain important for stability and multimerisation, (4) domain important for stability, (5) RNase H homologous domain that binds RNA-DNA hybrids and double-stranded RNA, (6) RNA-binding domain and multiple protein-binding domain that interacts with eukaryotic translational initiation factor (eIF3) and ribosomal protein L24 (RL24), (7) RNA-binding domain and (8) zinc-finger domain. Grey boxes indicate regions involved in the function or property indicated. The long yellow bar indicates the CaMV genome. In green the position of the P35S variants and below the open boxes indicates the different domains in P35S, as described by Benfey and Chua,¹³ and the hashed boxes indicate the enhancers described by Kay et al.¹⁴ and Fang et al.⁹ HVR, hypervariable regions.

Although information is available on the elements important for promoter activity and the functional domains of the overlapping gene VI this information has not been combined to investigate the possible impact of this overlap. In this article, we discuss the possible consequences of the overlap between gene VI and the 35S promoter, when variants of this promoter are introduced into plant nuclear genomes using stable transformation technology. More specifically we address whether potential expression of the ORFs contained by the P35S promoter overlapping with gene VI: (1) may affect the plant phenotype and (2) show similarity to known allergenic and toxic proteins.

Results

Identification of CaMV 35S promoter variants. The similarity searches against the Patent division of GenBank and information from the literature indicated that different variants of the CaMV P35S are used by plant biotechnologists. These 35S promoters

vary in length between -1329 to +45 and -300 to +8 (relative position to CAP).

Figure 1 shows a representation of the overlapping elements between the P35S, gene VI and the 35S terminator, illustrating that the 5' ends of the -300 and -343 P35S variants²² overlap with domain 4 of P6. The -941 P35S variant²² overlaps with domains 3 and 4 and in part domain 2 of P6. The -1329 P35S variant overlaps with domains 2–4 of P6.

Variants that contain one or more duplications of the 35S enhancer have also been created. Kay et al.¹⁴ fused to the -343 to +9 P35S to the -343 to -90 enhancer. These enhancers overlap with domain 4 of P6.

Determine if ORFs within P35S show similarity to allergenic proteins. The strategy used to search for similarities with toxic and allergenic proteins is in line with current risk assessment requirements in the European Union.^{20,21} The DNA sequences of two variants of the P35S were translated and used to search against allergen databases: (1) the -1329 to +60 P35S variant and

similarities were shown to known allergens using the different algorithms. The AlgPred SMV algorithms indicated that the ORF-encoding portion of the P6 yields a possible allergen. AlgPred is based on dipeptide composition and calculates the frequency of all possible dipeptide combinations. This approach is theoretical and needs to be used in combination with other methods. As no scientific literature has been reported on any allergenic properties of CaMV and no similarities have been shown to known allergens, it can be concluded that the P6 protein is most likely not an allergen. In addition, a toxin database was constructed, and no significant sequence similarity with the P35S variants was detected. These data suggest that the P35S variants do not contain ORFs that encode for proteins that have allergenic or toxic properties.

Clearly, the longer the P35S, the greater the overlap with the coding sequence of gene VI encoding P6 will be. Our literature survey shows that short versions of the P35S (up to position -522 relative to the CAP) overlap only with domain D4 of P6. This domain, when mutated, deleted, or inverted, reduces the rate of viral movement and influences viral host range.^{10,29-31} Thus, the D4 domain appears to be partially dispensable. For short P35S sequences that overlap only with the D4 domain of P6 and for promoters that harbour an additional 35S enhancer that overlaps only with the D4 domain, it is unlikely that chimeric proteins will have unintended effects.

The longest identified version of the P35S (-1329) overlaps with all P6 domains except domain D1. The P6 protein that lacks domain D1 localizes exclusively to the nucleus, because D1 contains residues that are required for P6-P6 intermolecular interactions and viroplasm formation.¹⁵ At least one of P6's nuclear functions is to suppress RNA silencing,³² and various abnormalities that are associated with overexpression of P6 have been suggested to correlate with inhibition of tasiRNA processing.³³ Variants in which the D1 domain has been deleted inhibit replication of the genome in single cells,¹² and De Tapia et al.³⁴ observed that this deleted protein transactivates translation of a polycistronic transcript. Therefore, it is clear that the D1 deletion variant of the P6 protein retains several functions. If a chimeric P6 that contains domains D2–D4 is generated in transgenic plants, it might suppress RNA silencing, affect viral infection through its transactivation activity, or result in an aberrant phenotype. Some of the phenotypes described are leaf chlorosis, vein clearing, plant stunting, late flowering and reduced fertility.^{30,35-38}

Although the P35S overlaps partially with gene VI, the likelihood of unintended effects occurring will depend on whether the partial gene VI is transcribed. We believe that if P35S is embedded in a transformation construct with another gene cassette at its 5' flank, it is unlikely that the partial gene VI will be transcribed. In contrast, when the P35S is inserted adjacent to plant genomic DNA, transcription from an endogenous plant promoter might take place and create a chimeric protein that contains part of P6. To assess these additional aspects a flowchart has been constructed in **Figure 2** to identify the potential unintended effects due to the overlap between the P35S and gene VI. The assessment begins with information on which variant has been used and considers the position of the P35S in constructs and the

insertion site. The impact of the insertion site can be determined, based on the phenotype of the transgenic plant and bioinformatic analyses. In case characteristics attributed to the expression of the P6 gene are observed it should be analyzed if the ORF is expressed.

In conclusion, different P35S variants are in use to express proteins in transgenic plants. Here, we detailed the overlap of P35S with the coding sequence of gene VI. Our bioinformatic analyses indicated that no ORFs are present in the P35S that are similar to known toxic and allergenic proteins. Possible unintended effects that are linked to the use of extended versions of the P35S have been determined. The -343 variant, identified by Odell and colleagues,²² contains all of the necessary elements for full promoter activity and does not appear to result in the presence of an ORF with functional domains, rendering it and its related variants the most appropriate promoter variants for avoiding unintended effects.

Materials and Methods

Identification of P35S variants. Similarity searches were performed against the Patent division of GenBank (on 04/08/2010: 15619638 sequences) using the BLASTn algorithm using default parameters. The search was conducted using sequences from the CaMV genome (GenBank accession number V00140.1) from position 6001 to 7500 to retrieve the largest fragment used as P35S.

Allergenicity assessment by sequence analysis. DNA sequences of the P35S variants were translated using the ExPASy tool (expasy.org). The translated sequences were studied in all six frames to determine the similarity to known allergenic proteins using several databases and the search algorithms that were provided.

Similarity searches were performed against: (1) the Food Allergy Research and Resource Program database (FARRP; www.allergenonline.org Version 12, February 2012: 1603 sequence entries). The 80-amino acid (AA) sliding window method using the criterion of > 35% identity as recommended by the FAO/WHO 2001 expert panel, was employed.^{19,25} In addition, potential identities of 8-AA stretches of identity were investigated; (2) the combined AllerMatch database (December 2005, 792 sequences entries) using the 80-AA sliding window approach and full FASTA, searching for 8-AA exact word matches (allermatch.org); (3) the Allergome database (performed 06/09/2010: 1844 sequence entries) using a full FASTA search (www.allergome.org); (4) the Allergen Database for Food Safety database (ADFS February 2010 version: 1,285 sequences, 91 epitopes, 77 structures and 88 sugar attached entries) using the 80-AA sliding window approach, searching for 8-AA exact word matches and using the MEME motif discovery tool;²³ (5) the AlgPred database using all of the provided search tools: (a) IgE epitope and PID search for IgE epitopes, (b) MEME/MAST motif, (c) SVM method based on AA composition, (d) SVM method based on dipeptide composition and (e) BLAST search of representative allergen peptides.²⁴

Sequence similarity to known toxic proteins. The toxin database was obtained by retrieving a subset of sequences from the GenBank non-redundant protein database (extracted 04/10/2010) using a string search (toxin and toxic). It contains 140774 sequences and was searched using BLASTx with default settings.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. Franck A, Guilley H, Jonard G, Richards K, Hirth L. Nucleotide sequence of cauliflower mosaic virus DNA. *Cell* 1980; 21:285-94; PMID:7407912; [http://dx.doi.org/10.1016/0092-8674\(80\)90136-1](http://dx.doi.org/10.1016/0092-8674(80)90136-1).
2. Hohn T, Richards K, Geneviève-Lebeurier. Cauliflower mosaic virus on its way to becoming a useful plant vector. *Curr Top Microbiol Immunol* 1982; 96:194-236; PMID:6276092; http://dx.doi.org/10.1007/978-3-642-68315-2_12.
3. Covey SN, Lomonosoff GP, Hull R. Characterisation of cauliflower mosaic virus DNA sequences which encode major polyadenylated transcripts. *Nucleic Acids Res* 1981; 9:6735-47; PMID:6174946; <http://dx.doi.org/10.1093/nar/9.24.6735>.
4. Guilley H, Dudley RK, Jonard G, Balázs E, Richards KE. Transcription of Cauliflower mosaic virus DNA: detection of promoter sequences, and characterization of transcripts. *Cell* 1982; 30:763-73; PMID:7139714; [http://dx.doi.org/10.1016/0092-8674\(82\)90281-1](http://dx.doi.org/10.1016/0092-8674(82)90281-1).
5. CERA. Center for Environmental Risk Assessment: GM Crop Database. 2010; http://cera-gmc.org/index.php?action=gmc_crop_database.
6. Odell JT, Knowlton S, Lin W, Mauvais CJ. Properties of an isolated transcription stimulating sequence derived from the Cauliflower mosaic-virus 3ks promoter. *Plant Mol Biol* 1988; 10:263-72; <http://dx.doi.org/10.1007/BF00027403>.
7. Benfey PN, Ren L, Chua NH. The CaMV 35S enhancer contains at least two domains which can confer different developmental and tissue-specific expression patterns. *EMBO J* 1989; 8:2195-202; PMID:16453896.
8. Bhullar S, Datta S, Advani S, Chakravarthy S, Gautam T, Pental D, et al. Functional analysis of cauliflower mosaic virus 35S promoter: re-evaluation of the role of subdomains B5, B4 and B2 in promoter activity. *Plant Biotechnol J* 2007; 5:696-708; PMID:17608668; <http://dx.doi.org/10.1111/j.1467-7652.2007.00274.x>.
9. Fang RX, Nagy F, Sivasubramaniam S, Chua NH. Multiple cis regulatory elements for maximal expression of the cauliflower mosaic virus 35S promoter in transgenic plants. *Plant Cell* 1989; 1:141-50; PMID:2535461.
10. Turner DS, McCallum DG, Covey SN. Roles of the 35S promoter and multiple overlapping domains in the pathogenicity of the pararetrovirus cauliflower mosaic virus. *J Virol* 1996; 70:5414-21; PMID:8764052.
11. Li YZ, Leisner SM. Multiple domains within the Cauliflower mosaic virus gene VI product interact with the full-length protein. *Mol Plant Microbe Interact* 2002; 15:1050-7; PMID:12437303; <http://dx.doi.org/10.1094/MPMI.2002.15.10.1050>.
12. Kobayashi K, Hohn T. Dissection of cauliflower mosaic virus transactivator/viroplasm reveals distinct essential functions in basic virus replication. *J Virol* 2003; 77:8577-83; PMID:12857928; <http://dx.doi.org/10.1128/JVI.77.15.8577-8583.2003>.
13. Benfey PN, Chua NH. The Cauliflower Mosaic Virus 35S Promoter: Combinatorial Regulation of Transcription in Plants. *Science* 1990; 250:959-66; PMID:17746920; <http://dx.doi.org/10.1126/science.250.4983.959>.
14. Kay R, Chan AMY, Daly M, McPherson J. Duplication of CaMV 35S Promoter Sequences Creates a Strong Enhancer for Plant Genes. *Science* 1987; 236:1299-302; PMID:17770331; <http://dx.doi.org/10.1126/science.236.4806.1299>.
15. Haas M, Geldreich A, Bureau M, Dupuis L, Leh V, Vetter G, et al. The open reading frame VI product of Cauliflower mosaic virus is a nucleocytoplasmic protein: its N terminus mediates its nuclear export and formation of electron-dense viroplasms. *Plant Cell* 2005; 17:927-43; PMID:15746075; <http://dx.doi.org/10.1105/tpc.104.029017>.
16. Hapiak M, Li YZ, Agama K, Swade S, Okenka G, Falk J, et al. Cauliflower mosaic virus gene VI product N-terminus contains regions involved in resistance-breakage, self-association and interactions with movement protein. *Virus Res* 2008; 138:119-29; PMID:18851998; <http://dx.doi.org/10.1016/j.virusres.2008.09.002>.
17. Kobayashi K, Hohn T. The avirulence domain of Cauliflower mosaic virus transactivator/viroplasm is a determinant of viral virulence in susceptible hosts. *Mol Plant Microbe Interact* 2004; 17:475-83; PMID:15141951; <http://dx.doi.org/10.1094/MPMI.2004.17.5.475>.
18. Palanichelvam K, Schoelz JE. A comparative analysis of the avirulence and translational transactivator functions of gene VI of Cauliflower mosaic virus. *Virology* 2002; 293:225-33; PMID:11886242; <http://dx.doi.org/10.1006/viro.2001.1293>.
19. FAO/WHO. Evaluation of Allergenicity of Genetically Modified Foods. Report of a Joint FAO/WHO Expert Consultation on Allergenicity of Foods Derived from Biotechnology. Food and Agricultural Organization/World Health Organization. Rome, Italy, 2001.
20. EFSA. Scientific Opinion on the assessment of allergenicity of GM plants and microorganisms and derived food and feed. *EFSA Journal* 2010; 8:1700.
21. EFSA. Guidance for risk assessment of food and feed from genetically modified plants. *EFSA Journal* 2011; 9:2150.
22. Odell JT, Nagy F, Chua NH. Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature* 1985; 313:810-2; PMID:3974711; <http://dx.doi.org/10.1038/313810a0>.
23. Stadler MB, Stadler BM. Allergenicity prediction by protein sequence. *FASEB J* 2003; 17:1141-3; PMID:12709401.
24. Saha S, Raghava GPS. AllPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res* 2006; 34(Web Server issue):W202-9; PMID:16844994; <http://dx.doi.org/10.1093/nar/gkl343>.
25. Codex Alimentarius. Alinorm 03/34: Joint FAO/WHO Food Standard Programme, Codex Alimentarius Commission, Twenty-Fifth Session, Rome, 30 June–5 July, 2003. Appendix III, Guideline for the conduct of food safety assessment of foods derived from recombinant-DNA plants and Appendix IV, Annex on the assessment of possible allergenicity. 2003;p. 47–60.
26. Goodman RE, Hefle SL, Taylor SL, van Ree R. Assessing genetically modified crops to minimize the risk of increased food allergy: a review. *Int Arch Allergy Immunol* 2005; 137:153-66; PMID:15947471; <http://dx.doi.org/10.1159/000086314>.
27. Poulsen LK. Allergy assessment of foods or ingredients derived from biotechnology, gene-modified organisms, or novel foods. *Mol Nutr Food Res* 2004; 48:413-23; PMID:15508176; <http://dx.doi.org/10.1002/mnfr.200400029>.
28. Spök A, Gaugitsch H, Laffer S, Pauli G, Saito H, Sampson H, et al. Suggestions for the assessment of the allergenic potential of genetically modified organisms. *Int Arch Allergy Immunol* 2005; 137:167-80; PMID:15947472; <http://dx.doi.org/10.1159/000086315>.
29. Cecchini E, Gong ZH, Geri C, Covey SN, Milner JJ. Transgenic Arabidopsis lines expressing gene VI from cauliflower mosaic virus variants exhibit a range of symptom-like phenotypes and accumulate inclusion bodies. *Mol Plant Microbe Interact* 1997; 10:1094-101; PMID:9390424; <http://dx.doi.org/10.1094/MPMI.1997.10.9.1094>.
30. Daubert S, Routh G. Point mutations in cauliflower mosaic virus gene VI confer host-specific symptom changes. *Mol Plant Microbe Interact* 1990; 3:341-5; PMID:2134858; <http://dx.doi.org/10.1094/MPMI-3-341>.
31. Noad RJ, Turner DS, Covey SN. Expression of functional elements inserted into the 35S promoter region of infectious cauliflower mosaic virus replicons. *Nucleic Acids Res* 1997; 25:1123-9; PMID:9092619; <http://dx.doi.org/10.1093/nar/25.6.1123>.
32. Haas G, Azevedo J, Moissiard G, Geldreich A, Himber C, Bureau M, et al. Nuclear import of CaMV P6 is required for infection and suppression of the RNA silencing factor DRB4. *EMBO J* 2008; 27:2102-12; PMID:18615098; <http://dx.doi.org/10.1038/emboj.2008.129>.
33. Shivaprasad PV, Rajeswaran R, Blevins T, Schoelz J, Meins F Jr., Hohn T, et al. The CaMV transactivator/viroplasm interferes with RDR6-dependent trans-acting and secondary siRNA pathways in Arabidopsis. *Nucleic Acids Res* 2008; 36:5896-909; PMID:18801846; <http://dx.doi.org/10.1093/nar/gkn590>.
34. De Tapia M, Himmelbach A, Hohn T. Molecular dissection of the cauliflower mosaic virus translation transactivator. *EMBO J* 1993; 12:3305-14; PMID:8344266.
35. Stratford R, Covey SN. Segregation of cauliflower mosaic virus symptom genetic determinants. *Virology* 1989; 172:451-9; PMID:2572087; [http://dx.doi.org/10.1016/0042-6822\(89\)90187-6](http://dx.doi.org/10.1016/0042-6822(89)90187-6).
36. Goldberg KB, Young MJ, Schoelz JE, Kiernan JM, Shepherd RJ. Single gene of CaMV induces disease. *Phytopathology* 1987; 77:1704.
37. Zijlstra C, Schärer-Hernández N, Gal S, Hohn T. Arabidopsis thaliana expressing the cauliflower mosaic virus ORF VI transgene has a late flowering phenotype. *Virus Genes* 1996; 13:5-17; PMID:8938975; <http://dx.doi.org/10.1007/BF00576974>.
38. Yu WC, Murfett J, Schoelz JE. Differential induction of symptoms in Arabidopsis by P6 of Cauliflower mosaic virus. *Mol Plant Microbe Interact* 2003; 16:35-42; PMID:12580280; <http://dx.doi.org/10.1094/MPMI.2003.16.1.35>.