A novel baculovirus-derived promoter with high activity in the Baculovirus Expression System

María Martínez-Solís [1, 2], **Silvia Gómez-Sebastián** [3], **José M Escribano** [4], **Agata K Jakubowska** [1], **Salvador Herrero** Corresp. [1, 2]

[1] Department of Genetics, Universitat de València, Burjassot, Spain

[2] Estructura de Recerca Interdisciplinar en Biotecnologia i Biomedicina (ERI BIOTECMED), Universitat de València, Burjassot, Valencia, Spain

[3] Alternative Gene Expression S.L. (ALGENEX), Madrid, Spain

[4] Departamento de Biotecnología, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Madrid, Spain

Corresponding Author: Salvador Herrero
Email address: sherrero@uv.es

The baculovirus expression vector system (BEVS) has been widely used to produce a large number of recombinant proteins, and is becoming one of the most powerful, robust, and cost-effective systems for the production of eukaryotic proteins. Nevertheless, as in any other protein expression system, it is important to improve the production capabilities of this vector. The *orf46* viral gene was identified among the most highly abundant sequences in the transcriptome of *Spodoptera exigua* larvae infected with its native baculovirus, the *S. exigua* multiple nucleopolyhedrovirus (SeMNPV). Different sequences upstream of the *orf46* gene were cloned, and their promoter activities were tested by the expression of the GFP reporter gene using the *Autographa californica* nucleopolyhedrovirus (AcMNPV) vector system in different insect cell lines (Sf21, Se301, and Hi5) and in larvae from *S. exigua* and *Trichoplusia ni*. The strongest promoter activity was defined by a 120 nt sequence upstream of the ATG start codon for the *orf46* gene. On average, GFP expression under this new promoter was more than two fold higher than the expression obtained with the standard polyhedrin (polh) promoter. Additionally, the *orf46* promoter was also tested in combination with the pph promoter, revealing an additive effect over the polh promoter activity. In conclusion, this new characterized promoter represents an excellent alternative to the most commonly used baculovirus promoters for the efficient expression of recombinant proteins using the BEVS.

A novel baculovirus-derived promoter with high activity in the Baculovirus Expression System

María Martínez-Solís[1,2]; Silvia Gómez-Sebastián[3]; José M. Escribano[4]; Agata K. Jakubowska[1]; Salvador Herrero[1,2]

[1]Department of Genetics, Universitat de València, Burjassot, Spain

[2]Estructura de Recerca Interdisciplinar en Biotecnologia i Biomedicina (ERI BIOTECMED), Universitat de València, 46100-Burjassot, Valencia, Spain

[3]Alternative Gene Expression S.L. (ALGENEX), 28223 Madrid, Spain

[4]Departamento de Biotecnología, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), 28040 Madrid, Spain

Corresponding author:

Salvador Herrero

Dr. Moliner 50, Burjassot, Valencia, 46100, Spain

Email address: sherrero@uv.es

Abstract

The baculovirus expression vector system (BEVS) has been widely used to produce a large number of recombinant proteins, and is becoming one of the most powerful, robust, and cost-effective systems for the production of eukaryotic proteins. Nevertheless, as in any other protein expression system, it is important to improve the production capabilities of this vector. The *orf46* viral gene was identified among the most highly abundant sequences in the transcriptome of *Spodoptera exigua* larvae infected with its native baculovirus, the *S. exigua* multiple nucleopolyhedrovirus (SeMNPV). Different sequences upstream of the *orf46* gene were cloned, and their promoter activities were tested by the expression of the GFP reporter gene using the *Autographa californica* nucleopolyhedrovirus (AcMNPV) vector system in different insect cell lines (Sf21, Se301, and Hi5) and in larvae from *S. exigua* and *Trichoplusia ni*. The strongest promoter activity was defined by a 120 nt sequence upstream of the ATG start codon for the *orf46* gene. On average, GFP expression under this new promoter was more than two fold higher than the expression obtained with the standard polyhedrin (polh) promoter. Additionally, the *orf46* promoter was also tested in combination with the polh promoter, revealing an additive effect over the polh promoter activity. In conclusion, this new characterized promoter represents an excellent alternative to the most commonly used baculovirus promoters for the efficient expression of recombinant proteins using the BEVS.

Introduction

Baculoviruses are enveloped, double-stranded DNA viruses pathogenic to invertebrates, preferably Lepidoptera. Their specificity to kill a narrow spectrum of insects and their safety for humans, plants, and non-target insects, make them a good biological control agent. In addition, since 1983, baculoviruses have been extensively used as protein expression vectors in insect cells (Smith, Summers, and Fraser 1983). The baculovirus expression vector system (BEVS) has been widely used to produce a large number of recombinant proteins, and several systems using different strategies for the generation of the recombinant viruses have been developed (Li et al. 2012; van Oers, Pijlman, and Vlak 2015). The high popularity reached by this system is due to its ability to produce large amounts of active proteins, together with its ability to introduce post-translational modifications in the expressed protein, similar to mammalian cells, such as glycosylation or phosphorylation (O'Reilly, Miller, and Luckow 1994).

Similar to most viruses, the baculovirus gene expression has a temporal regulation which can be divided into 3 main phases: the early, late, and very late phases (Friesen 1997; Lu and Miller 1997; Jarvis 2009). The expression of the early genes does not require prior viral protein synthesis and precedes viral DNA replication. The late phase is a period for viral DNA replication, and the very late phase is characterized by the production of viral particles. In this final phase of infection the expression of the polyhedrin and p10 structural proteins predominate, and these comprise the major proportion of the cell protein mass. The high transcription yield of the promoters of these two proteins has been exploited in the BEVS to express foreign proteins (G. F. Rohrmann 1999). The baculovirus of *Autographa californica* (*A*. c*alifornica* nucleopolyhedrovirus, AcMNPV) is the main viral species used as an expression vector for recombinant protein expression using the BEVS. The polyhedrin and the p10 promoters from AcMNPV have been extensively used for the expression of foreign proteins with this system. However, recombinant protein expression yields not only depend on the promoter used, but also on the host cell line, as well as the characteristics of the foreign gene (Morris and Miller 1992). Several strategies have been developed to improve the production of functional proteins in insect cells. For instance, modification of the expression vectors by the addition of DNA elements involved in protein expression processes can enhance the production yields of recombinant proteins (Lo et al. 2002; Venkaiah et al. 2004; Manohar et al. 2010; Tiwari et al. 2010; Gómez-

Sebastián, López-Vidal, and Escribano 2014). Nevertheless, one of the main cis-regulatory elements affecting the protein expression levels is the promoter. To date, different types of promoters have been tested in the BEVS to improve recombinant protein expression. Viral promoters such as vp39 or 39K, and promoters derived from insect larvae such as the hexamerin-derived promoter pB2 from *Trichoplusia ni* (López-Vidal et al. 2013) showed high levels of expression of recombinant proteins. In other cases, the combination of some of these promoters with the conventional promoters exhibited higher expression levels of the recombinant proteins than the standard late promoters alone (Thiem and Miller 1990; Morris and Miller 1992; Ishiyama and Ikeda 2010; Lin and Jarvis 2012).

In a previous work, the transcriptional pattern of the *Spodoptera exigua* multiple nucleopolyhedrovirus (SeMNPV) during the infective process in its natural host revealed very high levels of expression for the *orf46* viral gene (Pascual et al. 2012). Since the *orf46* gene codes for the structural protein polyhedron envelope protein (PEP), we hypothesized that its expression could be regulated by a strong promoter. In this study, we have determined the core regulatory sequence for the gene (*orf46*) derived from the SeMNPV and we have examined its ability to drive the expression of recombinant proteins in insect cells using the BEVS. Different sequences upstream of the ATG start codon of the *orf46* gene were cloned, and their promoter activities were tested by the expression of GFP as a reporter gene using the AcMNPV system in different insect cell lines. In addition, the promoter activity of this region was tested when combined with the standard polyhedrin promoter derived from the AcMNPV.

Materials and Methods

Culture cells and insects

The *Spodoptera exigua* (Se301) and *Spodoptera frugiperda* (Sf21) cell lines were cultured at 25 °C in Gibco® Grace's Medium (1X) (Life technologies™) supplemented with 10% heat-inactivated fetal bovine serum (FBS). The *Trichoplusia ni* (High Five, Hi5) cell line was cultured at 27 °C in TNMFH medium supplemented with 10% FBS and gentamicin (50 μg/ml). *S. exigua* larvae were maintained in the laboratory, reared on an artificial diet at 25 ± 3 °C with 70 ± 5% relative humidity and a photoperiod of 16/8 hours (light/dark). *Trichoplusia ni* (cabbage looper)

larvae were reared on an artificial insect diet and were kept in growth chambers at 22 ± 1 °C under controlled humidity (50%) and light period (8 h/day) conditions.

Sequence identification

The transcriptional regulatory region was determined by *in silico* analysis of the sequences derived from the Roche 454 FLX and Sanger methods obtained from the transcriptome of *S. exigua* larvae which included samples of SeMNPV-infected larvae (at the latest stage of the infection) (Pascual et al. 2012). First, the ten ORFs with the highest expression levels were obtained based on their maximum coverage. Then, the upstream region from the ATG start codon of *orf46* was analyzed *in silico* and manually for the prediction of the transcriptional regulatory region. Using promoter prediction software (http://www.fruitfly.org/seq_tools/promoter.html), we identified a transcription start site and other motifs characteristic for baculovirus promoters. A sequence of 300 bp upstream of the predicted start site was selected as an initial candidate region to act as a promoter.

Construction of recombinant baculoviruses

Several baculovirus-transfer plasmids containing different fragments of the 5' region of the *orf46* gene driving the expression of GFP were generated using the AcMNPV vector system (Fig. 1). The GFP gene was initially cloned under the control of the polyhedrin promoter (polh) to generate the pFB-PL-GFP vector (López-Vidal et al. 2013) (from now, polh-GFP). The initial pSeL and pSeS promoter sequences were obtained by PCR amplification using SeMNPV genomic DNA as template. PCR amplifications were performed using specific primers which added *Bstz17*I and *Spe*I restriction sites. The polh promoter was then replaced by the pSeL or pSeS fragments into the *Bstz17*I and *Spe*I sites, to generate the pSeL-GFP and pSeS-GFP vectors, respectively. The pSeL140 and pSeL120 sequences were amplified by PCR from the pSeL-GFP vector using two specific primers. The first primer included the corresponding 5' region of pSeL and a *BstZ17*I restriction site, and the second primer was designed to amplify from a 3' region of the GFP gene containing an *Avr*II restriction site. These sequences were cloned into the *BstZ17*I and *Avr*II sites of the polh-GFP vector by replacement of the polh promoter, generating the pSeL140-GFP and pSeL120-GFP vectors.

The vector combining two promoters (polh-pSeL-GFP) was constructed by modification of the polh-GFP vector. The pSeL120 promoter fragment was obtained by PCR using specific primers which added *Xho*I and *Avr*II restriction sites. The resulting fragment was inserted into the *Xho*I and *Avr*II sites of the polh-GFP vector, generating the polh-pSeL-GFP vector containing both of the polh and pSeL120 promoters in tandem. Additionally, the DNA sequence corresponding to the p131 (homolog to *orf46* in AcMNPV) promoter was chemically synthesized (GenScript) and flanked by *BstZ17*I and *Spe*I restriction sites. This was cloned into a pFB vector to control the expression of the GFP gene, generating the p131-GFP vector. Figure 1 shows a schematic representation of all of the different recombinant baculoviruses generated in the present work. The sequences of the primers employed for the cloning of the different constructs are summarized at table S1.

The recombinant baculoviruses were obtained using the Bac-To-Bac® baculovirus expression system (Invitrogen, USA) following the manufacturer's instructions. Plasmids generated in the previous step were used to transform *E. coli* DH10Bac™ heat-shock competent cells and generate the corresponding recombinant bacmids. Bacmids were purified and used to transfect Sf21 cells using Cellfectin® II Reagent (Invitrogen) following the manufacturer's instructions. The resulting baculoviruses were collected after 4-5 days of incubation at 27 °C. These baculoviruses were amplified once to obtain high-titer stocks for further experiments, and the viral titers were determined by quantitative PCR (qPCR) using specific primers (Table S1). For that purpose, viral DNAs were treated using Prepman reagent (Applied Biosystems) following the manufacturer's instructions and were quantified by comparing the obtained Ct values against a standard curve of known viral concentration. Viral titers used in the standard curve were obtained by end point dilution, a method that does not consider non-infective viruses. The viral titers were expressed as baculoviruses per milliliter (BVs/ml).

Infection assays in culture cells and insects

Cells (Se301, Sf21, and Hi5) were cultured in 24-well plates at a confluence of 70%, then the cells were infected with the different recombinant baculoviruses at a multiplicity of infection (MOI) of 5. The cells were collected at different times post-infection by low speed centrifugation

(3000 rpm, 5 min) to avoid cell lysis, and kept at –20 °C until the quantification of GFP expression. Last instar *S. exigua* and *T. ni* larvae were injected with 5 µl of recombinant baculoviruses containing $5x10^4$ BVs. Larvae were maintained at 25 °C and 28 °C, respectively, and after 72 hours post-infection (hpi) were frozen at –20 °C until they were processed for GFP quantification.

Analysis of GFP expression

Frozen cells from the infection assays were resuspended in a lysis buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM DTT, 5% glycerol), incubated for 5 min at room temperature, and centrifugated at 16,000x*g* for 1 min. The supernatant was collected to measure GFP expression by fluorescence in a microplate reader (Infinite® 200 PRO NanoQuant, TECAN) (excitation 485 nm, emission 535 nm). Each value was obtained by measuring each sample 4 times. Occasionally, the production of GFP was confirmed by direct observation of the GFP protein band in SDS-PAGE, suggesting a good correlation between GFP intensity and protein abundance. Frozen larvae were homogenized in 1 ml of extraction buffer (0.01% de Triton X-100, 1 mM de PMSF, and DTT 25 mM in PBS 1x). Homogenates were centrifuged at 1800 x *g* for 30 min at 4 °C, and the supernatant was collected to measure GFP as described above. The values correspond to at least two independent replicates for all of the experiments. Statistical analyses were performed by Dunnett's Multiple Comparison Test using the GraphPad Prism program (GraphPad software Inc., San Diego, CA, USA).

Results

Expression of viral genes and promoter selection

Expression levels of the SeMNPV genes were monitored by mapping of the viral reads on the transcriptome of *S. exigua* infected larvae (Pascual et al., 2012). As expected, the most abundant reads were mapping on the *orf1* which corresponded to the polyhedrin gene (*Table 1*). The second most abundant ORF mapped, corresponded to the *orf46* gene of SeMNPV. *Orf46* codes for the polyhedron envelope protein (PEP), a structural protein that surrounds the polyhedra of the viral particles. In addition, other genes highly expressed during the infection were *orf127* and

179 *orf122*. Given the high expression observed for the *orf46* gene under our experimental
180 conditions, and its role as a structural protein, we decided to explore the possibility of using its
181 regulatory sequence as a promoter for foreign gene expression using the BEVS.

182 Detailed analysis of the 454-derived reads mapping to the SeMNPV genome predicted the
183 transcription start site (site +1) of *orf46* at position 89150 (which referred to the reverse
184 complementary SeMNPV genome, GenBank acc: AF169823.1). A region of 301 nt upstream of
185 the start codon of *orf46* from SeMNPV was initially selected as the promoter sequence. The *in*
186 *silico* analysis predicted a promoter between nucleotides 224-269 from the selected sequence that
187 revealed the presence of a TAAG motif. This TAAG motif was in an AT rich region, and it was
188 described as a typical transcriptional initiation site of late and very late baculovirus promoters
189 (Lu and Miller 1997).

190 **Orf46 promoter activity in insect cells**

191 To determine the promoter activity of the region upstream of the *orf46* gene from SeMNPV and
192 its homologous equivalent region in AcMNPV (p131), different constructs were obtained and
193 tested for their ability to drive the expression of the GFP reporter gene (Figs. 2 and 3) using the
194 BEVS. Se301, Sf21, and Hi5 cells were infected with recombinant AcMNPV baculoviruses
195 expressing GFP under the different promoter regions, and their activities were compared to the
196 activity obtained with the standard polyhedrin (polh) promoter from AcMNPV. The GFP
197 expression yields obtained for the 300 bp fragment upstream of *orf46* (pSeL) was equivalent to
198 that obtained with the polh promoter in the Se301 and Sf21 cells. Interestingly, for the Hi5 cells,
199 the GFP expression mediated by the pSeL promoter was about two-fold higher than that obtained
200 using the polh promoter. The deletion of the 25 nucleotides in the 3' region of the pSeL sequence
201 (pSeS) strongly affected its promoter activity in the three cell types tested, revealing the
202 importance of this region in the activity of the *orf46* promoter. The homolog promoter in
203 AcMNPV (p131) showed a significantly lower expression level than pSeL and the control polh
204 in all of the cell lines tested (Fig. 2).

205 In order to further delimit the promoter region, two additional constructs containing 120 and 140
206 nucleotides upstream of the *orf46* gene were also tested (pSeL120 and pSeL140, respectively).
207 The reduction in the promoter size had a positive impact on the promoter activity in most cases,

as the GFP expression was double that seen when compared to the polh promoter (Figs. 2 and 3). The highest expression levels were observed for the region consisting of the 120 nt upstream of the *orf46* gene. When compared to the polh promoter, pSeL120 showed an increase in expression of more than two fold in all of the cell lines tested. These results strongly suggest that pSeL120 could be considered a useful promoter with the capacity to significantly increase the expression yields obtained with the conventional polyhedrin promoter in the BEVS.

Activity of the pSeL120 in combination with standard promoter in insect cells

In a subsequent analysis, a recombinant baculovirus expressing GFP under the control of a promoter combining the pSeL120 and polh in tandem (polh-pSeL) was generated and tested for its expression levels. After infecting insect cells with this recombinant baculovirus, we observed an additive effect over the two promoters used separately in different recombinant baculoviruses, increasing the polh-pSeL promoter GFP expression to around 3-fold of the levels obtained with the polh or pSeL120 promoters alone (Figs. 3 and 4). This additive effect was observed with small variations in the three insect cell lines tested. These results revealed the potential of pSeL120 to be combined with other promoters in order to produce increased amounts of recombinant proteins in the BEVS.

Activity of the new promoters in baculovirus-infected insect larvae

Although BEVS is mainly used for protein production in insect cell cultures, they can also be used to efficiently produce recombinant proteins in a cost-effective manner by using Lepidoptera larvae. We tested the activity of several of the above described new promoters in larvae from two species of Lepidoptera, the specific host of the SeMNPV, *S. exigua* and *T. ni*, commonly used for protein production using AcMNPV-based vectors. Last instar of *S. exigua* and *T. ni* larvae were infected by intrahemocelical injection with the recombinant baculoviruses expressing GFP under the control of every promoter tested. After 48 hpi, the protein production was estimated by measuring the GFP fluorescence of the larval extracts (Fig. 5). For all of the viruses tested in both insect species, the GFP production using pSeL or pSeL120 was equivalent to that obtained with a baculovirus expressing this protein under the control of the polyhedrin promoter. For the baculovirus comprised of both the polh and pSeL120 promoters, the expression was similar to the polh control promoter in *T. ni* larvae, and slightly lower in *S. exigua* larvae.

237

238 **Discussion**

239 Despite the wide use of the BEVS since the early 1980s (Smith, Summers, and Fraser 1983), the
240 system remains in terms of productivity very similar to the one originally developed. It is worth
241 pointing out the need for research in the improvement of the productivity by different
242 approaches, as was previously shown for other eukaryotic and prokaryotic production platforms.
243 Several strategies have been attempted to increase the production yields by introducing
244 modifications and improvements at different levels. Some of the improvements in the BEVS
245 have been focused on the modification of viral promoters (Manohar et al. 2010), or the
246 introduction of regulatory sequences (Sano et al. 2002; Tiwari et al. 2010; Ge et al. 2014;
247 Gómez-Sebastián, López-Vidal, and Escribano 2014). Other strategies were based on the
248 deletion of non-essential genes of the vector (Hitchman et al. 2010; Hitchman et al. 2011). One
249 standard strategy is the search for promoters which are stronger than those commonly used, such
250 as the p10 and polyhedrin (polh) promoters, or chimeras of them employed in laboratory and
251 industrial production (Thiem and Miller 1990; Ishiyama and Ikeda 2010; Lin and Jarvis 2012;
252 López-Vidal et al. 2013). However, often the efficiency of the promoter also depends on the
253 regulatory sequences around them and the type of cellular lines in which they are acting
254 (Matsuura et al. 1987; Morris and Miller 1992; C H Gross and Rohrmann 1993; Lo et al. 2002).
255 Thus, the development of new promoters to implement the cost-efficient production of
256 recombinant proteins and to provide alternatives to the traditional promoters, still remains of
257 interest.

258 Viral genes coding for structural proteins are usually regulated by strong promoters, since they
259 need to be highly translated to produce the viral particles. Thus, they are good candidates to
260 explore in the improvement of the BEVS. The promoter studied in this work regulates the
261 expression of the *orf46* gene from SeMNPV, which codes for the calyx/polyhedron envelope
262 protein (PEP). The polyhedron envelope is an electron-dense structure that forms a smooth,
263 seamless surface that surrounds polyhedra. The function of calyx/PE is to seal the surface of
264 polyhedra and to enhance their stability (George F. Rohrmann 2013). Homologs of the PEP are
265 found in the genomes of all lepidopteran nucleopolyhedroviruses. The PEP is associated with
266 p10 fibrillar structures, and both proteins appear to be important for the proper formation of the

polyhedron envelope (van Lent et al. 1990; Russell, Pearson, and Rohrmann 1991; C. H. Gross, Russell, and Rohrmann 1994; Lee et al. 1996). PEP from AcMNPV was shown to be associated with BV but not with ODV. It is abundantly produced during the late phase of infection (Wang et al. 2010).

In the present study, we have described a new viral promoter sequence derived from the gene that codes for the structural PEP from SeMNPV, showing better performance than the polh promoter in the BEVS in different cell lines. By testing different sequences upstream of the ATG start codon from the *orf46* gene driving the expression of GFP, we have limited the essential promoter sequence. The sequence corresponding to the 120 nt just before the ATG start codon (pSeL120) showed the strongest promoter activity when it was functioning in cultured cell lines. On average, the expression under the pSeL120 promoter was at least 2 times higher than the maximum expression levels reached using the standard polh promoter. Other groups have investigated the characterization of new promoters for increased expression yields. Lin & Jarvis (Lin and Jarvis 2012) showed that the delayed early 39K promoter from AcMNPV produced 4-fold more SEAP protein than the polyhedrin promoter in Sf21 cells. López-Vidal et al. (López-Vidal et al. 2013) isolated the pB2 promoter (promoter region of the Basic juvenile hormone-suppressible protein 2, BJHSP-2) from the Lepidoptera *T. ni* with activity in Sf21 cells. The pB2 promoter can drive the expression of GFP earlier in time, but it is not as strong as the polyhedrin promoter. Ishiyama & Ikeda (Ishiyama and Ikeda 2010) reported that the expression of GFP was increased using the vp39 late promoter in comparison to the polyhedrin promoter in *Bombix mori* cultured cells.

Despite the high level of conservation and similarity between sequences from different virus species, the homolog p131 sequence from AcMNPV showed the lowest promoter activity, even lower than the control polh promoter. Such discrepancy could be explained by the fact that the p131 transcription start site (predicted *in silico)* is not located in the TAAG region, and this region seems to be very important in order to obtain high expression levels (as mentioned above). Alternatively, it could also be possible that the activity of p131 in AcMNPV is not as crucial as the orf46 activity in SeMNPV. This hypothesis is supported by some gene expression data in AcMNPV in the literature. It has been published that the gene expression levels of pp34 (gene whose expression is controlled by p131 in AcMNPV) were considerably lower than the

297 polyhedrin and p10 expression in infected Sf9 cells (Iwanaga et al. 2004). The analysis of the
298 transcriptome of AcMNPV-infected *T. ni* cells also showed lower expression levels of pp34 in
299 comparison to the polyhedrin and p10 genes (Chen et al. 2013).

300 An additional improvement with regard to the protein expression was obtained when the
301 pSeL120 promoter was combined with the polh promoter (polh-pSeL), resulting in increases of
302 about 3-fold over the polh promoter and 1.5 fold over the pSeL120 promoter alone. Increases in
303 protein production have also been reported by the combination of different promoters. Thiem &
304 Miller (Thiem and Miller 1990) showed that the combination of the vp39 and the polyhedrin
305 promoter enhanced the expression of foreign genes compared to using those promoters alone in
306 Sf cells, because this hybrid promoter showed regulation patterns of late and very late promoters.
307 López-Vidal et al. (López-Vidal et al. 2013) also demonstrated an increase in GFP production of
308 more than 20% at early times post-infection, and similar expression levels at very late times post-
309 infection in Sf21 cells using a pB2-p10 promoter combination, with respect to conventional late
310 promoters.

311 Although our results have shown a clear improvement of the pSeL promoter activity in different
312 cell types, we could not observe such improvement when it was used for protein production in *S.*
313 *exigua* and *T. ni* larvae. The difference in the promoter activity between the cell lines and larvae
314 could be due to additional factors affecting the replication dynamics and/or promoter activity of
315 the virus, as well as the timing selected for the processing of the larvae. Nevertheless, the
316 pSeL120 promoter activity in larvae is equivalent to that obtained using the polh promoter, and
317 no significant differences were observed, demonstrating that the promoter exhibits versatility and
318 can be utilized in both cell lines (with high activity for a wide range of cell types) and insect
319 larvae (with activity equal to the polh promoter).

320 When compared with homologous sequences in other viral species, we found a region of 50 nt
321 upstream of the ATG start codon that was highly conserved between them. Interestingly,
322 removal of 25 nt of this sequence downstream of the +1 start transcription site in mRNA
323 abolishes the activity of the pSeS promoter. This observation suggests that this region is essential
324 for the strong promoter activity as already proposed in previous studies. Weyer and Possee
325 (Weyer and Possee 1988) showed that the 5'UTR regions are necessary for the maximum
326 activity of the polyhedrin and p10 promoters. In agreement with that, expression levels for

foreign proteins are related to the integrity of the 5'UTR region of the polyhedrin gene (Matsuura et al. 1987; Luckow and Summers 1988). The sequence located between the TAAG motif and the translation initiation site is known in baculoviruses as the burst sequence (BS) (Weyer and Possee 1988). This is a sequence of about 50 nt required for the efficient expression of viral genes during the very late phase of infection. Studies of mutational analysis regarding the BS region have demonstrated that BS are essential for efficient protein expression (Ooi, Rankin, and Miller 1989; Weyer and Possee 1988), which agrees with our results. If we take into consideration that most of those highly conserved 50 nucleotides are included in the 5'UTR of the ORF46 transcript, it seems that the increase in expression found with the pSeL-derived promoters is likely influenced by the effect of such sequences with respect to the access provided to the RNApol, which affects the transcription and translation rates, and even increases mRNA stability. The reason in this case is not known, but it has already been described that an upstream sequence of the AcMNPV polyhedrin gene has an important function for mRNA transcription and translation efficiencies (Min and Bishop 1991).

In conclusion, the sequence derived from the SeMNPV genome described in this work represents a new promoter which is able to express, in most cases, higher yields of foreign proteins than the polh promoter in the BEVS. Moreover, the combination of pSeL with the conventional polh promoter showed higher activity for the expression of GFP than the pSeL or polh promoters alone. Although additional validations of this promoter for the expression of recombinant proteins other than GFP would be needed, these results represent a new improvement in the production of recombinant proteins using the BEVS, with potential application in the cost-efficient large-scale industrial production of biologics.

Acknowledgments

We want to thank Rosa Maria González-Martínez for her excellent help with insect rearing and laboratory management.

353 **Figure legends**

354 **Figure 1: Schematic summary of the recombinant baculoviruses carrying different**
355 **promoter regions employed in this study.**

356 (A) Nucleotide sequence upstream of the *orf46* gene from SeMNPV selected as a regulatory
357 region (nucleotides 45417-47500 from AF169823). The arrows indicate the range of the
358 fragments from the 5' to 3' sequence that were tested for promoter activity and the transcription
359 initiation site (+1). (B) Schematic representation of the recombinant baculoviruses generated
360 which carry different fragments of the sequence upstream of the *orf46* gene to test their promoter
361 activity using GFP as a reporter. The white open arrows (polh) represent the polyhedrin
362 promoter. The white boxes represent the GFP gene. The numbers indicate the first and the last
363 nucleotides (from 5' to 3') of the sequence that was cloned as a promoter. The dotted line in the
364 pSeS construct represents the 5' fragment that is absent.

365

366 **Figure 2: Promoter activity of the sequences upstream of the *orf46* gene.**

367 GFP expression, measured as relative fluorescence intensity, in different insect cell lines infected
368 with the different baculoviruses at a multiplicity of infection (MOI) of 5. The fluorescence was
369 measured at different time points after infection of Se301 (A), Sf21 (B), and Hi5 (C) cells. The
370 results are expressed as the relative percentage of GFP fluorescence intensity, taken as 100% of
371 the value corresponding to the maximum intensity obtained with the polh promoter. The values
372 are the means of at least two independent assays. The error bars represent the standard error of
373 the mean.

374

375 **Figure 3: Fluorescence microscopy of Sf21 cells infected with the different baculoviruses.**

376 A representative image of Sf21 cells infected with a selected baculovirus at a MOI of 5. The
377 images were taken at 48 hours post-infection.

378

Figure 4: Promoter activity of pSeL120 when combined with the polh promoter.

GFP expression, measured as the relative fluorescence intensity, in different insect cell lines infected with the different baculoviruses at a MOI of 5. The fluorescence was measured at different time points after the infection of Se301 (A), Sf21 (B), and Hi5 (C) cells. The results are expressed as the relative percentage of GFP fluorescence intensity, taken as 100% of the value corresponding to the maximum intensity obtained with the polh promoter. The values are the means of at least two independent assays. The error bars represent the standard error of the mean.

Figure 5: Promoter activity in insect larvae.

GFP expression, measured as relative fluorescence intensity, was obtained in insect larvae infected with the different recombinant baculoviruses. The results are expressed as the relative percentage of GFP fluorescence intensity, taken as 100% of the value corresponding to values for the control sequence with the polh promoter. The values are the means of at least two independent assays. The error bars represent the standard error of the mean.

References

ChenY-R, Zhong S, Fei Z, Hashimoto Y, Xiang JZ, Zhang S, Blissard GW. 2013. "The Transcriptome of the Baculovirus *Autographa Californica* Multiple Nucleopolyhedrovirus in *Trichoplusia Ni* Cells." *Journal of Virology* 87 (11): 6391–6405. doi:10.1128/JVI.00194-13.

Friesen PD 1997. "Regulation of Baculovirus Early Gene Expression." In *The Baculoviruses*, 141–70. doi:10.1007/978-1-4899-1834-5_6.

Ge J, Jin L, Tang X, Gao D, An Q, Ping W. 2014. "Optimization of eGFP Expression Using a Modified Baculovirus Expression System." *Journal of Biotechnology* 173 (1): 41–46. doi:10.1016/j.jbiotec.2014.01.003.

Gómez-Sebastián S, López-Vidal J, Escribano JM. 2014. "Significant Productivity Improvement of the Baculovirus Expression Vector System by Engineering a Novel Expression Cassette." Edited by Yi Li. *PloS One* 9 (5). Public Library of Science: e96562. doi:10.1371/journal.pone.0096562.

Gross CH, Rohrmann GF. 1993. "Analysis of the Role of 5' Promoter Elements and 3' Flanking Sequences on the Expression of a Baculovirus Polyhedron Envelope Protein Gene." *Virology*. doi:10.1006/viro.1993.1030.

Gross CH, Russell RLQ, Rohrmann GF. 1994. "*Orgyia Pseudotsugata* Baculovirus p10 and Polyhedron Envelope Protein Genes: Analysis of Their Relative Expression Levels and Role in Polyhedron Structure." *Journal of General Virology* 75 (5): 1115–23. doi:10.1099/0022-1317-75-5-1115.

Hitchman RB, Locanto E, Possee RD, King LA. 2011. "Optimizing the Baculovirus Expression Vector System." *Methods* 55 (1). Elsevier Inc.: 52–57. doi:10.1016/j.ymeth.2011.06.011.

Hitchman RB, Possee RD, Crombie AT, Chambers A, Ho K, Siaterli E, Lissina O, Sternard H, Novy R, Loomis K, Bird LE, Owens RJ, King LA. 2010. "Genetic Modification of a Baculovirus Vector for Increased Expression in Insect Cells." *Cell Biology and Toxicology* 26 (1): 57–68. doi:10.1007/s10565-009-9133-y.

Ishiyama, Seiji, and Masahiro Ikeda. 2010. "High-Level Expression and Improved Folding of Proteins by Using the vp39 Late Promoter Enhanced with Homologous DNA Regions." *Biotechnology Letters* 32 (11): 1637–47. doi:10.1007/s10529-010-0340-7.

Iwanaga M, Takaya K, Katsuma S, Ote M, Tanaka S, Kamita SG, Kang W, Shimada T, Kobayashi M. 2004. "Expression Profiling of Baculovirus Genes in Permissive and Nonpermissive Cell Lines." *Biochemical and Biophysical Research Communications* 323 (2): 599–614. doi:10.1016/j.bbrc.2004.08.114.

Jarvis DL. 2009. "Baculovirus-Insect Cell Expression Systems." *Methods in Enzymology* 463: 191–222. doi:10.1016/S0076-6879(09)63014-7.

Lee SY, Poloumienko A, Belfry S, Qu X, Chen W, MacAfee N, Morin B, Lucarotti C, Krause M. 1996. "A Common Pathway for p10 and Calyx Proteins in Progressive Stages of Polyhedron Envelope Assembly in AcMNPV-Infected *Spodoptera Frugiperda* Larvae." *Archives of Virology*, 2057–76.

Li S, Wang H, Hu Z, Deng F. 2012. "Genetic Modification of Baculovirus Expression Vectors." *Virologica Sinica* 27 (2): 71–82. doi:10.1007/s12250-012-3236-y.

Lin C-H, Jarvis DL. 2012. "Utility of Temporally Distinct Baculovirus Promoters for Constitutive and Baculovirus-Inducible Transgene Expression in Transformed Insect Cells." *Journal of Biotechnology* 29 (6): 997–1003. doi:10.1016/j.biotechadv.2011.08.021.Secreted.

Lo HR, Chou CC, Wu TYY, Yuen JPY, Chao YC. 2002. "Novel Baculovirus DNA Elements Strongly Stimulate Activities of Exogenous and Endogenous Promoters." *Journal of Biological Chemistry* 277 (7): 5256–64. doi:10.1074/jbc.M108895200.

López-Vidal J, Gómez-Sebastián S, Sánchez-Ramos I, Escribano JM. 2013. "Characterization of a *Trichoplusia Ni* Hexamerin-Derived Promoter in the AcMNPV Baculovirus Vector." *Journal of Biotechnology* 165 (3-4). Elsevier B.V.: 201–8. doi:10.1016/j.jbiotec.2013.03.012.

Lu A, Miller LK. 1997. "Regulation of Baculovirus Late and Very Late Gene Expression." The Baculoviruses, Plenum, New York/London. 193–216.

Luckow VA, Summers MD. 1988. "Signals Important for High-Level Expression of Foreign Genes in *Autographa Californica* Nuclear Polyhedrosis Virus Expression Vectors." *Virology* 167 (1): 56–71. doi:10.1016/0042-6822(88)90054-2.

Manohar SL, Kanamasa S, Nishina T, Kato T, Park EY. 2010. "Enhanced Gene Expression in Insect Cells and Silkworm Larva by Modified Polyhedrin Promoter Using Repeated Burst Sequence and Very Late Transcriptional Factor-1." *Biotechnology and Bioengineering* 107 (6): 909–16. doi:10.1002/bit.22896.

Matsuura Y, Possee RD, Overton HA, Bishop DHL. 1987. "Baculovirus Expression Vectors : The Requirements for High Level Expression of Proteins , Including Glycoproteins." *Journal of General Virology* 68: 1233–50. doi:10.1099/0022-1317-68-5-1233.

Min MK, Bishop DHL. 1991. "Transcriptional Analyses of Baculovirus Polyhedrin and Foreign Gene Expression Relative to Baculovirus p10 mRNA Levels." *The Journal of General Virology* 72 (Pt 10 (1991): 2551–56.

Morris TD, Miller LK. 1992. "Promoter Influence on Baculovirus-Mediated Gene Expression in Permissive and Nonpermissive Insect Cell Lines." *Journal of Virology* 66 (12): 7397–7405.

O'Reilly DR, Miller LK, Luckow VA. 1994. *Baculovirus Expression Vectors: A Laboratory Manual*. Oxford University Press.

Ooi BG, Rankin C, Miller LK. 1989. "Downstream Sequences Augment Transcription from the Essential Initiation Site of a Baculovirus Polyhedrin Gene." *Journal of Molecular Biology* 210 (4): 721–36.

Pascual L, Jakubowska AK, Blanca JM, Cañizares J, Ferré J, Gloeckner G, Vogel H, Herrero S. 2012. "The Transcriptome of *Spodoptera Exigua* Larvae Exposed to Different Types of Microbes." *Insect Biochemistry and Molecular Biology* 42 (8): 557–70. doi:10.1016/j.ibmb.2012.04.003.

Rohrmann GF. 1999. "Nuclear Polyhedrosis Viruses." In *Encyclopedia of Virology*, 2nd ed.,

146–52. Academic Press, London, United Kingdom.

Rohrmann GF. 2013. *Baculovirus Molecular Biology*. Third Edit. Bethesda (MD): National Center for Biotechnology Information (US).

Russell RL, Pearson MN, Rohrmann GF. 1991. "Immunoelectron Microscopic Examination of *Orgyia Pseudotsugata* Multicapsid Nuclear Polyhedrosis Virus-Infected *Lymantria Dispar* Cells: Time Course and Localization of Major Polyhedron-Associated Proteins." *The Journal of General Virology* 72 (Pt 2) (2). Society for General Microbiology: 275–83. doi:10.1099/0022-1317-72-2-275.

Sano KI, Maeda K, Oki M, Maéda Y. 2002. "Enhancement of Protein Expression in Insect Cells by a Lobster Tropomyosin cDNA Leader Sequence." *FEBS Letters* 532 (1-2): 143–46. doi:10.1016/S0014-5793(02)03659-1.

Smith GE, Summers MD, Fraser MJ. 1983. "Production of Human Beta Interferon in Insect Cells Infected with a Baculovirus Expression Vector." *Molecular and Cellular Biology* 3 (12): 2156–65. doi:10.1128/MCB.3.12.2156.Updated.

Thiem SM, Miller LK. 1990. "Differential Gene Expression Mediated by Late, Very Late and Hybrid Baculovirus Promoters." *Gene* 91 (1): 87–94. doi:10.1016/0378-1119(90)90166-O.

Tiwari P, Saini S, Upmanyu S, Benjamin B, Tandon R, Saini KS, Sahdev S. 2010. "Enhanced Expression of Recombinant Proteins Utilizing a Modified Baculovirus Expression Vector." *Molecular Biotechnology* 46 (1): 80–89. doi:10.1007/s12033-010-9284-3.

van Lent JWM, Groenen JTM, Klinge-Roode EC, Rohrmann GF, Zuidema D, Vlak JM. 1990. "Localization of the 34 kDa Polyhedron Envelope Protein in *Spodoptera Frugiperda* Cells Infected with *Autographa Californica* Nuclear Polyhedrosis Virus." *Archives of Virology* 111 (1-2): 103–14. doi:10.1007/BF01310508.

van Oers MM, Pijlman GP, Vlak JM. 2015. "Thirty Years of Baculovirus-Insect Cell Protein Expression: From Dark Horse to Mainstream Technology." *Journal of General Virology* 96: 6–23. doi:10.1099/vir.0.067108-0.

Venkaiah B, Viswanathan P, Habib S, Hasnain SE. 2004. "An Additional Copy of the Homologous Region (hr1) Sequence in the *Autographa Californica* Multinucleocapsid Polyhedrosis Virus Genome Promotes Hyperexpression of Foreign Genes." *Biochemistry* 43 (25): 8143–51. doi:10.1021/bi049953q.

Wang R, Deng F, Hou D, Zhao Y, Guo L, Wang H, Hu Z. 2010. "Proteomics of the *Autographa Californica* Nucleopolyhedrovirus Budded Virions." *Journal of Virology* 84 (14): 7233–42. doi:10.1128/JVI.00040-10.

Weyer U, Possee RD. 1988. "Functional Analysis of the p10 Gene 5' Leader Sequence of the *Autographa Californica* Nuclear Polyhedrosis Virus." *Nucleic Acids Research* 16 (9): 3635–53. doi:10.1093/nar/16.9.3635.

Figure 1(on next page)

Schematic summary of the recombinant baculoviruses carrying different promoter regions employed in this study.

(A) Nucleotide sequence upstream of the *orf46* gene from SeMNPV selected as a regulatory region (nucleotides 45417-47500 from AF169823). The arrows indicate the range of the fragments from the 5’ to 3’ sequence that were tested for promoter activity and the transcription initiation site (+1). (B) Schematic representation of the recombinant baculoviruses generated which carry different fragments of the sequence upstream of the *orf46* gene to test their promoter activity using GFP as a reporter. The white open arrows (pph) represent the polyhedrin promoter. The white boxes represent the GFP gene. The numbers indicate the first and the last nucleotides (from 5’ to 3’) of the sequence that was cloned as a promoter. The dotted line in the pSeS construct represents the 5’ fragment that is absent.


A
B
C
+1
D
E
pSeL : ATCATGTGTTTGAACGGCGACTGTTTGGCGGTATGCAGGAGCGTCGCTACCTCGACCGACGAACCGGCGGCCTTTTTGGATTTTGTCAAAAAGAACGACGCAGCGAAAACGCCAATGTTG : 120
pSeL : CAAATTGTGGACCGGCACGAACACGTTGAATTGTTACTGGACCGAATGTACAATATTGTTGAAATGTTCAATAATCAATAAACCTCTCTATTATGCTTTGTAAATTTTTTATTTACTTTG : 240
pSeL : TTGTCCATTGCTGATGATTTATTAAGTCTTTTGGCAAAAATTAAATTTTTGCTACAATATA : 301

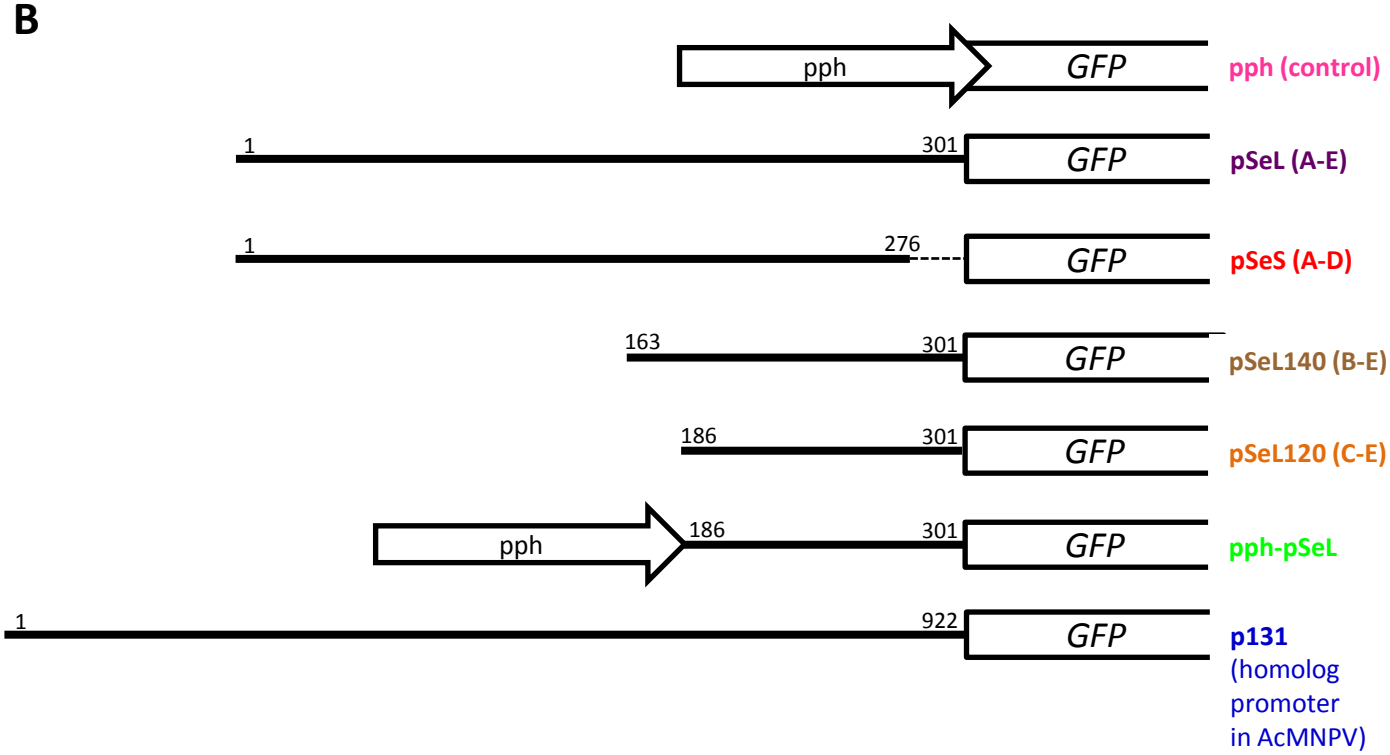

B
pph
GFP
pph (control)
1
301
pSeL (A-E)
276
pSeS (A-D)
163
pSeL140 (B-E)
186
pSeL120 (C-E)
pph-pSeL
922
p131
(homolog
promoter
in AcMNPV)

Table 1(on next page)

ORFs from the SeMNPV highly expressed during infection of *S. exigua* larvae. (Pascual et al., 2012).

Table 1. ORFs from the SeMNPV highly expressed during infection of *S. exigua* larvae. (Pascual et al., 2012).

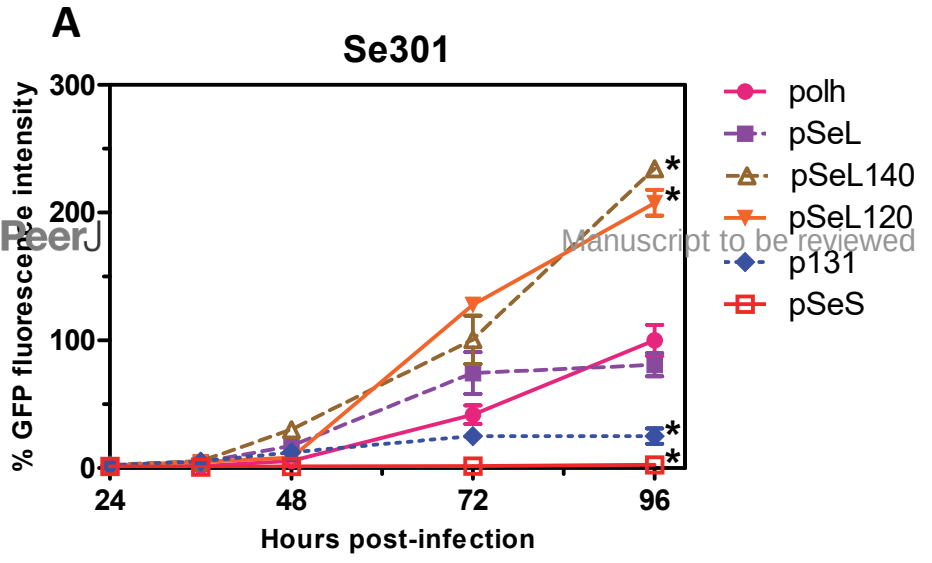
ORF	Description	Coverage[1]
ORF1	Polyhedrin	674
ORF46	Calyx/polyhedron envelope protein	590
ORF127	lef6	516
ORF122	-	416
ORF94	-	347
ORF71	odv-e25	344
ORF65	p6.9 DNA binding protein	262
ORF136	odv-e18	255
ORF32	pkip	224
ORF124	-	204

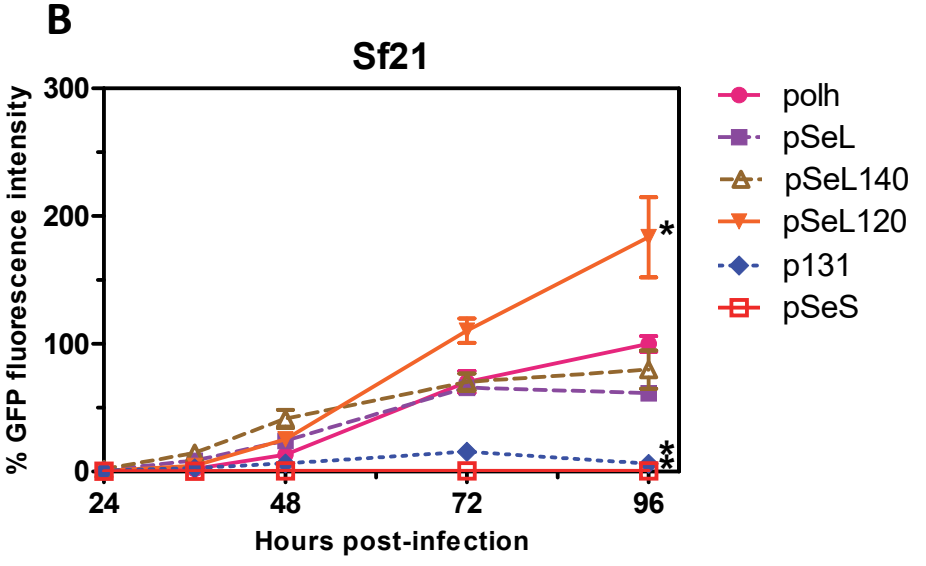
[1]Coverage reported as the maximum coverage (number of reads) for a given ORF after mapping of the SeMNPV genome with transcriptional data.

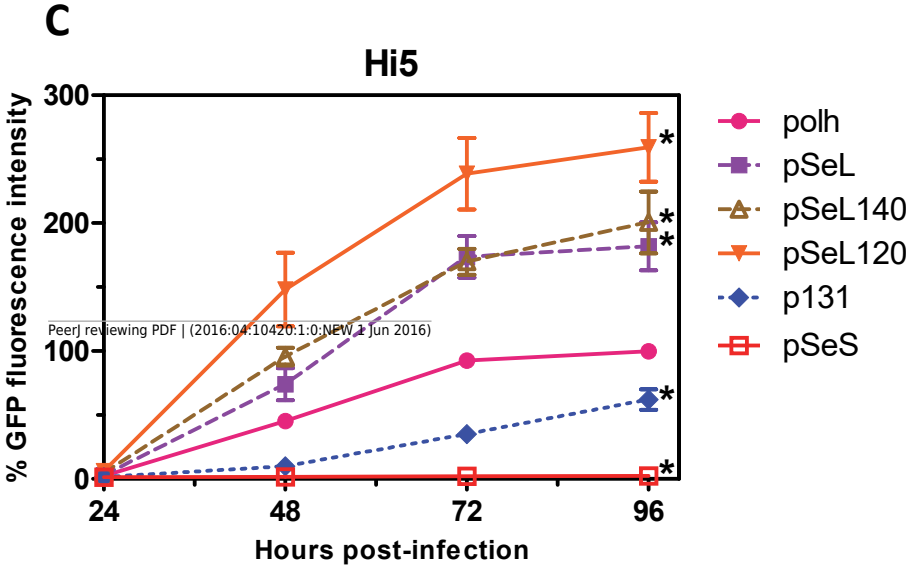
Figure 2(on next page)

Promoter activity of the sequences upstream of the *orf46* gene.

GFP expression, measured as relative fluorescence intensity, in different insect cell lines infected with the different baculoviruses at a multiplicity of infection (MOI) of 5. The fluorescence was measured at different time points after infection of Se301 (A), Sf21 (B), and Hi5 (C) cells. The results are expressed as the relative percentage of GFP fluorescence intensity, taken as 100% of the value corresponding to the maximum intensity obtained with the pph promoter. The values are the means of at least two independent assays. The error bars represent the standard error of the mean.







3

Fluorescence microscopy of Sf21 cells infected with the different baculoviruses.

A representative image of Sf21 cells infected with a selected baculovirus at a MOI of 5. The images were taken at 48 hours post-infection.

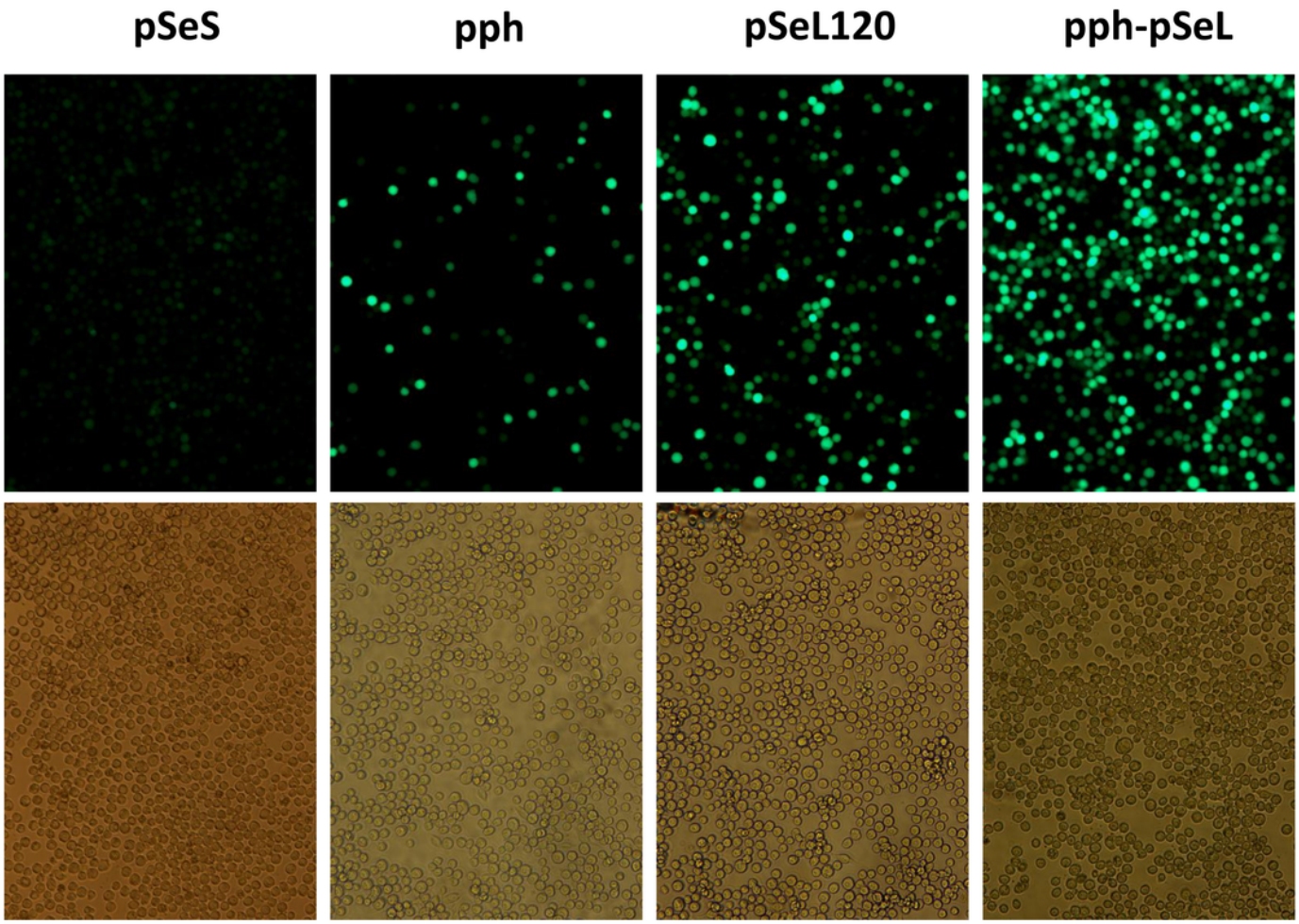
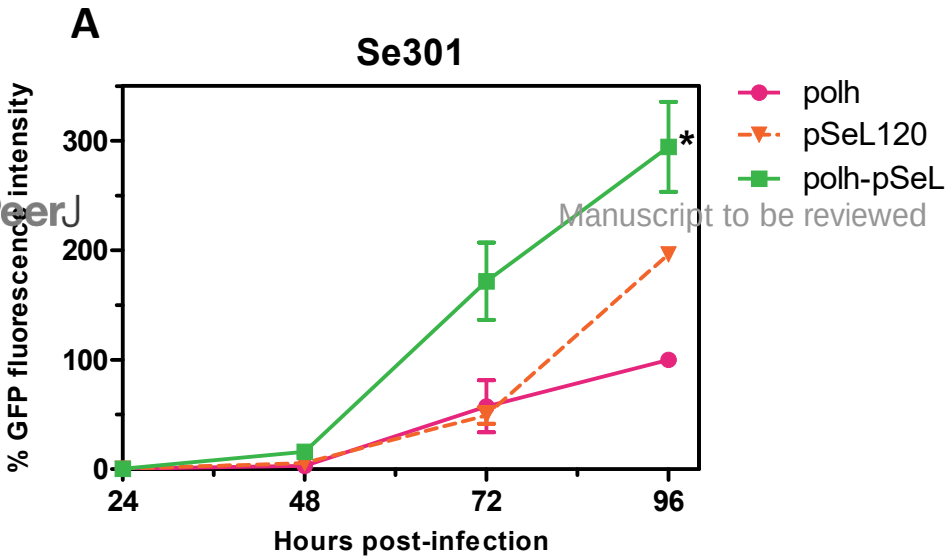
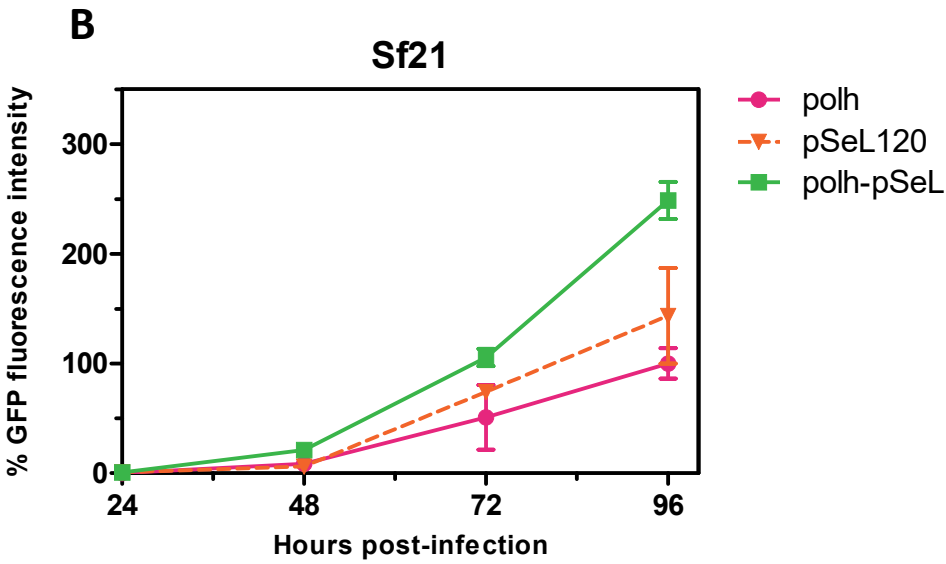

pSeS
pph
pSeL120
pph-pSeL

Figure 4(on next page)

Promoter activity of pSeL120 when combined with the pph promoter.

GFP expression, measured as the relative fluorescence intensity, in different insect cell lines infected with the different baculoviruses at a MOI of 5. The fluorescence was measured at different time points after the infection of Se301 (A), Sf21 (B), and Hi5 (C) cells. The results are expressed as the relative percentage of GFP fluorescence intensity, taken as 100% of the value corresponding to the maximum intensity obtained with the pph promoter. The values are the means of at least two independent assays. The error bars represent the standard error of the mean.





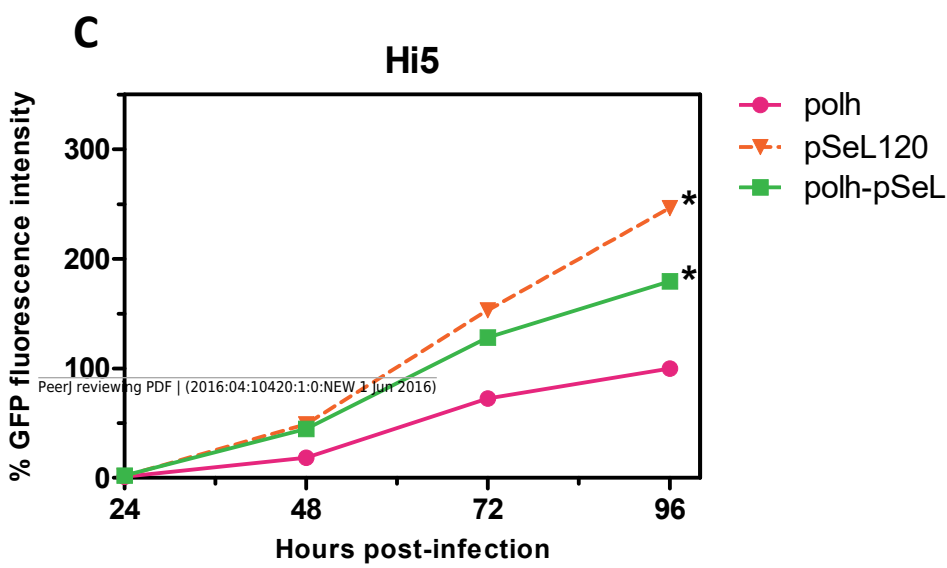
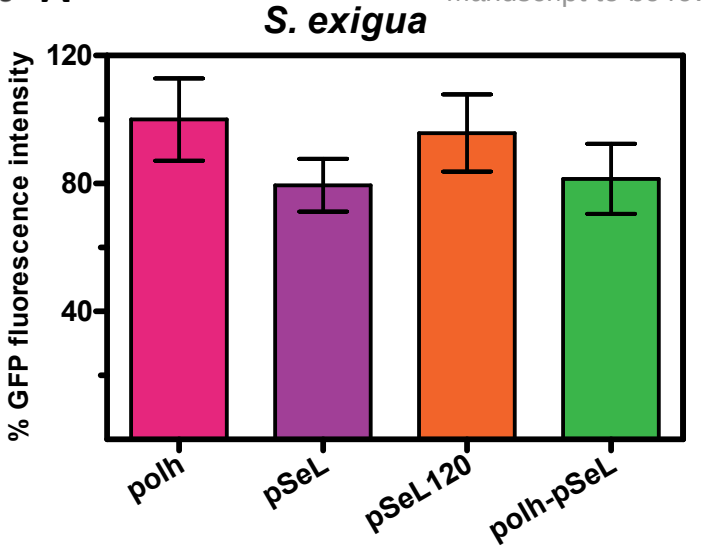


Figure 5(on next page)

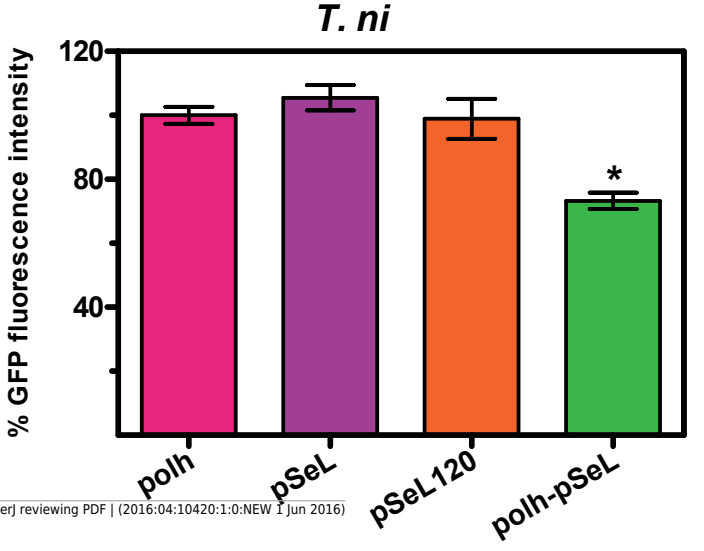
Promoter activity in insect larvae.

GFP expression, measured as relative fluorescence intensity, was obtained in insect larvae infected with the different recombinant baculoviruses. The results are expressed as the relative percentage of GFP fluorescence intensity, taken as 100% of the value corresponding to values for the control sequence with the pph promoter. The values are the means of at least two independent assays. The error bars represent the standard error of the mean.

A


S. exigua
% GFP fluorescence intensity
120
80
40
polh
pSeL
pSeL120
polh-pSeL

B


T. ni
% GFP fluorescence intensity
120
80
40
*
polh
pSeL
pSeL120
polh-pSeL