Identification of essential core machineries for baculovirus hyper expression system in insect cells

Susumu Katsuma

The University of Tokyo

Research objective

Baculoviruses are large DNA viruses that infect lepidopteran insects. Notably, they produce significant amounts of proteinaceous crystals, known as "polyhedra," which contain more than hundreds of progeny virions during the very late stages of infection. The main component of these polyhedra is the virus-encoded polyhedrin protein (POLH), which constitutes up to 50% of the total protein in the infected cells. Such a system, in which a single protein accounts for a significant portion of the total cellular protein, is unique to the baculovirus-insect cell system among multicellular eukaryotes. This high expression of the *polyhedrin* gene (*polh*) led to the development of the "Baculovirus Expression Vector System (BEVS)." Currently, BEVS is widely used to produce virus-like particle (VLP) vaccines, veterinary medicines, and reagents. Despite this, the mechanisms by which baculoviruses ultimately produce large quantities of polyhedra remain largely unexplained, even approximately 40 years after BEVS development. This study aimed to elucidate the core mechanisms of BEVS by identifying the *cis*-acting sequences and *trans*-acting factors that support the high expression of POLH and to develop an industrially applicable, virus-free foreign gene expression system that mimics these mechanisms.

Methods

In baculovirus-infected cells, *polh* is selectively transcribed during the late stages of infection. With the development and improvement of BEVS, the *polh* promoter and transcriptional regulatory regions, including enhancers, have been identified. Consequently, it was determined that an eight-base sequence, TAAGTATT, which includes the common late promoter sequence TAAG of baculovirus-encoded RNA polymerase, is essential for the transcription of *polh*[1]. Additionally, a sequence of approximately 50 bases up to the methionine codon (referred to as the burst sequence) has been found to be essential. It is also known that the BEVS product increases when expressed in fusion with the *polh* sequence (corresponding to approximately 20 amino acids) following the methionine. However, despite *polh* having the same baculoviral RNA polymerase recognition sequence as many other late genes, a unified understanding of why *polh* alone exhibits significantly higher transcript levels has not yet been achieved. Furthermore, this information is primarily based on research conducted up to the

1990s, and because of the limitations of molecular biology techniques at that time, the precise identification of *cis*-regulatory sequences has not been performed. In this study, we used the Bombyx mori nucleopolyhedrovirus (BmNPV), which infects the silkworm *Bombyx mori*, to accurately identify *cis*-regulatory sequences in the BmNPV genome that contribute to the transcription, translation, and stability of *polh*. This will be achieved using methods that utilize endogenous *polh* to measure POLH and *polh* expression and methods that use luciferase as a reporter.

Results

To identify *cis*-regulatory elements within the coding sequence of *polh*, we synthesized a codon-optimized version of *polh* (*co-polh*) for the host *B. mori* and replaced wild-type *polh* (*wt-polh*) in BmNPV to generate T3-CO. The *co-polh* sequence was 81% identical to *wt-polh* at the nucleotide level (100% identical at the amino acid level), with substitutions distributed throughout the gene.

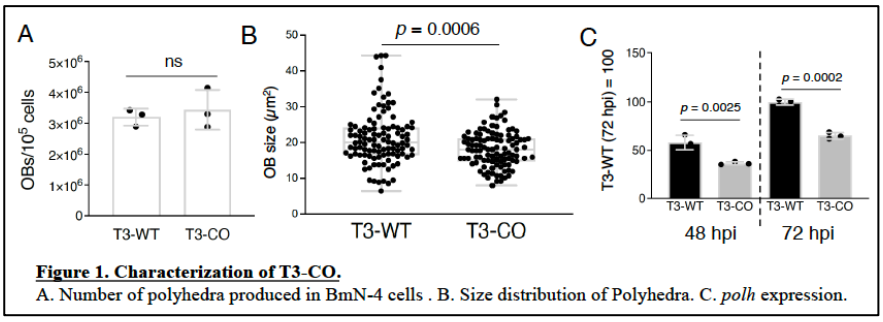

Figure 1. Characterization of T3-CO.
A. Number of polyhedra produced in BmN-4 cells . B. Size distribution of Polyhedra. C. *polh* expression.

Comparing the properties of the control virus expressing *wt-polh* (T3-WT) and T3-CO, we found that, while the number of polyhedra remained unchanged, their sizes were reduced (Fig. 1A, 1B). SDS-PAGE and RT-qPCR results indicated that this reduction was due to decreased mRNA levels rather than decreased translation efficiency (Fig. 1C). We then generated BmNPV with the chimeric genes of *wt-polh* and *co-polh* (Fig. 2A) and investigated *polh* mRNA levels, revealing that the first 30 nucleotides (cds-30nt) of *polh* contributed to mRNA abundance. Specifically, the presence of cds-30nt in *wt-polh* maintained high levels of *polh* mRNA (Fig. 2B). Reporter assays using luciferase showed that adding *polh* cds increased activity, peaking with the addition of cds-30nt and decreasing with longer sequences (Fig. 3A, 3B). There are five nucleotide substitutions within the cds-30nt between

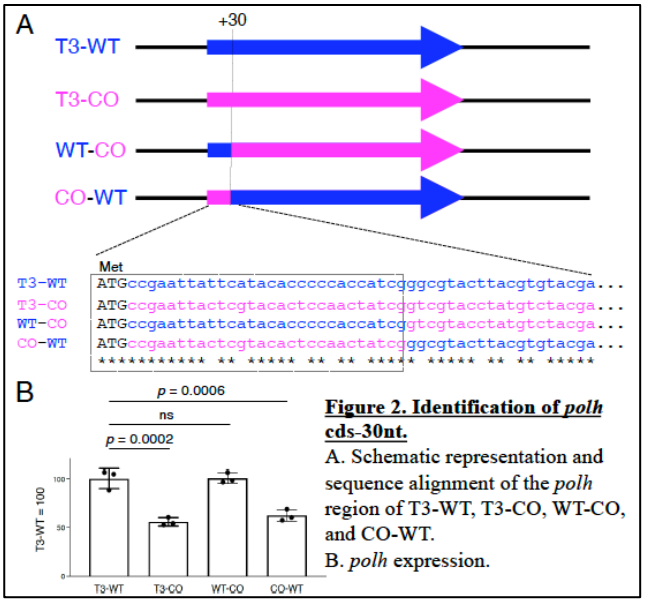

Figure 2. Identification of *polh* cds-30nt.
A. Schematic representation and sequence alignment of the *polh* region of T3-WT, T3-CO, WT-CO, and CO-WT.
B. *polh* expression.

wt-polh and *co-polh*, four of which are conserved in AcMNPV, another virus frequently used in BEVS.

During the generation of the control virus with *wt-polh*, we obtained a virus that produced very small polyhedra despite having *wt-polh*. This virus contained 19 base-long vector-derived sequences between the burst sequence and the start methionine. Shortening this sequence improved the polyhedron size, and complete removal restored it to wild-type levels. Similar phenomena were observed in reporter viruses and were attributed to differences in mRNA levels depending on the *polh* promoter.

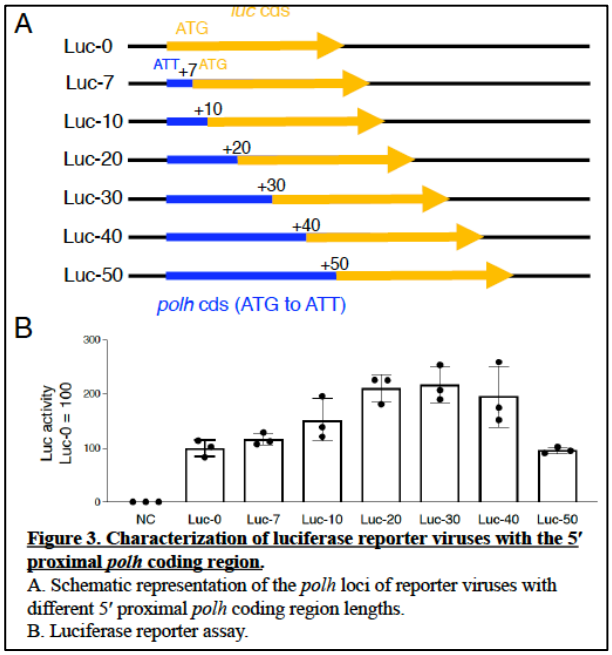

Figure 3. Characterization of luciferase reporter viruses with the 5′ proximal *polh* coding region.
A. Schematic representation of the *polh* loci of reporter viruses with different 5′ proximal *polh* coding region lengths.
B. Luciferase reporter assay.

Conclusion

We identified critical *cis*-regulatory elements within the cds-30nt of *polh* that influence *polh* mRNA levels. Future research should explore the relationship between this sequence and *trans*-acting factors as well as *polh* mRNA stability. Additionally, we found that minimizing the distance between the burst sequence and the starting methionine was crucial when using BEVS.

References

1. Rankin, C., Ooi, B.G., and Miller, L.K. (1988) Eight base pairs encompassing the transcriptional start point are the major determinant for baculovirus *polyhedrin* gene expression. *Gene* **70:**39-49.

2. Katsuma, S., and Matsuda-Imai, N. (2024) Codon optimization-based whole-gene scanning identifies hidden nucleotides essential for Bombyx mori nucleopolyhedrovirus *polyhedrin* hyperexpression. *J. Mol. Biol.* **436:**168595.

3. Katsuma, S., and Matsuda-Imai, N. (2023) A seamless connection from the burst sequence to the start codon is essential for *polyhedrin* hyperexpression in alphabaculoviruses. *Biochem. Biophys. Res. Commun.* **679**:1-5.