Baculovirus diversity and replication

George F. Rohrmann
Department of Microbiology, Oregon State University, USA

The genomes of the *Lymantria dispar* multinucleocapsid nucleopolyhedrovirus (LdMNPV) and the *Orgyia pseudotsugata* MNPV were sequenced and although they both have similar G + C contents (57.5 and 55 %, respectively), they differed in genome size (161 vs 132 kb, respectively) and gene content. Both genomes contain a similar set of genes involved in deoxyribonucleotide metabolism not found in the *Autographa californica* MNPV (AcMNPV). These include genes with homology to *dutpase* and both large and small subunits of *ribonucleotide reductase*. Over 9 % of the LdMNPV genome was occupied by 16 repeated genes related to AcMNPV ORF-2. In addition, LdMNPV encodes a *ligase* gene homolog and a second gene with helicase domains not found in OpMNPV. Neither the *helicase* nor the *ligase* gene appeared to be essential or stimulatory for DNA replication in transient assays suggesting that they may be involved in DNA recombination or repair. Both viruses contain multiple homologous regions (*hrs*) interspersed throughout their genomes that share both sequence homology and in some instances genome position with their counterparts in AcMNPV. Homologs of several genes that have been reported to play important roles in the life cycle of OpMNPV and other baculoviruses are not present in the LdMNPV genome. These include *ie*-2, a transcriptional transactivator, and *gp64*, a major envelope fusion protein of the non-occluded form of the virus. Investigations of LdMNPV infected cells suggest that it produces an acid activated envelope fusion protein during its replicative cycle. Computer analysis identified a single open reading frame in the LdMNPV with features predicted for an envelope fusion protein including an N-terminal signal sequence and a transmembrane domain.

Introduction

The Baculoviridae, a diverse family of viruses with large double-stranded, circular, DNA genomes, are pathogenic for invertebrates, particularly insects of the order Lepidoptera. Several hundred different insect species are reported to be infected by baculoviruses and viral diversity is reflected in their host specificity, the pathology of their infection cycle, and major differences in genome size, and G + C content. To date the genome sequences of four baculoviruses have been described including *Autographa californica* multinucleocapsid nucleopolyhedrovirus (AcMNPV),[1] *Bombyx mori* NPV (BmNPV) (a close relative to AcMNPV),[2] *Orgyia pseudotsugata* MNPV (OpMNPV)[3] and *Lymantria dispar* MNPV (LdMNPV).[4] In addiiton, the sequence of a granulosis virus pathogenic for *Xestia c-nigrum* (XcGV) has been made available to the author (Hayakawa, Ko, and Maeda, unpublished). These viruses have genome sizes of 128 to 178 kb, and whereas AcMNPV, BmNPV, and XcGV have G + C contents of about 41 %, OpMNPV and LdMNPV have a G + C content which is significantly higher at over 56 %.

Genome organization: OpMNPV

Of the 152 OpMNPV ORFs characterize, 125 had homologs in AcMNPV which collectively account for about 80 % (100 kb) of the genomes. The average amino acid identity between homologous ORFs of these two viruses was 56 %. The most conserved ORF is polyhedrin (ORF3) (89 % amino acid sequence identity) followed by the viral ubiquitin homolog (ORF25) (84 %), although the latter has a 3' extension of 16 amino acids in OpMNPV that is not included in this calculation. Other well-conserved ORFs show about 80 % identity and include ORF20 (AcMNPV ORF22); ORF79 (AcMNPV ORF76); *vlf-1*, (ORF80, AcMNPV ORF77); *chitinase*, (ORF124, AcMNPV ORF126); *cathepsin*, (ORF125, AcMNPV ORF127); *gp64*, (ORF126, AcMNPV ORF127); and *p74*, (ORF134, AcMNPV ORF138). In addition to about 50 ORFs that are not shared between OpMNPV and AcMNPV, several homologs showed low levels of amino acid sequence identity, such as ORF87, which has 24 % identity with AcMNPV ORF85. All the genes required for transient DNA replication and late gene transcription[5,6] are present in both genomes except the anti-apoptotic genes (OpMNPV *iap-3* [ORF35] and AcMNPV *p35* [ORF135]), which are stimulatory in replication assays, and *hcf-1*. In addition, most of the genes implicated as encoding virion structural proteins[7] are also present.

The relatedness between many OpMNPV and AcMNPV ORFs is also reflected in a similar organization over much of the two genomes. However, there are two extensive inversions present in the genomes.[3]

Genome organization: LdMNPV

Analysis of the sequence of LdMNPV led to the identification of 163 putative ORFs and 13 homologous regions (*hrs*). We identified homologs of 94 AcMNPV ORFs and when repeated ORFs are included, 114 of the LdMNPV ORFs are related to AcMNPV ORFs. Although there is extensive genetic relatedness, the overall order of LdMNPV genes relative to AcMNPV is limited. About 55 of the LdMNPV ORFs are organized into two regions which, although they contain a number of inversions, deletions, and insertions, have a pattern that is somewhat similar to AcMNPV. These groups include 10 ORFs homologous to AcMNPV ORFs 139–148 (LdMNPV ORFs14–23) and 45 ORFs organized similarly to the AcMNPV ORF 26–113 region (LdMNPV ORFs36–109).

The average amino acid identity of the AcMNPV homologs in LdMNPV was about 41 %. The most conserved ORFs were polyhedrin and an AcMNPV *ORF2* homolog (called Ld-*bro*-n [Ldorf153]) both of which show 82 % amino acid identity. Three genes likely to be involved in late/very late transcription (*lef-8*, *lef-9* and *vlf-1*; LdMNPV ORFs 51, 64, and 86,

respectively) were also highly conserved (60–70 % identical). The *chitinase* (Ld-ORF70) and *cathepsin* (Ld-ORF78) homologs were conserved (over 65 % identical) along with *p74* (Ld-ORF27), 61 %; *ubiquitin* (Ld-ORF43), 71 % (note: the LdMNPV ORF is almost twice as long as that of AcMNPV; *sod* (Ld-ORF145), 75 %; and *ctl* (Ld-ORF149), 77 %. Two additional copies of AcMNPV *ORF2* homologs, ld-*bro*-j and -p, (Ld-ORFs 114 and 161, respectively) at about 60 % identity are also among the most conserved genes. The 41 % overall identity of LdMNPV and AcMNPV ORFs is less than reported for AcMNPV and OpMNPV (56 %)[3] and with the exception of the three AcMNPV ORF2 homologs, the genes most conserved in LdMNPV were also well-conserved in OpMNPV. With the exception of the ORFs encoding the ribonucleotide reductase subunits which are highly conserved between LdMNPV and OpMNPV, but lacking in AcMNPV, the LdMNPV and OpMNPV demonstrated similar amino acid sequence identities as LdMNPV and AcMNPV ORFs. This suggests that the common ancestor of AcMNPV and OpMNPV diverged from LdMNPV prior to the separation of AcMNPV and OpMNPV.

Genome content

The LdMNPV genome is approximately 30 kb larger than the three other sequenced baculovirus genomes (AcMNPV, BmNPV, OpMNPV). The sources of the DNA that contribute to the large size of the LdMNPV genome are summarized in Fig. 1. The LdMNPV genome contains 163 putative ORFs. These include homologs of 94 AcMNPV genes. Exclusive of the repeated *bro* genes, these ORFs represent about 90 kb of DNA. A major source of additional genetic information results from gene amplification. The 16 copies of AcMNPV ORF2 homologs called *bro* genes contribute about 14.5 kb or 9 % of the DNA in the LdMNPV genome. Genes homologous to those from organisms other than AcMNPV are also major contributors to the LdMNPV genome. Genes encoding proteins such as virus enhancing factor (VEF), ribonucleotide reductase subunits, and *dutpase* contribute another 16 kb (10 %) to LdMNPV. The LdMNPV genome contains about 35 ORFs with no homologs in the database. These require about 12 % of the coding capacity. In addition, LdMNPV *hrs* comprise another 4.6 kb (2.8 %) of DNA. The remaining DNA (about 17.5 kb), if divided by the number of ORFs (163), results in about 106 bp of DNA/gene for 5' and 3' regulatory

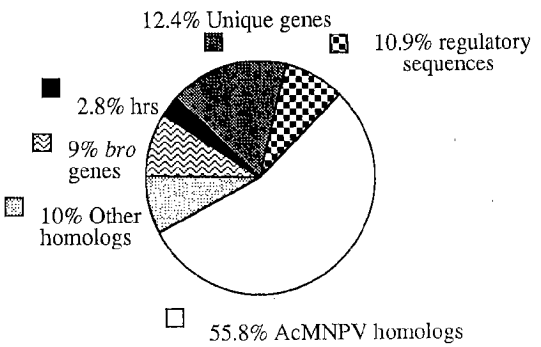

Fig. 1. The sources of LdMNPV genome content related to AcMNPV.

sequences. This is about twice the amount present in AcMNPV. Therefore, the large size of the LdMNPV genome relative to AcMNPV, OpMNPV and BmNPV can be accounted for by genes derived from other organisms, gene amplification, expanded intergenic regions, and unique genes either recruited by LdMNPV, or lost by the other viruses.

AcMNPV ORFs with no homologs in the LdMNPV genome

Homologs of 61 AcMNPV genes are not present in LdMNPV. Of these, 22 are also not present in the OpMNPV genome. Particularly evident is the lack of two transactivators of early transcription, *ie-2* and *pe38*. *Ie-2* is not essential for BmNPV viability in cell culture. When it is deleted, viral growth and late gene expression appear unaltered, but DNA replication is reduced.[8]

A major unexpected finding of this comparison is the absence of an envelope fusion protein gene homologous to the AcMNPV/OpMNPV *gp64* genes. In these viruses, GP64 appears to be a critical protein for the spread of infection from cell-to-cell. Production of an AcMNPV *gp64* minus mutant (using cells expressing *gp64*), led to the conclusion that *gp64* was required both for cell to cell virus spread and for spread of the infection from the insect gut to the haemocoel.[9] This suggests that LdMNPV either has other genes that are involved in membrane fusion, or that it utilizes a different mechanism for cell entry. The elucidation of LdMNPV cell entry mechanisms may contribute significantly to understanding the different approaches baculoviruses have evolved to infect cells.

In order to determine if LdMNPV encodes a protein that could function as a replacement for *gp64*, the genome was scanned for gene products containing N-terminal signal and transmembrane domains that are indicative of transmembrane receptor-like proteins. A single candidate gene was identified (LdMNPV-ORF130). This gene has a low level of homology to AcMNPV ORF23. Similar analyses of the AcMNPV and OpMNPV genomes identified both the *gp64* proteins and AcMNPV ORF23 (OpMNPV ORF21) as the only gene products predicted to contain both these domains. The lack of *gp64* in XcGV prompted a comparison between XcGV orfs and those of LdMNPV. In this analysis, an orf, Xc27, was identified that showed a higher degree of identity to LdMNPV ORF130 than it did to AcMNPV ORF23. In addition, computer analysis predicted that it had both an N-terminal signal sequence and a transmembrane domain.

OpMNPV and LdMNPV genes homologous to genes not found in AcMNPV or BmNPV

Although, a similar set of genes required for DNA replication has been identified by transient DNA replication assays in both AcMNPV, OpMNPV,[6,10–12] and critical components of this process such as ligases and topoisomerases have not been found in baculovirus genomes, and therefore, are likely supplied by the host cell. Two ORFs were identified in the LdMNPV genome with homology to genes involved in DNA replication in other systems. These include a *dna ligase* homolog (ORF22) which has about 35 % amino acid sequence identity to the *ligase* of vaccinia virus. A similar gene was

found in the XcGV genome (Table 1). The LdMNPV ligase gene was cloned and expressed in a bacterial system, and the resulting purified protein product had catalytic properties characteristic of a type III DNA ligase.[12] In addition, an ORF (ORF50) was identified with 31 % amino acid sequence identity to a yeast mitochondrial *helicase*, called *pif1*.[13] We call this orf *helicase b*. A homolog of *helicas eb* is also present in XcGV (Table 1).

Table 1. Amino acid sequence identity of selected baculovirus genes*

Organism	Gene	length (aa)	% identity
Ld : Xc	helicaseb	455 : 461	50 %
Ld : Xc	ligase	548 : 527	28 %
Ld : Xc	efp	626 : 599	26 %
Ld : Op	RR1	596 : 594	84 %
Ld : Se	RR1	360 : 348	28 %
Ld : Op	RR2	349 : 360	83 %
Ld : Ld	RR2a,b	360 : 348	22 %
Ld : GH	RR2b	348 : 386	72 %
Ld : Op	dutpase	147 : 316	29 %

* The following are abbreviations used in this table: Ld, LdMNPV; Op, OpMNPV; Xc, XcG[unpublished data]; GH, golden hamster; efp, putative envelope fusion protein; RR1, ribonucleotide reductase large subunits; RR2, ribonucleotide reductase small subunits; RR2a,b, RR2 variants found in LdMNPV.

Cotransfecton of the LdMNPV *ligase* or *pif1* homologs either individually or together, along with the six essential replication genes did not stimulate transient DNA replication above the level seen with the six replication genes alone.[12] Since type III DNA *ligase* and *pif1* are associated with DNA repair and recombination, they may be involved in a LdMNPV DNA recombination or repair system.

Genes involved in nucleotide metabolism

Several ORFs that are homologous to genes involved in nucleotide metabolism are present in both the OpMNPV and LdMNPV genome. These include a homolog of the large small subunits of ribonucleotide reductase (*rnr*) and a gene with homology to *dutpase*. An additional homolog of the *rnr* small subunit is also found in LdMNPV. Although homologs of these genes are not found in AcMNPV, all three are present in the OpMNPV genome.[3] Two adjacent genes (ORFs 147 and 148) encode predicted ribonucleotide reductase subunits in the LdMNPV genome. Both genes showed very high homology to the large (R1) and small (R2) subunits from OpMNPV[3] (84 and 83 % amino acid sequence identity, respectively), and like the OpMNPV RNR ORFs showed only low levels of identity to RNR ORFs from other organisms. In addition, R1, but not R2 homologs have been reported for two additional baculoviruses, *S. exigua* MNPV and *S. littoralis* MNPV.[14] R1 from these viruses show 58–68 % amino acid sequence identity with other eukaryotic RNR R1submunits and are distantly related to OpMNPV, HSV, and *E. coli* R1 (25–27 % identity) (Table 1). The R2 subunit from OpMNPV was even less well-conserved than R1 showing only 10–16 % sequence identity with other R2 peptides. In addition, a second R2 homolog (ORF120) was identified elsewhere in the LdMNPV genome. This ORF shows only limited amino acid sequence identity to both the R2 from OpMNPV and the other LdMNPV R2 (24 and 22 %, respectively) (Table 1). In contrast, the second LdMNPV R2 ORF shows about 70 % amino acid sequence identity to R2's of vaccinia virus and golden hamster (Table 1). This indicates that there are two distinct lineages of baculovirus *rnr* genes including one closely related to other eukaryotic genes and a second category of distantly related genes.

Dutpase is a critical enzyme because it converts dUTP to dUMP thereby preventing dUTP from incorporation into DNA where it is mutagenic. In addition, dUMP is a precursor for the synthesis of dTTP.[15] In OpMNPV, *dutpase* is fused to a homolog of Ld-ORF138 to form a reading frame of 319 amino acids. These ORFs are separate and in different locations in LdMNPV. The LdMNPV *dutpase* homolog shows 29 % amino acid sequence identity to the C-terminal 115 amino acids of the OpMNPV ORF. It is similar in size to that from vaccinia virus (149 vs 148 amino acids, respectively) to which it shows 45 % identity. Therefore this gene appears to be more closely related to those from the poxviridae than to the OpMNPV ORF. A summary of the relatedness of selected genes is indicated in Table 1.

Repeated genes

A striking feature of the LdMNPV genome is the presence of 16 ORFs that are related to AcMNPV ORF2. We have named these ORFs baculovirus repeated ORFs (BRO) and have sequentially lettered them in the genome. Whereas five copies of *bro* are present in the BmNPV genome,[2] in AcMNPV there is a single copy, and in OpMNPV there is a small truncated 88-amino acid representative and two other smaller related.[3] Some of the *bro* genes are located in contiguous clusters. Two clusters are located near *hrs* indicating that they may have been duplicated along with the *hrs*. Most these genes share a related core sequence, and they demonstrate differing degrees of similarity in other domains. Outside of the shared core region, members of the *bro* gene family were separated into four groups based on the relationship of different domains. Group I, the largest group, appears to have three subgroups of related genes. Several sets of LdMNPV *bro* genes are related and appear to be the result of recent gene duplication. These include Ld-*bro*-c (orf71) and -d (orf72) which are distinguished by their length and a common 3' region lacking in the other bro's. Also Ld-*bro*-a (orf32), -l (orf146), and -o (orf154) likely evolved from rela-

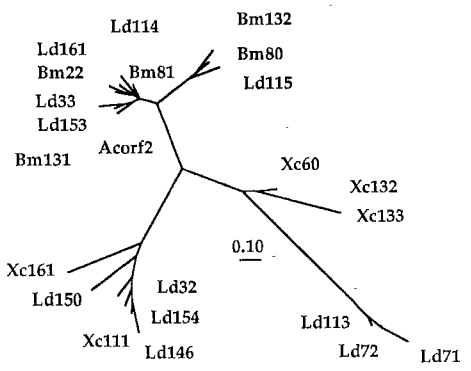

Fig. 2. An unrooted phylogenetic tree of selected baculovirus *bro* genes.

tively recent duplication events. There is a partial *bro* (*bro*-h) (orf112) located upstream of the -i (orf113), -j (orf114), -k (orf115) cluster that appears to be a duplication of part of bro-i (orf113). The Ac-*bro* (ORF2), Bm-*bro*-d (orf131) and Ld-*bro*-n (orf153) are the most highly conserved genes between these viruses (82 % identical in amino acid sequence) suggesting that they may have similar functions in the infection cycle. An unrooted phylogenetic tree for selected members of this family is shown in Fig. 2.

Repeated regions

Homologous regions

A novel feature of many baculovirus genomes is the presence of homologous regions (*hrs*) that are located throughout the genome.[1,3,16-19] *Hrs* are composed of repeated sequences encompassing both direct repeats and imperfect palindromic sequences and have closely related counterparts elsewhere in the genome.

AcMNPV has eight *hrs*, whereas the OpMNPV genome has five. In these viruses each *hr* contains one to ten 30-bp imperfect palindromes within a directly repeated sequence genome.[1,3] In both AcMNPV and OpMNPV, *hrs* can act as enhancers of RNA polymerase II-mediated transcription[18-20] and can also behave as origins of DNA replication in transient replication assays.[21-23] In addition, evidence suggests that the baculoviral transactivator, *ie-1*, binds to *hr* sequences.[24]

Thirteen homologous regions (*hr* 1, 2, 3a, 3b, 3c, 4, 5, 6, 7a, 7b, 7c, 7d, 8) were identified in the LdMNPV genome that contain from 1–7 palindrome repeats for a total of 53 repeats. A number of partial repeats are also present in some *hrs*. As previously reported, different LdMNPV *hrs* may be closely related to one another (e.g. *hr* 1 and *hr* 4 are 89 % identical).[25] There are variations between individual *hr* palindromes and differences within each palindrome which confer on these structures a direction. Therefore, some clusters of *hrs* are oriented in the same direction. This directionality is reflected by the larger direct repeats within which *hr*'s are often imbedded. *Hr* 7b is limited to a single palindrome and is not embedded in a larger direct repeat. In contrast to AcMNPV *hrs* which can act as origins of replication in transient assays, it was found that the LdMNPV *hrs* had to be linked to an AT-rich sequence before they were capable of undergoing replication.[25] Similar to the *hrs* of AcMNPV and OpMNPV, LdMNPV *hrs* can act as enhancers of early gene transcription.[26] The LdMNPV *hrs* are composed of a series of 72-bp DNA repeats containing a 30-bp palindrome with MluI sites in each arm. This palindrome flanks a combined XhoI-SacI site. Except for the four variations, all nt in the palindrome are over 92 % conserved, with the variable nt being about 60–75 % conserved with the exception of the 5' G which is 42 % conserved. In addition, when compared to the AcMNPV consensus *hr* palindrome, 50 % of the sequences are identical. The similarity between LdMNPV and OpMNPV *hrs* was not so clearly evident showing only 40 % identity if single gaps are inserted. However, when a consensus profile derived from a number of baculoviruses was used to scan the LdMNPV genome, all the *hrs* in the LdMNPV genome were identified. This is further evidence that the *hrs* from different baculoviruses share a common ancestor.

The position of at least some *hrs* appears to be conserved relative to specific baculovirus genes. In particular, it was found that an *hr* is conserved immediately downstream of AcMNPV ORF83 and the homologous genes in BmNPV and OpMNPV.[3] LdMNPV *hr* 5 is located immediately downstream of LdORF91, the LdMNPV homolog of AcMNPV ORF83. In addition, immediately downstream of this *hr* is located LdORF92 the homolog of AcMNPV vp39-capsid (AcMNPV ORF89). Since no homologs of the AcMNPV ORFs (84–88) are present in the LdMNPV genome, this indicates that the position of LdMNPV *hr* 4 is conserved relative to both the upstream and downstream genes. The regions downstream of this *hr* in OpMNPV and BmNPV are similar with the homologs of ORFs 84–86 missing in BmNPV and ORFs 84 and 86 missing in OpMNPV.

Given that *hr*'s share higher similarity within a virus strain than any given positionally conserved cluster of *hr*'s between species, the amplification process is tightly linked to functional conservation. Within a virus species the *hr*'s may co-evolve with the proteins they interact with, such as the transcriptional activator, *ie*-1. The evidence that *hrs* arose with a progenitor baculovirus and are maintained in diverse genomes further indicates that they play a fundamental role in the viral life cycle and support data suggesting that they are transcriptional enhancers and may also serve as origins of DNA replication *in vivo* (reviewed in Ref. 27).

References

1) M. D. Ayres, S. C. Howard, J. Kuzio, M. Lopez-Ferber, and R. D. Possee: Virology **202**, 586 (1994).
2) S. Gomi, K. Majima, and S. Maeda: J. Gen. Virol. **80**, 1323 (1999).
3) C. H. Ahrens, R. Russell, C. J. Funk, J. T. Evans, S. H. Harwood, and G. F. Rohrmann: Virology **229**, 381 (1997).
4) J. Kuzio, M. N. Pearson, S. H. Harwood, C. J. Funk, J. T. Evans, J. Slavicek, and G. F. Rohrmann: Virology **253**, 17 (1999).
5) M. Kool, C. Ahrens, R. W. Goldbach, G. F. Rohrmann, and J. M. Vlak: Proc. Natl. Acad. Sci. USA **91**, 11212 (1994).
6) A. Lu and L. K. Miller: J. Virol. **69**, 975 (1995).
7) C. J. Funk, S. Braunagel, and G. F. Rohrmann: in *The Baculoviruses*, edited by L. K. Miller (Plenum Press, New York, 1997), p. 468.
8) S. Gomi, C. E. Zhou, W. Yih, K. Majima, and S. Maeda: Virology **230**, 35 (1997).
9) S. A. Monsma, A. G. P. Oomens, and G. W. Blissard: J. Virol. **70**, 4607 (1996).
10) C. A. Ahrens, D. J. Leisy, and G. F. Rohrmann: *DNA Replication in Eukaryotic Cells*, edited by M. De Pamphilis (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1996).
11) M. Kool, C. Ahrens, R. W. Goldbach, G. F. Rohrmann, and J. M. Vlak: Proc. Natl. Acad. Sci. USA **91**, 11212 (1994).
12) M. N. Pearson and G. F. Rohrmann: J. Virol. **72**, 9142 (1998).
13) F. Foury and A. Lahaye: EMBO J. **6**, 1441 (1987).
14) E. A. van Strien, O. Faktor, Z. H. Hu, D. Zuidema, R. W. Goldbach, and J. M. Vlak: J. Gen. Virol. **78**, 2365 (1997).
15) C. K. Mathews and K. E. van Holde: *Biochemistry* (The Benjamin/Cummings Publishing Co., Inc., Menlo Park, CA, 1995).
16) M. A. Cochran and P. Faulkner: J. Virol. **45**, 961 (1983).
17) A. Garcia-Maruniak, O. H. O. Pavan, and J. E. Maruniak: Virus Res. **41**, 123 (1996).
18) D. A. Theilmann and S. Stewart: Virology **187**, 97 (1992).
19) W.-D. Xie, B. Arif, P. Dobos, and P. J. Krell: Virology **209**, 409 (1995).
20) L. A. Guarino and M. D. Summers: J. Virol. **60**, 215 (1986).

21) C. H. Ahrens, M. N. Pearson, and G. F. Rohrmann: Virology **207**, 572 (1995).
22) M. Kool, C. H. Ahrens, J. M. Vlak, and G. F. Rohrmann: J. Gen. Virol. **76**, 2103 (1995).
23) M. N. Pearson, R. M. Bjornson, G. D. Pearson, and G. F. Rohrmann: Science **257**, 1382 (1992).
24) J. Choi and L. A. Guarino: J. Virol. **69**, 4548 (1995).
25) M. N. Pearson and G. F. Rohrmann: J. Virol. **69**, 213 (1995).
26) M. N. Pearson and G. F. Rohrmann: Virology **235**, 153 (1997).
27) A. Lu, P. Krell, J. M. Vlak, and G. F. Rohrmann: in *The Baculoviruses*, edited by L. K. Miller (Plenum Press, New York, 1997).