Virology **253,** 17–34 (1999)
Article ID viro.1998.9469, available online at http://www.idealibrary.com on IDEAL®

Sequence and Analysis of the Genome of a Baculovirus Pathogenic for *Lymantria dispar*

John Kuzio,* Margot N. Pearson,† Steve H. Harwood,† C. Joel Funk,† Jay T. Evans,†
James M. Slavicek,‡ and George F. Rohrmann†,[1]

*National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894;
†Department of Microbiology, 220 Nash Hall, Oregon State University, Corvallis, Oregon 97331-3804;
and ‡USDA Forest Service, Northeastern Research Station, Forestry Sciences Laboratory,
359 Main Road, Delaware, Ohio 43015

Received August 4, 1998; returned to author for revision August 27, 1998; accepted October 12, 1998

The genome of the *Lymantria dispar* multinucleocapsid nucleopolyhedrovirus (LdMNPV) was sequenced and analyzed. It is composed of 161,046 bases with a G + C content of 57.5% and contains 163 putative open reading frames (ORFs) of ≥150 nucleotides. Homologs were found to 95 of the 155 genes predicted for the *Autographa californica* MNPV (AcMNPV) genome. More than 9% of the LdMNPV genome was occupied by 16 repeated genes related to AcMNPV ORF2. Readily identifiable homologs of several genes that have been reported to play important roles in the AcMNPV life cycle are not present; these include *ie-2,* a transcriptional transactivator, and *gp64*, a major envelope glycoprotein of the nonoccluded form of the virus. A number of genes lacking in AcMNPV but present in other baculoviruses were identified; these include two viral enhancing factor homologs, a second copy of a conotoxin-like gene, and a *dutpase* homolog. Although a single gene predicted to encode a large subunit of ribonucleotide reductase was found, two different copies of the small subunit gene were present. In addition, homologs of genes not previously reported for baculoviruses were identified, including a predicted protein with homology to DNA ligases and another that has motifs most closely related to a yeast mitochondrial helicase. Thirteen homologous regions (*hr*s) containing 54 repeated sequences that include 30-bp imperfect palindromes were identified. The imperfect palindromes are related to those from other baculoviruses.

Key Words: baculovirus; nuclear polyhedrosis virus; genome sequence.

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. The most extensively characterized baculoviruses are *Autographa californica* multinucleocapsid nucleopolyhedrovirus (AcMNPV) (Ayres *et al.,* 1994), *Bombyx mori* NPV (BmNPV) (GenBank accession no. L33180, a close relative to AcMNPV), and *Orgyia pseudotsugata* MNPV (OpMNPV) (Ahrens *et al.,* 1997). These three viruses have genome sizes of 128–134 kb, and although AcMNPV and BmNPV have G + C contents of ~40%, OpMNPV has a G + C content that is significantly higher (56%).

Lymantria dispar (the gypsy moth) is infected by a baculovirus (LdMNPV) that has a genome size of >160 kb (Riegel *et al.,* 1994; Smith *et al.,* 1988). Because of its large genome size and its potential importance in *L. dispar* control programs, we have undertaken a project to characterize the LdMNPV genome. This report describes the sequence and organization of the LdMNPV genome and compares it with sequence data from other baculoviruses.

[1] To whom reprint requests should be addressed. Fax: (541) 737-0496. E-mail: rohrmann@bcc.orst.edu.

RESULTS AND DISCUSSION

Genome Organization

The LdMNPV genome is 161,046 bp, which is significantly larger than those reported for AcMNPV, BmNPV, and OpMNPV (~130 kb). In addition, the G + C content (57.5%) is similar to that of OpMNPV (55%) but higher than those of AcMNPV and BmNPV (both ~40%) (see summary below). Because of the lack of overall colinearity with the AcMNPV genome, we designated the 0 point in the genome to be the beginning of a run of Ts located just upstream of a homologous region (*hr*1). This results in the polyhedrin gene designated ORF1. An analysis of the sequence led to the identification of 163 putative ORFs (Fig. 1, Table 1). Also present are 13 homologous regions (*hr*s) (see description below). The relative locations of the ORFs and *hr*s are diagramed in Figs. 1 and 2, and their names, coordinates, and other details are shown in Table 1. We identified homologs of 95 AcMNPV ORFs, and when repeated ORFs (see below) are included, 114 of the LdMNPV ORFs are related to AcMNPV ORFs. Although there is extensive genetic homology, the overall order of LdMNPV genes relative to AcMNPV is

0042-6822/99 $30.00

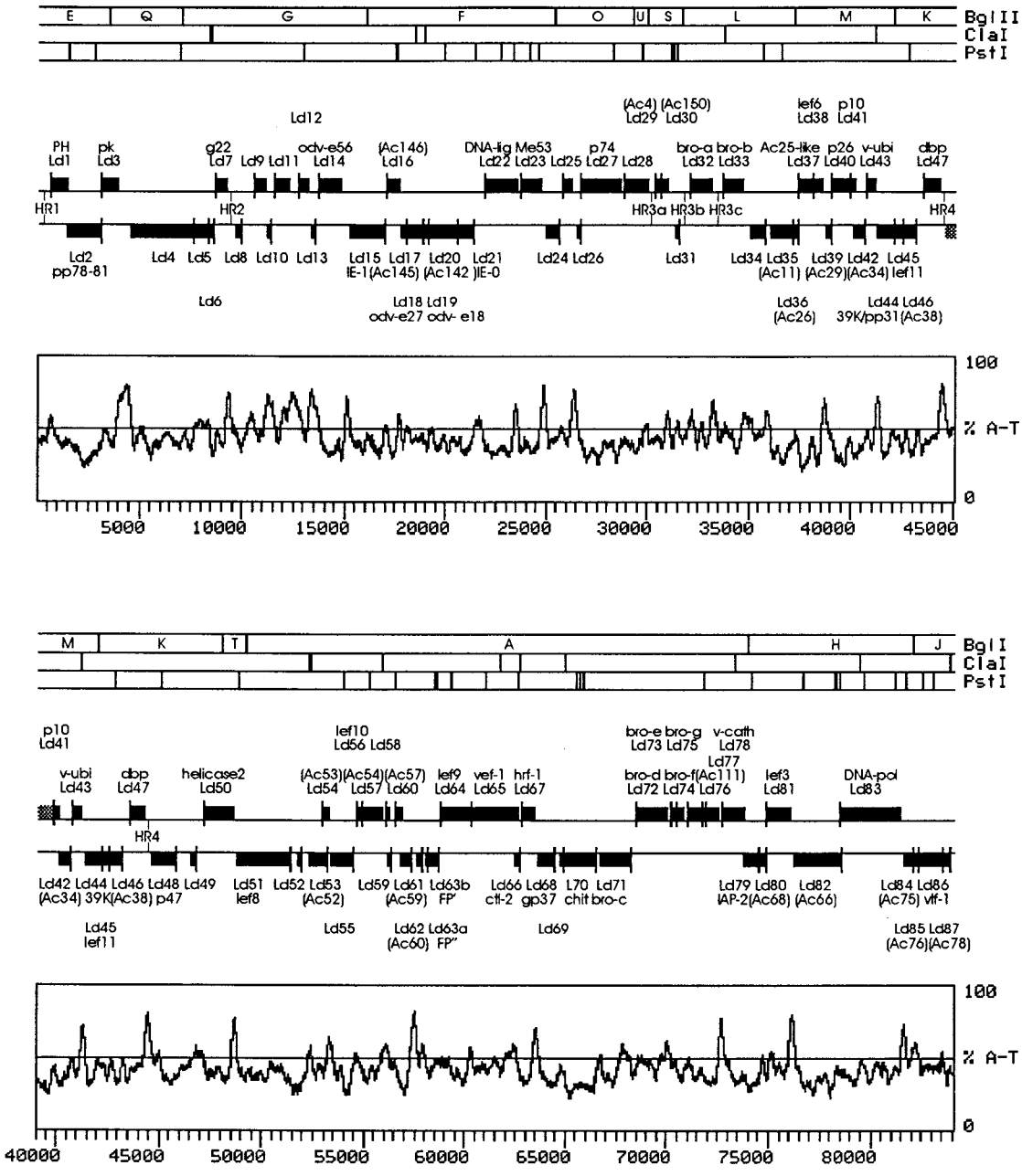


FIG. 1. Organization of the LdMNPV genome. The circular LdMNPV genome is shown in an overlapping linear format. The top part of each panel represents restriction endonuclease recognition sites in the virus genome. In the middle part of each panel, ORFs are shown as black rectangles proportional to the size of the ORF beginning with the polyhedrin gene designated as ORF1. The names of characterized gene homologs are also indicated. Homologs of an AcMNPV gene that have not been given generally accepted names are indicated with the AcMNPV ORF number (Ahrens *et al.,* 1997; Ayres *et al.,* 1994). ORFs in the 5′→3′ direction are indicated on the top line, and the bottom line indicates ORFs in the 3′→5′ direction; *hr* locations are shown between the two lines. The bottom part of each panel indicates the percent adenine and thymine (A-T) for that region. The units at the bottom are base pairs.

limited. About 55 of the LdMNPV ORFs are organized into two regions that, although they contain a number of inversions, deletions, and insertions, have a pattern that is somewhat similar to that of AcMNPV. These groups include 9 ORFs homologous to AcMNPV ORF139–ORF148 (LdMNPV ORF14–ORF23) and 45 ORFs organized similarly to the AcMNPV ORF26–ORF113 region (LdMNPV ORF36–ORF109). These data are shown in column 2 of Table 1.

The average amino acid identity of the AcMNPV homologs in LdMNPV was ~41% (Table 1). The most conserved ORFs were polyhedrin and an AcMNPV *ORF2* homolog (called Ld-*bro*-n, see below), 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 ORF51, ORF64, and ORF86, respectively) were also highly conserved (60–70% identical). The *chitinase* (Ld-ORF70) and *cathepsin* (Ld-ORF78) homologs were conserved

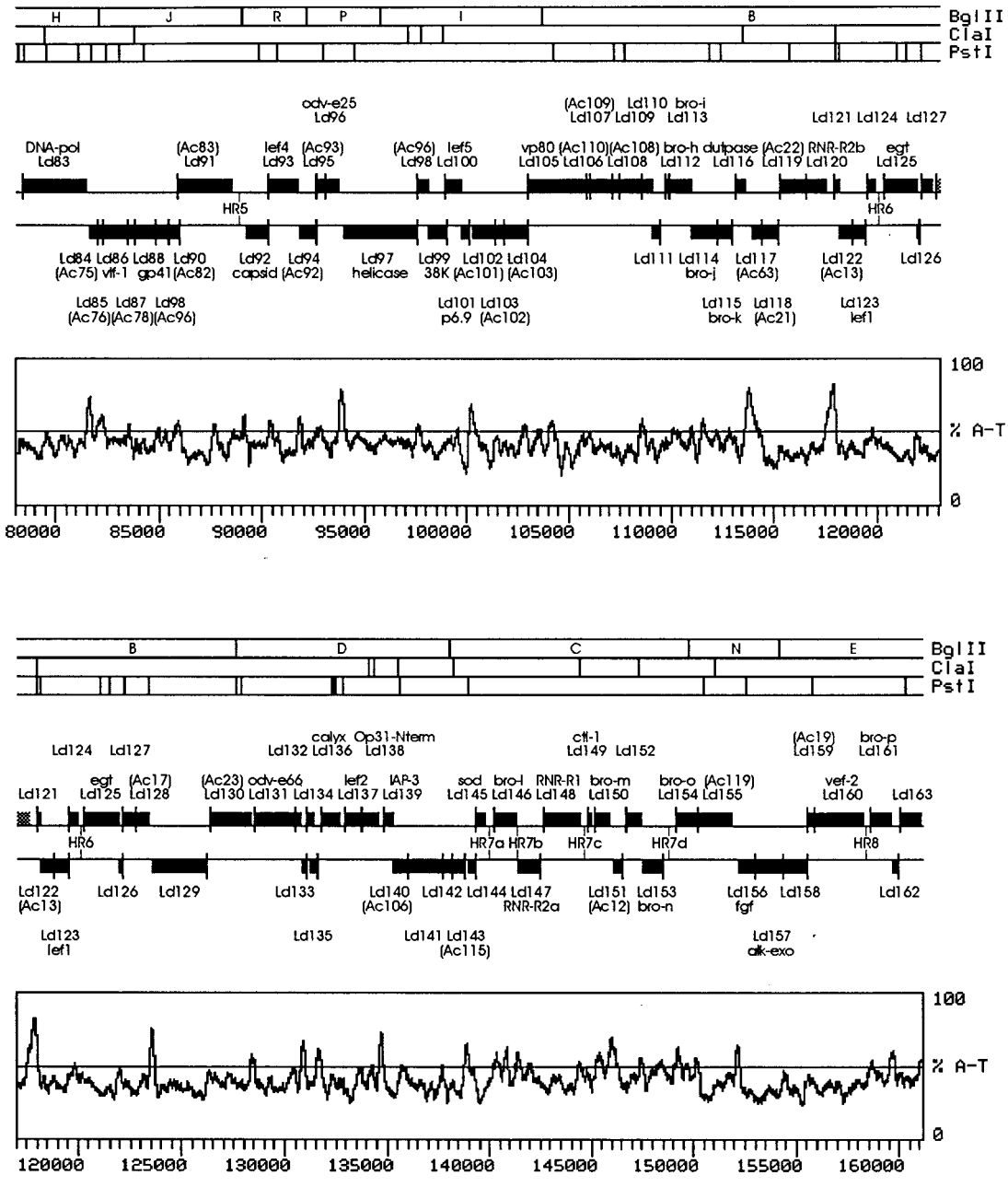


FIG. 1—*Continued*

(>65% identical) along with *p74* (Ld-ORF27), 61%; *ubiquitin* (Ld-ORF43), 71% (note that 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-ORF114 and -ORF161, respectively), at ~60% identity are also among the most conserved genes. The 41% overall identity of LdMNPV and AcMNPV ORFs is less than that reported for AcMNPV and OpMNPV (56%) (Ahrens *et al.,* 1997), 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 ORFs demonstrated similar amino acid sequence identities to the LdMNPV and AcMNPV ORFs. This suggests that the common ancestor of AcMNPV and OpMNPV diverged from LdMNPV before the separation of AcMNPV and OpMNPV.

A COMPARISON OF ORF CONTENT BETWEEN LdMNPV AND AcMNPV

LdMNPV homologs of AcMNPV ORFs

Of the 163 ORFs predicted for the LdMNPV genome, 95 AcMNPV homologs are represented (Table 2). This is in contrast to OpMNPV, which contains 126 homologs of AcMNPV. Of these 94 ORFs, only 3 (Ac-ORF12, -ORF58, and

TABLE 1

LdMNPV ORFs

orf	Ac ORF[1]	name	left	right	dir	aa	Mr	Ac:#aa %ID range[2]	prm[3]	reference[4]
repeat		*hr1*	39	523						Pearson and Rohrmann, 1995
1	ac8	*polyhedrin*	673	1410	>	245	28772	245 82% 242	L	
2	ac9	*orf1629 pp78-81*	1418	3085	<	555	61056	543 24% 393	-	Russell et al, 1997
3	ac10	*pk1*	3087	3911	>	274	32438	272 41% 266	L	Bischoff and Slavicek, 1994
4		*mucin-like*	4501	7590	<	1029	109196		L	GenBank P08640 Lambrechts et al, 1996
5			7722	8291	>	189	21763		-	
6			8282	8524	<	80	8960		-	
7		*g22*	8619	9194	>	191	22146		L	Bischoff and Slavicek, 1995
repeat		*hr2*	9496	10019						
8			10115	10474	<	119	13876		-	
9			10567	11085	>	172	19624		E, L	
10			11115	11336	<	73	8543		-	
11			11530	12264	>	244	27768		-	
12			12647	13165	>	172	20383		L	
13			13282	13482	<	66	7643		-	
14	ac148	*odv-e56*	13685	14755	>	356	38636	376 57% 346	L	Theilmann et al, 1996
15	ac147	*ie-1*	15221	16921	<	566	65076	582 31% 401	E	Pearson and Rohrmann, 1997
16	ac146		16978	17604	>	208	23360	201 31% 195	L	
17	ac145		17692	17970	<	92	10339	77 44% 77	L	
18	ac144	*odv-ec27*	17983	18834	<	283	32983	290 50% 285	L	Braunagel et al, 1996
19	ac143	*odv-e18*	18852	19118	<	88	9573	62 48% 60	L	Braunagel et al, 1996
20	ac142		19090	20541	<	483	56073	477 49% 472	L	
21	ac141	*ie-0*	20538	21314	<	258	29395	261 31% 225	E, L	Pearson and Rohrmann, 1997
22		*dna-ligase*	21745	23391	>	548	61974		-	Pearson and Rohrmann, in press
23	ac139	*me53*	23670	24698	>	342	40449	449 27% 352	L	
24			24921	25547	<	208	24201		-	
25			25720	26184	>	154	18391		L	
26			26415	26633	<	72	8417		-	
27	ac138	*p74*	26645	28663	>	672	74712	645 61% 579	L	
28			28711	29850	>	379	44008		-	
repeat		*hr3a*	29915	30399						
29	ac4		30469	30909	>	146	16686	151 22% 124	E, L	
30	ac150		31027	31311	>	94	10262	99 30% 88	L	
31			31305	31484	<	59	6727		-	
repeat		*hr3b*	31566	31959						
32	ac2	*bro-a*	32061	33113	>	350	39342	328 22% 277	-	
repeat		*hr3c*	33359	33586						
33	ac2	*bro-b*	33689	34660	>	323	36653	328 44% 290	-	
34			34977	35738	<	253	28940		-	
35	ac11		35898	36977	<	359	41044	340 30% 339	-	
36	ac26		36911	37282	<	123	13743	129 38% 101	L	
37	ac25	*dbp*	37327	38046	>	239	27255	316 31% 121	L	Mikhailov et al, 1998
38	ac28	*lef-6*	38053	38532	>	159	19250	173 28% 170	L	Passarelli and Miller, 1994
39	ac29		38727	38933	<	68	8393	71 31% 64	-	
40	ac136	*p26*	39013	39774	>	253	27792	240 27% 150	L	
41	ac137	*p10*	39809	40042	>	77	8118	94 24% 75	L	
42	ac34		40046	40612	<	188	20966	215 35% 156	-	
43	ac35	*v-ubi*	40690	41142	>	150	17092	77 71% 76	L	
44	ac36	*39K pp31*	41323	42117	<	264	29383	275 30% 265	L	
45	ac37	*lef-11*	41906	42469	<	187	21008	112 44% 87	-	Todd et al, 1995
46	ac 38		42388	43131	<	247	28494	216 55% 203	E, L	
47	ac25	*dbp*	43493	44266	>	257	29926	316 28% 257	-	Mikhailov et al, 1998
48	ac40	*p47*	44618	45790	<	390	45555	401 55% 395	-	
repeat		*hr4*	45899	46359						Pearson and Rohrmann, 1995
49			46446	46766	<	106	12487		L	
50		*helicase-2*	47126	48508	>	460	51498		-	Pearson and Rohrmann, in press
51	ac50	*lef-8*	48689	51313	<	874	99211	876 60% 881	-	Passarelli et al, 1994
52			51312	52211	<	299	33551		-	
53	Ac52		52246	53148	<	300	35249		-	
54	ac53		52826	53254	>	142	16826	139 43% 134	L	
55			53306	54391	<	361	39665		L	
56	ac53a	*lef-10*	54593	54823	>	76	8413	78 42% 64	L	Lu and Miller, 1994
57	ac54	*VP1054*	54684	55682	>	332	38552	365 38% 362	L	Olszewski and Miller, 1997
58	Ac55		55806	56000	>	64	7573		-	
59			56092	56253	<	53	6106		-	
60	ac57		56255	56749	>	164	18842	161 42% 159	-	
61	ac59		56732	57301	<	189	22314	69 55% 55	-	
62	ac60		57540	57827	<	95	10880	87 46% 85	L	
63*	ac61	*few polyhedra*	57955	58119	<	54	6077	214 31% 54	-	Bischoff and Slavicek, 1996
63*	ac61	*few polyhedra*	58079	58609	<	176	20062	214 53% 71	L	Bischoff and Slavicek, 1996
64	ac62	*lef9*	58726	60216	>	496	56693	516 67% 497	-	

-ORF139) are not present in OpMNPV (Ahrens *et al.,* 1997). Included in this set of conserved genes are all those implicated as being required for DNA replication (*lef*-1, -2, and -3; *helicase*; *dna polymerase*; and *ie-1*) (Kool *et al.,* 1994; Lu and Miller, 1995). Indeed, we have demonstrated that plasmids containing these genes are capable of supporting transient DNA replication when transfected into uninfected *L. dispar* cells (Pearson and Rohrmann, 1998), indicating

TABLE 1—*Continued*

65		*vef-1*	60268	62619	>	783	89336				L	Bischoff and Slavicek, 1997
66	ac3	*ctl-2*	62436	62714	<	92	9949	53	42%	38	-	Ahrens et al, 1997
67		*hrf-1*	62739	63395	>	218	25675				-	Thiem et al, 1996
68	ac64	*gp37*	63583	64392	<	269	30213	302	49%	249	L	
69			64648	64800	<	50	5766				-	
70	ac126	*chitinase*	64801	66477	<	558	60634	551	66%	551	L	Hawtin et al, 1997
71	ac2	*bro-c*	66612	68198	<	528	60852	328	23%	164	L	
72	ac2	*bro-d*	68445	69977	>	510	59388	328	22%	177	-	
73	ac2	*bro-e*	70133	70405	>	90	10451				-	
74	ac2	*bro-f*	70405	70794	>	129	14810				L	
75	ac2	*bro-g*	70965	71633	>	222	25786				-	
76	ac111		71665	71928	>	87	9829	67	41%	37	-	
77			71879	72523	>	214	24755				-	
78	ac127	*v-cath*	72684	73754	>	356	39770	323	68%	311	L	
79	ac71	*iap-2*	73757	74461	<	234	25966	249	31%	237	L	
80	ac68		74471	74857	<	128	14692	192	40%	112	L	
81	ac67	*lef3*	74856	75980	>	374	42659	385	30%	351	-	
82	ac66		76171	76507	<	778	87841	808	22%	452	-	
83	ac65	*dna-polymerase*	78389	81433	>	1014	115822	984	45%	992	-	Bjornson et al, 1992
84	ac75		81605	81991	<	128	15106	133	28%	120	L	
85	ac76		81993	82253	<	86	9920	84	37%	86	L	
86	ac77	*vlf-1*	82303	83439	<	378	43479	379	71%	361	L	McLachlin and Miller, 1994
87	ac78		83453	83794	<	113	12368	109	37%	97	L	
88	ac80	*gp41*	83791	84762	<	323	35879	409	60%	302	L	
89	ac81		84743	85402	<	219	25394	233	54%	199	L	
90	ac82		85287	85958	<	223	24759	180	32%	193	-	
91	ac83	*vp91 capsid*	85828	88422	>	864	95890	847	38%	883	L	Russell and Rohrmann, 1997
repeat		*hr5*	88457	89058								
92	ac89	*vp39 capsid*	89155	90213	<	352	38670	347	42%	352	L	Bjornson and Rohrmann, 1992b
93	ac90	*lef-4*	90212	91669	>	485	54707	464	44%	469	L	
94	ac92		91787	92542	<	251	29744	259	47%	259	-	
95	ac93		92541	93020	>	159	18144	161	46%	155	L	
96	ac94	*odv-e25*	93022	93675	>	217	23965	228	45%	223	-	
97	ac95	*helicase*	93899	97555	<	1218	141548	1221	41%	1227	L	
98	ac96		97512	98033	>	173	19289	173	45%	163	L	
99	ac98		98027	98995	<	322	38015	320	42%	318	L	
100	ac99	*lef-5*	98888	99724	>	278	32148	265	51%	244	L	
101	ac100	*p6.9*	99718	100017	<	99	11352	55	56%	66	L	
102	ac101		100203	101348	<	381	43418	361	39%	380	L	
103	ac102		101368	101733	<	121	13077	122	36%	78	L	
104	ac103		101726	102895	<	389	44514	387	42%	396	L	
105	ac104	*vp80*	102919	105813	>	964	107538	691	25%	174	-	
106	ac110		105815	105985	>	56	6595	56	26%	53	-	
107	ac109		105991	107091	>	366	42001	390	48%	390	L	
108	ac108		107094	107387	>	97	10670	105	36%	73	-	
109	ac112 113		107421	108431	>	336	39335				-	
110	ac24	*pkip-1*	108506	109045	>	179	21206	169	31%	102	L	Fan et al, 1998
111			109070	109372	<	100	11547				L	
112	ac2	*bro-h*	109738	109995	>	85	9423				-	
113	ac2	*bro-i*	109902	110942	>	346	39779	328	21%	234	-	
114	ac2	*bro-j*	110976	112187	<	403	46698	328	62%	222	-	
115	ac2	*bro-k*	112243	112959	<	238	27409	328	57%	127	-	
116		*dutpase*	113118	113567	>	149	16093				L	Ahrens et al, 1997
117	ac63		113878	114342	<	154	18735	155	22%	145	E	
118	ac21	*arif-1*	114395	115204	<	269	28932	319	23%	111	E	Roncarati and Knebel-Morsdorf, 1997
119	ac22		115244	116467	>	407	45498	382	59%	374	-	
120		*rnr r2b*	116489	117535	>	348	40154				-	
121			117907	118143	>	78	8990				E	
122	ac13		118140	118742	<	200	22678	327	26%	182	L	
123	ac14	*lef1*	118724	119428	<	234	26911	266	43%	227	-	
124			119473	119874	>	133	14888				L	
repeat		*hr6*	119863	120171								
125	ac15	*egt*	120183	121865	>	560	63255	506	43%	497	-	
126			121887	122054	<	55	6002				-	
127			122084	122662	>	192	22412				E	
128	ac17		122670	123350	>	226	25461	164	32%	102	-	
129			123478	126132	<	884	99009				-	
130	ac23		126267	128297	>	676	76738	690	22%	293	E, L	
131	ac46	*odv-e66*	128398	130395	>	665	73212	704	45%	553	E, L	Hong et al, 1994
132			130473	130718	>	81	9153				-	
133			130799	130963	<	54	6000				E	
134			130991	131359	>	122	12903				-	

that their function is conserved. Of 12 additional genes that have been implicated in late gene expression or in stimulating DNA replication (Kool *et al.,* 1994; Lu and Miller, 1995), homologs of 9 late transcription genes are present (*lef* 4–6 and 8–11, *39K*, and *p47*), whereas 3 that are stimulatory for replication (*lef*-7, *p35,* and *ie-2*) are not present. Most of the genes implicated as being structural components of virions or polyhedra are also conserved between the two viruses. Genes that likely modulate insect growth or development, such as ecdysteroid UDP-glu-

TABLE 1—*Continued*

135			131179	131550	<	123	13857				L	
136	ac131	*pe pp34 calyx*	131665	132606	>	313	33825	252	27%	245	L	Bjornson and Rohrmann, 1992a
137	ac6	*lef-2*	132917	133567	>	216	24127	210	38%	216	-	
138	Op31		133652	134527	>	291	33893				-	Ahrens et al, 1997
139		*iap-3*	134781	135248	>	155	17421				-	Ahrens et al, 1997
140	ac106		135240	135980	<	246	27729				L	
141			136023	137651	<	542	61879				L	
142			137720	138076	<	118	13234				-	
143	ac115		138085	138696	<	203	22177	204	53%	159	L	
144			138840	139181	<	113	13330				-	
145	ac31	*sod*	139262	139726	>	154	16106	151	75%	151	-	
repeat		*hr7a*	139818	140086								
146	ac2	*bro-l*	140189	141250	>	353	39770	328	18%	293	-	
147		*rnr r2a*	141380	142459	<	359	41350				-	Ahrens et al, 1997
148		*rnr r1*	142566	144356	>	596	65740				-	Ahrens et al, 1997
repeat		*hr7b*	144466	144693								
149	ac3	*ctl-1*	144764	144925	>	53	5513	53	77%	53	L	
150	ac2	*bro-m*	145098	145829	>	243	28158	328	33%	97	-	
151	ac12		145994	146500	<	168	19679	217	31%	114	-	
152			146602	147354	>	250	29284				-	
153	ac2	*bro-n*	147421	148437	<	338	38724	328	82%	317	-	
repeat		*hr7c*	148561	148945								
154	ac2	*bro-o*	149058	150068	>	336	37903	328	28%	112	-	
155	ac119		150170	151762	>	530	59185	530	46%	543	L	
156	ac32	*fgf*	152133	152990	<	285	31574	181	33%	133	L	
157	ac133	*alk-exo*	153023	154285	<	420	46852	419	37%	413	L	
158	ac18		154323	155444	<	373	41848	353	24%	249	-	
159	ac19		155431	155787	>	118	13575	108	27%	100	L	
160		*vef-2*	155812	158178	>	788	88408				L	Hashimoto et al, 1991
repeat		*hr8*	158178	158385								
161	ac2	*bro-p*	158489	159502	>	337	37643	328	64%	156	L	
162	ac12		159594	159869	<	91	10424	217	30%	50	E	
163			159993	160982	>	329	38472				L	

Note. In this viral strain, there is a frame shift in this gene (Bischoff and Slavicek, 1996). The two resulting genes are indicated by the same ORF number (ORF63).

[1] This column indicates the AcMNPV homologs (Ayres *et al.,* 1994). OpMNPV homologs (op) (Ahrens *et al.,* 1997) are shown in the few cases in which AcMNPV homologs are lacking.

[2] This column indicates the percentage amino acid identity between LdMNPV and AcMNPV homologs. The values shown are the number of amino acids in the AcMNPV ORF, the percentage identity with the LdMNPV homolog, and the overlap used to calculate the percentage identity.

[3] This column indicates the presence of early (E) and late (L) promoter elements within 120 nucleotides of the ATG. E indicates a TATA sequence with a CAGT or CATT mRNA start site sequence 25–35 nucleotides downstream. L indicates the presence of an ATAAG, a GTAAG, or a TTAAG.

[4] Only selected references are indicated in this column. For additional references, see Ahrens *et al.* (1997), Ayres *et al.* (1994).

cosyltransferase (*egt*) (O'Reilly and Miller, 1989) and a homolog to fibroblast growth factor (Ld-ORF156), are also conserved. Two additional conserved genes, *chitinase* (Ld-ORF70) and *cathepsin* (Ld-ORF78) (Hawtin *et al.,* 1997; Ohkawa *et al.,* 1994), are thought to be involved in virion dissemination by causing disintegration of insects late in infection. Therefore, genes encoding proteins involved in DNA replication, late transcription, virion structure, and dissemination account for about a third of the conserved genes.

AcMNPV ORFs with no homologs in the LdMNPV genome

Homologs of 61 AcMNPV genes are not present in LdMNPV (Table 3). Of these, 22 are also not present in the OpMNPV genome (Table 3). In particular, two transactivators of early transcription, *ie-2* and *pe38,* are lacking. *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 (Gomi *et al.,* 1997).

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 hemocel (Monsma *et al.,* 1996). This suggests either that a *gp64* homolog is present but highly diverged or that LdMNPV has other genes that are involved in membrane fusion or uses 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.

To determine whether 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 (see Materials and Methods) that are indicative of transmembrane receptor-like proteins. A single candidate gene was identified (ORF130). This gene has a low level of homology to AcMNPV ORF23. Similar analyses

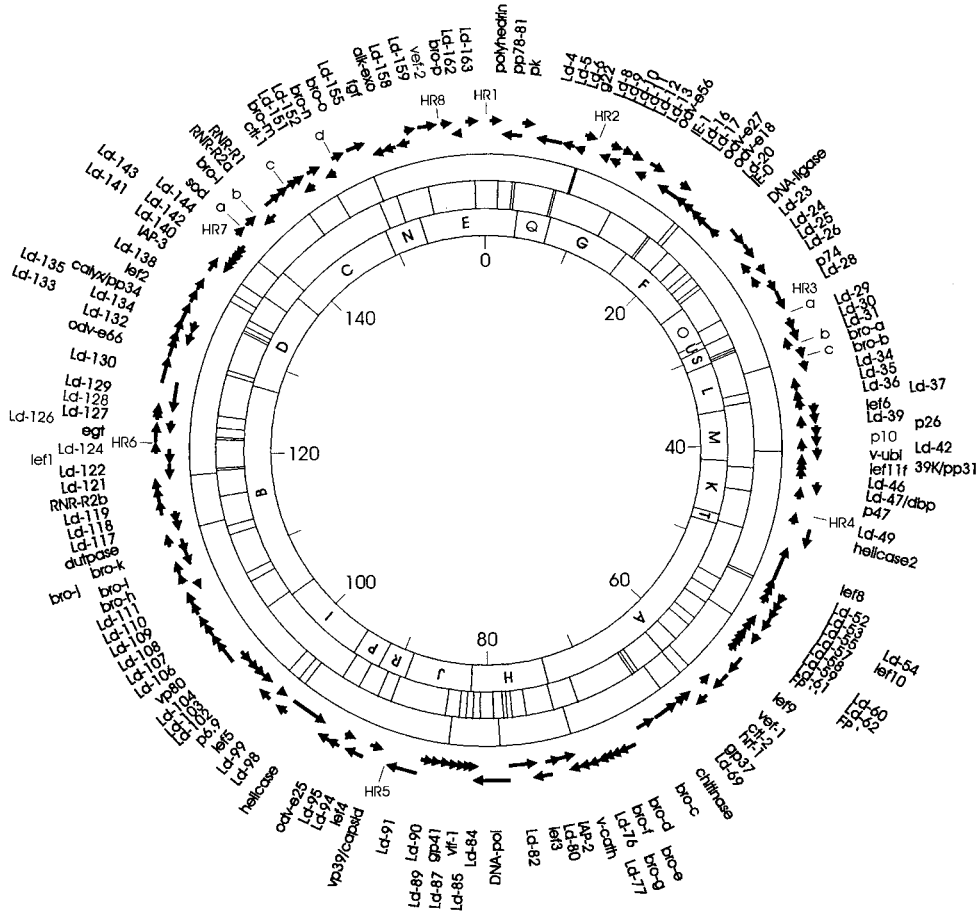


FIG. 2. Circular map of LdMNPV genome. The center circle shows a scale in kb. ORFs and their orientations are indicated by arrows. The names are described in the legend to Fig. 1. The inner circles indicate the position of sites for the following enzymes: outer circle, *Cla*I; middle circle, *Pst*I; and inner circle, *Bgl*II.

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.

LdMNPV genes homologous to non-AcMNPV genes

A number of LdMNPV genes are lacking AcMNPV homologs but are related to genes found in other organisms. These are summarized below and in Table 4.

TABLE 2

AcMNPV Genes with Homologs in LdMNPV

2 [16][a]	19	37	62(*lef*-9)	83	104(*vp80*)	138(*p74*)
3 [2]	21	38	64(*gp37*)	89(*vp39-capsid*)	106	139(*me53*)
4	22	40(*p47*)	65 (*DNApol*)	90(*lef-4*)	108	141(*ie-0*)
6(*lef*-2)	23	46(*odv-e66*)	66	92	109	142
8(polh)	24(*pkip*)	50(*lef*-8)	67(*lef*-3)	93	110	143(*odv-e18/35*)
9(*1629*)	25 [2]	52	68	94(*odv-e25*)	111	144(*odv-ec27*)
10(*pk-1*)	26	53	71(*iap-2*)	95(*helicase*)	112/113	145
11	28(*lef*-6)	53a	75	96	115	146
12 [2]	29	54	76	98	119	147(*ie-1*)
13	31(*sod*)	57	77(*vlf-1*)	99(*lef*-5)	126(*chi*)	148(*odv-e56*)
14(*lef*-1)	32(*fgf*)	58	78	100(*p6.9*)	127(*vcath*)	
15(*egt*)	34	59	80(*gp41*)	101	131(*pe*)	
17	35(*ubi*)	60	81	102	133	
18	36(*39K*)	61(*fp*)	82	103	137(*p10*)	

Note. Total of 95.

[a] The numbers in square brackets indicate the number of copies of homologs in LdMNPV.

TABLE 3

AcMNPV Genes[a] without Homologs in LdMNPV

1(*ptp*)	42(*gta*)	**63**	87	120	**134**
5	43	**69**	88(*cg30*)	**121**	**135(*p35*)**
7	44	**70(*hcf-1*)**	91	122	136(*p26*)
16	**45**	72	**97**	**123(*pk-2*)**	**140**
20	47	73	**105**	124	**149**
27(*iap-1*)	48	74	107	125(*lef-7*)	**150**
30	49(*pcna*)	79	114	128(*gp64*)	151(*ie-2*)
33	51	**84**	**116**	129	**152**
39	55	85	117	130(*gp16*)	153(*pe38*)
41	44	**86(*pnk*)**	**118**	132	**154**

Note. Total of 60.

[a] The numbers that are bold and underlined indicate genes also absent in OpMNPV.

Genes involved in DNA replication or DNA repair. Although a similar set of genes required for DNA replication has been identified by transient DNA replication assays in both AcMNPV and OpMNPV (Ahrens *et al.,* 1996; Kool *et al.,* 1994; Lu and Miller, 1995), 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) that has ~35% amino acid sequence identity to the LIGASE of vaccinia virus. This 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 (Pearson and Rohrmann, 1998). In addition, an ORF (ORF50) was identified with 31% amino acid sequence identity to a yeast mitochondrial *helicase*, called *pif1* (Foury and Lahaye, 1987). 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 (Pearson and Rohrmann, 1998). Because type III DNA *ligase* and *pif1* are associated with DNA repair and recombination, they may be involved in a LdMNPV DNA repair system.

Genes involved in nucleotide metabolism. Four ORFs that are homologous to genes involved in nucleotide metabolism are present in the LdMNPV genome. These include a homolog of the large subunit and two homologs of the small subunit of ribonucleotide reductase and a gene with homology to *dutpase*. Although homologs of these genes are not found in AcMNPV, all three are present in the OpMNPV genome (Ahrens *et al.,* 1997). The expression of ribonucleotide reductase by viruses is thought to allow them to synthesize a source of deoxyribonucleotides and thereby permit replication in nondividing cells (Goldstein and Weller, 1988). 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 (Mathews and van Holde, 1995). The LdMNPV-encoded homologs of ribonucleotide reductase are described below. The *dutpase* homolog differs from that found in the OpMNPV genome. 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 and 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 mucin-like homolog. Ld-ORF4 encodes a 1029-amino-acid polypeptide that contains 11 imperfect copies of the tandem repeat SSSTVKPKSSSDKPSDKAK. A modified BLAST search with the low complexity filtering default enabled produced one top scoring record (GenBank accession no. P08640) that showed a significant alignment with 22% amino acid sequence identity and 34% similarity over 902 amino acids. This polypeptide, encoded by the yeast DNA regions S1 and S2, encodes a mucin-like yeast protein (Lambrechts *et al.,* 1996).

REPEATED GENES

A number of LdMNPV genes are present in multiple copies. Some of these have homologs in AcMNPV or OpMNPV, whereas others are lacking in these viruses.

TABLE 4

LdMNPV Genes with Homology to Non-AcMNPV Genes in the Database

Viral enhancing factor 1	*iap-3*[a]
Viral enhancing factor 2	*dUTPase*
Ribonucleotide reductase, large subunit	*DNA ligase*
Ribonucleotide reductase, small subunit 1	*helicase 2*
Ribonucleotide reductase, small subunit 2	*hrf-1*[b]

[a] Birnbaum *et al.* (1994).

[b] Thiem *et al.* (1996).

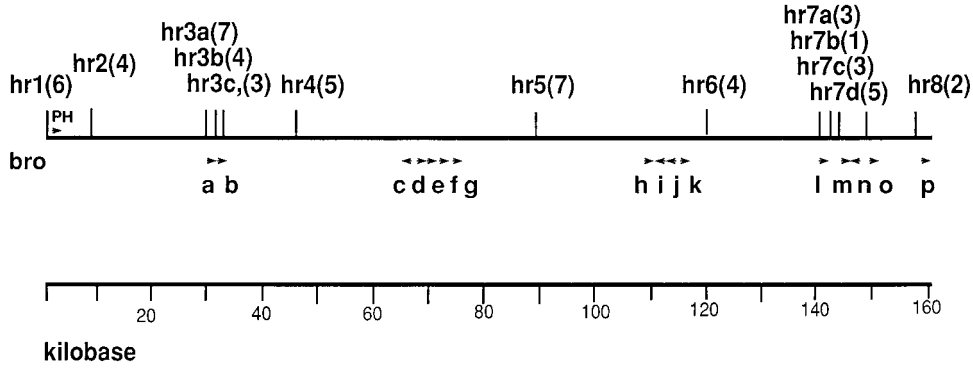


FIG. 3. LdMNPV genome map showing the location of homologs to AcMNPV ORF2 (bro genes) and *hr*s. The location and orientation of the AcMNPV ORF2 homologs (bro) are shown below the line, and the *hr* sequences are shown above the line. The numbers in brackets indicate the number of complete repeats in the *hr* palindromes. The *hr* data were generated by scanning the genome using a consensus sequence that finds all baculovirus *hr* 30-bp palindrome sequences, including LdMNPV, AcMNPV, BmNPV, CfMNPV, OpMNPV, and AgMNPV, in the database (Ahrens *et al.,* 1997; Ayres *et al.,* 1994; Cochran and Faulkner, 1983; Garcia-Maruniak *et al.,* 1996; Majima *et al.,* 1993; Theilmann and Stewart, 1992; Xie *et al.,* 1995).

Homologs of AcMNPV ORF2: A baculovirus repeated ORF (*bro*) family

A striking feature of the LdMNPV genome is the presence of 16 ORFs that are related to AcMNPV ORF2 (Fig. 3). We have named these ORFs baculovirus repeated ORFs (BRO) and have lettered them sequentially in the genome (Fig. 1, Table 1). Although five copies of *bro* are present in the BmNPV genome (GenBank accession no. L33180), in AcMNPV there is a single copy, and in OpMNPV there is a small truncated 88-amino-acid representative and two other smaller related ORFs (Ahrens *et al.,* 1997). Some of the *bro* genes are located in contiguous clusters (Fig. 3). Two clusters are located near *hr*s, indicating that they may have been duplicated along with the *hr*s. An analysis of the relatedness of *bro* genes from LdMNPV, AcMNPV, and BmNPV is shown in Fig. 4. In this analysis, related domains have been given similar colors. Most of these genes share a related core sequence, and they demonstrate differing degrees of similarity in other domains. Outside the shared core region, members of the *bro* gene family were separated into four groups based on the relationship of different domains. Alignment of the core sequences for groups 1–3 are shown in Fig. 5. 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 and -d, which are distinguished by their length and a common 3′ region lacking in the other *bro*s (Fig. 4). Also, Ld-*bro*-a, -l, and -o likely evolved from relatively recent duplication events. There is a partial *bro* (*bro*-h) located upstream of the -i/-j/-k cluster that appears to be a duplication of part of *bro*-l. Two of the genes (*bro*-e and -f) (ORF73 and ORF74) contain contiguous overlapping translational stop start signals (TGATG) that result in predicted polypeptides closely related to the 5′ and 3′ regions of Ld-*bro*-g. This is in a region of five contiguous *bro* sequences and appears to be an area of intense recombination (see Materials and Methods), which may account for the generation of this duplication and apparent frameshift. The Ac-*bro* (ORF2), Bm-*bro*-d, and Ld-*bro*-n 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.

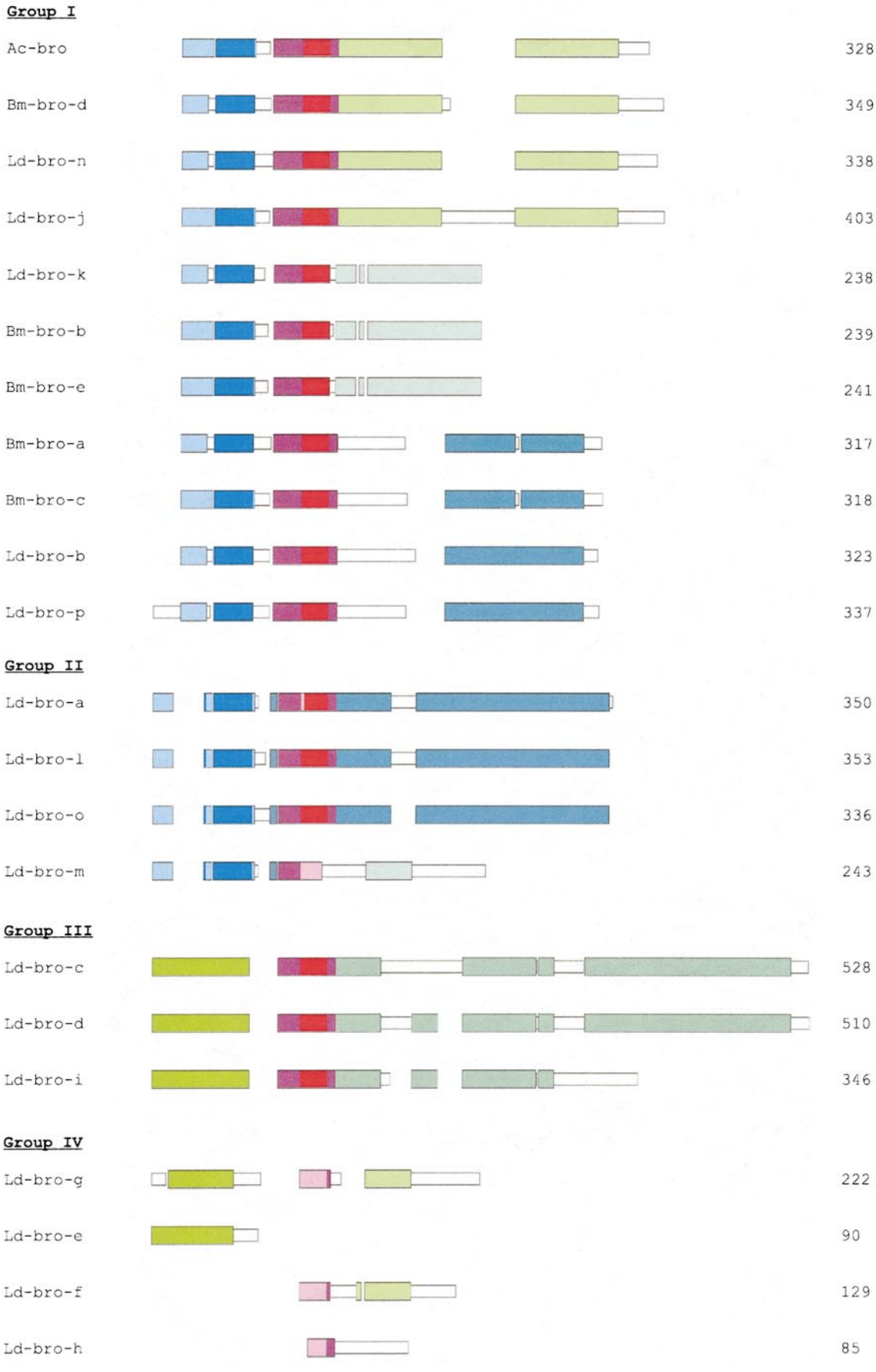
Families of related, highly variable genes occur in other virus species, including African swine fever virus (Yozawa *et al.,* 1994) and *Molluscum contagiosum* virus (MCV) (Senkevich *et al.,* 1996). In these viruses, these genes are grouped at the terminal replicative ends of the DNA. BLAST searches of intravirus proteins clearly identify these families, but no evidence of similarity exists when families of proteins are compared between different viruses.

LdMNPV ORF37 and ORF47: Duplicated homologs of AcMNPV ORF25

Two copies (ORF37 and ORF47) of homologs to AcMNPV ORF25 were identified in the LdMNPV genome. They show ~26% identity to each other, with the C-terminal 40 amino acids being almost 45% identical. LdMNPV ORF47 is most closely related to AcMNPV ORF25 (30%), with the C-terminal 40 amino acids being over 55% identical. LdMNPV ORF37 is ~26% identical to AcMNPV ORF25. The BmNPV homolog of AcMNPV ORF25 encodes a protein that is capable of binding to single-stranded DNA and unwinding partial DNA duplexes *in vitro* (Mikhailov *et al.,* 1998).

Viral enhancing factor (VEF)

Although VEF homologs are not found in AcMNPV (Ayres *et al.,* 1994) or OpMNPV (Ahrens *et al.,* 1997), a homolog has previously been reported from the LdMNPV genome (ORF65) (Bischoff and Slavicek, 1997). Deletion of the LdMNPV *vef*-1 gene caused a decrease


Group I
Ac-bro
328
Bm-bro-d
349
Ld-bro-n
338
Ld-bro-j
403
Ld-bro-k
238
Bm-bro-b
239
Bm-bro-e
241
Bm-bro-a
317
Bm-bro-c
318
Ld-bro-b
323
Ld-bro-p
337
Group II
Ld-bro-a
350
Ld-bro-l
353
Ld-bro-o
336
Ld-bro-m
243
Group III
Ld-bro-c
528
Ld-bro-d
510
Ld-bro-i
346
Group IV
Ld-bro-g
222
Ld-bro-e
90
Ld-bro-f
129
Ld-bro-h
85

```
Group I          x      •      •• •   •    •     x      x
Ac-bro           PLYLQPHTVLITKSGVIQLIMKSKLPYAIELQEWLLEEVIP
Ld-bro-n         PLYLQPHTVLITKSGVIQLIMKSKLPYAVELQEWLLEEVIP
Bm-bro-d         PLYLQPHTVLITKSGVIQLIMKSKLPYAIELQEWLLEEVIP
Ld-bro-j         PLYLHPHTVLVTKEGVIQLIMKSKLPYAVELQAWLLEEVIP
Ld-bro-k         PLYLHKSTILIDKIGVIQLFMRSKLHNAAELQNWFYERVLP
Bm-bro-b         PLYLSPQTILLDKIGVIQLFMRSKMHNAAELQNWFYEHVLP
Bm-bro-e         PLYLQTQTILLDKIGVIQLFMRSKMTNAAELQNWFYEHVLP
Bm-bro-a         PLYLQPHTVLITKEGVIQLIMKSKLPYAVELQAWLLEEVIP
Bm-bro-c         PLYLHPHTVLITKSGVIQLIMKSKLPYAIELQEWLLEEVIP
Ld-bro-b         PLYLHPHTVLITKEGVIQLIMKSKLPYAVELQAWLLEEVIP
Ld-bro-p         PLYLHPHTVLITKEGVIQLIMKSKLPYAVELQAWLLEEVIP

Group II
Ld-bro-a         TSSLHPQTKFINRAGVFELISASEMPAAKRFKQWNANDLLP
Ld-bro-l         PRNIQAKTKFINQAGVFELIGASEMPAAKRFKTWNTNDLLP
Ld-bro-o         PRNVQAKTKFINRAGVFELISASEMPAAKRFKTWNTNDLLP
Ld-bro-m         PRNVQAKTKFINMNGVIELLLASQMQQAKEFRYWMTNVFKA

Group III
Ld-bro-c         PANWHPETLFVLEPGVYALMARSTKPMAKEKMKFVYETILP
Ld-bro-d         PANWQPETLFVLEPGVYALMARSTKPMAKEKMKYVYETILP
Ld-bro-i         PANWHPETLFVLEPGVYALLARSNKPLAKERMKFVYETILP
```

FIG. 5. Alignment of 41-amino-acid core of LdMNPV BRO sequences. This region is shaded red in Fig. 4. The groups are the same as in Fig. 4. The dots above the sequence indicate identical amino acids. The xs indicate that amino acids in this position are highly conserved.

in viral potency, suggesting that VEF-1 may be involved in viral infectivity (Bischoff and Slavicek, 1997). A second copy of the gene was identified in the LdMNPV genome (ORF161, Table 1). Both genes differ significantly from other *vef*s and from each other, demonstrating only ~30% amino acid sequence identity. VEF is a large protein that forms ~5% of the mass of occlusion bodies of the *Trichoplusia ni* granulosis virus (TnGV) (Hashimoto *et al.,* 1991). The VEF protein appears to facilitate baculovirus infection by disrupting the peritrophic membrane, thereby allowing virions access to the surface of gut cells (Derksen and Granados, 1988).

Conotoxin-like (*ctl*) genes

Conotoxins are small disulfide-rich ion channel antagonists isolated from the genus *Conus* (Olivera *et al.,* 1994). A single *ctl* gene is present in AcMNPV (AcMNPV ORF3). In contrast, the OpMNPV genome encodes two ORFs called *ctl*-1 and *ctl*-2. LdMNPV also has two *ctl* genes. LdMNPV *ctl*-1 (ORF149) is ~78% identical in amino acid sequence to AcMNPV and OpMNPV *ctl*-1. In contrast, LdMNPV *ctl*-2 (ORF66) is predicted to be 40 amino acids longer than the other baculovirus *ctl*s and shows only 30–40% identity to the *ctl*-1 genes. However, it shows higher identity (~53%) to OpMNPV *ctl*-2. Therefore, in this report, we grouped it with the *ctl*-2 lineage. In a study examining the AcMNPV *ctl*-1 gene, no differences in mortality, motility, or weight gain were observed when neonate or late instar *Spodoptera frugiperda* larvae were infected with an AcMNPV *ctl*-1 deletion mutant compared with infection with wild-type virus (Eldridge *et al.,* 1992).

Genes involved in nucleotide metabolism: Two homologs of the ribonucleotide reductase (RNR) small subunit

Two adjacent genes (ORF147 and ORF148) that encode predicted ribonucleotide reductase subunits were identified in the LdMNPV genome. Both genes showed very high homology to the large (R1) and small (R2) subunits from OpMNPV (Ahrens *et al.,* 1997) (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 (van Strien *et al.,* 1997). R1 from these viruses show 58–68% amino acid sequence identity with other eukaryotic R1s and are distantly related to OpMNPV, HSV, and *Escherichia coli* R1 (25–27% identity). The R2 subunit from OpMNPV was even less well conserved than R1, showing only 10–16% sequence identity with other R2 polypeptides. 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). In contrast, the second LdMNPV R2 ORF shows 70% amino acid sequence identity to R2s of vaccinia virus and golden hamster. 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. The identification of two *r2* genes in an organism is not unprecedented. *Saccharomyces cerevisiae* has recently been reported to encode two R2 subunits that are ~54% identical (Huang and Elledge, 1997; Wang *et al.,* 1997). However, one of these subunits lacks critical amino acids associated with iron binding and appears to be catalytically inactive. Evidence suggests that it interacts with the other R2 subunit and may play a structural role in the RNR complex. In contrast to yeast, both LdMNPV RNR2 subunits contain all the conserved amino acids involved in iron binding.

Inhibitor of apoptosis (*iap*) genes

In OpMNPV and the *Cydia pomonella* granulovirus (CpGV), a gene has been identified that is capable of blocking apoptosis induced by infection with AcMNPV

FIG. 4. Relationships of BROs. AcMNPV, BmNPV, and LdMNPV BRO alignments were carried out using the program Macaw with guidance from Gap BLAST pairwise alignments. The polypeptide sequences are represented as rectangles, with the expanded rectangles of the same color corresponding to blocks of high sequence similarity ($P < 10^{-8}$). The BROs have been grouped into four classes (groups I, II, III, and IV) corresponding to observed sequence similarity of their N-terminal and core domains. The core region is shaded red. Core flanking regions, which have a lesser degree of conservation, are shaded purple. Degenerate cores are shaded pink. Group IV is composed of BROs with degenerate or missing core domains (see text). The numbers on the right are the sizes of the ORFs in amino acids.

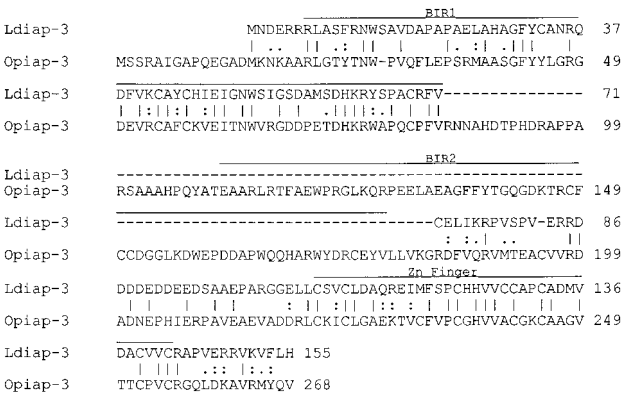


FIG. 6. Alignment of LdMNPV ORF139, a possible *iap*-3 homolog. Dashes (-) indicate gaps; amino acid number is indicated on the right; identical amino acids are indicated by the vertical line; and colons indicate conserved amino acids. The OpMNPV *iap*-3 sequence is from Birnbaum *et al.* (1994). The baculovirus IAP repeat (BIR) and zinc finger domains are overlined and labeled.

(Birnbaum *et al.,* 1994; Crook *et al.,* 1993). These are called *iap* genes, and they subsequently have been shown to belong to a large gene family with members in a variety of eukaryotes (reviewed in Clem, 1997). IAP homologs often contain two tandem baculovirus IAP repeats (BIR), as described by Birnbaum *et al.* (1994), and a zinc (RING) finger-like Cys/His motif. BIR domains are associated with binding to apoptosis-inducing proteins (Harvey *et al.,* 1997; Vucic *et al.,* 1997). Analysis of OpMNPV revealed three *iap* subfamilies with one, *iap*-3, capable of blocking apoptosis in AcMNPV-infected *S. frugiperda* cells (Birnbaum *et al.,* 1994). LdMNPV contains two ORFs (ORF79 and ORF139) that are members of the *iap* family. LdMNPV ORF79 shows ~30% identity with AcMNPV and OpMNPV IAP-2s but only 20–24% identity with AcMNPV and OpMNPV IAP-1s. This suggests that LdMNPV ORF79 may be a member of the IAP-2 subfamily. LdMNPV ORF139 at 155 amino acids is smaller than most *iap* homologs. It appears to have a major internal deletion that eliminated the BIR-2 domain while retaining the zinc (RING) finger domain (Fig. 6). When these domains are aligned separately to accommodate the internal deletion, they show 26–31% identity to AcMNPV and OpMNPV IAP1s and ~23% identity to IAP-2s. In contrast, LdMNPV ORF139 shows 42% identity with OpMNPV *iap*-3 (Fig. 6). This suggests that LdMNPV ORF139 may be a member of the *iap*-3 subfamily and be active in blocking apoptosis during LdMNPV infection.

LdMNPV ORF151 and ORF162 are related to AcMNPV ORF12

AcMNPV ORF12 is predicted to encode a peptide of 217 amino acids. Two LdMNPV ORFs share an overlapping region of similarity with this protein; these are Ld-ORF151 (31% identity) and Ld-ORF162 (30% identity). Both LdMNPV ORFs are smaller than Ac-ORF12 (Ld-ORF151, 168 amino acids; Ld-ORF162, 91 amino acids).

REPEATED REGIONS

Homologous regions

A novel feature of many baculovirus genomes is the presence of *hr*s that are located throughout the genome (Ayres *et al.,* 1994; Cochran and Faulkner, 1983; Garcia-Maruniak *et al.,* 1996; Theilmann and Stewart, 1992; Xie *et al.,* 1995). *hr*s are composed of repeated sequences encompassing both direct repeats and imperfect palindromic sequences and are related to one another. AcMNPV has eight *hr*s, whereas the OpMNPV genome has five. In these viruses, each *hr* contains 1–10 imperfect palindromes of 30 bp within a directly repeated sequence (Ahrens *et al.,* 1997; Ayres *et al.,* 1994). In both AcMNPV and OpMNPV, *hr*s can act as enhancers of RNA polymerase II-mediated transcription (Guarino and Summers, 1986; Theilmann and Stewart, 1992) and can behave as origins of DNA replication in transient replication assays (Ahrens *et al.,* 1995; Kool *et al.,* 1995; Pearson *et al.,* 1992). In addition, evidence suggests that the baculoviral transactivator, *ie-1*, binds to *hr* sequences (Choi and Guarino, 1995).

Thirteen homologous regions (*hr*s 1, 2, 3a, 3b, 3c, 4, 5, 6, 7a, 7b, 7c, 7d, and 8) were identified in the LdMNPV genome that contain 1–7 palindrome repeats for a total of 53 repeats (Fig. 3). A number of partial repeats are also present in some *hr*s. As previously reported, different LdMNPV *hr*s may be closely related to one another (e.g., *hr*1 and *hr*4 are 89% identical) (Pearson and Rohrmann, 1995). There are variations between individual *hr* palindromes and differences within each palindrome that confer on these structures a direction. Therefore, some clusters of *hr*s are oriented in the same direction. This directionality is reflected by the larger direct repeats within which *hr*s are often embedded (Fig. 7). *hr*7b is limited to a single palindrome and is not embedded in a larger direct repeat (Fig. 3). In contrast to AcMNPV *hr*s, which can act as origins of replication in transient assays, it was found that the LdMNPV *hr*s had to be linked to an AT-rich sequence before they were capable of undergoing replication (Pearson and Rohrmann, 1995). Similar to the *hr*s of AcMNPV and OpMNPV, LdMNPV *hr*s can act as enhancers of early gene transcription (Pearson and Rohrmann, 1997). The LdMNPV *hr*s are composed of a series of 72-bp DNA repeats containing a 30-bp palindrome with *Mlu*I sites in each arm. This palindrome flanks a combined *Xho*I--*Sac*I site (Fig. 7a). Except for the four variations shown in Fig. 7b, all nucleotides in the palindrome are >92% conserved, with the variable nucleotides being ~60–75% conserved, with the exception of the 5′ G, which is 42% conserved. The nucleotides comprising the palindrome are designated with asterisks in Fig. 7b and include 24 of 30 positions. In

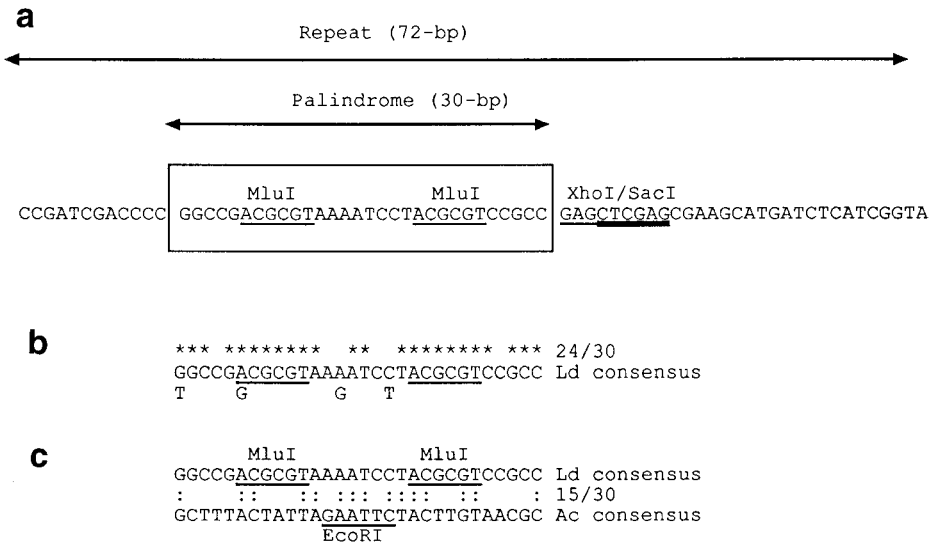


FIG. 7. LdMNPV hr sequences. (a) The organization of LdMNPV *hr*s is shown, including the repeated sequence and the location of the imperfect palindrome (boxed). Conserved restriction enzyme sites are indicated. (b) The LdMNPV *hr* palindrome consensus sequence. The asterisks above the sequence are nucleotides that contribute to the palindrome. The nucleotides below the line indicate positions conserved <92%. (c) A comparison of the LdMNPV and AcMNPV *hr* consensus sequences. The colons indicate identity.

addition, compared with the AcMNPV consensus *hr* palindrome, 50% of the sequences are identical (Fig. 7c). The similarity between LdMNPV and OpMNPV *hr*s was not so clearly evident, showing only 40% identity if single gaps are inserted (not shown). However, when a consensus profile derived from a number of baculoviruses was used to scan the LdMNPV genome, all the *hr*s in the LdMNPV genome were identified (Fig. 3). This is further evidence that the *hr*s from different baculoviruses share a common ancestor.

The position of at least some *hr*s 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 (Ahrens *et al.,* 1997). LdMNPV hr5 is located immediately downstream of Ld-ORF91, the LdMNPV homolog of AcMNPV ORF83 (Table 1). In addition, immediately downstream of this *hr* is located Ld-ORF92, the homolog of AcMNPV vp39-capsid (AcMNPV ORF89). Because no homologs of the AcMNPV ORFs (ORF84–ORF88) are present in the LdMNPV genome, this indicates that the position of LdMNPV *hr*5 is conserved relative to both the upstream and downstream genes. The ORF organization downstream of this *hr* in OpMNPV and BmNPV are similar to LdMNPV; homologs of ORF84–ORF86 are missing in BmNPV and ORF84 and ORF86 are missing in OpMNPV.

Given that *hr*s share higher similarity within a virus strain than do any *hr*s between species, the amplification process appears to be tightly linked to functional conservation. Within a virus species, the *hr*s may coevolve with the proteins with which they interact, such as the transcriptional activator, *ie-1*. The evidence that *hr*s 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 supports data suggesting that they are transcriptional enhancers and may also serve as origins of DNA replication *in vivo* (reviewed in Lu *et al.,* 1997).

GENOME COMPOSITION AND EVOLUTION

G + C content

Although the LdMNPV genome has a relatively high G + C content (57.5%), there is a markedly uneven distribution of nucleotide composition, with most of the ge-

TABLE 5

Nucleotide Content and G + C Composition of Baculovirus Genome Components[a]

Virus	Genome Size (% GC)	Coding (% GC)	hr[b] (% GC)	Intergenic (% GC)
LdMNPV	161,046 (57.5)	138,908 (59.6)	4,590 (57.4)	17,548 (40.4)
OpMNPV	131,993 (55.1)	118,000 (56.9)	1,619 (58.1)	12,374 (38.2)
AcMNPV	133,894 (40.7)	122,527 (41.6)	3,103 (33.5)	8,264 (29.7)
BmNPV	128,413 (40.4)	114,231 (41.6)	3,418 (31.5)	10,764 (30.5)

[a] The values indicate the number of nucleotides in each category followed by the percent G + C content of each category.
[b] hr sizes were determined from a profile search.

TABLE 6

Frequency of Late Promoter Sequences: Actual (and Predicted[a])

	ATAAG	CTAAG	GTAAG	TTAAG	Total (NTAAG)
LdMNPV	114 (190) 60%	30 (256) 12%	77 (256) 30%	62 (190) 33%	283 (892) 32%
AcMNPV	132 (421) 31%	55 (289) 19%	85 (289) 29%	125 (421) 30%	397 (1420) 28%
OpMNPV	92 (185) 50%	33 (226) 15%	71 (226) 31%	106 (184) 58%	302 (821) 37%

[a] Predicted NTAAG sequences are shown in brackets and were calculated from 8192 randomly shuffled genome sequences for each virus. The predicted frequency is preceded by the actual frequency and followed by the percent of predicted frequency. In all instances, the standard deviation was <8.2%. Note: The predicted values differ due to the G + C content of the virus and the size of the genome (see Table 5).

nome showing high G + C, interspersed with abrupt peaks of A + T-rich regions (Fig. 1). Analysis of these data (Table 5) shows that coding sequences have a G + C content of ~60% and the *hr*s are only slightly lower at 57.4%. In contrast, the sequences located between reading frames have a much lower G + C content (40.4%). Similar patterns of G + C distribution are found in AcMNPV, BmNPV, and OpMNPV. AcMNPV and BmNPV have a much lower G + C content (~40%) and show significantly lower intergenic G + C contents (~30%). OpMNPV also follows this pattern (coding sequences, ~57%; intergenic sequences, ~38% G + C) (Table 5). Baculovirus genes normally have AT-rich promoter sequences. Early genes are transcribed by the host cell RNA polymerase II, which uses TATA promoter sequences, and late genes are transcribed from the late promoter TAAG sequence (most often ATAAG) and are frequently surrounded by A + T-rich sequences. In addition, translational initiation and termination sequences are A + T rich. Baculoviruses appear to use the host cell 3′ processing machinery, which often requires signals such as AATAAA. In addition, it was found that an A + T-rich sequence was required for the utilization of an LdMNPV *hr* as an origin of DNA replication in a transient replication assay (Pearson and Rohrmann, 1995). Therefore, although LdMNPV has evolved a very high G + C content in the coding sequences, it appears that it may have maintained A + T-rich intergenic regions to preserve regulatory signals or because enzymes involved in processes using these regions function more efficiently in low G + C environments. In contrast, the translation of coding sequences may not be under such constraints.

Late promoter frequency

A unique feature of baculovirus late transcription is the late promoter element that serves as both the promoter and the mRNA start site. This sequence contains TAAG preceded by A, G, or T (Rankin et al, 1988; Rohrmann, 1986). To examine the conservation and distribution of this sequence, the genome sequences of three baculoviruses were randomly shuffled, and the frequency of each variant of the NTAAG sequence was calculated and compared with its actual occurrence (Table 6). In all instances, the sequence occurs less frequently than predicted; for example, in LdMNPV, 892 NTAAGs are predicted, but only 283, or 32%, of the predicted number are present. Similar values were calculated for AcMNPV and OpMNPV. CTAAG, which has not been documented to be a late promoter, is present at ~15% that predicted. This suggests that they are not used as late promoters and are selected against because of the disruptive effect they may have on viral gene expression or replication. The very low frequency of the CTAAG sequence suggests that it may also have a disruptive effect on late transcription. The viruses had similar frequencies of GTAAG (~30%), but contrary to the prediction, ATAAG occurred more often in the G + C-rich viruses (50–60%). The frequency of TTAAG showed no clear pattern.

Genome content

The LdMNPV genome is ~30 kb larger than the three other sequenced baculovirus genomes (AcMNPV, BmNPV, OpMNPV) (Table 5). The sources of the DNA that contribute to the large size of the LdMNPV genome are summarized in Fig. 8. The LdMNPV genome contains 163 putative ORFs; these include homologs of 94 AcMNPV genes (Table 2). Excluding the repeated *bro* genes, these ORFs represent ~90 kb or ~56% of the DNA. A major source of additional genetic information results from gene amplification. The 16 copies of AcMNPV ORF2 homologs called *bro* genes contribute ~14.5 kb, or 9%, of

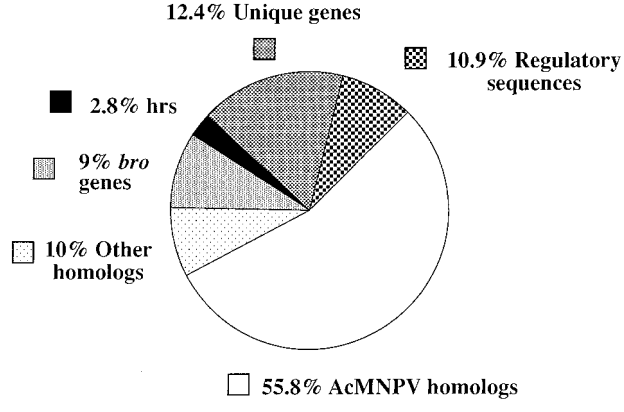


FIG. 8. The sources of LdMNPV genome content. The chart shows the relative contribution in genetic content from a variety of sources.

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 VEF, ribonucleotide reductase subunits, and dutpase, contribute an additional 16 kb (10%) to LdMNPV. The LdMNPV genome contains ~35 ORFs with no homologs in the database. These require ~12% of the coding capacity. In addition, LdMNPV *hr*s compose an additional 4.6 kb, or 2.8%, of DNA. Dividing the remaining DNA (~17.5 kb) (Table 5) by the number of ORFs (163) results in ~106 bp of DNA/gene for 5′ and 3′ regulatory sequences (Fig. 8). This is about twice the amount present in AcMNPV (Table 5). 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.

MATERIALS AND METHODS

Source and cloning of DNA

LdMNPV cosmids were constructed using partial *Pst*I or *Cla*I digests of DNA from the LdMNPV clonal isolate Cl 5–6 (Slavicek, 1991) and then cloned into the cosmid vector pHC79 (Hohn and Collins, 1980). LdMNPV strain Cl 5–6 is a few-polyhedra (FP) mutant (Bischoff and Slavicek, 1996), and this is reflected in a frameshift in Ld-ORF63, the *fp* locus. All plasmid subcloning was done in pBluescribe$^-$ (pBS−) or pBluescript$^-$ (pKS−) (Stratagene). pBS− was modified by the addition of a *Bgl*II site (Gombart *et al.,* 1989).

DNA sequence determination

Plasmid templates for sequencing were prepared using Qiagen columns (Qiagen Inc.). Sequencing reactions were performed using the Taq DyeDeoxy Terminator Cycle Sequencing Kit (Applied Biosystems, Inc.) according to the manufacturer's protocol, with the exception that the reactions were performed in 5% DMSO and cycled to a higher denaturation temperature (97°C). A Perkin-Elmer Cetus model 480 or model TC1 thermal cycler was used. Reactions were electrophoresed and analyzed on an ABI Model 377 Automated DNA Sequencer. Primers were designed to generate overlapping sequences of ~300–400 nucleotides in one direction. This sequence information was then confirmed using primers spaced at ~400–450 nucleotides apart for sequencing in the opposite direction. Sequencing was done using templates of plasmid, cosmid, and viral genomic DNA.

Sequencing problems encountered

The sequences that we determined correspond closely to the sizes estimated from restriction enzyme fragments. However, difficulty was encountered in sequencing the region extending from 67433 to 68298. The original cosmid derived from the genome lacked this region, and we were unable to clone it into plasmid DNA. The position of the deleted region was preliminarily located by sequencing a cosmid (c77) from a related strain (LdMNPV A21) of the virus (Bischoff and Slavicek, 1996; Slavicek *et al.,* 1996). The deleted sequence was found to be located in a large palindromic region that was produced by two oppositely oriented *bro* members related to AcMNPV ORF2. This palindromic region is ~3.5 kb with ~70% palindrome matches within the sequence. The sequence that was deleted from the cosmid DNA, which we were unable to clone, contained an area that was located between the two oppositely oriented genes. We attempted to sequence through the region using primers to flanking regions and viral genomic DNA as the template. Additional sequence information was obtained using this strategy, but overlaps were not possible because the sequence stopped, possibly due to template heterogeneity at specific positions. We used PCR to amplify this region from genomic DNA using primers containing the sequences 67234–67253 (called aeb38) and 68847–68856 (called aeb161) that flanked two known *Xho*I sites and another one in the internal unsequenced region. PCR yielded a ladder of fragments with a major band at 1.6 kb. The 1.6-kb band was gel purified and, when digested with *Xho*I, produced two fragments that were cloned. The sequence from these clones overlapped the ends of the sequence that we had already derived from direct viral genome sequencing. The sequence we determined from the cloned PCR products closely matched the size of the original PCR product. In addition, the sequence we generated conforms in size to a *Not*I–*Hin*dIII restriction fragment from genomic DNA that spans this region. Therefore, we believe that the sequence that we obtained fully represents this region of the genome. However, the region appears to be unstable, and cloned virus may contain a mixture of different sequences in this region.

Polymerase chain reaction

PCR was performed according to the Perkin-Elmer GeneAmp protocol using Amplitaq DNA polymerase in 2 mM $MgCl_2$. Template DNA (10 ng) was amplified using 20 μM of each of the 5′ and 3′ primers. One cycle of 95°C for 120 s was followed by 35 cycles of 60 s at 95°C, 60 s at 60°C, and a final cycle of 7 min at 72°C. The resulting products were purified on a 1.5% TAE agarose gel and extracted using the QIAquick gel extraction kit (Qiagen).

DNA sequence analysis

DNA sequence data were analyzed using the GCG suite of sequence analysis programs (Devereux *et al.,* 1984). BLAST searching was performed using the Gap-BLAST search engine (Altschul *et al.,* 1997). For percent identity between homologous proteins, default BLAST parameters were used. Because the alignment was be-

tween single pairs of known homologs, the Gap-BLAST two-pass method and multiple-hits-only option were not used. Affine gaps were set to gap'open 9; gap'extend 2. Percent identities indicate the percent identical residues in the aligned portions of two sequence. The following considerations led to the designation of ORFs as likely coding sequences. ORFs with significant homology to genes in the database were selected using BLAST searches. ORFs with no homologs in the database and >50 amino acids were selected based on whether their codon use reflected the composition of a set of large LdMNPV ORFs that are conserved with other baculoviruses and, therefore, likely to be expressed (e.g., DNA polymerase, helicase, *lef*-8, etc.). If two ORFs overlapped and showed similar codon use, the larger was selected. ORFs with significant overlap of *hr*s were also excluded. Splicing was not taken into account because only one baculovirus gene (*ie-1*) has been shown to be spliced (Chisholm and Henner, 1988). Alternate start codon use is not suspected in baculoviruses; therefore, the possibility of alternate start codons was not investigated.

Searches for motifs within predicted peptides were performed using the Prosite database (Bairoch *et al.,* 1997). N-terminal signal and transmembrane domains were screened (Kyte and Doolittle, l982; Von Heijne, 1986) using SignalP (Nielsen *et al.,* 1997) and PHD (Rost and Sander, 1993).

Protein alignments were carried out using the program MACAW (Schuler *et al.,* 1991) with guidance from Gap BLAST pairwise alignments (Altschul *et al.,* 1997). Related blocks of sequence were scored when the observed sequence similarity had a probability of $<10^{-8}$ of occurring at random.

The *hr* profiles were made by recursive alignment of the 72-bp repeated region in LdMNPV *hr*1 to the entire LdMNPV genome. The 72-bp profile is a long quasisymmetrical repeat containing the *Mlu*I–*Mlu*I core palindrome of the LdMNPV *hr* regions. This is reflected by the appearance of similarity regardless of which strand was scanned. Similarly, the 30-bp baculovirus *hr* profile was produced by aligning to the published sequences of AcMNPV, BmNPV, CfMNPV, and OpMNPV. The final profiles were scanned against the LdMNPV sequence.

Nucleotide sequence accession number

The nucleotide sequence data reported in this paper will appear in the GSDB, DDBJ, EMBL, and NCBI nucleotide sequence databases under accession no. AF 081810.

ACKNOWLEDGMENTS

The authors thank the following for their assistance: Nancy Hayes-Plazolles and Melissa Mercer, cloning; Andrei Chabes and Isaac Pretorius, data interpretation; Ian Smith, editorial comments; and the Central Services Laboratory at Oregon State University, DNA sequencing. This research was supported by Cooperative Agreement 23–193 from the U.S. Forest Service and grants from the U.S. Department of Agriculture (95 37302-1920), National Science Foundation (MCB 960769), and National Institutes of Health (GM 53233C). This is Technical Report 11402 from the Oregon State University Agricultural Experiment Station.

REFERENCES

Ahrens, C. A., Leisy, D. J., and Rohrmann, G. F. (1996). Baculovirus DNA replication. *In* "DNA Replication in Eukaryotic Cells" (M. De Pamphilis, Ed.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Ahrens, C. H., Pearson, M. N., and Rohrmann, G. F. (1995). Identification and characterization of a second putative origin of DNA replication in a baculovirus of *Orgyia pseudotsugata*. *Virology* **207,** 572–576.

Ahrens, C. H., Russell, R., Funk, C. J., Evans, J. T., Harwood, S. H., and Rohrmann, G. F. (1997). The sequence of the *Orgyia pseudotsugata* multinucleocapsid nuclear polyhedrosis virus genome. *Virology* **229,** 381–399.

Altschul, S. F., Madden, T., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402.

Ayres, M. D., Howard, S. C., Kuzio, J., Lopez-Ferber, M., and Possee, R. D. (1994). The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. *Virology* **202,** 586–605.

Bairoch, A., Bucher, P., and Hofmann, K. (1997). The PROSITE database: Its status in 1997. *Nucleic Acids Res.* **25,** 217–221.

Birnbaum, M. J., Clem, R. J., and Miller, L. K. (1994). An apoptosis-inhibiting gene from a nuclear polyhedrosis virus encoding a polypeptide with Cys/His sequence motifs. *J. Virol.* **68,** 2521–2528.

Bischoff, D. S., and Slavicek, J. M. (1994). Identification and characterization of a protein kinase gene in the *Lymantria dispar* multinucleocapsid nuclear polyhedrosis virus. *J. Virol.* **68**, 1728–1736.

Bischoff, D. S., and Slavicek, J. M. (1995). Identification and characterization of an early gene in the *Lymantria dispar* multinucleocapsid nuclear polyhedrosis virus. *J. Gen. Virol.* **76**, 2933–2940.

Bischoff, D. S., and Slavicek, J. M. (1996). Characterization of the *Lymantria dispar* nucleopolyhedrovirus 25K FP gene. *J. Gen. Virol.* **77,** 1913–1923.

Bischoff, D. S., and Slavicek, J. M. (1997). Molecular analysis of an *enhancin* gene in the *Lymantria dispar* nuclear polyhedrosis virus. *J. Virol.* **71,** 8133–8140.

Bjornson, R. M., Glocker, B., and Rohrmann, G. F. (1992). Characterization of the nucleotide sequence of the *Lymantria dispar* nuclear polyhedrosis virus DNA polymerase gene region. *J. Gen. Virol.* **73**, 3177–3183.

Bjornson, R. M., and Rohrmann, G. F. (1992). Nucleotide sequence of the p39-capsid gene region of the *Lymantria dispar* nuclear polyhedrosis virus. *J. Gen. Virol.* **73**, 1505–1508.

Bjornson, R. M., and Rohrmann, G. F. (1992). Nucleotide sequence of the polyhedron envelope protein gene region of the *Lymantria dispar* nuclear polyhedrosis virus. *J. Gen. Virol.* **73**, 1499–1504. [Author's correction (1994). *J. Gen. Virol*. **75,** 3687.]

Braunagel, S. C., He, H., Ramamurthy, P., and Summers, M. D. (1996). Transcription, translation, and cellular localization of three *Autographa californica* nuclear polyhedrosis virus structural proteins: ODV-E18, ODV-E35 and ODV-EC27. *Virology* **222**, 100–114.

Chisholm, G. E., and Henner, D. J. (1988). Multiple early transcripts and splicing of the *Autographa californica* nuclear polyhedrosis virus IE-1 gene. *J. Virol.* **62,** 3193–3200.

Choi, J., and Guarino, L. A. (1995). The baculovirus transactivator IE1 binds to viral enhancer elements in the absence of insect cell factors. *J. Virol.* **69,** 4548–4551.

Clem, R. J. (1997). Regulation of programmed cell death by baculoviruses. *In* "The Baculoviruses" (L. K. Miller, Ed.). Plenum Press, New York.

Cochran, M. A., and Faulkner, P. (1983). Location of homologous DNA sequences interspersed at five regions in the baculovirus AcMNPV genome. *J. Virol.* **45,** 961–970.

Crook, N. E., Clem, R. J., and Miller, L. K. (1993). An apoptosis-inhibiting baculovirus gene with a zinc finger-like motif. *J. Virol.* **67,** 2168–2174.

Derksen, A. C. G., and Granados, R. R. (1988). Alteration of a lepidopteran peritrophic membrane by baculoviruses and enhancement of viral infectivity. *Virology* **167,** 242–250.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12,** 387–395.

Eldridge, R., Li, Y., and Miller, L. K. (1992). Characterization of a baculovirus gene encoding a small conotoxin-like polypeptide. *J. Virol.* **66,** 6563–6571.

Fan, X., McLachlin, J. R., and Weaver, R. F. (1998). Identification and characterization of a protein kinase-interacting protein encoded by the *Autographa californica* nuclear polyhedrosis virus. *Virology* **240,** 175–182.

Foury, F., and Lahaye, A. (1987). Cloning and sequencing of the PIF gene involved in repair and recombination of yeast mitochondrial DNA. *EMBO J.* **6,** 1441–1449.

Garcia-Maruniak, A., Pavan, O. H. O., and Maruniak, J. E. (1996). A variable region of *Anticarsia gemmatalis* nuclear polyhedrosis virus contains tandemly repeated DNA sequences. *Virus Res.* **41,** 123–132.

Goldstein, D. J., and Weller, S. K. (1988). Herpes simplex virus type 1-induced ribonucleotide reductase activity is dispensable for virus growth and DNA synthesis: Isolation and characterization of an ICP6 lacZ insertion mutant. *J. Virol.* **62,** 196–205.

Gombart, A. F., Blissard, G. W., and Rohrmann, G. F. (1989). Characterization of the genetic organization of the *Hin*dIII-M region of the multicapsid nuclear polyhedrosis virus of *Orgyia pseudotsugata* reveals major differences among baculoviruses. *J. Gen. Virol.* **70,** 1815–1828.

Gomi, S., Zhou, C. E., Yih, W., Majima, K., and Maeda, S. (1997). Deletion analysis of four of eighteen late gene expression factor gene homologues of the baculovirus, BmNPV. *Virology* **230,** 35–47.

Guarino, L. A., and Summers, M. D. (1986). Interspersed homologous DNA of *Autographa californica* nuclear polyhedrosis virus enhances delayed-early gene expression. *J. Virol.* **60,** 215–223.

Harvey, A. J., Bidwai, A. P., and Miller, L. K. (1997). Doom, a product of the Drosophila mod(mdg4) gene, induces apoptosis and binds to baculovirus inhibitor-of-apoptosis proteins. *Mol. Cell. Biol.* **17,** 2835–2843.

Hashimoto, Y., Corsaro, B. G., and Granados, R. R. (1991). Location and nucleotide sequence of the gene encoding the viral enhancing factor of the *Trichoplusia ni* granulosis virus. *J. Gen. Virol.* **72,** 2645–2651.

Hawtin, R. E., Zarkowska, T., Arnold, K., Thomas, C. J., Gooday, G. W., King, L. A., Kuzio, J. A., and Possee, R. D. (1997). Liquefaction of *Autographa californica* nucleopolyhedrovirus-infected insects is dependent on the integrity of virus-encoded chitinase and cathepsin genes. *Virology* **238,** 243–253.

Hohn, B., and Collins, J. (1980). A small cosmid for efficient cloning of large DNA fragments. *Gene* **11,** 291–298.

Hong, T., Braunagel, S. C., and Summers, M. D. (1994). Transcription, translation, and cellular localization of PDV-E66: A structural protein of the PDV envelope of *Autographa californica* nuclear polyhedrosis virus. *Virology* **204**, 210–224.

Huang, M., and Elledge, S. J. (1997). Identification of *RNR4*, encoding a second essential small subunit of ribonucleotide reductase in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **17,** 6105–6113.

Kool, M., Ahrens, C., Goldbach, R. W., Rohrmann, G. F., and Vlak, J. M. (1994). Identification of genes involved in DNA replication of the *Autographa californica* baculovirus. *Proc. Natl. Acad. Sci. USA* **91,** 11212–11216.

Kool, M., Ahrens, C. H., Vlak, J. M., and Rohrmann, G. F. (1995). Replication of baculovirus DNA. *J. Gen. Virol.* **76,** 2103–2118.

Kyte, J., and Doolittle, R. F. (l982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol. 157,* 105–132.

Lambrechts, M. G., Bauer, F. F., Marmur, J., and Pretorius, I. S. (1996). Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohyphal differentiation in yeast. *Proc. Natl. Acad. Sci. USA* **93,** 8419–8424.

Lu, A., Krell, P., Vlak, J. M., and Rohrmann, G. F. (1997). Baculovirus DNA replication. *In* "The Baculoviruses" (L. K. Miller, Ed.). Plenum Press, New York.

Lu, A., and Miller, L. K. (1994). Identification of three late expression factor genes within the 33.8- to 43.4-map-unit region of *Autographa californica* nuclear polyhedrosis virus. *J. Virol.* **68**, 6710–6718.

Lu, A., and Miller, L. K. (1995). The roles of eighteen baculovirus late expression factor genes in transcription and DNA replication. *J. Virol.* **69,** 975–982.

Majima, K., Kobara, R., and Maeda, S. (1993). Divergence and evolution of homologous regions of *Bombyx mori* nuclear polyhedrosis virus. *J. Virol.* **67,** 7513–7521.

Mathews, C. K., and van Holde, K. E. (1995). "Biochemistry." The Benjamin/Cummings Publishing Co., Inc., Menlo Park, California.

McLachlin, J. R., and Miller, L. K. (1994). Identification and characterization of *vlf*-1, a baculovirus gene involved in very late gene expression. *J. Virol.* **68,** 7746–7756.

Mikhailov, V. S., Mikhailova, A. L., Iwanaga, M., Gomi, S., and Maeda, S. (1998). *Bombyx mori* nucleopolyhedrovirus encodes a DNA-binding protein capable of destabilizing duplex DNA. *J. Virol.* **72,** 3107–3116.

Monsma, S. A., Oomens, A. G. P., and Blissard, G. W. (1996). The gp64 envelope fusion protein is an essential baculovirus protein required for cell-to-cell transmission of infection. *J. Virol.* **70,** 4607–4616.

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Prot. Eng.* **10,** 1–6.

O'Reilly, D. R., and Miller, L. K. (1989). A baculovirus blocks insect molting by producing ecdysteroid UDP-glucosyl transferase. *Science* **245,** 1110–1112.

Ohkawa, T., Majima, K., and Maeda, S. (1994). A cysteine protease encoded by the baculovirus *Bombyx mori* nuclear polyhedrosis virus. *J. Virol.* **68,** 6619–6625.

Olivera, B. M., Miljanich, G. P., Ramachandran, J., and Adams, M. E. (1994). Calcium channel diversity and neurotransmitter release: The omega-conotoxins and omega-agatoxins. *Annu. Rev. Biochem.* **63,** 823–867.

Olszewski, J., and Miller, L. K. (1997). Identification and characterization of a baculovirus structural protein, VP1054, required for nucleocapsid formation. *J. Virol.* **71**, 5040–5050.

Passarelli, A. L., and Miller, L. K. (1994). Identification and transcriptional regulation of the baculovirus *lef*-6 gene. *J. Virol.* **68**, 4458–4467.

Passarelli, A. L., Todd, J. W., and Miller, L. K. (1994). A baculovirus gene involved in late gene expression predicts a large polypeptide with a conserved motif of RNA polymerases. *J. Virol* **68,** 4673–4678.

Pearson, M. N., Bjornson, R. M., Pearson, G. D., and Rohrmann, G. F. (1992). The *Autographa californica* baculovirus genome: Evidence for multiple replication origins. *Science* **257,** 1382–1384.

Pearson, M. N., and Rohrmann, G. F. (1998). Characterization of a baculovirus encoded ATP-dependent DNA ligase. *J. Virol.* **72,** 9142–9149.

Pearson, M. N., and Rohrmann, G. F. (1995). *Lymantria dispar* nuclear polyhedrosis virus homologous regions: Characterization of their ability to function as replication origins. *J. Virol.* **69,** 213–221.

Pearson, M. N., and Rohrmann, G. F. (1997). Splicing is required for transactivation by the immediate early gene 1 of the *Lymantria dispar* multicapsid nuclear polyhedrosis virus. *Virology* **235,** 153–165.

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.

Riegel, C. I., Lanner-Herrera, C., and Slavicek, J. M. (1994). Identification and characterization of the ecdysteroid UDP-glucosyl transferase gene of the *Lymantria dispar* multinucleocapsid nuclear polyhedrosis virus. *J. Gen. Virol.* **75,** 829–838.

Rohrmann, G. F. (1986). Polyhedrin structure. *J. Gen. Virol.* **67,** 1499–1513.

Roncarati, R., and Knebel-Morsdorf, D. (1997). Identification of the early

actin-rearrangement-inducing factor gene, arif-1, from *Autographa californica* multicapsid nuclear polyhedrosis virus. *J. Virol.* **71**, 7933–7941.

Rost, B., and Sander, C. (1993). Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* **232,** 584–599.

Russell, R. L. Q., Funk, C. J., and Rohrmann, G. F. (1997). Association of a baculovirus encoded protein with the capsid basal region. *Virology* **227**, 142–152.

Russell, R. L. Q., and Rohrmann, G. F. (1997). Characterization of baculovirus P91, a virion-associated protein. *Virology* **233**, 210–223.

Schuler, G. D., Altschul, S. F., and Lipman, D. J. (1991). A workbench for multiple alignment construction and analysis. *Proteins* **9,** 180–190.

Senkevich, T. G., Bugert, J. J., Sisler, J. R., Koonin, E. V., Darai, G., and Moss, B. (1996). Genome sequence of a human tumorigenic poxvirus: Prediction of specific host response-evasion genes. *Science* **273,** 813–816.

Slavicek, J. M. (1991). Temporal analysis and spatial mapping of *Lymantria dispar* nuclear polyhedrosis virus transcripts and in vitro translation products. *Virus Res.* **20,** 223–236.

Slavicek, J. M., Mercer, M. J., Kelly, M. E., and Hayes-Plazolles, N. (1996). Isolation of a baculovirus variant that exhibits enhanced polyhedra production stability during serial passage in cell culture. *J. Invertebr. Pathol.* **67,** 153–160.

Smith, I. R. L., van Beek, N. A. M., Podgwaite, J. D., and Wood, H. A. (1988). Physical map and polyhedrin gene sequence of *Lymantria dispar* nuclear polyhedrosis virus. *Gene* **71,** 97–105.

Theilmann, D. A., Chantler, J. K., Stewart, S., Flipsen, H. T. M., Vlak, J. M., and Crook, N. E. (1996). Characterization of a highly conserved baculovirus structural protein that is specific for occlusion-derived virions. *Virology* **218**, 148–158.

Theilmann, D. A., and Stewart, S. (1992). Tandemly repeated sequence at the 3′ end of the IE-2 gene of the baculovirus *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus is an enhancer element. *Virology* **187,** 97–106.

Thiem, S. M., Du, X., Quentin, M. E., and Berner, M. L. (1996). Identification of a baculovirus gene that promotes *Autographa californica* nuclear polyhedrosis virus replication in a nonpermissive insect cell line. *J. Virol.* **70**, 2221–2229.

Todd, J. W., Passarelli, A. L., and Miller, L. K. (1995). Eighteen baculovirus genes, including lef-11, p35, 39K, and p47, support late gene expression. *J. Virol.* **69**, 968–974.

van Strien, E. A., Faktor, O., Hu, Z. H., Zuidema, D., Goldbach, R. W., and Vlak, J. M. (1997). Baculoviruses contain a gene for the large subunit of ribonucleotide reductase. *J. Gen. Virol.* **78,** 2365–2377.

Von Heijne, G. (1986). A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* **14,** 4683–4690.

Vucic, D., Kaiser, W., J, Harvey, A. J., and Miller, L. K. (1997). Inhibition of reaper-induced apoptosis by interaction with inhibitor of apoptosis proteins (IAPs). *Proc. Natl. Acad. Sci. USA* **94,** 10183–10188.

Wang, P. J., Chabes, A., Casagrande, R., Tian, X. D., Thelander, L., and Huffaker, T. C. (1997). Rnr4p, a novel ribonucleotide reductase small-subunit protein. *Mol. Cell. Biol.* **17,** 6114–6121.

Xie, W.-D., Arif, B., Dobos, P., and Krell, P. J. (1995). Identification and analysis of a putative origin of DNA replication in the *Choristoneura fumiferana* multinucleocapsid nuclear polyhedrosis virus genome. *Virology* **209,** 409–419.

Yozawa, T., Kutish, G. F., Afonso, C. L., Lu, Z., and Rock, D. L. (1994). Two novel multigene families, 530 and 300, in the terminal variable regions of African swine fever virus genome. *Virology* **202,** 997–1002.