Research article

Relatedness of baculovirus and gypsy retrotransposon envelope proteins

George F. Rohrmann*[1] and P. Andrew Karplus[2]

Address: [1]Departments of Microbiology, Oregon State University, Corvallis Oregon, USA and [2]Biochemistry and Biophysics, Oregon State University, Corvallis Oregon, USA

E-mail: George F. Rohrmann* - rohrmann@bcc.orst.edu; P. Andrew Karplus - karplusp@ucs.orst.edu
*Corresponding author

Abstract

Background: Current evidence suggests that lepidopteran baculoviruses may be divided into two phylogenetic groups based on their envelope fusion proteins. One group utilizes gp64, a low pH-dependent envelope fusion protein, whereas the other employs a protein family (e.g. LD130 in the *Lymantria dispar* nucleopolyhedrovirus) unrelated to gp64, but that is also low pH-dependent. Database searches with members of the LD130 protein family often record significant levels of homology to envelope proteins from a number of insect retrovirus-like transposable elements of the *gypsy* class. In this report, the significance of the homology between these two types of envelope proteins is analyzed.

Results: The significance of the alignment scores was evaluated using Z-scores that were calculated by comparing the observed alignment score to the distribution of scores obtained for alignments after one of the sequences was subjected to 100 random shuffles of its sequence. These analyses resulted in Z-scores of >9 for members of the LD130 family when compared to most *gypsy* envelope proteins. Furthermore, in addition to significant levels of sequence homology and the presence of predicted signal sequences and transmembrane domains, members of this family contain a possible a furin cleavage motif, a conserved motif downstream of this site, predicted coiled-coil domains, and a pattern of conserved cysteine residues.

Conclusions: These analyses provide a link between envelope proteins from a group of insect retrovirus-like elements and a baculovirus protein family that includes low-pH-dependent envelope fusion proteins. The ability of gypsy retroelements to transpose from insect into baculovirus genomes suggests a pathway for the exchange of this protein between these viral families.

Background

Baculoviruses are a diverse family of insect viruses that are pathogenic for insects particularly members of the Lepidoptera, Diptera and Hymenoptera. They have large double-stranded, circular, supercoiled, DNA genomes of 100-180 kb and are characterized by the occlusion of their virions in crystalline protein matrices. There are two genera of baculoviruses, the nucleopolyhedroviruses (NPVs) in which many virions are occluded in large polyhedron-shaped occlusion bodies, and the granuloviruses (GVs) which normally occlude a single nucleocapsid per small granular occlusion body. A novel feature of these viruses is the production of two types of virions [1]. One, the occlusion derived virion (ODV) type, is present

in occlusion bodies and spreads the infection between insects, whereas the other, the budded virion (BV) type, spreads the infection between cells within insects or in cell culture. A major difference between these two types of virions is their envelope proteins. The envelope of ODV is not well characterized; and although it may be composed of multiple proteins, the mechanism by which it facilitates the initiation of infection of insect midgut cells is unclear [reviewed in [2]]. In contrast, BV envelopes contain an envelope fusion protein that causes the merging of the virion envelope and the membrane of cellular endocytic vesicles when exposed to low pH.

Current evidence suggests that lepidopteran baculoviruses may be divided into two phylogenetic groups based on the envelope fusion proteins of budded virions [3]. One group that includes *Autographa californica* multinucleocapsid nucleopolyhedrovirus (AcMNPV) and *Orgyia pseudotsugata* MNPV (OpMNPV), contain GP64 in their budded virion envelopes [4, 5]. Homologs of *gp64* are also found in a genus of orthomyxoviruses, the thogotoviruses [6]. Recently a number of complete baculovirus genome sequences have been described and were found to lack an open reading frame with homology to *gp64*. Analysis of the genome of the *Lymantria dispar* MNPV (LdMNPV), revealed a single orf (*ld130*) with predicted signal and transmembrane domains. Characterization of LD130 indicated that it localizes to the membrane of infected cells, is a component of budded virions, and is N-glycosylated. Uninfected cells transiently transfected with a plasmid encoding LD130, showed localization of the protein to the cell membrane and low-pH mediated cell fusion suggesting that LD130 is the envelope fusion protein of LdMNPV [3]. Similar results have been reported for the LD130 homolog in SeMNPV [7].

Homologs of *ld130* are found in the genomes of all the sequenced *gp64*-minus viruses including LdMNPV, SeMNPV [8], *Plutella xylostella* GV (PxGV) [9] and *Xestia c-nigrum* GV (XcGV) [10]. In addition, homologs of *ld130* are also found in *gp64*-containing viruses, although they do not appear to be capable of mediating low-pH-dependent membrane fusion (Pearson et al, unpublished). In contrast to the close relatedness of baculovirus gp64 homologs, the LD130 homologs are highly variable suggesting that *gp64* was recently incorporated into a baculovirus genome where it displaced the envelope fusion function of the *ld130* homologs [3].

Database searches with LD130 and its homologs routinely identify, not only homologous baculovirus proteins, but some members also showed significant levels of homology with the predicted envelope proteins of a number of insect retrovirus-like elements from *Drosophila* and Lepidoptera [11], (also see genbank documentation for AcMNPV orf23 in the AcMNPV sequence [12]). Evidence suggests that at least some of these retrovirus-like elements are infectious and have been classified as insect retroviruses [13,14,15] or errantiviruses [16]. In this report, we provide evidence for an evolutionary link between this newly characterized family of baculovirus envelope proteins and the envelope proteins predicted for a variety of insect retrovirus-like retrotransposons. These relationships suggest a possible common origin of the envelope fusion proteins for these two groups of viruses.

Results
Database searches with members of the LD130 family
Routine BLAST searches of the database with individual members of the LD130 family of baculovirus proteins revealed high levels of similarity with a set of predicted envelope proteins from a number of insect retrovirus-like retrotransposons. For example, an LD130 homolog from PxGV, PX26 [9] showed significance values of 10^{-10} with the envelope protein of TED, a lepidopteran retrovirus-like retrotransposon [17], and in the 10^{-4} to 10^{-10} range for a number of other predicted envelope proteins from retrovirus-like retrotransposons from other insects, particularly members from the genus *Drosophila*. Searches with other members of the ld130 family showed values of 10^{-4} to 10^{-6} to these proteins. A *Drosophila* orf (see below) that is not part of a retrovirus-like element was also identified. These values suggested that the relatedness was significant and may be evidence of an exchange of genes encoding this important category of proteins between insect cells and members of the baculovirus and retrovirus families. To investigate this relationship further, we performed BLAST searches of the database with all members of the baculovirus LD130 family and also with all sequences that showed alignments with E values of less than 10^{-3} to members of this family. When sequences within this category were identified that were over 90% identical, only one was included. Similarly, fragments from within complete sequences that were identified by this search were not included. This resulted in a total of 17 proteins used in our analyses (Table 1).

With one exception, all the non-baculovirus sequences showing the best scores are derived from retrovirus-like retrotransposons from insects that fall in the T3-*gypsy* group [18]. The exception is a *Drosophila* orf called cg4715 (Table 1). The orfs adjacent to cg4715 (two upstream and two downstream) appear to be non-homologous to retrovirus orfs (data not shown) suggesting that it may either be an insect gene or a remnant of an integrated retroelement.

Table 1: Sequences used for analysis

	Name/Source	Accession number	LearnCoil-VMF score	Reference
Baculovirus				
1.	ld130	af081810	0.57	[32]
2.	se8	af169823	0.56	[8]
3.	xc27	af162221	0.51	[10]
4.	px26	af270937	0.37	[9]
5.	ac23	l22858	0.26* and 0.63	[12]
6.	op21	u75930	0.23	[33]
Dipteran, Retrovirus-like				
7.	gypsy Ds	s64735	0.26	[34]
8.	gypsy Dv	s26841	0.47	[35]
9.	gypsy Dm	m12927	0.31	[36]
10.	yoyo	u60529	1.0	[37]
11.	tom	s34640	0	[38]
12.	17.6	x01472	0.44	[39]
13.	297	p20829	0	[40]
14.	idefix	aj009736	0.07	[41]
15.	zam	t13994	0.69* and 0.17	[42]
Lepidopteran, Retrovirus-like				
16.	ted	c36329	0.54*	[17]
Drosophila genome				
17.		cg4715	0	[43]

*Predicted coiled-coil domains are in an upstream location relative to the others.

Significance of the relatedness of the members of the Ld130 family

The potential relationships uncovered by the BLAST searches were further assessed for significance by the calculation of Z-scores for each of the sequences compared to each of the other sequences (see Methods). This led to the grouping of the sequences into six subfamilies based on Z-score cutoff values of ≥ 20 (Figs. 1, 2, Table 1). These include the baculovirus subfamily, two major divisions of the dipteran *gypsy* retrotransposon group with four members each, and three subfamilies each with a single member. The latter included one dipteran *gypsy* sequence, the lepidopteran *gypsy*-like element TED [19],

and the *Drosophila* cellular protein cg4715. We then compared each subfamily using the conservative criterion of a Z-score of ≥ 6 as an indicator of homology.

Furthermore, guided by the assumption that sequence features conserved over long evolutionary distances should also be conserved over short distances, a secondary criteria used for the more distant relationships was that Z-scores derived from comparing families of sequences should show enhanced significance over that of the individual sequences. The Z-scores revealed strong evidence for the homology of the seventeen sequences (Fig. 1). Members of the LD130 family showed Z-scores
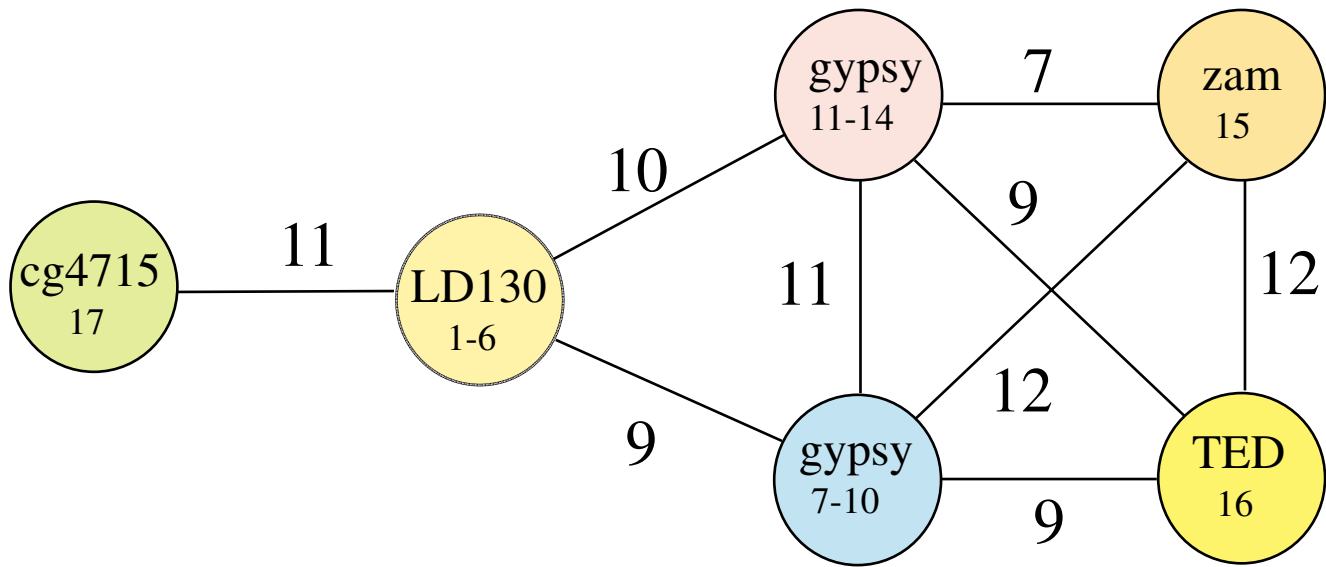
Figure 1
Significant sequence similarities link putative homologs of the LD130 family. Each circle represents one subfamily of sequences (defined as groups linked by Z-scores ≥ 20) and is labeled with a representative family member's name and the numbers from Table 1. The lines linking the circles are shown for pairs of subfamilies that give comparison Z-scores ≥ 6. The Z-scores between each subfamily are shown adjacent to the connecting line. Circles that are not connected by lines give comparison Z-scores < 6.

of 9 and 10 to two groups containing eight *gypsy* proteins. These are similar to scores determined for comparisons between different gypsy envelope proteins of from 9 to12. In addition, only one other protein [called rhoptry, from *Plasmodium yoelii* (accession # T28677)], fell within our 10^{-3} BLAST search cutoff in two instances (with T13994 and p20829 [Table 1]). However, when aligned to each of the 17 sequences, only one alignment resulted in a Z-score of 4 and all the other z-scores were ≤2. Also when members of each subfamily were grouped together and compared with this protein, the scores were not enhanced (data not shown). A matrix comparing the levels of sequence identities among these sequences is shown in Fig. 2, and reveals that in contrast to Z-scores, sequence identity is not a reliable indicator of evolutionary relatedness at the longer distances. For example, although the proteins with the groups composed of #1-4, #7-10 and #11-14 (Table 1) are clearly related to one another, their relationship with other members of the data set are unclear.

Features of the LD130 protein family
In addition to the overall sequence homology, our analyses revealed a number of other conserved features (Fig. 3). They all are predicted to be moderately large proteins of 348 to 649 amino acids and all contain predicted signal peptides at their amino terminus and transmembrane domains near their carboxyl terminus.

Another feature of these sequences is the conservation of cysteine residues. In the baculovirus sequences, 11 cysteines are conserved (Fig. 4). In TED, the lepidopteran retrovirus-like element, seven of the C-terminal cysteine residues are conserved. Likewise in the Drosophila sequence, cg4715, that is not a component of a retrotransposon, six of the C-terminal cysteines are conserved. With the two other categories of retroelements represented by *gypsy* and *tom* in Fig. 4, four of the cysteines are conserved. The consensus pattern of cysteine conservation is shown at the bottom of Fig. 4.

It has recently been reported that SE8, a homolog of LD130, is the likely membrane fusion protein for the *Spodoptera exigua* MNPV [7]. Furthermore, by N-terminal sequence analysis it was found that in budded virions, SE8 was truncated downstream of a motif that is a furin cleavage consensus sequence in other viruses. This motif has the sequence K/RXXK/R. One of the striking features of our alignments is that this site and downstream sequences appear to be conserved in these proteins (Fig. 5). The downstream sequence has the 21-amino acid motif gXΦBΦΦGXΦXKΦΦΦGXΦDXXD where Φ represents sites with predominantly hydrophobic amino acids, B represents D or N, and X represents any amino acid. The lower case of the initial g is related to its absense in one group of *gypsy* sequences (#7-10, Table 1). This motif is present in all the sequences except those represented by Ac23 and Op21 (Figs. 3, 5) and cg4715. Both AcMNPV and OpMNPV utilize GP64 as

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1)	100																
2)	38	100															
3)	24	25	100														
4)	24	23	35	100													
5)	13	15	14	15	100												
6)	13	15	15	15	37	100											
7)	14	10	13	11	10	9	100										
8)	11	10	13	11	9	9	86	100									
9)	11	9	11	12	9	11	74	75	100								
10)	11	10	12	9	6	7	22	20	19	100							
11)	16	13	14	16	10	10	13	13	12	15	100						
12)	16	12	13	15	10	8	14	14	13	16	63	100					
13)	15	12	12	15	10	8	16	15	14	15	62	83	100				
14)	13	12	12	12	10	7	15	14	14	13	32	34	33	100			
15)	17	15	12	16	12	9	15	15	15	14	19	19	18	15	100		
16)	14	12	14	15	12	13	14	12	12	10	14	15	14	12	20	100	
17)	13	13	15	17	13	12	11	10	10	9	15	13	13	12	17	11	100

Figure 2
Comparison matrix of the per cent identities of Ld130 family members. The members are represented by numbers from Table 1. The groups are boxed based on Z-scores of ≥ 20. The scores were determined from alignments using a gap penalty of 10 and the BLOSUM62 program [31].
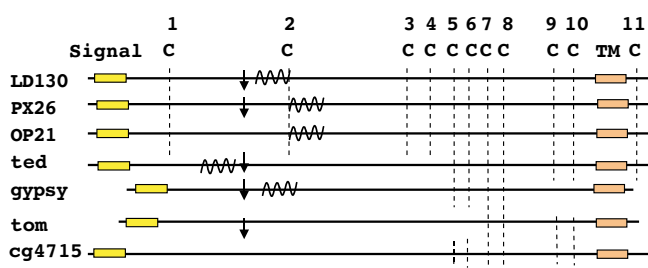

Figure 3
Diagram of the main features of representative members of the Ld130 protein family. The predicted signal and trans-membrane (TM) domains are shown. Conserved cysteines (C) are indicated by the vertical dashed line intersecting the maps of each protein group. The consensus cleavage site (K/RXXK/R) along with the downstream conserved sequences are shown by the vertical arrow (↓). The predicted coiled-coil domains are indicated by the coil-like structure. Representations are not drawn to scale.

their envelope fusion protein and neither OP21 nor AC23 appear to function as an envelope fusion protein in cell fusion assays (Pearson et al, unpublished). Likewise, cg4715 is not a component of a virus-like element. In contrast, the sequence is conserved in the other proteins that may play a role in membrane fusion. It has previously been suggested that for members of the *gypsy* family of env proteins the protease cleavage site may be downstream of this sequence [13, 20]. A similar cleavage site location was suggested for TED [21]. However, the major env components were never N-terminally sequenced and

the data could also be interpreted to be the result of the removal of about 12 kda from the N-termini of these proteins at the cleavage site shown in Fig. 5.

Coiled-coil-like motifs are a distinguishing feature of a number of viral fusion proteins. A program, Learncoil-VMF, has been developed to predict the location and the probability of these domains occurring within a protein [22]. We applied this program to the LD130 family of proteins. A summary of the scores is shown in Table 1 and the location of the coiled-coil domains is indicated in Fig. 6. Coiled-coil-like domains were predicted for many of these proteins immediately downstream of the predicted cleavage site. With two exceptions in which two regions were identified (Table 1), the program predicted a single coiled-coil region in a similar location, or in three instances, no such region. In the protein from the two NPVs that lack GP64 (LD130 and SE8), the same location was identified. However, the program predicted these regions slightly downstream for the two granulovirus proteins, XC27 and PX26, (these two viruses also lack gp64) placing them in the same position as predicted for proteins from NPVs that contain GP64. The baculovirus proteins had a probability of this motif occurring from 0.23 (Op21) to 0.63 (Ac23). In one set of gypsy retrovirus-like elements (#7-10), the probability of the coiled-coil motif downstream of the cleavage site ranged from 0.26 to 1.0. The rest of the sequences showed a high degree of variability. Two sequences (#15 and #16 [TED]), showed a strong probability (0.69 and 0.54, re-

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
Ac23		123	121	4	29	9	13	16	37	25	130
Op21		126	121	4	29	9	13	16	36	24	126
Ld130		125	117	5	29	7	14	18	34	27	128
Se8		123	115	5	27	7	14	17	34	31	130
Px26		101	115	5	25	7	14	12	34	21	115
Xc27		112	115	5	26	7	14	17	34	20	121
7						9	16	11			
8						9	16	11			
9						9	16	11			
10						8	12	11			
11								11	32	19	
12								11	32	19	
13								11	32	19	
14								11	32	19	
15						8	12				
16						8	13	13	32	17	101
17						8	15	12	32	20	
Spacing Consensus											
Baculovirus											
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
		101	115	4	25	7	13	12	34	20	115
		126	121	5	29	9	14	18	37	31	130
Retroelement											
						8	12	11	32	17	101
						9	16	13		20	

Figure 4
Spacing of conserved cysteines. The protein is indicated on the left (From Table 1) and the cysteine residues are shown on the top. The numbers indicate the number of amino acids between each conserved cysteine. The baculovirus and retrotransposon spacing consensus is shown at the bottom.

spectively) just upstream of the predicted cleavage site (Fig. 3). No.15 also showed a broad region of 0.17 probability just downstream of the cleavage site. With the exception of #12 (0.44), the remaining three sequences (#11,13, and 17) had a probability of 0. This included the cellular protein cg4715.

Discussion

In this report we have investigated the evolutionary relatedness between three categories of proteins, the retrovirus-like *gypsy* envelope proteins, the baculovirus LD130 group of envelope-associated proteins, and an insect cellular protein. Globally optimal sequence alignments in combination with assessment of the significance of the alignment scores resulted in striking patterns of relatedness. Not only did our analyses link the *gypsy*-derived envelope proteins together with a high level of significance as would be predicted from studies of other *gypsy* genes [18], but it also linked two of the major categories of the *gypsy* sequences to the baculovirus ld130 group with similar levels of significance. In addition, a cell-encoded protein from *Drosophila* showed relatedness at a high level of significance to the baculovirus ld130 protein group. The parameters we used as a basis for these analyses were conservative with cut-off Z-score values representative of at least 6 standard deviations above the value for the randomized sequences. In addition to the overall homology, a number of other features common to members of this class of proteins were identified. These included the predicted signal sequences and transmembrane domains and a striking conservation surrounding a possible protease cleavage site in the proteins that likely serve as envelope fusion proteins [7]. In addition, predicted coiled-coil domains in similar locations relative to the cleavage site in a number of the sequences further suggest both functional and structural similarities between members of this protein group.

Concurrent with the preparation of this research for publication, a similar study by Malik et al was reported [23]. They also describe a relationship between the envelope protein of gypsy retrovirus-like elements and the ld130 group of baculovirus envelope proteins. They identified the predicted signal and transmembrane domains that we describe in Fig. 3 and the conserved region that we compare in Fig. 5. Their alignments also highlight the conservation of the cysteine residues that we describe in Fig. 4.

Although our analyses convincingly indicate a relationship between the insect cellular, *gypsy*, and baculovirus membrane proteins, we can only speculate on the evolutionary pathway leading to the presence of members of this set of proteins in all three types of organisms. If homologs of the cg4715 gene are conserved among Diptera and Lepidoptera, the interrelationship between members of the baculoviridae and TED, the lepidopteran *gyp-*
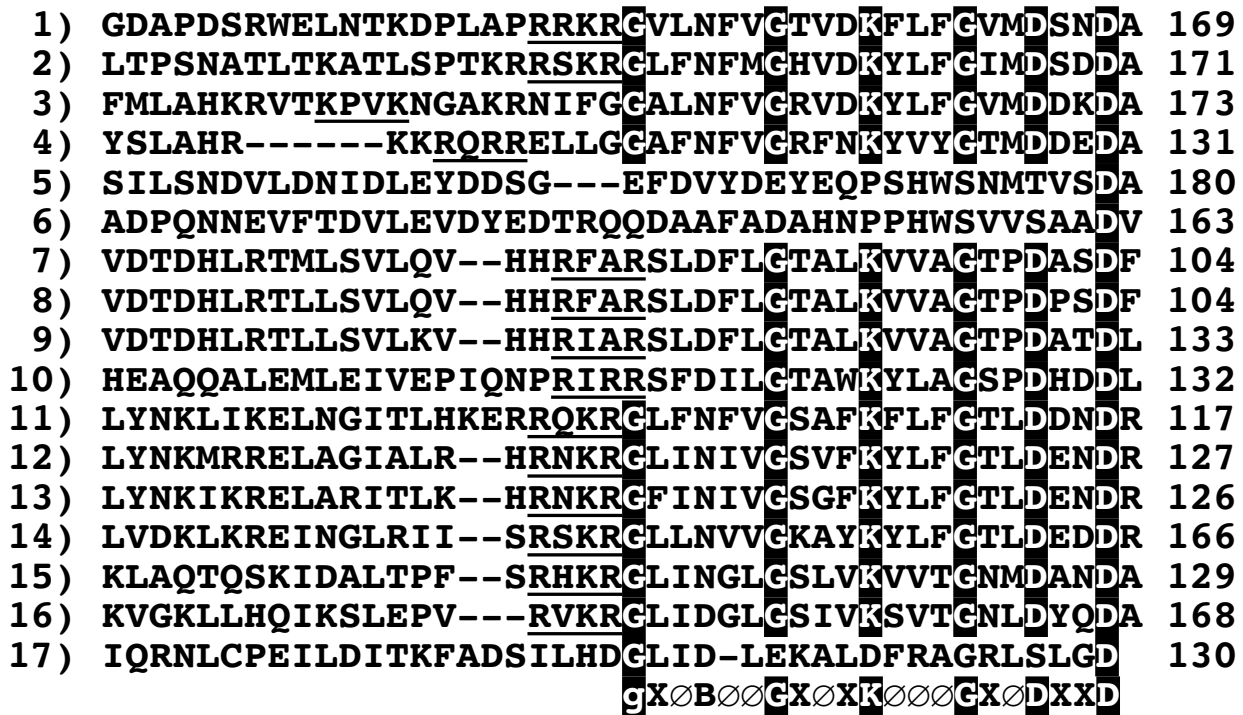
```
 1)  GDAPDSRWELNTKDPLAPRRKRGVLNFVGTVDKFLFGVMDSNDA  169
 2)  LTPSNATLTKATLSPTKRRSKRGLFNFMGHVDKYLFGIMDSDDA  171
 3)  FMLAHKRVTKPVKNGAKRNIFGGALNFVGRVDKYLFGVMDDKDA  173
 4)  YSLAHR------KKRQRRELLGGAFNFVGRFNKYVYGTMDDEDA  131
 5)  SILSNDVLDNIDLEYDDSG---EFDVYDEYEQPSHWSNMTVSDA  180
 6)  ADPQNNEVFTDVLEVDYEDTRQQDAAFADAHNPPHWSVVSAADV  163
 7)  VDTDHLRTMLSVLQV--HHRFARSLDFLGTALKVVAGTPDASDF  104
 8)  VDTDHLRTLLSVLQV--HHRFARSLDFLGTALKVVAGTPDPSDF  104
 9)  VDTDHLRTLLSVLKV--HHRIARSLDFLGTALKVVAGTPDATDL  133
10)  HEAQQALEMLEIVEPIQNPRIRRSFDILGTAWKYLAGSPDHDDL  132
11)  LYNKLIKELNGITLHKERRQKRGLFNFVGSAFKFLFGTLDDNDR  117
12)  LYNKMRRELAGIALR--HRNKRGLINIVGSVFKYLFGTLDENDR  127
13)  LYNKIKRELARITLK--HRNKRGFINIVGSGFKYLFGTLDENDR  126
14)  LVDKLKREINGLRII--SRSKRGLLNVVGKAYKYLFGTLDEDDR  166
15)  KLAQTQSKIDALTPF--SRHKRGLINGLGSLVKVVTGNMDANDA  129
16)  KVGKLLHQIKSLEPV---RVKRGLIDGLGSIVKSVTGNLDYQDA  168
17)  IQRNLCPEILDITKFADSILHDGLID-LEKALDFRAGRLSLGD   130
                           gXØBØØGXØXKØØØGXØDXXD
```

Figure 5
Sequence alignment near the protease cleavage site. The cleavage site consensus sequence is underlined and the conserved amino acids are boxed. The consensus sequence is shown at the bottom of the figure. B represents D or N and Φ indicates a hydrophobic amino acid. The numbers on the left are from Table I and the numbers on the right indicate the amino acid sequence coordinates.

sy element capable of integrating into both a host insect and a baculovirus genome, provides a clear pathway via DNA recombination for the exchange of members of the ld130 gene family between host cells, baculovirus genomes, and the transposable element. Indeed, a phylogenetic analysis of the envelope protein of these elements reported by Malik et al [23] suggests that they are derived from a monophyletic event that resulted from their incorporation of a baculovirus envelope protein gene in an ancestral member of these retrovirus-like elements.

TED has features that suggest that it may have specifically evolved to exploit a relationship with a baculovirus and this may have facilitated its acquisition of a baculovirus envelope protein gene. It is present as a mid-level repeated element in its host insect genome (about 50 copies/genome) [19, 24], it encodes a set of proteins (gag, pol, env) with features similar to those of well-characterized retrovirus proteins [17] and it is capable of producing virus-like particles [25]. A specific example of TED's adaptation to baculovirus biology is the presence of palindromes containing the baculovirus late promoter element in its long terminal repeat [17, 24]. Evidence suggests that this results in the high level transcription of the TED genome by the viral RNA polymerase [17] and could lead to the production of TED virus-like particles concomitant with the baculovirus late gene expression. This would provide a novel two-pronged escape mechanism from insects suffering from a fatal baculovirus infection; TED might be passed to other insects as an integrated transposable element in an infectious baculovirus, or as an infectious retrovirus. Oral transmission is the normal route for baculovirus infection and has been reported for *gypsy* retrovirus-like retrotransposons [15]. This strategy could elevate the retrotransposon's potential to exchange genetic material with the genomes of its viral and cellular hosts.

Members of the Ld130 family of baculovirus envelope proteins have been demonstrated to mediate low-pH membrane fusion [3]. In contrast, retroviruses normally enter cells by receptor-mediated fusion at the cell surface (reviewed in [26]) or by a combination of both receptor and low-pH mediated events [27]. Despite the differing mechanisms of cell entry, members of these two groups have been related by a number of structural criteria [28].

```
 1)  170  ----RE HDLAKT NALNEQ KEVTDE VNIAKFEEHKQCLERQRD--DLCGYATAKMALITEQLTQLDLLYTNLDRAVDDAL-D-NRINSLIMTPQRLYEE  263
 2)  172  ----HE HMLANT NSLNSQ KQLNDE IVLADY DHEFHASRMRDADKQCRYIIENYNILCKQLDEVATLYNKLDLAVDNAK-L-NHLNSFVVSPERLLNE  267
 3)  174  ----ELLYK LAIRENSTHYRIKQLTVDTIKITNIIE----ALKPELIQVSCMAVERK QHLRDT QGVARA TKLTNA NLSL-NKQRGSTSILSPLQLIDT  266
 4)  132  ----ELLYDVVRHENNTEYRVVELSQQTLKIANYIE----KVM-QSEAVTCFEINRK NYLRDS EDIEST NKIINA QMAL-NTNRVSPHIISPIQILEQ  223
 5)  181  ----QALLR-NPPKDRVMFLDTVTTSDVS--SKYEEYINCIVSNRTVENECMFLANM NVLNDK DDAAAL KMLERI KQTRKNKLNISNTVIDDDTLLTE  275
 6)  164  ----KELLAVAPPRDRVRVLPHISTANVA--DKYLKYEACINEERSADNECLYLTEM GVMASK ADAASF NTLDRL KQTRRN LNMTNN IDDDELLLRE  259
 7)  105  LKVRITEAQLVES SKQIII SETQKQ NRLTDT NKIISS KGDLVDTPHLFETLLARNRILNTEIQNLILTITLAKANI-----------  185
 8)  105  LRIRVTETQLVEA SKQILI SETQKQ NRLTDT NKIISS KGDLVDTPHLFETLLARNRILNSEIQNLVLTITLAKASI-----------  185
 9)  114  FKIKITEAQLVES SRQIAI SETQKQ NKLTDT NKVINA KGDLVDTPHLYEALLARNRMLSTEIQNLILTITLVKSNI-----------  194
10)  113  NMINNNLDKLITNNNNQIFINNAVKNQ NKLTEI NNILQS KNDNMI NEIALN QNRLRL KEE INIKYAIQWARLGI-----------  193
11)  118  IQFEEKLNSEAENSIKIHEFNEVMQFVNDGLQRIKKYENNRNSIDTLVYELMQFIEYIE-----------------------------------------  177
12)  128  VDIQRKLET AHNSVN HELNDA QLINDG QKIQNYENNSNII SLLYELMQFTEYIE-----------------------------------------  187
13)  127  VEIQKKLEINVHNSVKLHELNDAIRLINDGMQKIQNYENNHTIIDSLLFELMQFTEYIE-----------------------------------------  186
14)  167  EELEEKINNMSEDSVKTHDLNT LDVINS CIDIINK KVDKEQ QQIAVL FNLEQFTEYIE-----------------------------------------  229
15)  130  KNIETE NHLKSQ TTISDN EIQNSF DEVQLRFKNL RHINNE NLIKNF ENTQNT VTKIYNNEEEIKKLQYINRLNYNIDLLVSHL  221
16)  139  LKYDEALKTLQTNEGKLTSEFNSHLSLCKEWMSQHNKVLEQLTLNQIRVNATLELLLQKEAYRDYSLIKFAKFAQILGIITNNVEDLMLEI  230
17)  131  ------------------------------------DDVSIELGMGTSCIDSSINVINVILQEPFHEAYEPENLIMMKPYLYLMGSRL  182
```

Figure 6

Alignment of sequences that include a predicted coil-coil domain. The coiled-coil domain was predicted by Learncoil-VMF and the 'a' residue of the motif is boxed and the 'd' residue is underlined [22]. The probability for each domain is shown in Table 1. The numbers used to identify the sequences are defined in the Fig. 5 legend.

The link that we have described in this report and that of Malik et al [23] suggests the possibility of a genetic relationship as well.

Conclusions

The relatedness of a class of low pH-dependent baculovirus envelope fusion proteins to envelope proteins of *gypsy* retrovirus-like elements was demonstrated to be highly significant. Transposon mediated exchange provides a documented pathway for the movement of this gene between insect cells, and two different types of viruses.

Materials and Methods

Blast searches were performed using the Gap-Blast search engine [29]. Sequence alignments were performed with the program SEQUOIA by C.M Bruns which is available on the web at http://www.scripps.edu/~bruns/sequoia.html. SEQUOIA uses a conventional dynamic programming algorithm based on the work of Needleman and Wunsch [30] to find the globally optimal alignment given a particular residue comparison matrix and gapping model. The statistical significance of the optimal alignment was assessed by calculation of their Z-score. The Z-score is derived by comparison of an observed alignment score with the distribution of scores obtained for alignments after one of the sequences was shuffled to create a random sequence with the same amino acid composition. 100 shuffles were carried out for each sequence comparison and the significance was measured relative to the mean and standard deviation of scores from the shuffled sequences. Quantitatively, the Z-score is defined as the number of standard deviations of the observed score above the mean of the shuffled scores and is determined by the following equation:

Z-score = Observed score - mean shuffled score/Standard deviation of shuffled scores

In the work we describe here, alignments were done using a gap penalty of 10 and the BLOSUM62 scoring matrix [31]. An additional important feature of SEQUOIA is that a "sequence" need not be just a single sequence, but can be a prealigned family of sequences that are treated as a unit.

References

1. Volkman LE, Summers MD, Hsieh CH: **Occluded and nonoccluded nuclear polyhedrosis virus grown in Trichoplusia ni: comparative neutralization, comparative infectivity, and in vitro growth studies.** *J. Virol.* 1976, **19**:820-832
2. Funk CJ, Braunagel S, Rohrmann GF: **Baculovirus Structure.** *In: The Baculoviruses Edited by Miller LK. pp. 2-32: Plenum; 1997, 2-32*
3. Pearson MN, Groten C, Rohrmann GF: **Identification of the Lymantria dispar nucleopolyhedrovirus envelope fusion protein provides evidence for a phylogenetic division of the Baculoviridae.** *J. Virol.* 2000, **74**:6126-6131
4. Whitford M, Stewart S, Kuzio J, Faulkner P: **Identification and sequence analysis of a gene encoding gp67 an abundant envelope glycoprotein of the baculovirus, Autographa californica nuclear polyhedrosis virus.** *J. Virol.* 1989, **63**:1393-1403
5. Blissard GW, Rohrmann GF: **Location, sequence, transcriptional mapping, and temporal expression of the gp64 envelope glycoprotein gene of the Orgyia pseudotsugata multicapsid nuclear polyhedrosis virus.** *Virology* 1989, **170**:537-555
6. Nuttall PA, Morse MA, Jones LD, Portela A: **Adaptation of members of the Orthomyxoviridae family to transmission by ticks.** *In: Molecular basis of virus evolution Edited by Gibbs AJ, Calisher CH, García-Arenal F. pp. 416-425. New York: Cambridge University Press; 1995, 416-425*
7. Ijkel WFJ, Westernberg M, Goldbach RW, Blissard GW, Vlak JM, Zuidema D: **A novel baclovirus envelope fusion protein with a proprotein convertase cleavage site.** *Virology* 2000, **274**:30-41

8.   Ijkel WFJ, van Strien EA, Jeldens JGM, Broer R, Zuidema D, Goldbach RW, Vlak JM: **Sequence and organization of the Spodoptera exigua multicapsid nucleopolyhedrovirus genome.** *J. Gen. Virol.* 1999, **80**:3289-3304

9.   Hashimoto Y, Hayakawa T, Ueno Y, Fugita T, Sano Y, Matsumoto T: **Sequence analysis of the Plutella xylostella granulovirus genome.** *Virology* 2000, **275**:358-372

10. Hayakawa T, Ko R, Okano K, Seong S, Goto C, Maeda S: **Sequence analysis of the Xestia c-nigrum granulovirus genome.** *Virology* 1999, **262**:277-297

11. Hayakawa T, Rohrmann GF, Hashimoto Y: **Patterns of genome organization and content in lepidopteran baculoviruses.** *Virology* 2000, **278**:1-12

12. Ayres MD, Howard SC, Kuzio J, Lopez-Ferber M, Possee RD: **The complete DNA sequence of Autographa californica nuclear polyhedrosis virus.** *Virology* 1994, **202**:586-605

13. Song SU, Gerasimova T, Kurkulos M, Boeke JD, Corces VG: **An env-like protein encoded by a Drosophila retroelement: evidence that gypsy is an infectious retrovirus.** *Genes Dev* 1994, **8**:2046-2057

14. Pelisson A, Teysset L, Chalvet F, Kim A, Prud'homme N, Terzian C, Bucheton A: **About the origin of retroviruses and the co-evolution of the gypsy retrovirus with the Drosophila flamenco host gene.** *Genetica* 1997, **100**:29-37

15. Kim A, Terzian C, Santamaria P, Pelisson A, Purd'homme N, Bucheton A: **Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of Drosophila melanogaster.** *Proc Natl Acad Sci U S A* 1994, **91**:1285-1289

16. Boeke JD, Eickbush TH, Sandmeyer SB, Voytas DF: **Virus Taxnomy: ICTV VIIth report.** *In: Edited by Murphy FA. New York: Spirnger-Verlag;* 1999,

17. Friesen PD, Nissen MS: **Gene organization and transcription of TED, a Lepidopteran retrotransposon integrated within the baculovirus genome.** *Mol. Cell. Biol.* 1990, **10**:3067-3077

18. Boeke JD, Stoye JP: **Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements.** *In: Retroviruses Edited by Coffin JM, Hughes SH, Varmus HE. pp. 343-435. Cold Spring Harbor, USA: Cold Spring Harbor Press;* 1997, 343-435

19. Miller DW, Miller LK: **A virus mutant with an insertion of a copia-like transposable element.** *Nature (London)* 1982, **299**:562-564

20. Song SU, Kurkulos M, Boeke JD, Corces VG: **Infection of the germ line by retroviral particles produced in the follicle cells: a possible mechanism for the mobilization of the gypsy retroelement of Drosophila.** *Development* 1997, **124**:2789-2798

21. Ozers MS, Friesen P: **The Env-like open reading frame of the baculovirus-integrated retrotransposon TED encodes a retrovirus-like envelope protein.** *Virology* 1996, **226**:252-259

22. Singh M, Berger B, Kim PS: **LearnCoil-VMF: computational evidence for coiled-coil-like motifs in many viral membrane-fusion proteins.** *J. Mol. Biol.* 1999, **290**:1031-1041

23. Malik HS, Henikoff S, Eickbush TH: **Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses.** *Genome Res.* 2000, **10**:1307-1318

24. Friesen PD, Rice WC, Miller DW, Miller LK: **Bidirectional transcription from a solo long terminal repeat of the retrotransposon TED: Symmetrical RNA start sites.** *Mol. Cell. Biol.* 1986, **6**:1599-1607

25. Lerch RA, Friesen PD: **The baculovirus-integrated retrotransposon TED encodes gag and pol proteins that assemble into viruslike particles with reverse transcriptase.** *J. Virol.* 1992, **66**:1590-1601

26. Hunter E: **Viral Entry and Receptors.** *In: Retroviruses Edited by Coffin JM, Hughes SH, Varmus HE. Cold Spring Harbor, USA: Cold Spring Harbor Press;* 1997,

27. Mothes W, Boerger AL, Narayan S, Cunningham JM, Young JA: **Retroviral entry mediated by receptor priming and low pH triggering of an envelope glycoprotein.** *Cell* 2000, **103**:679-689

28. Hughson FM: **Enveloped viruses: A commmon mode of membrane fusion?** *Curr. Biol.* 1997, **7**:565-569

29. Altschul SF, Madden T, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res.* 1997, **25**:3389-3402

30. Needleman SB, Wunsch CD: **A general method applicable to the search for similarities in the amino acid sequence of two proteins.** *J. Mol. Biol.* 1970, **48**:443-453

31. Henikoff S, Henikoff JG: **Amino acid substitution matrices from protein blocks.** *Proc Natl Acad Sci U S A* 1992, **89**:10915-10919

32. Kuzio J, Pearson MN, Harwood SH, Funk CJ, Evans JT, Slavicek J, Rohrmann GF: **Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria dispar.** *Virology* 1999, **253**:17-34

33. Ahrens CH, Russell R, Funk CJ, Evans JT, Harwood SH, Rohrmann GF: **The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear polyhedrosis virus genome.** *Virology* 1997, **229**:381-399

34. Alberola TM, de Frutos R: **Molecular structure of a gypsy element of Drosophila subobscura (gypsyDs) constituting a degenerate form of insect retroviruses.** *Nucleic Acids Res.* 1996, **4**:914-923

35. Mizrokhi LJ, Mazo AM: **Cloning and analysis of the mobile element gypsy from D. virilis.** *Nucleic Acids Res.* 1991, **19**:913-916

36. Marlor RL, Parkhurst SM, Corces VG: **The Drosophila melanogaster gypsy transposable element encodes putative gene products homologous to retroviral proteins.** *Mol. Cell. Biol.* 1986, **6**:1129-1134

37. Zhou Q, Haymer DS: **Molecular structure of yoyo, a gypsy-like retrotransposon from the mediterranean fruit fly, Ceratitis capitata.** *Genetica* 1998, **101**:167-178

38. Tanda S, Mullor JL, Corces VG: **The Drosophila tom retrotransposon encodes an envelope protein.** *Mol. Cell. Biol.* 1994, **14**:5392-5401

39. Saigo K, Kugimiya W, Matsuo Y, Inouye S, Yoshioka K, Yuki S: **Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in Drosophila melanogaster.** *Nature* 1984, **312**:659-661

40. Inouye S, Yuki S, Saigo K: **Complete nucleotide sequence and genome organization of a Drosophila transposable genetic element.** *Eur. J. Biochem.* 1986, **154**:417-415

41. Desset S, Conte C, Dimitri P, Calco V, Dastugue B, Vaury C: **Mobilization of two retroelements, ZAM and Idefix, in a novel unstable line of Drosophila melanogaster.** *Mol. Biol. Evol.* 1999, **16**:54-66

42. Leblanc P, Desset S, Dastugue B, Vaury C: **Invertebrate retroviruses: ZAM a new candidate in D.melanogaster.** *EMBO J.* 1997, **16**:7521-7531

43. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, P.W. L, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD: **The genome sequence of Drosophila melanogaster.** *Science* 2000, **287**:2185-2195