The structure of an insect parvovirus (*Galleria mellonella* densovirus) at 3.7 Å resolution

Alan A Simpson[1], Paul R Chipman[1†], Timothy S Baker [1], Peter Tijssen[2] and Michael G Rossmann[1*]

Background: Parvoviruses infect vertebrates, insects and crustaceans. Many arthropod parvoviruses (densoviruses) are highly pathogenic and kill approximately 90% of the host larvae within days, making them potentially effective as selective pesticides. Improved understanding of densoviral structure and function is therefore desirable. There are four different initiation sites for translation of the densovirus capsid protein mRNA, giving rise to the viral proteins VP1 to VP4. Sixty copies of the common, C-terminal domain make up the ordered part of the icosahedral capsid.

Results: The *Galleria mellonella* densovirus (*Gm*DNV) capsid protein consists of a core β-barrel motif, similar to that found in many other viral capsid proteins. The structure most closely resembles that of the vertebrate parvoviruses, but it has diverged beyond recognition in many of the long loop regions that constitute the surface features and intersubunit contacts. The N termini of twofold-related subunits have swapped their positions relative to those of the vertebrate parvoviruses. Unlike in the vertebrate parvoviruses, in *Gm*DNV there is no continuous electron density in the channels running along the fivefold axes of the virus. Electron density corresponding to some of the single-stranded DNA genome is visible in the crystal structure, but it is not as well defined as in the vertebrate parvoviruses.

Conclusions: The sequence of the glycine-rich motif, which occupies each of the channels along the fivefold axes in vertebrate viruses, is conserved between mammalian and insect parvoviruses. This motif may serve to externalize the N-terminal region of the single VP1 subunit per particle. The domain swapping of the N termini between insect and vertebrate parvoviruses may have the effect of increasing capsid stability in *Gm*DNV.

Addresses: [1]Department of Biological Sciences, Purdue University, West Lafayette, IN 47907-1392, USA and [2]Institut Armand-Frappier, Université du Québec, 531 Boulevard des Prairies, Laval, Québec, Canada H7V 1B7.

[†]Present address: Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue, Seattle, WA 98109, USA.

*Corresponding author.
E-mail: mgr@indiana.bio.purdue.edu

Key words: domain swapping, evolution, *Galleria mellonella*, parvovirus, X-ray diffraction

Received: **11 May 1998**
Revisions requested: **29 June 1998**
Revisions received: **5 August 1998**
Accepted: **7 August 1998**

Structure 15 November 1998, **6**:1355–1367
http://biomednet.com/elecref/0969212600601355

Introduction

The Parvoviridae family is a group of small viruses, of 250–280 Å diameter, that have single-stranded DNA genomes and are icosahedral. There are 60 protein subunits in each virion. An approximately 5.5 kb genome encodes the capsid protein and from one to three non-structural proteins. Members of the Parvovirinae and Densovirinae subfamilies infect vertebrates and arthropods, respectively [1]. The Densovirinae contain four genera distinguished by their genome organization and sequence similarities: Densovirus, Brevidensovirus, Iteravirus and Bidensovirus. Although many viruses either encode their own replication machinery or induce replication in the host cell, parvoviruses only replicate in dividing cells, such as in insect larvae, fetuses and tissues with a high turnover. For example: porcine parvovirus causes embryonic death [2]; the human virus, B19, infects erythrocyte precursor cells [3]; and insect parvoviruses infect the larvae of their host. Densoviruses, like other parvoviruses, are highly species-specific, possibly as a result of their extreme reliance on host function; this has prompted their use as highly specific pest-control agents [4,5]. The larvae of the greater wax moth, *Galleria mellonella* (*Gm*), are parasitic on honey bee colonies and are frequently infected with *Gm* densovirus (*Gm*DNV). The virus typically kills the host larvae within several days of infection. Some densoviruses infect only the mid-gut and are transmitted through the intestinal tract, whereas others (e.g. *Gm*DNV) infect nearly all tissues except those of the mid-gut [1].

The parvovirus genome encodes a single capsid protein, but transcriptional regulation and post-transcriptional modification produce proteins with differing lengths at their N termini. The relative levels of expression of these variants is controlled by alternative splicing in the vertebrate viruses and by a leaky scanning mechanism of alternative ribosome start sites in the insect viruses. In *Gm*DNV, the viral proteins VP1, VP2, VP3 and VP4 are present in the virion in an approximate 1:9:9:41 ratio for a

total of 60 subunits per particle [6,7]. These proteins all have the same C terminus and contain 811, 535, 488 and 437 amino acids, respectively.

Parvoviruses have local regions of conserved sequence [1], confirming their common ancestry, but the insect and vertebrate viruses differ in terms of genome organization and control of gene expression. The most conserved sequences are an initiator protein motif and a helicase superfamily III motif [1], which have been identified in the nonstructural proteins of all parvoviruses sequenced to date. In the largest capsid protein, VP1, there is a highly conserved, 39-residue motif within which the densoviruses and the vertebrate parvoviruses have approximately 70% sequence identity. This level of conservation is particularly noteworthy because there is only about 8% sequence identity between the insect and vertebrate viruses in the structurally ordered part of the capsid protein. Conservation of the 39-residue motif in VP1 could indicate a critical function, or it may signify a recent horizontal gene transfer between the vertebrate and insect viruses. The capsid proteins also contain a conserved, glycine-rich sequence that allows this motif to be located in channels along the virion fivefold axes, thus permitting externalization of the N-terminal end of the capsid proteins [8].

The crystal structures of three vertebrate parvoviruses, canine parvovirus (CPV) [8], feline panleukopenia virus (FPV) [9] and minute virus of mice (MVM) [10], have been determined to near-atomic resolution. The capsid protein of these parvoviruses consists of a β-barrel domain similar to that found in many other viruses. Long loops between the β strands form most of the intersubunit contacts and define the surface features of the virus. The polypeptide chain can be seen in the electron-density maps, beginning with the conserved polyglycine region. Although the polyglycine motif occurs in each of the 60 subunits, the size of the fivefold channels permits only one out of every five peptides to be present. The polypeptide preceding the polyglycine region is disordered and could be either situated outside the capsid or associated with the internal DNA. The structures of CPV and FPV have allowed the mapping of mutations that control host range to specific regions on the surface of the capsid [11], implying that these regions interact with host-specific factors at some point in the viral life cycle. Icosahedrally ordered single-stranded DNA (up to 23% of all bases in MVM [12]) is observed in the vertebrate parvoviruses. The icosahedral ordering of the single DNA strand may result from tight, non-sequence-specific binding of the DNA to the capsid protein.

Little is known about how parvoviruses enter cells or how the genome is subsequently transported to the nucleus. The human parvovirus B19 binds to the cell-surface glycolipid, globoside [13], at the protruding region on the threefold axis [14]. CPV and FPV can, however, enter certain nonpermissive cell types, showing that tissue tropism is not determined solely by the cellular receptor [15].

Particles of *Gm*DNV can be separated on CsCl gradients into three distinct fractions with densities of 1.44, 1.40 and 1.32 g/cm^3. The denser fractions contain infectious virions, whereas the lightest fraction consists of empty capsids. The 1.40 g/cm^3 form is stable in the presence of divalent cations, but removal of these cations induces a change in density to 1.44 g/cm^3. Presumably, a conformational change accompanies the binding of divalent ions, although the two denser types of capsid appear to have similar morphologies when negatively stained virions are viewed with an electron microscope [16]. The change in virus density, mediated by the presence or absence of calcium and magnesium ions, implies that the virions have a specific divalent-cation-binding site.

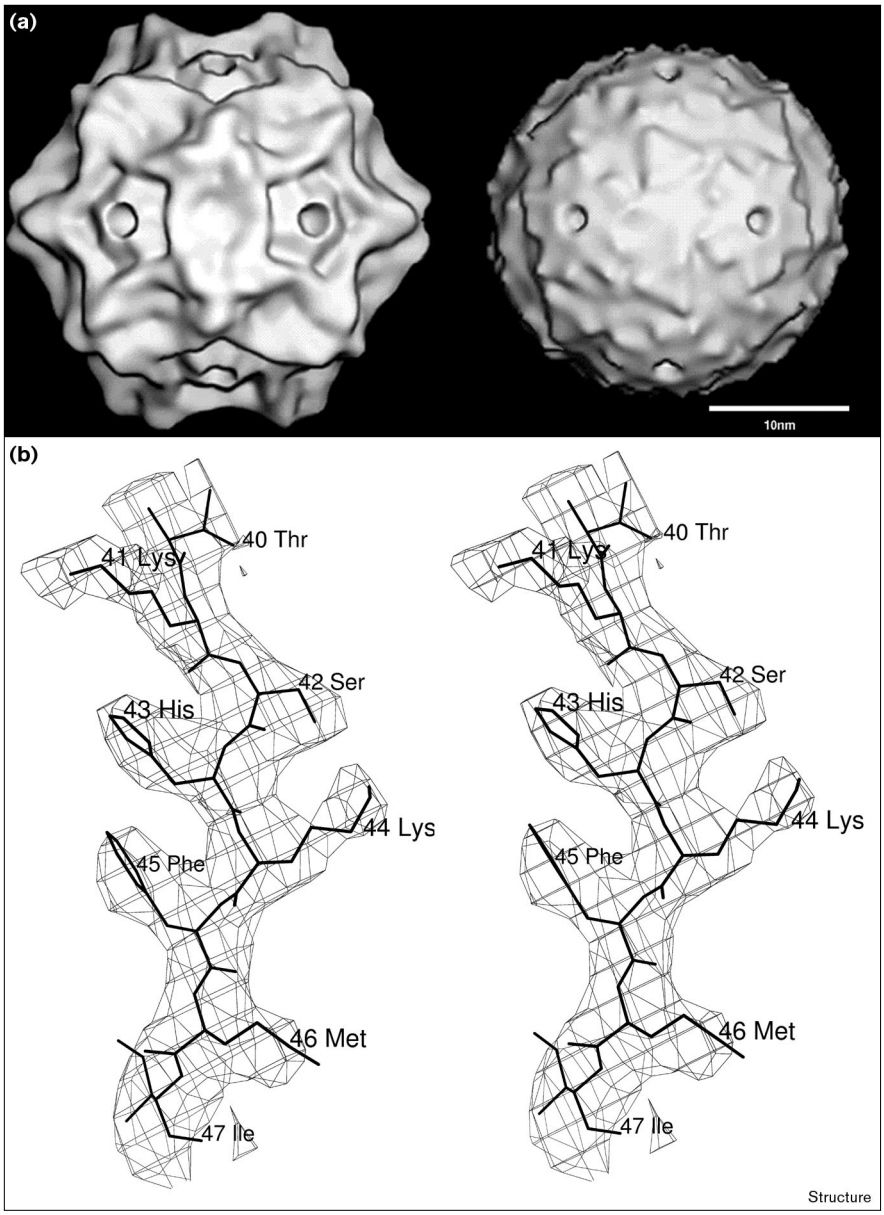
Here, we present the first near-atomic-resolution structure of an insect parvovirus, and show that its structure is completely different from vertebrate parvoviruses in most of the large loops connecting the β strands of the conserved β-barrel domain.

Results and discussion

Description of the structure

The *Gm*DNV structure was readily interpreted from a 3.7 Å, icosahedrally averaged, electron-density map. The mainchain was unambiguously traced from residue 22 to 436 of VP4 (Figure 1). The fold (Figure 2) consists of a core β-barrel domain, as found in many other virus capsid proteins, and many long loop insertions. A search of the Brookhaven Protein Data Bank (PDB), using the program DALI [17], showed that *Gm*DNV had the greatest similarity to the vertebrate parvoviruses, with a score of 9.7σ (where σ is a standard deviation) above a random hit. Other viral capsid proteins with β-barrel folds gave scores of 4.1σ or less above background (Table 1). The backbone of *Gm*DNV superimposes on that of CPV with a root mean square (rms) deviation of 2.2 Å for 148 equivalenced Cα atoms (or 36% of the structure), of which 100 are in the β barrel. When the icosahedral symmetry axes of *Gm*DNV and CPV are superimposed, the CPV β barrel must be rotated by 7.4° and translated radially outwards by 9.7 Å in order to superimpose it onto the β barrel of *Gm*DNV. The β-barrel domains in other viruses have somewhat greater differences in position and orientation in comparison to *Gm*DNV. In *Gm*DNV, the β strands are connected by long loops (up to 100 residues), which form most of the intersubunit contacts as well as the surface features of the capsid. Loops 1, 2 and 3 exist in both CPV and *Gm*DNV, although their structures bear little relationship to one another (Figures 2 and 3). In CPV, loop 3

Figure 1



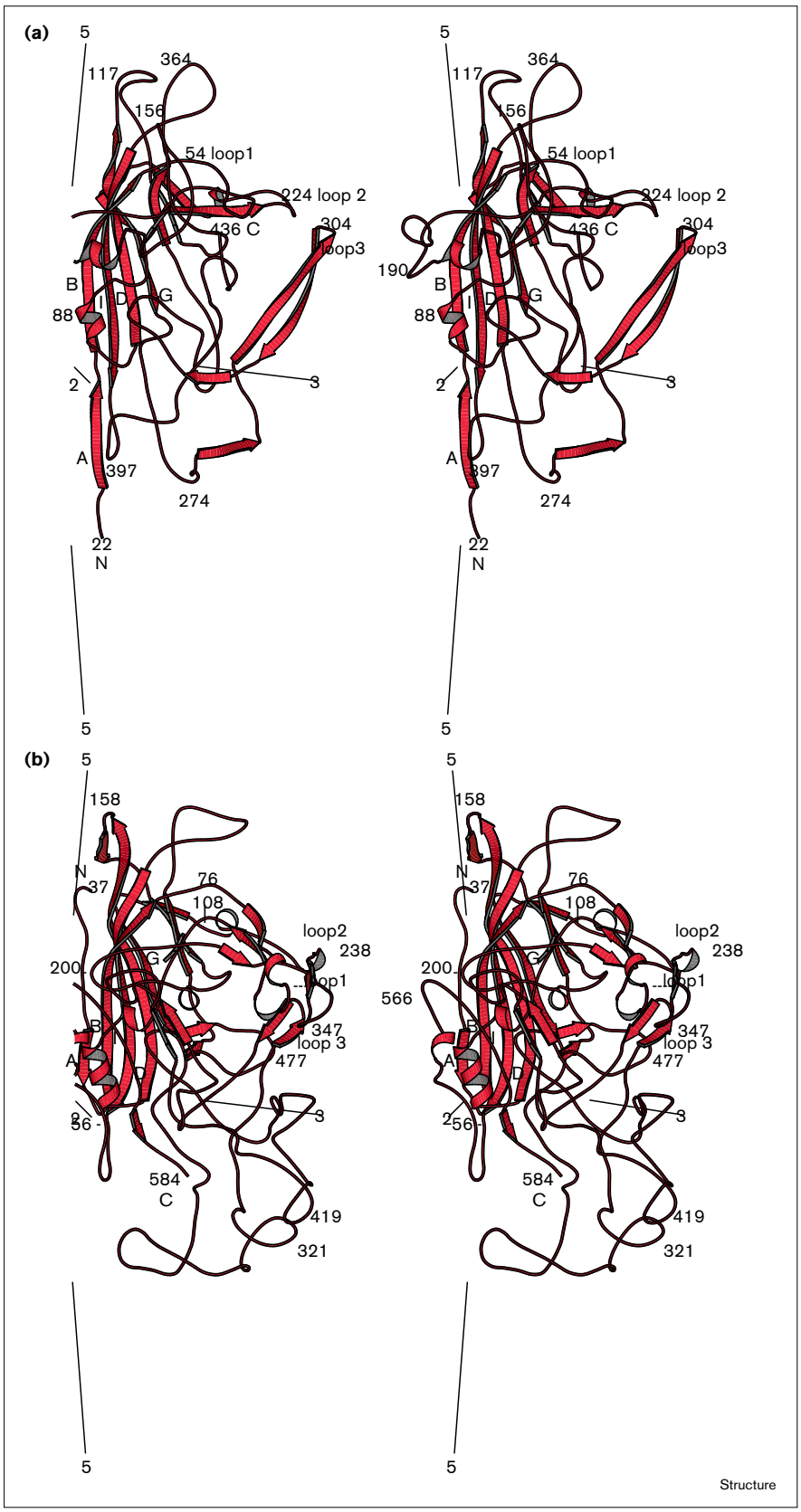
Parvovirus electron densities. **(a)** A comparison of a surface-rendered image of CPV (left), produced from model-derived structure factors, at 25 Å resolution with a cryo-electron microscopy reconstruction of *Gm*DNV (right) at the same resolution. Note the absence of the spike on the threefold axes in *Gm*DNV. **(b)** An example of the quality of the 3.7 Å electron-density map showing the region between residues 40 and 47. The map was contoured at a height of 2σ above zero density. Figures were drawn with the programs ROBEM (a) (R Ashmore and TS Baker, unpublished results) and O (b) [41].

forms the base of a spike at the threefold vertices (Figure 1a), whereas loop 4 forms the tip of the spike. In *Gm*DNV (Figure 1a), loop 4 is entirely missing, whereas loop 3 is much shorter and forms a β-annulus-type structure (Figure 3b) around the threefold axis, similar to that observed in tomato bushy stunt virus [18] and southern bean mosaic virus [19]. The center of the annulus forms an ~10 Å-diameter hole lined by residues Met289, Arg290, Gly291, Ala292, Thr312, Arg313 and Asn314, with Arg290 pointing towards the annulus center. These features form a depression around the threefold axis, where there is a protruding spike in CPV. The sidechains of residues that contribute to the β annulus and loop 3 are poorly defined in the electron-density map and the sequence is especially variable among densoviruses in this region (Figure 4). These observations suggest that the β annulus is a flexible part of the structure and might provide a portal for DNA or ligand exchange.

Evolution of parvoviruses

An alignment (Figure 4) was made between the VP1 capsid proteins of *Gm*DNV, *Junonia coenia* (*Jc*DNV) and *Pseudoplusia includens* (*Pi*DNV) densoviruses. These viruses all infect larval tissue other than that of the mid-gut. The

Figure 2



Comparison of the folds of **(a)** *Gm*DNV and **(b)** CPV. The conformation of the two β sheets, consisting of the β strands BIDG (labeled) and CHEF (not labeled), are well conserved between the two viruses, but the connecting loops are very different. In particular, βA adopts a domain-swapped conformation in *Gm*DNV, and the loops involved in threefold interactions are much extended in CPV. In *Gm*DNV, the most important threefold contact is made by residues 270–319 in loop 3, whereas in CPV loops 3 and 4, consisting of residues 279–336 and 409–449, respectively, make extensive additional threefold contacts within the threefold spike. This figure was drawn with the program MOLSCRIPT [43].

Table 1

The β-barrel structures similar to that of the *Gm*DNV capsid protein*†.

Chain	Z‡	rms (Å)§	Len¶	Lenseq#	Fragments¥	Id**	Name
	9.7	4.7	258	548	27	2cas	Canine parvovirus
B	4.1	3.5	138	199	15	4sbv	Southern bean mosaic virus
A	4.1	4.5	143	308	15	2bbv	Black beetle virus
A	3.8	4.3	136	283	14	2tbv	Tomato bushy stunt virus
2	3.8	3.7	150	266	24	1pvc	Poliovirus
2	3.4	3.6	138	210	17	1bbt	Foot-and-mouth disease virus
1	3.3	6.2	148	426	18	2bpa	Bacteriophage ϕX174
A	3.3	3.7	111	141	15	1stm	Satellite panicum mosaic virus
1	3.0	6.4	137	235	19	1pov	Poliovirus
3	3.0	7.9	135	220	19	1bbt	Foot-and-mouth disease virus
1	2.8	7.9	147	283	23	1r1a	Rhinovirus
3	2.8	5.8	139	235	23	1pvc	Poliovirus
2	2.5	5.6	119	374	15	1bmv	Bean pod mottle virus (middle component)
1	2.5	4.2	115	185	14	1bmv	Bean pod mottle virus (middle component)
1	2.4	7.9	132	268	21	2mev	Mengo encephalomyocarditis virus
	2.3	3.9	114	184	17	2stv	Satellite tobacco necrosis virus
A	2.2	2.2	103	149	15	1cwp	Cowpea chlorotic mottle virus

*The search was made with the program DALI [17]. †Structures are listed in decreasing order of significance of similarity. ‡Z is the significance of the similarity to *Gm*DNV expressed in standard deviations. §rms, the root mean square deviation of the structurally equivalenced regions in Å. ¶Len, the total number of residues in the equivalenced region. #Lenseq, total number of residues, including unequivalenced regions. ¥Fragments, number of stretches of equivalenced structure. **Id, the PDB accession number.

sequence of the CPV capsid protein was aligned with that of the densoviruses on the basis of its structural equivalence with *Gm*DNV and of the conserved sequence motifs in the N-terminal region, where no structural data are available. The sequences of several representative vertebrate parvoviruses were also aligned with that of CPV.

The VP1 residues 1–134 and 277–375 (the N termini of VP2 and VP4, respectively) of *Gm*DNV have diverged from those of the other insect viruses (Table 2). The intervening, more conserved region (135–276) contains the 39-residue motif that is conserved in most vertebrate parvoviruses, but which is absent in the distantly related adeno-associated and mink viruses. The sequence of VP4, starting with the conserved polyglycine motif and constituting all of the ordered part of the capsid protein, is as conserved as are the residues 135–276 of VP1.

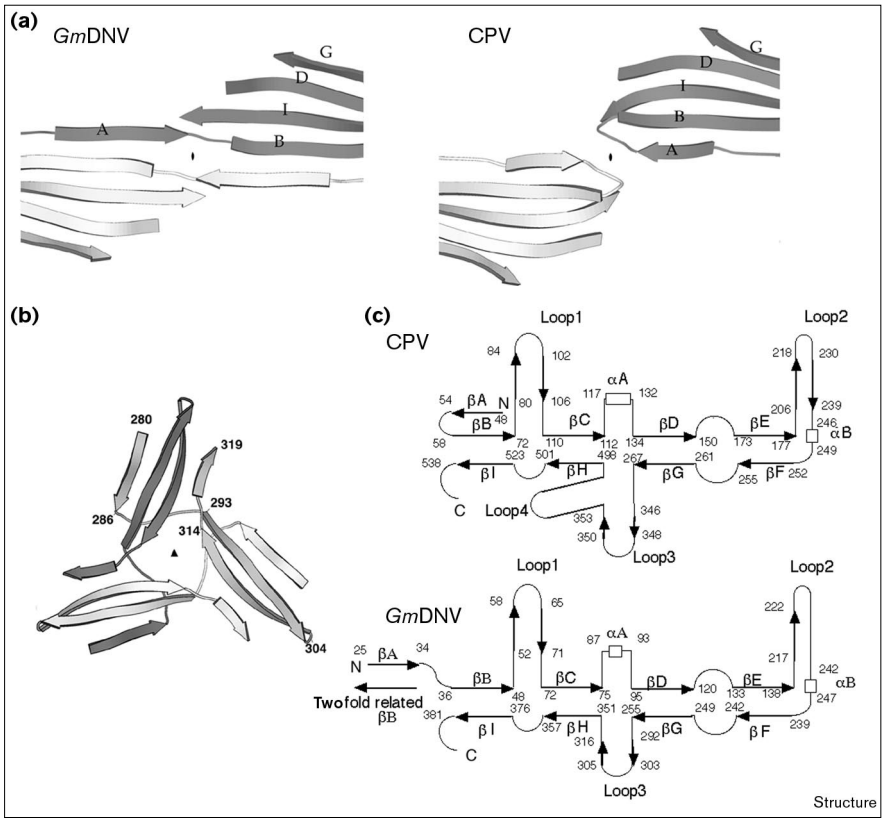
The capsid protein of *Gm*DNV as a whole is more closely related to that of *Jc*DNV than it is to that of *Pi*DNV. However, different parts of the protein show considerable variation in the level of sequence conservation. In particular, residues 277–375 of VP1 in *Gm*DNV have diverged substantially from those residues in *Pi*DNV and *Jc*DNV. In contrast, in the highly variable loop 3, the divergence of *Pi*DNV is greater than elsewhere in the sequence. The variable pattern of sequence conservation suggests that various insect parvoviruses may have recombined while replicating simultaneously in the same host. The gene transfer possibility is given support in that the regions with differing sequence relationships are bounded by the N termini of VP2 and VP4 and are at the ends of loop 3, defining the β annulus. Acceptance of the chimera may be the result of functional advantages within specific hosts.

The sequence alignment of the conserved, 39-residue motif and also of the structurally equivalenced regions of the capsid protein were each used independently to calculate the most parsimonious evolutionary tree (Figure 5). The two resultant trees have similar topologies, although the tree based on the structurally equivalenced capsid region shows a much larger divergence between the insect and vertebrate groups. The tree derived solely from consideration of the 39-residue motif would imply that the insect parvoviruses are as closely related to the vertebrate viruses as the vertebrate viruses are to each other.

Domain swapping

One side of the conserved β barrel is composed of the antiparallel βB, βI, βD and βG strands, which form the 'BIDG' sheet. In vertebrate parvoviruses, the amino end of βA is close to a fivefold axis; its carboxy end turns into βB near a twofold axis and forms an antiparallel ABIDG sheet. This topology is similar to that of the C subunit in icosahedral plant viruses with a triangulation number of 3 and in VP2 of the picornaviruses [20]. In *Gm*DNV, however, βB is essentially a linear extension of βA, the latter hydrogen bonding with a neighboring twofold-related subunit (Figures 3a and 6a). An example of domain swapping [21], in which one domain in a protein is substituted by the same domain from an identical, symmetry-related protein, is the relationship of the βA strand in *Gm*DNV (the 'swapped' structure) to that in CPV (the 'unswapped'

Figure 3



Secondary structural elements. **(a)** Domain swapping of βA between *Gm*DNV and CPV. **(b)** The β annulus about the threefold axis. **(c)** Diagrammatic representation of the polypeptide folds of *Gm*DNV and CPV. Numbering corresponds to the beginnings and ends of secondary structural elements. Figures (a) and (b) were drawn with the program MOLSCRIPT [43].

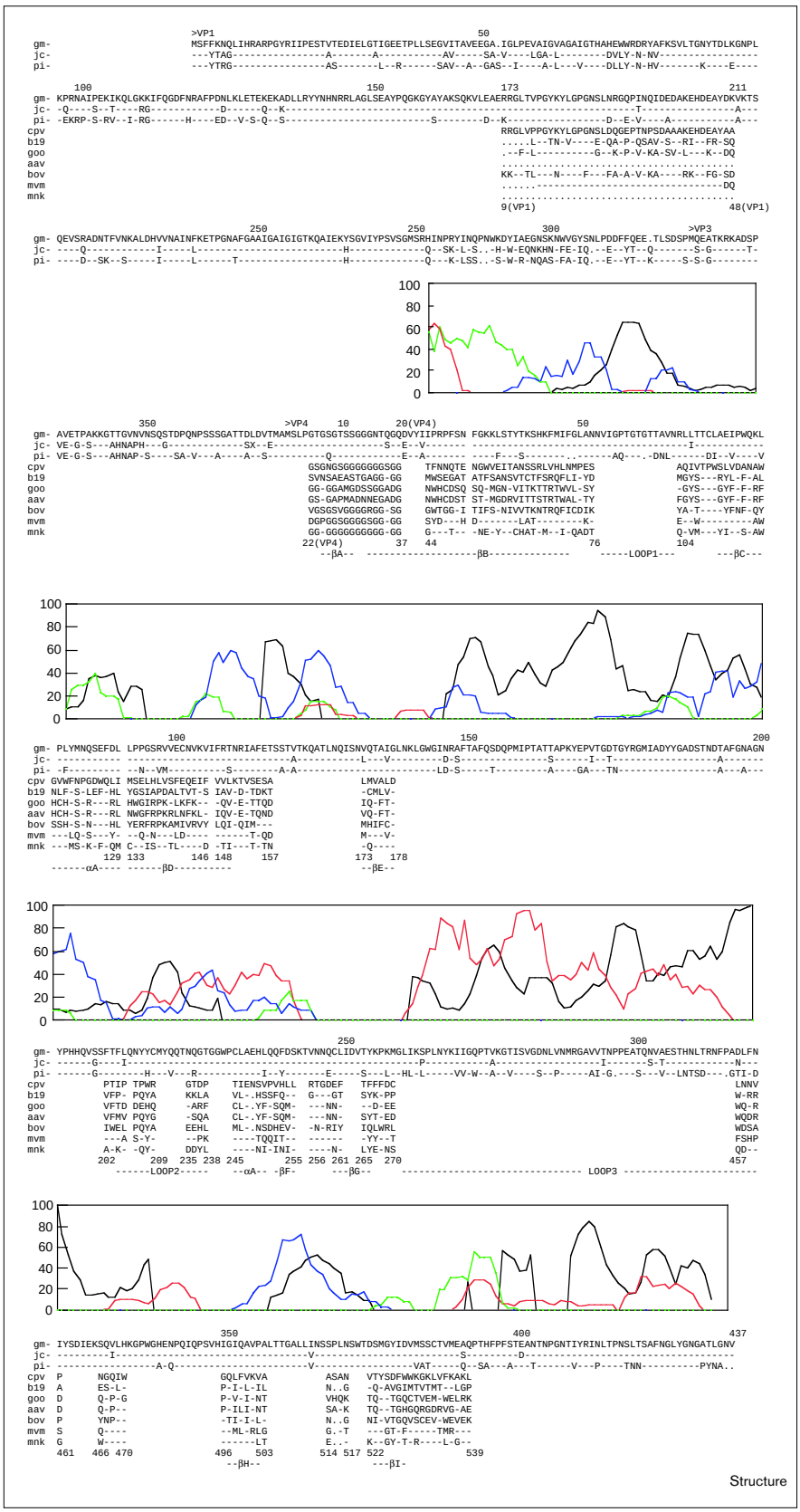
structure). Domain swapping can promote oligomerization of subunits "by modification of an existing interface, rather than evolution of a new one" [22]. The CPV structure has five additional residues relative to *Gm*DNV between βA and βB. For CPV to adopt domain swapping, two of these residues would need to be deleted or the chain would need to distort into a bulge. In contrast, the *Gm*DNV structure has the potential to form an unswapped conformation without the need for insertions.

The fivefold-axis channel

*Gm*DNV, like the vertebrate parvoviruses CPV, FPV and MVM, has a channel along each fivefold axis of the capsid. In the vertebrate parvoviruses, the channel shows continuous electron density along its length. This density has been interpreted as belonging to the conserved, glycine-rich motif near the N terminus of VP2 [8,23]. In contrast, the electron density in the channel of *Gm*DNV is weak and discontinuous. In CPV, the channel is ~45 Å long and is lined mainly with small amino acids. In *Gm*DNV, the channel is only ~20 Å long (Figure 7) and is lined by large, mostly hydrophobic residues. In both CPV and *Gm*DNV, the channel is ~9 Å in diameter. The glycine-rich sequence in CPV consists of 16 residues, 12 of which are glycines, whereas in *Gm*DNV the corresponding sequence contains 11 residues, of which 6 are glycines; this difference is in proportion to the lengths of the respective fivefold channels. If the polypeptide backbone of the *Gm*DNV capsid protein were to extend from the first visible residue (Val22) backwards towards the fivefold axis, three consecutive glycine residues would lie in the most constricted region of the fivefold channel. The well-conserved glycine-rich motif in vertebrate and insect parvoviruses probably serves a common function, which is presumably to facilitate the externalization of some of the N termini of the subunits. The absence of continuous electron density in the *Gm*DNV channel suggests that the glycine-rich peptide might occupy this region only transiently (e.g. during assembly). Alternatively, only a small fraction of the N termini, for instance only the VP1 polypeptides, might pass through the fivefold channel.

It would be extremely difficult for a long and bulky polypeptide (374 amino acids for the *Gm*DNV VP1) to be threaded through this channel after the virus has assembled. It is more likely that the glycine-rich region is located along the fivefold axis prior to association of the surrounding five subunits; this could happen easily in CPV, where the glycine-rich sequence belongs to a subunit

Figure 4



The sequences of the insect and vertebrate parvoviruses aligned on the basis of a superposition of the structures of *Gm*DNV and CPV. The other insect and vertebrate virus sequences are aligned with *Gm*DNV and CPV, respectively, based on sequence. *Gm*DNV is numbered according to the sequence of VP1 in the N-terminal, unstructured part of the protein, but VP4 sequence numbers are used for the common C-terminal region. The initiation of VP4 translation can occur at either of the two methionines in the sequence MAMSLP... (PT, unpublished results). The second of the two methionines has been used as a starting point for numbering VP4. Residues 22–436 of VP4 are ordered in the structure. Given that βA adopts a domain-swapped conformation in *Gm*DNV, it has been structurally aligned with residues in CPV that belong to a twofold-related subunit. The surface area (Å^2 per residue) buried by the twofold, threefold and fivefold interactions are shown in black, red and green, respectively. The surface area exposed on the outside of the capsid is plotted in blue. These curves have been smoothed over a window of five residues. The 39-residue conserved motif (residues 173–211 of VP1) and the glycine-rich region (residues 5–20 of VP4) have been aligned between the insect and vertebrate viruses based only on sequence. The recently determined sequences of *Diatraea saccharalis* (Y Boublik and M Bergoin, personal communication) and *Mythimna loreyi* (G Fediere and PT, unpublished results) densoviruses confirm the trends shown here. Abbreviations: gm, *Galleria mellonella* densovirus; jc, *Junonia coenia* densovirus; pi, *Pseudoplusia includens* densovirus; cpv, canine parvovirus; b19, human B19 parvovirus; goo, goose parvovirus; aav, adeno-associated virus; bov, bovine parvovirus; mvm, minute virus of mouse; mnk, mink parvovirus. Sequence alignments were generated using the DARWIN system [44].

Table 2

Sequence identity between the capsid proteins of insect densovirus* (%).

*Gm*DNV segment numbers	*Gm vs Jc*	*Gm vs Pi*	*Jc vs Pi*
1 –134 (VP1)	78	69	80
135 –276 (VP1)	95	88	91
173 –211 (VP1)†	87	87	97
277 –375 (VP1)‡	57	56	77
All VP4	95	83	79
270 –319 (loop 3; VP4)	90	56	55

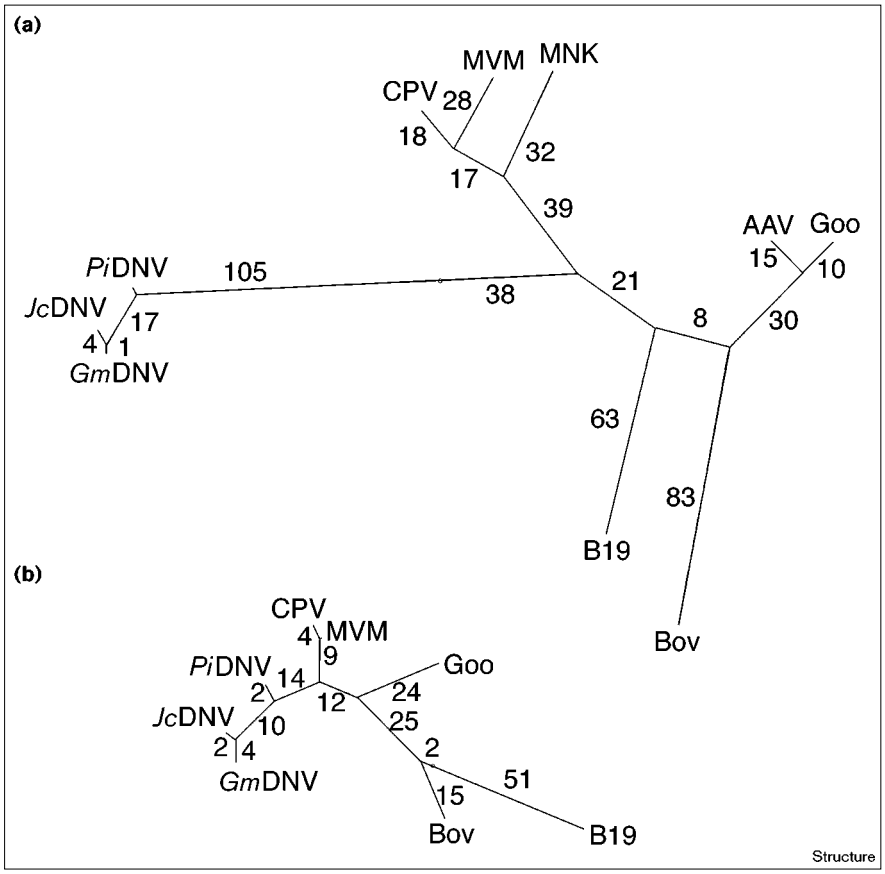
**Gm* is *Galleria mellonella* DNV; *Jc* is *Junonia coenia* DNV; *Pi* is *Pseudoplusia includens* DNV. †The conserved 39-residue motif present in most parvoviruses. ‡This sequence is from the N terminus of VP2 to the N terminus of VP4.

that contributes to the formation of a fivefold channel into which the glycine-rich sequence is placed. In *Gm*DNV, the N termini adopt a swapped conformation that would require them to be placed into a twofold-related fivefold channel accomplished by the prior formation of dimers to position the chain correctly in the channel. Another possibility is that the small number of N termini passing through the fivefold channels adopt the unswapped conformation, as seen in CPV. Domain swapping may, therefore, be a mechanism for selecting only one or a few VP1s to be assembled into the virion.

Subunit interactions

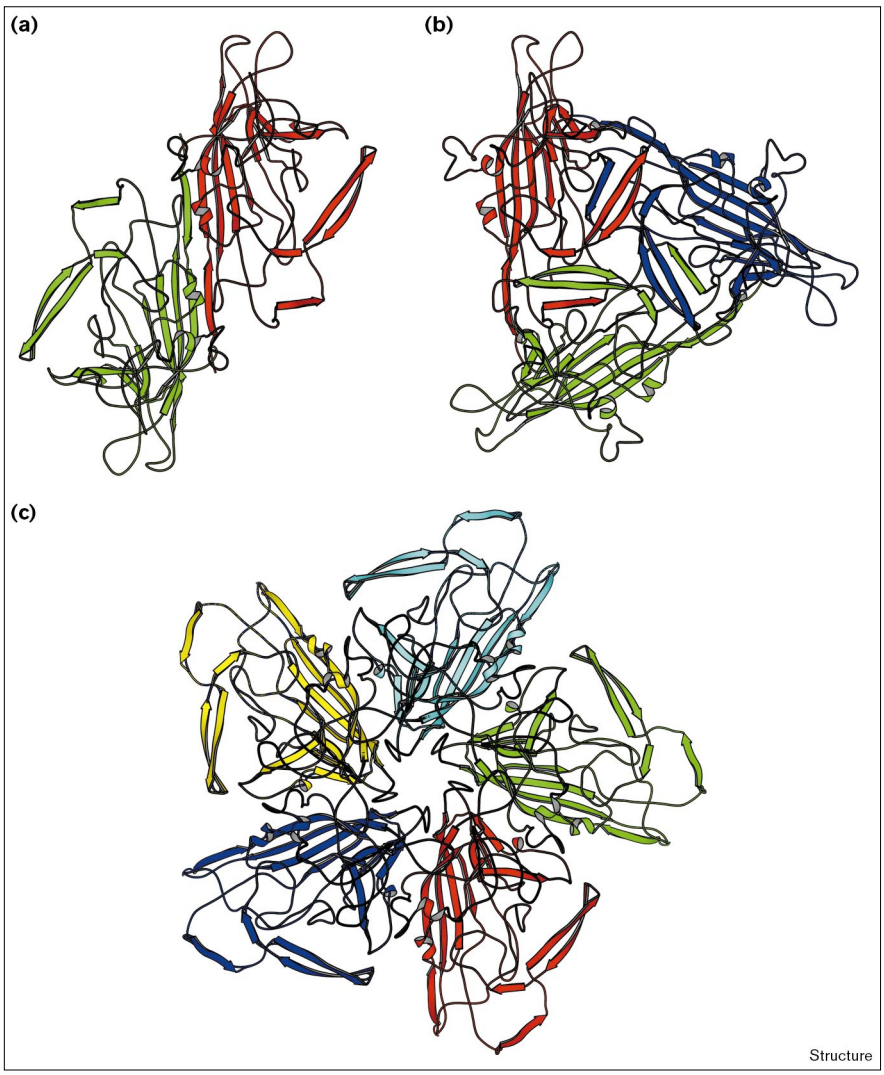
The most prominent surface feature of the vertebrate parvoviruses is the protruding spike, consisting of loops 3 and 4 interleaved with their threefold, symmetry-related equivalents, at each of the threefold axes (Figure 6). These spikes are absent from *Gm*DNV because there are 133 fewer residues in these loops in GmDNV relative to CPV. In human B19 parvovirus, the threefold spikes form the binding site for the receptor on the host cell [14]. In the assembled capsid of *Gm*DNV, 45% of the solvent-accessible surface area of the monomer (25,381 Å^2) is buried. This is distributed between the contacts created by the fivefold, threefold and twofold interactions, which bury 15, 20 and 10% of the monomer's surface, respectively. The N-terminal arm accounts for most of the contacts generated by the twofold axes (Figure 6). The contacts generated by the threefold axes mainly involve loop 3 and its flanking sequences (residues 270–319), which form a

Figure 5



Evolutionary trees for the parvovirus capsid proteins, derived from **(a)** the ordered part of the capsid protein and **(b)** the highly conserved, 39-residue motif in the unique N-terminal domain of VP1. The number of point accepted mutations (PAM) per 100 residues is shown between nodes. Phylogenetic trees were generated using the DARWIN system [44].

Figure 6



Association of subunits in the *Gm*DNV virion. The orientation of the red subunit is approximately the same in each view and also as in Figure 2. **(a)** The main contact between twofold-related subunits is the βA N-terminal arm. **(b)** Loop 3 of neighboring subunits interlocks around the β annulus at the threefold axis. **(c)** The fivefold-related contacts are less extensive than the twofold or threefold contacts. The channel along the fivefold axis is discussed in the text. This figure was drawn with the program MOLSCRIPT [43].

four-stranded, intersubunit β sheet around the β annulus (Figures 3b and 6).

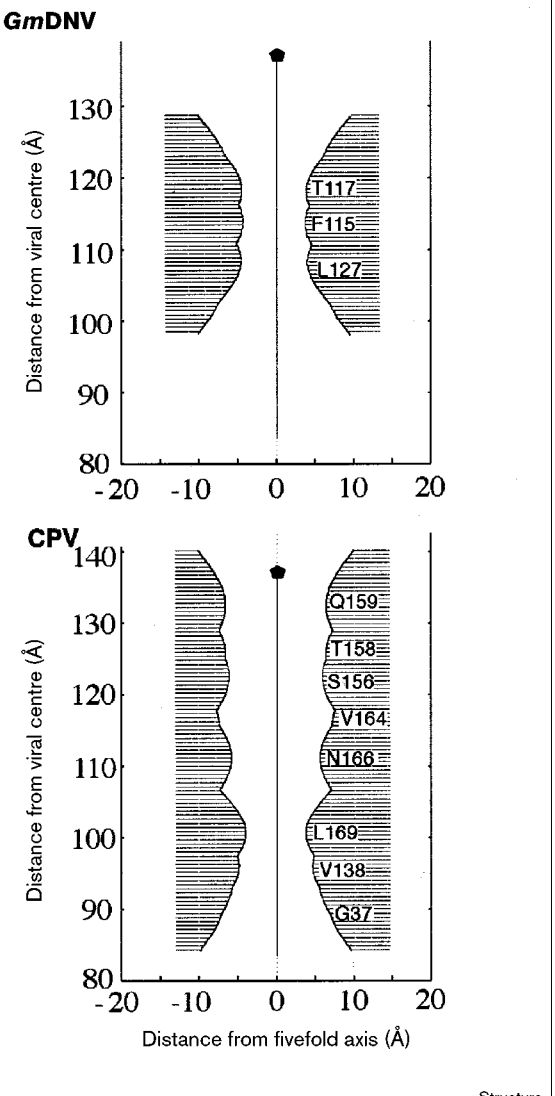
Putative divalent-cation-binding site

As divalent cations are known to stabilize the virus, a search was conducted for clusters of three or more negatively charged sidechains that could serve as possible cation-binding sites. The *Gm*DNV structure contains one such site, which is similar to that in other structures that coordinate Ca^{2+} or Mg^{2+} [24,25]. Residues Glu116, Asp156, Gln157 and Asp375 surround a central position that is located 2.3 Å from the coordinating sidechains. The electron density for the sidechains is weak, however, and there is no significant density in the central position, implying that the site is at best poorly occupied. Glu116 is part of the loop that makes the wall of the fivefold channel, but its sidechain faces away from the channel, placing the putative cation-binding site about 15 Å from the fivefold axis. Nevertheless, conformational changes at the cation-binding site might alter the properties of the channel along the fivefold axis. Hence, the presence of the glycine-rich polypeptide in the channel would be likely to cause conformational changes in the residues coordinating the putative divalent-cation-binding site.

DNA structure

The electron density in the solvent regions of the crystal does not exceed 0.3σ from zero. However, there are features at a level of 1.2σ, close to the internal surface of the capsid, which may represent partially icosahedrally ordered protein or DNA. Near the threefold axis, there is a gap of approximately 4 Å between the protein and the uninterpreted density. This contrasts with the well-ordered single-stranded DNA seen in CPV (11 bases)

Figure 7



The dimensions of the fivefold channel in *Gm*DNV and CPV. The width of the channel at its narrowest point is similar in the two structures, but the channel is much longer in CPV.

[26] and MVM (21 bases) [12]. Unlike CPV, *Gm*DNV exhibits no specific contacts between the DNA and the capsid protein. In plant viruses, the presence of polyamines correlates with the absence of positively charged N termini of the capsid protein subunits [27]. The absence of well-ordered DNA in DNVs, in contrast to the mammalian parvoviruses, could reflect the significant amount of polyamines that they often contain; these neutralize the nucleic-acid charge [28]. Alternatively, the DNA in the single frozen crystal of *Gm*DNV may have been damaged during the long data collection. In contrast,

Table 3

X-ray diffraction data.

	Frames 2–128		Frames 2–66	
Resolution limits (Å)	Completeness (%)	R_{sym}*	Completeness (%)	R_{sym}*
100.00–8.63	67.4	0.064	44.6	0.044
8.63–6.85	69.9	0.104	47.5	0.075
6.85–5.98	69.4	0.149	47.6	0.109
5.98–5.44	69.1	0.160	47.6	0.111
5.44–5.05	68.0	0.168	47.0	0.117
5.05–4.75	67.2	0.165	45.4	0.111
4.75–4.51	64.3	0.174	42.8	0.115
4.51–4.32	61.6	0.195	40.1	0.132
4.32–4.15	57.5	0.237	33.9	0.156
4.15–4.01	52.7	0.281	30.2	0.179
4.01–3.88	49.5	0.316	27.8	0.189
3.88–3.77	46.6	0.395	25.4	0.249
3.77–3.67	43.1	0.411	20.3	0.254
3.67–3.58	38.8	0.571	17.5	0.466
3.58–3.50	36.6	0.656	15.2	0.582
All data 100–3.50 Å	57.6	0.182	35.6	0.110
No. of measurements		405,646		164,912
No. of unique reflections		174,343		107,965

*$R_{sym} = \Sigma_h \Sigma_i |(I_{hi} - I_h)| / \Sigma_h \Sigma_i |I_h|$,

where I_{hi} is the measured integrated intensity for reflection (h) and I_h is the mean for all the i observations of a reflection. The integrated intensities of partial reflections on consecutive images were added to give the intensity of the complete reflection. No reflections were rejected. All images were collected over a 0.2° oscillation range. The crystal-to-detector distance was set to 460 mm.

in the structural determinations of vertebrate parvoviruses, numerous crystals were exposed to X-rays for shorter times at room temperature. The possibility of radiation damage is supported by the improved visualization of the single-stranded DNA when using only the less damaged first half of the data (Table 3). Presumably, the nucleic acid is more susceptible to radiation damage than is the protein.

Biological implications

The structure of the insect parvovirus *Galleria mellonella* densovirus (*Gm*DNV) has some structural features in common with vertebrate parvoviruses, such as canine parvovirus (CPV), including the core β-barrel fold that is also found in many viral capsid proteins. Nevertheless, there is no detectable sequence similarity in the structurally defined part of the capsid protein. The large loops connecting the strands of the β barrel differ mostly in their conformations. When compared to CPV, the β barrel of *Gm*DNV is shifted by 9.7 Å outwards and its orientation is changed by 7.4° with respect to the icosahedral symmetry axes. In contrast to the single-stranded DNA of CPV and of the minute virus of mouse (MVM), the DNA of *Gm*DNV is not icosahedrally well ordered, possibly because it is not in close association with the capsid protein or because it

might have suffered radiation damage in the single, frozen crystal used for the data collection.

In *Gm*DNV, the βB strand is a linear extension of the N-terminal βA strand, allowing βA to hydrogen bond with βB of the neighboring, twofold-related subunit and thereby increasing the number of interactions between subunits. In CPV, however, βA folds back and hydrogen bonds with the βB strand from its own subunit. Comparison of these two structures provides an example of 'domain swapping'. The 'swapped' structure in *Gm*DNV may regulate the inclusion of only one or two VP1s in the capsid and thereby allow externalization of the unique, functionally important, N-terminal domain of VP1.

The surface of *Gm*DNV lacks the 20 Å high, antigenic, threefold spike that is present in the mammalian parvoviruses; this spike is comprised of intertwined loops from threefold-related subunits. In *Gm*DNV, there is an antiparallel β-annulus structure around a hole at the threefold axis. The receptor-recognition function of this spike in mammalian viruses is apparently not required in insect viruses.

A channel along the fivefold axes exists in both *Gm*DNV and mammalian parvoviruses. Whereas in the mammalian parvoviruses a highly conserved, glycine-rich motif is located in the channel, the channels appear to be mostly empty in the insect virus. The glycine-rich motif allows the externalization of the N termini of up to one fifth of all the capsid proteins in the mammalian viruses. As this motif is also conserved in insect parvoviruses, it is probable that its function is conserved. There is no continuous electron density within the fivefold axis channels of *Gm*DNV, however, suggesting that, at most, a few N termini are located in the channels. Perhaps the glycine-rich sequence in the insect parvovirus is required to allow the one or two copies of VP1 to externalize its N-terminal region, which contains a highly conserved 39-residue motif.

Removal of divalent cations from *Gm*DNV results in a change in the density of the capsid. This may be mediated by a putative calcium- or magnesium-binding site near the fivefold axis. Externalization of the N-terminal region through the fivefold axis is likely to alter the ability to bind a cation.

The common β-barrel domain, the simple icosahedral structure and the localized amino acid similarities strongly suggest that insect and mammalian parvoviruses have had a common origin. However, extensive structural differences outside the β-barrel motif and differences in the transcription processes indicate a lengthy divergent evolutionary development of these parvoviruses. As insects lack an adaptive immune response, it is possible that the absence of an evolutionary pressure may account for the large structural differences in the surface of the capsid.

Materials and methods

Preparation of virus and crystals

Larvae of the wax moth, *Galleria mellonella*, were grown and infected with densovirus by feeding with dead, infected larvae. After several days, the virus was extracted [29] by blending the dead larvae in a mixture of two-thirds phosphate-buffered saline (PBS) at pH 7.5 and one-third CCl_4. Differential centrifugation followed by CsCl equilibrium density gradient centrifugations produced two bands with densities of 1.40 g/cm^3 and 1.44 g/cm^3 that contained infectious virions and a third of density 1.32 g/cm^3 that contained empty capsids. The 1.40 g/cm^3 band was transferred into 10mM tris-buffered saline (TBS), containing 0.1 M NaCl at pH 7.5 and 1.0 mM Ca^{2+} and Mg^{2+}. Virus stability was enhanced in the presence of these ions. Crystals of up to 0.3 mm in length were grown at 4°C in TBS with approximately 5% PEG8000 as a precipitant. The crystals were flash-frozen for data collection by placing them in a nylon loop and immersing them in a stream of N_2 at approximately 100K. Crystals were soaked for at least 10 min in the presence of 25% glycerol as a cryoprotectant prior to freezing, although it was also possible to grow the crystals in the presence of glycerol.

Cryo-electron microscopy and three-dimensional image reconstruction

The cryo-electron microscopic data of *Gm*DNV were recorded as previously described for B19 parvovirus [14]. An initial attempt to use B19 human parvovirus as a model [30] failed. Subsequently, the orientation of the particles was determined using the common-lines technique [31]. A total of 38 particle images were then combined to compute a three-dimensional reconstruction that was used for further model-based refinement of the view angle and origin parameters [30]. The final reconstructed density map was calculated to 22 Å resolution. The threshold level for rendering the isosurface view (Figure 1a) was selected to give a molecular volume consistent with the expected volume of the GmDNV capsid. The view of CPV in Figure 1a was computed to 25 Å resolution using atomic coordinates (PDB accession number 2cas).

Structure determination

A dataset was collected to 3.5 Å resolution (Table 3) from a single frozen crystal at the Cornell High Energy Synchrotron Source (CHESS) on the F1 beam line ($\lambda = 0.9180$ Å), using Fuji imaging plates and a BAS 2000 scanner. The initial structure determination was performed with data from the first 127 frames. Subsequently, only data from the first 65 images were used because the crystal was noticeably radiation-damaged during the second half of data collection. Although this dataset was less complete, its quality was superior. The data were processed and scaled using the DENZO and SCALEPACK [32] programs, respectively. The space group was found to be either $P4_32_12$ or $P4_12_12$ with $a = 264.44$ and $c = 683.45$ Å. The Matthews coefficient, V_M [33], was 2.2 Å^3/Da, assuming four particles in the unit cell. This requires the particle to be situated on a crystallographic twofold axis with half a particle per crystallographic asymmetric unit. Packing considerations eliminated all other primitive space groups based on a 422 point group and showed that a particle must be positioned close to (0,0,1/4), assuming the origin as defined in the International Tables [34].

The self-rotation function [35,36] confirmed the icosahedral symmetry of the virus and its location on a crystallographic twofold axis along the **ab** face diagonal. The icosahedral symmetry creates 30-fold noncrystallographic redundancy. A particle in the standard orientation [37] requires a rotation of 45° about the *z*-axis, followed by a second rotation of 54.7° about the $[\bar{1},1,0]$ axis to have the orientation as found in

the crystal unit cell. The final relationship found between a position ($X\,Y\,Z$) in the standard particle and the point ($X'Y'Z'$) in a chosen molecule of the crystal was

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} =$$

$$\begin{pmatrix} (1-\cos\kappa)u^2+\cos\kappa & (1-\cos\kappa)vu-w\sin\kappa & (1-\cos\kappa)wu+v\sin\kappa \\ (1-\cos\kappa)uv+w\sin\kappa & (1-\cos\kappa)v^2+\cos\kappa & (1-\cos\kappa)wv-u\sin\kappa \\ (1-\cos\kappa)uw-v\sin\kappa & (1-\cos\kappa)vw+u\sin\kappa & (1-\cos\kappa)w^2+\cos\kappa \end{pmatrix} \times$$

$$\begin{pmatrix} 1/\sqrt{2} & -(1/\sqrt{2}) & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} 3.31 \\ -3.31 \\ c/4 \end{pmatrix}$$

$$= \begin{pmatrix} 0.40870 & -0.70711 & 0.57703 \\ 0.40870 & 0.70711 & 0.57703 \\ -0.81604 & 0.00000 & 0.57800 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} 3.31 \\ -3.31 \\ 170.86 \end{pmatrix}$$

where the direction cosines of the $[\bar{1},1,0]$ axis are

$$u = -\frac{1}{\sqrt{2}},\ v = \frac{1}{\sqrt{2}}, w = 0 \text{ and } \kappa = 54.7°$$

and where ($X\,Y\,Z$) and ($X'Y'Z'$) are measured in an orthogonal coordinate system in Å.

To determine the position of the particle along the $[\bar{1},1,0]$ direction in the unit cell, the electron density of a search structure was placed systematically at positions along $[\bar{1},1,0]$ close to (0,0,1/4) and on space-group-equivalent positions for both of the two possible space groups. The resultant electron density was then Fourier back-transformed to give structure-factor amplitudes for comparison with the observed data between 45 and 25 Å resolution. Three different search models were used: a hollow sphere with internal and external radii optimized to 78 Å and 128 Å [38]; an atomic model of CPV (PDB accession code 2cas); and the cryo-electron microscopic reconstruction of *Gm*DNV at 25 Å resolution (Figure 1a). The hollow-shell model was unable to differentiate between the two space groups because of its spherical symmetry. The highest correlations with observed amplitudes were obtained with the cryo-electron microscopic model, presumably because the reconstructed density more closely matched the actual *Gm*DNV structure.

Cycles of electron-density averaging [39] were computed using 45 to 25 Å resolution data in both possible space groups, the particle orientation derived from the self-rotation function and the positions determined from the search using the spherical-shell model. Averaging was performed within a shell defined by an outer radius of 135 Å and inner radius of 70 Å. The density was set to its average value in the external (solvent) and the internal (nucleic-acid) regions. Unit weights were applied to all structure amplitudes in the Fourier synthesis. There was no attempt to replace unobserved reflections with calculated values. The overall correlation coefficient increased from 0.50 to 0.51 over 17 cycles of averaging for space group $P4_32_12$, but increased from 0.54 to 0.69 for space group $P4_12_12$. The phases were then extended from 25 to 3.7 Å resolution in space group $P4_12_12$ using steps no larger than 1/**a** with five cycles of averaging at each step. Calculated structure factors were included for unobserved reflections after 5 Å resolution had been reached. The program ENVELOPE [39] was used to refine the position and the orientation of the particle several times during phase extension by systematically changing the orientation and position of the noncrystallographic symmetry (NCS) operators to maximize the correlation between NCS-related electron densities [40]. The position refined from $\mathbf{x} = 0.0080$ to 0.0125, which corresponds to a movement of 1.68 Å along the $[\bar{1},1,0]$ direction, but the orientation did not change. After the phases had been extended to 4.7 Å resolution, the later radiation-damaged frames were omitted from the dataset (Table 3). A further 316 cycles of density averaging were performed in order to achieve complete convergence.

Residues 22–464 were built into the electron-density map using the program O [41]. The atom positions in the icosahedral asymmetric unit were refined using the program X-PLOR [42] while applying strict, non-crystallographic symmetry. A final *R* factor of 27.1% for all measured reflections using frames 2–66 (Table 3) was obtained after restrained positional and restrained individual temperature factor refinement, using data from 8–3.6 Å. Many of the external sidechains appeared disordered. The deviations from ideality for the bond lengths and angles were 0.02 Å and 2.9°, respectively.

Accession numbers

The coordinates and structure factors have been deposited with the PDB (accession numbers 1dnv and 1dnvsf, respectively).

Acknowledgements

We are grateful for many helpful discussions with Terje Dokland, and for help in the preparation of the manuscript from Cheryl Towell and Sharon Wilder. The work was supported by National Institutes of Health and National Science Foundation grants to TSB, a National Science Foundation grant to MGR, a National Institutes of Health Program Project grant, a Lucille P Markey award in support of structural biology at Purdue University and a Purdue University reinvestment grant.

References

1. Tijssen, P. & Bergoin, M. (1995). Densonucleosis viruses constitute an increasingly diversified subfamily among the parvoviruses. *Semin. Virol.* **6**, 347-355.
2. Mengeling, W.L. & Cutlip, R.C. (1975). Pathogenesis of *in utero* infection: experimental infection of five-week old porcine fetuses with porcine parvovirus. *Am. J. Vet. Res.* **36**, 1173-1177.
3. Anderson, L.J. & Török, T.J. (1989). Human parvovirus B19. *New Engl. J. Med.* **321**, 536-538.
4. Tal, J. & Attathom, T. (1993). Insecticidal potential of the insect parvovirus GmDNV. *Arch. Insect Biochem. Physiol.* **22**, 345-356.
5. Belloncik, S. (1990). Potential use of densonucleosis viruses as biological control agents of insect pests. In *CRC Handbook of Parvoviruses.* (Tijssen, P., ed.), pp. 29-39, CRC Press Inc., Boca Raton.
6. Tijssen, P., van den Hurk, J. & Kurstak, E. (1976). Biochemical, biophysical and biological properties of densonucleosis virus (parvovirus). I. Structural proteins. *J. Virol.* **17**, 686-691.
7. Tijssen, P. & Kurstak, E. (1981). Biophysical and biological properties of densonucleosis virus (parvovirus). III. Common sequences of structural proteins. *J. Virol.* **37**, 17-23.
8. Tsao, J., *et al.*, & Parrish, C.R. (1991). The three-dimensional structure of canine parvovirus and its functional implications. *Science* **251**, 1456-1464.
9. Agbandje, M., McKenna, R., Rossmann, M.G., Strassheim, M.L. & Parrish, C.R. (1993). Structure determination of feline panleukopenia virus empty particles. *Proteins* **16**, 155-171.
10. Llamas-Saiz, A.L., Agbandje-McKenna, M., Wikoff, W.R., Bratton, J., Tattersall, P. & Rossmann, M.G. (1997). Structure determination of minute virus of mice. *Acta Cryst. D* **53**, 93-102.
11. Chapman, M.S. & Rossmann, M.G. (1993). Structure, sequence and function correlations among parvoviruses. *Virology* **194**, 491-508.
12. Agbandje-McKenna, M., Llamas-Saiz, A.L., Wang, F., Tattersall, P. & Rossmann, M.G. (1998). Functional implications of the structure of the murine parvovirus, minute virus of mice. *Structure* **6**, in press.
13. Brown, K.E., Anderson, S.M. & Young, N.S. (1993). Erythrocyte P antigen: cellular receptor for B19 parvovirus. *Science* **262**, 114-117.
14. Chipman, P.R., *et al.*, & Rossmann, M.G. (1996). Cryo-electron microscopy studies of empty capsids of human parvovirus B19 complexed with its cellular receptor. *Proc. Natl Acad. Sci. USA* **93**, 7502-7506.

15. Horiuchi, M., Ishiguro, N., Goto, H. & Shinagawa, M. (1992). Characterization of the stage(s) in the virus replication cycle at which the host-cell specificity of the feline parvovirus sub-group is regulated in canine cells. *Virology* **189**, 600-608.
16. Tijssen, P. & Kurstak, E. (1979). Studies on the structure of the two infectious types of densonucleosis virus. *Intervirology* **11**, 261-267.
17. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138.
18. Harrison, S.C., Olson, A.J., Schutt, C.E., Winkler, F.K. & Bricogne, G. (1978). Tomato bushy stunt virus at 2.9 Å resolution. *Nature* **276**, 368-373.
19. Rossmann, M.G., Abad-Zapatero, C., Hermodson, M.A. & Erickson, J.W. (1983). Subunit interactions in southern bean mosaic virus. *J. Mol. Biol.* **166**, 37-83.
20. Rossmann, M.G. & Johnson, J.E. (1989). Icosahedral RNA virus structure. *Annu. Rev. Biochem.* **58**, 533-573.
21. Bennett, M.J., Schlunegger, M.P. & Eisenberg, D. (1995). 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci.* **4**, 2455-2468.
22. Bax, B., *et al.*, & Slingsby, S. (1990). X-ray analysis of βB2-crystallin and evolution of oligomeric lens proteins. *Nature* **347**, 776-780.
23. Xie, Q. & Chapman, M.S. (1996). Canine parvovirus capsid structure, analyzed at 2.9 Å resolution. *J. Mol. Biol.* **264**, 497-520.
24. McPhalen, C.A., Strynadka, N.C.J. & James, M.N.G. (1991). Calcium-binding sites in proteins: a structural perspective. *Adv. Protein Chem.* **42**, 77-144.
25. Zhao, R., Hadfield, A.T., Kremer, M.J. & Rossmann, M.G. (1997). Cations in human rhinoviruses. *Virology* **227**, 13-23.
26. Chapman, M.S. & Rossmann, M.G. (1995). Single-stranded DNA-protein interactions in canine parvovirus. *Structure* **3**, 151-162.
27. Nickerson, K.W. & Lane, L.C. (1977). Polyamine content of several RNA plant viruses. *Virology* **81**, 455-459.
28. Kelly, D.C. & Elliott, R.M. (1977). Polyamines contained by two densonucleosis viruses. *J. Virol.* **21**, 408-410.
29. Bailey, L., Gibbs, A.J. & Woods, R.R. (1970). A simple way of purifying several insect viruses. *J. Gen. Virol.* **6**, 175-177.
30. Baker, T.S. & Cheng, R.H. (1996). A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy. *J. Struct. Biol.* **116**, 120-130.
31. Crowther, R.A. (1971). Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs. *Philos. Trans. R. Soc. Lond. B* **261**, 221-230.
32. Otwinowski, Z. (1993). Oscillation data reduction program. In *Proceedings of the CCP4 Study Weekend: Data Collection and Processing.* (Sawyer, L., Isaacs, N. & Bailey, S., eds), pp. 56-62, SERC Daresbury Laboratory, England.
33. Matthews, B.W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.
34. Hahn, T. (1983). *International Tables for Crystallography*. D. Reidel Publishing Company, Boston.
35. Rossmann, M.G. & Blow, D.M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst.* **15**, 24-31.
36. Tong, L. & Rossmann, M.G. (1997). Rotation function calculations with GLRF program. *Methods Enzymol.* **276**, 594-611.
37. Tsao, J., Chapman, M.S., Wu, H., Agbandje, M., Keller, W. & Rossmann, M.G. (1992). Structure determination of monoclinic canine parvovirus. *Acta Cryst. B* **48**, 75-88.
38. Tsao, J., Chapman, M.S. & Rossmann, M.G. (1992). *Ab initio* phase determination for viruses with high symmetry: a feasibility study. *Acta Cryst. A* **48**, 293-301.
39. Rossmann, M.G., *et al.*, & Lynch, R.E. (1992). Molecular replacement real-space averaging. *J. Appl. Cryst.* **25**, 166-180.
40. Muckelbauer, J.K., *et al.*, & Rossmann, M.G. (1995). Structure determination of coxsackievirus B3 to 3.5 Å resolution. *Acta Cryst. D* **51**, 871-887.
41. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A* **47**, 110-119.
42. Brünger, A.T. (1992). *X-PLOR, Version 3.1 Manual: A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.
43. Kraulis, P. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946-950.
44. Gonnet, G.H. (1994). A tutorial introduction to computational biology using DARWIN. ETH, Zurich, Switzerland.