9

Atomic structure of viral particles

MICHAEL S. CHAPMAN AND MAVIS AGBANDJE-McKENNA

Since 1991, we have learned much about the structure of parvoviruses. Those inclined towards outrageous claims might declare parvoviruses to be the first completed structural genomics project. We know structures for products of all (two!) genes. In fact the story is far from complete. Some components still have structures that are uncharacterized, and much remains in associating function with structure. However, much about function and mechanism can be learned from the available structures. This chapter will focus exclusively on the capsid, leaving structures of the replication proteins to be discussed elsewhere (Hickman *et al.*, 2002; James *et al.*, 2003, 2004; Yoon-Robarts *et al.*, 2004). This chapter will start from the molecular anatomy of the capsid, working towards function. The basis of the next chapter will be molecular virology and function as interpreted through a structural perspective.

ICOSAHEDRAL VIRUS STRUCTURE

Parvoviruses are among the smallest of the spherical or isometric viruses at a diameter of ~255 Å. A modest single-strand ssDNA genome of 5000 nucleotides is surrounded by a protein shell that is the main focus of this chapter. This protein capsid provides a protective coat to the DNA as it encounters environmental challenges in transmission from host to host, and cell to cell. A capsid may have several other functions: recognition of appropriate host cells, entry, intracellular transport, release of nucleic acid at the appropriate time and place, assembly of progeny virus and their release.

Enclosing itself in a proactive protein coat presents a challenge to a viral genome. The number of amino acids that can be encoded by a small viral genome is not sufficient to enclose it unless the same proteins are used repeatedly. The capsid must be modular, and largely self-assembling. These considerations led to the expectation that viruses would be assembled with the geometry of the platonic solids with multiple copies of the same facet. If the interactions of a subunit with several neighbors were self-encoded, formation of new associations could lead to self-directed assembly. The largest platonic solid is the icosadeltahedron with 60 identical faces (Figure 9.1). Caspar and Klug predicted that the simplest, so-called T = 1 isometric viruses would be icosadeltahedra, and that larger viruses could be assembled with 'quasi-equivalent' modules with higher T numbers and greater numbers of protein subunits per icosadeltahedral facet (Caspar and Klug, 1962). These icosadeltahedra have the same point group rotational symmetry elements as the 20-sided icosahedron, leading to the common terminology of 'icosahedral viruses', although they contain 60 not 20 identical units. Although previously suspected, the icosahedral nature of parvoviruses was unequivocally established from symmetry detected in the preliminary characterization of canine parvovirus crystals (Luo *et al.*, 1988).

The 5-3-2 point group symmetry contains 31 rotational symmetry elements that intersect at the center: six 5-folds, ten 3-folds and fifteen 2-folds (double these numbers if you count separately the views from opposite directions). If the structure is defined for all atoms between two neighboring 3-fold axes, and one 5-fold, then all atoms of the capsid

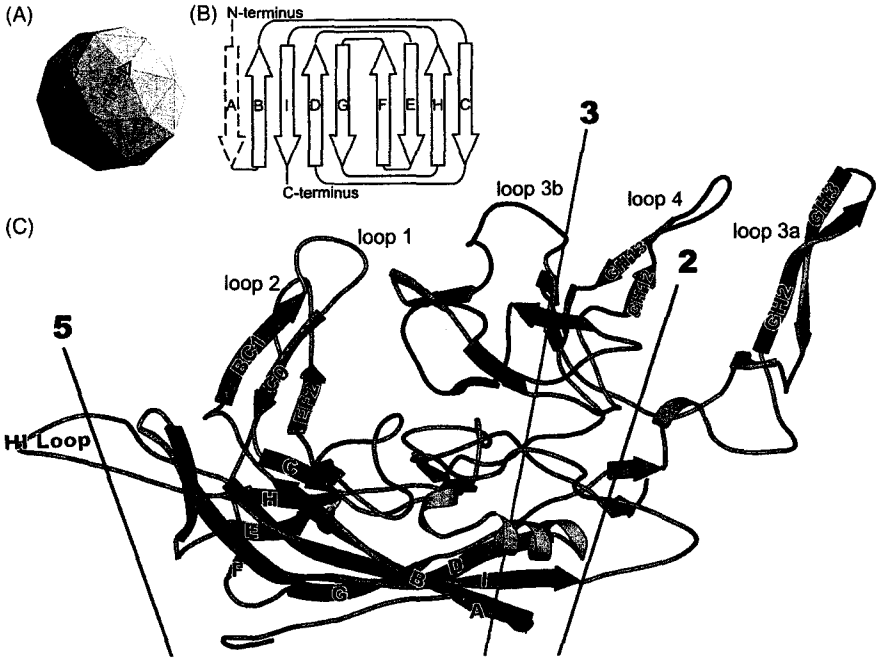

Figure 9.1 *The architecture of a parvovirus. (A) Parvoviruses are T = 1 icosahedral viruses, meaning that their near spherical capsid is comprised of 60 copies of the capsid protein. The surface can be divided into an icosadeltahedron of 60 equal triangles, each designating one protein. Of course the capsid protein is not triangular, so the surface triangle is comprised of a set of unique atoms from each of several neighboring subunits that combined would complete one triangle. The icosadeltahedron has vertices where either 5-fold or 3-fold axes pass towards the center of the virus, examples shown as a pentagon and triangles respectively. Midway between each neighboring pair of 3-fold axes passes a 2-fold (elipse). Successive application of the rotational symmetries can generate an entire capsid from a single subunit. (B) The topology of each subunit is a viral jellyroll β-barrel. The connectivity of the strands (arrows) is indicated in a schematic representation that is not supposed to indicate the length or structure of the intervening loops. Strand A is dashed, because it is present in parvo- and a number of other virus families, but not all. One can imagine the nine strands as consisting of two antiparallel sheets of five and four strands respectively. There is a discontinuity in the hydrogen-bonding between strands F and G as the second sheet is folded over the first. The schematic is simplified by omitting embellishing secondary structures from the loops. (C) The 3D fold of the AAV-2 subunit. The β-sheets that are highly conserved among parvoviruses are shown in fuschia at the bottom with strands labeled as in (B). The loops between the strands are very different from the loops of other viruses, and even differ significantly between the different genuses of parvovirus. The loops are particularly convoluted in parvoviruses with the GH loop accounting for ~250 residues. Most of these loops have extended chain structure, but there are additional secondary structural elements (red, orange, and purple), but these are not fully conserved, even among parvoviruses, and account for their unique properties in terms of antigenicity and association with cellular receptors. Loop secondary structures are labeled by the two letters of the bounding barrel strands and a number that increases towards the C-terminus. Other secondary structures are labeled with a single letter/number if they are nearly contiguous with a barrel strand. Thus CO immediately precedes strand C. ([C] is reproduced with permission from Xie* et al., *2002, copyright of National Academy of Sciences, USA, 2002.) See also Color Plate 9.1.*

can be generated by systematic application of all of the symmetry elements. These symmetry axes are often used as frames of reference for describing the locations of features in the structures. It is understood that the same features would be repeated near each of the identical symmetry axes.

Icosahedral virus structure determination

At atomic resolution, parvoviral capsid structures have been the exclusive domain of X-ray crystallography. Electronmicroscopy has the potential for near-atomic resolution but, with parvoviruses, has been applied only at lower resolution – 10 Å or less, sufficient to determine the surface topology, localization of components, and complexes of the virus with antibody or receptor fragments.

While the principles of viral crystallography are the same as protein crystallography, differences in the application of methods have implications for what can and cannot be visualized. The diffraction signal is weaker in proportion to the size of the viral complex. Averaging of the electron density between symmetry-related parts of the capsid is needed to ameliorate the poor signal/noise, and to help solve the phase problem. Diffraction amplitudes are Fourier amplitudes that,

Table 9.1 *Parvoviral atomic structure*

Genus	Structure	Abbreviation	Host	PDB accession number	Resolution	R-factor	Citation(s)
Densovirus	*Galleria mellonella* densovirus	*Gm*DNV	Wax moth	1DNV	3.7 Å	0.271	(Simpson *et al.*, 1998)
Parvovirus	Canine parvovirus	CPV	Dog	1DPV	2.9 Å	0.29	(Tsao *et al.*, 1991; Xie and Chapman, 1996) (Agbandje *et al.*, 1993; Simpson *et al.*, 2000)
	Feline panleukopenia virus	FPV	Cat	1FPV/1C8F	3.3/3.0 Å	0.285 (1CF8)	
	Minute virus of mouse	MVM	Mouse	1MVM	3.5 Å	0.327	(Agbandje-McKenna *et al.*, 1998)
	Porcine parvovirus	PPV	Pig	1K3V	3.5 Å	0.286	(Simpson *et al.*, 2002)
Erythrovirus	B19	B19	Human	1S58	3.5 Å	0.309	(Kaufmann *et al.*, 2004)
Dependovirus	Adeno-associated virus-2	AAV-2	Human	1LP3	3.0 Å	0.338	(Xie *et al.*, 2002)

if transformed, yield an image of the sample's electron density. Phases are also needed to synchronize the sinusoidal waves in the Fourier transformation, but conventional methods of experimentally determining phases by binding heavy atoms and/or changing the X-ray wavelength (Smith, 1991; McPherson, 2002) are difficult to apply to viruses (Tsao *et al.*, 1992). Fortunately the icosahedral symmetry can be applied as a constraint to improve crudely approximated phases. The starting approximations are usually calculated from related structures, electron microscopic images, or *ab initio* calculations (Rossmann 1972, 1995; Chapman, 1998; Chapman *et al.*, 1998). The implication of being forced to use the icosahedral symmetry during structure determination is that generally only those parts of the structure that adhere to the 60-fold symmetry will be seen.

PHYLOGENY AND ATOMIC STRUCTURES

The family *Parvoviridae* has two subfamilies – Parvovirinae whose members infect vertebrates, and Densovirinae whose members infect arthropods. There are four Densovirinae genera of which the wax moth (*Galleria mellonella*) densovirus (*Gm*DNV) from the genus *Densovirus* is the sole structural representative (Simpson *et al.*, 1998). All Parvovirinae genera infect vertebrates. Genus *Parvovirus* is the best characterized with structures of canine parvovirus (CPV) (Tsao *et al.*, 1991), feline panleukopenia virus (FPV) (Agbandje *et al.*, 1993), minute virus of mouse (MVM) (Agbandje-McKenna *et al.*, 1998) and porcine parvovirus (PPV) (Simpson *et al.*, 2002). Genus *Dependovirus* is currently represented solely by adeno-associated virus serotype 2 (AAV-2) (Xie *et al.*, 2002), but good progress is being made on other serotypes which we hope will follow soon. The most recent addition is the structure of a B19 capsid, recombinantly expressed in insect cells from a baculovirus vector (Kaufmann *et al.*, 2004), representing the *Erythrovirus* genus. Just when it was thought that all three Parvovirinae genera were represented by structures, taxonomic reclassification based on genomic differences has introduced two new Parvovirinae genera: *Amdovirus* (Aleutian mink disease virus, AMDV) and *Bocovirus* (bovine parvovirus and canine minute virus) (see Chapter 1). AMDV has been visualized by cryo-electron microscopy at 22 Å resolution (McKenna *et al.*, 1999), but there are no atomic structures for these new genera. Statistics for the atomic structures are summarized in Table 9.1. Unlisted are several variant CPV and FPV structures – full versus empty particles, mutants and different pHs (Wu and Rossmann, 1993; Llamas-Saiz *et al.*, 1996; Simpson *et al.*, 2000; Govindasamy *et al.*, 2003). CPV was the first structure to be determined (Tsao *et al.*, 1991), was pursued to the highest resolution refinement (Chapman and Rossmann, 1996), and most exhaustively interpreted (Xie and Chapman, 1996). It is therefore natural in this chapter to frame an overview around the CPV structure, and then to describe the differences of each of the other parvoviral structures and their implications relating to the viral life cycle.

Within each *Parvovirinae* genus, the capsid sequence identity may be as low as ~50 percent (e.g. Chapman and Rossmann, 1993; Rutledge *et al.*, 1998). This is a similar level of identity to that between different picornaviral genera. Even within a parvoviral genus, sequence might suggest an array of structural variations and diversity in host interactions as great as between, say, enteroviruses and rhinoviruses. Between the *Parvoviridae* genera, the capsid sequence identity ranges between 10 and 27 percent, much lower than between picornaviral genera. While overall subunit folds would be expected to be conserved, the low sequence identity might lead one to expect substantially different variations on a common structural theme, and perhaps several distinct ways that the virus might handle receptor-binding, immune evasion *etc.* In this light, the level of structural conservation that is seen is quite remarkable.

OVERVIEW OF THE STRUCTURE

Elements of the structures visualized

In the CPV structure, the capsid proteins are resolved starting at residue 22 of viral protein 2 (VP2) (Tsao *et al.*, 1991; Chapman and Rossmann, 1996; Xie and Chapman, 1996). The capsid gene actually encodes three capsid proteins that differ at their N-terminal ends. VP1 at 737 residues is the largest and is the product of different mRNA splicing that give it an additional N-terminal 153 residues (in CPV) relative to VP2, while VP3 is either another splicing product, or, as in CPV, a proteolytic cleavage product of VP2 (Clinton and Hayashi, 1976; Tattersall *et al.*, 1977), missing (in CPV) the N-terminal 15–20 residues. Whether VP2 or VP3 is the majority form in mature capsids depends on the species, but VP1 is always at about 10-fold lower concentration. In crystal structures, the protein becomes visible a few residues after the N-terminus of VP3 (or VP4 in Densoviridae), so the N-terminal regions distinctive to VP1 or VP2 are not seen.

Are the VP1/2 N-terminal regions actually present in the crystallographic samples? This has been most carefully examined for the adeno-associated virus-2 (AAV-2) crystals, which, by denaturing gel electrophoresis, has an identical 1:1:8–10 complement of VP1, 2, and 3 to that of fresh viral preparations (Rose *et al.*, 1971; Johnson, 1984; Xie *et al.*, 2004). VP1 is likely present in many of the parvoviral crystals with the exception of those, like B19, that are recombinantly expressed. Successful crystallization suggests an internal location for VP1, because there are too few copies to be symmetrically arranged on the outer surface, and any heterogeneity in lattice contacts usually foils crystallization. The C-terminal ~570 residues of VP1, 2, and 3 are of the same sequence and structurally indistinguishable. The distinctive parts of the VP1/2 N-termini are present at well below the 30 percent occupancy generally needed to observe crystallographically, and thus it is not surprising that these 'minority' components are not seen.

The DNA that is enclosed by the capsid does not have 60-fold symmetry in its sequence. For the most part, its 3D structure will not conform to the 60-fold symmetry constrained during structure determination. It is likely to be seen only where the capsid imposes its symmetry in DNA-binding sites. Up to 28 percent of the genomic DNA has been seen in CPV and MVM, bound to the capsid, but in other structures, no DNA has been seen.

CAPSID SUBUNIT FOLD

Viral jellyroll

One of the surprises from early viral atomic structures was the prevalence of the 'viral jellyroll' fold. It was first revealed in several plant, then insect virus structures (reviewed in Rossmann and Johnson, 1990; Harrison *et al.*, 1996). It is an eight-stranded β-barrel, sometimes known as a beta sandwich to emphasize that it contains two curved (usually) 4-stranded sheets facing each other (Figure 9.1). In the prevalent configuration, the sheet containing strands B, I, D, and G is closest to the nucleic acid. Strands are labeled alphabetically, starting from the N-terminus. The strands are all antiparallel with strands B and D pointing towards the 5-fold axis, I and G away. Like all β-sheets, the sheet has a right-handed twist looking along the direction of the chains. The effect (for parvoviruses) is that the end of the BIDG sheet farthest from the 5-fold is tangential to the viral surface, but by the time it approaches the 5-fold the sheet is nearly edge on. Like many other viruses, parvoviruses have a fifth 'A' strand hydrogen-bonded to the B strand. The sheet containing strands C, H, E, and F is shorter and closer to viral surface.

Comparing viral β-barrels from different viral families, it is at once surprising that the same motif can appear in so many guises, and to be expected, considering that there is no recognizable sequence similarity between many viruses that appear to share structural homology. The barrels have strands that vary in length up to 3-fold. Although they share the same topology, the strand lengths of CPV and bacteriophage ϕX174 (McKenna *et al.*, 1992), another ssDNA virus with comparable genome size, are poorly correlated (CC = 0.3) (Xie and Chapman, 1996). By convention loops are named according to the strands between which they bridge, so that the BC loop is between strands B and C. The loops of different viruses have little in common, other than generally: loops BC, HI, DE, and FG, which are close to the 5-fold axis, tend to be shorter than the CD, EF, and GH loops. More surprising is the number of ways that this fold can be used. The prevalent orientation (as in parvoviruses) is with the strands approximately tangential to the virus. Within this prevalent form, the overall orientation varies by as much as 19° (Chapman and Liljas, 2003). Other examples use the same fold in a different context, such as the adenovirus hexon with its β-strands radial to the virus (Roberts *et al.*, 1986), or have added or tandem-repeat domains (reviewed in Chapman and Liljas, 2003). There have been several attempts to explain why the jellyroll has been found in so many viral structures, but none of these rationalizations has survived extended scrutiny.

When parvoviral structures first appeared, available virus structures were mostly restricted to small RNA viruses. The jellyroll barrel was near-universal, but the number of exceptions was beginning to grow. There were questions as to whether a single-stranded DNA virus would share homology with RNA viruses. There was also speculation, based on the length of the primary sequence, that the capsid protein could encode perhaps three barrels in a pseudo-T = 3 viral architecture. As we shall see, the latter proved to be incorrect.

β-barrel of CPV and other parvoviruses

The CPV structure showed a single β-barrel domain per capsid protein (Tsao *et al.*, 1991) oriented in a standard way (Figure 9.1). For the most part, the barrels of parvoviral structures superimpose closely, but in *Gm*DNV it is shifted outwards 10 Å and rotated 7° (Simpson *et al.*, 1998). Parvoviral VP3 is about three times the size of the smallest jellyrolls from other viruses, and the reason is long loops between the strands, as has also proved to be the case in microphage bacteriophages (McKenna *et al.*, 1992, 1996). Several loop nomenclatures are in use. Tsao *et al.* defined loops 1 through 4 according to their location in 3D space. In the AAV-2 structure, loop 3 needs to be divided into 'a' and 'b' regions. Russell and colleagues (Rutledge *et al.*, 1998) introduced 'variable domains' I through IV where the sequence varied most between different AAV serotypes. Regions I, II, and IV correspond respectively to loops 3a, 3b, and 4. From RNA virus literature, there is a more detailed nomenclature that will be used here, where loops are labeled according to the strands that flank them. Thus the BC loop links β-strands B and C.

The β-barrel is less regular than text-book schematics might imply, with strand lengths varying from 3 to 22 residues. Of the two sheets facing each other in the β-barrel, it is the BIDG sheet near the inner surface of the capsid that is more regular and with longer strands, averaging 15 compared with six for the CHEF sheet. The barrels, irregularities and all, are well conserved through the species of parvovirus. Lists of secondary structural elements in the protein data bank appear to indicate inexact structural homology, but this is an artifact. The lists are generated automatically through prediction of hydrogen-bonding (Hutchinson and Thornton, 1996), a process that works well for the 2.9 Å CPV structure (Chapman and Rossmann, 1996; Xie and Chapman, 1996), but not for the structures at lower (~3.5 Å) resolution. The ribbon diagrams drawn from manual designations of secondary structure give a truer impression of the high structural conservation of the barrel. There are a few detailed differences: βG is continuous in AAV-2, but broken into two in CPV with the insertion of a threonine; strand C is shorter in B19 than in other parvoviruses. However, these are small differences in a highly conserved structure. The structural conservation is remarkable considering the relatively low sequence identity which, for example, is 21 percent for the barrel strands of AAV-2 and CPV, no greater than the overall identity for VP2.

EMBELLISHMENTS UPON THE BASIC BARREL

As in several other, but not all jellyroll virus structures, the BIDG sheet is actually 5-stranded with an N-terminal edge strand, βA. It is present in all parvoviruses. In all but *Densovirus*, there is a tight U-turn between βB and βA. *Densovirus* differs with domain swapping of the first strand (Simpson *et al.*, 1998; Liu and Eisenberg, 2002). βA is an N-terminal extension of βB, extending, without a tight turn, to become the first strand of the BIDG sheet of a 2-fold related subunit, i.e. the first (βA) strands of two subunits have been swapped. The exchange of the βA-βB intrasubunit hydrogen bonds of other parvoviruses for their intersubunit equivalents in *Densovirus* helps to cement the association of dimers in the mature capsid. Domain-swapped is certainly the dominant configuration in most of the *Gm*DNV VP4 subunits. If there were an unswapped configuration for minor components such as VP1, this would likely not be seen in the averaged electron density.

In CPV, the CHEF sheet is actually 6-stranded. At the N-terminal end of βA, the chain runs in an extended configuration without a turn, becoming, after a 3-residue gap, a 3-residue strand (βA0) that is a 6th strand of the CHEF sheet on the inner surface of the capsid (Xie and Chapman, 1996). On the outer edge, βEF2 from the EF loop 2 forms a first strand, an addition that is conserved in many parvoviruses.

LOOPY STRUCTURES

VP3 is about three times larger than the most compact β-barrels of other viruses. Only ~80 of the ~530 VP3 residues are within the strands of the β-barrel. Before the CPV structure determination, it was expected that each subunit might contain tandem repeats of the barrel as in the pseudo-T = 3 insect family Comoviridae (Chen *et al.*, 1989; Chapman and Liljas, 2003), or might contain N- or C-terminal domains decorating the surface, as in the plant family Luteoviridae (Harrison *et al.*, 1978; Chapman and Liljas, 2003). It was a complete surprise, with the structure of CPV (Tsao *et al.*, 1991), to find just a single domain with the core secondary structure accounting for only 20 percent of the primary sequence. The rest of the sequence was contained in the loops between the strands, one of which, the GH loop, was an unprecedented 220 amino acids. This insertion within the β-barrel fold is itself equal to a medium-sized protein. Subsequently, large insertions have been found in capsid proteins of Microviridae bacteriophages (McKenna *et al.*, 1992, 1996; Bernal *et al.*, 2003; Chapman and Liljas, 2003), but none quite as big.

Taking CPV as an example, the loop sizes are 35 residues for the BC loop, 19 for CD, 17 for DE, 72 for EF, 5 for FG, 221 for GH, and 18 for HI. Thus, parvoviruses agree with the general finding that the turns between β-strands near the 5-fold axis have shorter insertions than those pointing away.

Among the turns near the 5-fold, the BC loop or 'loop 2' is the longest (35 residues in CPV). The B and C strands are on the edges of their respective sheets, and near the 5-fold axes they are twisted nearly edge on, such that the BC turn is the closest to the viral surface. The loop insertion turns

90° towards the capsid surface and away from the 5-fold to associate with the EF loop (loop 1) originating from the opposite ends of the sheets (Figure 9.1, p. 108). The tips of both of these loops (near CPV VP2 residues 93 and 225 respectively) are highly exposed on the outer surface. Several of the loops contain secondary structural elements additional to the barrel strands. The BC loop contains a β-ribbon – two antiparallel strands hydrogen-bonded together. Some parvoviruses have a relatively tight turn between the ribbon strands, whereas others have a longer loop containing a short α-helix.

Next down from the surface is the HI turn. In all of the parvoviruses, the loop forms a loose turn of about 18 residues that interacts with a 5-fold-related subunit. The tip of this loop (near CPV VP2 520) forms much of the bottom of CPV's canyon, a ring of lower surface elevation encircling each 5-fold about 20 Å from the symmetry axis (Figures 9.2, 9.3 and 9.4). Although residues here are solvent accessible, they are not as exposed as the four protruding loops, and are perhaps not accessible to antibodies, so the HI loop was not one of the numbered loops in th CPV structure.

The ends of the D and E strands are closer to the vira center. The ~17 residue loop connecting them forms β-ribbon that extends parallel to the 5-fold axis to the vira surface where there is a tight turn near CPV VP2 residu 161. These two antiparallel strands are hydrogen-bonde with each other. Five of these ribbons surround each 5-fol axis. Although they have been referred to as a β-cylinder the five ribbons are not quite close enough for the ribbon to hydrogen bond to each other. It is the ribbons of the symmetry-equivalent DE loops that line the 5-fold pore, which will be discussed later. Closest to the inner surface is the FG loop which is the shortest at five residues.

The 5-fold ends of the barrel β-strands are roughly aligned. As the CHEF sheet strands are much shorter than those of the BIDG sheets, the loops have to bridge a larger gap at the 2-fold end of the barrel. The short CHEF sheet remains mostly edge-on, but the longer BIDG sheet has twisted such that the strands are running tangentially to

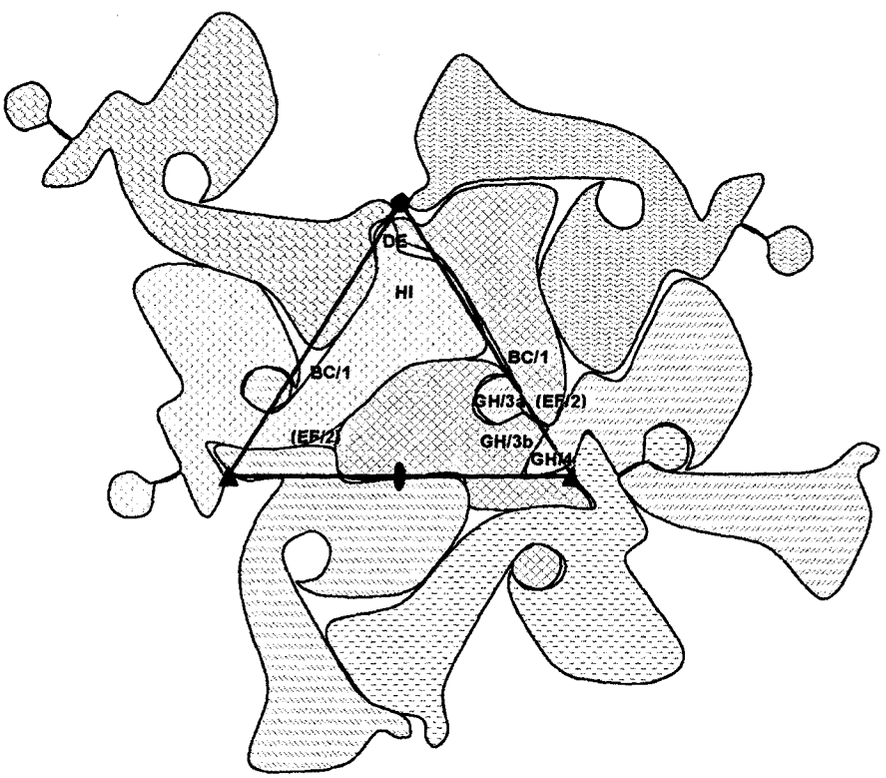

Figure 9.2 *Surface composition of parvoviruses. The triangle shows one of the 60 identical faces of an icosadeltahedron. As the subunit is not triangular, each face has contributions from several neighboring subunits, each shown with different shading. Near the 3-fold axes, the subunits are wrapped around each other intimately, helping to cement the assembly. Labeled are the barrel loops that contribute most to different parts of the surface. In many viruses these are named according to the flanking β-strands. In parvoviruses, the most surface-exposed loops have been given numbers 1–4 which are also shown. This schematic representation relates differences in the subunit primary/tertiary structure to the surface topology of the assembly. In AAV-2 and B19, loop 3a is particularly long, leading to prominent spikes (Xie* et al., *2002; Kaufmann* et al., *2004). In genus Parvovirus, the individual protrusions are not resolved, because an insertion in loop 4 adds bulk between the (shorter) symmetry equivalents of loop 3, filling the valleys of AAV-2. Loops 1 and 2 are longer in CPV and other members of the* Parvovirus *genus, leading to extension of the elevated region with a shoulder running towards the 5-fold axis.*

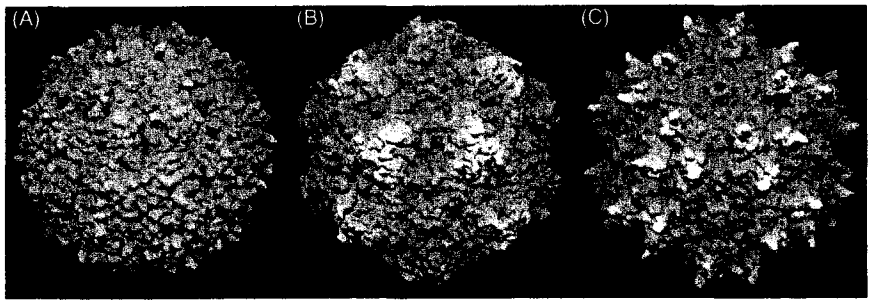

Figure 9.3 *Surface topology of parvoviruses (A) Densovirus; (B) Parvovirus; (C) Dependovirus and likely Erythrovirus. GRASP (Nicholls, 1992) surface representations calculated from the atomic structures are shown to scale for representatives of parvoviral genera (A) (insect) densovirus* Gm*DNV (Simpson* et al., *1998); CPV (Tsao* et al., *1991) as a representative of genus Parvovirus – MVM (Agbandje-McKenna* et al., *1998) and others would be very similar; and AAV-2 (Xie* et al., *2002). B19, the* Erythrovirus *of known structure (Kaufmann* et al., *2004) has an unseen disordered surface loop whose absence would distort a surface representation, but, in all other respects, its structure is most homologous to AAV-2 (Kaufmann* et al., *2004), so it likely shares similar surface topology. The view is down a 2-fold axis (like Figure 9.2) with 3-fold axes left and right of center, and a 5-fold above center. See also Color Plate 9.3.*

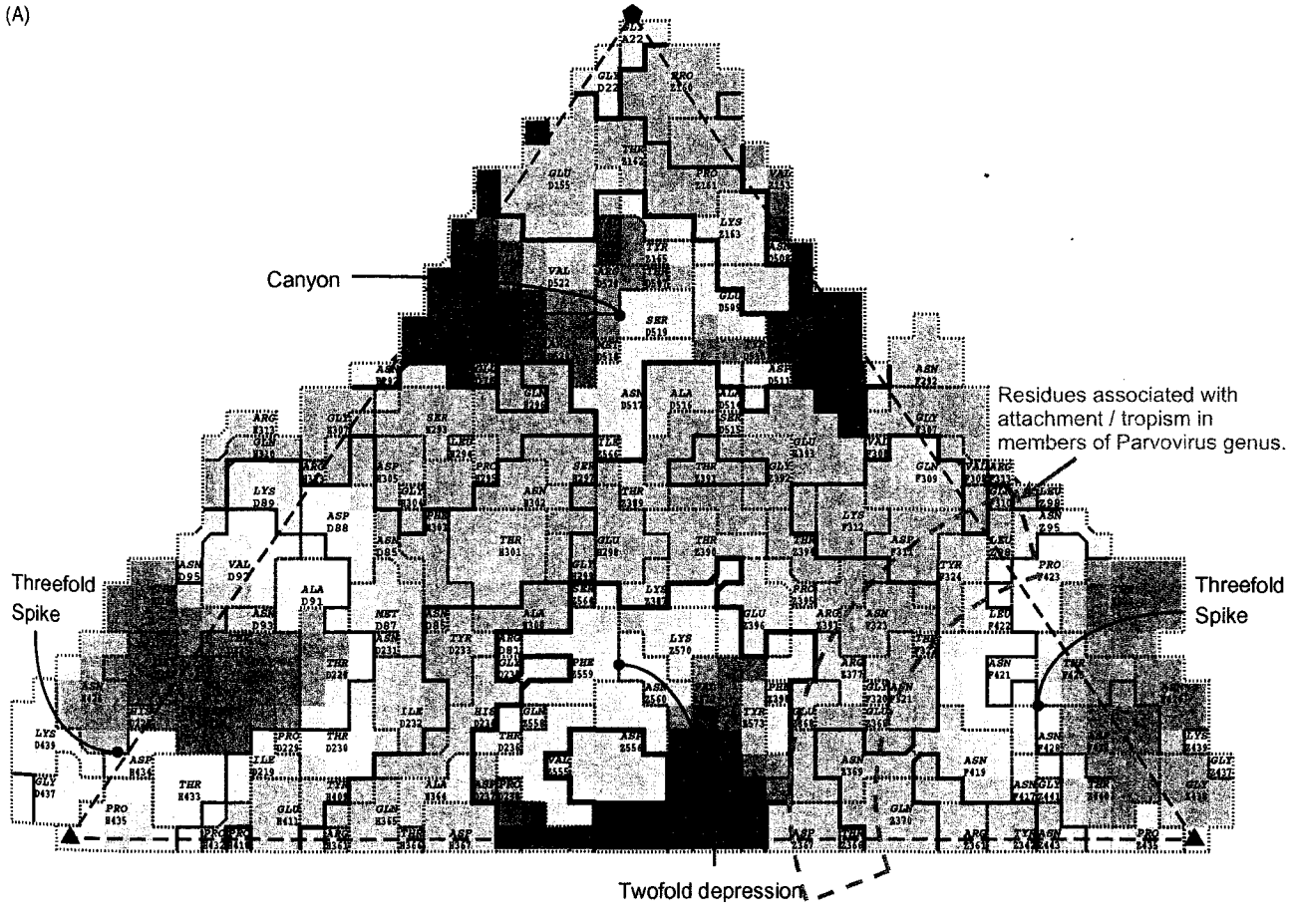

Figure 9.4 *Surface topologies and receptor-attachment of (A) CPV (Tsao* et al., *1991) and (B–D) AAV-2 (Xie* et al., *2002). (A and B) show Roadmap schematics (Chapman, 1993) of one of the 60 icosadeltahedral faces, bounded by the 5-fold (*top*) and two 3-folds (*left and right*), and viewed parallel to the 2-fold (*bottom center*). The solvent-accessible surface is colored like a topographical map with lower regions blue, elevated regions red. Surface residues are numbered with letter prefixes designating the subunit. (Different letters are used for CPV and AAV-2.) The region highlighted in (A) as associated with cell attachment and cell tropism represents a compilation from related parvoviruses aligned to the CPV surface. They include residues implicated in the cell tropism of MVM (Ball-Goodrich* et al., *1991), PPV (Vasudevacharya and Compans, 1992), ADV (Bloom* et al., *1988), and in the transferrin receptor-binding, cell specificity and hemaglutination of CPV and FPV (Parrish* et al., *1988a,b; Barbis* et al., *1992; Llamas-Saiz* et al., *1996; Govindasamy* et al., *2003; Hueffer* et al., *2003). See also Color Plate 9.4.*

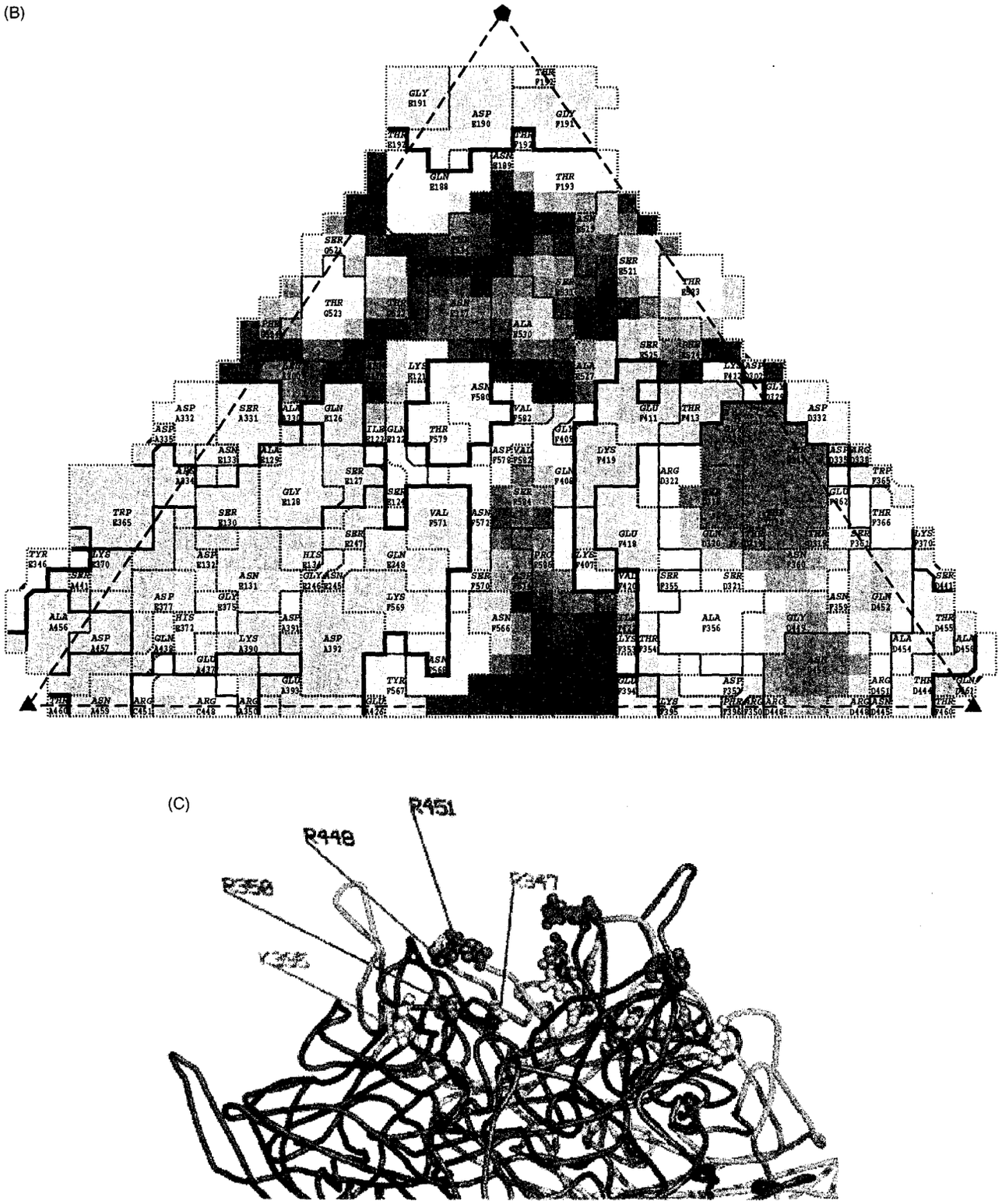

Figure 9.4 *(continued)* *(C) Residues implicated by mutagenesis in the binding of AAV-2 to analogs of the heparan sulfate proteoglycan receptor and in cell entry (Kern* et al., *2003; Opie* et al., *2003) are shown (VP2 numbering) on a trimer from the AAV-2 structure (Xie* et al., *2002) in which the backbones of subunits are in different colors. (D) The locations of these residues are mapped to a Roadmap surface of the region surrounding a 3-fold projected onto a plane. The region enclosing the implicated residues is shown with dashed blue lines, and lies between pairs of spikes, in the valley and on the side of one of the protrusions. These regions are repeated by the 3-fold symmetry, and one could imagine a heparan polysaccharide chain passing in through one valley, over the 3-fold and out through another. See also Color Plate 9.4.*

Figure 9.4 *(continued)*

the viral center and the 2-fold ends of the BIDG sheet are all roughly the same distance from the viral center. Of the three loops, the CD loop is the shortest (~19 residues) forming a relatively direct connection with two α-helices. The second and longer of these is buried at the dimer interface between subunits and among the most conserved parts of the capsid protein (Chapman and Rossmann, 1993). In the vertebrate parvoviruses, the CD loop runs below the GH loop that dominates the surface near the 3-fold axes.

The EF loop is 72 residues long in CPV and contains up to two α-helices and four β-strands. Only one of the strands is conserved among all parvoviruses – a strand that hydrogen-bonds with the ribbon of the BC loop to form a three-stranded antiparallel sheet, cementing the association between these two loops. The loop follows a similar path in all parvoviruses and its tip is a prominent part of the surface, but there are significant insertions and deletions in this loop as detailed later.

The GH loop

The GH loop has held several surprises. As described above, it was of unprecedented length (221 residues) in

CPV. The GH loop of CPV can be divided into loop 3 and loop 4, reflecting that after ~80 residues, the chain has returned back close to the barrel before heading out to the surface again as loop 4 (Tsao *et al.*, 1991). Together they account for the large elevated area near the 3-fold axes. The second surprise in the CPV structure was that subloops from neighboring, 3-fold related subunits were interdigitated, greatly increasing the surface shared between adjacent subunits, and that the interdigitated subloops were stacked one on top of another. This indicated a sequential folding process that could only occur after the 3-fold related subunits had come together during assembly (Xie and Chapman, 1996). Glycines at the bases of the subloops, some of which are conserved, suggest flexible hinge points as these associations are made dynamically during assembly (Chapman and Rossmann, 1993; Xie and Chapman, 1996).

The third surprise came with the *Densovirus* structure (Simpson *et al.*, 1998). The reason that the capsid protein is smaller than the other parvoviruses is because it lacks loop 4. Loop 4 forms much of the bulk of what was called the 3-fold spike in Tsao *et al.* (1991), but might be more properly described as a flat elevated massif surrounding each 3-fold axis. Its absence in the arthropod parvoviruses leads to a much flatter surface topology (Figure 9.3, p. 113). The relative lack of surface features could perhaps be rationalized by a much different insect immune system that might not enforce as much selective pressure for change in the surface to escape immune neutralization. The fourth surprise came with the structures of AAV-2 and B19 (Xie *et al.*, 2002; Kaufmann *et al.*, 2004), and the extent of differences between the dependo- and erythroviruses compared with genus *Parvovirus*, as exemplified by CPV and MVM. In the vertebrate parvoviruses, the GH loops meander with little secondary structure other than some β-ribbons and small sheets that likely help stabilize the irregular fold. AAV-2 and CPV share a total of eight homologous β-strands, within buried regions of the loop, where the structures are very similar. AAV-2 has five additional β-strands, near the more exposed parts of the GH loop. Particularly striking is the inserted β-ribbon near AAV-2 VP2 317 that launches a ~20 residue 'finger' to the tip of a prominent protrusion, about 25 Å from the 3-fold.

SURFACE TOPOLOGY

Three distinct outside surface topologies have been seen in the parvoviral structures. Common to two of these are the following features: an elevated region near the 3-fold axis, a depression between the 3-fold elevated regions at the 2-fold axis, and another depressed region encircling the 5-fold axis (Figures 9.3 and 9.4, p. 113). These two groups of viruses differ only in the shape of the 3-fold elevated region (see below). By contrast, a third type of topology was seen in *Densovirus* which is relatively spherical, flat, and featureless (Simpson *et al.*, 1998). The absence of surface features that are antigenic in other parvoviruses has been suggested to be the result of the lack of an adaptive immune response (and its associated evolutionary pressure) in arthropods (Simpson *et al.*, 1998).

The characteristic feature of genus *Parvovirus* (CPV, MVM, PPV etc.) is the '3-fold spike' or elevated massif, 70 Å wide and 22 Å high, centered on each 3-fold axis. In comparing CPV (Tsao *et al.*, 1991), FPV (Agbandje *et al.*, 1993), MVM (Agbandje-McKenna *et al.*, 1998), and PPV (Simpson *et al.*, 2002), one is first struck by the overall similarity in the surface shape of all of these structures from the *Parvovirus* genus. The most antigenic regions of CPV (near residues 93 and 300 [Parrish *et al.*, 1988a]) lie on a spur of the massif running towards the 5-fold axis (Tsao *et al.*, 1991). The tips of the DE loops form a smaller elevated feature, forming a ring of 22 Å diameter surrounding each 5-fold axis that is about half the height of the massif (Tsao *et al.*, 1991). These features give rise to two regions of lower elevation. The first is an 11 Å wide valley, between the 5-fold ring and the 3-fold massifs, that, by analogy to the picornaviral structures, has been named the canyon (Figure 9.4, p. 113) (Tsao *et al.*, 1991). The second, named the 2-fold depression, is centered between two adjacent 3-fold massifs. The depression is implicated as the cell-receptor-binding locus in MVM through a recent crystallographic structure of MVM complexed with a fragment of its receptor, sialic acid (Lopez-Bueno *et al.*, submitted). This is the first direct visualization of virus-receptor interactions in genus *Parvovirus*. There is genetic evidence implicating neighboring, but distinct regions in several species. This includes mutants affecting CPV-binding to the transferrin receptor and affecting CPV/FPV host range (Govindasamy *et al.*, 2003). It also includes residues implicated in the cell tropism of MVM (Ball-Goodrich *et al.*, 1991), PPV (Vasudevacharya and Compans, 1992), and AMDV (Bloom *et al.*, 1988). The majority of the implicated amino acids cluster on the edge of the massif where it faces the 2-fold depression and 5-fold axis (Figure 9.4, p. 113). There are outliers to this clustering, as might be expected with the combined effects of:

- incomplete genetic characterization;
- different receptors for each virus; and
- cell tropism phenotypes mediated by primary or secondary external receptors, or those involved with intracellular targeting/transport.

More detailed discussions of the evidence and caveats are presented in the next chapter and elsewhere (Chapman and Rossmann, 1993; Hueffer and Parrish, 2003), but, overall, there is a compelling case that the same general region of the massif may be involved with receptor-binding in many members of the *Parvovirus* genus.

Genera *Dependovirus* and likely *Erythrovirus* have a similar surface topology that is distinct from that of *Parvovirus*.

In the B19 structure, a key surface loop is disordered and unseen, but with close homology in all other parts of the structure to AAV-2, all indications are that the *Erythrovirus* surface is similar to that of *Dependovirus* (Kaufmann *et al.*, 2004) with the minor exceptions detailed below. In both, the broad massif of CPV-like viruses is replaced by three distinct spike-like protrusions that rise to greater elevation in AAV and B19. Furthermore, the immediate vicinity of the 3-fold is not part of the elevated region as in CPV and MVM, but is at the lower 'average' elevation. Therefore the valleys between the symmetry-related spikes are connected at the 3-fold (Figure 9.3, p. 113). The B19 surface is subtly different from the AAV-2 (Xie *et al.*, 2002; Kaufmann *et al.*, 2004) with spur-like extensions from the spikes leading to more of a continuous rim at medium elevation surrounding the canyon (Kaufmann *et al.*, 2004).

Receptor-virus interactions have not yet been visualized directly for dependoviruses, but genetic evidence combined with structural inferences strongly implicate a region on the side of the spikes for AAV-2. Of panels of insertional and other mutants prepared and phenotypically characterized, a number were found to have reduced heparin-binding (Rabinowitz *et al.*, 1999; Wu *et al.*, 2000). Several disparate loci within the primary sequence had impact. Not all, but the preponderance of these loci cluster as different regions of the primary sequence are folded together and associated with neighboring subunits in the structure of AAV-2 (Xie *et al.*, 2002). The clustering is not perfect, but there are additional indicators that the cluster is the site of receptor-binding, and that other loci might have remote conformational effects. The dominant linear epitope of monoclonal C37-B attachment-inhibiting antibody was a neighboring surface peptide (Girod *et al.*, 1999; Xie *et al.*, 2002). Also, calculation of the Poisson–Boltzmann electrostatic potential shows the implicated surface on the side of the spike to be strongly positively charged (Nicholls, 1992; Gerstein *et al.*, 2001; Xie *et al.*, 2002), as expected for a protein binding a heparan-sulfate receptor (Summerford and Samulski, 1998; Mulloy and Linhardt, 2001).

As for B19, there is a ~10 Å discrepancy between the EM-visualized globoside (receptor) binding site (Brown *et al.*, 1993; Chipman *et al.*, 1996) and the location of the B19 neutralizing epitopes (Sato *et al.*, 1991) that align structurally with residues implicated in AAV-2 receptor binding (Xie *et al.*, 2002; Kaufmann *et al.*, 2004). It is not yet clear whether the antibody footprint might (not implausibly) extend over the 3-fold to block the binding of the B19 receptor, or whether higher resolution imaging of the complex might lead to closer agreement. Abundantly clear is that the sides of the spikes, and the valleys running between the spikes and over the 3-fold axes, are a common general location of receptor-binding in dependoviruses and erythroviruses. As with genus *Parvovirus*, the surface topology of exposed protrusions appears to be key to understanding cell attachment and the antigenic characteristics of all parvoviruses.

One enigma might have been resolved recently. Members of the *Parvovirus* genus have similar surface topologies that are distinct from those of *Dependovirus* and *Erythrovirus*. AMDV's surface, as seen in the cryo-electron microscopy reconstruction at 22 Å resolution (McKenna *et al.*, 1999) appears more like AAV-2 and B19 than the other members of the *Parvovirus* genus in which it was previously classified. The new taxonomy (Tattersall, reviewed in Chapter 1), reflecting genomic differences, has AMDV in a separate *Amdovirus* genus. The three types of surface topology in the *Parvoviridae* family are now more consistent with phylogenetic trees. The Densovirus subfamily has a relatively flat surface. Most of the Parvovirinae subfamily have three distinct peaks surrounding each 3-fold, with the exception of the *Parvovirus* genus where these are merged into a single massif centered about the 3-fold.

RELATION OF SUBUNIT STRUCTURE TO SURFACE TOPOLOGY

The differences in the four surface loops, especially loops 3 and 4, lead to the three different surface topologies (Figures 9.2, p. 112, 9.3, p. 113, and 9.5, p. 118). The relatively featureless surface of densoviruses is due to the absence of a loop 4 that is the foundation for the elevated regions in other parvoviruses. For the elevated massif of genus *Parvovirus*, all four surface loops contribute. Subloops from the GH-loop (or loops 3 and 4) from adjacent subunits intertwine to form the center of the massif. In the dependoviruses and erythroviruses, subloops from adjacent subunits also interdigitate. However, with a number of sequence insertions and deletions, these subloops are deployed in a different manner. The subloops interdigitate with more of a radial orientation, not tangential. The contacts are similarly intimate, stabilizing the 3-fold associations, but the loops are not dynamically stacked on top of each other. The more radial subloop orientation leads to a well-separated spike that is repeated three times around the symmetry axis.

The exposed tips of the subloops come to the CPV and AAV-2 surfaces in slightly different positions. The most prominent part of the CPV-like massif superimposes on the valley between spikes in AAV-2 (Figure 9.3, p. 113). The prominence of the AAV-2 spikes is not only due to the extension of loops 3 and 4, but also to the truncation of loops 1 and 2, that in genus *Parvovirus* would build the edges of and fill in the valley between the spikes. The modest differences in surface topology between AAV-2 and B19 are due to longer BC and EF loops (1 and 2) in B19. Although of length more similar to CPV, the BC and EF loops of B19 come to the surface in positions different from CPV. They join a 10-residue insertion near the C-terminus to extend the spikes with a spur towards the 2-fold depression, helping to form a rim around the canyon (Kaufmann *et al.*, 2004).

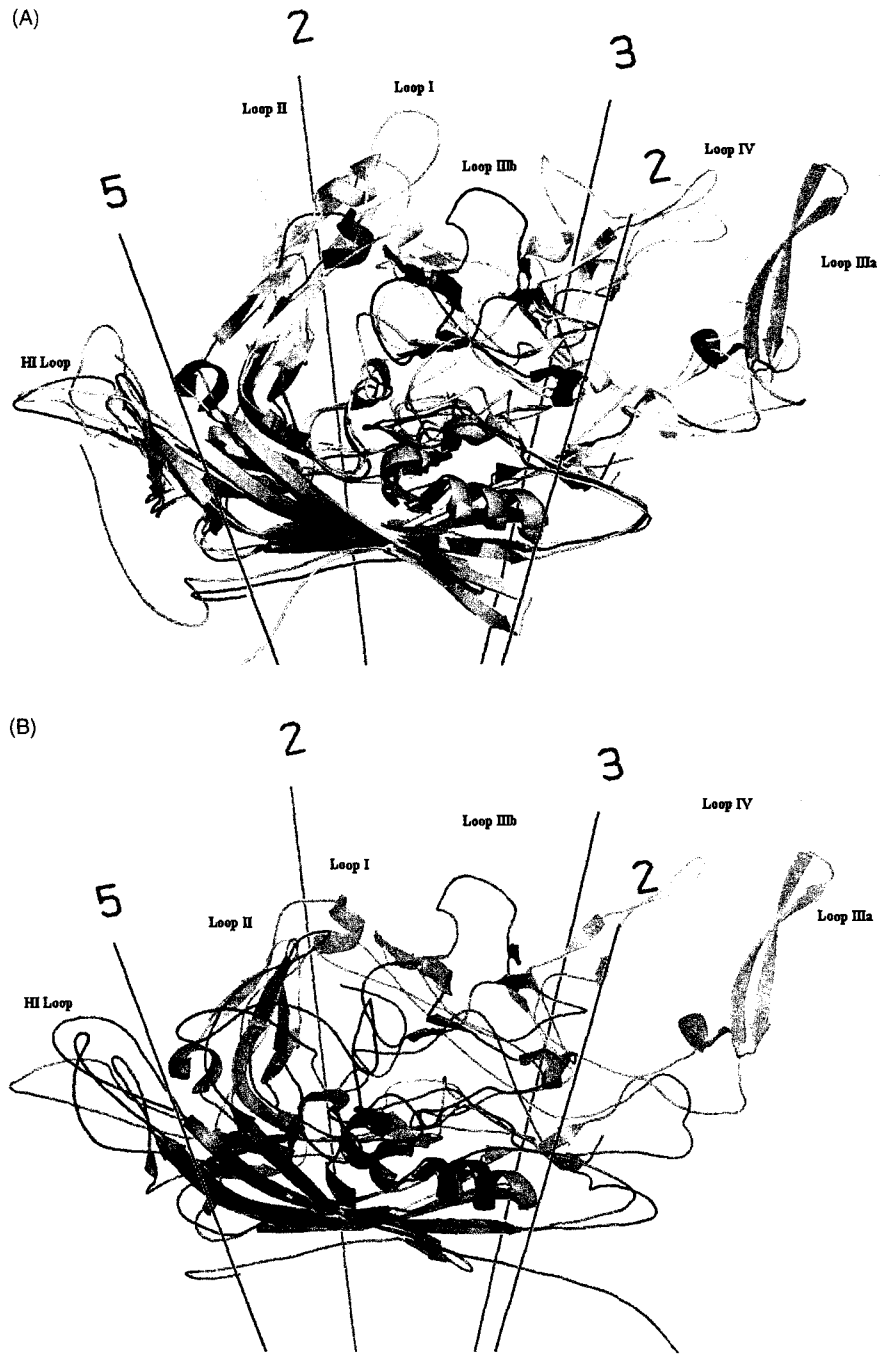

Figure 9.5 *Superimposition of the subunit structures of: (A) CPV (green, Tsao* et al., *1991), AAV-2 (red, Xie* et al., *2002), and B19 (yellow, Kaufmann* et al., *2004), and (B) DNV (blue, Simpson* et al., *1998) and AAV-2 (red). Strands of the β-barrel superimpose well, as do the loops of AAV-2 and B19. The biggest differences are in loops 3 and 4, between CPV, DNV and a group including both AAV-2 and B19. DNV superimposes less well than the other parvoviruses* (panel b)*. See also Color Plate 9.5.*

EMBELLISHING SECONDARY STRUCTURES

While the long barrel loops do not have recognizable folds and have less secondary structure than most protein domains, they have a few α-helices, and a number of β-strands. In several cases (Figure 9.1, p. 108) the bases of subloops are a pair of strands forming a β-ribbon, tying the ends together. In some cases the strands are part of small β-sheets, often bridging between neighboring subunits. Some of the details differ, but consider CPV where hydrogen-bonding and secondary structure has been analyzed exhaustively (Xie and Chapman, 1996). GH1, a small β-sheet with

strand lengths of 3–6 residues, consists of strands βGH2/βGH3 from one subunit and βGH8/βGH9 from a 3-fold neighbor. Sheet GH2 has strands of 3–5 residues: βGH4/βGH6 from one subunit and βGH1 from a 3-fold neighbor. Sheets BCE and GH3 are short 3- and 2-stranded sheets within their own subunits.

These additional secondary structures are not conserved in *Densovirus*, where the loops are absent or of very different structure. However, they are mostly conserved in our examples of *Erythrovirus* and *Dependovirus* (Xie *et al.*, 2002; Kaufmann *et al.*, 2004). They likely have two important functions. First they help solidify a foundation for the more exposed and variable tips of the loops that decorate the surface in a variety of configurations. Second, the hydrogen-bonding between strands from different subunits, helps to rigidify the interfaces.

SUBUNIT INTERACTIONS IN THE CAPSID ASSEMBLY

Polar interactions, including hydrogen bonding, are unlikely to contribute to free energies of association, because lost solvation energy has to be subtracted from any favorable interaction energy (Kyte, 1995). Generally, assembly is driven by hydrophobic interactions that can be estimated semi-empirically from the change in (non-polar) molecular surface energy (Gerstein *et al.*, 2001). Calculations from the CPV structure (Xie and Chapman, 1996) show that the interactions between 3-fold related subunits is very favorable. The association appears very stable and unlikely to dissociate. It does not indicate whether trimers are, or are not, an assembly intermediate. The intertwining of the GH subloops accounts for more than half of the calculated interaction energy. The unassembled subunit structure is unknown, but without their neighbors, the loop structures would likely be very different, and the net change in solvent accessible area likely much less than when estimated just from the subunit structure in its assembled configuration. When the flexible GH subloops are completely omitted from the calculations, 3-fold interactions appear slightly more favorable than 5-fold, and then 2-fold. This is no more than a crude guide, not only because of the uncertain effect of flexible loops, but because estimated interaction energies speak to binding equilibria, but assembly pathways are likely kinetically controlled, and there is no data on the dynamics of relevant loop conformational changes. Experimental data for MVM is beginning to emerge from the laboratories of Almendral and others, so perhaps assembly pathways will soon be put on a firmer footing.

The 5-fold pore

A tube of partially ordered electron density was found to run exactly along the 5-fold axis in CPV (Xie and Chapman, 1996), down a channel formed by the DE loop β-ribbons of five symmetry-related subunits. The β-barrel starts on the inside surface of the capsid near the 5-fold axis. Following βA0 backwards towards the N-terminus, the backbone trace diverges at Ser35-Gly34 with some conformers leading inwards and disappearing. About 1 in 8 conformers turn tightly and proceed outwards along the 5-fold axis through the disordered electron density (Figure 9.6). The pore is large enough to accommodate the peptide chain from one of the five surrounding subunits. The density level corresponds to two-thirds of the pores filled, but after icosahedral averaging, we cannot tell which, and details of the density are lost. All parvoviruses have a region rich in glycines, suitable for the tight turns, flexible configuration, and small size required to fit through the pore. In CPV, the chain could be traced towards the N-terminus to the site of VP2/3 cleavage, which is at the external surface near the 5-fold.

The length of the channel varies in different parvoviruses, but is correlated to the length of the glycine-rich sequence. In *Gm*DNV, the glycine-rich sequence is five residues shorter than in CPV, but the channel is also only about half the length (20 versus 45 Å). B19 is intermediate (Kaufmann *et al.*, 2004). Among the various parvovirus structures, the pore is not uniformly filled. Structures determined

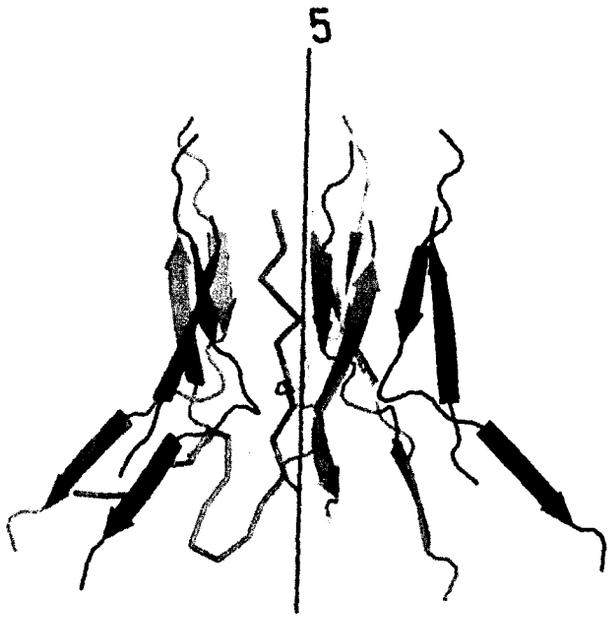

Figure 9.6 *The 5-fold pore of CPV (Xie and Chapman, 1996). The β-ribbon of the DE loop is shown along with its symmetry equivalents (in different colors) surrounding the 5-fold. The tight turns of the ribbons on the external viral surface (top) are disordered and can not be modeled, but are visible in AAV-2. Up through the center, along the 5-fold, runs the glycine-rich sequence from one of the five neighboring subunits (green), so that up to one of five capsid protein N-termini can be on the external surface (top). Residues are numbered according to VP2, though it is not known whether it is VP1, 2 or 3 that is visualized crystallographically. See also color plate 9.6.*

of empty capsids (including all that are recombinantly expressed) lack any electron density along the 5-fold. Such cases include B19 (Kaufmann *et al.*, 2004), the empty capsid forms of CPV (Wu and Rossmann, 1993), and FPV (Agbandje *et al.*, 1993), indicating a relationship between the presence of the DNA and the protein contents of the 5-fold pore. In B19 (Kaufmann *et al.*, 2004) the pore was narrower than in full capsids of CPV, with a constriction down to 9 Å (atom center–atom center). As is, this would be too narrow for passage of a polypeptide chain, but there is likely enough flexibility for room to be made. However, even among structures of DNA-containing capsids, the contents of the pore have not been uniform. The channel was filled in the structure of MVMi (Agbandje-McKenna *et al.*, 1998) as in CPV. The pore in AAV-2 contained only weak discontinuous density that could not be modeled, and the same was true in *Gm*DNV (Simpson *et al.*, 1998). The pore density is expected to be weak, because at most one of the five surrounding subunits can have its N-terminal region extending up the pore to the outside. Within the pore, the structure visualized is the average of five different orientations, so parts of its density will be smeared. At the external surface, the polypeptide could take one of five equivalent (non-overlapping) routes, each of which would have occupancy less than detectable crystallographically. More mysterious is our failure to see the four of five N-terminal regions that remain internal at copy number that should be detectable. It suggests that the extreme N-termini of the VP2/3 that remain internal are of variable configuration.

A consequence of failing to resolve the VP2 N-termini is that there is not direct crystallographic evidence of whether it is VP1, 2, or 3 (or a mixture) that have external N-termini. It is assumed that VP1 must be externalized for its phospholipase activity to be put to use, but this could be the result of a receptor or entry-triggered conformational change. A variety of evidence indicates that some VP1 can be external in at least some parvoviruses. This included VP1-specific antigenicity in particles of CPV and B19 (Rimmelzwaan *et al.*, 1990; Kajigaya *et al.*, 1991; Rosenfeld *et al.*, 1992). Cryoelectron microscopy of AAV-2 empty particles gave a contrary indication, with 'fuzzy globules' suggesting a location on the inner capsid surface near the 2-fold (Kronenberg *et al.*, 2001). Weak and disordered density is expected for a flexible feature present in only a proportion of cases, but caution is needed in interpreting such features. Also, the location in these empty particles might differ from that in DNA-containing particles.

Crystal packing considerations argue that the VP1 are internal. Denatured crystals of AAV-2 show the expected ratios of VP1, 2, and 3 (Xie *et al.*, 2002, 2004) with insufficient VP1 to populate every 5-fold axis. If VP1-unique domains decorated some of the outside surface in a heterogeneous way, it is inconceivable that they would not disrupt the exact crystalline arrays in some of the now many parvovirus crystal forms. If the biophysical evidence for internal localization is to be reconciled with biochemical and immunological evidence that it is external, the inescapable conclusion is that under some conditions the location can change dynamically. The pore is wide enough only for an extended peptide with small side chains. None of the existing structures give much indication of how the assembly could become plastic enough for the extrusion of an entire domain. (Structures solved under varying conditions, including DNA-containing and empty particles, show structures that are only subtly different [Wu and Rossmann, 1993; Xie and Chapman, 1996]). Similar structural reconfigurations are accomplished by other viruses, such as the externalization of VP4 in picornaviruses (Li *et al.*, 1994), so it is likely that parvoviruses are likely to reveal some interesting secrets in the future. Mutational studies on the phenotypes of pore-lining residues are beginning to give insights into their relevance to N-terminal location, DNA encapsidation, and capsid assembly (Farr and Tattersall, 2004; Reguera *et al.*, 2004).

There is independent evidence of dynamic conformational changes in parvoviruses. This includes tryptic susceptibility of VP2 in full, but not empty heat-treated MVM particles (in which it is assumed that the VP2-unique region might be sequestered internally) (Clinton and Hayashi, 1976; Tattersall *et al.*, 1977; Carreira *et al.*, 2004; Farr and Tattersall, 2004; Reguera *et al.*, 2004). This is consistent with the observation that only DNA-containing particles have filled 5-fold pores (see above). An interesting twist comes from considering the impact of domain-swapping in *Gm*DNV (Simpson *et al.*, 1998). If all subunits are in the domain swapped configuration (and there is not a minority unswapped component), then the N-termini that fill *Densovirus* 5-folds must be threaded following at least partial assembly of the capsid, and can not be in place as pentamers are formed. One might also expect conformational changes associated with attachment, endosomal entry, nuclear targeting, and initiation of uncoating (Vihinen-Ranta *et al.*, 1998; Bartlett *et al.*, 2000; Seisenberger *et al.*, 2001). Attempts to characterize any conformational changes are just beginning. The structure of CPV at low pH showed conformational changes only in a surface loop (Simpson *et al.*, 2000), but there are indications that low pH is not the only endosomal trigger required for productive viral entry (Vihinen-Ranta *et al.*, 1998).

DNA BINDING SITE

Although the inside surfaces of some other viruses are positively charged for interactions with DNA or RNA (Ban and McPherson, 1995), it was close to neutral in CPV (Xie and Chapman, 1996), consistent perhaps with ssDNA that remained (at least partly) single-stranded (rather than locally hairpin/duplex) (Chapman and Rossmann, 1995). With the icosahedral averaging, DNA was expected to be seen only when the capsid enforced this symmetry upon an otherwise unsymmetric genome (see above). Eleven nucleotide

fragments of CPV ssDNA were seen (Tsao *et al.*, 1991; Chapman and Rossmann, 1995), which, when multiplied over the 60 symmetry equivalent positions, accounted for about 13 percent of the genome. Its footprint consisted of mostly polar, but uncharged amino acids (Xie and Chapman 1996). The presence of DNA causes small conformational changes to the inside capsid surface (Wu and Rossmann, 1993). These are local to the DNA binding site, and there is no wholescale rearrangement communicating the presence of DNA to the outer surface. Of course we cannot comment about the unseen N-termini of VP1–3 that could have such a role. In MVMi, 23 nucleotides or 28 percent of the genome is visualized, bound in a similar location (Agbandje-McKenna *et al.*, 1998).

In CPV, the ssDNA has an unusual inverted-loop structure with phosphates chelated inside by 2 Mg^{++}, and bases pointing outwards to interact with the capsid protein, four of them in a single-stranded stack. Characteristic features in the electron density showed that there was a weak but detectable preference for particular base types in some positions, and that the average image of 60 DNA fragments did not contain a random mixture of all possibilities (Chapman and Rossmann, 1995). A profile of sequences (Gribskov *et al.*, 1990), estimated to be consistent with the crystallographic electron density, was not an exact match to any one locus of the CPV genome, but was a closer match than random to a number of sequence elements (Chapman and Rossmann, 1995). Efficient packing dictates that 60 different regions of the sequence must be compatible with the symmetry-equivalent capsid-binding sites, and it may be that CPV selects its own DNA for encapsidation, not with a single highly specific recognition sequence, but through the multiplicative effect of many different regions of its genome each having modest similarity to a consensus binding sequence. A requirement to match only approximately a binding sequence would severely constrain the coding function of the genome.

Ten of MVM's 23 nucleotides seen correspond to those in CPV (Agbandje-McKenna *et al.*, 1998) and many of the others corresponded to weak uninterpretable density in CPV, suggesting that the DNA binding site is a conserved feature. In *Gm*DNV, there is at best disordered density for some of the DNA, and it is not in a homologous position (Simpson *et al.*, 1998). Several of the structures were determined as empty capsids, including B19, FPV, and PPV, so of course no DNA is expected. One can only speculate why ordered DNA has not yet been seen in AAV – perhaps because unlike CPV and MVM, both positive and negative strands of the genomic DNA are packaged.

CONCLUSION

Structural studies have now extended to many of the parvovirus genera. This has opened a new chapter of comparative analysis from which further functional inferences will emerge. Over the decade-plus of parvovirus structure, the combination of genetic and structural characterizations has greatly extended our understanding of selected examples, and now we can expect similar advances throughout the family. The atomic structures are providing a framework for the planning and interpretation of functional studies, which will be reviewed in the next chapter. Emerging studies of receptor and immune complexes offer particularly exciting prospects of advancing our understanding of the molecular bases and dynamic processes of viral-host interactions.

REFERENCES

Agbandje, M., McKenna, R., Rossmann, M. G., Strassheim, M. L. and Parrish, C. R. 1993. Structure determination of feline panleukopenia virus empty capsids. *Proteins* **16**: 155–71.

Agbandje-McKenna, M., Llamas-Saiz, A. L., Wang, F., Tattersall, P. and Rossmann, M. G. 1998. Functional implications of the structure of the murine parvovirus, minute virus of mice. *Structure* **6**: 1369–81.

Ball-Goodrich, L. J., Moir, R. D. and Tattersall, P. 1991. Parvoviral target cell specificity: acquisition of fibrotropism by a mutant of the lymphocyte strain of minute virus of mice involves multiple amino acid substitutions within the capsid. *Virology* **184**: 175–86.

Ban, N. and McPherson, A. 1995. The structure of satellite panicum mosaic virus at 1.9 Å resolution. *Nature Structural Biology* **2**: 882–90.

Barbis, D. P., Chang, S. F. and Parrish, C. R. 1992. Mutations adjacent to the dimple of the canine parvovirus capsid structure affect sialic acid binding. *Virology* **191**: 301–8.

Bartlett, J. S., Wilcher, R. and Samulski, R. J. 2000. Infectious entry pathway of adeno-associated virus and adeno-associated virus vectors. *Journal of Virology* **74**: 2777–85.

Bernal, R. A., Hafenstein, S., Olson, N. H. *et al.* 2003. Structural studies of bacteriophage alpha3 assembly. *Journal of Molecular Biology* **325**: 11–24.

Bloom, M. E., Alexandersen, S., Perryman, S., Lechner, D. and Wolfinbarger, J. B. 1988. Nucleotide sequence and genomic organization of Aleutian mink disease parvovirus (ADV): Sequence comparisons between a non-pathogenic and a pathogenic strain of ADV. *Journal of Virology* **62**: 2903–15.

Brown, K. E., Anderson, S. M. and Young, N. S. 1993. Erythrocyte P antigen: cellular receptor for B19 parvovirus. *Science* **262**: 114–17.

Carreira, A., Menendez, M., Reguera, J., Almendral, J. M. and Mateu, M. G. 2004. In vitro disassembly of a parvovirus capsid and effect on capsid stability of heterologous peptide insertions in surface loops. *Journal of Biological Chemistry* **279**: 6517–25.

Caspar, D. L. D. and Klug, A. 1962. Physical principles in the construction of regular viruses. *Cold Spring Harbor Symposium in Quantitative Biology* **27**: 1–24.

Chapman, M. S. 1993. Mapping the surface properties of macromolecules. *Protein Science* **2**: 459–69.

Chapman, M. S. 1998. Introduction to the use of non-crystallographic symmetry in phasing. In: S. Fortier (ed.), *Direct Methods for*

Solving Macromolecular Structures. Dordrecht, Netherlands: Kluwer, pp. 99–108.

Chapman, M. S. and Liljas, L. 2003. Structural folds of viral proteins. *Advances in Protein Chemistry* **64**: 125–96.

Chapman, M. S. and Rossmann, M. G. 1993. Structure, sequence and function correlations among parvoviruses. *Virology* **194**: 491–508.

Chapman, M. S. and Rossmann, M. G. 1995. Single-stranded DNA-protein interactions in canine parvovirus. *Structure* **3**: 151–62.

Chapman, M. S. and Rossmann, M. G. 1996. Structural refinement of the DNA-containing capsid of canine parvovirus using *RSRef*, a resolution-dependent stereochemically restrained real-space refinement method. *Acta Crystallographica* **D52**: 129–42.

Chapman, M. S., Blanc, E., Johnson, J. E. *et al.* 1998. Use of non-crystallographic symmetry for *ab initio* phasing of virus structures. In: S. Fortier (ed.), *Direct Methods for Solving Macromolecular Structures.* Dordrecht, Netherlands: Kluwer, pp. 433–42.

Chen, Z., Stauffacher, C., Li, Y. *et al.* 1989. Protein-RNA Interactions in an icosahedral virus at 3. 0 Å resolution. *Science* **245**: 154–59.

Chipman, P. R., Agbandje-McKenna, M., Kajigaya, S. *et al.* 1996. Cryo-electron micorscopy studies of empty capsids of human parvovirus B19 complexed with their receptor. *Proceedings of the National Academy of Sciences, USA* **93**: 7502–6.

Clinton, G. M. and Hayashi, M. 1976. The parvovirus MVM: a comparison of heavy and light particle infectivity and their density conversion *in vitro*. *Virology* **74**: 57–63.

Farr, G. A. and Tattersall, P. 2004. A conserved leucine that constricts the pore through the capsid fivefold cylinder plays a central role in parvoviral infection. *Virology* **323**: 243–56.

Gerstein, M., Richards, F., Chapman, M. S. and Connolly, M. 2001. Protein surfaces and volumes: measurement and use. In: M. G. Rossmann and E. Arnold (eds), *International Tables for Crystallography. Crystallography of Biological Molecules.* Dordrecht, Netherlands: Kluwer, pp. 531–45 (Cpt. 522.531).

Girod, A., Ried, M., Wobus, C. *et al.* 1999. Genetic capsid modifications allow efficient re-targeting of adeno-associated virus type 2. *Natural Medicine* **5**: 1052–6.

Govindasamy, L., Hueffer, K., Parrish, C. R. and Agbandje-McKenna, M. 2003. Structures of host range-controlling regions of the capsids of canine and feline parvoviruses and mutants. *Journal of Virology* **77**: 12211–21.

Gribskov, M., Luthy, R. and Eisenberg, D. 1990. Profile analysis. *Methods in Enzymology* **183**: 146–59.

Harrison, S. C., Olson, A., Schutt, C. E., Winkler, F. K. and Bricogne, G. 1978. Tomato bushy stunt virus at 2. 9 Å resolution. *Nature* **276**: 368–73.

Harrison, S. C., Skehel, J. J. and Wiley, D. C. 1996. Virus structure. In: B. N. Fields, D. M. Knipe, P. M. Howley *et al.* (eds), *Fundamental Virology.* Philadelphia: Lippincott-Raven Press, pp. 59–99.

Hickman, A., Ronning, D., Kotin, R. and Dyda, F. 2002. Structural unity among viral origin binding proteins. Crystal structure of the nuclease domain of adeno-associated virus Rep. *Molecular Cell* **10**: 327.

Hueffer, K., Parker, J. S., Weichert, W. S., Geisel, R. E., Sgro, J. Y. and Parrish, C. R. 2003. The natural host range shift and subsequent evolution of canine parvovirus resulted from virus-specific binding to the canine transferrin receptor. *Journal of Virology* **77**: 1718–26.

Hueffer, K. and Parrish, C. R. 2003. Parvovirus host range, cell tropism and evolution. *Current Opinion in Microbiology* **6**: 392–8.

Hutchinson, E. G. and Thornton, J. M. 1996. PROMOTIF – a program to identify and analyze structural motifs in proteins. *Protein Science* **5**: 212–20.

James, J. A., Escalante, C. R., Yoon-Robarts, M., Edwards, T. A., Linden, R. M. and Aggarwal, A. K. 2003. Crystal structure of the SF3 helicase from adeno-associated virus type 2. *Structure Camb* **11**: 1025–35.

James, J. A., Aggarwal, A. K., Linden, R. M. and Escalante, C. R. 2004. Structure of adeno-associated virus type 2 Rep40-ADP complex: insight into nucleotide recognition and catalysis by superfamily 3 helicases. *Proceedings of National Academy of Sciences USA* **101**: 12455–60.

Johnson, F. B. 1984. Parvovirus proteins. In K. I. Berns (ed.), *The Parvovirses.* New York: Plenum, p. 259.

Kajigaya, S., Fujii, H., Field, A. *et al.* 1991. Self-assembled B19 parvovirus capsids, produced in a baculovirus expression system, are antigenically and immunogenically similar to native virions. *Proceedings of the National Academy of Sciences, USA* **88**: 4646–50.

Kaufmann, B., Simpson, A. A. and Rossmann, M. G. 2004. The structure of human parvovirus B19. *Proceedings of National Academy of Sciences USA* **101**: 11628–33.

Kern, A., Schmidt, K., Leder, C. *et al.* 2003. Identification of a heparin-binding motif on adeno-associated virus type 2 capsids. *Journal of Virology* **77**: 11072–81.

Kronenberg, S., Kleinschmidt, J. A. and Bottcher, B. 2001. Electron cryo-microscopy and image reconstruction of adeno-associated virus type 2 empty capsids. *EMBO Report* **2**: 997–1002.

Kyte, J. 1995. *Structure in Protein Chemistry.* New York: Garland.

Li, Q., Yafal, A. G., Lee, Y. M., Hogle, J. and Chow, M. 1994. Poliovirus neutralization by antibodies to internal epitopes of VP4 and VP1 results from reversible exposure of these sequences at physiological temperature. *Journal of Virology* **68**: 3965–70.

Liu, Y. and Eisenberg, D. 2002. 3D domain swapping: as domains continue to swap. *Protein Science* **11**: 1285–99.

Llamas-Saiz, A. L., Agbandje-McKenna, M., Parker, J. S., Wahid, A. T., Parrish, C. R. and Rossmann, M. G. 1996. Structural analysis of a mutation in canine parvovirus which controls antigenicity and host range. *Virology* **225**: 65–71.

López-Bueno, A., Bryant, N., Kontou, M. *et al.* 2005. Genetic and structural analysis of the parvovirus MVM capsid binding site to the sialic component of a primary pathogenic receptor. *Journal of Virology* (in preparation).

Luo, M., Tsao, J., Rossmann, M. G., Bassak, S. and Compans, R. W. 1988. Preliminary X-ray crystallographic analysis of canine parvovirus crystals. *Journal of Molecular Biology* **200**: 209–11.

McKenna, R., Xia, D., Willingham, P. *et al.* 1992. Atomic structure of single-stranded DNA bacteriophage fX 174 and its functional implications. *Nature* **355**: 137–43.

McKenna, R., Bowman, B. R., Ilag, L. L., Rossmann, M. G. and Fane, B. A. 1996. Atomic structure of the degraded procapsid particle of the bacteriophage G4: induced structural changes in the presence of calcium ions and functional implications. *Journal of Molecular Biology* **256**: 736–50.

McKenna, R., Olson, N. H., Chipman, P. R. *et al.* 1999. Three-dimensional structure of Aleutian mink disease parvovirus: implications for disease pathogenicity. *Journal of Virology* **73**: 6882–91.

McPherson, A. 2002. *Introduction to Macromolecular Crystallography.* Hoboken, NJ: Wiley-Liss.

Mulloy, B. and Linhardt, R. J. 2001. Order out of complexity – protein structures that interact with heparin. *Current Opinion in Structural Biology* **11**: 623–8.

Nicholls, A. 1992. GRASP: Graphical representation and analysis of surface properties. New York: Columbia University.

Opie, S. R., Warrington Jr, K. H., Jr., Agbandje-McKenna, M., Zolotukhin, S. and Muzyczka, N. 2003. Identification of amino acid residues in the capsid proteins of adeno-associated virus type 2 that contribute to heparan sulfate proteoglycan binding. *Journal of Virology* **77**: 6995–7006.

Parrish, C. R., Aquadro, C. F. and Carmichael, L. E. 1988a. Canine host range and a specific epitope map along with variant sequences in the capsid protein gene of canine parvovirus and related feline, mink and raccoon parvoviruses. *Virology* **166**: 293–307.

Parrish, C. R., Burtonboy, G. and Carmichael, L. E. 1988b. Characterization of a nonhemaggulinating mutant of canine parvovirus. *Virology* **163**: 230–2.

Rabinowitz, J. E., Xiao, W. and Samulski, R. J. 1999. Insertional mutagenesis of AAV2 capsid and the production of recombinant virus. *Virology* **265**: 274–85.

Reguera, J., Carreira, A., Riolobos, L., Almendral, J. M. and Mateu, M. G. 2004. Role of interfacial amino acid residues in assembly, stability and conformation of a spherical virus capsid. *Proceedings of National Academy of Sciences USA* **101**: 2724–9.

Rimmelzwaan, G. F., Poelen, M. C. M., Meloen, R. H., Carlson, J., UytdeHaag, F. G. C. M. and Osterhaus, A. D. M. E. 1990. Delineation of canine parvovirus T cell epitopes with peripheral blood mononuclear cells and T cell clones from immunized dogs. *Journal of General Virology* **71**: 2321–9.

Roberts, M. M., White, J. L., Grutter, M. G. and Burnett, R. M. 1986. Three-dimensional structure of the adenovirus major coat protein hexon. *Science* **232**: 1148–51.

Rose, J. A., Maizel Jr., J. V., Inman, J. K. and Shatkin, A. J. 1971. Structural protein of adenovirus-associated viruses. *Journal of Virology* **8**: 766–70.

Rosenfeld, S. J., Yoshimoto, K., Kajigaya, S. *et al.* 1992. Unique region of the minor capsid protein of human parvovirus B19 is exposed on the virion surface. *Journal of Clinical Investigation* **89**: 2023–9.

Rossmann, M. G. 1972. *The Molecular Replacement Method.* New York: Gordon and Breach.

Rossmann, M. G. 1995. *Ab initio* phase determination and phase extension using non-crystallographic symmetry. *Current Opinion in Structural Biology* **5**: 650–5.

Rossmann, M. G. and Johnson, J. E. 1990. Icosahedral RNA virus structure. *Annual Review of Biochemistry* **58**: 533–73.

Rutledge, E., Halbert, C. and Russell, D. 1998. Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2. *Journal of Virology* **72**: 309–19.

Sato, H., Hirata, J., Kuroda, N., Shiraki, H., Maeda, Y. and Okochi, K. 1991. Identification and mapping of neutralizing epitopes of human parvovirus B19 by using human antibodies. *Journal of Virology* **65**: 5485–90.

Seisenberger, G., Ried, M. U., Endress, T., Buning, H., Hallek, M. and Brauchle, C. 2001. Real-time single-molecule imaging of the infection pathway of an adeno-associated virus. *Science* **294**: 1929–32.

Simpson, A. A., Chipman, P. R., Baker, T. S., Tijssen, P. and Rossmann, M. G. 1998. The structure of an insect parvovirus (*Galleria mellonella* densovirus) at 3. 7 Å resolution. *Structure* **6**: 1355–67.

Simpson, A. A., Chandrasekar, V., Hebert, B., Sullivan, G. M., Rossmann, M. G. and Parrish, C. R. 2000. Host range and variability of calcium binding by surface loops in the capsids of canine and feline parvoviruses. *Journal of Molecular Biology* **300**: 597–610.

Simpson, A. A., Hebert, B., Sullivan, G. M. *et al.* 2002. The structure of porcine parvovirus: comparison with related viruses. *Journal of Molecular Biology* **315**: 1189–98.

Smith, J. L. 1991. Determination of three-dimensional structure by multiwavelength anomalous diffraction. *Current Opinion in Structural Biology* **1**: 1002–11.

Summerford, C. and Samulski, R. J. 1998. Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. *Journal of Virology* **72**: 1438–45.

Tattersall, P., Shatkin, A. J. and Ward, D. C. 1977. Sequence homology between the structural polypeptides of minute virus of mice. *Journal of Molecular Biology* **111**: 375–94.

Tsao, J., Chapman, M. S., Agbandje, M. *et al.* 1991. The three-dimensional structure of canine parvovirus and its functional implications. *Science* **251**: 1456–64.

Tsao, J., Chapman, M. S., Wu, H., Agbandje, M., Keller, W. and Rossmann, M. G. 1992. Structure determination of monoclinic canine parvovirus. *Acta Crystallographica* **B48**: 75–88.

Vasudevacharya, J. and Compans, R. W. 1992. The NS and capsid genes determine the host range of porcine parvovirus. *Virology* **187**: 515–24.

Vihinen-Ranta, M., Kalela, A., Makinen, P., Kakkola, L., Marjomaki, V. and Vuento, M. 1998. Intracellular route of canine parvovirus entry. *Journal of Virology* **72**: 802–6.

Wu, H. and Rossmann, M. G. 1993. The canine parvovirus empty capsid structure. *Journal of Molecular Biology* **233**: 231–44.

Wu, P., Xiao, W., Conlon, T. *et al.* 2000. Mutational analysis of the adeno-associated virus type 2 (AAV2) capsid gene and construction of AAV2 vectors with altered tropism. *Journal of Virology* **74**: 8635–47.

Xie, Q. and Chapman, M. S. 1996. Canine parvovirus capsid structure, analyzed at 2.9 Å resolution. *Journal of Molecular Biology* **264**: 497–520.

Xie, Q., Bu, W., Bhatia, S., Hare, J., Somasundaram, T., Azzi, A. and Chapman, M. S. 2002. The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. *Proceedings of National Academy of Sciences USA* **99**: 10405–10.

Xie, Q., Hare, J., Bu, W., Jackson, W., Turnigan, J. and Chapman, M. S. 2004. Large-scale preparation, purification and crystallization of wild-type adeno-associated virus 2. *Journal of Virological Methods* **122**: 17–27.

Yoon-Robarts, M., Blouin, A. G., Bleker, S. *et al.* 2004. Residues within the B' motif are critical for DNA binding by the SF3 helicase Rep40 of adeno-associated virus type 2. *Journal of Biological Chemistry* **279**: 50472–81.