Single-stranded DNA–protein interactions in canine parvovirus

Michael S Chapman[1,2] and Michael G Rossmann[1]*

[1]Department of Biological Sciences, Purdue University, West Lafayette, IN 47907-1392, USA and [2]Department of Chemistry & Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306-3006, USA

Background: Parvoviruses are small icosahedral single-stranded (ss) DNA viruses which replicate in rapidly proliferating cells, causing a variety of serious and often lethal diseases in mammals, including humans. The structure of canine parvovirus (CPV) showed an 11-nucleotide oligomeric fragment of its genome bound to 60 equivalent binding sites on the inside surface of the capsid. This provides an opportunity to study the conformation of ssDNA, its interactions with protein, and its role in viral assembly.

Results: The icosahedrally ordered part of CPV ssDNA has an unusual loop conformation with the bases pointing outwards and the phosphates surrounding metal ions on the inside. The protein interacts with the bases, making 15 putative hydrogen bonds. The DNA electron density indicates preferences for particular base types in parts of the binding site. Statistical analysis of the genome yields ~30 regions with sequences similar to that observed in the structure, demonstrating a low level of sequence specificity for binding to capsid protein.

Conclusions: ssDNA can adopt unusual conformations upon association with protein by using phosphoribose backbone rotamers that are found in tRNA, but not in DNA duplexes. The CPV DNA–protein interactions differ from the non-specific backbone interactions seen in some plant and insect viruses. The sequence specificity, albeit low level, of the protein for CPV DNA may contribute both to distinguishing the viral DNA from other nucleic acids and to the DNA packaging process during viral assembly.

Structure 15 February 1995, **3**:151–162

Key words: canine parvovirus, DNA–protein interactions, single-stranded DNA, virus structure, virus assembly

Introduction

The structure determination of canine parvovirus (CPV) [1] showed 11-nucleotide fragments of the single-stranded (ss) genome bound at equivalent positions on the inside surface of each copy of the capsid protein. The conformation of the ssDNA, and its interactions with the protein, are relevant to other cellular processes involving protein–nucleic acid interactions of low sequence specificity, such as replication and transcription.

Many studies have been conducted in the attempt to elucidate the conformational restraints on DNA (reviewed in [2]), and on DNA in complex with protein (reviewed in [3]). Almost all of these have focused on double-stranded (ds) DNA. Structural information on ssDNA comes from complexes of protein with short segments of ssDNA or ssRNA, such as pancreatic ribonuclease [4], the Klenow fragment [5], and tRNA synthetases [6,7].

Viral nucleic acid was first seen as repeating sets of three ribonucleotides in the structure of the helical tobacco mosaic virus (reviewed in [8]). Earlier studies of spherical viruses did not show any icosahedrally ordered nucleic acid, although the interactions of double-helical B-RNA with southern bean mosaic virus capsid protein had been proposed [9]. More recently, seven or eight well ordered bases per capsid protein have been seen in two plant viruses, bean pod mottle virus (BPMV) [10] and satellite tobacco mosaic virus (STMV) [11,12], and in an insect virus, flock house virus (FHV) [13]. In BPMV the RNA has the conformation of a single strand of an A-like DNA structure, whereas in STMV and FHV the RNA was found to have a double-stranded A-like conformation. Besides CPV, four nucleotides of ssDNA were found in the structure of the bacteriophage ϕX174 [14,15]. In each of these examples, the relatively few interactions with the capsid protein were predominantly hydrophilic and involved the phosphate backbone. On the other hand, the interaction of a short fragment of viral RNA with the capsid of the bacteriophage MS2 shows specific base–protein interactions [16].

Parvoviruses are small T=1 icosahedral viruses which contain ~5 kilobases (kb) of ssDNA [17]. CPV packages positive-sense DNA, but some other parvoviruses package either positive-sense or negative-sense DNA [17]. The virus carries only two genes, but different mRNA splicings lead to several variants of both the capsid protein and a non-structural protein that has several functions in replication.

Parvoviruses may be split into two families: the autonomous parvoviruses (e.g. CPV), which infect only the replicating tissues of its host; and the adeno-associated viruses (AAVs) which are dependent for replication upon co-infection with a helper adenovirus or herpes virus [17]. The non-pathogenic AAV, whose capsid amino acid sequence is 20% identical to that of CPV [18], has

*Corresponding author.

 ISSN 0969-2126

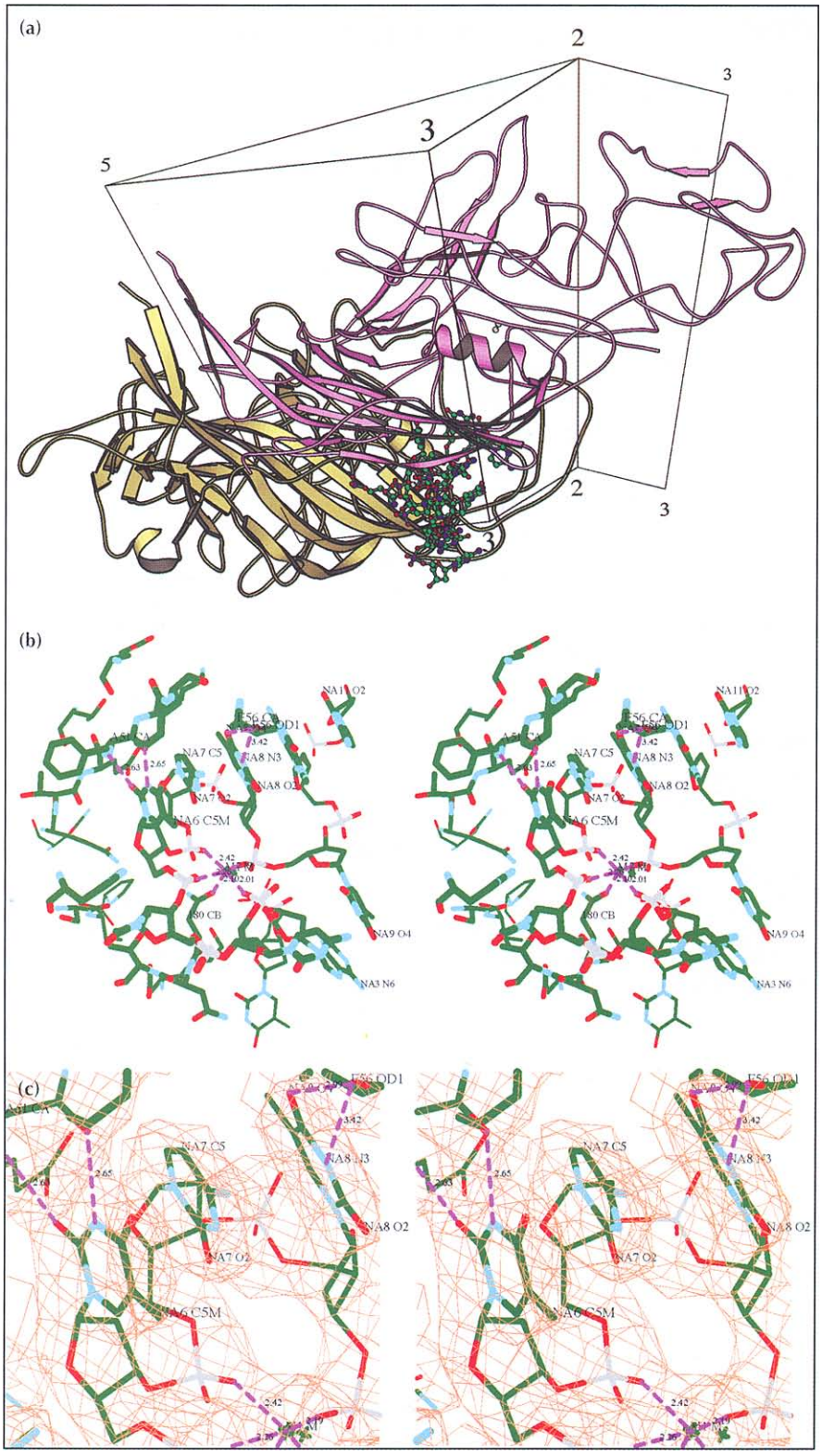

Fig. 1. **(a)** Diagramatic view showing the structure of the ssDNA bound to the inside of the CPV capsid. The DNA (ball-and-stick model) is shown nestled between two of the 60 subunits (pink and yellow ribbons) that comprise the capsid. The viewing direction is tangential to the roughly spherical capsid shell, with the inside surface at the bottom. The symmetry operators (five-, three-, three- and two-fold) that surround the pink subunit radiate from the center of the virus and are drawn between 85 Å and 145 Å from the center. A complete icosahedral capsid can be generated from all atoms within the volume that is outlined. The yellow subunit (behind) is related to the pink subunit by the five-fold axis shown to the left. These two subunits account for most of the contacts with the DNA. A third subunit also makes contact, but is omitted for clarity. It is related to the pink subunit by the two-fold axis and would contact the DNA on the right-hand side. **(b)** A stereoscopic close-up of the DNA. Green designates carbon, blue nitrogen, red oxygen, gray phosphorus and some of the putative hydrogen bonds and ionic interactions are added in magenta. Nucleotides (NA) 2 (bottom rear) to 11 (top right) are shown looping clockwise around the central metal ion (M1). The latter is liganded by phosphates 4, 6, 7 and 9 and Asn180. Bases 6, 7, 8, 10 and 11 are stacked left to right. T9 loops out to join a smaller stack, part of which is shown (bottom right). For clarity only a few of the surrounding amino acids have been drawn to illustrate typical interactions such as those between T6 and the main chain of amino acid 51. The view is away from the center of the capsid protein which is slightly to the left. **(c)** Part of the view above, magnified and superimposed on the averaged electron density. At base 6, the bulge at atom C5M is distinctive for T; however, the hydrogen bonding does not distinguish between T and C. The electron density bulges at the 2 and 4 ring positions of base 7 suggest a *syn* conformation, and the lack of a bulge at the 5 position suggests C. In all cases the density for the bases is quite planar. The density for the phosphates is nearly spherical, but, as can be seen for phosphates 7 and 8, there are slight bulges for O1P and O2P. (Figure prepared using the program O [36].)

potential as a transducing vector for human gene therapy [19], because of its ability to package foreign DNA efficiently. Capsid–DNA interactions may be of importance for such AAV vectors, although wild-type DNA appears to have no *cis*-acting packaging signal [19].

The only places where DNA is likely to be seen in a crystallographic structure determination of a virus are those where it has a similar conformation in the majority of the 60 icosahedral positions. This is most likely to be the case where the DNA binds to the capsid protein. Elsewhere, the electron density will be an uninterpretable average of 60 different nucleic acid conformations. Where ordered DNA is observed, variation of the base sequence might blur the electron density of the bases.

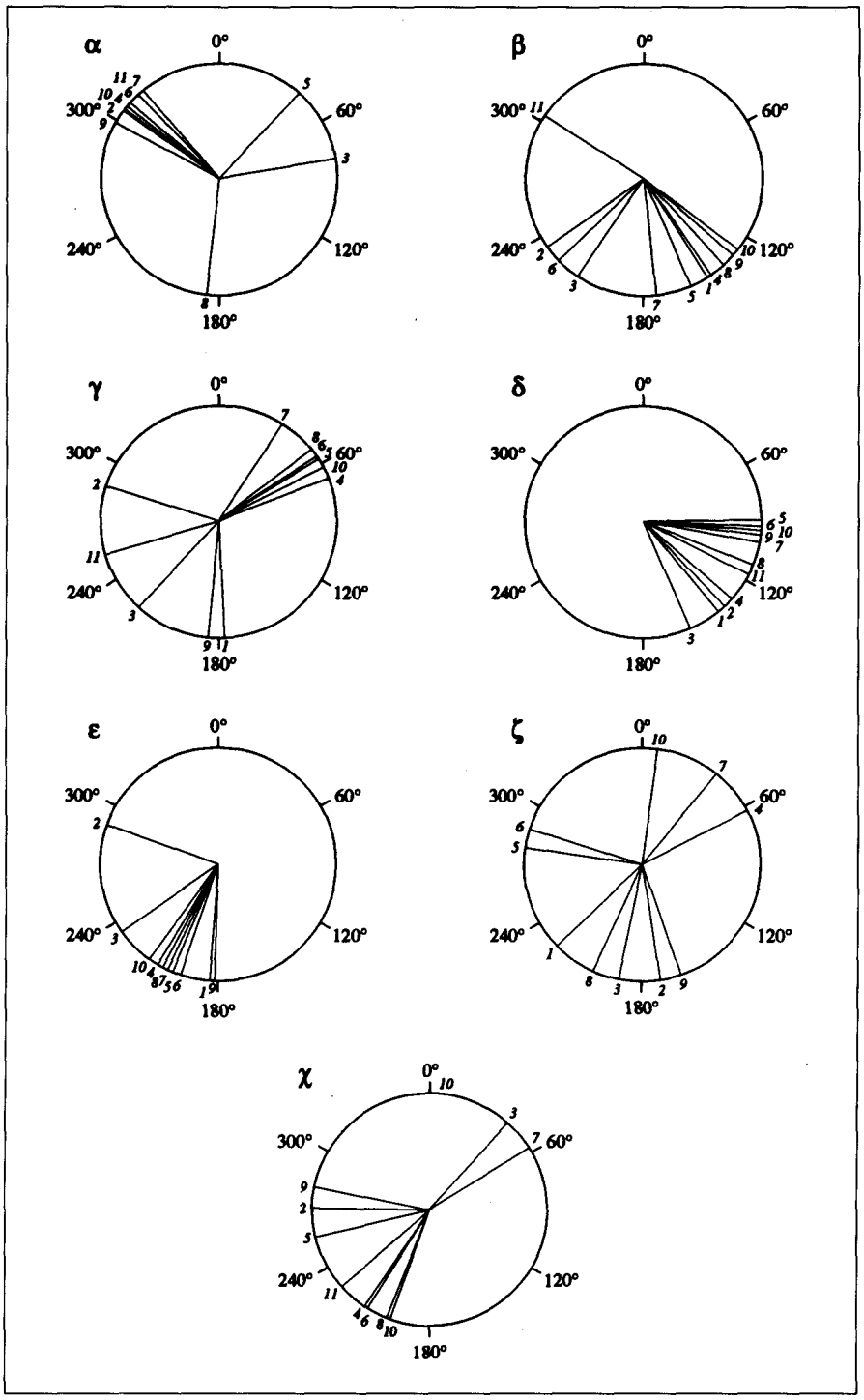

Fig. 2. CPV DNA torsion angles. The conformational angles of each of the nucleotides are shown as radii in the conformational wheels. Each of the radii is numbered at the circumference with the appropriate nucleotide number. Glycosidic linkage: χ of nucleotides 3 and 7 are *syn*, but otherwise the conformation is usually close to *anti*. Pyrimidines 2 and 5 have higher than usual χ values, with pyrimidine 9 extending slightly into the high-*anti* conformation. Similar values have been found in nucleoside crystals and are within 1 kcal mol^{-1} of the optimal value [54]. P-$O_{5'}$ ester: α is tightly distributed around the three staggered angles. A strong preference is observed for –60°, where an oxygen lone pair is *anti* to the P-$O_{3'}$ bond (the *gauche* effect [2]). Other rotamers are found occasionally. The distribution about each rotamer is tighter than that found in tRNA [22]. $O_{5'}$-$C_{5'}$ ester: β has a very broad distribution about *anti*, as found in tRNA [22]. β of C11 is an exception at –60°, but is also less reliable as the backbone could not be built beyond C11. $C_{5'}$-$C_{4'}$ bond: with a couple of exceptions, γ is more tightly distributed about the three staggered rotamers than tRNA [22]. C11 is an exception once again, as is A3, which is near the largest protein conformational changes between full and empty particles (see text and Table 1). Also, Phe266 is inserted between the bases of T2 and A3, stacking with A3. To accommodate this, the phosphate backbone between the second and third deoxyriboses is distorted, with T2 ϵ and A3 γ absorbing most of the strain. Otherwise, as in tRNA, the rotamer at 60° is strongly favored. $C_{4'}$-$C_{3'}$ bond: δ shows a very similar distribution to that found in tRNA. $C_{3'}$-$O_{3'}$ bond: with the exception of T2 (discussed above) ϵ has a unimodal distribution about –150° as in tRNA. $O_{3'}$-P ester: the unusual conformation of the CPV DNA is reflected most in ζ. Either –60° or 180° satisfies the *gauche* effect, but in tRNA there is a strong bias for the –60° rotamer, which is found in helices. In CPV, all three rotamers are found, with –60° being the least populated. Furthermore, deviations by as much as 45° from the ideal values for the three staggered rotamers are observed.

In this paper, DNA conformation and interactions of DNA with the capsid are analyzed in detail, based on the refined structure of CPV (MS Chapman and MG Rossmann, unpublished data; the coordinates have been submitted to the Brookhaven Protein Data Bank). It is shown that the electron density of some of the bases is quite well defined, suggesting a sequence preference in the capsid–DNA interactions. The observed sequence preference is found to be consistent with sequences in the viral genomic DNA and, hence, may account for recognition of viral DNA by the capsid.

Results and discussion

DNA conformation

The DNA model was built with the base types that fitted the experimental electron density best and had the most favorable interactions with protein, with no regard to the actual sequence of the genome. The overall structure of the ssDNA is unusual, because the bases point outwards to interact with the amino acids of the capsid binding site, and the phosphoribose backbone is on the inside of a loop with the phosphates brought together by counterions (Fig. 1). Very few inter-nucleotide interactions that

Table 1. Conformational changes induced by DNA binding onto the protein capsid.

Region	Magnitude (in Å) of largest conformational changes within each region (Å)		Comments
51–57	Leu54 Cδ2	3.0	The main chain of residue 51 moves to H-bond with T6′. The side chain of Asn56
	Leu54 O	1.5	moves 0.6 Å for its Nδ2 to H-bond to G10 O6 and its Oδ2 to H-bond to T8′ O4.
179–180	Asn180 Oδ1	1.6	The Cα of Ser179 is moved away from close contact (3.1 Å) with T2 O2.
	Asn180 Cα	0.8	The side chain of Asn180 moves to interact with T9 O1P and C4 O2P.
266–267	Phe266 Cε1	1.2	Phe266 is moved from a close contact with A3: C5 (2.9 Å) and N7 (2.8 Å).
269–270	Asp269 Cδ1	4.7	This region is pushed by Asn492, but does not have a direct interaction with the DNA.
	Cys270 Cβ	1.3	
491–496	Gln491 Nε2	6.9	Asn493 moves to relieve a close contact (0.6 Å) between its Oδ1 and A3 N7.
	Asn493 Nδ2	7.0	Asn492 moves to relieve the contact (2.6 Å) between its main chain and C4.
	Gln491 O	2.1	Movement of Asn492 forces Glu491 to adopt a different rotamer.
	Asn492 O	3.0	The movement of Asn493 is propagated to Pro495. Glu496 moves
	Asn493 Cα	2.0	slightly to optimize the contact between its Cα and T2 O2.
	Cys494 N	1.6	
582	Lys582 Nζ	2.5	Moves to H-bond with C5′ O4′.

A prime following a nucleotide number indicates that the nucleotide is related by an icosahedral symmetry operation to the reference subunit.

Table 2. Secondary structures involved in DNA interaction.

Amino acid	Location
Asn51	β-Strand A
Asn56	Turn following β-strand A
Gln143	β-Strand D
Tyr244	N terminus of α-helix B
Phe266 and Phe267	Bulge on β-strand G
Gly496	Turn preceding β-strand H

might stabilize the unusual conformation are observed. One putative divalent cation, M1 (see below), brings together the phosphates of nucleotides 2, 6, 7 and 9 on the inside of the loop. Cation M2 ties the ends of the loop together, bridging between A3 and T9. No base-pairing is observed, the predominant determinant of the DNA conformation being the interactions with the capsid protein.

The initial model was built using deoxyribose ring puckers of either $C_{3'}$-*endo* (nucleotides 5–7) or $C_{2'}$-*endo* (nucleotides 1–4 and 8–11) to optimize the fit of bases to their electron densities. The pseudo-rotational angle, P [20], that defines the sugar pucker was not restrained in refinement, and the individual torsion angles around the ring were only loosely restrained. During refinement, T9 switched from $C_{2'}$-*endo* to $C_{3'}$-*endo*, and T8 and G10 changed to intermediate puckers of $C_{1'}$-*exo* and $O_{4'}$-*endo* respectively. Such intermediate puckers have been seen occasionally in ribonucleoside and deoxyribonucleoside structures [21]. With different base types superimposed at each position, the electron density may show the average of different ribose puckers. Surveys of known DNA structures (mostly double-helical) show a strong preference for $C_{2'}$-*endo* rings over $C_{3'}$-*endo* deoxyribose [21]. This preference is not seen in CPV, either because the structure models the average of superimposed nucleotides with distinct puckers, or possibly because, in the presence of protein, the distribution of puckers for single-stranded deoxyriboses resembles that usually found for riboses.

The root mean square (rms) deviation from ideal rotamers is 24° for the CPV DNA torsion angles. Helical DNA structures generally have a tighter distribution of torsion angles than extended nucleotide structures, presumably because of their inherent repeating structure. Thus, one would expect the torsion-angle distributions for CPV DNA to be broader than for structures containing mostly helical DNA. The CPV distributions (Fig. 2) are a little broader than those for tRNA, approximately two-thirds of which is helical [22]. Several staggered rotamers are observed for each torsion angle. Helical DNA structures, but not RNA structures, show strong bias in the selection between these rotamers [2], with a more equal sampling of the alternative rotamers being observed in CPV DNA than in RNA structures (Fig. 2). In summary, the DNA is able to adopt an unusual conformation through the use of all staggered rotamers, and through slight deviations from ideal torsion angles, especially in ζ.

Protein conformational changes

The atomic coordinates of full and empty [23] particles have an rms difference for all atoms of 0.78 Å. Calculation of the rms positional coordinates error according to Luzzati [24], using free R-factors, suggests that much of the difference can be attributed to experimental error. However, the larger differences in protein structure between empty and full particles occur near the

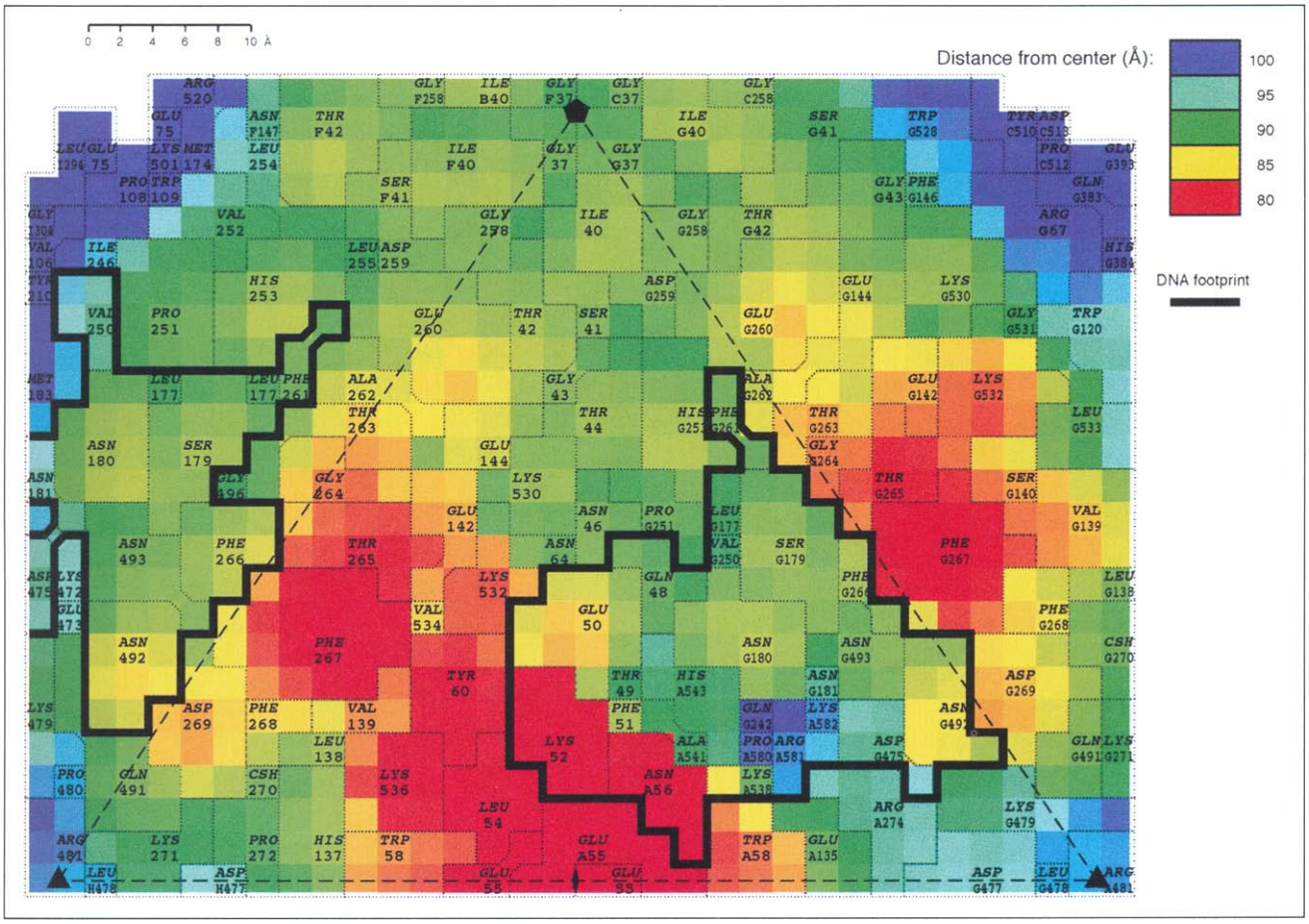

Fig. 3. The projected inner surface of the capsid is viewed from the DNA in the center of the virus. The triangular unit is one of 60 equivalent units that comprise the capsid. As the capsid protein is not triangular, parts of several symmetry equivalents are shown. In particular, residue numbers preceded by G denote a five-fold related protein. The two-, three- and five-fold axes are shown as an ellipse, triangles and a pentagon on the perimeter of the triangular unit. The inner surface is projected outwards along a two-fold axis and is colored according to distance from the viral center. All residues that come within 3.8 Å of a DNA atom are enclosed in a bold outline (the DNA 'footprint'). Part of a footprint in a five-fold related icosahedral asymmetric unit can also be seen. (Figure prepared using the program RoadMap [55].)

DNA-binding site (rms difference of 1.9Å) indicating significant conformational differences (Table 1). The largest changes are centered about residue 493, which is pushed away from nucleotide 3 (A3). Some neighboring residues (269–270) are affected, but in spite of main-chain displacements of up to 3 Å the conformational changes do not propagate far down the peptide backbone. In summary, the conformational changes are modest and are localized near the DNA.

DNA–protein interactions

The amino acids that interact with the DNA are distributed throughout the primary sequence of the protein, and are not associated with any known DNA-binding structural motif. Many of these amino acids are associated with secondary structural elements (Table 2) of the jelly-roll β-barrel that is the core of many viral capsid structures (reviewed in [25]). The DNA-binding site lies in a trough on the inner surface between two five-fold related capsid proteins (Fig. 3). This location may be functionally significant, as the binding site is complete only in the assembled particle. Of the amino acid side chains that hydrogen bond with the DNA, five (56, 143, 180, 475 and 493) are highly, but not completely, conserved among parvoviruses [18], but three (244, 543 and 582) are not. The inner surface of the capsid is more conserved than the outer surface, possibly because polar residues are required to interact with the DNA, or because the inner surface does not experience the immune pressure to mutate that may be present on the outer surface [18]. The difference between the conservation of amino acids that contact the DNA and that of other residues that line the inner surface of the capsid is insignificant [18].

The most striking feature of the DNA structure in the CPV capsid is the complete lack of interactions between phosphates and basic amino acids (Table 3) that are usually observed in protein–nucleic acid complexes. Instead, the phosphates interact with putative metal ions. The first metal ion site (M1) is coordinated octahedrally, with four equatorial phosphate ligands (nucleotides 4, 6, 7 and 9) and one protein ligand (Asn180 amide oxygen) (Fig. 1). Without any restraint, the metal to phosphate-oxygen distances refined to 2.2±0.15 Å, intermediate between the

Table 3. Nucleic acid–protein interactions.

Nucleotide	Nucleotide atom(s)	Interaction	Protein/ nucleotide atom	Distance (Å)
X1	Phosphoribose			
	backbone	SA		
	Base	Not modeled		
T2	Phosphoribose			
	backbone	SA		
	O2	Acceptor	496N	2.7
	N3	Donor	496O	3.2
	O4	Acceptor	Gln143 Nε3	3.1
A3	O2P	Ionic	M2 Me^{2+}	2.6
	O3′	Acceptor	Asn493†	
	N6	Donor	267O	2.7
	O6 (if guanine)	Acceptor	267N	2.4
	O1P, O4′, N7, N1	SA		
	O5′, N3	Buried		
	Base	Stacking	Phe266	3.6
	Base	Stacking	T9	3.8
C4	O2P	Ionic	M1 Me^{2+}	2.0
	O4′	Acceptor	Asn493†	
	O1P, O5′, O3′, O2, N3, N4	SA		
C5	O4′	Acceptor	Lys582* N2	2.8
	O2	Acceptor	Lys582* N3	3.4
	O2	Acceptor	Tyr244 OH	3.4
	N4	Donor	580* O	3.4
	O1P, O2P, O5′, N3	SA		
	O3′	Buried		
T6	O2P	Ionic	M1 Me^{2+}	2.3
	O3′	Acceptor	His543†	
	O2	Acceptor	51*N	2.5
	N3	Donor	51*O	2.7
	O1P, O4′, O3′, O4	SA		
	O5′	Buried		
	Base	Stacking	C7	3.5
C7	O1P	Ionic	M1 Me^{2+}	2.4
	O2P	Acceptor	Asn180 Nδ2	2.7
	O3′, O2, N3, N4	SA		
	O3′, O4′	Buried		
	Base	Stacking	T6/T8	3.5/3.6
T8	N3	Donor	Asn56* Nδ1	3.4
	O4	Acceptor	Asn56* Oδ2	3.4
	O1P, O2P, O5′, O3′, O2	SA		
	Base	Stacking	C7/G10	3.6/3.8
T9	O1P	Ionic	M1 Me^{2+}/	2.2
	O5′, O4′, O2, N3, O4	SA	M2 Me^{2+}	
	O3′	Buried		
	Base	Stacking	A3	3.8
G10	O6	Acceptor	Asn56* Nδ2	3.0
	PO_4, O6, N7, N1, N2, N3	SA		
	O4′	Buried		
	Base	Stacking	T8/C11	3.7/3.8
C11	PO_4, O2, N3, N4	SA		
	O4′	Buried		
	Base	Stacking	G10	3.8

The amino acid type is not mentioned where the DNA interacts with the polypeptide main chain. Atoms in the DNA structure that do not make an interaction with the protein were designated as 'accessible' or 'buried' by inspection of the atomic structure. A protein residue following an asterisk (*) denotes an icosahedral symmetry equivalent residue to that given in the PDB. A residue followed by a dagger (†) indicates that the side chain has two conformations; the contact described here is with the less occupied rotamer. Probable nucleotides are indicated by A, C, G or T, with X indicating a completely indeterminate residue. Me^{2+}, metal ion; SA, solvent accessible; Acceptor, H-bond acceptor; Donor, H-bond donor.

expected values of 2.0 Å for Mg^{2+} and 2.3 Å for Ca^{2+} or Na^{+} [26]. When assumed to be Mg^{2+}, the temperature factor refined to the minimum allowed (0.0 $Å^2$), but Ca^{2+} proved to be too electron dense as the temperature factor rose to 50 $Å^2$, more than twice that of the average protein atom and larger than the mean DNA temperature factor. It is possible that equivalent sites may be occupied by a mixture of ions of different atomic number.

The second putative metal ion site (M2, which is also likely to be Mg^{2+}, Ca^{2+} or a mixture thereof) stabilizes an unusual conformation of the DNA, bridging between the phosphates of A3 and T9. Although the interactions of phosphates 3, 4, 6, 7 and 9 with metal ions M1 and M2 are the only ones to be modeled explicitly, it is likely that the other phosphates interact with counterions, water or polyamines. Indeed, all of the other electronegative O1P and O2P atoms are either accessible to solvent or adjacent to weak unmodeled electron density, namely between O2P of C7 and Asn180 Nδ2. The protein makes only one direct interaction with any of the O1P or O2P atoms. The O3′ and O5′ atoms of each base are less polar than the phosphate oxygens and the only direct interactions appear to involve A3 O3′ and T6 O3′. These O3′ atoms might form hydrogen bonds with lower occupancy alternative conformations of Asn493 or His543′ respectively. Only one direct interaction is observed with a deoxyribose O4′, where there is a hydrogen bond between CO4′ and Lys582′ Nζ. About one-third of the O3′, O4′ and O5′ atoms are in positions inaccessible to water.

Structures of protein–nucleic acid complexes exhibit great variability. In the exonuclease site of *Escherichia coli* polymerase I [5], the bases point away from the protein, permitting few interactions and preventing specific hydrogen-bonding. At the other extreme, the three bases of the anticodon loop of *E. coli* $tRNA^{Gln}$ are inserted into three separate complementary pockets of its synthetase, each involving several hydrogen-bonding interactions [7]. It is between these two extremes of low and high sequence specificity that the CPV binding site is found. Six of the bases, spanning T2 to G10, have some specific interaction with the protein, though none of these is as close as those found in $tRNA^{Gln}$ synthetase. All but one (A3 atom N3) of the base hetero-atoms either interacts with protein or is accessible to water atoms. Of the 15 base–protein interactions that have been identified, half involve protein backbone atoms (three amide nitrogens and four carbonyl oxygens). Most of the base hetero-atoms interact with a protein atom. Amino acid side-chain amide nitrogens (Asn2 and Gln1) figure prominently in the formation of DNA interactions, along with one example each of a lysyl amine, aspartate and tyrosyl oxygen ligands. Three of the bases (2, 3 and 6) interact with the peptide nitrogen and carbonyl oxygen of the same amino acid (496, 267 and 51′ respectively). In the first of these cases, the amino acid mimics an adenine in a Watson–Crick AT base pair. In the last, the amino acid mimics an adenine in the reversed base-pair orientation.

Cross-validation tests using free R-factors [27] showed that explicit water molecules could not be reliably

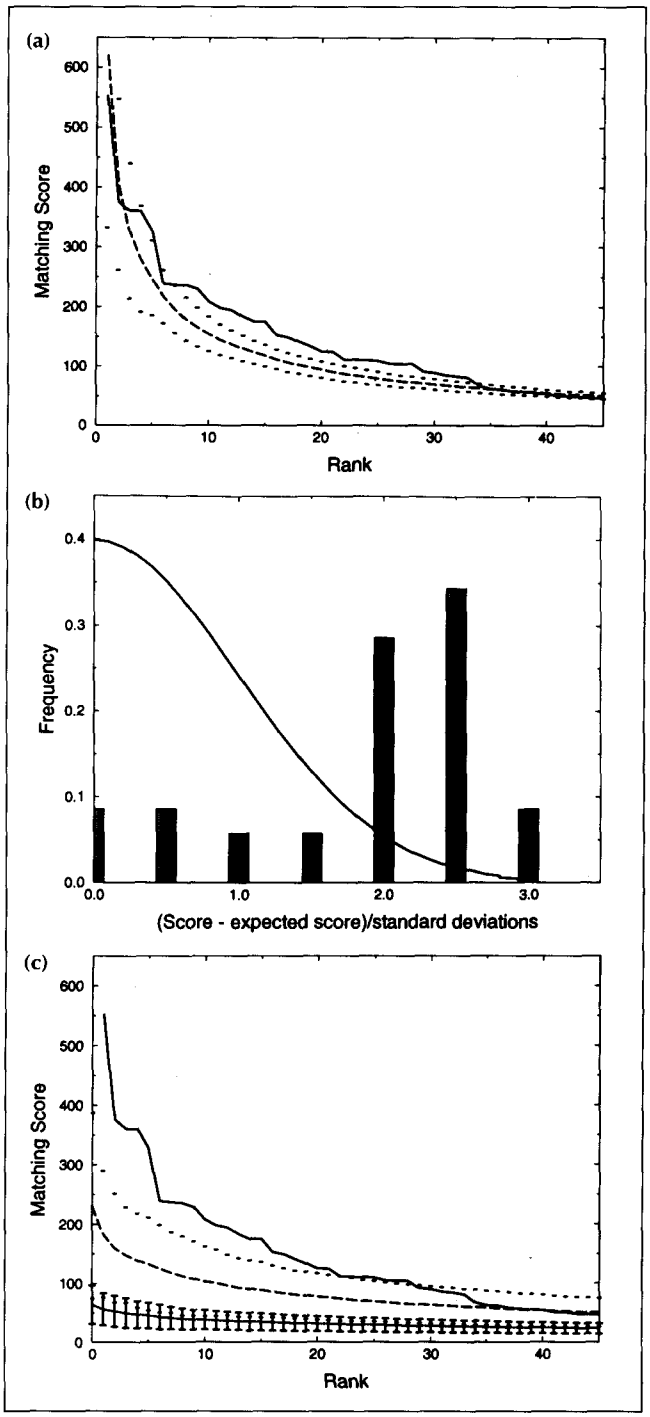

Fig. 4. Sequence matching scores of the crystallographically observed nucleotides compared with the genomic DNA sequence. **(a)** The solid line shows the scores of the best matching fragments of the CPV genome when compared with the observed profile of the DNA-binding site. Long dashes show the expected scores averaged from 100 trials with randomly shuffled sequences. Short dashes are drawn at one standard deviation from the mean of the random matching scores. **(b)** The histogram shows that the scores are significantly higher than expected. The differences between the actual and expected scores for the top 30 matches were normalized by the standard deviations of the expected scores of the same rank. These are plotted as a frequency histogram that shows that most of the differences exceed 2σ. The smooth curve is a Gaussian with a mean of 0.0 and a standard deviation of 1.0. It shows the frequency distribution that would be expected if the differences were random. However, the histogram shows a significant number of large differences. **(c)** A dotted line with standard deviation error bars shows the scores expected from random profiles calculated from 30 randomly selected fragments of the CPV genome. The long-dashed line shows the scores expected from random profiles calculated with just 10 fragments (far fewer than seen in the binding site). The short dashes show single standard deviations about these means.

refined against the 2.9 Å map of CPV (see the Materials and methods section). However, previous high-resolution structural analyses of nucleic acids have shown that most of the hetero-atoms are involved in hydrogen-bond interactions with other nucleic acid atoms or solvent molecules (for example, see [28]). Although the modeling of explicit water molecules is not warranted in the structure of CPV, it is likely to be rare that a hetero-atom is unable to form polar interactions with other atoms. Allowing for experimental error in the positions of atoms, all interactions between nucleic acid hetero-atoms and protein atoms that are within 0.5 Å of ideal separations are listed in Table 3.

DNA-binding specificity

If only a single, highly preferred DNA sequence were able to bind to the capsid protein, it is unlikely that it could be recognized in the electron-density map, as its density would be averaged with 59 other unoccupied sites. If, on the other hand, a number of similar sequences could bind to the capsid protein with lower specificity, analysis of the electron-density map might be informative. The height of the DNA density suggests that at least 30 fragments bind to the interior of the capsid.

A DNA sequence profile was constructed to determine which, if any, regions of the genomic sequence were consistent with the observed DNA bound to the capsid's interior surface (see the Materials and methods section). The profile contained 'probabilities' that the base at each nucleotide position was each of the four possible types (A, G, C or T), estimated according to the shape of the observed electron density and potential interactions with the protein. It was then possible to search the CPV genomic sequence for good matches to the profile. The region of the CPV genome that gives the best alignment has a matching score of 522, which is lower, but not significantly lower, than that expected from random sequences (619±219) (Fig. 4a). Similarly, the second-best score is not as high as the average of second-best scores obtained by aligning the profile with a large number of randomly shuffled CPV sequences. However, from the fourth to the 33rd ranked scores (Figs 4a and 4b), the scores are consistently ~2 standard deviations higher than those of corresponding rank from alignments to randomized sequences. This suggests that ~30 fragments of the CPV genome fit the binding site better than non-CPV sequences. A second way of evaluating the significance of the scores is by comparing them to random profiles instead of randomized sequences. Random profiles can be calculated by selecting, at random, 11-nucleotide fragments from the genome, overlaying the sequences and calculating the frequency of

Table 4. Sequences in the canine parvovirus genome that best match the observed profile.

Rank	Position	Sequence	Rank	Position	Sequence
1	2805–2815	T T A C C T C T C C T	21	3100–3110	G T G T C T C T A A G
2	3324–3334	C T G T C T C T T A T	22	1437–1447	G T A G C T C C T T C
3	633–643	A T A C C T C C A A T	23	1709–1719	A T A T C T C C A T G
4	1023–1033	T T A C C T C C A A T	24	3669–3679	G T A T T T C T T T T
5	3438–3448	T T A C C T T T T T G	25	157–167	G C A T C T C C C A C
6	3132–3142	A T A C C T T C T T G	26	1374–1384	G T A T T T C C C A T
7	744–754	T T A C C T T T C C A	27	1112–1122	A T A T C T T C C T G
8	4308–4318	C T A T C T T C T G C	28	4124–4134	A T A C T T C A T A A
9	4930–4940	A T A C C T T T A A C	29	2502–2512	T T A G C T C T A A A
10	2349–2359	T T G T C T C T T T T	30	4960–4970	C C A C C T T T T C C
11	389–399	C T A T T T C T T A C	31	3873–3883	G T A T T T T T T A A
12	3891–3901	T C A C C T C C T G G	32	2895–2905	A T G G C T C T C A G
13	2421–2431	T T A C T T C T T T T	33	3801–3811	T T A T C T G C T T T
14	251–261	A T G T C T T T T A T	34	1404–1414	T T A T C T T G T T G
15	4921–4931	T T G T C T T T T A T	35	344–354	A T A C A T T T T A T
16	2277–2287	G T A G C T C T T T C	36	3178–3188	G T A C T T T T C C C
17	2494–2504	T T G C C T T T T T A	37	11–21	C T G T C T T T A A G
18	4684–4694	T T A G C T C T T C A	38	1803–1813	C T A T C T A A T G C
19	3429–3439	T T A C T T C C T T T	39	4036–4046	T T G T T T C C T G T
20	2908–2918	C T G C C T C T A T T	40	4589–4599	T T G C T T T T T G G

The position of the sequences refers to the encapsidated positive strand, not the published negative strand [49].

Table 5. Normalized frequencies of nucleotides in the 33 best-matching sequences.

Nucleotide number	Nucleotide frequencies			
	A	G	C	T
1	0.24	0.21	0.15	0.39
2	0.00	0.00	0.09	0.91
3	0.75	0.24	0.00	0.00
4	0.00	0.15	0.42	0.42
5	0.00	0.00	0.79	0.21
6	0.00	0.00	0.00	1.00
7	0.00	0.03	0.64	0.33
8	0.03	0.00	0.36	0.61
9	0.21	0.00	0.18	0.61
10	0.43	0.06	0.12	0.39
11	0.18	0.21	0.21	0.40

base type at each position. Fig. 4c shows that the matching scores from the observed profile are significantly greater than those from random profiles, confirming the earlier analysis. The scores for the random profiles depend on the number of fragments used to construct each random profile. Tests were initially performed using 30 fragments per random profile, but were checked using profiles constructed from just 10 fragments. This is much lower than the 30–60 occupied binding sites suggested by the strength of the electron density. When the profile matching was repeated using the complement of the packaged DNA, the scores were not significantly higher than those for randomly shuffled sequences. This not only provides an additional control to the analysis, but suggests an explanation for the strand specificity of encapsidation in CPV.

The genomic sequences with the highest matching scores are distributed throughout the genome (Table 4) without any significant clustering. The frequencies of base types at each position (Table 5) agree well with the probabilities of the profile that were estimated from the electron density (compare Tables 5 and 6). Differences between the estimated base probabilities for the profile (Table 6) and the actual frequencies of bases in matched sequences (Table 5) are no more than 20%, with the exception of base 6. The latter was estimated to be 60% thymine in the profile, but is invariantly thymine in the best matching sequences. This is consistent with the high electron density at the C5 methyl position of base 6, suggesting that the estimated profile should have had a higher probability for thymine at this position.

Biological implications

The structure of single-stranded (ss) DNA as found in canine parvovirus (CPV) has implications for the incorporation of viral DNA into capsids and is relevant to processes such as replication and transcription, in which the strands of duplex DNA are separated. The recently determined structure of *HhaI* methyltransferase showed that a single nucleotide can 'flip out' of a B-DNA duplex conformation to interact with an enzyme [29]. Similarly, in CPV DNA, the whole loop structure can be considered to be 'flipped' inside out, with bases pointing outwards towards the protein and phosphates surrounding metal ions on the inside. The DNA achieves this conformation through the use of staggered (non-eclipsed) rotamers that are not found in the structures of DNA duplexes, but do occur in tRNA structures.

Table 6. Profile of oligonucleotide base types consistent with the icosahedrally averaged electron density.

Nucleotide number	Base-type probabilities				Weight	Comments
	A	G	C	T		
1	0.25	0.25	0.25	0.25	0.1	Electron density very poor.
2	0.00	0.0	0.2	0.8	1.0	A or G would give steric clashes. Size of electron density is consistent with C or T.
3	0.55	0.3	0.1	0.05	1.3	Lack of electron density for N2 suggests A rather than G.
4	0.1	0.1	0.5	0.3	1.0	Little electron density for C5 methyl of T.
5	0.1	0.1	0.6	0.2	1.3	
6	0.075	0.075	0.25	0.6	1.3	Good electron density shows C5 methyl for T.
7	0.15	0.15	0.45	0.25	1.3	
8	0.15	0.15	0.3	0.4	1.1	
9	0.125	0.125	0.3	0.45	0.8	Electron density is diffuse.
10	0.4	0.3	0.15	0.15	0.6	Poor electron density.
11	0.2	0.2	0.3	0.3	0.2	Size of poor electron density suggests C or T.

Base-type probabilities represent estimates for the probability of finding each base type at a binding site. The weights giving the confidence of the estimates are based on the shape of the electron density, steric conflicts and potential hydrogen bonds.

The number of interactions between the CPV DNA and the capsid protein is fewer than seen in other DNA–protein complexes. Approximately half of the DNA base hetero-atoms are involved in direct interactions with the protein, so there is the potential for the protein to have a low level of DNA sequence specificity. Overall, the CPV structure shows that DNA is flexible when it is not tied to the restraints of a double helix but is interacting as a single strand with protein.

The assembly of infectious progeny viruses often requires the specific recognition of the viral nucleic acid by the capsid or its components. The structure of CPV shows a small section of ssDNA bound to the inside of the capsid protein coat. The bases, rather than the ribose-phosphate backbone, make most of the interactions with the capsid and show that some sequences are preferred. The requirements for satisfying such a preference in DNA recognition impose restraints on codon selection, as the preferred homologous sequence must be repeated frequently (at least 30 times in the case of CPV). The present analysis should be applicable to other viruses in which the capsid recognizes the nucleic acid in a number of analogous sequences within the genome.

The mechanism for specific encapsidation of viral genomes into the assembled mature virion is poorly understood. Some viruses (but not CPV) require the genomic nucleic acid for assembly of the capsid, but many plant [30] and animal [19,23] viral capsids can be correctly assembled in the presence of non-genomic oligonucleotides, suggesting a lack of specific recognition. In these cases, the high concentration of genomic nucleic acid in infected cells may assure its encapsidation. However, specific recognition of the genomic RNA by capsid protein has been clearly demonstrated for tobacco mosaic virus and for the RNA phage MS2 [16], where the RNA:capsid protein complexes form an initiation signal for viral assembly [8]. Similar mechanisms of initiation, recognition and assembly occur in, for example, alfalfa mosaic virus [31–33] and alphaviruses [34,35]. In most of these cases, little is known about either the specific RNA sequence or the protein structure involved in complex formation. With icosahedral viruses exhibiting various levels of specificity, it is likely that several mechanisms of encapsidation exist.

The observed nucleic acids within some icosahedral virus structures [10–13] have interactions between the ribose-phosphate backbone and capsid protein that are not specific for the nucleic acid sequence. These structures suggest that such binding may be a useful mechanism for nucleic acid packaging, a scaffold for capsid assembly, or

both. In contrast, the low specificity of binding of genomic DNA to the icosahedral interior of the CPV capsid suggests mechanisms for the recognition of genomic DNA by capsid protein and DNA packaging.

Materials and methods

Model building and structure refinement

The unrefined model of CPV [1] was improved (MS Chapman and MG Rossmann, unpublished data) by four rounds of automatic refinement, alternated with model building using the program O [36]. The relative locations of phosphate and ribose peaks in the electron density specified the chain direction of the DNA unambiguously (Fig. 1). The electron density at each DNA base was the average of all of the bases at the 60 symmetry-equivalent positions. The ssDNA structure model was built with bases that fitted the electron density best and minimized steric conflict with the protein.

Several virus structures have previously been refined in reciprocal space using either a modified version of PROLSQ [37,38] or X-PLOR [12,39]. Reciprocal-space methods are favored for non-viral structures, because they are independent of poorly determined phases. However, icosahedral symmetry averaging permits an accurate determination of phases [38]. Thus, it was reasonable to use real-space refinement methods in which the atoms in only one of the non-crystallographic asymmetrical units (1/60 of a crystallographic asymmetric unit) needed to be refined.

Real-space methods [40] have previously been used for the refinement of satellite tobacco necrosis virus [41], poliovirus and Theiler virus [42]. The method used for the refinement of CPV [43] incorporated geometric restraints [44] and took into consideration the resolution limit of the map. The geometric restraints (Table 7) were the defaults for TNT [44], except that the main chain ϕ and ψ torsion angles were loosely restrained to the favorable areas of a Ramachandran plot [45], restraints were added for the N1–N9 torsional angles between deoxyriboses and bases [2], and the target value for the phosphoribose backbone torsion angle ζ was changed from 260° (appropriate for B-DNA) to 290° (intermediate between A and B forms). The value of the free R-factor, R_T^{free}, based on reflections not used in the refinement [27], did not change with the addition of 75 explicit water molecules, suggesting that the placement of solvent atoms was unreliable. Therefore, solvent atoms were not included in the final model.

The model was initially refined into an electron-density map [1] that had been obtained through extensive phase refinement and extension starting from low-resolution single isomorphous replacement phases, using the 60-fold non-crystallographic redundancy [46]. During the latter process, phases in local regions of reciprocal space might have converged to values that are consistent with the non-crystallographic symmetry, but represent a different hand or choice of origin from other parts of reciprocal space [15,47]. Therefore, phase refinement [48] was repeated, starting with phases calculated from a partially refined atomic model. Although the phases started out 25° different on average, they refined in 10 cycles to within 12° of the previously refined phases. The electron-density map computed with the original phases was very similar to that computed with the new phases.

Table 7. Phosphoribose backbone and selected peptide torsion angle restraints.

Angle	Definition				Target values[b]			σ[c]
α	O3′	+P[a]	+O5′	+C5′	60	120	300	20
β	P	O5′	C5′	C4′	50	170	290	25
γ	O5′	C5′	C4′	C3′	55	175	295	20
δ	C5′	C4′	C3′	O3′	0	120	240	30
ε	C4′	C3′	O3′	+P	10	190	310	30
ζ	C3′	O3′	+P	+O5′	50	170	290	30
ν_0	C4′	O4′	C1′	C2′	0	120	240	40
ν_1	O4′	C1′	C2′	C3′	0	120	240	40
ν_2	C1′	C2′	C3′	C4′	0	120	240	40
ν_3	C2′	C3′	C4′	O4′	0	120	240	40
ν_4	C3′	C4′	O4′	C1′	0	120	240	40
χ (Pu)	O4′	C1′	N9	C5	30	210		40
χ (Py)	O4′	C1′	N1	C2	40	220		40
ψ	N	CA	C	+N	135	315		30
ω	Cα	C	+N	+Cα	0	180		10
ϕ	C	+N	+Cα	+C	75	255		30
χ	Aliphatic side chains				60	180	300	15

[a]+ indicates an atom in the subsequent residue. [b]The two or three target values are alternatives that are equally acceptable. [c]σ is an estimate of the usual variation of the parameter. Its inverse is used as a weight in refinement [44]. The restraints on the nucleotide angles are about as tight as on the polypeptide backbone (ψ, ω, ϕ), but looser than on the protein side chains (χ).

The final overall value of R_T^{free} [27] was 28.3% for ~4400 randomly chosen reflections from data above 5σ(F) extending to a resolution of 2.9 Å. This is comparable with the R_T^{free} calculated for well-refined protein structures at a similar resolution [27]. The deviations from ideal geometries for bond lengths were ±0.02 Å, for bond angles ±2.7°, and for torsion angles (including ϕ, ψ and χ) ±15°. Each non-crystallographic asymmetric unit contained 167 close contacts with an rms deviation from ideality of 0.1 Å. Statistics derived from the program PROCHECK [49] for other stereochemical parameters, such as peptide torsion angles and side-chain conformations, compared favorably with other well-refined structures at this resolution.

DNA-binding specificity

A DNA sequence profile (Table 6) was constructed with subjective estimates of probability that a specific base (B) is A, G, C or T at each nucleotide [Pi(B) for i=1 to 11]. Thus, at a particular nucleotide, the base might be judged to be predominantly C (at say 60% of the superimposed binding sites), possibly G (20%), but improbably A or T (5% each). The estimates were based on the shape of the electron density, potential steric overlap and the possibilities of hydrogen bonding. Each nucleotide was also assigned a weight (w) based on the confidence with which its electron density could be explained by the assigned mixture of bases.

The profile was used to search for sequences within the packaged DNA, i.e. the opposite strand of the published negative-sense DNA sequence [50]. The Program Find Matches was written using subroutines from the Genetics Computer Group library [51,52]. The profile was compared with every 11-residue fragment of the genomic DNA. At each starting position j of the CPV genome, by analogy to conditional

probability, a score S_j was calculated for the 11 bases, $B_j=1, 11$ in the profile, where

$$Sj = \frac{\Pi_{i=1}^{11}[P_i(B=B_{i+j-1})]^{w_i}}{\Pi_{i=1}^{11}0.25^{w_i}} \quad (1)$$

An alternative scoring system, s_j, was also used where

$$s_j = \frac{\Sigma_{i=1}^{11}w_iP_i(B=B_{i+j-1})}{\Sigma_{i=1}^{11}0.25w_i} \quad (2)$$

This is less sensitive to errors in the estimate of P_i and resembles the scores used in many sequences-alignment programs (e.g. [53]). Analysis using both scoring systems gave similar results. Hence, only S_j is presented. The highest scores were sorted and ranked.

To assess the significance of matches of the profile with CPV sequences, scores were compared with expected values. The expected scores were calculated in two different ways designed to test two slightly different questions. Firstly, does the profile resemble fragments of the genomic sequence more than expected from random sequences? Expected scores were calculated from randomized genomic DNA sequences. The CPV genomic sequence was shuffled randomly 100 times; for each shuffled sequence, a matching score was calculated for the profile aligned at each position within the sequence; the mean and standard deviations were calculated for scores of the same rank from different shuffled sequences. Scores for the real sequence could then be compared with the expected scores of corresponding rank.

Secondly, does the profile resemble fragments of the genomic sequence more than expected from randomly constructed profiles? Expected scores were calculated from randomized profiles. Thirty 11-nucleotide fragments were randomly selected from the CPV genome; the frequencies of each of the four bases at each position within the fragments were used as the 'probabilities' for the profile; weights from the real profile were used. Scores were calculated with respect to the alignment of the random profiles at each position of the genomic sequence. Mean expected scores and their standard deviations were obtained from scores of the same rank calculated with different random profiles. The expected scores for this test depend on the number of fragments used to make the random profile. In the trivial case of using one fragment, a (high scoring) exact match will be found to its own sequence. When using a small number of fragments, good matches will be found to each of the constituent sequences. As the number of fragments increases, the profile broadens and the highest expected matching scores are reduced. The height of the observed electron density suggested that at least 30 of the sites were occupied. However, the results of the analysis were the same even when the random profile was estimated from only 10 sequence fragments.

Coordinates have been deposited with the Brookhaven Protein Data Bank (PDB entry 3DPV).

Acknowledgements: We are grateful for the help of our colleagues Jun Tsao, Mavis Agbandje, Hao Wu and Walter Keller, during earlier phases of the structure determination. This work was supported by grants from the National Institutes of Health and the National Science Foundation to MGR. We are also grateful to the Lucille P Markey Foundation for the development of structural biology at Purdue University and Florida State University.

References

1. Tsao, J., *et al.*, & Parrish, C.R. (1991). The three-dimensional structure of canine parvovirus and its functional implications. *Science* **251**, 1456–1464.
2. Saenger, W. (1984). *Principles of Nucleic Acid Structure.* Springer-Verlag, New York.
3. Steitz, T.A. (1993). *Structural Studies of Protein–Nucleic Acid Interaction: The Sources of Sequence-Specific Binding.* Cambridge University Press, UK.
4. McPherson, A., Brayer, G. & Morrison, R.D. (1986). Crystal structure of RNase A complexed with $d(pA)_4$. *J. Mol. Biol.* **189**, 305–327.
5. Freemont, P.S., Friedman, J.M., Beese, L.S., Sanderson, M.R. & Steitz, T.A. (1988). Co-crystal structure of an editing complex of Klenow fragment with DNA. *Proc. Natl. Acad. Sci. USA* **85**, 8924–8928.
6. Moras, D., *et al.*, & Giegé, R. (1980). Crystal structure of yeast $tRNA^{Asp}$. *Nature* **288**, 669–674.
7. Rould, W.A., Perona, J.J. & Steitz, T.A. (1991). Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. *Nature* **352**, 213–218.
8. Stubbs, G. (1989). Protein–nucleic acid interactions in tobacco mosaic virus. In *Protein–Nucleic Acid Interactions.* (Saenger, W. & Heinemann, U., eds), pp. 87–109, CRC Press, Boca Raton.
9. Rossmann, M.G., Chandrasekaran, R., Abad-Zapatero, C., Erickson, J.W. & Arnott, S. (1983). Appendix I. RNA–protein binding in southern bean mosaic virus. *J. Mol. Biol.* **166**, 73–80.
10. Chen, Z., *et al.*, & Johnson, J.E. (1989). Protein–RNA interactions in an icosahedral virus at 3.0 Å resolution. *Science* **245**, 154–159.
11. Larson, S.B., *et al.*, & McPherson, A. (1993). Double-helical RNA in satellite tobacco mosaic virus. *Nature* **361**, 179–182.
12. Larson, S.B., *et al.*, & McPherson, A. (1993). Three-dimensional structure of satellite tobacco mosaic virus at 2.9 Å resolution. *J. Mol. Biol.* **231**, 375–391.
13. Fisher, A.J. & Johnson, J.E. (1993). Ordered duplex RNA controls capsid architecture in an icosahedral animal virus. *Nature* **361**, 176–179.
14. McKenna, R., Ilag, L.L. & Rossmann, M.G. (1994). Analysis of the single-stranded DNA bacteriophage ϕX174, refined at a resolution of 3.0 Å. *J. Mol. Biol.* **237**, 517–543.
15. McKenna, R., Xia, D., Willingmann, P., Ilag, L.L. & Rossmann, M.G. (1992). Structure determination of the bacteriophage ϕX174. *Acta Crystallogr. B* **48**, 499–511.
16. Valagård, K., Murray, J.B., Stockley, P.G., Stonehouse, N.J. & Liljas, L. (1994). Crystal structure of an RNA bacteriophage coat protein–operator complex. *Nature* **371**, 623–626.
17. Berns, K.I. (1990). Parvoviridae and their replication. In *Virology.* (Fields, B.N. & Knipe, D.M., eds), pp. 1743–1763, Vol. **2**, 2nd edn, Raven Press, New York.
18. Chapman, M.S. & Rossmann, M.G. (1993). Structure, sequence and function correlations among parvoviruses. *Virology* **194**, 491–508.
19. Carter, B.J. (1990). Parvoviruses as vectors. In *Handbook of Parvoviruses.* (Tijssen, P., ed), pp. 247–284, CRC Press, Boca Raton.
20. Altona, C. & Sundaralingam, M. (1972). Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. *J. Am. Chem. Soc.* **94**, 8502–8512.
21. de Leeuw, H.P.M., Haasnoot, C.A.G. & Altona, C. (1980). Empirical correlations between conformational parameters in β-D-furanoside fragments derived from a statistical survey of crystal structures of nucleic acid constituents. *Isr. J. Chem.* **20**, 108–126.
22. Holbrook, S.R., Sussman, J.L., Warrant, R.W. & Kim, S.-H. (1978). Crystal structure of yeast phenylalanine transfer RNA. II. Structural features and functional implications. *J. Mol. Biol.* **123**, 631–660.
23. Wu, H. & Rossmann, M.G. (1993). The canine parvovirus empty capsid structure. *J. Mol. Biol.* **233**, 231–244.
24. Luzzati,'V. (1953). Resolution d'une structure crystalline lorsque les positions d'une partie des atomes sont connues: traitement statistique. *Acta Crystallogr.* **6**, 142–152.
25. Liljas, L. (1991). Structure of spherical viruses. *Int. J. Biol. Macromol.* **3**, 273–279.
26. Scordari, F. (1992). Ionic crystals. In *Fundamentals of Crystallography.* (Giacovazzo, C., ed), pp. 420–421, Oxford University Press, UK.
27. Brünger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475.
28. Berman, H.M. (1986). Nucleic acid hydrations: a case study. *Trans. Am. Cryst. Assoc.* **22**, 107–109.
29. Klimasaukas, S., Kumar, S., Roberts, R.J. & Cheng, X. (1994). *Hhal* methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357–369.
30. Sorger, P.K., Stockley, P.G. & Harrison, S.C. (1986). Structure and assembly of turnip crinkle virus II. Mechanism of reassembly *in vitro*. *J. Mol. Biol.* **191**, 639–658.

31. Koper-Zwarthoff, E.C. & Bol, J.F. (1980). Nucleotide sequence of the putative recognition site for coat protein in the RNAs of alfalfa mosaic virus and tobacco streak virus. *Nucleic Acids Res.* **8**, 3307–3318.
32. Houwing, C.J. & Jaspars, E.M.J. (1980). Complexes of alfalfa mosaic virus RNA 4 with one and three coat protein dimers. *Biochemistry* **19**, 5255–5260.
33. Houwing, C.J. & Jaspars, E.M.J. (1980). Preferential binding of 3'-terminal fragments of alfalfa mosaic virus RNA 4 to virions. *Biochemistry* **19**, 5261–5264.
34. Weiss, B., Nitschko, H., Ghattas, I., Wright, R. & Schlessinger, S. (1989). Evidence for specificity in the encapsidation of Sindbis virus RNAs. *J. Virol.* **63**, 5310–5318.
35. Wengler, G., Wurkner, D. & Wengler, G. (1992). Identification of a sequence element in the alphavirus core protein which mediates interaction of cores with ribosomes and the disassembly of cores. *Virology* **191**, 880–888.
36. 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 Crystallogr. A* **47**, 110–119.
37. Hendrickson, W.A. (1985). Stereochemically restrained refinement of macromolecular structures. *Methods Enzymol.* **115**, 252–270.
38. Arnold, E. & Rossmann, M.G. (1988). The use of molecular-replacement phases for the refinement of the human rhinovirus 14 structure. *Acta Crystallogr. A* **44**, 270–282.
39. Brünger, A.T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science* **235**, 458–460.
40. Diamond, R. (1971). A real-space refinement procedure for proteins. *Acta Crystallogr. A* **27**, 436–452.
41. Jones, T.A. & Liljas, L. (1984). Crystallographic refinement of macromolecules having non-crystallographic symmetry. *Acta Crystallogr. A* **40**, 50–57.
42. Yeates, T.O., *et al.*, & Hogle, J.M. (1991). Three-dimensional structure of a mouse-adapted type 2/type 1 poliovirus chimera. *EMBO J.* **10**, 2331–2341.
43. Chapman, M.S. (1995). Restrained real-space macromolecular atomic refinement using a new resolution-dependent electron density function. *Acta Crystallogr. A*, in press.
44. Tronrud, D.E., Ten Eyck, L.F. & Matthews, B.W. (1987). An efficient general-purpose least-squares refinement program for macromolecular structures. *Acta Crystallogr. A* **43**, 489–501.
45. Ramachandran, G.N., Ramakrishnan, C. & Sasisekharan, V. (1963). Stereochemistry of polypeptide conformations. *J. Mol. Biol.* **7**, 95–99.
46. Tsao, J., *et al.*, & Rossmann, M.G. (1992). Structure determination of monoclinic canine parvovirus. *Acta Crystallogr. B* **48**, 75–88.
47. Valegård, K., Liljas, L., Fridborg, K. & Unge, T. (1991). Structure determination of bacteriophage MS2. *Acta Crystallogr. B* **47**, 949–511.
48. Rossmann, M.G., *et al.*, & Lynch, R.E. (1992). Molecular replacement real-space averaging. *J. Appl. Crystallogr.* **25**, 166–180.
49. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemistry quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
50. Parrish, C.R., Aquadro, C.F. & Carmichael, L.E. (1988). 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.
51. Genetics Computer Group, Inc. (1991). *Sequence Analysis Software Package.* Madison, WI.
52. Devereux, J., Haerberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387–395.
53. Gribskov, M., McLachlan, A.D. & Eisenberg, D. (1987). Profile analysis: detection of distantly related proteins. *Proc. Natl. Acad. Sci. USA* **84**, 4355–4358.
54. Saran, A., Perahia, D. & Pullman, B. (1973). Molecular orbital calculations on the conformation of nucleic acids and their constituents. VII. Conformation of the sugar ring in β-nucleosides. *Theor. Chim. Acta (Berlin)* **30**, 31–44.
55. Chapman, M.S. (1993). Mapping the surface properties of macromolecules. *Protein Sci.* **2**, 459–469.

Received: **4 Oct 1994**; revisions requested: **7 Nov 1994**; revisions received: **15 Dec 1994**. Accepted: **16 Dec 1994**.