Published in final edited form as:
Nature. 1992 January 9; 355(6356): 137–143. doi:10.1038/355137a0.

Atomic structure of single-stranded DNA bacteriophage ΦX174 and its functional implications

Robert McKenna[*], **Di Xia**[*], **Peter Willingmann**[*,†], **Leodevico L. Ilag**[*,‡], **S. Krishnaswamy**[*,†], **Michael G. Rossmann**[*,§], **Norman H. Olson**[*], **Timothy S. Baker**[*], and **Nino L. Incardona**[‡]

[*]Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, USA

[‡]Department of Microbiology and Immunology, Center for the Health Sciences, University of Tennessee, 858 Madison Avenue, Memphis, Tennessee 38163, USA

Abstract

The mechanism of DNA ejection, viral assembly and evolution are related to the structure of bacteriophage ΦX174. The F protein forms a $T = 1$ capsid whose major folding motif is the eight-stranded antiparallel β barrel found in many other icosahedral viruses. Groups of 5 G proteins form 12 dominating spikes that enclose a hydrophilic channel containing some diffuse electron density. Each G protein is a tight β barrel with its strands running radially outwards and with a topology similar to that of the F protein. The 12 'pilot' H proteins per virion may be partially located in the putative ion channel. The small, basic J protein is associated with the DNA and is situated in an interior cleft of the F protein. Tentatively, there are three regions of partially ordered DNA structure, accounting for about 12% of the total genome.

Bacteriophage ΦX174 is a small icosahedral virus that contains a single-stranded, closed circular DNA molecule with 5,386 nucleotide bases (for a recent review, see ref. 1). Of the 11 gene products, four (J, F, G and H) participate in the structure of the virion. There are 60 copies each of the J, F and G proteins and 12 copies of the H protein per virion[2–4] (Table 1). Electron microscopy of shadowed or negatively stained particles[5–8], as well as a three-dimensional reconstruction of frozen-hydrated virions (Fig. 1*a*), clearly show large spikes at the vertices of the icosahedral particles (N.H.O., T.S.B., P.W. and N.L.I., manuscript in preparation). These spikes are about 32 Å long and have a mean diameter of 70 Å. Edgell *et al.*[3] have shown that these spikes, which can be removed from the capsid with 4 M urea treatment, are composed of G and H proteins. Assuming all 12 spikes are identical, then each spike will contain five G and one H protein[2]. The capsid itself has an external diameter of roughly 260 Å, with a protein shell that is about 30 Å thick, and contains the F and J proteins as well as the single-stranded (ss) DNA. The phage recognizes a lipopolysaccharide (LPS) receptor on the outer cell membrane of *Escherichia coli*[8–11]. Host range mutants, which determine which strains of *E. coli* are infected by ΦX174, map into the F, G and H genes[12–15].

[§]To whom correspondence should be addressed.
[†]Present addresses: Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK (P.W.); Bioinformatics Center, Department of Biotechnology, Madurai Kamaraj University, Madurai 625 021, India (S.K.).

After attachment, which can occur at or above 15 °C, the phage DNA and at least one of the H proteins are injected into the *E. coli* periplasmic space on raising the temperature to 37 °C. As this temperature shift produces an irreversible loss of infectivity, it is called the eclipse reaction. The transfer of phage DNA into the cell requires energy and may be coupled to synthesis of the complementary DNA strand[16]. Eclipse rate, as measured by loss of infectivity, can be modulated by deletion of certain bases in the viral genome[17] as well as by mutations that map in genes F (L.L.I., J.K. Tuech, L. A. Beisner, R. A. Sumrada and N.L.I., manuscript in preparation) and H (ref. 18, and L.L.I, and N.L.I., manuscript in preparation). The lipid portion of the LPS receptor is required for the DNA injection process[10]. However, eclipse can also be initiated by incubation at 37 °C in the presence of 100 mM $CaCl_2$ (ref. 19). The eclipsed particles, which have a sedimentation coefficient of 90S, are sensitive to nucleases[13]. Electron microscopy[20,21] shows DNA is an extrusion from one of the spikes when the LPS is used and from many spikes when Ca^{2+} is used to initiate eclipse.

In later stages of the viral life cycle, five molecules of F protein form 9S particles and five molecules of G protein form 6S particles. These pentamers together with the H protein and the B and D scaffolding proteins form 108S proheads. The ssDNA is then packaged into proheads along with 60 J proteins to form, after removal of the B proteins, a 132S complex[22,23]. Conversion to the infective 114S particles involves removal of the D protein, requires divalent cations and may occur in the final stages of morphogenesis and cell lysis[24]. Various temperature-sensitive (ts) and chain-termination mutants that block the assembly pathway have been isolated and characterized[1] (L.L.I, and N.L.I., manuscript in preparation). The small J protein is required for DNA packaging[23]. They contain six lysines and six arginines, typical of other viral capsid proteins that have a role in nucleic acid packaging (see, for instance, ref. 25).

The phage ΦX174 was the first example of a biological system in which the complete nucleotide sequence of a DNA genome was successfully determined[26,27] and shown to contain 11 genes with overlapping reading frames. Several similar phages have been isolated[28] of which only ΦX174, G4 (ref. 24) and S13 (ref. 30) have been sequenced. Comparisons of the protein sequences demonstrate that the spike protein G is the most variable. The H protein contains a much larger percentage of glycines and smaller percentage of aromatic residues than is usual for globular proteins. Such sequences often give rise to disordered peptide conformations and, hence, lack distinct structure[31].

The infectious virion has a relative molecular mass of 6.2×10^6 (M_r 6,200K) (26% is DNA)[32]. Lysates of infected *E. coli* contain two components which can be separated by sedimentation through a sucrose gradient[32,33]. The fast band contains the infectious 114S particles whereas the slow band contains noninfectious, 70S particles[32,34,35]. The latter contains about 20% of the normal DNA content with all of the genome represented in the population of particles. Both types of particles can be crystallized into isomorphous monoclinic crystals with space group $P2_1$, and cell dimensions a = 305.6, b = 360.8, c = 299.5 Å, β = 92.89° that diffract at least to 2.6-Å resolution[33]. The asymmetric unit of the crystal cell contains one complete particle. We report here the three-dimensional structure determination of ΦX174 and its implication for viral functions.

Protein structure

The structure determination of ΦX174 is summarized in Table 2 and will be described in detail elsewhere (R.M., D.X., P.W., L.L.I, and M.G.R., manuscript in preparation). The structure was initially interpreted at 3.5 Å resolution in terms of the amino-acid sequences of the F, J and G proteins[36], but the mutant ΦX174 used here has an arginine at position 216 in the F protein (Fig. 1*b*). An atomic model was built in an Evans and Sutherland PS390 graphics system with respect to a final 3.4 Å resolution map with the use of the program FRODO[37]. The overall appearance of the atomic structure of ΦX174 is remarkably similar to that found by electron microscopy of unstained, frozen-hydrated virions (Fig. 1). The maximum outer radius is 165 Å measured along the 5-fold axes. Each of the 12 spikes consists of 5 G proteins which make only minimal contact with the capsid itself (Fig. 2*d*, Table 3). The spikes have a pentagonal cross-section (Fig. 2*g*), with an average diameter of 70 Å, and a somewhat blunt extremity that extends about 32 Å above the mean capsid surface. The capsid has a number of protrusions on its surface, one of which extends to a radius of 138 Å at the 3-fold icosahedral axes, but at the 2-fold axes there is a depression in the surface that only extends to a radius of 111 Å. The protein coat has an average thickness of about 30 Å.

The capsid itself consists primarily of the F protein, with $T = 1$ (ref. 38) icosahedral symmetry, and shows protrusions on the 3-fold axes, in the centre of the triangle formed by adjacent 5- and 3-fold axes, and midway between the 5- and 3-fold axes. The F polypeptide was easily traced with confidence except for the first and last two residues which are disordered. The central motif of the F protein is an eight-stranded antiparallel β barrel similar to that found in many other virus coat proteins (Fig. 2*b*), although the usual 'αA' and 'αB' helices are missing (see, for example, ref. 39 for definition of secondary structure nomenclature). The organization of the β barrel with respect to the icosahedral symmetry is similar to that in other viruses: the β strands run roughly tangential to the viral surface and the B–C, D–E, F–G and H–I turns point toward the 5-fold axes. This β barrel projects into the nucleic-acid region of the virus. The contact between neighbouring β barrels is limited, occuring at the 5-fold vertices (Table 3), and, therefore, cannot be considered to be the dominating interactions required for assembly. Superposition of the F subunit β barrel onto VP1 of human rhinovirus 14 (HRV14)[40] and canine parvovirus (CPV)[41] shows that 101 and 92 amino acids of the F protein can be topologically superimposed onto the corresponding β barrels of HRV14 and CPV with an r.m.s. deviation of 3.5 and 3.4 Å, respectively, for equivalenced Cα coordinates. The insertions between the β strands (Fig. 2*a*) are almost as extensive as in CPV, accounting for more than 60% of the total F protein. These insertions form the exterior of the viral capsid and most of the contacts between neighbouring F subunits (Table 3). The insertions contain five α helices of which one has five-and-a-half turns. The long helix runs antiparallel to two short and nearly colinear helices from a 3-fold and a 2-fold related F subunit, respectively. The insertions also account for the protrusions on the viral surface.

The J protein is internal[23] and, because of its high content of basic residues, was expected to be associated with DNA. However, carboxy-terminal residues 28-RLWYVGGQQF-37 contain only one basic residue and are ordered. Their characteristic sequence was easy to

recognize in the electron density. They make a loop on an interior cleft of the capsid protein and contact the E–F insertion. Many plant viruses have very basic, rather variable, amino-terminal extensions to their capsid proteins. These basic regions are almost invariably 'disordered' (for example, tomato bushy stunt virus[42], southern bean mosaic virus[43] and satellite tobacco necrosis virus[44]), lacking a unique conformation. When there is no especially basic protein component in the capsid protein, there usually is a significant amount of polyamines to neutralize the nucleic acid[45]. The disordered basic part of the J protein, therefore, is probably involved in DNA neutralization and packaging and may account for the difference of the Raman spectrum between ΦX174-DNA when in solution and in the virion[46].

The G polypeptide was traced over its entire length with complete confidence. Its structure is a somewhat complex β barrel with its strands running roughly radially (Fig. 2*f, g*). It has considerable resemblance to the β-barrel domain of the F protein (Fig. 2*a, e*) and, indeed, to many other viral capsid proteins[39]. The β turns, BC, DE, FG and HI, that point toward the 5-fold axes in many viral capsid proteins, are at the blunt extremity of the G protein spike. Of the 175 amino acids in the G protein, 73 residues can be topologically superimposed onto the F protein with a 3.0 Å r.m.s. difference in the Cα positions. There is a rather hydrophilic channel down the 5-fold axis of each spike (Fig. 2*h*), with a diameter which varies between 6 and 26 Å. There is diffuse electron density in this channel at a radius of about 135 Å. There is also another uninterpreted, but smaller, piece of electron density at a radius of 159 Å. The total cavity in the channel is large enough to accommodate at least 15 amino acids. The quality of the G-protein electron density is particularly surprising as the crystallographic environment of each spike is different and might, therefore, have asymmetrical lattice forces causing local deviation from icosahedral symmetry and, hence, blurring of electron density. Furthermore, the quality of the map is even more surprising because the spikes make only a few contacts with the capsid itself (Table 3).

There is no compelling evidence in the electron density map for the presence of the H protein. This is not surprising, given the unusually large number of glycines and the low percentage of aromatic residues in the H protein, which suggest that this protein may not have a distinct fold[31]. The absence of any defined structure for the H protein may be due to the effect of electron density averaging, to different orientations of the protein at each spike or, most likely, to different folds of each polypeptide. Nevertheless, denaturing gels of virus dissolved from crystals show clear evidence of its presence[33]. In light of the 'pilot' function of the H protein (Table 1), it is possible that the H protein would bind to the outside extremity of the spikes. In the crystal, the spikes interdigitate, which leaves little room for a 35K protein at the vertices of the spikes. It is unlikely that DNA accounts for the diffuse density in the large cavity on the 5-fold axis of each spike as there is insufficient space for a strand of DNA to enter and exit this cavity. Genetic evidence suggests that Asn 113 of the G protein, which is in the channel facing the diffuse density (Fig. 2*h*), interacts with Ala 68 of the H protein (L.L.I., *et al.*, manuscript in preparation). Therefore, the diffuse density might be part of the H protein. However, it is not clear whether the remainder of the H protein is internal to the capsid or external to the virus, though both possibilities would be consistent

with the electron microscope observation that DNA is ejected through the spikes[21] and that the H protein is injected into *E. coli* with DNA[47].

DNA structure and G protein ion channel

A difference electron density map was calculated between the partial diffraction data (Table 2) of the full 114S particles, F_{114S}, and the mostly empty 70S particles, F_{70S}, based on the phases, $\alpha_{combined}$, determined for the 3.4 Å resolution map. This shows regions where there is ordered structure not present in the electron density map of the 70S particles. This structure is probably attributable to DNA (Fig. 3*a*) but it could be also a part of the H protein. The basic protein environment and characteristic electron density strongly suggested this to be nucleic acid. As in bean pod mottle virus[48] and CPV[41], some of the nucleic acid is bound to inside cavities of the icosahedral shell and, therefore, has itself attained some icosahedral order. The most extensive ordered DNA structure occurs in three regions of the cavities between the F protein α barrel and the rest of the capsid, including a region close to the ordered part of the J protein. A total of 11 nucleotides per icosahedral asymmetric unit could be built into the electron density while maintaining standard dihedral angles[49]. There are extensive polar interactions of the DNA backbone with the F and J proteins including Arg 216, Arg 233 and Arg 420 in the F protein as well as Arg 28 in the J protein (Fig. 3*b*). The order of the protein in this region is greater in the 114S than in the 70S particles. An electron density map using $(F_{114S} - kF_{70s})e^{i\alpha_{combined}}$, where $k = 0.9$, showed the DNA density to have the same height as the protein. Full occupancy of DNA would be expected for $k = 0.5$ (ref. 50), implying that the ordered DNA is present in about one out of five icosahedral positions.

A large negative peak occurs in the difference map in the G protein channel at the site of the smaller piece of diffuse density, suggesting that a cation is bound here in the 70S particles. This density is liganded by the five symmetry-related Asp117 residues. Hence, binding of a putative Ca^{2+} ion at this site may trigger DNA ejection[19–21] through the ion channel. Additional density in the full 114S particles occurs also on the inside of the F capsid but around the 5-fold axis, which could be either DNA or more ordered H protein poised for ejection through the channel in the spike. In addition, the J protein shows a higher degree of order in the 114S particles.

Mutations controlling infectivity stages

Both ts and cold-sensitive (cs) mutants have been selected and characterized. The cs mutations at residues Arg 216 and Arg 233 in the F protein affect the eclipse rate and are also implicated in functional interactions with DNA[17] (L.L.I., *et al.*, manuscript in preparation). One of the more highly ordered segments of DNA structure is sandwiched between these residues and the ordered region of the J packaging protein (Fig. 3). Because a change of the above arginines to cysteines or histidine should alter protein-DNA interactions, it is reasonable to expect that these and other residues adjacent to this segment of DNA should have a role in determining the rate at which infectivity is lost because of DNA ejection.

Nearly all ts mutants have defects in late stages of viral infection, presumably in the assembly process (L.L.I. and N.L.I., manuscript in preparation). The mutations in the G and F proteins map primarily to surface features on the virus (Table 4). In addition, there is one mutation on the inside cavity of the spike facing the 5-fold axes. These mutations are mostly of the type that tend to destabilize the local protein environment[51], thus explaining why elevated temperatures might disrupt assembly.

Evolution

Comparisons of the protein sequencies of ΦX174 and G4 show the greatest variability in the extremity of the G protein and the greatest conservation in the contacts around the 5-fold axes in the F protein (particularly loop F–G). The higher rate of accepted mutation in the G protein would suggest that its function (perhaps formation of the channel through which the DNA is ejected) is a relatively new feature, whereas the capsid may have evolved earlier to a more stable structure. Superposition of tertiary protein structures permits amino-acid alignments when there is no readily recognizable memory of a common primordial amino-acid sequence[52]. The alignment of the virus β-barrel fold in the F and G proteins of ΦX174 with some other viral capsid proteins, as well as the protruding domain of tomato bushy stunt virus, is shown in Fig. 4. There is a tendency to conserve hydrophobic residues in the β barrel, but there is no obvious specific sequence, as sometimes occurs, by which the virus capsid fold might be readily recognized. The degree of similarity of the ΦX174 protein folds to the other capsid protein structures is similar to that between, for instance, HRV14 VP1 and VP2 which occur in the same virus capsid.

The considerable similarity of the F protein to other viral capsid proteins suggests that many viral capsid proteins have probably diverged from a common ancestral structure[53]. The F gene has, over time, captured insertions which are important for forming the capsid exterior. As the virion exterior determines the interaction with its host, and as the host presumably already contains molecules that modulate its metabolism, such gene fusion processes in viral evolution may be important in the function of the resultant viral particle.

The eight-stranded antiparallel virus fold has now been found in many RNA and DNA viral capsids[39], yet its presence is not entirely universal (for example in the plant RNA tobacco mosaic virus[54,55], in the RNA phage MS2[56] and in the animal RNA Sindbis virus[57]). Hence, the eight-stranded antiparallel β-barrel fold is not a prerequisite for assembly of icosahedral structures[58]. Furthermore, the ΦX174 structure, in contrast to other icosahedral viruses, shows only little contact of the β-barrel fold with neighbouring subunits, as it essentially projects into the DNA core (Table 3). It has also been suggested[59,60] that the fold might be conserved to modulate viral stability by virtue of its hydrophobic interior.

Acknowledgments

We thank T. Smith and M. Chapman for discussions, the many participants in data collection trips (including M. Agbandje, J. Bibler, M. Chapman, H.-K. Choi, V. Giranda, A. Hadfield, W. Keller, M. Oliveira, A. Prongay, T. Smith, L. Tong and H. Wu) and the support at the various synchrotron facilities (the National Synchrotron Light Source at Brookhaven, the Cornell High Energy Synchrotron Source, the Scientific Research Council synchrotron in Daresbury, the Deutsches Elektronen-Synchrotron in Hamburg and the Stanford Synchrotron Research Laboratory) and H. Prongay, S Wilder and M. O'Neil for help in preparation of the manuscript. The work was supported by a NSF grant and a Lucille P. Markey Charitable Trust grant for the development of structural studies

at Purdue (to M.G.R.). an NSF grant (to N.L.I.), a travel award from the Institute for Biological Recognition and Catalysis (to L.L.I.), a NIH grant (to T.S.B.) and a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft (to P.W.). Unrefined coordinates have been deposited with the Brookhaven Protein Data Bank.

References

1. Hayashi, M.; Aoyama, A.; Richardson, DL., Jr; Hayashi, MN. The Bacteriophages (The Viruses). Calendar, R., editor. Vol. Vol. 2. New York: Plenum; 1988. p. 1-71.
2. Burgess AB. Proc. natn. Acad Sci. U.S.A. 1969; 64:613–617.
3. Edgell MH, Hutchison CA III, Sinsheimer RL. J. molec. Biol. 1969; 42:547–557. [PubMed: 5804158]
4. Siden EJ, Hayashi M. J molec. Biol. 1974; 89:1–16. [PubMed: 4613851]
5. Hall CE, Maclean EC, Tessman I. J. molec. Biol. 1959; 1:192–194.
6. Thomas WJ, Horne RW. Virology. 1961; 15:1–7. [PubMed: 13778169]
7. Stouthamer AH, Daems WT, Eigner J. Virology. 1963; 20:246–250. [PubMed: 13984471]
8. Brown DT, Mackenzie JM, Bayer ME. J. Virol. 1971; 7:836–846. [PubMed: 4105119]
9. Incardona NL, Selvidge L. J. Virol. 1973; 11:775–782. [PubMed: 4575285]
10. Jazwinski SM, Lindberg AA, Kornberg A. Virology. 1975; 66:268–282. [PubMed: 1094681]
11. Feige U, Stirm S. Biochem. biophys. Res. Commun. 1976; 71:566–573. [PubMed: 786289]
12. Sinsheimer RL. Prog. Nucleic Acid Res. molec. Biol. 1968; 8:115–169. [PubMed: 4874231]
13. Newbold JE, Sinsheimer RL. J. Virol. 1970; 5:427–431. [PubMed: 4916322]
14. Weisbeek PJ, Van de Pol JH, Van Arkel GA. Virology. 1973; 52:408–416. [PubMed: 4574511]
15. Dowell CE, Jansz HS, Zandberg J. Virology. 1981; 114:252–255. [PubMed: 6269288]
16. Newbold JE, Sinsheimer RL. J. molec. Biol. 1970; 49:49–66. [PubMed: 5450517]
17. Incardona NL, Müller UR. J. molec. Biol. 1985; 181:479–486. [PubMed: 3158743]
18. Doniger J, Tessman I. Virology. 1969; 39:389–394. [PubMed: 5358076]
19. Incardona NL. J. Virol. 1974; 14:469–478. [PubMed: 4604733]
20. Yazaki K. J. virol. Meth. 1981; 2:159–167.
21. Mano Y, Kawabe T, Komano T, Yazaki K. Agric. Biol. Chem. 1982; 46:2041–2049.
22. Fujisawa H, Hayashi M. J. Virol. 1977; 23:439–442. [PubMed: 886652]
23. Aoyama A, Hamatake RK, Hayashi M. Proc. natn. Acad. Sci. U.S.A. 1981; 78:7285–7289.
24. Mukai R, Hamatake RK, Hayashi M. Proc. natn. Acad. Sci. U.S.A. 1979; 76:4877–4881.
25. Liljas L. Prog Biophys. molec. Biol. 1986; 48:1–36. [PubMed: 3544053]
26. Sanger F, et al. Nature. 1977; 265:687–695. [PubMed: 870828]
27. Shaw DC, et al. Nature. 1978; 272:510–515. [PubMed: 692656]
28. Godson, GN. The Single-Stranded DNA Phages. Denhardt, DT.; Dressler, D.; Ray, DS., editors. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 1978. p. 103-112.
29. Godson, GN.; Fiddes, JC.; Barrell, BG.; Sanger, F. The Single-Stranded DNA Phages. Denhardt, DT.; Dressler, D.; Ray, DS., editors. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 1978. p. 51-86.
30. Lau PCK, Spencer JH. Gene. 1985; 40:273–284. [PubMed: 3007293]
31. Huber R, Bennett WS Jr. Biopolymers. 1983; 22:261–279. [PubMed: 6673759]
32. Sinsheimer RL. J. molec. Biol. 1959; 1:37–42.
33. Willingmann P, et al. J. molec. Biol. 1990; 212:345–350. [PubMed: 2138678]
34. Eigner J, Stouthamer AH, van der Sluys I, Cohen JA. J. molec. Biol. 1963; 6:61–84.
35. Weisbeek PJ, Van de Pol JH, Van Arkel GA. Virology. 1972; 48:456–462. [PubMed: 4554264]
36. Sanger F. J. molec Biol. 1978; 125:225–246. [PubMed: 731693]
37. Jones TA. J. appl. Crystallogr. 1978; 11:268–272.
38. Caspar DLD, Klug A. Cold Spring Harb. Symp quant. Biol. 1962; 27:1–24. [PubMed: 14019094]
39. Rossmann MG, Johnson JE. A Rev. Biochem. 1989; 58:533–573.

40. Rossmann MG. Nature. 1985; 317:145–153. [PubMed: 2993920]
41. Tsao J, et al. Science. 1991; 251:1456–1464. [PubMed: 2006420]
42. Harrison SC, Olson AJ, Schutt CE, Winkler FK, Bricogne G. Nature. 1978; 276:368–373. [PubMed: 19711552]
43. Abad-Zapatero C, et al. Nature. 1980; 286:33–39. [PubMed: 19711553]
44. Liljas L, et al. J molec. Biol. 1982; 159:93–108. [PubMed: 7131560]
45. Cohen SS, McCormick FP. Adv. Virus Res. 1979; 24:331–387. [PubMed: 389006]
46. Benevides JM, Stow PL, Hag LL, Incardona NL, Thomas GJ Jr. Biochemistry. 1991; 30:4855–4863. [PubMed: 1827990]
47. Jazwinski SM, Marco R, Kornberg A. Virology. 1975; 66:294–305. [PubMed: 1094683]
48. Chen Z, et al. Science. 1989; 245:154–159. [PubMed: 2749253]
49. Saenger, W. Principles of Nucleic Acid Structure. Cantor, CR., editor. New York: Springer; 1984. p. 51-104.
50. Chapman MS, Minor I, Rossmann MG, Diana GD, Andries K. J. molec. Biol. 1991; 217:455–463. [PubMed: 1847215]
51. Argos P, et al. Biochemistry. 1979; 18:5698–5703. [PubMed: 518863]
52. Rossmann MG, Moras D, Olsen KW. Nature. 1974; 250:194–199. [PubMed: 4368490]
53. Matthews BW, Rossmann MG. Meth. Enzym. 1985; 115:397–420. [PubMed: 4079794]
54. Bloomer AC, Champness JN, Bricogne G, Staden R, Klug A. Nature. 1978; 276:362–368. [PubMed: 19711551]
55. Namba K, Stubbs G. Science. 1986; 231:1401–1406. [PubMed: 3952490]
56. Valegård K, Liljas L, Fridborg K, Unge T. Nature. 1990; 345:36–41. [PubMed: 2330049]
57. Choi HH. Nature. 1991; 354:37–43. [PubMed: 1944569]
58. Ladenstein R, et al. J. molec. Biol. 1988; 203:1045–1070. [PubMed: 3145341]
59. Smith TJ, et al. Science. 1986; 233:1286–1293. [PubMed: 3018924]
60. Rossmann MG. Proc. natn. Acad. Sci. U.S.A. 1988; 85:4625–4627.
61. Rossmann MG, Blow DM. Acta crystallogr. 1962; 15:24–31.
62. Tong L, Rossmann MG. Acta crystallogr. 1990; A46:783–792.
63. Stauffacher, CV., et al. Crystallography in Molecular Biology. Moras, D.; Drenth, J.; Strandberg, B.; Suck, D.; Wilson, K., editors. Plenum; New York: 1987. p. 293-308.
64. Rossmann MG. Ada crystallogr. 1990; A46:73–82.
65. Rossmann MG. J. appl. Crystallogr. (in the press).
66. Smith TJ. J. appl. Crystallogr. 1990; 23:141–142.
67. Gibson TJ, Argos P. J. molec. Biol. 1990; 212:7–9. [PubMed: 2319600]

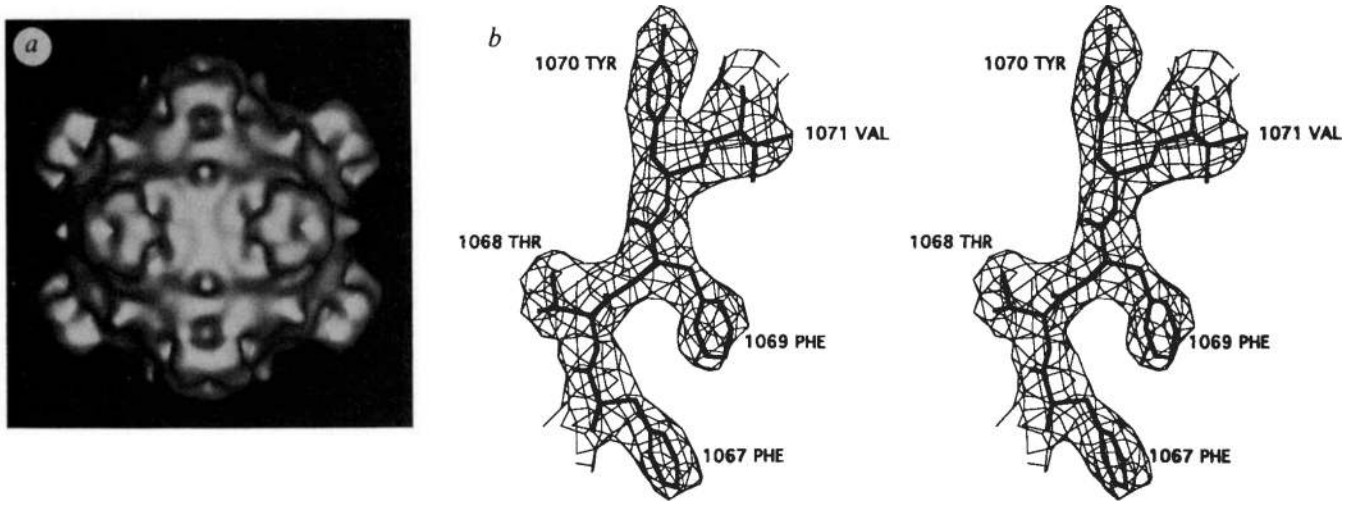

FIG. 1.

a, Shaded surface representation of three-dimensional reconstruction of ΦX174 calculated from particle images recorded in the electron microscope. Unstained virus samples were frozen in a layer of vitreous ice and images were recorded under minimal electron dose conditions (~20 e/Å^2). Twenty-five ΦX174 particle images were selected, their centres of density (origins) and orientation positions in the layer of ice were determined, and reconstructions were calculated to 21 Å (see N.H.O., T.S.B., P.W. and N.L.I., manuscript in preparation, for details on procedures). The view is down an icosahedral 2-fold axis, *b*, Stereo view of fit of residues 67-FTFYV-71 in the βE strand of the F protein electron density[66]. The amino-acid numbers in the F protein have been increased by 1,000.

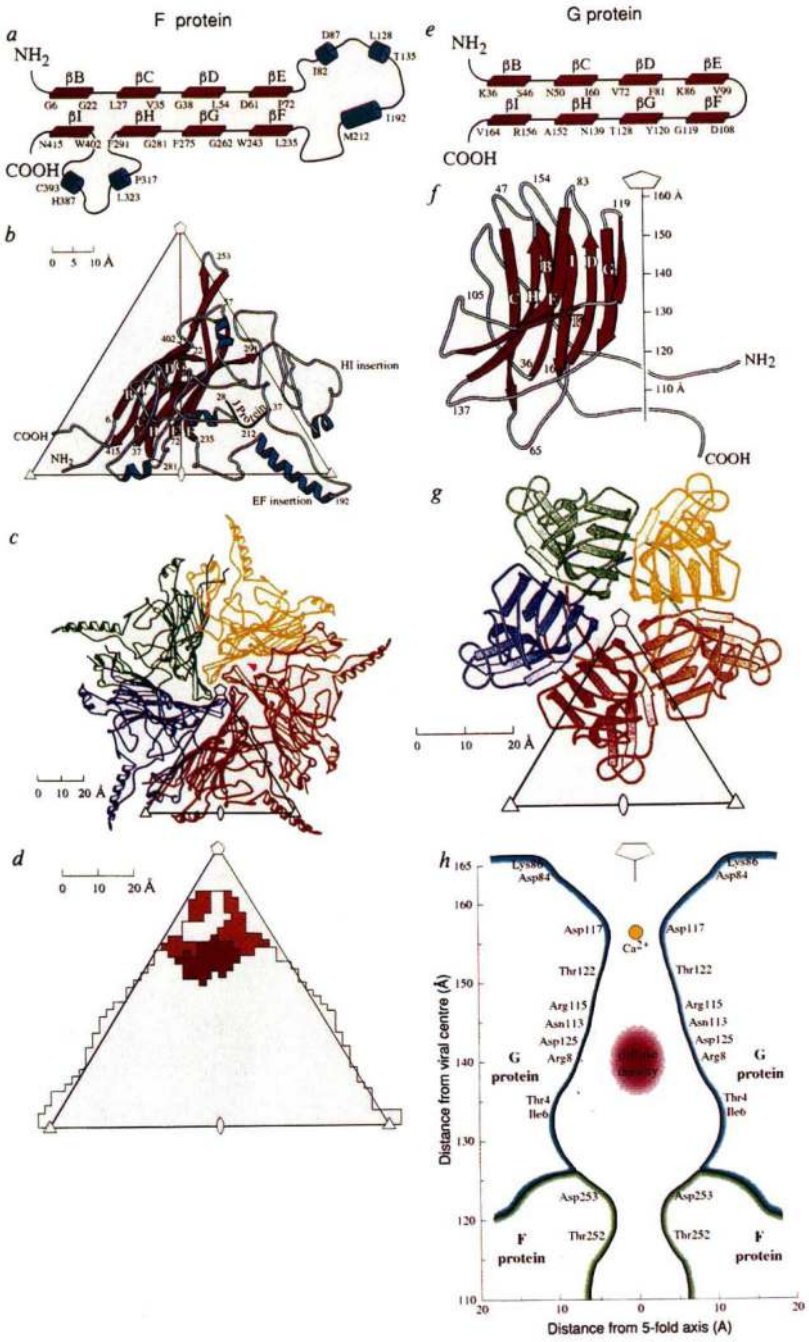

FIG. 2.

Structure of the viral proteins G and F. *a, e*, Topological depiction. The antiparallel sheet is subsequently wound into a right-handed helix to make a β barrel, *b, f*, Ribbon diagrams showing secondary structure nomenclature and amino-acid numbering. The orientation of the F protein is shown, viewed down a 2-fold axis, relative to an icosahedral asymmetric unit formed by adjacent 5-, 3- and 2-fold axes, whereas the G protein is shown as a side view with the spike pointing along the 5-fold axis. *c, g*, 9S and 6S pentameric units of the F and G proteins, respectively, viewed down a 5-fold axis. *d*, Footprint of the spike onto the F

protein. The colours indicate different sub-units of G corresponding to the code shown in *g*. *h*, Cross-section through the spike showing the amino acids that line the hydrophilic channel. The larger diffuse density may be part of the H protein and a smaller density on the 5-fold axis is a putative Ca^{2+} ion.

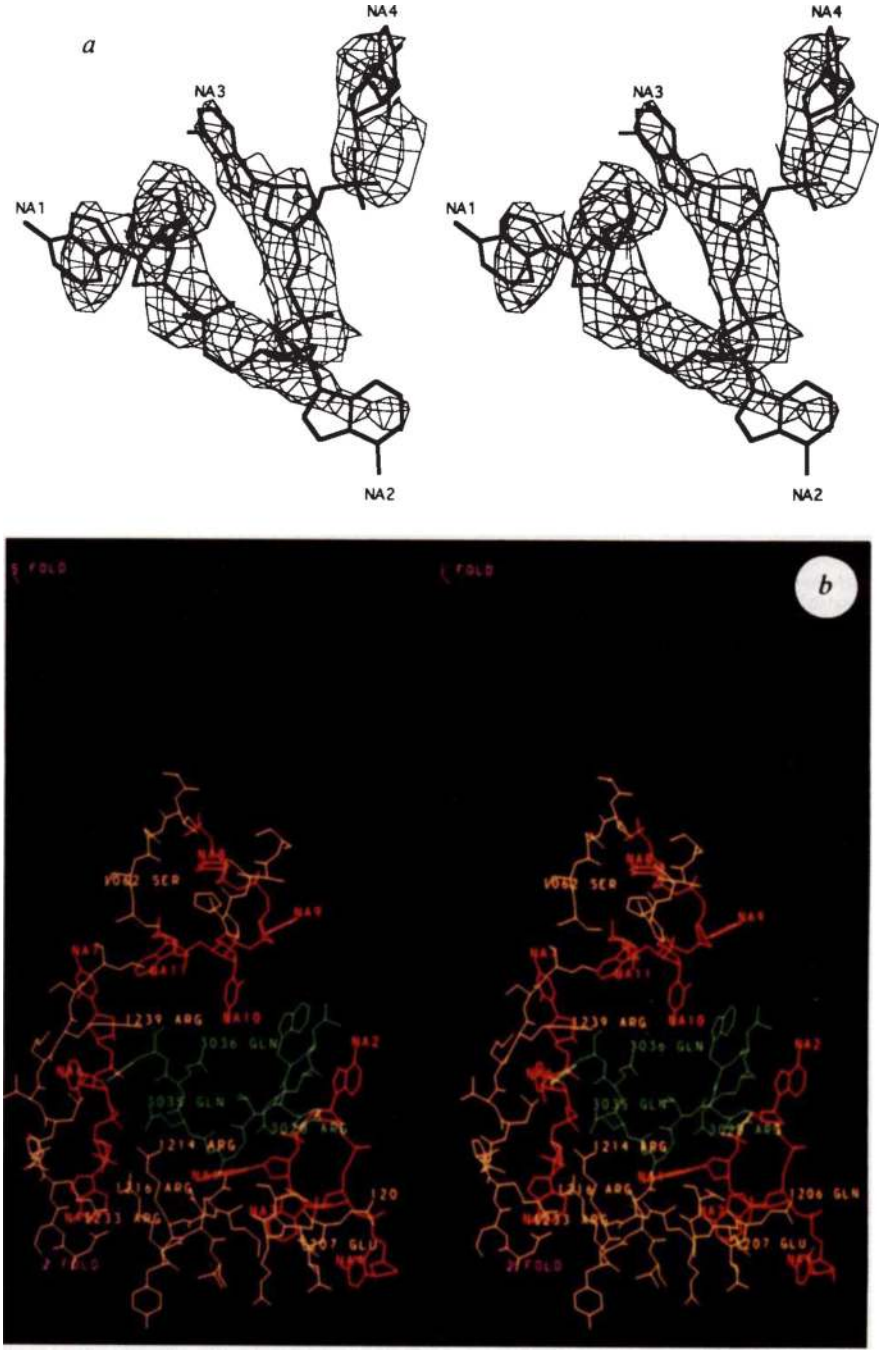

FIG. 3.

a, Stereo view of fit of DNA bases NA1 to NA4 to the electron density. The choice of nucleotide bases shown is arbitrary, *b*, Stereo diagram of environment of the J protein (green). The DNA structure is in red and the F protein is in yellow. Orientation is given by the purple 5- and 2-fold symmetry axes. The polar residues on F and J have been identified. The amino-acid numbers have been increased by 1,000 and 3,000 for the F and J proteins, respectively. The nucleotides are numbered NA1 to NA11 although they occur in three noncontiguous sequences.

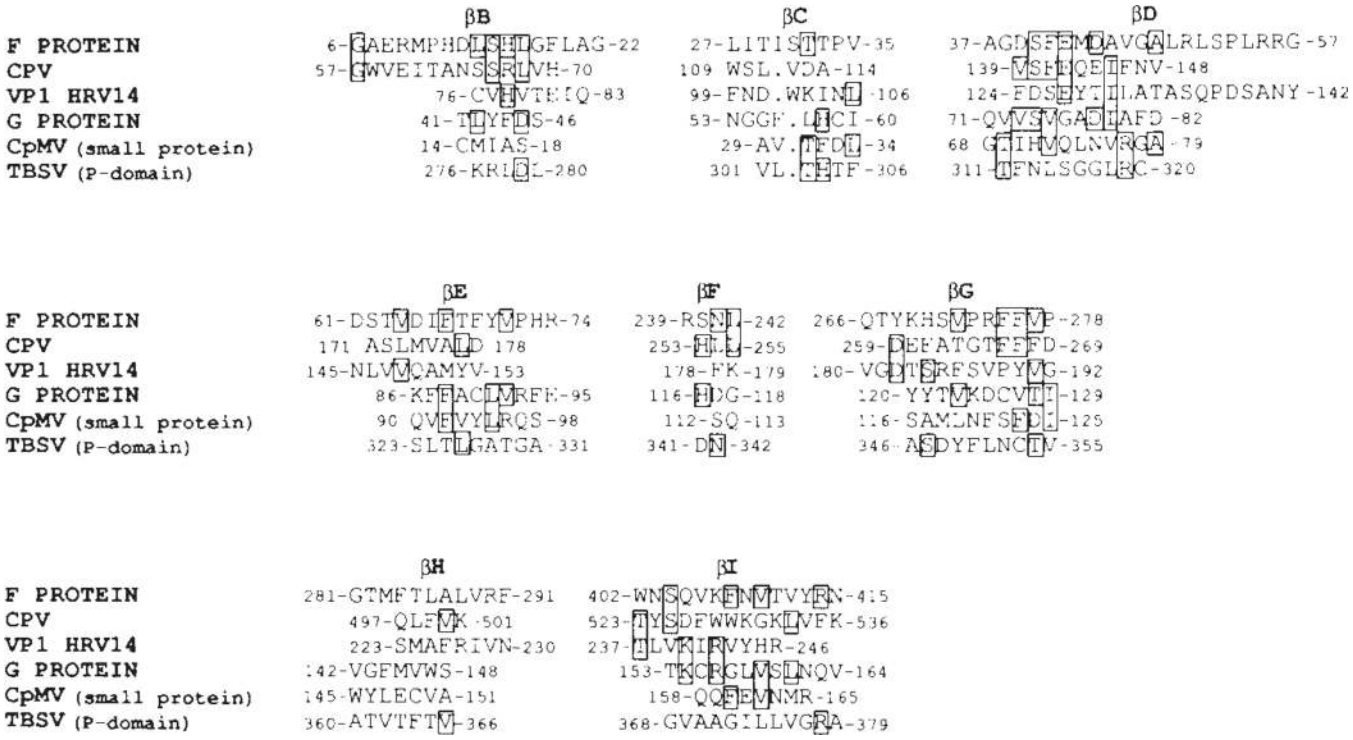

FIG. 4.

Alignment of amino-acid sequences based on structure superpositions (single-letter amino-acid code). Shown are only those residues in the β-barrel structures. Residues that are boxed are conserved in at least two sequences. The F and G proteins are aligned with each other as well as with the similar virus β barrels found in canine parvovirus (CPV), VP1 of HRV14, the small protein of CpMV and the protruding domain of TBSV. Note that the helical hand of all these structures is the same (see Fig. 2*a* and *e*) contrary to the claim by Gibson and Argos[67] concerning the TBSV protruding domain. The orientation and position of the G protein in ΦX174 are similar to that of the small protein in CpMV accounting for the similar external appearance of these two viruses (Fig. 1).

TABLE 1

Proteins in the virion

Protein	M_r	Number of amino acids	Amino-acid characteristics	Function
F	48,400	426	Normal	DNA injection and protection
G	19,050	175	Less conserved than other capsid proteins among related phages	Attachment
H	34,400	328	High Gly and low aromatic content	Attachment, DNA ejection, penetration (‘the pilot protein’)
J	4,200	37	Highly basic amino terminus	DNA packaging

TABLE 2

X-ray diffraction data collection

Resolution range	Combined data (from 70S and 114S particles)		Data from 70S particles		Data from 114S particles	
(Å)	Number	%	Number	%	Number	%
∞-30.0	940	73	924	72	95	7
30.0-15.0	8,133	91	7,976	89	1,044	12
15.0-10.0	22,240	92	21,794	90	3,039	12
10.0-7.5	40,725	86	39,467	83	5,870	12
7.5-5.0	148,401	76	145,712	75	24,130	12
5.0-4.0	193,012	73	189,722	72	31,897	12
4.0-3.5	177,528	67	168,282	64	28,447	11
3.5-3.2	148,155	60	142,153	58	22,990	9
3.2-3.0	115,489	51	112,079	49	17,295	8
3.0-2.9	53,573	39	50,945	37	4,777	3
2.9-2.7	92,108	27	86,295	25	8,754	3
Independent reflections	1,000,304		965,349		148,338	
Number of films	299		279		20	
R_{merge}	12.90		12.04		12.49	

Number of observations ($F^2 > 2\sigma(F^2)$) for each data set and percentage of possible total.

R_{merge} is defined as $\frac{\Sigma_h \Sigma_j |(\langle I_h \rangle - I_{hi})|}{\Sigma_h \Sigma_j \langle I_h \rangle} \times 100$ where $\langle I_h \rangle$ is the mean intensity of the *i* observations I_{hi} for reflection *h*. The virus was propagated, purified and crystallized as described by Willingmann *et al.*[33]. Crystals of the partially empty, slower sedimenting particles were easiest to grow. A complete X-ray diffraction data set was collected for these crystals, whereas only a fraction of the data was collected from crystals of the faster-sedimenting virus. In addition, partial data sets were collected for two heavy atom derivatives of the slower-sedimenting particles. X-ray diffraction data were obtained from non-oriented crystals as described by Willingmann *et al.*[33]. Data from crystals of full 114S particles and from partially empty 70S particles were combined for the initial structure determination. Later, these were separated to determine the presence of ordered DNA in the crystal structure. A self-rotation function[61], calculated initially at 10 Å resolution, showed icosahedral symmetry and gave an orientation of a particle in the asymmetric unit. This was refined using a locked rotation function[62] at 3.5 Å resolution. Packing considerations indicated that a particle position must be close to $(\frac{1}{4}, y, \frac{1}{4})$ in the unit cell. Several initial phasing sets, for data between 50 and 25 Å resolution, derived from electron microscopy image reconstructions and spherical shell models, failed to give useful phase extensions. In hindsight, these failures were probably caused by a lack of knowledge of the precise particle position. An insightful observation (J. E. Johnson, personal communication) was made that the overall external shape, appearance and size of ΦX174 was rather similar to that of cowpea mosaic virus (CpMV). As the atomic structure of CpMV was known[63], it could be used as a search model at 12 Å resolution. This yielded a particle position at (0.2440, 0.2500, 0.2480) which was 2 Å away from any previously used centre. By contrast, the structure factors derived from the ΦX174 electron microscopy reconstruction had essentially zero amplitude beyond 25 Å resolution. Phases were then slowly extended by averaging the 60 noncrystallographically related electron densities in the viral envelope (see, for instance, ref. 64). The particle centre, redetermined several times during the phase extension process, refined to a final position of (0.2505, 0.2500, 0.2505). The markedly nonspherical particle envelope was

also redetermined at regular intervals[65]. Some heavy atom derivative data had been collected (R.M., D.X., P.W., L.L.I. and M.G.R., manuscript in preparation). Difference Fouriers at various resolutions showed that the Babinet opposite solution had been obtained[41,56]. The phases were, therefore, flipped ($\alpha \rightarrow \pi + \alpha$) before plotting the final electron density map. The CpMV phasing model had not been used in the usual molecular replacement manner to solve the ΦX174 structures as the atomic structures of CpMV and ΦX174 are almost entirely different. The CpMV model served only to obtain a set of initial, very low resolution, structure factors. These structure amplitudes were used in an *R*-factor search for the particle centre that had phases modulated by the envelope shape but not its contents.

TABLE 3

Number of amino acid contacts of reference subunit with other protein subunits or DNA

	Reference subunits				
	F_{barrel}*	**F_{else}†**	**G**	**J**	**DNA**
F_{barrel}	7(F–G turn)	13	7	1	8
F_{else}	11(βD,βG)	78 (44 to E–F insertion) (24 H–I insertion)	5	8	6
G	4 (F–G turn) (D–E turn)	4 (H–I insertion)	56	0	0
J	1	17 (E–F insertion)	0	0	2
DNA	21 (8 βF)	13	0	2	—

Contacts are defined as any amino acid which approaches another amino acid to within 3.8 Å.

*F_{barrel} is defined as all amino acids in the F protein β barrel excluding the E–F and H–I insertions.

†F_{else} is defined as all other amino acids excluded from F_{barrel}.

TABLE 4

Sites of mutations

Location of ts mutations that may block folding or assembly*	
G Protein	
(1) On spike surface	S39→L(βB), G52→S(βC), P100→L(βE), P100→S(βE), S148→F(βH)
(2) On spike internal cavity	T4→A (amino end), H116→Y(βF)
F Protein	
(1) On exterior surface of the capsid	A91→V (E-F loop), P120→S (E–F loop), P148→H (E–F loop), P175→L (E–F loop), G337→S (H–I loop), L384→I (H–I loop)
(2) On interior surface of the capsid	H16→Y(βB), M43→I(βD)
Location of cs mutations that may block eclipse/DNA injection	
F Protein	R216→H (near J), R216 → C (near J), R233 → C (near J) in interior cleft in close proximity to J

* L.L.I. and N.L.I., manuscript in preparation.