The crystal structure of bacteriophage Qβ at 3.5 Å resolution

Roshan Golmohammadi[1], Kerstin Fridborg[1], Maija Bundule[2], Karin Valegård[1] and Lars Liljas[1]*

Background: The capsid protein subunits of small RNA bacteriophages form a T=3 particle upon assembly and RNA encapsidation. Dimers of the capsid protein repress translation of the replicase gene product by binding to the ribosome binding site and this interaction is believed to initiate RNA encapsidation. We have determined the crystal structure of phage Qβ with the aim of clarifying which factors are the most important for particle assembly and RNA interaction in the small phages.

Results: The crystal structure of bacteriophage Qβ determined at 3.5 Å resolution shows that the capsid is stabilized by disulfide bonds on each side of the flexible loops that are situated around the fivefold and quasi-sixfold axes. As in other small RNA phages, the protein capsid is constructed from subunits which associate into dimers. A contiguous ten-stranded antiparallel β sheet facing the RNA is formed in the dimer. The disulfide bonds lock the constituent dimers of the capsid covalently in the T=3 lattice.

Conclusions: The unusual stability of the Qβ particle is due to the tight dimer interactions and the disulfide bonds linking each dimer covalently to the rest of the capsid. A comparison with the structure of the related phage MS2 shows that although the fold of the Qβ coat protein is very similar, the details of the protein–protein interactions are completely different. The most conserved region of the protein is at the surface, which, in MS2, is involved in RNA binding.

Addresses: [1]Department of Molecular Biology, Uppsala University, BMC, Box 590, S-751 24 Uppsala, Sweden and [2]Biomedical Research and Study Centre, University of Latvia, A. Kirhensteina 1, Riga, Latvia.

*Corresponding author. E-mail: lars@xray.bmc.uu.se

Key words: crystal structure, MS2, Qβ

Received: 30 Jan 96
Revisions requested: 14 Feb 1996
Revisions received: 16 Feb 1996
Accepted: 15 Mar 1996

Structure 15 May 1996, 4:543–554

Introduction

Bacteriophage Qβ is one of many small RNA bacteriophages infecting *Escherichia coli*. The molecular biology of this group of viruses has been reviewed by Fiers [1] and van Duin [2]. Small RNA bacteriophages are classified as belonging to one of four groups, (I, II, III and IV). Qβ belongs to group III. It has an icosahedrally symmetric capsid with T=3 quasi-symmetry [3] consisting mainly of 180 copies of the coat protein. It has been shown that the coat protein subunits of Qβ are linked together by disulfide bonds in covalent pentamers and hexamers with a stoichiometric ratio of 12:20, which is consistent with icosahedral symmetry [4]. Most of the coat protein subunits have 132 residues, but three to five subunits are extended an additional 196 residues by suppression of the termination of translation. The coat protein subunits in this form are denoted A1. In addition to this, the capsid contains, or is accompanied by, one copy of the maturation protein, A2 (corresponding to the A protein in MS2), which participates in attachment of the phage to bacterial pili and in lysis of the bacteria [5]. In group I and II small RNA bacteriophages lysis is performed by a separate lysis protein encoded in another reading frame, partly overlapping the genes of the coat protein and replicase. The genome of Qβ consists of one positive sense single-stranded RNA molecule of 4220 bases [6].

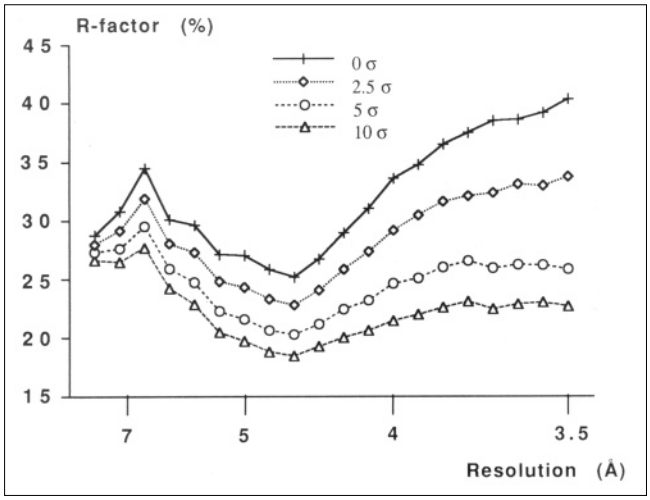
The crystal structures of two group I bacteriophages, MS2 and fr, have been determined [7–9] and the structure of a dimeric mutant form of the MS2 capsid protein is known [10]. The crystal structure of the Qβ particle shows that the capsid protein assumes three different conformations (denoted A, B and C [7]) and forms two different dimer types with the same topology as those present in MS2. The capsid structure provides us with an opportunity to try to determine which structural elements are not necessary for the overall folding and packing by comparing it with the previously determined structures, with which it has less than 25% sequence identity.

Results and discussion

Quality of the model

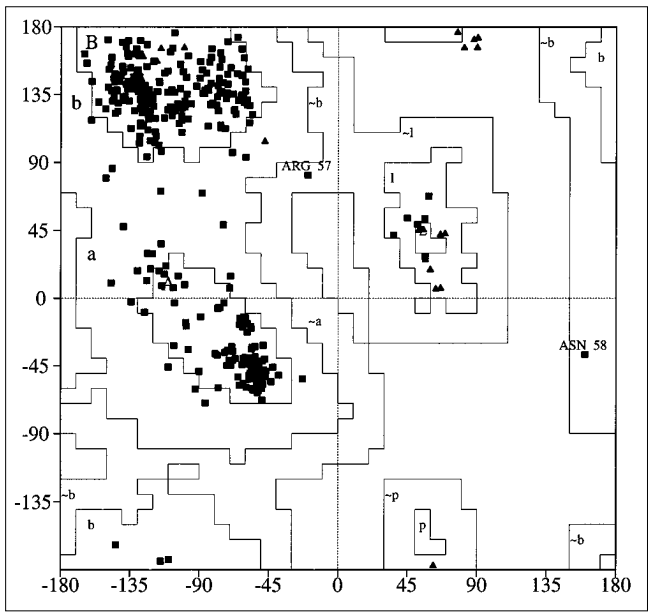
The model of the three independent polypeptide chains contains 378 amino acid residues and has an R-factor of 0.307 for the 328 697 reflections between 15.0 and 3.5 Å resolution used in the refinement. All measured reflections were used in the refinement. The high R-factor is partly due to poor data resulting from difficulties in processing images that had diffuse streaks of non-integer periodicity in the direction of the I index of the diffraction pattern (see Materials and methods section). In addition, a great majority of reflections were weak. The scaling R-factor is accordingly as high as 0.201. Disordered regions

Figure 1



R-factor versus resolution for the final model. Data for a number of different σ cut-offs are shown; no cut-off (329 000 reflections), 2.5σ (250 000 reflections), 5σ (183 000 reflections) and 10σ (143 000 reflections).

Figure 2



Ramachandran plot for the final model. Triangles denote glycine residues and boxes denote non-glycine residues. The two residues found in the 'generously allowed regions' are ArgB57 and AsnB58.

covering 18 amino acids could not be modeled and six other residues were not modeled beyond the Cβ atom. No solvent atoms were included in the model. The R-factor versus resolution for different choices of σ cut-off is presented in Figure 1. With a 5σ cut-off the R-factor was 0.245, and with a 10σ cut-off the value was 0.227.

The density was interpretable for most of the structure with the exception of two loops in the A subunit and two loops in the C subunit. Residues 56–60 in the EF loop of both subunits are located on the inside of the capsid and have similar densities but the density is weaker than it is for the rest of the structure. Although there is a certain degree of disorder, which is likely to be caused by interaction with the non-symmetric RNA, there appears to be two main alternative conformations of the EF loop. Four residues in the FG loop have been left out of the models of both the A and C subunits as it was not possible to unequivocally determine which part of the weak density belonged to the side chains and which part to the main chain. These residues (76–79) also have similar densities in both the A and C subunits, but suffer from what appears to be a smearing of the density due to radial movement.

The refinement procedure did not easily permit the inclusion of disulfide bonds between symmetry-related subunits and consequently there is a clearly visible deviation from ideality. The density at the cysteine side chains is very strong but unfortunately lacking in definition. Therefore, it has not been possible to identify the exact positions of the disulfide sulfur atoms

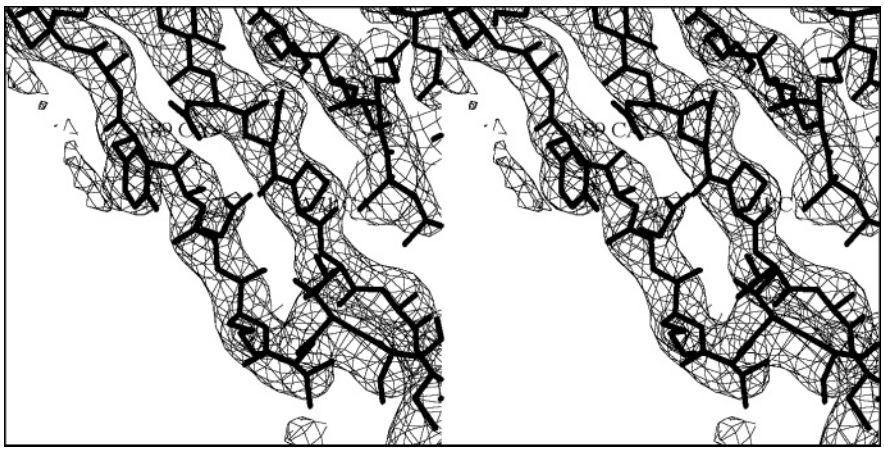
The root mean square (rms) deviation from ideality for bond lengths and angles was 0.007 Å and 1.5° respectively. A Ramachandran diagram is shown for all residues included in the model (Fig. 2). All non-glycine residues fall in energetically allowed regions. The model quality was analyzed with the program PROCHECK [11] and found to be reasonable for all the properties tested. Representative electron density covering residues in the F and G strands of the A subunit (Fig. 3) showed that the model side chains are in register with the correct side chain density in the β sheet on both sides of the FG loop.

Structure of the capsid

The conformations of the three structurally independent subunits A, B, and C is similar, with small dissimilarities arising from differences in packing interactions within the capsid. Figure 4a shows a Cα trace of the polypeptide chains of the three subunits. In Figure 4b a B subunit is chosen to illustrate the general fold of a subunit within the capsid. From the N terminus, two antiparallel β strands (A and B) fold over the triangular five-stranded antiparallel β sheet (strands C–G) which is followed by two contiguous α helices at the C-terminal end. The rms deviation between the subunit conformations is 0.7 Å for either 122 or 123 Cα, atoms depending on which two subunits are compared. This fold is not likely to be especially stable for the monomeric polypeptide; rather, it is believed to be obtained as a consequence of dimerization. The subunits form two types of dimer (AB and CC) in which each subunit contributes

Figure 3

Electron density stereodiagram for the F and G β strands of the A subunit. The Cα atom of residue A89 is marked.



five β strands to a ten-stranded antiparallel β sheet. Figure 4c shows three dimers organised around the quasi-threefold axis. The conformations of the ordered parts of the two types of dimer present in the capsid are similar to each other, as are the conformations of the individual subunits; both dimers and subunits show about 0.7 Å rms deviation for superimposed Cα coordinates. In both dimer types, the AB loop of one subunit embraces the first α helix of the other subunit and vice versa. This configuration of the AB loops is likely to confer added stability on the capsid. The core of the dimer between the helices and the β sheet largely contains hydrophobic side chains.

The capsid protein contains two cysteines, one at position 74 and one at position 80. These cysteines form disulfide bonds connecting adjacent FG loops. The dimers are held together in the icosahedral lattice by the disulfide bonds located around the fivefold and threefold axes. These bonds covalently connect five subunits at the fivefold axis and six subunits at the threefold axis. This finding is consistent with experiments done by Takamatsu and Iso [4], which showed that subunits are covalently linked in arrangements of five and six with the stoichiometric ratio of 12:20. At the threefold axis CysA80 (Cys80 of the A subunit) forms a disulfide bond with CysC′74, and CysC80 forms a disulfide bond with CysA′74. CysB80 is bonded to CysB′74 at the fivefold axis. The Cα atoms of CysA74 and CysC74 are located 16–17 Å from the threefold axis and for CysA80 and CysC80 the distance is 14 Å. The cysteines are closer to the fivefold axis: CysB74 Cα is positioned roughly 12.5 Å from the axis and CysB80 Cα is about 2.5 Å closer.

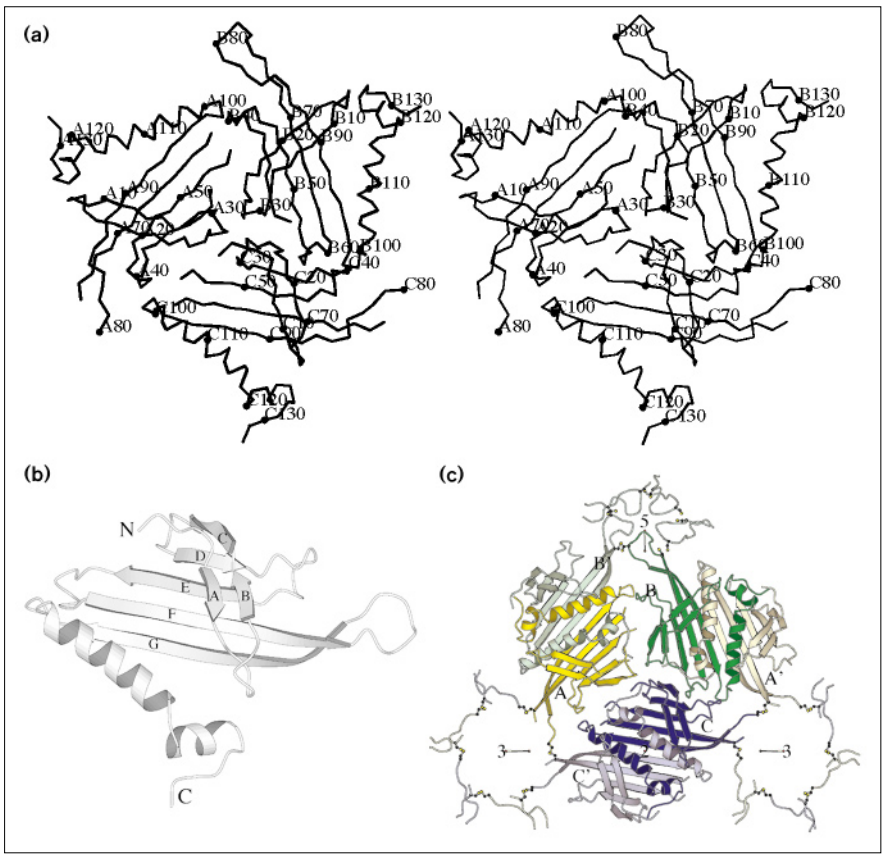
The model has two unique interdimer salt bridges and 13 unique hydrogen bonds in the ordered parts of the structure. The C strands of the three different subunits interact with the C-terminal regions of subunits in symmetry related protomers near the quasi-threefold axis. The interacting parts are largely similar, with differences only in the finer details. In β strand C of the A subunit backbone, nitrogen A23 and oxygen A25 hydrogen bond to AsnC′129 Oδ1 and Nδ2, respectively, in the C-terminal region of subunit C. In other subunit types these residues in β-strand C are not involved in similar interactions, but all three subunit types have other interactions between these regions. Interaction between Arg24 and the C-terminal region of a subunit in an adjacent dimer is common to all subunits. In the A subunit, the Arg24 Nϵ hydrogen bonds to GlnC′127 O. In the B and C subunits this arginine makes different contacts: in the B subunit Arg24 Nϵ binds AsnB′129 Oδ1 and the Nη2 atom interacts with GlnB′127 O; in the C subunit the arginine makes similar contacts with the A′ subunit.

Interdimer interactions are sparse between CD loops at the quasi-threefold axis. The CD loop in the A subunit may have one atom within van der Waals binding range of the CD loop of the B subunit, and one atom within range of the CD loop of the C subunit. No hydrogen bonds have been identified between these structural elements.

Only one set of three quasi-equivalent interdimer hydrogen bonds has been identified: N A43, N B43 and N C43 in the DE loops bind O C98, O A98 and O B98 respectively. Atoms in and around residues in the DE loops and the turns, which connect the G strands to the A helices in the individual subunits, also make hydrophobic van der Waals contacts.

Lys2 Nζ in the C subunit forms a salt bridge with the C-terminal carboxylate of the A′ subunit. In the A and B subunits the N termini are located farther from the quasi-threefold axis. In the C subunit, Arg47 Nη2 (in the E strand) hydrogen bonds to Tyr62 Oη in the F strand of the B subunit; this is the only interdimer hydrogen bond

Figure 4



Structure of the Qβ subunits.
(a) Stereo drawing showing the three structurally independent subunits A, B, and C. (b) Schematic drawing of a B subunit. The strands are denoted A–G. The secondary structure is defined by DSSP [35]. Strand A (residues 17–20), B (21–27), C (32–36), D (47–53), E (62–74), F (83–96). The two helices comprise residues 102–117 and 119–126. (c) Schematic drawing showing the subunit arrangement in Qβ. The three subunits A (yellow), B (green), and C (blue) representing an icosahedral asymmetric unit are shown in dark colors, and the three symmetry-related subunits A′, B′ and C′ completing the dimers are shown in light colors. All the loops forming the interactions at the fivefold (5) and threefold (3) axes are included. The view is from the outside of the particle down a twofold axis. (Drawings made using the program MOLSCRIPT [36].)

between β strands. The second salt bridge between dimers is between Asp C81 Oδ1, in the N-terminal end of the FG loop, and Arg A′86 Nη2 in the G strand. In the A subunit Oη of Tyr99 (located in the turn connecting strand G and helix A) hydrogen bonds to N B84 in strand G and in the C subunit it forms a hydrogen bond with O A82 in the FG loop.

The A1 protein has not been located in the electron density. This is partly due to the averaging procedure employed and to the different orientations that a capsid, which is asymmetric with respect to the A1 protein can assume in the crystal. In addition, the A1 protein, could be occupying positions for either one of the three subunit types: A, B or C. Evidence supporting this has been presented in [4], where the A1 protein was reported to be identified in both hexamers and pentamers of subunits held together by intra hexamer or intra pentamer disulfide bonds. In the crystal structure the C-termini of the subunits in the capsid, where the read-through extension would be expected, are located at the exterior part of the capsid, but are shielded from the exterior by the N termini of the subunits with which they form dimers. A rearrangement of at least these N termini is necessary to accommodate the extra domain.

Comparison with MS2

Figure 5 shows a sequence alignment of Qβ, MS2 and several other small RNA phages. Seventeen positions are conserved in the sequences of all four groups of *E. coli* phages, and five positions are conserved in all small RNA phages. A comparison of the Qβ and MS2 structures shows that the Qβ capsid has a radius 5 Å higher than that of the MS2 capsid. Qβ also has about 20% more genomic RNA. Figure 6a shows the icosahedral asymmetric units of Qβ and MS2 as they would appear if the two particles had been superimposed, and Figure 6b shows superimposed Qβ and MS2 dimers. Apart from interactions within secondary structural elements, few of the interactions within and between dimers are structurally conserved between Qβ and MS2. Despite the lack of conservation of the detailed interactions, the phage particles are stabilized in a very similar way. The conformation of the secondary structural elements is conserved between Qβ and MS2 and the differences between the phages are concentrated in loop regions and in the C and N termini. The β sheet is very similar in both phages. The greater radius in Qβ is manifested in part by all the strands in the β sheet being situated farther away from the quasi-threefold axis than in MS2. The difference in radius thus appears to be caused by translation of the dimers along the twofold or quasi-twofold axes.

Figure 5

Sequence alignments of a number of small RNA phages. The sequences of the *E. coli* phages MS2 [37], fr [38], GA [39], Qβ [40] and SP [41] as well as the *Pseudomonas* phages PRR1 [42] and PP7 [43] are shown. MS2 and fr belongs to group I, GA to group II, Qβ to group III and SP to group IV. The top row shows the position of secondary structural elements in Qβ. Conserved amino acid residues in the *E. coli* phages and in all phages are marked in rows *E. coli* and All, respectively. This alignment differs from earlier ones in that the Qβ and SP EF loops are 2 residues longer. In addition the FG loop is two residues shorter than in MS2. The alignment must be regarded as tentative for PRR1 and PP7, since the sequence similarity is very low.

```
                    aaaa aa     bbb bbccccccc     dddddd          eee
MS2       1 ASNFTQFVLV DNGG..TGDV TVAPSNFA.. NGVAEWISS. NSRSQAYKVT
fr        1 ASNFEEFVLV DNGG..TGDV KVAPSNFA.. NGVAEWISS. NSRSQAYKVT
GA        1 .ATLRSFVLV DNGG..TGNV TVVPVSNA.. NGVAEWLSN. NSRSQAYRVT
Qβ        1 .AKLETVTLG NIGKDGKQTL VLNPRGVNPT NGVASLSQAG AVPALEKRVT
SP        1 .AKLNQVTLS KIGKNGNQTL TLTPRGVNPT NGVASLSEAG AVPALEKRVT
E. coli             L    G           P        NGVA              +VT
PRR1      1 .AQLQNLVLK DREAT.PNDH TFVPRDIRD. NVGEVVEST. GVPIGESRFT
PP7       1 ...SKTIVLS VGEAT.RTLT EIQST..... ADRQIFEEKV GPLVGRLRLT
All               L                                             + T

            eee          ffffffff fffff           gggg gggggggg
MS2      46 CSVRQSSA.. QNRKYTIKVE VPKVATQTVG .GVELPVAAW RSYLNMELTI
fr       46 CSVRQSSA.. NNRKYTVKVE VPKVATQVQG .GVELPVAAW RSYMNMELTI
GA       46 ASYRASGA.. DKRKYAIKLE VPKIVTQVVN .GVELPGSAW KAYASIDLTI
Qβ       50 VSVSQPSRNR KNYKVQVKIQ NPTACTANG. ..SCDPSVTR QAYADVTFSF
SP       50 VSVAQPSRNR KNFKVQIKLQ NPTACTRDA. ...CDPSVTR SAFADVTLSF
E. coli     S           K   K    P            P
PRR1     47 ISLRKTSN.. GRYKSTLKLV VPVVQSQTVN .GIVTPVVVR TSYVTVDFDY
PP7      42 ASLRQNGAK. TAYRVNLKLD QADVVDCSTS VCGELPKVRY TQVWSHDVTI
All         S            +   K                    P

               AAAAA AAAAAAAAAA A   BBBBBB BB
MS2      93 PIFATNSDCE LIVKAMQGLL KD.GNPIPSA IAANSGIY.
fr       93 PVFATNDDCA LIVKALQGTF KT.GNPIATA IAANSGIY.
GA       93 PIFAATDDVT VISKSLAGLF KV GNPIAEA ISSQSGFYA
Qβ       97 TQYSTDEERA FVRTELAALL AS..PLLIDA IDQLNPAY.
SP       96 TSYSTDEERA LIRTELAALL AD..PLIVDA IDNLNPAY.
E. coli         -                         A I      Y
PRR1     94 DARSTTKERN NFVGMIADAL KADLMLVHDT IVNLQGVY.
PP7      91 VANSTEASRK SLYDLTKSLV ATSQVEDLVV NLVPLGR..
All
```

The AB loops of each Qβ subunit cover part of helix A in the subunit with which it forms a dimer. In MS2 the AB loops protrude radially outwards from the particle. In Qβ the CD loop region, close to the quasi-threefold axis, contains two extra residues. The presence of these residues enables the CD loops to reach further in towards the capsid center than is possible in MS2. The relationship between the CD loops in the A, B and C subunits is largely the same in Qβ and MS2; the B subunit is located at the highest radius, followed by the A subunit. A novel feature in Qβ is that proline residues A28, B28 and C28 have their prolyl ring roughly perpendicular to the quasi-threefold axis and, due to their different radii in the particle, appear as steps forming one turn of a spiral staircase.

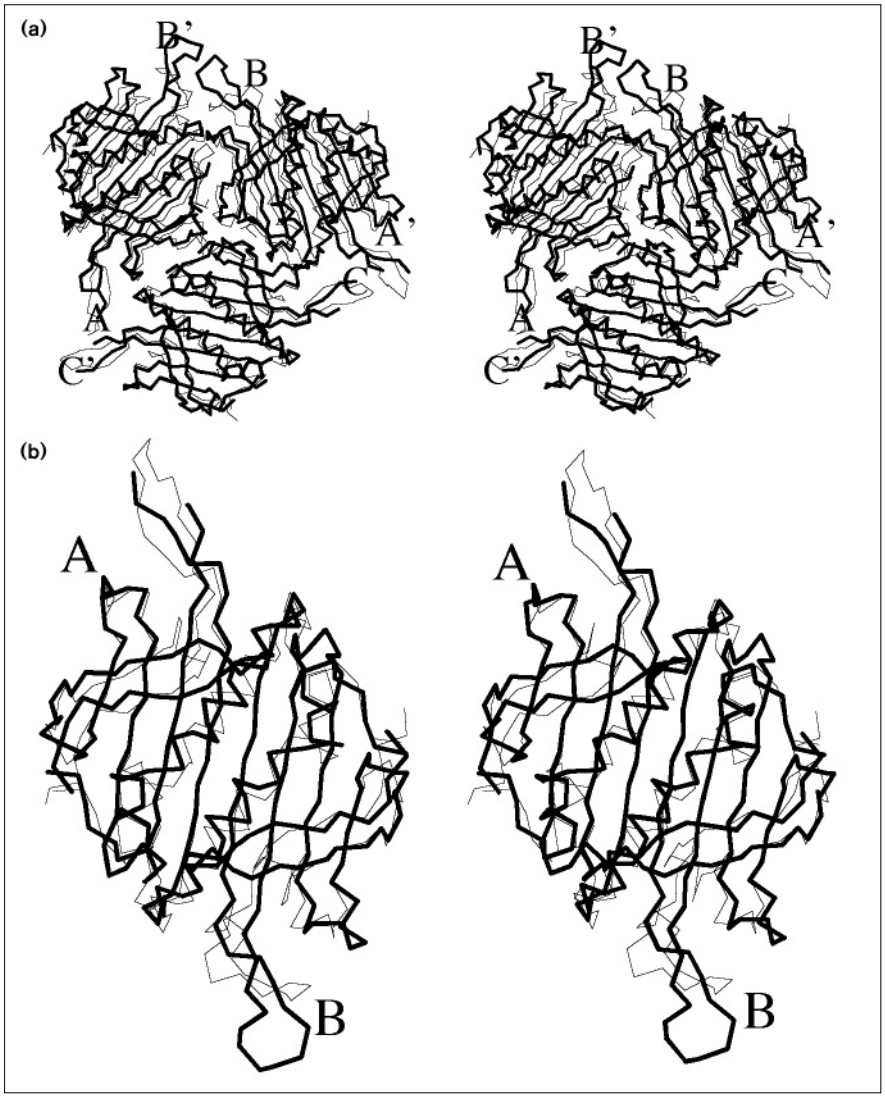
The DE loops of Qβ have one more residue than in MS2 and in the B and C subunits the loops are more extended. To a small degree these loops can, therefore, be viewed as spacers, placing the N-terminal ends of the A helices at a greater distance and in this way separating the two interacting dimers. The half of each DE loop which is closest to the E strand retains similarity to the corresponding part in MS2. There are only three interdimer interactions which are related to one another by quasi-threefold symmetry. These contacts take place between the DE loop of each and every subunit and the turn connecting strand G and helix A in an adjacent subunit. The three quasi-equivalent interdimer interactions occur between the atoms N A43 and O C98, N B43 and O A98 and between N C43 and O B98; these interactions are structurally conserved between the phages. In MS2 these contacts are formed by the main chain nitrogen of residue 39 and the peptide oxygen of residue 94.

The EF loop is two residues longer in Qβ than in MS2. The density for this loop could not be modelled in the A and C subunits, but the weak density in all subunits indicates that it is bent down towards the center of the particle. In the B subunit this loop was built as an alanine chain trace. In MS2 the loop has a β hairpin turn in the plane of the β sheet.

The greatest difference in the subunit by subunit comparison of the two phages, is by far the disulfide-linked FG loops in the B subunits of Qβ (see below for details). The FG loops of Qβ extends towards the fivefold axis, unlike the contracted conformation in MS2, which leaves a larger hole at the axis. Part of the FG loops at the fivefold axis have not been modeled, but it is still clear that the density for these loops does not follow the same course as in MS2.

Helix A is one residue longer in Qβ but the fold of the C-terminal regions is very similar in both phages. In Qβ

Figure 6



Comparison of Qβ and MS2. (a) Stereo drawing showing three dimers of Qβ and MS2. View down the twofold axis of the icosahedron. The MS2 subunits (thin lines) are found at a lower radius. (b) Stereo drawing showing the superposition of an AB dimer of MS2 (thin lines) on Qβ (thick lines). The superposition was based on the Cα coordinates of the residues in the ten-stranded sheet.

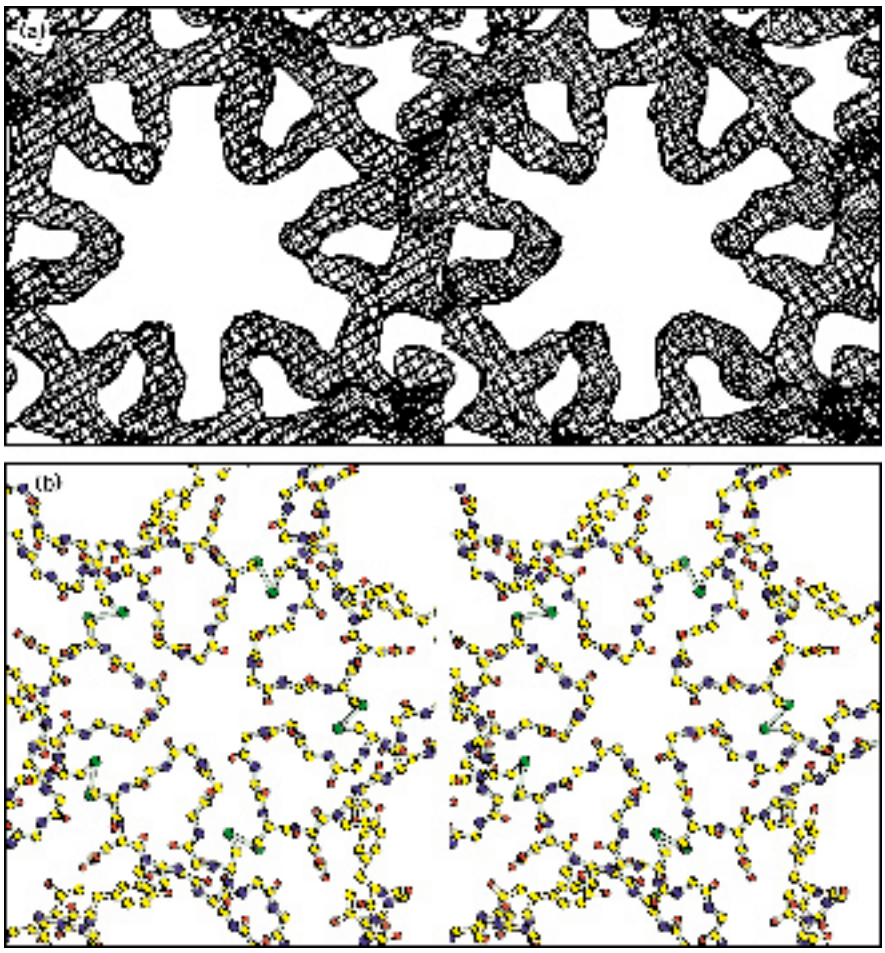
the carboxyl terminus of the A subunit forms a hydrogen bond with the hydroxyl group of Tyr B′132. The Qβ C terminus of the C subunit interacts with Leu3 N in the twofold symmetry related C′ subunit. No interdimer interactions between C termini are present in the MS2 structure. In both phages the C termini of all three subunits make intradimer contacts with residues in the N-terminal regions of the symmetry related subunits. In the Qβ B monomer the helix is positioned as if rotated around a pivot point at the N-terminal end, bringing the C terminus a little closer to the nearby quasi-threefold axis than it is in MS2.

The rms deviation of Cα positions between individual subunits in MS2 and Qβ is 1.5 Å for the A subunits (108 atoms), 1.6 Å for the B subunits (101 atoms) and 1.6 Å for the C subunits (107 atoms). When only Cα atoms from the more conserved β sheets are used, the rms deviation is 0.9 Å (45 atoms) between A subunits, 1.1 Å (43 atoms) between B subunits and 1.0 Å (45 atoms) between C subunits.

The FG loop

The FG loops are connected around the threefold and fivefold axes by disulfide bridges. The beginning of each FG loop is connected with the end of the adjacent FG loop and in this way forms fivefold or quasi-sixfold covalent hubs around the icosahedral fivefold and threefold axes. This locks the dimers covalently in the capsid. Chemical evidence for a T=3 capsid has been presented by Takamatsu and Iso [4], showing approximately 12 pentamers and 20 hexamers to be stabilized by intermolecular disulfide bonds. This is consistent with near complete engagement of the juxtaposed cysteine residues poised for participation in disulfide bonds. The uncoupling of

Figure 7



Conformation of the FG loops.
(a) Stereodiagram of the electron density around the threefold axis showing the disulphide bridges between the FG loops. Six FG loops, three from each of the A and C subunits, extend towards the axis. The disulphide bridges are seen as thick connections between these loops. There is a high degree of similarity between the densities of the FG loops in the A and C subunits.
(b) Conformation of the FG loops around the five-fold axis. Stereo drawing of residues B72 to B84 and A97 to A100. Carbon atoms are yellow, nitrogen atoms blue, oxygen atoms red and sulfur atoms green.

one initial dimer from the capsid would require the breaking of four disulfide bonds, two at each FG loop. This stabilization might compensate for the instability and movement at the apical position of the FG loops which is manifested in the apparent disorder of these regions in the A and C subunits. Figure 7a shows the electron density around the threefold axis. In the B subunit the density accounts for all atoms with the exception of those in the Asn 77 side chain, which are not included in the model. The temperature factors of atoms in the FG loop are very high.

The two cysteine residues involved in disulfide bond formation are present in all known sequences of coliphage belonging to group III and IV and also in the *Pseudomonas* phage PP7. This conservation indicates that this feature is more than an aberration in small RNA phages. The pore diameter of the threefold axis is 13–14 Å, as estimated from the electron density. The pore at the threefold axis is 3–4 Å wider in MS2. The FG loop in Qβ is two residues shorter than the FG loop of MS2. The pore diameter at the fivefold axis is 7 Å (Fig. 7b), which is considerably smaller than the corresponding 16 Å pore in MS2 [7]. With these pore sizes it is inconceivable that the intact folded A2 protein (about 420 residues) could pass through either pore. The A2 protein appears to be responsible for the binding of particles to bacterial pili during infection [12].

The covalently cross-linked configuration might not be the form that initially assembles inside the *E. coli* cell. This configuration might be assumed in a process of transient unfolding and folding of an MS2-like initial conformation, during which two moving loops interact and form a disulfide bond. These two loops would then be committed to the new conformation, increasing their likelihood of forming disulfide bonds with adjacent uncommitted loops would increase. In the case of the B subunit, the extended loop conformation seen in the crystal structure would exist in a dynamic equilibrium with a contracted conformation, similar to the corresponding FG loop in MS2. Initially only a small percentage of loops would have the extended conformation. However, the equilibrium would be shifted away from the contracted form as the disulfide

bonds of the extended form become engaged by neighbouring extended FG loops.

Assembly

In small RNA bacteriophages it has been shown that insertions of up to 25 amino acids can be tolerated in the AB loop [13]. It is also known that residue 51 can be replaced by up to nine other consecutive residues without complete loss of particle formation [14]. In fr it is also possible to delete up to five amino acids in the FG loop along with one in the F strand and still observe the assembly of dimers into capsids [14]. The regions of Qβ that show disorder or movement coincide with segments that can be altered in other phages.

One possibility is that these regions do not need to have special conformations in order to exist in a T=3 capsid; instead, they may need to adopt specific conformations when they are being packed. In the case of MS2, we proposed previously that the dimers, in solution prior to assembly, have their FG loops in one of two possible conformations: one contracted as in the B subunit and one extended as in A or C [8]. This would give rise to three possible types of dimers, AB, AA (or CC) and BB. The BB dimer would serve no obvious purpose, but might explain the small fraction of T=1 particles observed [15,16]. Initially this might also be the case for Qβ, the structure present in the crystal only being adopted after assembly. A case where each dimer to be added to the growing capsid first forms a disulfide bond with the correct cysteine prior to other protein–protein recognition is less conceivable.

The *cis*-proline ProB78 (corresponding to ProB82 in Qβ), previously suggested to be a switch regulating the equilibrium between AB and CC dimers in MS2 [8], has been modeled in Qβ as a *trans*-proline in all three subunits. An MS2 mutant Pro78→Asn forms capsids without assuming a *cis*-peptide conformation in that position [17], showing that the *cis*-proline is not necessary for capsid formation. This again raises the question of why this proline is one of the very few strictly conserved residues in all small RNA bacteriophages. Proline isomerization is known to be a rate limiting step in the folding of some proteins [18,19]. It may be possible that a limitation in assembly rate due to isomerization can serve to decrease capsid protein cooperativity in RNA binding, by limiting the rate at which AB dimers are made available for consumption by a growing capsid.

The RNA is known to promote capsid formation [20,21]. A limitation in assembly rate, due to a reduction in available AB dimers, could be advantageous if the effect of the RNA is largely due to charge neutralization and if the competitive advantage of the RNA lies mainly within one short operator segment (3500–4000 nucleotides) [21,22]. Whether or not the MS2 mutant (Pro78→Asn) has a greater propensity to package non-specific RNA than the wild type capsid is not known.

Conserved operator RNA binding surface

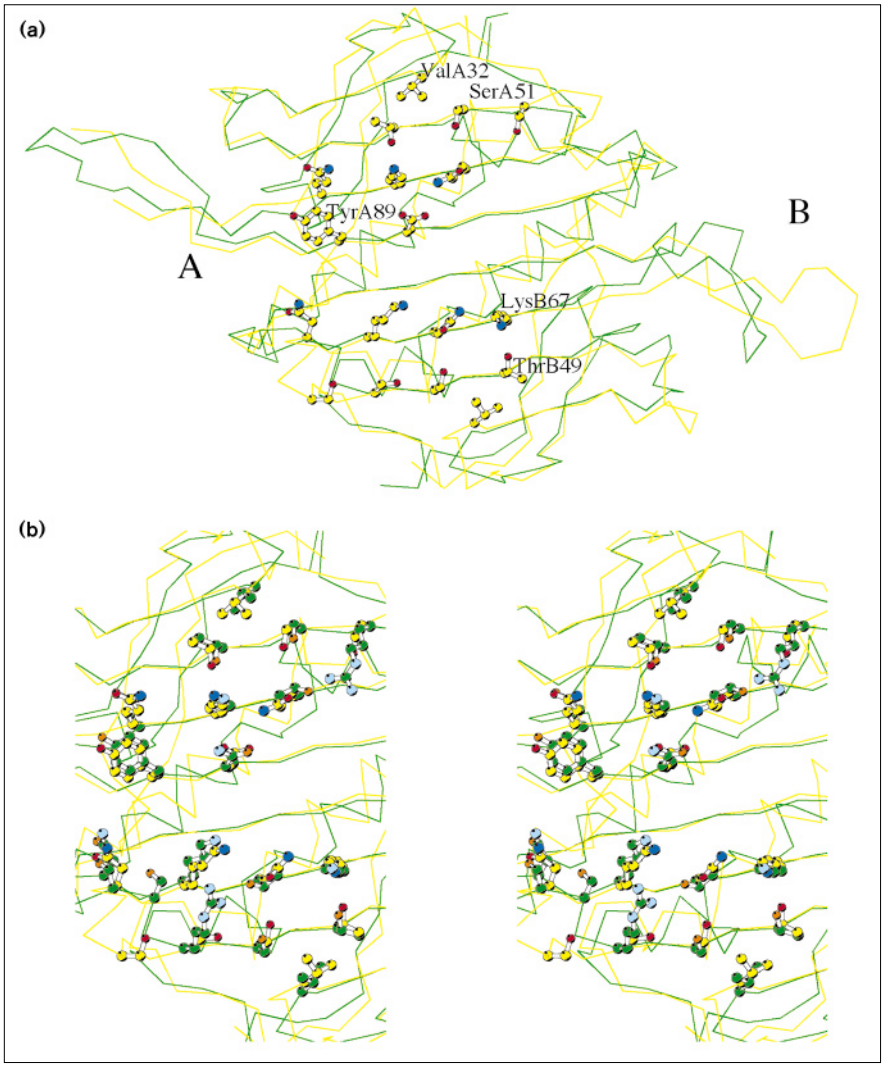
Small RNA bacteriophage coat proteins bind to a specific site on the viral RNA. This binding has two functions [23]. Firstly, since the binding site is an operator RNA stem-loop, and includes the initiation codon of the replicase, the coat protein acts as a translational repressor of the viral replicase. The second function is that the binding of coat protein dimers to the operator is the starting point of virus assembly. The coat proteins bind analogues of the viral stem-loop with varying affinities depending on the RNA sequence [24]. Normally around 20 nucleotides compose the stem-loop structure and certain base positions have been found to be essential for capsid protein binding. Two adenines which do not form base pairs are required in MS2 for optimal binding. One of them forms a bulge on the 5′ half of the stem, and the other one is found in the loop which has a length four nucleotides. These adenines are also present in the Qβ operator, but the loop has only three nucleotides and the bulged adenine on the 5′ half of the operator may be deleted with only 40% loss of binding [25].

The crystal structure of a complex between the MS2 capsid and a variant of the analogue has been determined [26]. The operator analogue binds to a dimer of the coat protein, and the two essential adenines bind to equivalent sites on the two monomers. There are three conserved residues in the β sheet of all small RNA phages (Fig. 5). These residues, Thr45, Ser47 and Lys61 (MS2 numbering), form a binding site where the adenine base forms hydrogen bonds to the serine and threonine side chains. The adenines bind differently to the two identical sites, which explains why the bulged adenine, but not the adenine in the loop, can be replaced by guanine.

The presumed RNA-binding region in the Qβ capsid is very similar to that of MS2 (Fig. 8). A superposition of an MS2 AB dimer on the corresponding Qβ dimer shows that the atoms interacting with the adenines in MS2 (the side chains of Val29, Thr45, Ser47 and Lys61) are within 1 Å of their equivalent positions in Qβ. In MS2 Ser59 is close to the adenine base but in Qβ this residue is replaced by a glutamine. The presence of this glutamine residue might add to the binding strength of this site by stacking against the base.

In the MS2 operator complex, RNA bases, beginning with one pyrimidine in the loop, are stacked on TyrA85. A hydrogen bond from AsnA87 to O2 of the pyrimidine explains the specificity at this position. The positions of these side chains are similar in Qβ. In Qβ the asparagine residue is replaced by an aspartate, and a similar binding of a pyrimidine is not favoured. It is likely that stacking of bases to the tyrosine occurs also in Qβ, but the fact

Figure 8



A comparison of the RNA-binding sites in MS2 and Qβ. (a) Overview of an AB dimer of Qβ (yellow) and MS2 (green). The side chains of a number of residues in Qβ, corresponding to those interacting with RNA in the repressor–operator complex in MS2 are shown. (b) Stereo drawing of the RNA-binding surface of the dimer. Side chains of both Qβ (yellow) and MS2 (green) are shown. The atom colours for the side chains of Qβ are yellow (carbon), dark blue (nitrogen) and red (oxygen). The corresponding atoms in the MS2 side chains are green, light blue and orange. Side chains included are (Qβ residue in parenthesis) ValA29 (A32), ThrA45 (A49), SerA47 (A51), ArgA49 (SerA53), ThrA59 (GlnA65), LysA61 (A67), GluA63 (GlnA69), TyrA85 (A89), AsnA87 (AspA91), ValB29 (B32), ThrB45 (B49), SerB47 (B51), ArgB49 (SerB53), SerB51 (not included in Qβ), SerB52 (B56), AsnB55 (B61), LysB57 (B63), ThrB59(GlnB65) and LysB61 (B67).

that the loop in the Qβ operator has only three nucleotides probably means that the interactions will be different in this region.

A number of arginine and lysine side chains interact with phosphates in the MS2 operator. Of these, Lys57 and Lys61 are conserved in Qβ, while Arg49 (MS2 numbering) is replaced by a serine. In this region, the nearby Lys57 side chain interacts with Glu89, which is replaced by a threonine in Qβ. This means that the net positive charge at this site is retained. Close to this region, residues in the EF loop of the B subunit (Ser51, Ser52 and Asn55) interact with phosphates of the stem-loop. In Qβ, the EF loop is extended and contains three positively charged side chains, possibly available for interaction with RNA. This is one of the regions for which it was possible to model only the B subunit; the side chains of these arginine and lysine residues are completely disordered.

Biological implications

Small RNA bacteriophages, such as **Qβ**, carry genetic material encoding only four proteins, one of which is the major component of the capsid. The spherical capsid is formed by 90 coat protein dimers. Small RNA phages provide simple systems for the study of capsid assembly and the encapsidation of RNA. Crystal structures of these phages provide a framework for probing the importance of different amino acid residues implicated in these functions by site directed mutagenesis.

Although the sequence identity between **Qβ** and the related bacteriophage MS2 is less than 25%, the structures are similar and the main differences are located in the loop regions. In both types of phage the regions interacting with one another are largely the same, despite large differences in the individual interacting residues. The conserved amino acid residues in the **β**-sheet

regions of the Qβ subunits have a configuration very similar to that of the corresponding residues in MS2. These residues are known to interact with the operator RNA hairpin, an important step in initiating capsid formation. This indicates that Qβ and MS2 use similar mechanisms for operator binding and priming of RNA encapsidation.

The crystal structure of bacteriophage Qβ shows that disulfide bonds covalently link five subunits around the fivefold icosahedral axis and six subunits around the threefold axis. The subunit dimers are held strongly together by a common hydrophobic core. The disulfide bonds link each dimer to the capsid by four covalent bonds. This implies that the Qβ particle should be more stable than related RNA phages which lack covalent intersubunit bonding.

A single loop in the MS2 coat protein subunit forms most of the interactions at the fivefold and threefold axes, and it has previously been suggested to regulate the correct assembly of the particle by adopting different conformations. The conformation of these loops in Qβ differs at the fivefold and threefold axes, but not to the same extent as in MS2, where they have a completely different conformation. The regions surrounding these loops are very similar and no far-ranging peptide-mediated effects caused by these loops appear to govern capsid assembly.

Material and methods

Crystals

The crystals were obtained by vapour diffusion essentially as described in Valegård *et al.* [27]. Crystals were obtained using slightly different conditions but the majority of crystals were grown using 10% PEG 2000, 0.3–0.45 M ammonium sulphate and 0–0.2 M lithium chloride in HEPES buffer (5 g l^{-1}) at pH 7.4. Hanging drops were equilibrated at room temperature against 0.4 M sodium chloride and 50 mM Tris pH 7.4. The crystals were mainly grown from complete phage particles but some recombinant capsid was also used which did not contain the A_2 protein or genomic phage RNA.

Data collection and processing

Diffraction data were collected on film with an Enraf-Nonius Arndt-Wonacott camera, an R-AXIS II C and a MAResearch image-plate detector system at the SERC Synchrotron Radiation Source in Daresbury, UK (station 9.6) and by using MAResearch image-plate systems at the European Synchrotron Radiation Facility (ESRF) in Grenoble (station ID2) and at the EMBL outstation in Hamburg (station D11). Data were processed using the HKL program package (Z Otwinowski and W Minor, personal communication). Scaling and merging were done using the CCP4 program suite [28].

The space group is $C222_1$ as previously described [27], with the cell edges being a=477.8 b=295.2 c=477.8. The degree of mosaicity varied substantially between images and was highest in the c direction. The centered ab layers pack in the c direction with a constant shift and alternating sign in the direction of the a axis, which gives rise to the space group $C222_1$. The disorder is probably due to random errors in the sign of these shifts [27]. The high mosaicity was visibly manifested as streaking within rows of reflections along along the direction of the l index. This we believe has contributed to the high scaling R-factor of 20.1% for 333 576 reflections in the resolution range 40.0 to 3.5 Å, as shown in Table 1. This data set included 80% of all possible reflections from a total of about 800 000 observations used. About 2% of the observed reflection intensities (17 370 measurements) were discarded in the scaling due to bad agreement.

Table 1

Scaling R-factor and degree of completeness as a function of resolution.

Resolution	R-factor	Number of reflections	Degree of completeness	$I/\sigma(I)$	Mulitiplicity
14.62	0.108	4326	77.3	3.4	2.6
10.71	0.089	8171	91.1	5.1	2.9
8.85	0.094	10 398	91.9	5.0	3.0
7.71	0.120	12 148	91.8	4.5	3.0
6.92	0.158	13 738	92.2	3.8	3.1
6.33	0.190	15 067	91.9	3.0	3.0
5.87	0.184	16 215	91.4	3.4	3.0
5.50	0.206	17 213	90.6	3.0	3.0
5.19	0.197	18 058	89.4	3.2	2.9
4.93	0.210	18 868	88.5	3.2	2.9
4.70	0.204	19 688	87.9	3.2	2.9
4.51	0.199	20 408	87.2	3.3	2.9
4.33	0.245	21 137	86.7	2.7	2.9
4.18	0.312	21 712	85.8	2.2	2.7
4.04	0.402	22 301	85.1	1.7	2.7
3.91	0.444	21 288	78.7	1.5	2.5
3.79	0.521	20 010	71.8	1.3	2.3
3.69	0.548	19 201	66.9	1.3	2.1
3.59	0.561	18 249	61.8	1.2	2.0
3.50	0.590	15 380	50.7	1.2	1.8
Total	0.201	333 576	80.5	–	2.7

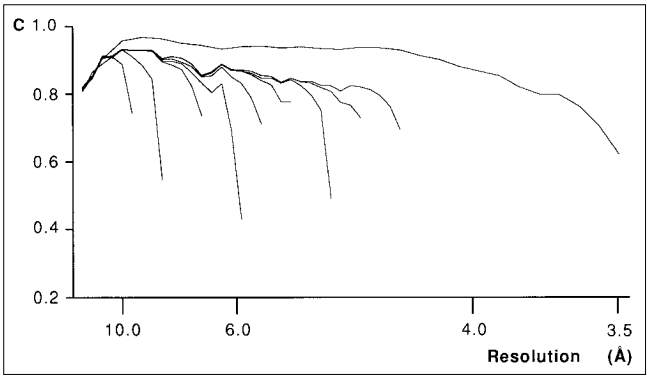
The R-factor column shows R_{merge} at scaling: $R_{merge}=\Sigma_h\Sigma_i|I_h-I_{hi}|/\Sigma_h\Sigma_i I_{hi}$, where I_h is the mean of intensity observations I_{hi} of reflection h.

Structure determination and refinement

The structure was solved by molecular replacement at low resolution with the program X-PLOR [29], using the MS2 capsid structure as a starting model. The MS2 structure was aligned with the icosahedral axes previously obtained from rotation searches [27]. One particle twofold axis coincides with the crystal a axis, and the other twofold axes are parallel to the b and c axes. The particle centre was initially at –0.1542, 0.0, 0.0, as obtained from packing considerations and from the Patterson function. At a later stage, when more high-resolution data were available, an improved position at –0.15464, 0.0, 0.0 was determined.

The MS2 model was used for initial phasing to 10.0 Å resolution. Map averaging using the 30-fold non-crystallographic symmetry was done with RAVE [30]. A mask, defined as a 10 Å radius around each unique capsid protein atom, was used for this averaging and the following 120 cycles of phase extension in steps to 6.0 Å resolution. Map and structure factor calculations were done using the CCP4 program system. A map was computed to 6.2 Å resolution with the resulting phases. This map was not directly interpretable but when contoured at a negative level, density for the expected α helices and β sheets appeared at a radius 5 Å greater than in MS2. The initial model was thus positioned about 5 Å from the correct radius, with the consequence that the procedure resulted in low resolution phases converging at values corresponding to the Babinet opposite of the correct structure. This phenomenon has also occurred in the case of MS2 [7,31] and several other virus structure determinations [32,33].

Figure 9



Correlation coefficients at each 50th cycle of phase refinement at the phase extension from 10–4.5 Å resolution, starting with the MS2 model at the correct radius. The upper curve shows the correlation coefficient after phase refinement at 3.5 Å resolution starting with the Qβ model.

The MS2 A, B and C subunits were moved individually radially outwards in O [34] in the negative density. The subunits were then refined as three individual rigid bodies in X-PLOR using 126 963 reflections between 14.0 and 4.0 Å resolution. This reduced the R-factor from 53.2 to 49.9. The side chains of the model were then truncated beyond Cβ and used to produce a mask with a volume described by a 7 Å radius around each remaining atom. This mask was manually edited in O so as to be more generous in regions that appeared to be different in the Qβ density, such as the AB loop. Stepwise phase extension was now done from 10 Å to 4.5 Å resolution yielding a map with traces of density for some of the side chains in the last three strands of the β sheets and in the α helices. The correlation coefficient versus resolution at some cycles of this procedure is shown in Figure 9.

After manual adjustment, 95% of the residues were included and modelled as alanines. This model was used to extend the phases to 3.8 Å resolution in one step. At this resolution many side chains could be included although most of the N-terminal regions, up to about residue 30, were modeled as alanines, as were other short stretches of inconclusive electron density.

At this stage a more complete data set to 3.5 Å resolution was available, and the model was improved by manual rebuilding in maps based on this data and in model phases refined by map averaging. The model was then used for conventional positional refinement in X-PLOR [29]. All measured reflections were used in the refinement.

After 320 cycles of conventional positional refinement interspersed with manual rebuilding, the R-factor had decreased from 46.3 to 39.4. New maps were produced using the refined model and a mask covering a volume defined by a 4 Å radius around all atoms. The density at the EF and FG loops was relatively weak in all subunits and a series of omit maps and masks altered in the loop regions were used in parallel. The cysteines were constrained during final refinement and in the A and C monomers, the EF loops as well as four residues of the FG loops were omitted. At this stage, the position of the particle centre was refined. A shift of –0.2 Å along the x-axis was obtained from the Patterson function and confirmed by stepwise changes in X-PLOR. This led to a large decrease of about 4% in the R-factor. The cell dimensions were also improved slightly by finding the minimal R-factor by stepwise changes. The correlation coefficient upon map averaging, after 10 cycles of phase refinement using the new parameters and model phases based on the improved parameters, was also much improved. The correlation coefficient versus resolution for this map is showed in Figure 9 (top curve). In spite of the significant change in model R-factor and correlation coefficient for the averaged map, the electron density map after averaging was virtually identical to the earlier map based on the slightly incorrect particle centre. This means that the coordinates of the subunits were essentially identical after this correction. With the optimized parameters, the R-factor converged at 34.4% for reflections between 15 Å and 3.5 Å resolution. Refinement of restrained individual temperature factors reduced the R-factor to 30.7%. The rms deviation of temperature factors for bonded atoms was 2.6 $Å^2$. The coordinates have been deposited in the Protein Data Bank with the accession code 1qbe.

Acknowledgements

We wish to thank Dr Nicola J Stonehouse for discussions. The final refinement of the particle centre, to improve the R-factor, was suggested by Dr David Stuart. This work was supported by the Swedish Natural Science Research Council.

References

1. Fiers, W. (1979). Structure and function of RNA bacteriophages. In *Comprehensive Virology*. (Fraenkel-Conrat, H. & Wagner, R.R., eds), vol. 13, pp. 69–204, Plenum Publishing Corporation, NY.
2. van Duin, J. (1988). Single-stranded RNA bacteriophages. In *The Bacteriophages*, (Calendar, R., ed), vol. 1, pp. 117–167, Plenum Press, NY.
3. Caspar, D.L.D. & Klug, A. (1962). Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* 27, 1–24.
4. Takamatsu, H. & Iso, K. (1982). Chemical evidence for the capsomeric structure of phage Qβ. *Nature* 298, 819–824.
5. Karnik, S. & Billeter, M. (1983). The lysis function of RNA bacteriophage Qβ is mediated by the maturation (A_2) protein. *EMBO. J.* 2, 1521–1526.
6. Mekler, P. (1981). Determination of nucleotide sequences of the bacteriophage Qβ genome: organization and evolution of an RNA virus [PhD Thesis], University of Zurich, Switzerland.
7. Valegård, K., Liljas, L., Fridborg, K. & Unge, T. (1990). The three-dimensional structure of the bacterial virus MS2. *Nature* 345, 36–41.
8. Golmohammadi, R., Valegård, K., Fridborg, K. & Liljas, L. (1993). The refined structure of bacteriophage MS2 at 2.8 Å resolution. *J. Mol. Biol.* 234, 620–639.
9. Liljas, L., Fridborg, K., Valegård, K., Bundule, M. & Pumpens, P. (1994). Crystal structure of bacteriophage fr capsids at 3.5 Å resolution. *J. Mol. Biol*. 244, 279–290.
10. Ni, C.-Z., Syed, R., Kodandapani, R., Wickersham, J., Peabody, D.S. & Ely, K. R. (1995). Crystal structure of the MS2 coat protein dimer: implications for RNA binding and virus assembly. *Structure* 3, 255–263.
11. Laskowski, R.A., MacArthur, M. W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26, 283–291.
12. Roberts, J.W. & Steitz, J.A. (1967). The reconstitution of infective bacteriophage R17. *Proc. Natl. Acad. Sci. USA* 58, 1416–1421.
13. Mastico, R.A., Talbot, S.J. & Stockley, P.G. (1993). Multiple presentation of foreign peptides on the surface of an RNA-free spherical bacteriophage capsid. *J. Gen. Virol.* 74, 541–548.
14. Pushko, P., *et al.*, & Grens, E. (1993). Analysis of RNA phage fr coat protein assembly by insertion, deletion and substitution mutagenesis. *Protein Engin*. 6, 883–891.
15. Knolle, P. & Hohn, T. (1975). Morphogenesis of RNA phages. In *RNA Phages.* (Zinder, N.D., ed), pp. 147–201. Cold Spring Harbor Laboratory, NY.
16. Stonehouse, N. J. & Stockley, P. G. (1993). Effects of amino acid substitutions on the thermal stability of MS2 capsids lacking genomic RNA. *FEBS Lett.* 334, 355–359.
17. Stonehouse, N.J., *et al.*, & Liljas, L. (1996). Crystal structure of MS2 capsids with mutations in the subunit FG loop. *J. Mol. Biol.* 256, 330–339.
18. Brandts, J.F., Halvorson, H.R. & Brennan, M. (1975). Consideration of the possibility that the slow step in protein denaturation reactions is due to *cis-trans* isomerism of proline residues. *Biochemistry* 14, 4953–4963.
19. Kelley, R. F. & Richards, F.M. (1987). Replacement of proline-76 with alanine eliminates the slowest kinetic phase in thioredoxin folding. *Biochemistry* 26, 6765–6774.

20. Hohn, T. (1969). Role of RNA in the assembly process of bacteriophage fr. *J. Mol. Biol.* 43, 191–200.
21. Hung, P.P., Ling, C.M. & Overby, L.R. (1969). Self-assembly of Qβ and MS2 phage particles: Possible function of initiation complexes. *Science* 166, 1638–1640.
22. Beckett, D., Wu, H.-N. & Uhlenbeck, O.C. (1988). Roles of operator and non-operator RNA sequences in bacteriophage R17 capsid assembly. *J. Mol. Biol.* 204, 939–947.
23. Witherell, G.W., Gott, J.M. & Uhlenbeck, O.C. (1991). Specific interaction between RNA phage coat proteins and RNA. *Progr. Nucl. Acid Res. Mol. Biol.* 40, 185–220.
24. Romaniuk, P.J., Lowary, P.T., Wu, H.N., Stormo, G. & Uhlenbeck, O.C. (1987). RNA binding site of R17 coat protein. *Biochemistry* 26, 1563–1568.
25. Witherell, G.W. & Uhlenbeck, O.C. (1989). Specific RNA binding by Qβ coat protein. *Biochemistry* 28, 71–76.
26. Valegård, 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.
27. Valegård, K., Fridborg, K. & Liljas, L. (1994). Crystallization and preliminary X-ray diffraction studies of the bacteriophage Qβ. *Acta Cryst.* D50, 105–109.
28. Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst D* 50, 760–763.
29. Brünger, A.T. (1990). X-PLOR Manual. Yale University, New Haven, CT.
30. Kleywegt, G.J. & Jones, T.A. (1994). Halloween... Masks and Bones. In *From first map to final model.* Proceedings of the CCP4 Study Weekend 1994. (Bailey, S., Hubbard, R., & Waller, D., eds), pp. 59–66. SERC Daresbury Laboratory, Warrington, UK.
31. Valegård, K., Liljas, L., Fridborg, K. & Unge, T. (1991). Structure determination of bacteriophage MS2. *Acta Cryst. B* 47, 949–960.
32. Tsao, J., Chapman, M.S., Wu, H., Agbandje, M., Keller, W., & Rossmann, M.G. (1992). Structure determination of monoclinic canine parvovirus. *Acta Cryst. B* 48, 75–88.
33. McKenna, R., Xia, D., Willingmann, P., Ilag, L.L. & Rossmann, M.G. (1992). Structure determination of the bacteriophage ΦX174. *Acta Cryst. B* 48, 499–511.
34. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Cryst. A* 47, 110–119.
35. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637.
36. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* 24, 946–950.
37. Min Jou, W., Haegeman, G., Ysebaert, M. & Fiers, W. (1972). Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature* 237, 82–88.
38. Wittmann-Liebold, B. & Wittmann, H.G. (1967). Coat proteins of strains of two RNA viruses: Comparison of their amino acid sequences. *Mol. Gen. Genet.* 100, 358–363.
39. Inokuchi, Y., Takahashi, R., Hirose, T., Inayama, S., Jacobson, A.B. & Hirashima, A. (1986). The complete nucleotide sequence for the group II RNA coliphage GA. *J. Biochem.* 99, 1169–1180.
40. Konigsberg, W., Maita, T., Katze, J. & Weber, K. (1970). Amino acid sequence of the Qβ coat protein. *Nature* 227, 271–273.
41. Inokuchi, Y., Jacobson, A.B., Hirose, T., Inayama, S. & Hirashima, A. (1988). Analysis of the complete nucleotide sequence of the group IV RNA coliphage SP. *Nucleic Acids Res.* 16, 6205–6221.
42. Dhaese, P., Vandekerckhove, J.S. & Van Montagu, M.C. (1979). The primary structure of the coat protein of the broad-host-range RNA bacteriophage PRR1. *Eur. J. Biochem.* 94, 375–386.
43. Olsthoorn, R.C.L., Garde, G., Dayhuff, T., Atkins, J.F. & van Duin, J. (1995). Nucleotide sequences of a single-stranded RNA phage from *Pseudomonas aeruginosa*: kinship to coliphages and conservation of regulatory RNA structures. *Virology* 206, 611–625.