ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

VIROLOGY

Virology 310 (2003) 287–297

www.elsevier.com/locate/yviro

The three-dimensional structure of cocksfoot mottle virus at 2.7 Å resolution

Kaspars Tars,[a,*] Andris Zeltins,[b] and Lars Liljas[a]

[a] *Department of Cell and Molecular Biology, Uppsala University, Box 596, S751 24 Uppsala, Sweden*
[b] *Biomedical Research and Study Centre, Ratsupites 1, LV 1067, Riga, Latvia*

Received 3 January 2003; returned to author for revision 21 January 2003; accepted 8 February 2003

Abstract

Cocksfoot mottle virus is a plant virus that belongs to the genus *Sobemovirus*. The structure of the virus has been determined at 2.7 Å resolution. The icosahedral capsid has $T = 3$ quasisymmetry and 180 copies of the coat protein. Except for a couple of stacked bases, the viral RNA is not visible in the electron density map. The coat protein has a jelly-roll β-sandwich fold and its conformation is very similar to that of other sobemoviruses and tobacco necrosis virus. The N-terminal arm of one of the three quasiequivalent subunits is partly ordered and follows the same path in the capsid as the arm in rice yellow mottle virus, another sobemovirus. In other sobemoviruses, the ordered arm follows a different path, but in both cases the arms from three subunits meet and form a similar structure at a threefold axis. A comparison of the structures and sequences of viruses in this family shows that the only conserved parts of the protein–protein interfaces are those that form binding sites for calcium ions. Still, the relative orientations and position of the subunits are mantained.

Keywords: Cocksfoot mottle virus; Sobemovirus; Virus structure; Virus assembly

Introduction

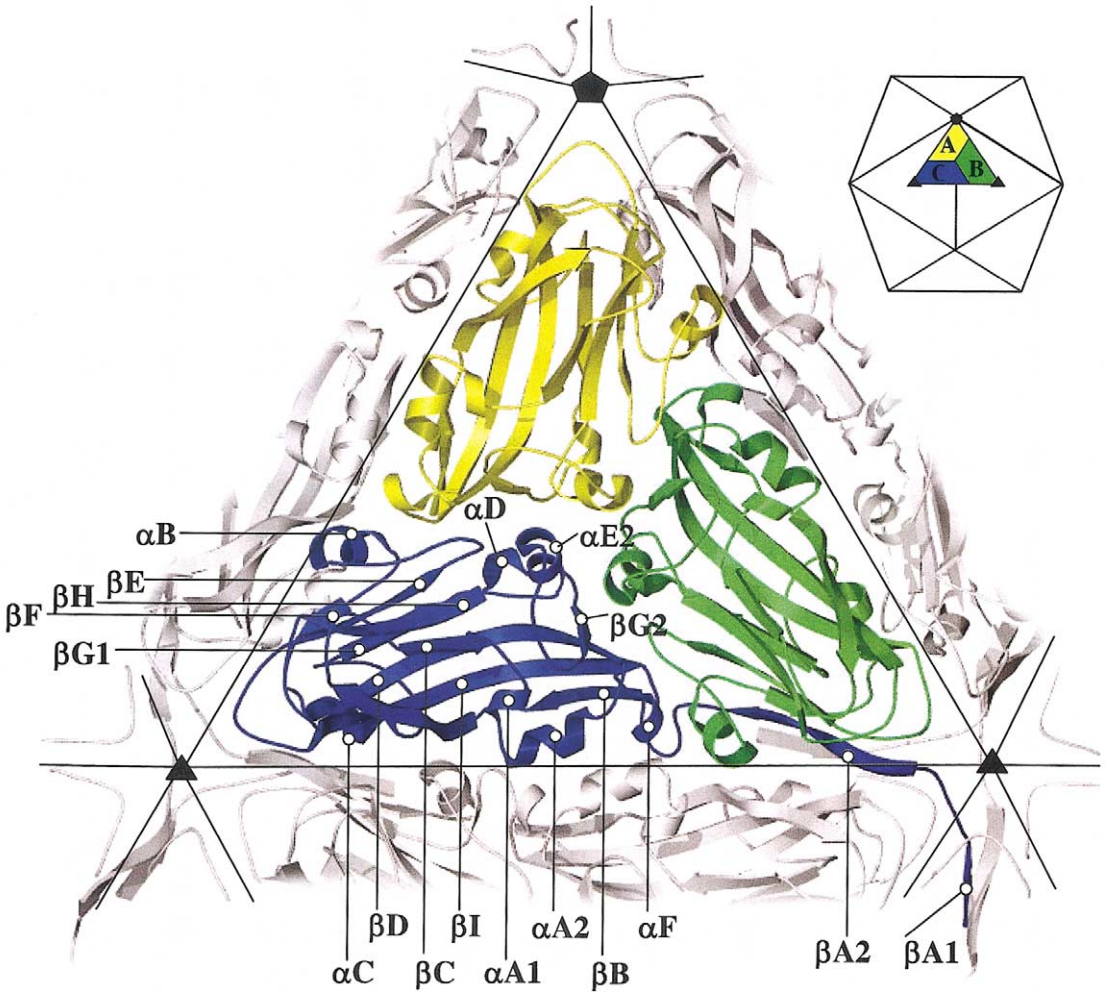
Cocksfoot mottle virus (CfMV) is a plant virus belonging to the genus *Sobemovirus*. The main host of CfMV is cocksfoot (*Dactylis glomerata*), an important herbage grass. Wheat, oat, and barley are experimental hosts. As most other sobemoviruses, CfMV is transmitted by insect (beetle) vectors. The genomes of the Russian, Norwegian, and Japanese isolates of CfMV have been sequenced and shown to be composed of a single-stranded positive-sense 4082-nucleotides-long nonpolyadenylated RNA molecule (Makinen et al., 1995; Ryabov et al., 1996). The CfMV genome contains four open reading frames (ORFs). ORF1 encodes the P1 protein, which is presumably envolved in virus movement. The coat protein (CP) is expressed from the 3′ proximal ORF. In other sobemoviruses a polyprotein is encoded by ORF2. In CfMV, however, there are two overlapping ORFs—ORF2a and ORF2b, and a polyprotein is expressed through a −1 ribosomal frameshift mechanism. CfMV also lacks the internal coding region corresponding to the ORF3 of other sobemoviruses. The polyprotein encodes three proteins, an RNA-dependent RNA polymerase, a serine protease used for the cleavage of the polyprotein, and a genome-linked VPg protein attached to the 5′ terminus of the viral RNA.

The closest known relative of CfMV is *Rice yellow mottle virus* (RYMV), which infects rice plants. The amino acid sequence of CfMV CP is about 30% identical to that of RYMV for which the crystal structure is known (Qu et al., 2000). Other members in the same genus with known 3D structures are southern cowpea mosaic virus (SCPMV, previously known as southern bean mosaic virus, cowpea strain; Abad-Zapatero et al., 1980) and Sesbania mosaic virus (SeMV; Bhuvaneshwari et al., 1995). The 3D structure is also known for tobacco necrosis virus (TNV; Oda et al., 2000), which belongs to a different genus *Necrovirus*, assigned to the *Tombusviridae* family, but has a coat protein of remarkable similarity to those of sobemoviruses.

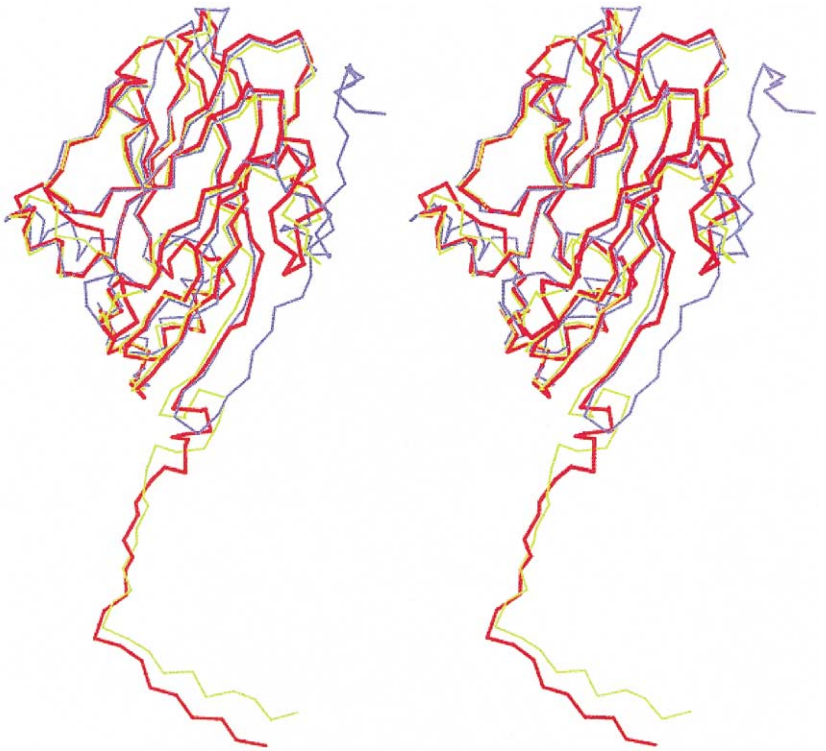
Virions of sobemoviruses have $T = 3$ quasisymmetry,

* Corresponding author.
E-mail address: kaspars@xray.bmc.uu.se (K. Tars).

0042-6822/03/$ – see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0042-6822(03)00148-X



1



2

Table 1
Primary sequence similarity and rms deviations in sobemoviruses and TBV

	CfMV	RYMV	SCPMV	SeMV	TNV	LTSV
CfMV	—	31	21	19	16	20
RYMV	1.4/97	—	23	23	19	23
SCPMV	1.4/87	1.4/88	—	60	25	27
SeMV	1.4/87	1.4/88	0.6/100	—	24	28
TNV	1.5/87	1.5/86	1.4/97	1.4/97	—	24

Note. Top right: percentage identity between the respective virus sequences. Gaps were ignored in the calculation. Bottom left: rms deviations (in Å) of superimposed Cα atoms of the respective 3D structures. The second number indicates the percentage of available amino acid residues used for calculations. The icosahedral asymmetric unit consisting of subunits A, B, and C was used as a rigid body in all cases. The default protocol in program O (Jones et al., 1990) was used for superpodition of the molecules. LTSV, lucern transient streak virus (TrEMBL Accession No. Q83095).

containing 180 coat protein monomers in three distinct environments. The monomers have a jelly-roll β-sandwich topology, common to most nonenveloped icosahedral viruses. The single-stranded RNA genome together with the genome-linked protein VPg is packaged inside the capsid.

Some details of the assembly process of sobemoviruses have been studied. It has been shown that the N-terminal part (known as the N-terminal arm) of the subunit plays a crucial role in determining the capsid size. The N-terminal arm is ordered only in one of the quasiequivalent positions (denoted C), but disordered in the other two (denoted A and B). The N-terminal arm of the C subunit is inserted between the interacting surfaces of B and C subunits at the threefold axis, making the contacts between subunits flat. The two quasiequivalent contacts are lacking the inserted arm and are therefore bent. In that manner the N-terminal arm acts as a molecular switch, controlling the subunit contacts, and the flat and bent contacts in turn regulate the curvature and thereby lead to the correct capsid side. Removal of the N-terminal arm by proteolytic cleavage (Savithri and Erickson, 1983) or by genetic engineering methods (Lokesh et al., 2002) results in the formation of $T = 1$ particles, where all contacts between subunits are bent. Compared to SCPMV, the N-terminal arm in RYMV is swapped across the twofold axis (Qu et al., 2000), but still regulates the same type of flat and bent contacts.

We have now determined the structure of CfMV at relatively high resolution to show the assembly control in this virus as well as conserved structural features among the sobemoviruses and other $T = 3$ plant viruses.

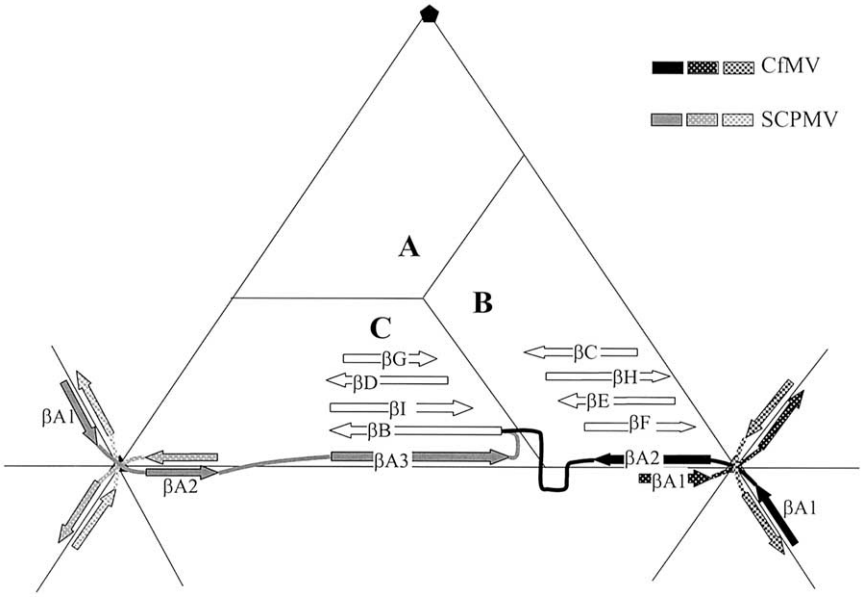


Fig. 3. A schematic representation of the position of the N-terminal arm in sobemoviruses. The position of β-strands common for all sobemoviruses are shown in white. Strands characteristic for CfMV and RYMV and SCPMV and SeMV are shown in black and gray, respectively. The threefold related N-terminal arms from neighboring C subunits are shown in slightly different patterns. The actual position and length of the β-strands is not correct in detail. In CfMV βA1 and βA2 from two different C subunits extend the CHEF sheet of a B subunit, but in SCPMV an extra strand (βA3) is added to the BIOG sheet in the C subunit.

Fig. 1. Arrangement of coat protein molecules in the CfMV capsid. Three monomers, A (yellow), B (green), and C (blue), form the icosahedral asymmetric unit. The boundaries to the subunits in other icosahedral asymmetric units (in gray) are marked by black lines. The icosahedral fivefold and threefold axes are marked by a pentagon and triangles, respectively. Inset shows the position of the three subunits A, B, and C in the icosahedron.

Fig. 2. A stereo view of superimposed Cα traces of C subunits of CfMV (red), RYMV (yellow-green), and SCPMV (blue). The view is shown from inside the particle.

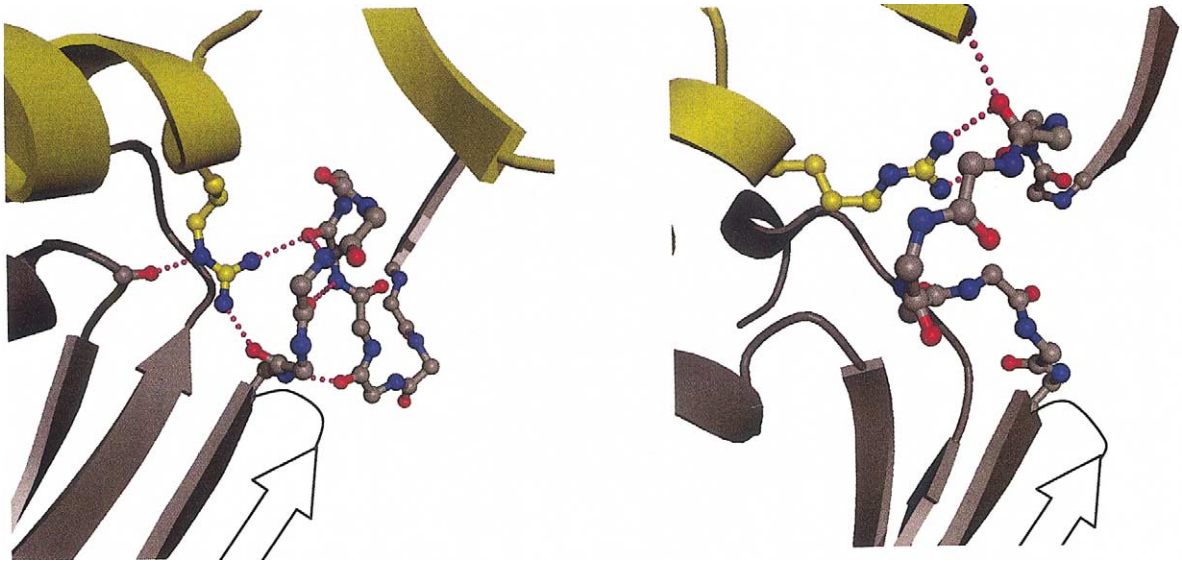


Fig. 4. The loop connecting βA and βB of the C subunit. In both CfMV (a) and RYMV (b) there is an arginine from the symmetry-related C subunit inserted in a pocket, formed by three loops. In the case of CfMV, the AB loop is composed of two turns, one of which is a β-turn stabilized by three main chain hydrogen bonds. In RYMV the conformation of the AB loop is very different and it lacks any internal stabilizing hydrogen bonds. There is, however, a hydrogen bond between carbonyl oxygen of residue C47 and nitrogen of residue 60 from the symmetry-related C subunit (the same from where Arg C87 comes). In the case of SCPMV there is a tryptophan in the position analogous to Arg 97 in CfMV. Therefore, the N-terminal arm is no longer attracted by stabilizing hydrogen bonds with residue 97 and can fold backward and make βA3 instead (schematically shown with a white block arrow).

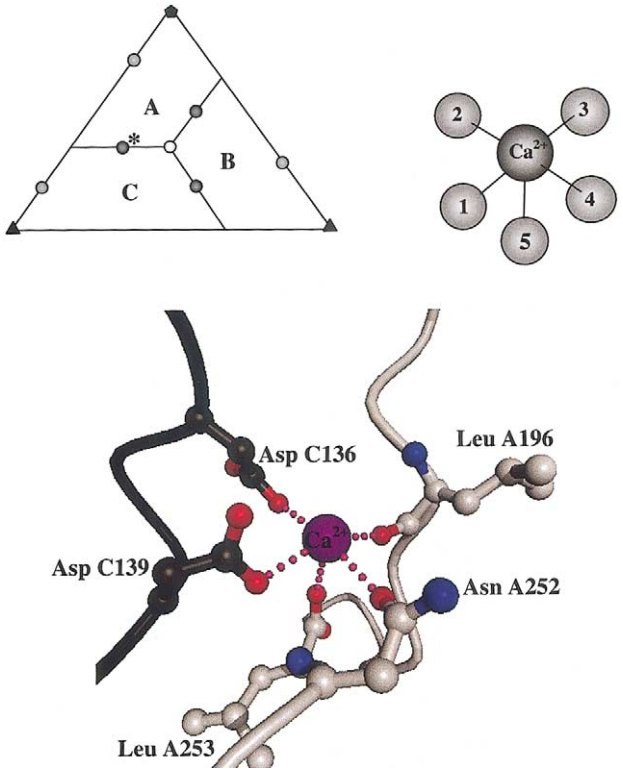


Fig. 5. (a) Location of calcium binding sites in sobemoviruses and TNV. The sites present in all sobemoviruses and TNV are shown in black. Sites found in only SeMV and TNV are shown in gray and white, respectively. (b) Coordination of the Ca^{2+} ion at the calcium binding site, marked with (*) in (a). The actual ligands 1–5 are shown in Table 2 for each virus. (c) The ball-and-stick model of the actual calcium binding site of CfMV, marked with (*) in (a).

Results and discussion

Quality of the model

The model of the three independent polypeptide chains contains 602 amino acid residues, 3 calcium ions, and 256 water molecules. The *R*-factor is 0.281 for 1,653,761 reflections between 30.0 and 2.7 Å resolution used in the refinement. All measured reflections were used in the refinement. The high *R*-factor is partly due to weak reflections, especially at high resolution.

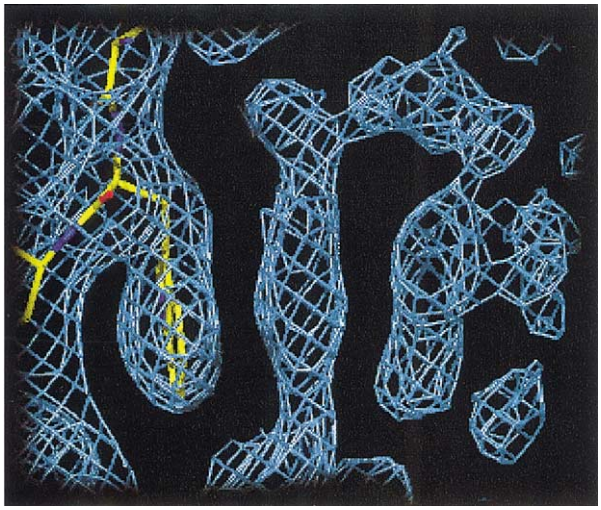


Fig. 6. The electron density of putative RNA bases stacked at the side chain of tryptophan C62.

The density was interpretable for most of the structure except the disordered N-terminal parts. Disordered regions include residues 1 to 61 for the A and B subunits and residues 1 to 35 for the C subunit. Additionally, residues 82 to 83 (in all three subunits) and C53–C55 appeared to be disordered and had little density for side chains.

The Ramachandran plot showed 98% of residues in core regions (Kleywegt and Jones, 1996). The rmsd of the bond lengths is 0.007 Å and the rmsd of the bond angles is 1.4°. The coordinates have been deposited in the Protein Data Bank (entry 1NG0).

The overall structure

The CfMV capsid consists of 180 chemically identical, 253 residues long monomers in three slightly different conformations, denoted A, B, and C (Fig. 1). The capsid has an icosahedral shape with a maximum dimension of 290 Å.

All subunits have the standard jelly-roll β-sandwich fold with strands denoted βB through βI. The C subunit has two further β strands, βA1 and βA2, which are formed by interactions with other subunits after assembly. The connecting loops contain a number of short helices. Two of them, αA2 in the CD and αB in the EF loop, have been found in most jelly-roll viral coat proteins, but the helix αC in the FG loop at the inner surface of the shell is only found in sobemoviruses and TNV.

A stereo view of superimposed Cα atoms of the C subunit from CfMV, RYMV, and SCPMV is shown in Fig. 2. Both the degree of sequence similarity and the root mean square (rms) differences between the superimposed coordinates of Cα atoms of CfMV, SCPMV, SeMV, and TNV are presented in Table 1. The structure of CfMV is very similar to that of the other sobemoviruses and almost all residues are superimposed except the N-terminal arm (discussed below). The rms difference values for individual subunits (data not shown) were at most 0.1 Å lower than those of the all three subunits, A, B, and C taken as a single rigid body, as given in the table. This suggests that no significant changes in relative orientations of individual subunits have occurred during the evolution of coat proteins of sobemoviruses and TNV. It can be noted that the TNV coat protein is slightly more similar in sequence and conformation to SCPMV than CfMV and RYMV, despite the virus belonging to a different genus.

The biggest differences between the CP structures of RYMV and CfMV are located in the loop regions, particularly in places where loops of both viruses are of different length. In CfMV, similar to in SCPMV, SeMV, and TNV, the loop between βG1 and βG2 contains a short helix αD. That helix is absent in RYMV. Other regions of significant structural differences include the surface-exposed loop between βG2 and αE and the loop between αC and βG1, which is facing the inner surface of virion.

Compared to other sobemoviruses, CfMV CP has an insertion after residue 67. This insertion is located in the middle of βB and produces a β-bulge. The inserted amino acid is glycine and has a psi/phi combination that is not favored for any nonglycine residue.

The N-terminal arm

Many icosahedral plant and insect viruses have an N-terminal part of the coat protein that contains a number of positively charged residues, probably responsible for the RNA binding (Lee and Hacker, 2001). When the structures were determined, these residues were disordered in all subunits and therefore not observed in the electron density map. In the sobemoviruses as well as in TNV, tomato bushy stunt virus (TBSV), and turnip crinkle virus (TCV), these residues are followed in the sequence by a region of about 30 residues that is ordered in the C-subunit but disordered in the A- and B-subunits. The ordered N-terminal arm is extended and forms a short β-strand, originally labeled βA. In SCPMV, SeMV, TNV, TCV (Hogle et al., 1986), TBSV (Harrison et al., 1978), βA is antiparallel to βB (Fig. 3, strand βA2 in gray). In contrast, both in CfMV and in RYMV, the N-terminal arm does not make a "U turn." Instead, it extends toward the B-subunit (Fig. 3, strand βA2 in black) and further forms a β-structure around the threefold axis (strand βA1) together with analogous β-arms from symmetry-related C subunits. This β-structure is again very similar to the corresponding structures in SCPMV, SeMV, and TNV. The rms difference between the superimposed CfMV and SCPMV coordinates of Cα atoms from residues C38 to C53 (CfMV numbering) in βA is only 1.1 Å. In CfMV and RYMV, the four-stranded β-sheet, composed of βF, βE, βH, and βC, is extended to a six-stranded sheet by addition of the βA1 and βA2 strands from neighboring subunits.

At the position in SCPMV where the N-terminal arm of the C subunit makes a U turn, there is a loop in CfMV and RYMV connecting β-strands A2 and B. However, the conformation of the loop is very different. In RYMV the loop appears to be of rather irregular conformation and without any stabilizing hydrogen bonds. In CfMV the same loop is composed of two turns, one of which appears to be an ordinary β-turn with several stabilizing hydrogen bonds. Qu et al. (2000) suggested that in RYMV the Pro-Gly-Pro sequence at the turn was important for the direction of the chain, but these residues are not conserved in CfMV. The first proline is conserved in other sobemoviruses, but not in CfMV. Important for the conformation is clearly Arg97 in αA, which is conserved in both RYMV and CfMV. The side chain of this residue from a twofold related subunit is buried between subunits and interacts with several oxygen atoms in the turn (Fig. 4). These interactions appear to stabilize the observed conformation. In SCPMV, SeMV, and TNV there is a tryptophan side chain occupying a hydrophobic cavity in the analogous position of the Arg97 in CfMV. Since a tryptophan side chain would be unable to make favorable interactions with main chain oxygen atoms, the N-terminal arm is free to fold backward and form βA3 instead. The

Table 2
The ligands for Ca^{2+} ion in sobemoviruses and TNV

Virus	1	2	3	4	5
CfMV	Oδ1	Oδ1	O	Oδ1	OT
	Asp 139	Asp 136	Leu 196	Asn 252	Leu 253
	2.3–2.4 Å	2.4 Å	2.4–2.5 Å	2.4–2.5 Å	2.4–2.6 Å
SCPMV	Oδ1	Oδ1	O	Oδ1	OT
	Asp 141	Asp 138	Phe 199	Asn 259	Leu 260
	2.4–2.8 Å	1.8–2.4 Å	2.4–2.5 Å	2.0–2.1 Å	2.0–2.4 Å
SeMV	Oδ1	Oδ1	O	Oδ1	OT
	Asp 141	Asp 138	Tyr 199	Asn 259	Leu 260
	2.3 Å	2.3–2.4 Å	2.2–2.4 Å	2.0–2.2 Å	2.0–2.2 Å
RYMV	Oδ1	Oδ1	O	Oδ1	OT
	Asp 129	Asp 126	Val 182	Asn 237	Thr 238
	2.5 Å	2.4–2.5 Å	2.4 Å	2.5–2.6 Å	2.4–2.5 Å
TNV	Oδ1	Oδ1	O	Oδ1	—
	Asp 163	Asp 160	Thr 219	Asn 275	
	2.8–2.9 Å	2.8–3.1 Å	2.5–2.6 Å	2.6–2.7 Å	

Note. Ligands 1, 2 and 3, 4, 5 are located in different subunits (for example, if 1 and 2 are in subunit A, then 3, 4, and 5 are in subunit B). The average distance between the respective ligand and Ca^{2+} ion is shown in Å. O, carbonyl oxygen; OT, C-terminal carboxyl oxygen.

different arrangement of the arms in RYMV and CfMV might lead to increased stability of the particles, as experimentally demonstrated for RYMV (Qu et al., 2000).

The calcium binding site

In a number of plant and insect viruses, bivalent metal ions—Ca^{2+} or Mg^{2+}—are involved in the particle assembly/disassembly process. The metal ions are located between the interacting surfaces of monomers and coordinated by several negatively charged residues. Upon treatment of particles by high pH or chelating agents, the metal ions can be removed. This leads to the swelling of particles (Hull, 1977) due to the repelling forces between the negatively charged residues that are otherwise involved in the metal binding. The swollen particles are very unstable and can easily be dissociated by increasing the ionic strength of the solvent. It is thought that release of metal ions is an initial step for virus uncoating upon infection, utilizing the low calcium concentration within a cell.

Although the incorporation of metal ions in viral capsids seems to serve the same purpose, the sites of metal binding and types of coordination are very different for different viruses. Therefore, it appears that the ability to use metal ions for efficient assembly/disassembly is an example of convergent evolution.

In sobemoviruses and the similar TNV three Ca^{2+} ions are inserted between the interacting interfaces of A, B, and C subunits related by the quasi-threefold axis. The Ca^{2+} ion appears to have octahedral coordination, but with only five ligands. The interactions between the Ca^{2+} ion and its ligands in sobemoviruses and TNV are summarized in Fig. 5 and Table 2. It can be seen that the interactions between ligands and the Ca^{2+} ion are identical in the case of sobemoviruses and very similar for TNV. In addition to these calcium ions, there is another Ca^{2+} ion at the quasi-threefold axis in SeMV. This ion, however, lacks any apparent ligands. In TNV there are additional Ca^{2+} ions between pairs of fivefold related A subunits and between neighboring B and C subunits close to the threefold axis.

In TBSV three calcium binding sites are located approximately at the same positions as in sobemoviruses, but two close-by calcium ions occupy each site. Some of the ligands to one of the ions appear to be the same as those found in sobemoviruses. In the nodavirus black beetle virus (BBV; Hosur et al., 1987), there are two close-by Ca^{2+} ions sitting on the quasi-three fold axis and coordinated by six aspartate residues (three for each ion). Three further calcium ions are located between pairs of fivefold related A subunits and between neighboring B and C subunits close to the threefold axis in the same way as in TNV. In another nodavirus, pariacoto virus (Tang et al., 2001), there appears to be only one calcium ion, located on the quasi-threefold axis. In the $T = 1$ satellite tobacco necrosis virus (STNV) calcium ions are located at and near the threefold axis and on the fivefold axis (Jones and Liljas, 1984).

The RNA binding

To pack the viral nucleic acid, the coat proteins of many viruses recognize a specific region within the genome. In the case of single-stranded RNA viruses, it is frequently a 20- to 30-nucleotides-long stem-loop sequence that is used as the packing signal (Hacker, 1995; Beckett et al., 1988). The same interaction may have additional functions, such as initiation of assembly (Hohn, 1969) and translational repression (Beckett et al., 1988). The coat protein binding region in the SCPMV genome has been localized (Hacker, 1995). It is a 30-nucleotide-long stem-loop sequence within a conserved region in ORF2, coding for the serine protease. Somewhat similar stem-loop sequences can be found within the same coding region of other sobemoviruses, including CfMV. It has, however, not been proven that the same kind of interaction is present in the other viruses.

In addition to specific interactions with RNA, coat proteins of a number of single-stranded RNA viruses make nonspecific interactions to the nucleic acid. These are mostly interactions between positively charged residues in the protein and phosphate groups or stacking interactions between aromatic residues and nucleotide bases. When expressed in *Escherichia coli*, capsids of several single-stranded RNA viruses, including SeMV (Lokesh et al., 2002), pack RNA of bacterial origin.

Earlier studies have shown that the CfMV coat protein does not appear to bind in vitro any specific sequence within the CfMV genome (Tamm and Truve, 2000). Instead, it seems to bind any kind of RNA with a rather strong affinity. To solve the problem of how the virus recognizes and packs its own genomic RNA instead of cellular RNA, the mechanism for RNA binding might be different in vivo. Possibly,

some kind of premade multimeric form of coat protein (e.g., dimer, trimer, or pentamer) is required for specific binding. It is also possible that there are several RNA binding sites within the coat protein. Only one of those sites might be specific, while others (for example, the arginine-rich N-terminal arm) might bind to any available RNA. Since sobemoviruses have the genome-linked protein VPg, it is also possible that specific interactions occur between VPg and the coat protein. In that case there would be no need for specific interactions between coat protein and RNA.

In several viruses, a part of RNA genome follows the icosahedral symmetry of the capsid leading to observable electron density for a number of consecutive nucleotides. This is the case in nodaviruses (Fisher and Johnson, 1993; Hosur et al., 1987; Tang et al., 2001) and two plant viruses (Chen et al., 1989; Larson et al., 1993). In most cases, however, the RNA chain does not follow the icosahedral symmetry. The RNA will be randomly oriented in the crystal and no significant electron density will be observed, with the possible exception of isolated stacked bases.

At least two nucleotide bases could be clearly seen in the electron density map of CfMV stacking at the side chain of residue Trp C62 (Fig. 6). However, none of the particular bases G, A, U, or C could be fitted into that density and no density for the RNA backbone was observed. Presumably different bases with different backbone positions occupy analagous sites in the particle and become averaged by the random orientation of particles in the crystal. There is no clear evidence of RNA bases stacked to residues A62 or B62. It should be noted that Trp 62 is not conserved among the sobemoviruses. Moreover, for SCPMV, SeMV, and RYMV nonaromatic residues are found in the same position, making similar stacking interactions unlikely in those viruses.

Another density that could be attributed to RNA was found close to Trp A160 near the fivefold axis (see below in the discussion about conserved residues), but in that case the density was not very clear.

Conserved residues and interactions in sobemoviruses

The aligned amino acid sequences of coat proteins of sobemoviruses and TNV are shown in Fig. 7. It can be seen that 22 residues are identical in all sobemoviruses. Fifteen of those residues are conserved also in TNV. Aspartates 137 and 140 and Asn 252 (CfMV numbering) are involved in the conserved calcium binding and are therefore invariant. The C-terminal residue also participates in calcium binding (except for TNV), but since it is the terminal carboxyl group that makes the interaction, conservation of the side chain is not necessary. Similarly, the carbonyl oxygen of residue 196 participates in the calcium binding, but the side chain is not conserved.

Four of the conserved residues are glycines, occupying positions 128, 162, 165, and 233. The phi/psi torsion angle values for glycines 128 and 162 are in regions unfavorable for nonglycine residues. Consequently, these residues might be conserved due to geometric constraints. In contrast, glycines 165 and 233 have phi/psi values in favored regions. The reasons for conservation of these residues are not evident from the structure, since there appears to be enough space for a small side chain.

All the five conserved prolines in the structure appear to produce a sharp bend in the protein chain. However, the requirement for proline is not absolute. For example, Pro 92 is conserved in all sobemoviruses. In TNV an alanine occupies the same position, but the conformation of that and the following residues is still similar to sobemoviruses.

The conserved residues Tyr 117, Met 131, Phe 133, and Val 159 fill a big hydrophobic volume in the middle of the β-sandwich of each monomer. In addition, the hydroxyl group of Tyr 117 makes a hydrogen bond to the carbonyl group of Val 159 in all four viruses.

As many as four of the conserved residues are alanines. Conservation of alanine residues normally would be expected in places where a hydrophobic residue is needed, but there is no room for a bigger side chain. This is exactly the case for Ala 102, which is in direct contact with the conserved Pro 92. Similarly, Ala 45 makes a close contact to another conserved residue, Trp 160. In TNV there is a tyrosine at position 160, leaving a bigger space for the side chain of residue 45, which is isoleucine in this case. For Ala 179 and Ala 216, however, there appears to be enough room for a larger side chain.

The side chain of the conserved Trp 160 lies near the fivefold axis in the A subunits. Together with the four symmetry-related tryptophans, it makes a ring around the fivefold axis. The tryptophan side chains are, however, not sufficiently close (4.8 Å) to make direct contacts. In the middle of the ring on the fivefold axis there is a relatively strong spherical electron density peak. This peak does not appear to be due to an ion, since there are no suitable ligands around. The density could be contributed by statically disordered nucleotide bases, making stacking interactions with the close-by tryptophans. At the same position, electron density is reported to be present in SeMV structure (Bhuvaneshwari et al., 1995), described as “probably representing a bound water molecule.” In the case of CfMV, however, the density blob is too big and strong and also too far from any hydrogen bonding partners to be a water molecule. A similar density on and around the fivefold axes has also been observed in the case of satellite panicum mosaic virus (Ban and McPherson, 1995).

Implications for assembly

It is generally accepted that the N-terminal arm present in the C-subunit of most $T = 3$ plant viruses acts as a molecular switch, determining whether interacting subunits are making flat or bent contacts and thereby regulating the assembly process. The N-terminal arm can be removed from SCPMV monomers by mild treatment with trypsin. Upon reassembly in vitro, the truncated monomers form $T = 1$

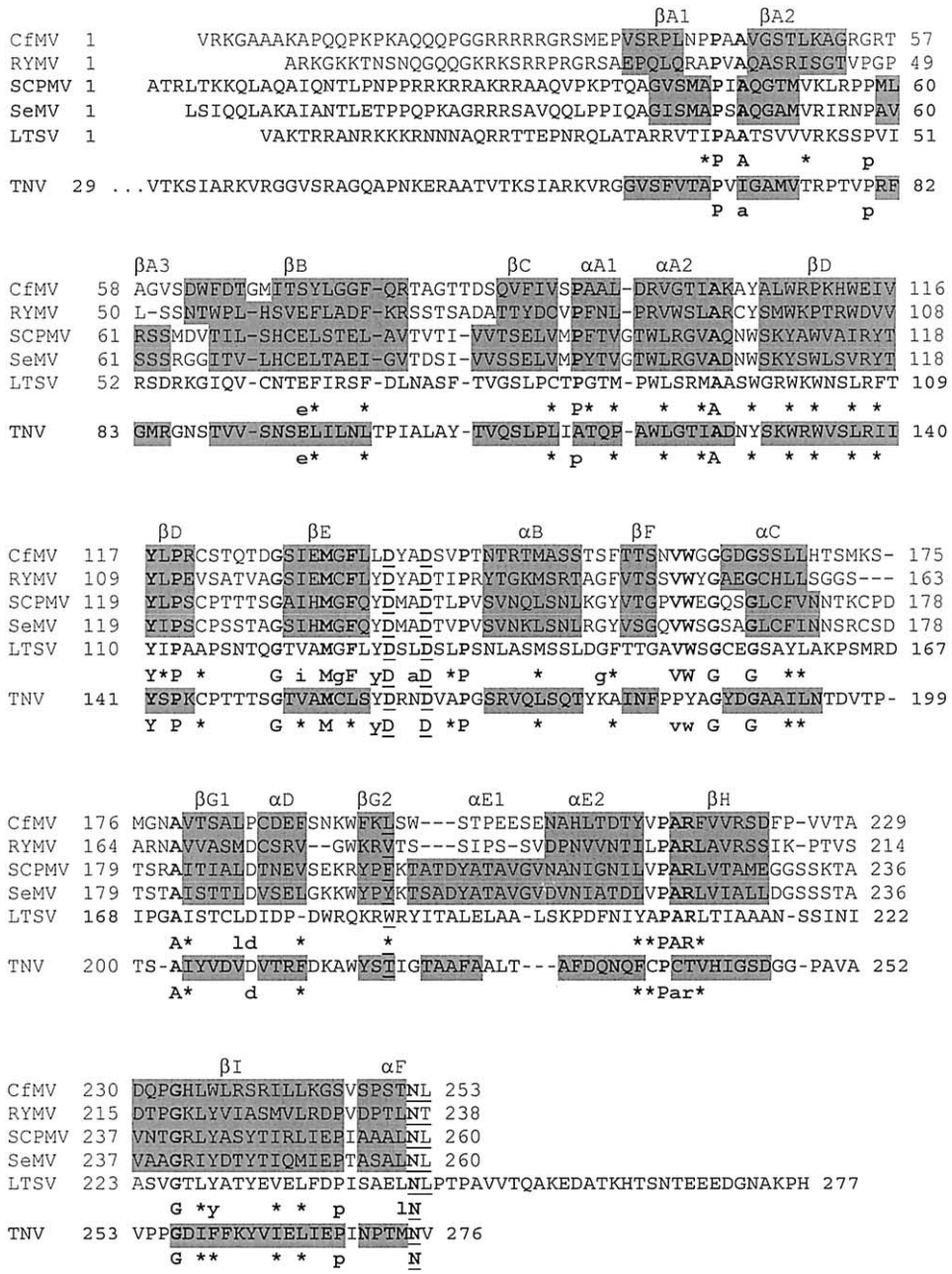


Fig. 7. The primary structure alignment of coat proteins from sobemoviruses and TNV. The secondary structure elements (assigned from the hydrogen bonding pattern of the known 3D structures) are highlighted and denoted in the first line. The conserved residues are shown in bold capital letters and separately for sobemoviruses only and sobemoviruses plus TNV. Residues conserved in all but one virus are shown in small bold letters. Residues involved in calcium binding are underlined. Stars stand for a conserved hydrophobic feature of the particular residue.

particles (Savithri and Erickson, 1983). Also, removal of N-terminal arm in recombinant SeMV particles leads to formation of $T = 1$ particles only (Lokesh et al., 2002).

While the N-terminal arm appears to determine the particle size, there are many other contacts between the monomers. Apart from the interactions mediated by calcium ion, there are no inter-subunit contacts that are conserved among sobemoviruses. A similar observation has been made earlier from comparisons of several related small RNA phages (Tars et al., 1997). It appears that subunit–subunit contacts with similar properties can result from many different interactions and residues forming the subunit–subunit interfaces are therefore not conserved. In contrast, folding seems to be far more demanding—of 22 conserved amino acids among the sobemoviruses, at least 11 are either prolines or glycines that might be important for the local main chain conformation, or big hydrophobic residues, forming the protein core. A similar pattern of sequence conservation has also been found from a comparison of picorna-like viruses (Liljas et al., 2002).

Only three of the conserved residues can be regarded as being important for assembly—the two aspartates and the asparagine forming the calcium binding site. It is also not very likely that these residues would be extremely important for assembly itself. Instead, it appears that carboxyl group–calcium interactions are just a convenient tool to efficiently regulate the disassembly mechanism upon virus uncoating and RNA release.

Materials and methods

Virus production, purification, and sequence verification

CfMV Russian isolate (CfMV-RU) infected oat leaves (kindly provided by Dr. K. Makinen, Helsinki, Finland) were used as a virus source. Oat plants cv. "Arta" were mechanically inoculated at the two to three leaves growth stage and cultivated for 4 weeks at temperatures fluctuating between 21 and 24°C under artificial light for 16 h/day.

For the virus purification (Makinen et al., 1995) harvested plants were homogenized in 0.075 M potassium phosphate buffer (pH 5.5) and filtered to remove the plant debris. The filtrate was extracted with 0.4 volumes of chloroform and centrifuged at 3000 *g* for 20 min. To the resulting virus containing solution ammonium sulfate was added to a final concentration of 30%. The precipitate was collected by centrifugation, solubilized in destilled water to 1/10 of original volume, and dialyzed against 200 volumes of distilled water.

Virus particles were collected by 1 h centrifugation at $RCF_{max} = 280{,}000\ g$. The virus pellet was resuspended in 0.075 M potassium phosphate buffer (pH 5.5) and centrifuged for 1 h at 280,000 *g* through a 20% sucrose cushion. The sucrose cushion steps were repeated three times and the virus particles were resuspended in potassium phosphate buffer (pH 5.5) to a final concentration of approximately 15 mg/ml.

To verify the sequence, virus RNA was isolated using TRI-reagent (Sigma, St. Louis, MO, USA). The RNA was reverse transcribed and a PCR fragment, corresponding to coat protein gene, was produced and sequenced. Several point mutations were found; however, none of them resulted in the amino acid change.

Crystallization and data collection

Crystals were obtained under two different conditions using the hanging drop technique. In both cases, 5 μl of virus solution (10 mg/ml) were mixed with an equal volume of bottom solution. Crystal type A (irregular, cucumber-like shape) was obtained using 0.25 M sodium succinate buffer at pH 3.3 and 0.5% PEG 8000 as an additive. Crystal type B (prisms) was obtained using 0.1 M sodium phosphate buffer, pH 5.5, and 2 M NaCl as precipitant. For data collection, crystals were grown to a size of about 0.2 mm, soaked for 1 min in mother liquor containing 30% PEG 400, and frozen in liquid nitrogen. Data were collected at −170°C on the ADSC Quantum4 CCD detector at beamline ID14-2 at ESRF synchrotron radiation source in Grenoble, France. The oscillation angle was 0.25°. Crystal form A (orthorhombic spacegroup $P2_12_12_1$) diffracted to about 2.6 Å resolution and crystal form B ($P4_x2_x2$) diffracted to about 3.2 Å resolution. Data for both crystal forms were processed and scaled using the HKL package (Otwinowski and Minor, 1996). The statistics from data collection and scaling is shown in Table 3.

Table 3
Scaling statistics for orthorhombic and tetragonal crystal forms

	Orthorhombic crystals	Tetragonal crystals
Unit cell dimensions		
a	292.1	537.1
b	537.2	537.1
c	555.6	517.1
Resolution limits	40–2.7 Å	40–3.3 Å
High-resolution bin	2.75–2.70 Å	3.36–3.30 Å
Completeness	79%	52%
Completeness in high-resolution bin	70%	49%
R_{merge}*	0.19	0.17
R_{merge} in high-resolution bin	0.99	0.56
I/σ	7.3	8.4
I/σ in high-resolution bin	1.3	2.0

* $R_{merge} = \Sigma_h\Sigma_j|I_{hj} - \langle I_h\rangle|/\Sigma\Sigma I_{hj}$.

Structure determination

The structure was solved by the molecular replacement method, using rice yellow mottle virus coordinates (PDB entry 1F2N) as the search model. In both crystal forms, a complete virus particle occupied a crystallographic asymmetric unit. This means that six parameters had to be determined. First, rotational parameters of particles in the unit cell were determined using the locked self-rotation function in program GLRF (Tong and Rossmann, 1990). Reflections between 6 and 2.7 Å (orthorhombic crystal form) or 6 and 3.3 Å (tetragonal crystal form) were used for calculations. To test the usefulness of the high-resolution data for the orthorhombic crystal form, only the data between 3.0 and 2.7 Å were used in a separate calculation. The results were the same as for 6–2.7 Å dataset, suggesting that there is useful information at a resolution higher than 3 Å, despite the bad data quality as judged from scaling statistics. The radius of integration was chosen to be 300 Å. Use of a smaller radius did not produce any clear peaks. The results showed that particle orientations relative to a standard icosahedral orientation were $\varphi = 89.6°$, $\psi = 80.5°$, and $\kappa = 86.2°$ for the orthorhombic crystal form A and $\varphi = 63.1°$, $\psi = 15.4°$, and $\kappa = 35.0°$ for the tetragonal crystal form B. Translational parameters were calculated using Patterson correlation search in the program TF (Tong, 1993) with

Table 4
Refinement statistics

High-resolution limit	Low-resolution limit	*N* reflections	*R*-factor
5.4	40.0	257,623	0.21
4.28	5.40	258,510	0.22
3.74	4.28	242,118	0.26
3.40	3.74	215,109	0.30
3.16	3.40	192,521	0.34
2.97	3.16	173,281	0.37
2.82	2.97	161,570	0.40
2.70	2.82	153,029	0.43
2.70	40.0	1,653,761	0.28

precalculated structure factors from CNS (Brünger et al., 1998). For the orthorhombic crystal form, a clear peak appeared at $a = 0.2336$, $b = 0.2690$, and $c = 0.2748$ (fractional coordinates). For the tetragonal crystal form, no consistent solution of the translation function was found, and only the orthorhombic crystal form was used for further calculations. A polyalanine model of RYMV was placed in the correct orientation and position in the unit cell and the initial phase was then calculated in CNS. The phases were refined by 30 cycles of real-space averaging by the program RAVE (Kleywegt and Jones, 1993, 1994), using the 60-fold noncrystallographic symmetry. Since only the low resolution (40 to 6 Å) data were used in translation function (otherwise the calculated structure factor files could not be handled by the available computers), the translational parameters were refined by shifting the particle center and calculating statistics after averaging using RAVE. This procedure is described earlier for the similar case of RNA bacteriophage PP7 (Tars et al., 2000). The new particle center was found to be moved by $\Delta y = 0.25$ Å and $\Delta z = 0.375$ Å. This minor shift improved dramatically the overall correlation coefficient in RAVE from 0.68 to 0.84.

The model was built in program O (Jones et al., 1990) starting from the RYMV coordinates where the respective residues were mutated to those of CfMV. In most places the map was very clear and the chain could be traced unambiguously. The model was refined by the simulated annealing procedure in the program CNS. Four hundred water molecules were picked from the ordinary 2Fo-Fc map, using the CCP4 programs Peakmax and Watpeak (Collaborative Computational Project, 1994). Only water molecules compatible with the noncrystallographic symmetry and making hydrogen bonds to protein atoms or other water molecules were kept. Forty water molecules were added manually, resulting in a total of 256 assigned waters. The final *R*-value after temperature factor and positional refinements was 0.281 for all reflections between 40 and 2.7 Å resolution. The R_{free} is very similar to the *R*-value due to the noncrystallographic symmetry (Kleywegt and Brünger, 1996). The refinement statistics are given in Table 4.

Acknowledgments

We thank Dr. Gunnar Berglund for collecting the orthorhombic data set, Dr. K. Makinen for providing the initial CfMV material, Prof. E. Truve for advice in virus propagation, and Prof. P. Pumpens for the critical reading of the manuscript. This work was financially supported by the Swedish Research Council, by O.E. and Edla Johanssons Scientific Foundation, by Grant 01.022 of Latvian Science Council, and by the EU INCO Copernicus ERBIC Grant 15-CT96-0907.

References

Abad-Zapatero, C., Abdel-Meguid, S.S., Johnson, J.E., Leslie, A.G.W., Rayment, L., Rossmann, M.G., Suck, D., Tsukihara, T., 1980. Structure of southern bean mosaic virus at 2.8 Å resolution. Nature 286, 33–39.

Ban, N., McPherson, A., 1995. The structure of satellite panicum mosaic virus at 1.9 Å resolution. Na. Struct. Biol. 2, 882–890.

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.

Bhuvaneshwari, M., Subramanya, H.S., Gopinath, K., Savithri, H.S., Nayudu, M.V., Murthy, M.R.N., 1995. Structure of *sesbania* mosaic virus at 3 Å resolution. Structure 3, 1021–1030.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., Warren, G.L., 1998. Crystallography and NMR System: a new software suite for macromolecular structure determination. Acta Crystallogr. D 54, 905–921.

Chen, Z., Stauffacher, C., Li, Y., Schmidt, T., Bomu, W., Kamer, G., Shanks, M., Lomonossoff, G., Johnson, J.E., 1989. Protein–RNA interactions in an icosahedral virus at 3.0 Å resolution. Science 245, 154–159.

Collaborative Computational Project, N., 1994. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D 50, 760–763.

Fisher, A.J., Johnson, J.E., 1993. Ordered duplex RNA controls capsid architecture in an icosahedral animal virus. Nature 361, 176–179.

Hacker, D.L., 1995. Identification of a coat protein binding site on the southern bean mosaic RNA. Virology 207, 562–565.

Harrison, S.C., Olson, A.J., Schutt, C.E., Winkler, F.K., Bricogne, G., 1978. Tomato bushy stunt virus at 2.9 Å resolution. Nature 276, 368–373.

Hogle, J.M., Maeda, A., Harrison, S.C., 1986. Structure and assembly of turnip crinkle virus. I. X-ray crystallographic structure analysis at 3.2 Å resolution. J. Mol. Biol. 191, 625–638.

Hohn, T., 1969. Role of RNA in the assembly process of bacteriophage fr. J. Mol. Biol. 43, 191–200.

Hosur, M.V., Schmidt, T., Tucker, R.C., Johnson, J.E., Gallagher, T.M., Selling, B.H., Rueckert, R.R., 1987. Structure of an insect virus at 3.0 Å resolution. Proteins Struct. Funct. Genet. 2, 167–176.

Hull, R., 1977. The stabilization of turnip rosette virus and of other members of the southern bean mosaic virus group. Virology 79, 58–66.

Jones, T.A., Bergdoll, M., Kjeldgaard, M., 1990. O: a macromolecule modeling environment, in: Bugg, C., Ealick, S. (Eds.), Crystallographic and Modeling Methods in Molecular Design, Springer-Verlag, New York, pp. 189–199.

Jones, T.A., Liljas, L., 1984. Structure of satellite tobacco necrosis virus after crystallographic refinement at 2.5 Å resolution. J. Mol. Biol. 177, 735–767.

Kleywegt, G.J., Brünger, A.T., 1996. Checking your imagination: applications of the free R value. Structure 4, 897–904.

Kleywegt, G.J., Jones, T.A., 1993. Masks made easy. CCP4/ESF-EACBM Newsl. Protein Crystallog. 29, 56–59.

Kleywegt, G.J., Jones, T.A., 1994. Halloween . . . masks and bones, in: Bailey, S., Hubbard, R., Waller, D. (Eds.), From First Map to Final Model. Proceedings of the CCP4 Study Weekend, SERC Daresbury Laboratory, Daresbury, pp. 59–66.

Kleywegt, G.J., Jones, T.A., 1996. Phi/psi-chology: Ramachandran revisited. Structure 4, 1395–400.

Larson, S.B., Koszelak, S., Day, J., Greenwood, A., Dodds, J.A., McPherson, A., 1993. Double-helical RNA in satellite tobacco mosaic virus. Nature 361, 179–182.

Lee, S.K., Hacker, D.L., 2001. In vitro analysis of an RNA binding site within the N-terminal 30 amino acids of the southern cowpea mosaic virus coat protein. Virology 286, 317–327.

Liljas, L., Lin, T., Tate, J., Christian, P., Johnson, J.E., 2002. Evolution of the picornavirus superfamily: implications of conserved structural motifs between picornaviruses and insect picorna-like viruses. Arch. Virol. 147, 59–84.

Lokesh, G.L., Gowri, T.D.S., Satheshkumar, P.S., Murthy, M.R.N., Savithri, H.S., 2002. A molecular switch in the capsid protein controls the particle polymorphism in an icosahedral virus. Virology 292, 211–223.

Makinen, K., Tamm, T., Naess, V., Truve, E., Puurand, U., Munthe, T., Saarma, M., 1995. Characterization of cocksfoot mottle sobemovirus genomic RNA and sequence comparison with related viruses. J. Gen. Virol. 76 (Pt. 11), 2817–25.

Oda, Y., Saeki, K., Takahashi, Y., Maeda, T., Naitow, H., Tsukihara, T., Fukuyama, K., 2000. Crystal structure of tobacco necrosis virus at 2.25 A resolution. J. Mol. Biol. 300 (1), 153–69.

Otwinowski, Z., Minor, W., 1996. Processing of X-ray diffraction data collected in oscillation mode, in: Carter Jr., C.W., Sweet, R.M. (Eds.), Methods in Enzymology, Vol. 276, Academic Press, New York, pp. 307–326.

Qu, C., Liljas, L., Opalka, N., Brugidou, C., Yeager, M., Beachy, R.N., Fauquet, C.M., Johnson, J.E., Lin, T., 2000. 3D domain swapping of a molecular switch for quasi-equivalent symmetry modulates the stability of an icosahedral virus. Structure 8, 1095–1103.

Ryabov, E.V., Krutov, A.A., Novikov, V.K., Zheleznikova, O.V., Morozov, S.Y., Zavriev, S.K., 1996. Nucleotide sequence of RNA from the sobemovirus found in infected cocksfoot shows a luteovirus-like arrangement of the putative replicase and protease genes. Phytopathology 86, 391–397.

Savithri, H.S., Erickson, J.W., 1983. The Self-assembly of the Cowpea strain of southern bean mosaic virus: formation of T = 1 and T = 3 nucleoprotein particles. Virology 126, 328–335.

Tamm, T., Truve, E., 2000. RNA-binding activities of cocksfoot mottle sobemovirus proteins. Virus Res. 66 (2), 197–207.

Tang, L., Johnson, K.N., Ball, L.A., Lin, T., Yeager, M., Johnson, J.E., 2001. The structure of pariacoto virus reveals a dodecahedral cage of duplex RNA. Nat. Struct. Biol. 8, 77–83.

Tars, K., Bundule, M., Fridborg, K., Liljas, L., 1997. The crystal structure of bacteriophage GA and a comparison of phages belonging to the major groups of *Escherichia coli* leviviruses. J. Mol. Biol. 271, 759–773.

Tars, K., Fridborg, K., Bundule, M., Liljas, L., 2000. Structure determination of phage PP7 from *Pseudomonas aeruginosa*: from poor data to a good map. Acta Crystallogr. D 56, 398–405.

Tong, L., 1993. REPLACE, a suite of computer programs for molecular-replacement calculations. J. Appl. Cryst. 26, 748–751.

Tong, L., Rossmann, M.G., 1990. The locked rotation function. Acta Crystallogr. A 46, 783–792.