JOURNAL OF VIROLOGY, Jan. 2000, p. 493–504 Vol. 74, No. 1
0022-538X/00/$04.00+0

Structural Fingerprinting: Subgrouping of Comoviruses by Structural Studies of Red Clover Mottle Virus to 2.4-Å Resolution and Comparisons with Other Comoviruses†

TIANWEI LIN,[1] ANTHONY J. CLARK,[2]‡ ZHONGGUO CHEN,[3]§ MICHAEL SHANKS,[2] JIN-BI DAI,[3]‖ YING LI,[3]§ TIM SCHMIDT,[3] PER OXELFELT,[4] GEORGE P. LOMONOSSOFF,[2] AND JOHN E. JOHNSON[1]*

Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037[1]; Department of Virus Research, John Innes Centre, Norwich NR4 7UK, United Kingdom[2]; Structural Studies, Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907[3]; and Department of Plant Pathology, Swedish University of Agricultural Sciences, S750 07 Uppsala, Sweden[4]

Received 12 April 1999/Accepted 20 September 1999

Red clover mottle virus **(RCMV) is a member of the comoviruses, a group of picornavirus-like plant viruses. The X-ray structure of RCMV strain S has been determined and refined to 2.4 Å. The overall structure of RCMV is similar to that of two other comoviruses, *Cowpea mosaic virus* (CPMV) and *Bean pod mottle virus* (BPMV). The sequence of the coat proteins of RCMV strain O were modeled into the capsid structure of strain S without causing any distortion, confirming the close resemblance between the two strains. By comparing the RCMV structure with that of other comoviruses, a structural fingerprint at the N terminus of the small subunit was identified which allowed subgrouping of comoviruses into CPMV-like and BPMV-like viruses.**

Red clover mottle virus (RCMV) is a member of the comoviruses, a group of plant viruses in the picornavirus superfamily with nonenveloped, icosahedral capsids and bipartite, single-stranded, positive-sense RNA genomes. As with other comoviruses, the two RNA molecules (termed RNA 1 and RNA 2) are encapsidated separately in isometric particles made up of 60 copies each of a large (L) and a small (S) coat protein (31, 34, 47). A number of strains of RCMV have been characterized (reference 28 and references therein), and the complete nucleotide sequence of the genome of one of them, RCMV strain S (34), has been determined (41, 43, 44). Inspection of the RNA sequence together with in vitro translation data (42, 45) indicates that RCMV has a mode of gene expression similar to that of the type member of the group, *Cowpea mosaic virus* (CPMV). Comparison of the RNA 2 sequences of RCMV strain S with those of CPMV, *Bean pod mottle virus* (BPMV) and *Cowpea severe mosaic virus* (CPSMV) appears to divide comoviruses into two subgroups, with RCMV and CPMV forming one subgroup and CPSMV and BPMV forming the other (9).

One notable feature of RCMV is the ability of different strains of the virus to form viable pseudorecombinants (35). Pseudorecombinants between strains S and O have proved useful in mapping symptom and host range determinants (11, 36). The viability of the pseudorecombinants indicates that the coat proteins encoded by the RNA 2 of one strain can be processed correctly by the RNA 1-encoded 24K proteinase of another and are able to efficiently encapsidate the RNA 1 of the heterologous strain. This implies that any differences in the amino acid sequences of the coat proteins from the different strains do not dramatically alter their structure and assembly characteristics.

Our basic knowledge of comovirus structure has been derived from crystallographic studies of CPMV and BPMV (10, 31, 48; T. Lin, Z. Chen, R. Usha, C. V. Stauffaacher, J.-B. Dai, T. Schmidt, and J. E. Johnson, submitted for publication). In both cases the structures reveal architectures typical of $P=3$ plant and animal viruses, with the asymmetric unit consisting of three eight-stranded antiparallel β-sandwich domains (38) (Fig. 1). The A domain is located around the fivefold axes and comprises the S subunit of either 24 kDa (CPMV) or 22 kDa (BPMV). The B and C domains, located alternately around threefold axes, are formed by the L subunit of 41 kDa (in both cases). Despite the apparent overall similarity between the CPMV and BPMV structures, the viruses show different physical properties. For example, by ultracentrifugation on CsCl gradients, the RNA 1-containing component (bottom component) of CPMV could be separated into subcomponents, termed bottom-upper (B_U) and bottom-lower (B_L) components (4). This behavior was associated with the Cs ion permeability of the particle (53). By contrast, the bottom component of BPMV could not be separated into subcomponents by CsCl gradient centrifugation (4). In terms of its permeability to Cs ions, RCMV appears to be more similar to CPMV than to BPMV (1). This finding is consistent with the subgrouping based on the sequence alignments. Another difference between CPMV and BPMV is that while the crystal structure of the RNA 2-containing component (middle component) of BPMV revealed ordered ribonucleotides in an RNA binding pocket (10), no such ordered RNA was seen in the corresponding CPMV structure. The diverse properties of CPMV and BPMV provided an incentive for the further investigation of comovirus structures. To this end, crystals of RCMV strain S were prepared and used to solve the structure of the virus by molecular replacement.

* Corresponding author. Mailing address: Department of Molecular Biology, The Scripps Research Institute, 10550 N. Torrey Pines Rd., La Jolla, CA 92037. Phone: (858) 784-2947. Fax: (858) 784-8660. E-mail: jackj@scripps.edu.

† Dedicated to the memory of J.-B. Dai.

‡ Present address: Department of Plant Pathology, University of Kentucky, Lexington, KY 40546-0091.

§ Present address: Merck Research Laboratory, West Point, PA 19486.

‖ Deceased.

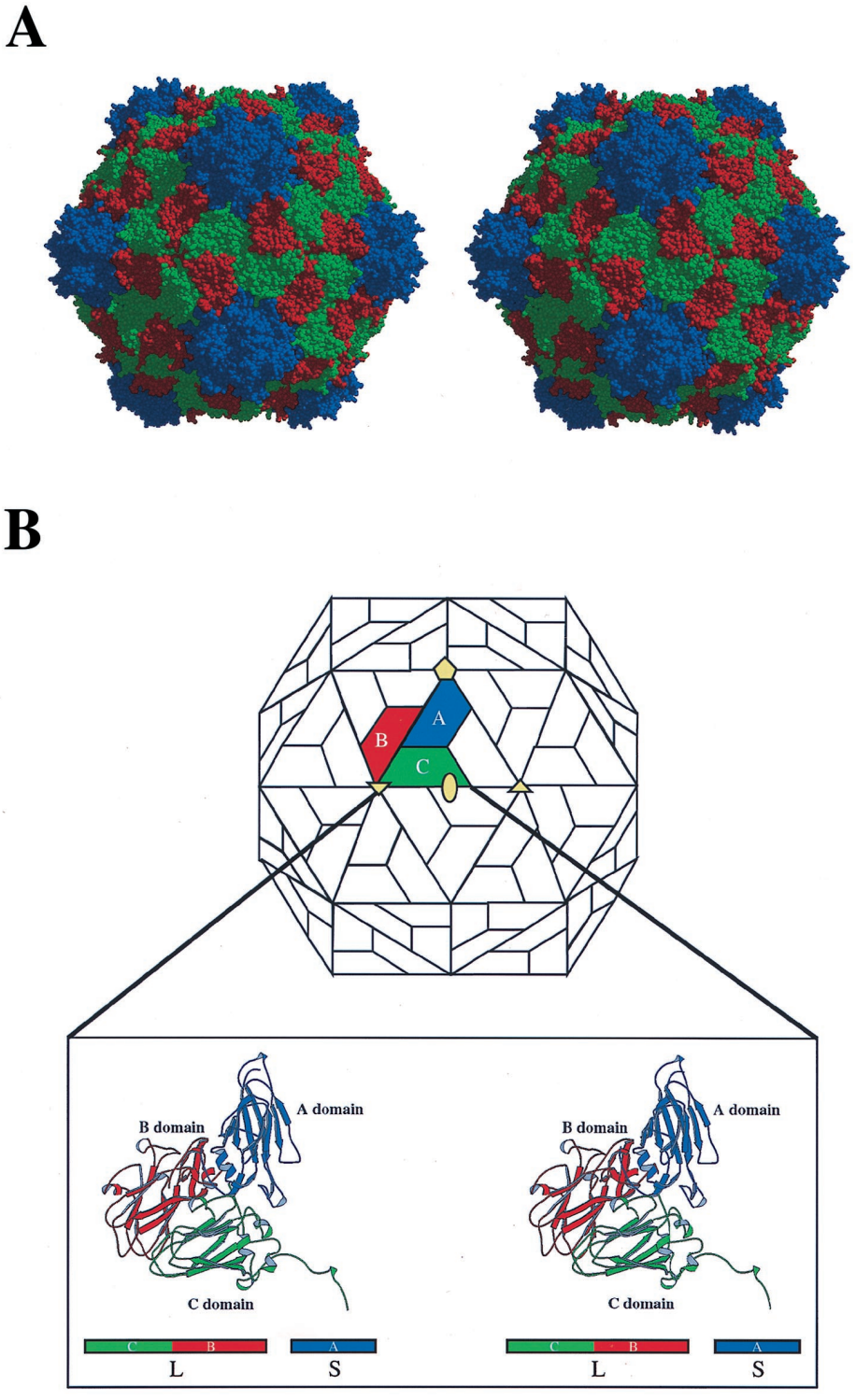


FIG. 1. The structures of viral capsid and the icosahedral asymmetric unit of RCMV. Two proteins (S and L subunits) are in the RCMV capsid. The S subunit forms the A domain (in blue), while the L subunit forms the B (red) and C (green) domains. (A) Stereoview of a space-filling drawing of the RCMV capsid. All atoms are shown as spheres corresponding to a diameter of 1.8 Å. The pentameric S subunits form the protrusion. The CPK presentation was made with MidasPlus (14, 20). (B) At the top, the icosahedral asymmetric unit of the capsid is color coded in the schematic presentation of the CPMV capsid. The S subunit occupies the A position, forming the A domain around the fivefold axis; the two domains of the L subunit occupy the B and C positions. Positions A, B, and C are quasiequivalent positions of identical gene products in a T=3 surface lattice. At the bottom, a stereoview of a ribbon diagram of the icosahedral asymmetric unit is shown. All three domains are variations of the jelly-roll β-sandwich structure. The schematic presentation of composite proteins, L and S subunits, is also shown. The ribbon diagrams were drawn with the program MOLSCRIPT (26).

MATERIALS AND METHODS

Virus propagation and purification. RCMV strains S and O (34) were propagated in *Pisum sativum* cv "Onward," and virus particles were purified as described by Shanks et al. (41). RNA was extracted from virus particles by the method of Zimmern (55).

Determination of the structure of RCMV strain S. Elongated RCMV crystals were produced by the sitting drop vapor diffusion method (33). The starting solution contained 10 mg of RCMV per ml in 10 mM sodium phosphate (pH 7). The reservoir solution contained 50 mM potassium phosphate (pH 7), 1.8% polyethylene glycol 8000, 0.3 M ammonium sulfate, 2 mM EDTA, and 1 mM sodium azide. Equal volumes of the virus and reservoir solution were mixed and equilibrated with the reservoir solution at room temperature. The crystals grew to 0.5 to 1 mm in all dimensions after 5 to 7 days.

The lifetime of RCMV crystals was about 40 h under CuKα radiation of a rotating anode operating at 35 kV and 40 mA. Screenless precession photographs with X-ray perpendicular to hk0 and h0l were taken with a μ angle of 1.5°. A complete data set, including 257 pairs of oscillation patterns, was recorded on photographic films in the A1 station of the Cornell High Energy Synchrotron Source with crystal-to-film distance of 90 or 100 mm, oscillation angle of 0.5°, and wavelength of 1.565 Å. The diffraction patterns were digitized at 50-μm intervals on a rotating-drum microdensitometer (Optronics model C-4100; Optronics International, Inc., Chelmsford, Mass.). The crystal orientations were determined by an autoindexing algorithm (25). The recorded reflection maxima were processed (37), scaled, and postrefined (39). A rotation function algorithm (50, 51) was used to resolve the ambiguity in the orientation. The initial structure factors were calculated from a poly(Ala) model of CPMV. The phase refinements were carried out as described previously (21, 40). The models were built by the programs FRODO and O (22, 23). The refinement scheme was similar to that adopted for the refinement of CPMV structures (Lin et al., submitted). Random shifts of 0.25 Å were applied to the coordinates before we calculated the structure factors from a refined model. These structure factors were used to calculate omit maps to differentiate ambiguities. Difference Fourier synthesis techniques with the resulting electron density map being averaged once were used to locate the water molecules by use of the programs PEAKMAX and WATPEAK of the CCP4 suite (12). Automatic procedures in program O were used to model the water molecules. The suite PROCHECK (29) was used to monitor the geometries of the models. The coordinates will be deposited in the Protein Data Bank.

Determination of the strain O coat protein sequences. The virus-specific inserts in plasmids pRCOM-C27 and pRCOM-D1 (36) were excised by digestion with *Eco*RI and subcloned in *Eco*RI-digested M13mp18. Two subclones of pRCOM-C27, with inserts in the opposite orientation, were sequenced by the exonuclease III deletion method of Henikoff (16) by using the −40 universal primer. Sequencing was carried out either manually with Sequenase (U.S. Biochemicals) and [α-^{35}S]dATP or automatically by using an Applied Biosystems 373 DNA sequencer. Gaps in the sequence were filled in by manual sequencing with specific primers within the virus-specific region. Additional sequence information in the region encoding the N terminus of the L protein was obtained from the subclones of pRCOM-D1 by priming with specific primers. Further RNA 2-specific cDNA clones were obtained by producing double-stranded cDNA to RNA 2 as described earlier (41), digesting the DNA with *Sau*3A, and ligating the products into *Bam*HI-digested M13mp18. The resulting recombinants were sequenced manually as described above.

N-terminal sequence analysis of the strain O coat proteins. A sample of purified strain O virions was denatured and electrophoresed on a polyacrylamide gel, and the proteins were transferred to Problott membranes as described by Achon et al. (2). The protein bands were visualized by staining with 1% (wt/vol) Amido Black, and individual bands were excised for N-terminal sequence analysis by using an Applied Biosystems model 492 gas-phase protein sequencer.

RESULTS AND DISCUSSION

Determination of the structure of RCMV strain S. The space group of RCMV strain S crystals was determined by screenless precession photography. The precession photographs of the hk0 and h0l planes of RCMV crystals showed two perpendicular mirror planes. The same symmetry elements were also found in the higher layers of the diffraction patterns, suggesting 222 symmetry. All the reflections with indices of h+k+l=odd were systematically absent, indicating a centered lattice. The cell dimensions were not the same along three axes, indicating the space group I222. Packing considerations required that the three noncrystallographic twofold axes of the particle were coincident with the crystallographic twofold axes. Consequently, there was 15-fold noncrystallographic redundancy. There was no translation problem, but the particle orientation remained to be determined from two possibilities.

TABLE 1. Data reduction and phase refinement[a]

Resolution (Å)	Data completeness (%)	CC in phase refinement
30.0		
15.0	100	0.963
10.0	100	0.991
7.50	100	0.987
5.00	98	0.986
3.50	96	0.979
3.00	90	0.925
2.75	80	0.813
2.50	52	0.624
2.40	26	0.415
Overall	72	0.960

[a] There were a total of 257 films (A-B pairs). There were 1,590,543 total observations and 451,303 unique observations with $1/\sigma = 2$. Overall R_{merg} was 8.83%. R_{merg} is defined as follows:

$$\frac{\sum_{hkl}\sum_{j=1}^{N}|I_{hkl} - I_{hkl}(j)|}{\sum_{hkl} NxI_{hkl}}$$

Correlation coefficient (CC) is defined as follows:

$$\frac{\sum_{h}\left(|F_h(obs)| - \overline{|F_h(obs)|}\right) x \left(\left|F_h(calc) - \overline{|F_h(calc)|}\right|\right)}{\left[\sum_{h}\left(\left|F_h(obs) - \overline{|F_h(calc)|}\right|\right)^2 x \sum_{h}\left(\left|F_h(calc) - \overline{|F_h(calc)|}\right|\right)^2\right]^{1/2}}$$

RCMV crystals diffracted X-rays to 2.2 Å with synchrotron radiation. A complete data set comprised of 257 A-B pairs of oscillation patterns recorded on photographic film was obtained. The lattice constants for the orthorhombic body-centered cell were a = 333.4, b = 305.0, and c = 315.0 Å. Fifty A films were initially processed, scaled, and postrefined. The cell dimensions were refined as a = 332.07, b = 303.87, and c = 314.31 Å. These parameters were used to reprocess the original 50 and the rest of the A films to 2.4 Å. Analysis of the intensity profile as a function of resolution (54) showed that reflections at the lower resolution were underestimated, probably due to the saturation of film by large numbers of high-intensity reflections. Subsequently, the B films were also processed, and the reflections were merged with the reflections of the A films. The final data set contained 451,303 reflections, with an R_{merg} of 8.8%. The completeness of the data is listed in Table 1.

A self-rotation function (50, 51) was used to resolve the ambiguity in the particle orientation. CPMV was chosen as the initial phasing model. The direct sequence homology is 53% between RCMV and CPMV coat proteins (41), and the two virus particles share similar physical properties (reference 31 and references therein). The starting phases were calculated from a poly(Ala) CPMV model placed in the RCMV cell, and the phases were refined at a 3-Å resolution with 12 cycles of averaging. The final map was reasonably free of model bias as judged by side-chain density that matched the RCMV sequence (Fig. 2A). Manual building of an RCMV model was carried out based on the amino acid sequence of the coat proteins deduced from the nucleotide sequence of RNA 2 (41), taking into account the recently verified cleavage sites used to release the L and S capsid proteins (28). With the introduction of an atomic mask (40) and calculation of the structure factors from the new model, the final round of averaging had a cor-

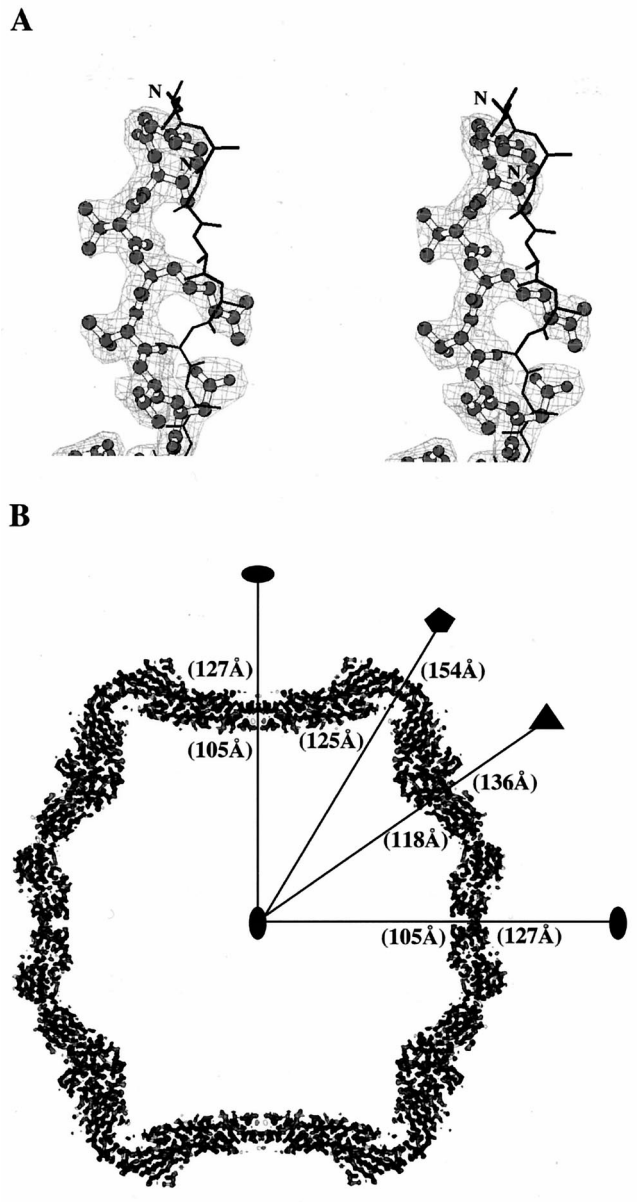


FIG. 2. Averaged electron density of RCMV. (A) Stereoview of electron density around the N terminus of the S subunit. The initial phasing model (in "black tracing") is overlapped with the averaged density (in "chickenwire") and the final RCMV model (in "ball-and-stick"). The initial phasing model is derived from a poly(Ala) structure of CPMV. There was no indication of bias toward the initial phasing model in the averaged electron density. The RCMV sequence was modeled into the averaged electron density without ambiguity. The electron density shown was contoured at 1.5σ. This figure was drawn by using the program BOBSCRIPT (13). (B) An equatorial section of the averaged electron density of the virus particle. The outer and inner dimensions of the particle along each symmetry element are indicated. The density was contoured at 2σ. This figure was generated with the program O (22, 23).

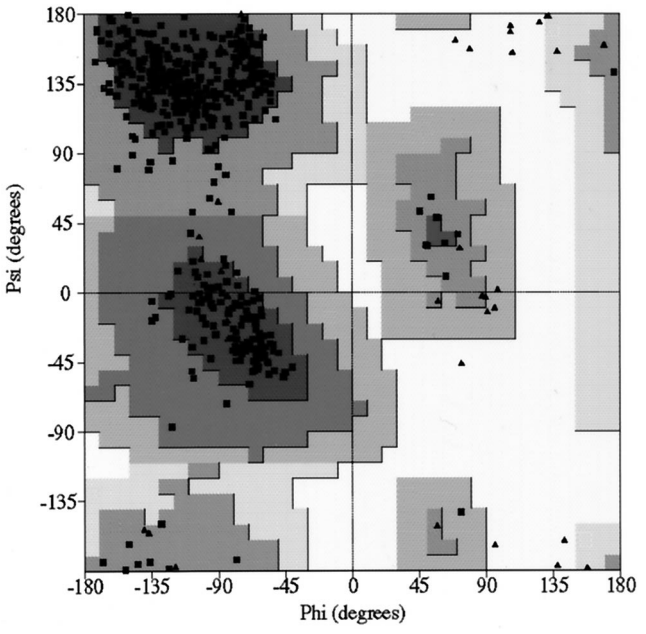


FIG. 3. The quality of the refined RCMV atomic model as shown in a Ramachandran plot of the main chain dihedral angles. Each glycine is represented as a black triangle; each nonglycine residue is represented as a black square. The areas from dark to light gray represent most-favored, allowed, generously allowed, and disallowed regions, respectively, for nonglycine residues. Ninety percent of the nonglycine residues were in the most-favored regions, and none were in the disallowed region. The atomic model was refined to a crystallographic R factor of 19.9% for 442,909 reflections between 10 and 2.4 Å and 19.3% for 412,704 reflections between 6 and 2.4 Å. There were 4,238 atoms from protein and 122 atoms from water molecules in the icosahedral asymmetric unit model. The root mean square deviations were as follows: for bond length, 0.014; for bond angle, 1.813; and for dihedral angle, 27.217. The crystallographic R factor is defined as follows:

$$\frac{\sum_{hkl}\left||F_{obs}| - k|F_{calc}|\right|}{\sum_{hkl}|F_{obs}|}$$

.

This figure was generated by the program PROCHECK (29).

relation coefficient of 96% for all the data between 30 and 2.4 Å (Table 1). An equatorial section of the averaged electron density of the virus particle is shown in Fig. 2B. All residues were assigned in the final model except 32 C-terminal residues (residues 183 to 214) of the S subunit and 8 C-terminal residues (residues 369 to 376) of the L subunit, which were disordered. In light of the CPMV structure, the undefined density of the C terminus of the S subunit can also be attributed to a partial degradation (27, 31; Lin et al., submitted).

Refinement and overall structure. The RCMV model was initially refined by simulated annealing (5–7), which decreased the R factor from 26.4 to 16.3% for data between resolutions of 10.0 and 5.0 Å. Further simulated annealing followed by conjugate gradient minimizations and manual modeling with data between resolutions of 8.0 and 3.0 Å decreased the R factor from 28.2 to 22.4%. A difference map was calculated and averaged once for the identification of water molecules. The resulting model was further refined by conjugate gradient minimization with data between resolutions of 10.0 and 2.4 Å to an R factor of 22.2%. Refinement of temperature factors was performed. The final structure includes 550 amino acid residues (4,238 atoms) and 122 water molecules in the icosahedral asymmetric unit with a crystallographic R factor of 19.9% for 442,909 reflections and with $I/\sigma(I) \geq 2$ in the resolution range of 10.0 to 2.4 Å. A Ramachandran plot of the main chain angles is shown in Fig. 3.

The overall structure of RCMV is similar to CPMV (48; Lin et al., submitted) and BPMV (10). Each icosahedral asymmetric unit is occupied by one S and one L subunit. By fivefold,

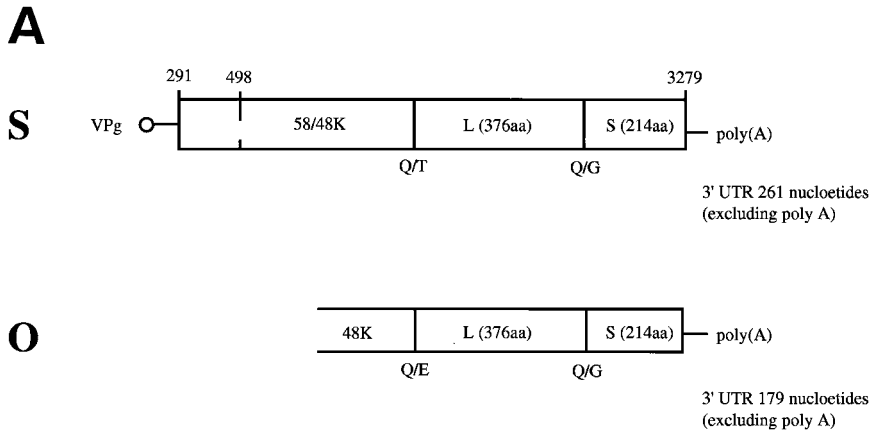
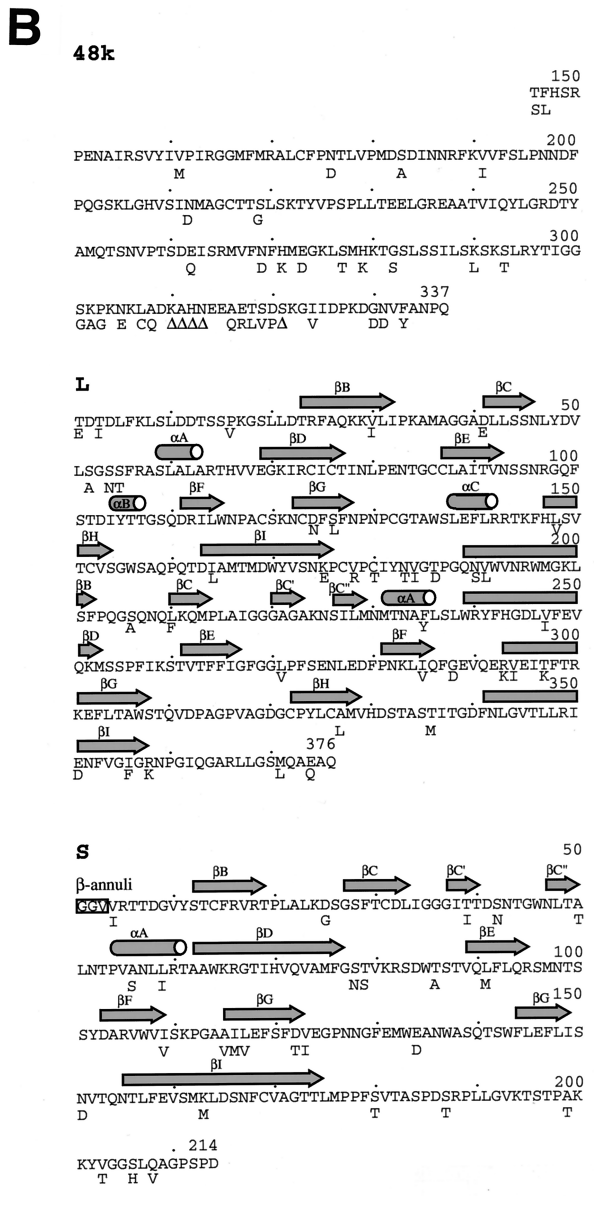


FIG. 4. (A) Organization of RCMV RNA 2. The top portion of the diagram represents the complete sequence of the RNA-2 of strain S (41), while the lower portion represents the region of strain O which has been sequenced to date. The genome-linked protein (VPg) and the 3′ polyadenylate [poly(A)] are indicated, as is the size of the 3′ untranslated region (3′ UTR) in each case. The open reading frame on each RNA 2 is shown as an open rectangle. For strain S the positions of the two initiation codons responsible for the synthesis of the 58K (position 291) and 48K (position 498) proteins are marked, as is the position of the termination codon at the end of the polyprotein (position 3279). The positions of the polyprotein processing sites are shown together with the dipeptide sequence which is cleaved (Q/T, Q/E, or Q/G). The identity of each final cleavage product (48K/58K; L and S) is also indicated. (B) Detailed comparisons of the amino acid sequences of the C-terminal portion of the 48K protein (48k), large coat protein (L), and small coat protein (S) of RCMV strains S and O. The upper sequence represents the sequence of strain S, with the differences found in strain O given underneath. Δ represents an amino acid which is deleted in the strain O protein. The secondary structure assignments are shown as arrows and cylinders.



threefold, and twofold symmetry operations, 60 copies of the icosahedral asymmetric units are generated that constitute the virus capsid with protrusions around the fivefold axes (Fig. 1). The outer radius at the fivefold axes is 154 Å; the outer radius across the twofold axes is 127 Å, and the outer radius across the threefold axes is 136 Å. The average thickness of the capsid is about 40 Å. The S subunit is a simple β-barrel, while the L subunit contains two β-barrels in a single polypeptide chain. Comparison of the RCMV capsid with the $T=3$ surface lattice shows that the S subunit occupies the A (pentamer) position, while the N-terminal and C-terminal β-barrels of the L subunit occupy the C and B (quasihexamer) positions, respectively.

Comparison with the coat proteins of RCMV strain O. The structural implications of amino acid substitutions between the coat proteins of closely related strains of tobacco mosaic virus (TMV) have proved to be invaluable in assigning functions to various residues (3). To carry out a similar analysis on a comovirus, the amino acid sequences of the two coat proteins of RCMV strain O were determined by analysis of the nucleotide sequences of cDNA clones of the 3′-terminal region of strain O RNA 2 (36). In total, the sequence of the 3′-terminal 2,565 nucleotides was determined. This sequence contained a single long open reading frame and the entire 3′ noncoding region of RNA 2 (Fig. 4A). The nucleotide sequence identity between strains S and O in the coding region was 75.9%, considerably higher than the 44% previously found in the 3′ noncoding regions (36). The amino acid sequences of the L and S subunits of RCMV strain O (Fig. 4B) were deduced by comparison of the amino acid sequence of the translation product of strain O RNA 2 with its strain S equivalent. The proposed cleavage sites for the release of the mature coat proteins were confirmed by Edman degradation with proteins isolated from a sodium do-

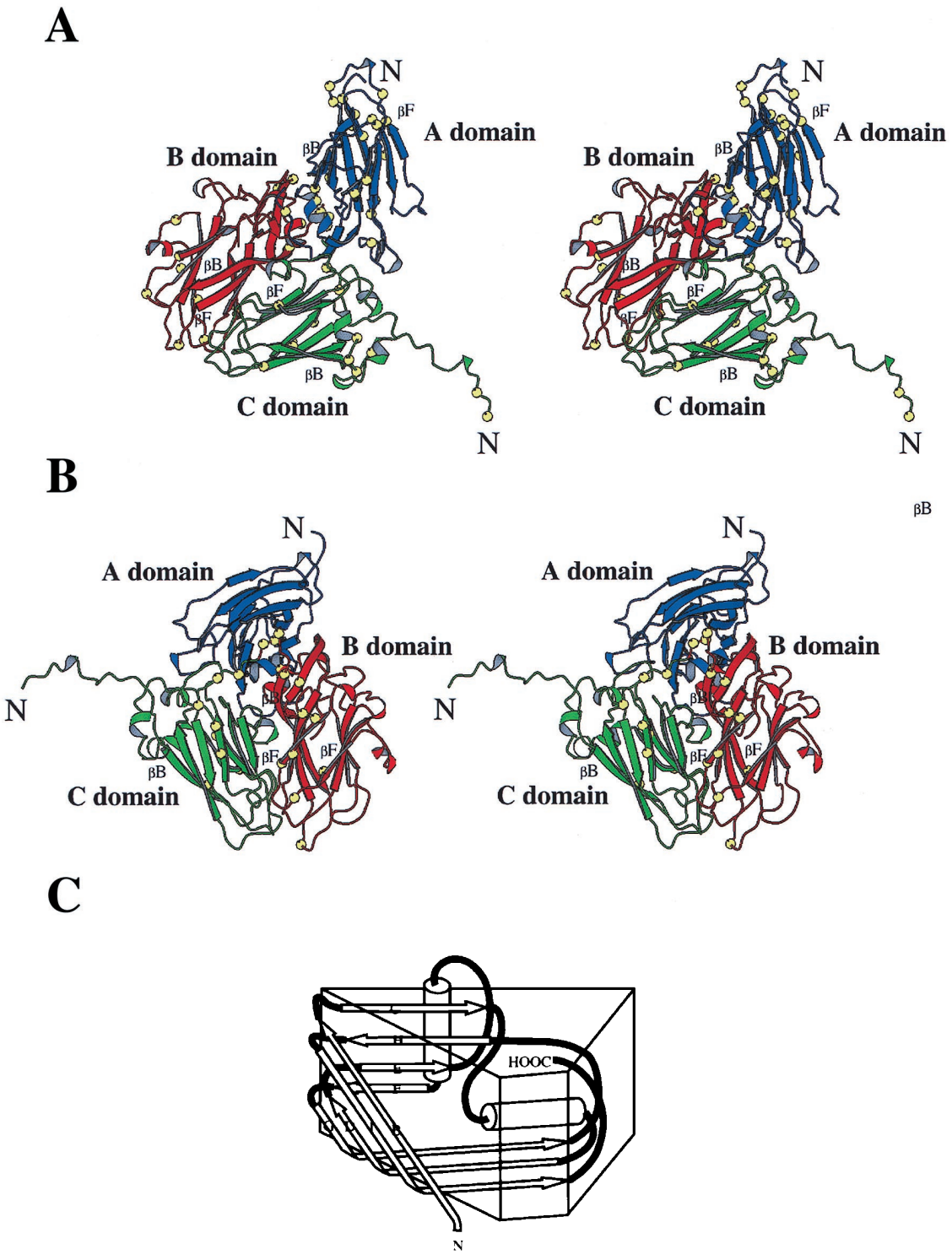


FIG. 5. Illustration of the difference between the strain S and O. The ribbon diagram is drawn through the Cα positions. The A domain is in blue, the B domain is in red, and the C domain is in green. (A) Stereoview of the icosahedral asymmetric unit of strain S, with all changes found in strain O labeled with yellow spheres. All of the differences can be accommodated in the context of the structure of strain S. (B) Stereoview of icosahedral asymmetric unit observed from the interior. Large numbers of changes (labeled as yellow spheres) are clustered around the cleft in which ordered ribonucleotides were observed in the crystal structure of BPMV. For clarity, only the changes around the cleft are shown. (C) Schematic presentation of the canonical β barrel for reference to the illustrations in panels A and B.

decyl sulfate-polyacrylamide gel. The sequence of the C-terminal 187 amino acids of the RCMV 48K protein, a protein involved in cell-to-cell movement of comoviruses, could also be deduced (Fig. 4B).

The L and S subunits of strain O consist of 376 and 214 amino acids (Fig. 4B), respectively, exactly the same sizes as their strain S counterparts. Overall, the coat proteins of strains S and O are 90.2% (L) and 87.9% (S) identical. Visual inspec-

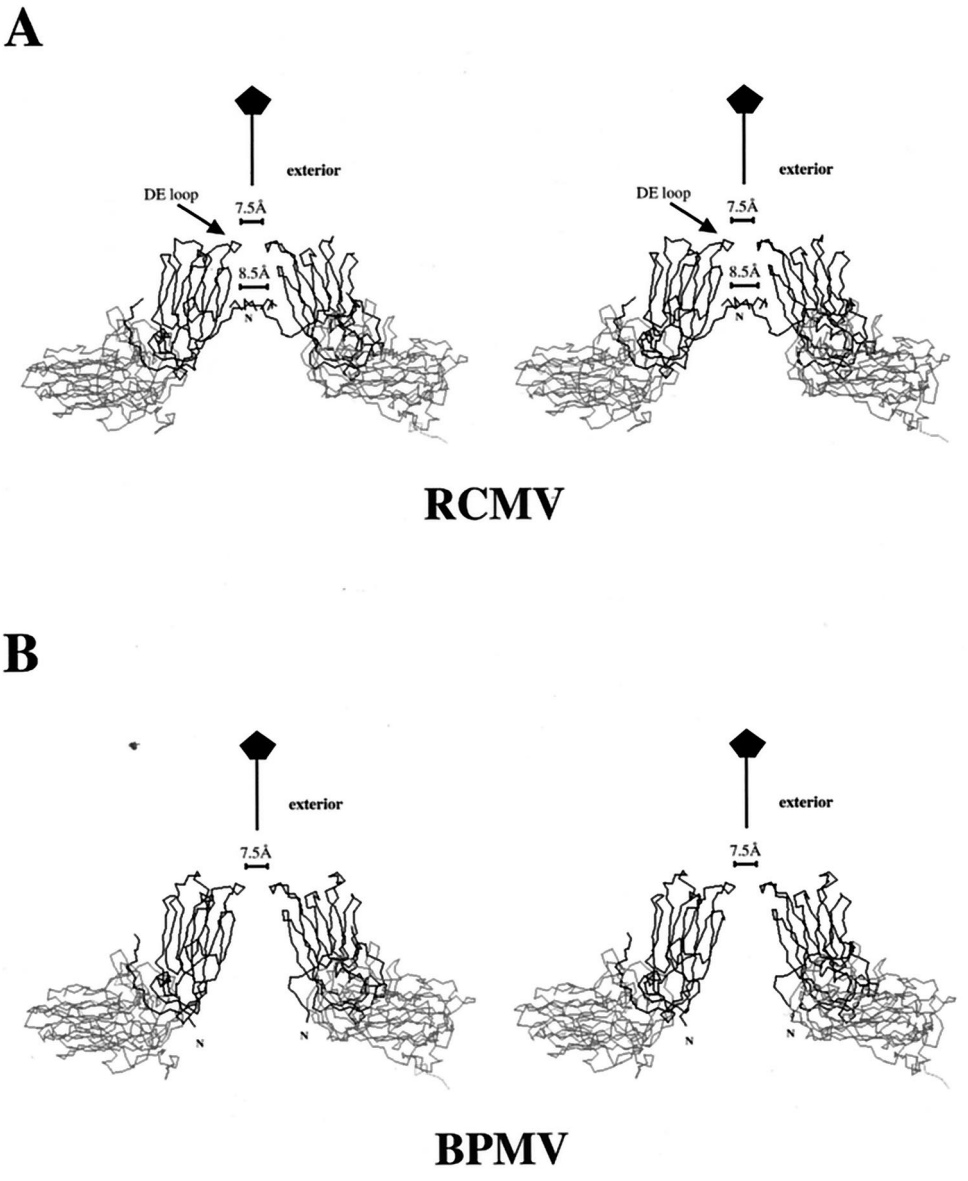


FIG. 6. Stereoviews of comovirus subunits in the formation of a channel along the fivefold symmetry axis. The S subunit is in dark gray tracing; the L subunit is in light gray tracing. (A) RCMV. Tracings of three N-terminal residues from all five S subunits, which form the pentameric annulus, are shown. For clarity, however, only tracings of two subunits related by 144° around the viral fivefold axis are drawn. The narrowest opening located on the surface has a diameter of ca. 7.5 Å and is comprised of the DE loops of the S subunits. The second constriction (~8.5 Å in diameter) is roughly in the middle of the channel and is formed by the pentameric annulus. A similar annulus was also identified in the CPMV structure. (B) BPMV. Two subunits related by 144° around the viral fivefold axis are shown. The N termini fold in the direction opposite that of RCMV and CPMV, and there is no annular structure.

tion of the alignment between the strain S and O coat proteins indicates that the changes are not evenly spread throughout the polypeptide chains but tend to be clustered (Fig. 4B). One particularly notable cluster occurs between amino acids 175 and 191 of the L protein, where 8 of 17 amino acids have changed, and not necessarily in a conservative manner.

To investigate the structural implications of the sequence variations in the coat proteins of strains S and O, the residues which were changed in strain O were mapped onto the three-dimensional (3D) structure of strain S. The changes in the strain O sequence can be comfortably accommodated in the atomic model of S strain with no obvious disruption of the structure. This finding is consistent with the observation that the two strains are serologically indistinguishable (P. Oxelfelt, unpublished data) and can form viable pseudorecombinants. Many of the changes are found on the exterior and interior surfaces of the viral capsid (Fig. 5). Several of the changes found among the buried residues are compensatory. For example, residues Ile285 and Val294 of the L subunit are packed against each other in strain S. The coordinated sequence changes to Val285 and Ile294 in strain O result in the occupation of the same volume and thus do not alter the capsid structure. As shown in Fig. 5B, many changes cluster around a cleft on which RNA was located in the BPMV structure (10). The most variable sequence (eight changes between amino acids 175 and 191 of the L subunit) comprises the top half of this cluster and is in the region connecting the C and B domains. Changes in the sequence in this nucleotide-binding re-

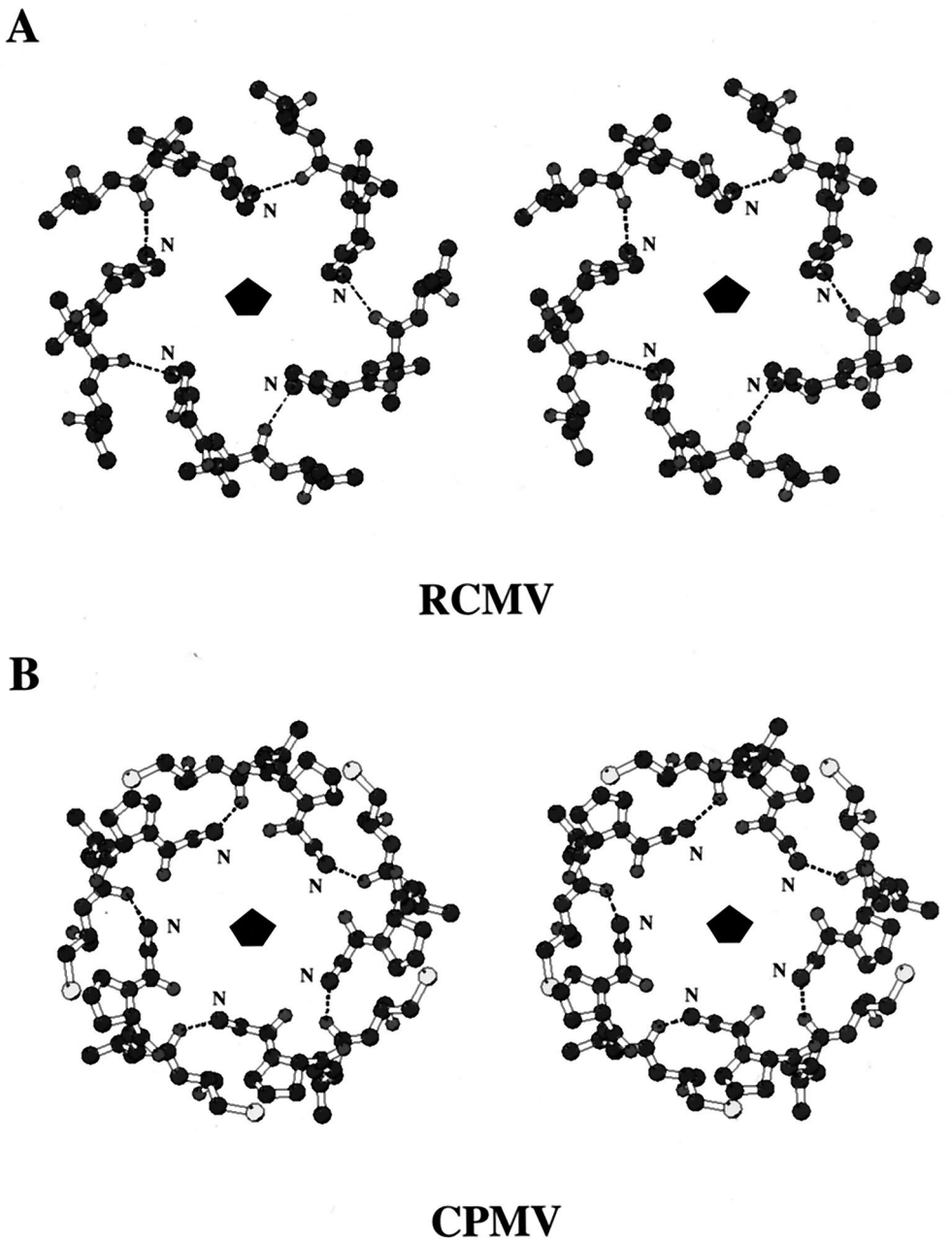


FIG. 7. Stereoviews of the pentameric annuli of comoviruses looking down the fivefold axis from the viral exterior. All of the atoms are drawn as black spheres, except for the sulfur atoms, which are gray spheres. Hydrogen bonds are shown as dashed lines. (A) Pentameric annulus of RCMV. It is formed by hydrogen bonding with the first three N-terminal residues from each of the pentameric S subunits. The amino group of each N terminus forms a hydrogen bond with the main chain carbonyl oxygen of the neighboring third residue. The annulus is star-shaped, with each peptide extending upward and then downward. The opening is about 8.5 Å. (B) Pentameric annulus of CPMV. It adopts a hydrogen bonding pattern similar to that found in RCMV. Each N terminus extends upward toward the exterior, and the annulus is crown-shaped.

gion may be related to the fact that strain O produces considerably less top component (empty, protein-only shells) than strain S (34).

The identity between the C-terminal portions of the 48K movement proteins of strains S and O is 60.4%, considerably less than that found in the coat proteins. Most of the changes are concentrated near the C terminus of the 48K protein and include a number of deletions (Fig. 4B). These changes occur in precisely the region that has been implicated in the interaction between the tubules containing, and probably formed by, the 48K protein and virus particles in the case of CPMV (30). These changes in the 48K protein may be needed to compensate for differences in the surface properties of the virus particles of strains S and O.

A structural fingerprint for subgrouping comoviruses. A noticeable surface feature of RCMV and other comoviruses is the protrusion centered around the viral fivefold axes (Fig. 1A). This protrusion is formed by the pentameric S subunits which are not tangential to the spherical capsid. Looking down the fivefold axis toward the viral interior, there is no part of the

FIG. 8. Sequence alignment of comovirus S subunits based on the 3D structures of RCMV, CPMV, and BPMV. The secondary structure assignments are shown at the top of the sequences. The boxed residues at the beginning of CPMV and RCMV sequence are those involved in the formation of the pentameric annuli. The boxed residues in the C termini of CPMV, RCMV, and BPMV are not visible in the X-ray structure. The uppercase letters represent identities. The overall similarity in the alignment suggests that all six viruses adopt similar capsid structures. The most notable differences between the CPMV and RCMV S subunits and those of the other comoviruses reside in the N and C termini (see text). The sequences used in the alignment are CPMV (52), BPMV (32), CPSMV (9), APMV (46), strain S of RCMV (41), and strain Z of SqMV (15).

S subunit that approaches the particle symmetry axis. A channel is apparent along the fivefold axis running from the exterior to the interior (Fig. 6). RCMV bears more similarity to CPMV than to BPMV in this part of the structure.

The channel has an overall funnel shape, with the narrow end at the outer surface and the wider end in the interior. The narrow end is comprised of the DE loops of the S subunits reaching over the innermost FG loop of the jelly-roll β-sandwich structure and clustering around the fivefold axis (Fig. 6A). The opening at this end is about 7.5 Å. Farther down the fivefold axis, a second constriction can be found in RCMV and CPMV, but not in BPMV (Fig. 6). This occurs as a result of the three N-terminal residues of the S subunits forming a pentameric annulus structure (Fig. 7). Similar annular structures were found in a number of icosahedral viruses (reference 17 and references therein). In CPMV and RCMV, the amino group of the N terminus forms a hydrogen bond with the main chain carbonyl oxygen of the neighboring third residue. The opening is ca. 8.5 Å. No constriction is found in BPMV since it lacks the pentameric annuli (Fig. 6B).

The original subgrouping of RCMV with CPMV and of BPMV with CPSMV was based on the degree of similarity between the proteins (48K, L, and S) encoded by their RNA 2 molecules (9). However, the extensions of this approach to other comoviruses were not particularly revealing, and no obvious discriminating criteria could be derived for subgrouping Andean potato mottle virus (APMV) (46) and several strains of squash mosaic viruses (SqMV) (15, 19). The results were especially ambiguous in the case of APMV, which shares a high degree of similarity with BPMV but a comparatively low degree of similarity with CPSMV (15). This discrepancy is inconsistent with the previous placement of BPMV and CPSMV in the same subgroup (9).

Structural alignment (i.e., protein sequence alignment based on superposition of 3D structures) of RCMV, CPMV, and BPMV indicates that one of the discriminating sequence differences between the two subgroups is at the N termini of the S subunits. The N-terminal residue of the RCMV and CPMV S subunits is glycine in both cases, instead of serine as found in BPMV. Moreover, the N-terminal sequences of both the CPMV and RCMV S proteins are three residues longer than that of the corresponding BPMV protein. These three residues form the pentameric annuli observed in the X-ray crystal structures. Glycine residues are preferred as the first residue to avoid space constraints in the formation of the annular structure. The shorter N termini of the S subunits of BPMV prevent the formation of the annulus, and they are folded in the opposite direction (Fig. 6B). Thus, the longer N termini, with a glycine as the terminal residue, can serve as a fingerprint for the subgrouping of comoviruses between CPMV-like and BPMV-like viruses.

The above fingerprint is used to verify the subgrouping of CPSMV with BPMV based on sequence homology criteria (9). Although protein sequence alignments suggest that the CPSMV S subunit would adopt an overall architecture similar to those of CPMV, RCMV, and BPMV (Fig. 8), the N terminus of the CPSMV S subunits resembles that of BPMV in both length and terminal residue. It is therefore reasonable to suggest that, as in the case of BPMV, a pentameric annular structure is not formed, supporting the previous subgrouping of CPSMV with BPMV. When the structural alignments are extended to include the sequences of the S proteins of APMV (46) and SqMV (15, 19), subgrouping with BPMV is predicted in each case (Fig. 8). The N terminus of the APMV S subunit is two residues shorter than the length required for the formation of the annular structure. Moreover, the N-terminal residue is phenylalanine, which would not be conducive to the formation of the annular structure. Similarly, the N terminus of the SqMV S subunit is one residue too short, and, like BPMV, the S subunit of SqMV has a serine residue at its N terminus. It can therefore be concluded that, based on structural criteria, there are two major subgroups of comoviruses. One group includes CPMV and RCMV; the other includes BPMV, CPSMV, APMV, and SqMV. This conclusion is consistent

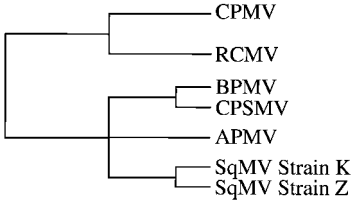


FIG. 9. Subgrouping of comoviruses. Comoviruses are placed in two major subgroups, CPMV-like viruses and BPMV-like viruses, based on a discriminating structural element, the pentameric annulus. A further classification of BPMV-like viruses is by the structural element, the pentameric annulus. Further classification of BPMV-like viruses is based on nucleotide sequence homology calculated by Haudenshield and Palukaitis (15). Similarity at the C termini, however, places APMV closer to BPMV and CPSMV in the BPMV-like subgroup.

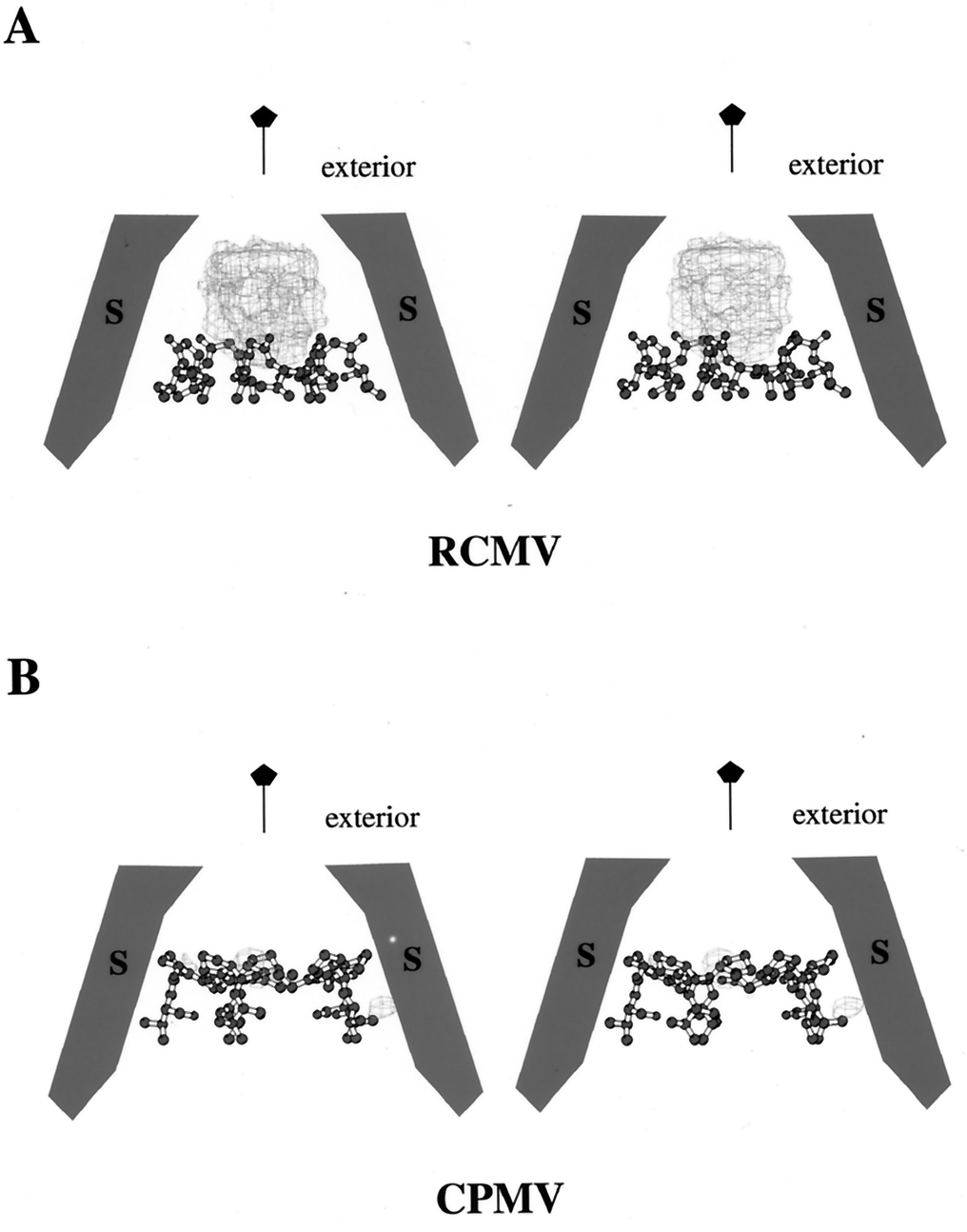


FIG. 10. Stereoviews of the nonprotein density and the pentameric annuli observed in the channel along the viral fivefold axis. The annuli are shown in "ball-and-stick," and the electron density is shown in "chickenwire." The difference electron density maps are calculated with Fourier coefficients of $F_o e^{i\alpha(ave)} - F_{model} e^{i\alpha(model)}$. Any density shown in these maps would be that which has not been included in the atomic model. (A) RCMV. A bulky density is observed atop the annulus. It appears that this bulky density cannot penetrate further into the capsid due to the blockage by the annular structure. The virus was prepared from infected plant leaves without ultracentrifugation in a CsCl gradient. (B) CPMV. The density is spherical, suggesting metal ion, and it is situated inside the opening of the annulus. There is no obvious barrier between the spherical density and the viral interior. The preparation of the virus used in the structure determination involved ultracentrifugation with CsCl gradients. Both electron density maps are contoured at 3.5σ. This figure was drawn by using the program BOBSCRIPT (13).

with the proposed subgrouping of SqMV based on the calculation of a parsimonious phylogenetic tree made by using RNA 2 nucleotide sequences (15). Moreover, structural alignment clearly places AMPV in the subgroup of BPMV-like viruses.

Another noticeable feature in the structural alignment of the S proteins of different comoviruses is that the length of sequence on the C-terminal side of the βI strand for BPMV, CPSMV, APMV, and SqMV (22, 22, 20, and 8 amino acids, respectively) is considerably less than it is for CPMV and RCMV (38 and 40 residues) (Fig. 8). Particularly striking is the fact that the C-terminal region of the SqMV S subunit is 32 residues shorter than its RCMV equivalent, having a total length equivalent to that of the ordered C-terminal polypeptide seen in the crystal structures of RCMV and BPMV. Though the full implications of these findings await further investigation, the C-terminal region of the S subunit of CPMV has been implicated in the packaging of the viral RNA (49). Taking into account all of the data from the structural and

nucleotide sequence alignments, a scheme for subgrouping comoviruses is suggested as shown in Fig. 9.

Comparison of the pentameric annuli of RCMV and CPMV. Despite sharing a similar pattern of hydrogen bonding, there is a noticeable difference between the annular structures of RCMV and CPMV (Fig. 7). The annulus of RCMV is star-shaped, with each N-terminal sequence extending first upward and then downward. By contrast, that of CPMV is crown-shaped, with each N-terminal sequence extending upward toward the exterior of the virion. Of greater significance, however, is the difference in the electron densities associated with the annuli in the two viruses, differences which seem to critically influence the annular structure. While spherical density, suggesting the presence of metal ions, is found in CPMV, a much bulkier density is found in RCMV (Fig. 10). This difference is most probably due to the fact that additional purification by CsCl gradient was employed in the preparation of CPMV for the structural studies (Lin et al., submitted). The conformation of the CPMV annular structure also allows the carbonyl oxygen of the N-terminal glycine to interact with the putative metal ion (Fig. 7B). The bulky density seen in RCMV probably represents some native material which can be replaced with metal ions, as seen in CPMV. The location of the putative metal ion in CPMV suggests that it can diffuse further into the interior of the capsid; it is situated inside the opening of the annulus, and there is no further barrier toward the interior (Fig. 10B). This is consistent with the findings that significant amounts of Cs ions diffuse into the CPMV B_L component (53) and that the annular structure was perturbed by the flux of large amounts of Cs ions into the CPMV capsid (Lin et al., submitted). By contrast, the bulky density found in RCMV is situated above the annulus and seems to be prevented from further diffusion into the capsid (Fig. 10A). It seems quite possible that the channel is a conduit, with the pentameric annulus playing a discriminatory role for the exchange of material between the viral interior and the environmental exterior. No density is found in the equivalent region of the BPMV capsid, which does not possess annuli (10) and is not Cs permeable.

Structural evolution. Structural studies have suggested that picorna-like viruses (nepoviruses, comoviruses, and picornaviruses) have evolved from $T=3$ viruses by gene triplication, independent evolution of each β-barrel sandwich domain, and the development of cleavage sites in the polyprotein (8, 10, 18). Consideration of the complexity of the capsid structures places nepoviruses in the lower, comoviruses in the middle, and picornaviruses in the higher part of the evolution tree. Structural alignments suggest that the N terminus of the BPMV S subunit is not much different in structure from the equivalent polypeptide in a nepovirus, tobacco ringspot virus (18). It points away from the fivefold axis and is structurally unsophisticated (Fig. 6B). By contrast, the N termini of the CPMV and RCMV S subunits point in the opposite direction, with the extra residues allowing the formation of the pentameric annular structure that seems to correlate with Cs permeability. It therefore seems reasonable to suggest that the BPMV-like comoviruses are closer to nepoviruses in evolution, with the separation of L and S subunits. On the other hand, CPMV-like viruses are closer to picornaviruses in evolution, with more sophisticated features developed at the N termini of the S subunits.

ACKNOWLEDGMENTS

We thank Pat Barker (Babraham Institute, Cambridge, United Kingdom) for N-terminal sequence analysis of the RCMV strain O coat proteins. T.L. thanks Anette Schneemann for reviewing the manuscript before publication and Lars Liljas for encouragement in the preparation of the manuscript.

A.J.C. was supported by a BBSRC (United Kingdom) research studentship, and G.P.L. acknowledges the support of an EMBO short-term fellowship for part of this work. The crystallographic studies were supported by grant GM54076 to J.E.J.

REFERENCES

1. **Abdelmoeti, M. A. H.** 1979. Red clover mottle virus (RCMV)—purification, stability, separation of components and genetic studies. Ph.D. thesis. Swedish University of Agricultural Sciences, Uppsala, Sweden.
2. **Achon, M. A., V. Medina, M. Shanks, P. G. Markham, and G. P. Lomonossoff.** 1994. Characterisation of a maize-infecting potyvirus from Spain. Eur. J. Plant Pathol. **100:**157–165.
3. **Altschuh, D., A. M. Lesk, A. C. Bloomer, and A. Klug.** 1987. Correlation of coordinated amino acid substitutions with function in virus related to tobacco mosaic virus. J. Mol. Biol. **193:**693–707.
4. **Bancroft, J. B.** 1962. Purification and properties of beanpod mottle virus and associated centrifugal and electrophoretic components. Virology **16:**419.
5. **Brunger, A. T.** 1992. X-PLOR version 3.1. A system for X-ray crystallography and NMR. Yale University Press, New Haven, Conn.
6. **Brunger, A. T., and L. M. Rice.** 1997. Crystallographic refinement by simulated annealing: methods and applications. Methods Enzymol. **277:**243–269.
7. **Brunger, A. T., A. Krukowski, and J. W. Erickson.** 1990. Slow-cooling protocols for crystallographic refinement by simulating annealing. Acta Crystallogr. **A46:**585–593.
8. **Chandrasekar, V., and J. E. Johnson.** 1997. The structure of tobacco ringspot virus: a link in the evolution of icosahedral capsids in the picornavirus superfamily. Structure **6:**157–171.
9. **Chen, X., and G. Bruening.** 1992. Nucleotide sequence and genetic map of cowpea severe mosaic virus RNA 2 and comparison with RNA 2 of other comoviruses. Virology **187:**682–692.
10. **Chen, Z., C. Stauffacher, Y. Li, T. Schmidt, W. Bomu, G. Kamer, M. Shanks, G. Lomonossoff, and J. E. Johnson.** 1989. Protein-RNA interactions in an icosahedral virus at 3.0 Å resolution. Science **245:**154–159.
11. **Clark, A. J.** 1997. Production and characterization of hybrid comoviruses. Ph.D. thesis. University of East Anglia, East Anglia, United Kingdom.
12. **Collaborative Computational Project, Number 4.** 1994. The CCP4 suites: programs for protein crystallography. Acta Crystallogr. **D50:**760–763.
13. **Esnouf, R. M.** 1997. An extensively modified version of Moscript that includes greatly enhanced coloring capacities. J. Mol. Graphics **15:**132–134.
14. **Ferrin, T. E., C. C. Huang, L. E. Jarvis, and R. Langridge.** 1988. The MIDAS display system. J. Mol. Graphics **9:**13–27.
15. **Haudenshield, J. S., and P. Palukaitis.** 1998. Diversity among isolates of squash mosaic virus. J. Gen. Virol. **79:**2331–2341.
16. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28:**351–359.
17. **Hogle, J. M., M. Chow, and D. J. Filman.** 1985. Three-dimensional structure of poliovirus at 2.9 Å resolution. Science **229:**1358–1365.
18. **Hosur, M. V., T. Schmidt, R. C. Tucker, J. E. Johnson, T. M. Gallagher, B. H. Selling, and R. R. Rueckert.** 1987. Structure of an insect virus at 3.0 Å resolution. Proteins **2:**167–176.
19. **Hu, J. S., S. Z. Pang, P. G. Nagpala, D. R. Siemieniak, J. L. Slightom, and D. Gonsalves.** 1993. The coat protein genes of squash mosaic virus: cloning, sequence analysis, and expression in tobacco protoplasts. Arch. Virol. **130:** 17–31.
20. **Huang, C. C., E. F. Petterson, T. E. Klein, T. E. Ferrin, and R. Langridge.** 1991. Conic: a fast renderer for space-filling molecules with shadows. J. Mol. Graphics **9:**230–236.
21. **Johnson, J. E.** 1978. Averaging of electron density maps. Acta Crystallogr. **B34:**576–577.
22. **Jones, T. A., J. Y. Zou, S. W. Cowan, and M. Kjeldgaard.** 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. **A47:**110–119.
23. **Jones, T. A., and M. Kjeldgaard.** 1997. Electron density map interpretation. Methods Enzymol. **277:**173–208.
24. **Kabsch, W., and S. Sander.** 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers **22:**2577–2637.
25. **Kim, S.** 1989. Auto-indexing oscillation photographs. J. Appl. Crystallogr. **22:**53–60.
26. **Kraulis, P. J.** 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. **24:**946–950.
27. **Kridl, J. C., and G. Bruening.** 1983. Comparison of capsids and nucleocapsids from cowpea mosaic virus-infected cowpea protoplasts and seedlings. Virology **129:**369–380.
28. **Lapchic, L. G., A. J. Clark, G. P. Lomonossoff, and M. Shanks.** 1998. Red clover mottle virus from Ukraine is an isolate of RCMV strain S. Eur. J. Plant Pathol. **104:**409–412.
29. **Laskowski, R. W., M. W. MacArthur, D. S. Moss, and J. M. Thornton.** 1993. PROCHECK: a program to check the stereochemical quality of protein

structures. J. Appl. Crystallogr. **26:**345–364.
30. **Lekkerkerker, A., J. Wellink, P. Juan, J. van Lent, R. Goldbach, and A. B. van Kammen.** 1996. Distinct functional domains in the cowpea mosaic virus movement protein. J. Virol. **70:**5658–5661.
31. **Lomonossoff, G. P., and J. E. Johnson.** 1991. The synthesis and structure of comovirus capsids. Prog. Biophys. Mol. Biol. **55:**107–137.
32. **MacFarlane, S. A., M. Shanks, J. W. Davies, A. Zlotnick, and G. P. Lomonossoff.** 1991. Analysis of the nucleotide sequence of bean pod mottle virus middle component RNA. Virology **183:**405–409.
33. **McPherson, A.** 1982. The preparation and analysis of protein crystals. John Wiley and Sons, New York, N.Y.
34. **Oxelfelt, P.** 1976. Biological and physico-chemical characteristics of three strains of red clover mottle virus. Virology **74:**73–80.
35. **Oxelfelt, P., and M. Abdelmoeti.** 1978. Genetic complementation between natural strains of red clover mottle virus. Intervirology **10:**78–86.
36. **Oxelfelt, P., M. Shanks, A. K. Widmark, and G. P. Lomonossoff.** 1992. Identification and characterization of pseudo-recombinants of red clover mottle comovirus. J. Gen. Virol. **73:**2121–2124.
37. **Rossmann, M. G.** 1979. Processing oscillation diffraction data for very large unit cells with an automatic convolution technique and profile fitting. J. Appl. Crystallogr. **12:**225–238.
38. **Rossmann, M. G., and J. E. Johnson.** 1989. Icosahedral RNA virus structure. Annu. Rev. Biochem. **58:**533–573.
39. **Rossmann, M. G., A. G. W. Leslie, S. S. Abdel-Meguid, and T. Tsukihara.** 1979. Processing and post-refinement of oscillation camera data. J. Appl. Crystallogr. **12:**570–581.
40. **Rossmann, M. G., R. McKenna, L. Tong, D. Xia, J.-B. Dai, H. Wu, and H. K. Choi.** 1992. Molecular replacement real-space averaging. J. Appl. Crystallogr. **25:**166–180.
41. **Shanks, M., J. Stanley, and G. P. Lomonossoff.** 1986. The primary structure of red clover mottle virus middle component RNA. Virology **155:**687–706.
42. **Shanks, M., K. Tomenius, D. Clapham, N. S. Huskisson, P. J. Barker, I. G. Wilson, A. J. Maule, and G. P. Lomonossoff.** 1989. Identification and subcellular localization of a putative cell-to-cell transport protein from red clover mottle virus. Virology **173:**400–407.
43. **Shanks, M., and G. P. Lomonossoff.** 1990. The primary structure of the 24K protease from red clover mottle virus: implications for the mode of action of comoviral proteases. J. Gen. Virol. **71:**735–738.
44. **Shanks, M., and G. P. Lomonossoff.** 1992. The nucleotide sequence of red clover mottle virus bottom component RNA. J. Gen. Virol. **73:**2473–2477.
45. **Shanks, M., J. T. Dessens, and G. P. Lomonossoff.** 1996. The 24-kDa proteinases of comoviruses are virus-specific in *cis* as well as in *trans*. J. Gen. Virol. **77:**2365–2369.
46. **Shindo, N., A. C. P. Vicente, R. Krengiel, and D. E. Oliveira.** 1993. Nucleotide sequence analysis of an Andean potato mottle virus middle component RNA cDNA clone: comparison of the encoded proteins with those of other comoviruses. Intervirology **36:**169–180.
47. **Sinha, R. C.** 1960. Red clover mottle virus. Ann. Appl. Biol. **48:**742–748.
48. **Stauffacher, C. V., R. Usha, M. Harrington, T. Schmidt, M. Hosur, and J. E. Johnson.** 1987. The structure of cowpea mosaic virus at 3.5 Å resolution, p. 293–308. *In* D. Moras, J. Drenth, B. Strandberg, D. Suck, and K. Wilson (ed.), Crystallography in molecular biology. Plenum Publishing Corporation, New York, N.Y.
49. **Taylor, K. M., V. E. Spall, P. J. G. Butler, and G. P. Lomonossoff.** 1999. The cleavable C-terminus of the small coat protein of cowpea mosaic virus is involved in RNA encapsidation. Virology **255:**129–137.
50. **Tong, L., and M. G. Rossmann.** 1990. The locked rotation function. Acta Crystallogr. **A46:**783–792.
51. **Tong, L., and M. G. Rossmann.** 1997. Rotation function calculation with GLRF program. Methods Enzymol. **276:**594–611.
52. **van Wezenbeek, P., J. Verver, J. Harmsen, P. Vos, and A. van Kammen.** 1983. Primary structure and gene organization of the middle-component RNA of cowpea mosaic virus. EMBO J. **2:**941–946.
53. **Virudachalam, R., M. Harrington, J. E. Johnson, and J. L. Markley.** 1985. ^{1}H, ^{13}C and ^{31}P nuclear magnetic resonance studies of cowpea mosaic virus: detection and exchange of polyamines and dynamics of the RNA. Virology **141:**43–50.
54. **Wilson, A. J. C.** 1942. Determination of absolute from relative X-ray intensity data. Nature **150:**151.
55. **Zimmern, D.** 1975. The 5′ end group of tobacco mosaic virus RNA is m7G5′ppp5′Gp. Nucleic Acids Res. **2:**1189–1201.